

## Cluster analysis of a world collection of red clover germplasm

C. N. Kouamé<sup>1</sup> & K. H. Quesenberry<sup>2</sup>

<sup>1</sup>National Agricultural Research Institute Savannah Zones (IDESSA) BP 633 Bouaké, Côte d'Ivoire;

<sup>2</sup>University of Florida Agronomy Department, Gainesville, FL 32611 USA

Received 20 August 1992; accepted 28 February 1993

*Key words:* *Trifolium pratense* L., red clover, cluster analysis, core collection, germplasm

### Summary

Germplasm evaluation data on more than 800 accessions of red clover (*Trifolium pratense* L.) representing 41 countries of origin were examined from the Germplasm Resource and Information Network of the National Plant Germplasm System. Large variations exist for most characters over all origins. Clustering accessions using standardized values of 15 morphological and physiological descriptors produced three distinct groups that correspond to early, medium, and late maturity groups. Within each maturity group, different similarity groups were produced. The analysis revealed a large range of diversity among the red clover accessions over all origins with the most diversity occurring in accessions from Eastern and Northern Europe. Accessions with similar attributes often were from countries with contrasting agroecological conditions. Categorizing red clover accessions into similarity groups provided a structure for sampling a limited number of entries to represent the diversity of the germplasm into a core collection. The utilization of this core collection should facilitate genetic resource management and identification of genotypes for breeding and other agronomic uses.

### Introduction

Plant germplasm collections have multiplied around the world to preserve genetic diversity of crop species and ensure a wide genetic base in plant improvement programs (Astley, 1987; Brush, 1989). The large size and heterogenous structure of many collections has often hindered the efficient maintenance and utilization of gene bank materials. The implementation of long- and short-term management strategies by gene bank curators with limited resources is a concern which must be addressed to improve germplasm utilization (Holden, 1984; Brown, 1989). This has called for the establishment of a core collection (Frankel,

1984), a severely limited sample which represent the genetic diversity of a crop species and its relatives. Categorizing germplasm accessions into morphologically similar, and presumably genetically similar, groups is the initial step of the core collection concept of genetic resource management. Relationships among plant genotypes of various plant species have been well established using cluster analysis of several quantitative and/or qualitative characters (Souza & Sorrells, 1991; Kouamé & Quesenberry, 1992). The similarity groups derived from such analyses may then be sampled either randomly or by some stratification such as geographical origin to form the core germplasm collection. The objectives of this

analysis were to examine the red clover germplasm collection in the US National Plant Germplasm System and to identify accessions with similar morphological and physiological characters. It is intended that this classification will provide a basis for the development of a core collection of red clover.

### Materials and methods

The data base for this analysis was the passport and descriptor evaluation data of approximately 800 accessions of red clover maintained by the US Plant Genetic Resource Unit at Geneva, NY. The passport data available indicated that accessions originated from 41 countries (possible errors in country of origin due to germplasm exchange could not be ruled out). Germplasm data were accessed using the Germplasm Resource and Information Network (GRIN), a computer database of the US National Plant Germplasm System (NPGS). For each accession, records included geographical origin, year of acquisition, and mean values for various numbers of 23 plant descriptors. Not all descriptors had been evaluated on all accessions, and only those accessions with complete data were used. Frequency distribution, mean, and variance were calculated for each character over all origins. Simple correlation coefficients between all characters were calculated for the whole collection.

Zohary & Heller (1984) classify red clover into six botanical varieties, however, this classification is not generally used by plant breeders. Botanical variety was not a descriptor in the GRIN data base and was not included in these analyses. Many recently released European cultivars are tetraploid, however virtually all wild introductions of red clover are diploid. Although ploidy level information may be available in some of the narrative information received as cultivars enter the plant introduction system, it was not one of the GRIN descriptors and thus was not used in these analyses.

Data in the GRIN system had been collected over many growing seasons since about 1965, but only a limited number of accessions were evaluated

in any one season. It is recognized that for some of the traits genotype by environment interactions may be important. Nevertheless, the purpose of this effort was to attempt to use the available data (recognizing that some errors may exist) to stratify the large number of accessions for sampling to produce a preliminary core collection. The intention then would be to produce seed, under isolation conditions, from original source seed, and use this seed for a more complete evaluation of all accessions in the core collection across various environments.

A cluster analysis was conducted using 15 selected plant descriptors (Table 1). A total of 463 accessions which had complete passport data for all 15 descriptors were used in the cluster procedure. In general, procedures outlined in the SAS Users Guide: Statistics, Statistical Analysis System (SAS Institute, 1985) were followed in the data analysis. Because different scales were used for the descriptors, data were first standardized to a mean of 0 and a variance of 1 using PROC STANDARD. A hierarchical cluster procedure was run on the standardized value of the original variables according to the average linkage method as executed by PROC CLUSTER, option AVERAGE. This agglomerative method begins with the computation of a matrix of euclidean distances among groups means and produces a dendrogram showing successive fusions of individuals, which culminates at the stage where all individuals belong to the same cluster. The relationship among the clusters formed was then assessed through a measure of the distance between them using the Mahalanobis  $D^2$  statistic of the canonical discriminant procedure: PROC CANDISC, option MAHALANOBIS. The distance  $D^2 = (Z_i - Z_j)'R^{-1}(Z_i - Z_j)$  is the square difference between the *i*th and *j*th cluster' vectors ( $Z_i, Z_j$ ) of the character. Each character distance is weighted by the inverse of the variance-covariance matrix (*R*) of the character among all accessions in the *i*th and *j*th cluster. Weighing by the inverse of the covariance matrix reduces the emphasis on highly correlated traits and increases the emphasis on independent traits. A multivariate analysis of variance was used to test the distances measured and the *F*-statistic (calculated using Wilk's Lambda criteria) associated with cluster comparison was used to measure the probability of cluster identity.

Table 1. Selected morphological and agronomic characters recorded on red clover accessions in the US National Plant Germplasm System

1-Growth habit (GHB)	1 = prostrate, 5 = medium, 9 = very erect.
2-Number of heads (NHD)	1 = few (1-3), 2 = medium (4-6), 3 = many (> 7).
3-Number of stems (NST)	1 = few, 2 = medium, 3 = most.
4-Fall recovery (FRC)	1 = little, 5 = medium, 9 = much growth.
5-Spring recovery (SRC)	1 = little, 5 = medium, 9 = much growth.
6-Winter hardiness (WHD)	1 = little, 5 = medium, 9 = complete survival.
7-Leafiness (LFS)	1 = few, 2 = medium, 3 = many.
8-Maturity (MTY)	1 = very late, 5 = medium, 9 = very early.
9-Vigor (VGR)	1 = poor, 5 = medium, 9 = most.
10-Uniformity (UNI)	1 = poor, 5 = medium, 9 = very.
11-Plant height (PLH)	cm
12-Plant width (PLW)	cm
13-Plant type (PLT)	% of medium and mammoth plants in the sample.
14-Reaction to mildew (MLD)	% of plants showing symptoms in the sample.
15-Reaction to virus (VRS)	% of plants showing symptoms in the sample.

## Results and discussion

The frequency distributions observed for selected plant descriptors in the whole collection are shown in Figs. 1 and 2. A wide range of variation for all traits were found in the collection of red clover germplasm. An analysis of variance indicated significant differences among accessions for all 15 selected characters. Many pairs of characters were correlated in these genotypes. Maturity data was significantly correlated to all but four characters (data not shown). Uniformity was correlated to the fewest number of characters. The high coefficient of correlation between spring recovery and winter hardiness ( $r^2 = 0.88$ ) suggests that evaluation of only one of these traits would be sufficient for germplasm records.

The statistics generated by the cluster procedure indicated significant division of the red clover germplasm into 3, 10, or 18 classes. The cubic clustering criterion had a local peak at 3-cluster but higher peaks at 10- and 18-clusters. The pseudo F statistic indicated 3, 10, and 18 clusters but the 10- and 18-cluster had lower and less sharper peaks than the 3-cluster peak. The  $t^2$  statistic had a low value at 3, 6, 10, 15, 18-clusters but had a particularly low value at 10- and 18-clusters. All three statistics suggested 3, 10, or 18-cluster partitions. We limited our discussion to the 3- and 10-cluster partitions which are most useful for the purpose of the study.

For the 3-cluster partition, three groups were formed. These were designated as A, B, and C and

appear to correspond somewhat to Early, Medium, and Late Maturity date, respectively (Fig. 3). Each group had distinctly different maturity date, heads number, and plant type. Group A contained the largest number of accessions (364) and was comprised of medium-type red clover (plants that flower in the first year after seeding), with 4–6 heads and relatively early maturity date (4.9–7.6). Group B is composed of 97 accessions of which all but 12 accessions were mammoth type red clover. This group is characterized by a medium to late maturity date (1.5–4.4) and a low head number (1–3) per accession. Group C contained only 2 mammoth red clover accessions with low head number and very late maturity date (1.0).

When considering the 10-cluster partition, groups A and B were subdivided into 5 and 4 subgroups, respectively. The 2 accessions of group C remained together and constituted the 10th cluster (cluster X). Mean and standard deviation of the cluster means are presented in Tables 2 and 3. Based on the  $D^2$  distance of the standardized variables, all clusters were significantly different from each other (Table 4). Clusters IV and II were the most similar ( $D^2 = 13.3$  units) whereas clusters X and VII were the most distantly related ( $D^2 = 108.1$  units). The 5 clusters of group A are I, II, IV, VII, and VIII; containing 155, 124, 71, 8, and 6 accessions, respectively. Cluster I comprised accessions with the highest head number ( $2.7 \pm 0.5$ ) and highest vigor ( $7.4 \pm 1.0$ ). Clusters II and IV were very closely related ( $D^2 = 13.3$  units) differing (cluster II vs. cluster IV) in spring

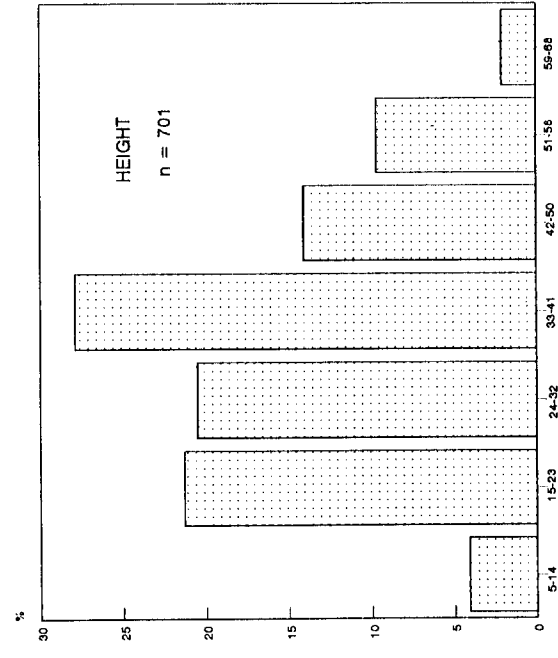
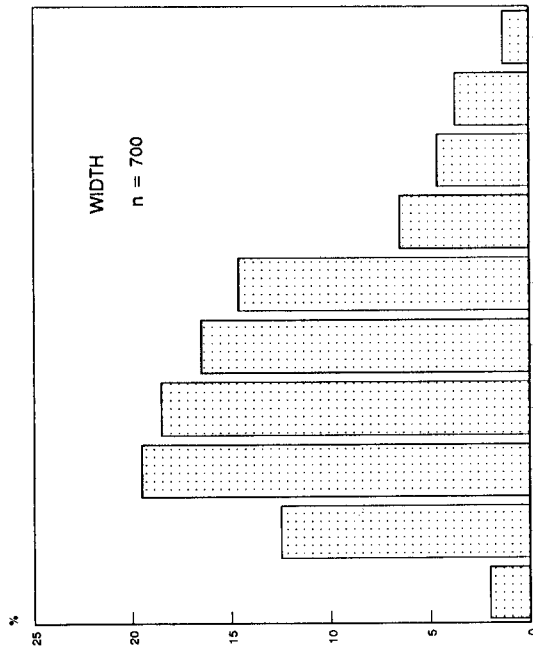
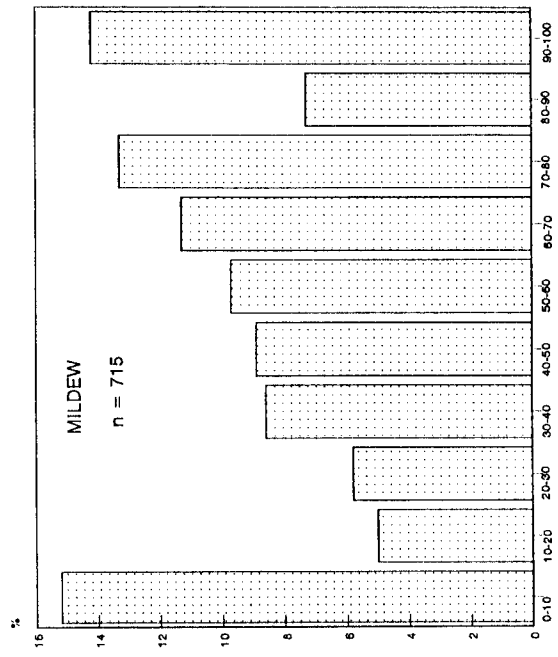
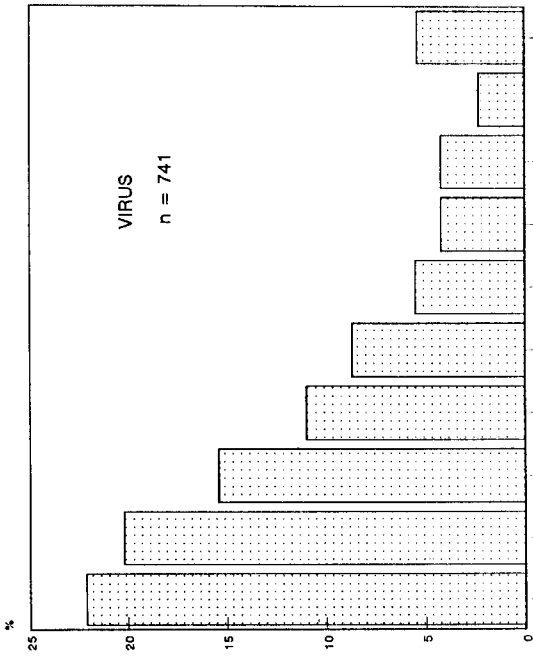


Fig. 1. Frequency distribution of red clover plant width, height, virus reaction, and mildew reaction in the US National Plant Germplasm System. Classes were established according to records legend of Table 1.

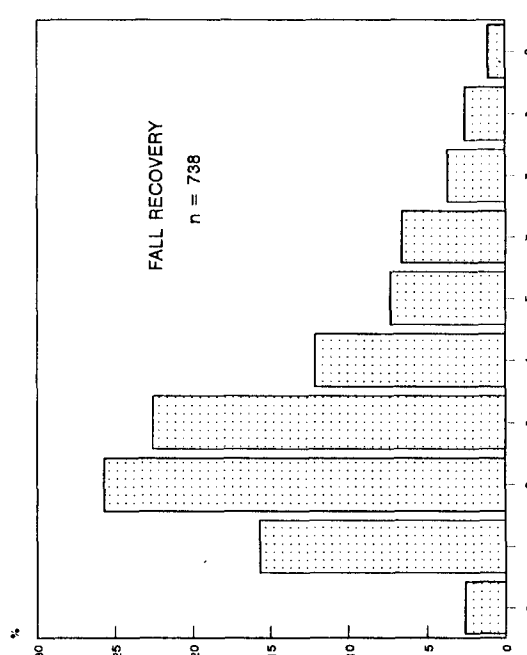
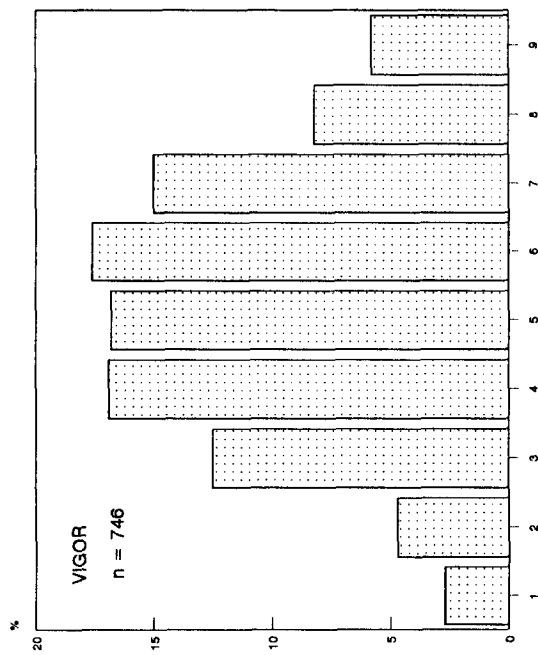
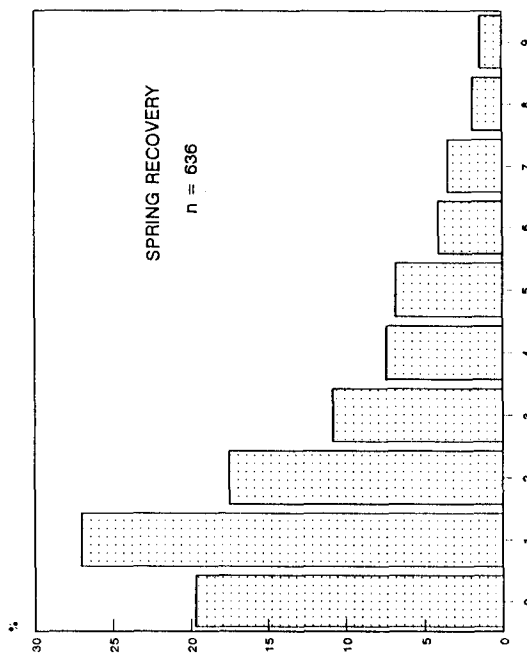
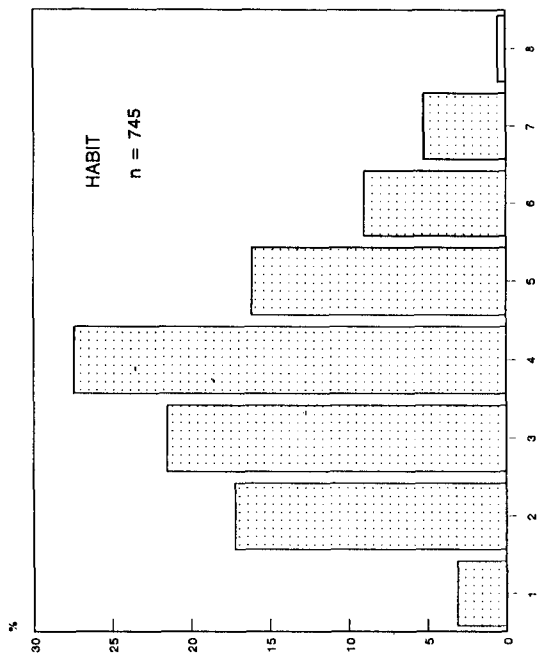


Fig. 2. Frequency distribution of red clover plant vigor, growth habit, Spring recovery, and Fall recovery in the US National Plant Germplasm System. Classes were established according to records legend of Table I.

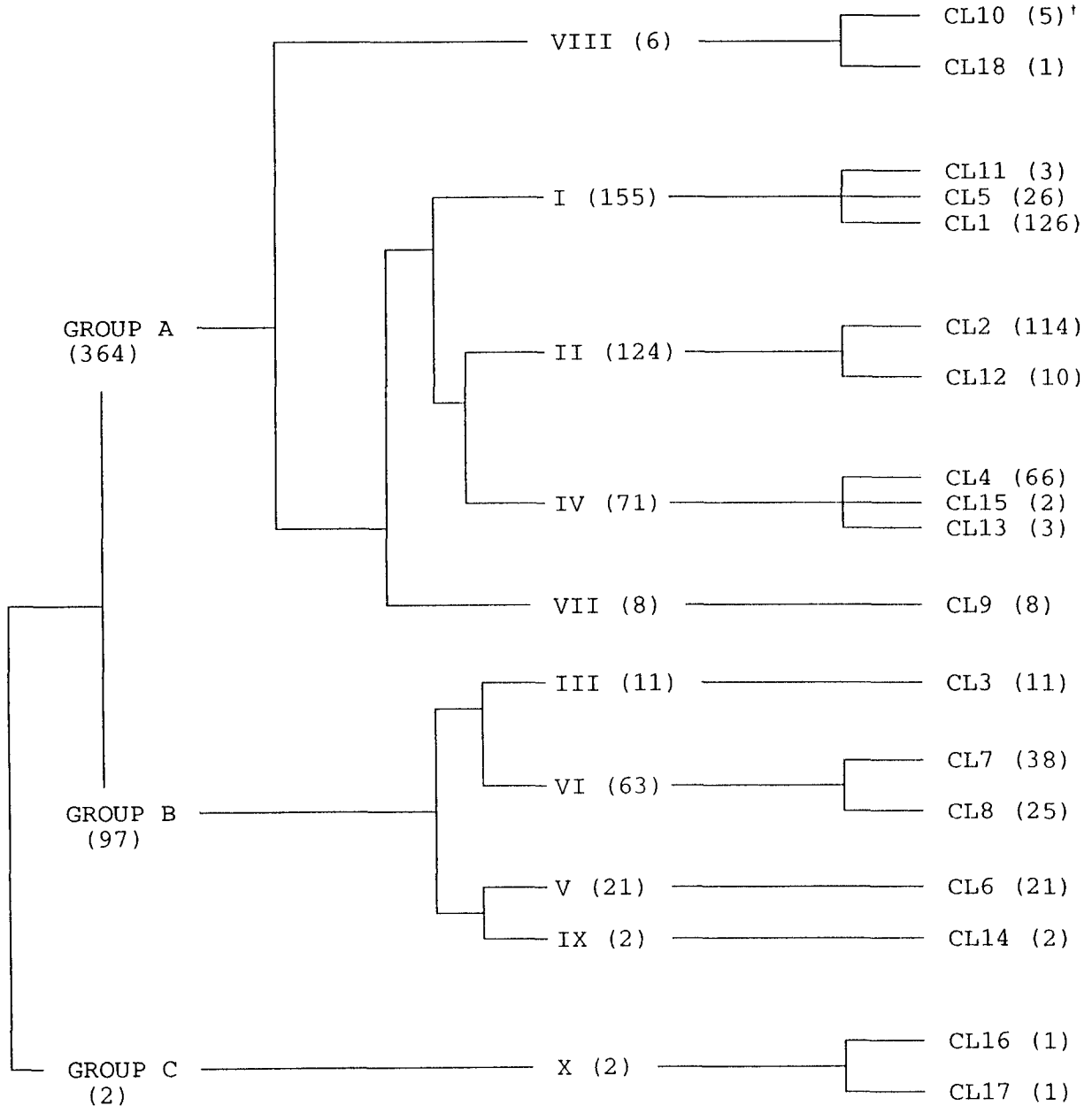


Fig. 3. Cluster analysis of 15 red clover plant descriptors in the US National Plant Germplasm System. †Numbers in parentheses are the total number of accessions in a cluster.

recovery rate ( $0.9 \pm 0.9$  vs.  $5.0 \pm 1.9$ ), winterhardiness ( $0.9 \pm 0.9$  vs.  $6.0 \pm 1.9$ ), and virus reactions ( $41.1 \pm 25.2$  vs.  $21.1 \pm 17.2$ ). Cluster VII comprised accessions with the lowest fall recovery rate ( $1.4 \pm 0.9$ ), tallest height ( $57.4 \pm 10.8$  cm), and greatest reaction to mildew ( $91.2 \pm 9.9$ ) and virus ( $93.7 \pm 10.2$ ). On the other hand, cluster VIII contained the shortest ( $13.1 \pm 5.9$  cm) and most

prostrate ( $1.5 \pm 0.5$ ) genotypes with the least reaction to mildew ( $16.6 \pm 26.0$ ) and virus ( $5.8 \pm 10.2$ ).

The four clusters of group B, designated as III, V, VI, and IX, contained 11, 21, 63, and 2 accessions, respectively. Cluster III comprised medium-type red clover which differed from the other medium-type clusters (group A) by their single or few heads ( $1.4 \pm 0.7$ ) and late maturity date

Table 2. Means and standard deviation of red clover plant characteristics in each cluster of early maturity group

Characters	Clusters				
	I	II	IV	VII	VIII
	Mean $\pm$ SD				
GHB†	3.9 $\pm$ 1.1	4.5 $\pm$ 1.7	4.1 $\pm$ 1.3	4.9 $\pm$ 1.2	1.5 $\pm$ 0.5
NHD	2.7 $\pm$ 0.5	1.9 $\pm$ 0.7	1.7 $\pm$ 0.6	1.7 $\pm$ 0.4	1.8 $\pm$ 0.4
NST	2.4 $\pm$ 0.5	1.8 $\pm$ 0.5	1.9 $\pm$ 0.4	2.4 $\pm$ 0.7	2.6 $\pm$ 0.5
FRC	3.2 $\pm$ 1.9	2.9 $\pm$ 1.6	3.5 $\pm$ 1.5	1.4 $\pm$ 0.9	3.8 $\pm$ 2.2
SRC	1.9 $\pm$ 1.4	0.9 $\pm$ 0.9	5.0 $\pm$ 1.9	1.0 $\pm$ 0.5	3.6 $\pm$ 1.6
WHD	2.3 $\pm$ 1.7	0.9 $\pm$ 0.9	6.0 $\pm$ 1.9	2.0 $\pm$ 1.5	4.3 $\pm$ 1.7
MTY	6.7 $\pm$ 1.4	6.5 $\pm$ 1.4	4.9 $\pm$ 1.8	7.1 $\pm$ 2.2	7.6 $\pm$ 0.5
VGR	7.4 $\pm$ 1.0	5.0 $\pm$ 1.2	4.9 $\pm$ 1.1	5.6 $\pm$ 0.9	6.6 $\pm$ 0.8
PLH	43.3 $\pm$ 9.8	30.6 $\pm$ 9.3	36.2 $\pm$ 7.9	57.4 $\pm$ 10.8	13.1 $\pm$ 5.9
PLW	95.8 $\pm$ 24.5	86.7 $\pm$ 23.9	69.2 $\pm$ 14.9	101.2 $\pm$ 20.5	60.0 $\pm$ 25.2
PLT	62.6 $\pm$ 28.4	86.7 $\pm$ 23.9	93.3 $\pm$ 12.2	97.5 $\pm$ 3.8	78.3 $\pm$ 40.2
MLD	62.6 $\pm$ 28.4	60.1 $\pm$ 24.4	63.9 $\pm$ 21.3	91.2 $\pm$ 9.9	16.6 $\pm$ 26.0
VRS	21.4 $\pm$ 13.2	41.1 $\pm$ 25.2	21.1 $\pm$ 17.2	93.7 $\pm$ 10.2	5.8 $\pm$ 10.2

†Plant descriptors abbreviations: GHB = Growth habit, NHD = Number of heads, NST = Number of stems, FRC = Fall recovery, SRC = Spring recovery, WHD = Winter hardiness, MTY = Maturity, VGR = Vigor, PLH = Plant height, PLW = Plant width, PLT = Plant type, MLD = reaction to mildew, VRS = Reaction to virus.

(4.4  $\pm$  1.6), common features of mammoth red clover. Although cluster III was classified in group B, it was most closely related to clusters II ( $D^2 = 26.5$  units) and IV ( $D^2 = 29.6$  units) of group A than to any cluster of group B. Cluster V is characterized by vigorous accessions (7.6  $\pm$  0.7), high stem number (2.8  $\pm$  0.3), and low spring recovery rate (0.8  $\pm$  0.9) while cluster VI accessions

had the smallest width (44.1  $\pm$  11.9 cm). Cluster IX comprised 1 medium and 1 mammoth red clover. Both accessions had the most erect growth habit (6.0  $\pm$  0) and the least reaction to mildew (5.0  $\pm$  7.1) and virus (5.0  $\pm$  7.1). Group C is a single-cluster group in the 10-cluster partition, comprising 2 accessions of mammoth red clover. These accessions had distinctly high fall recovery

Table 3. Means and standard deviation of red clover plant characteristics in each cluster of medium and late maturity groups

Characters	Clusters				
	III	V	VI	IX	X
	Mean $\pm$ SD				
GHB†	2.3 $\pm$ 1.1	2.5 $\pm$ 0.9	3.4 $\pm$ 1.3	6.0 $\pm$ 0.0	3.0 $\pm$ 2.8
NHD	1.4 $\pm$ 0.7	1.3 $\pm$ 0.6	1.0 $\pm$ 0.4	1.0 $\pm$ 0.0	1.0 $\pm$ 0.0
NST	1.9 $\pm$ 0.5	2.8 $\pm$ 0.3	1.4 $\pm$ 0.5	2.0 $\pm$ 0.0	1.5 $\pm$ 0.7
FRC	1.9 $\pm$ 0.5	3.5 $\pm$ 2.0	2.9 $\pm$ 1.5	2.0 $\pm$ 0.0	7.0 $\pm$ 0.0
SRC	2.9 $\pm$ 1.2	0.8 $\pm$ 0.9	2.4 $\pm$ 1.9	1.0 $\pm$ 1.4	7.5 $\pm$ 0.7
WHD	3.4 $\pm$ 1.2	1.0 $\pm$ 1.2	2.8 $\pm$ 2.5	1.0 $\pm$ 1.4	9.0 $\pm$ 0.0
MTY	4.4 $\pm$ 1.6	3.3 $\pm$ 1.9	1.5 $\pm$ 1.0	2.0 $\pm$ 1.4	1.0 $\pm$ 0.0
VGR	3.2 $\pm$ 0.7	7.6 $\pm$ 0.7	4.7 $\pm$ 1.6	7.5 $\pm$ 2.1	4.5 $\pm$ 2.1
PLH	15.8 $\pm$ 5.4	31.1 $\pm$ 10.5	21.1 $\pm$ 5.1	48.5 $\pm$ 3.5	21.5 $\pm$ 16.2
PLW	57.9 $\pm$ 7.0	60.5 $\pm$ 17.9	44.1 $\pm$ 11.9	85.0 $\pm$ 12.7	85.0 $\pm$ 5.6
PLT	88.0 $\pm$ 19.7	3.6 $\pm$ 10.0	4.6 $\pm$ 11.2	57.5 $\pm$ 60.1	12.5 $\pm$ 17.7
MLD	98.9 $\pm$ 3.5	68.8 $\pm$ 28.4	64.6 $\pm$ 24.5	5.0 $\pm$ 7.1	40.0 $\pm$ 21.2
VRS	91.4 $\pm$ 10.2	46.7 $\pm$ 16.4	43.4 $\pm$ 26.9	5.0 $\pm$ 7.1	0.0 $\pm$ 0.0

†Plant descriptors abbreviations: GHB = Growth habit, NHD = Number of heads, NST = Number of stems, FRC = Fall recovery, SRC = Spring recovery, WHD = Winter hardiness, MTY = Maturity, VGR = Vigor, PLH = Plant height, PLW = Plant width, PLT = Plant type, MLD = reaction to mildew, VRS = Reaction to virus.

Table 4. Mahalanobis  $D^2$  distances between the 10 clusters of red clover germplasm in the US collection

Clusters	I	II	III	IV	V	VI	VII	VIII	IX
	$D^2$ distances								
II	16.9	—							
III	40.6	26.5	—						
IV	15.3	13.3	29.6	—					
V	44.9	44.2	62.6	53.9	—				
VI	50.5	31.9	45.5	38.4	18.5	—			
VII	30.4	22.5	40.9	29.9	65.6	65.5	—		
VIII	25.0	31.0	50.2	24.2	46.4	53.8	59.4	—	
IX	31.3	41.8	74.3	42.0	40.4	42.4	60.7	60.2	—
X	67.7	77.3	76.8	48.9	58.6	38.4	108.1	56.2	64.2

( $7.0 \pm 0$ ), high spring recovery ( $7.5 \pm 0.7$ ), high winterhardiness ( $9.0 \pm 0$ ), and showed no reaction to virus ( $0.0 \pm 0$ ).

The 10-cluster classification did not necessarily correspond to the country of origin. Instead, the grouping revealed large areas of diversity in Northern and Eastern Europe. A potential source of discrepancy with the data base is that accessions received from the former Soviet Union only carry this generalized identity, rather than a specific state or region. About 33% of accessions in the data base were only listed as the former Soviet Union. In some instances, red clover accessed from different countries fell within the same cluster whereas accessions from a common country were found in several clusters. The greatest diversity was found among accessions from the Soviet Union, Sweden, Switzerland, and Turkey with accessions scattered in a total of 8, 7, 6, and 5 clusters, respectively. This suggests that with the exception of Turkey, the region of North Central Europe is a good source to obtain variability in red clover germplasm. Smith et al. (1985) suggested that southeastern Europe and Asia Minor was the area of origin of red clover, thus, Turkey and areas of the southwestern Soviet Union would fall in this region and would also be expected to contribute variability to the germplasm collection. Although cluster I contained the largest number of accessions, it encompassed only nine countries with more than 63% of these accessions originating from the former Soviet Union. Clusters II and IV were comprised of accessions from 25 and 22 countries, respectively. The close resemblance between germplasm from countries climatically and geographically different such as Sweden and Spain (cluster X) or Norway and Turkey (cluster V) may indicate that exchange

of material has occurred between scientists. Thus, the two accessions of cluster X which records show as originated from Sweden and Spain, may not have actually evolved in these ecological zones. The occurrence of germplasm exchange, with subsequent loss of original collection location, points to the need for complete passport data maintenance by germplasm banks.

### Application

Clustering from this study provided a structure for sampling red clover accessions for genetic, breeding, and agronomic studies where a range of diversity is desired. A major application of this method is to aid in the development of a core collection of red clover. Such a collection could be established by judiciously sampling a limited number of entries from each cluster to represent the ecological and morphological diversity of the germplasm. For example, a stratified random sampling of 10% of each cluster accessions in the 10-cluster division could constitute such a core. The core population would allow the gene bank manager to efficiently use its available resources by maintaining sufficient seed reserve of the core entries to respond to various seed requests (short term) and keep minimal seed supply of the other accessions of the gene pool (long term). The clusters obtained in this study represent a first step in the development of a core collection of red clover.

A stratified sample by country of origin within each core was developed, reviewed by the Clover and Special Purpose Crop Advisory Committee of the National Plant Germplasm System and approved as a preliminary core collection in red



clover. This preliminary core contains at least one accession from every country of origin in the collection, and at least two accessions from each cluster group. A few introductions with known attributes of disease or pest resistance were added to the core. This core was recommended to the red clover germplasm curator and initial seed multiplication is in progress. Once seed of the entire core is available, further investigations of range of variability represented by the core will be conducted. It is intended that the composition of the core collection will be fluid and that it should be adjusted as the range of variability for traits could be extended when additional introductions are evaluated (Brown, 1989).

#### Acknowledgements

We are grateful to Drs S. Littell and K. Portier, IFAS Statistics, University of Florida, for their statistical guidance and to Dr J. McFerson, USDA ARS Plant Genetic Resource Unit Geneva, NY, for his assessment of the collection records.

#### References

- Astley, D., 1987. Genetic resource conservation. *Expl. Agric.* 23:245-257.
- Brown, A. H. D., 1989. Core collections: a practical approach to genetic resources management. *Genome* 31:818-824.
- Brush, S. B., 1989. Rethinking crop genetic resource conservation. *Conservation Biology* 3:19-29.
- Frankel, O. H., 1984. Genetic perspectives of germplasm conservation. pp. 161-170. In: W. K. Arber, K. Llimense, W. J. Peacock & P. Startinger, (Eds.). *Genetic manipulation: Impact on man and society*. Cambridge University Press, Cambridge.
- Holden, J. H. W., 1984. The second ten years. p. 276-285 In: J. H. W. Holden & J. T. Williams (Eds.). *Crop genetic resources: Conservation and evaluation*. Allen and Unwin, Inc. Winchester, MA.
- Kouamé, C. N. & K. H. Quesenberry, 1992. Relationships among accessions of red clover in the US collection. pp. 54-55. In: D. S. Wofford & K. H. Quesenberry (Eds.). *Proceed. 12th Trifolium Conference*. University of Florida, Gainesville, FL.
- SAS Institute, 1985. *SAS User's Guide: Statistics*. NC State Univ. Press. Cary, NC.
- Souza, E. & M. E. Sorrells. 1991. Relationships among 70 North American oat germplasms: I. Cluster Analysis using the quantitative characters. *Crop Sci.* 31:599-605.
- Smith, R. R., N. L. Taylor & S. R. Bowley, 1985. Red clover. In: N. L. Taylor (Ed.). *Clover Science and Technology*. *Agronomy* 25:457-470.
- Zohary, M. & D. Heller, 1984. *The Genus Trifolium*. Israeli Academy of Sciences and Technology, Jerusalem.