Genetic Resources and Crop Evolution **49:** 271–283, 2002.

Genetic variability in cultivated common bean beyond the two major gene pools

F.M.A. Islam[1], K.E. Basford[1,*], R.J. Redden[2], A.V. Gonzalez[3], P.M. Kroonenberg[4] and S. Beebe[3]

[1]*School of Land and Food Sciences, The University of Queensland, Brisbane QLD 4072, Australia;* [2]*Hermitage Research Station, QDPI, Warwick QLD 4370, Australia;* [3]*CIAT, Cali, Colombia;* [4]*Department of Education, Leiden University, Wassenaarseweg 52, Leiden, The Netherlands;* [*]*Author for correspondence*

Received 14 February 2001; accepted in revised form 27 July 2001

Key words: Characterization, Common bean, Core collection, Gene pools, Multivariate analysis

Abstract

It is generally accepted that two major gene pools exist in cultivated common bean (*Phaseolus vulgaris* L.), a Middle American and an Andean one. Some evidence, based on unique phaseolin morphotypes and AFLP analysis, suggests that at least one more gene pool exists in cultivated common bean. To investigate this hypothesis, 1072 accessions from a common bean core collection from the primary centres of origin, held at CIAT, were investigated. Various agronomic and morphological attributes (14 categorical and 11 quantitative) were measured. Multivariate analyses, consisting of homogeneity analysis and clustering for categorical data, clustering and ordination techniques for quantitative data and nonlinear principal component analysis for mixed data, were undertaken. The results of most analyses supported the existence of the two major gene pools. However, the analysis of categorical data of protein types showed an additional minor gene pool. The minor gene pool is designated North Andean and includes phaseolin types CH, S and T; lectin types 312, Pr, B and K; and mostly A5, A6 and A4 types α-amylase inhibitor. Analysis of the combined categorical data of protein types and some plant categorical data also suggested that some other germplasm with C type phaseolin are distinguished from the major gene pools.

Introduction

Common bean (*Phaseolus vulgaris* L.) is one of the most important food legumes worldwide. As stated by van Schoonhoven and Voysest (1991), it is adapted to a variety of climatic conditions, being grown from around 52° north latitude to 32° south latitude in the humid tropics, in the semiarid tropics, and even in cold-climate regions. Latin America, from Mexico to the southern Andes, is now accepted to be the centre of origin for the common bean and is the leading bean-producing region in the world.

Genetic diversity of common bean is not randomly distributed geographically (Tohme et al. 1995). Evidence based on morphological traits (Singh et al. 1991a), seed proteins (Gepts et al. 1986), isozymes (Singh et al. 1991b) and DNA markers (Becerra Velásquez and Gepts 1994) indicates that two major gene pools exist in cultivated common bean, a Middle American and an Andean one.

Seed protein and isozyme variability have been used extensively in many crops, including common bean, to detect patterns and levels of genetic diversity within a species (Loveless and Hamrick 1984). Phaseolin is the major seed storage protein of common bean and represents up to 49% of the total protein concentration of the seed (Ma and Bliss 1978). Phaseolin protein diversity has contributed to the evidence for two gene pools (Gepts et al. 1986; Singh et al. 1991a; Gepts 1998). The genetic structure of cultivated common bean parallels that of wild bean. Gepts et al. (1986) showed that both landraces and

wild races from Mexico display phaseolin morphotypes that are distinct from those found in South America.

In the Middle American gene pool, the S-type (Sanilac) phaseolin predominates, while the Andean gene pool expresses primarily the T-type (Tendergreen) phaseolin (Gepts and Bliss 1986; Gepts 1993; Becerra Velásquez and Gepts 1994) followed by C and H type (Koenig and Gepts 1989). Based on a unique phaseolin and unusual isozyme patterns, Debouck et al. (1993) reported another gene pool of wild bean in Ecuador-Northern Peru. The seeds of common bean also synthesize a number of proteins other than phaseolin, including true lectins or lectin-like proteins such as arcelin (Osborn et al. 1988) which are thought to protect seeds against insect predation (Gatehouse et al. 1984; Osborn et al. 1988). Bean seeds also contain another type of protein, i.e. α-amylase inhibitor. This protein inhibits insects and is thought to be a plant defense protein (Ishimoto et al. 1995a).

The genetic structure of a core collection of wild bean was studied by Tohme et al. (1996) by analyzing DNA fingerprints with amplified restriction fragment polymorphism (AFLP) technology (Lin and Kuo 1995). The study found that in addition to the major Andean and Middle American gene pools, there was a distinct minor gene pool from Ecuador-Northern Peru and another minor gene pool from Colombia. The Ecuadorian-Northern Peruvian and Colombian pools were regarded as genetically unique and not simply the product of hybridization between Middle American and Andean gene pools. Therefore, although most studies support the hypothesis that there are two major gene pools, as many as two more groups could be expected based on AFLP analysis, phaseolin and isozymes of wild bean.

In cultivated common bean, some evidence has also been reported that a North Andean gene pool exists. Gepts and Bliss (1986) reported finding a 'CH' type phaseolin in the wild bean and a 'B' type phaseolin (later identified with 'S' type phaseolin) in both wild and landraces of native Colombian germplasm and suggested that domestication had occurred in this region. Chacón et al. (1996), Beebe et al. (1997) reported 'CH' and 'L' phaseolin types in both Colombian wild bean and a few landraces from the Northern Andes. The latter types were not represented in this study. Chacón et al. (1996) established the genetic similarity of the respective wild and cultivated common bean using AFLP technology. However, a North Andean gene pool or any other minor gene pool of cultivated common bean has not been well substantiated nor characterized.

The objective of this study was to analyze the structure of the bean core collection with different combinations of data types (categorical, quantitative and mixed) to assess the evidence for the existence of distinct gene pools. It also of interest to see if there was any effect of different data types on gene pool classifications with an emphasis on the non-human selected attributes.

Materials and methods

Accessions

The International Centre for Tropical Agriculture (CIAT), Cali, Colombia holds the world's largest collection of common bean germplasm, around 40000 accessions collected worldwide. A core collection has been composed consisting of a stratified random sample of 1441 accessions obtained as genetically representative of the total collection (Tohme et al. 1995). Amongst the 1441 accessions, 1072 with complete information were from Middle America (Mexico, Belize, Costa Rica, El Salvador, Guatemala, Honduras and Nicaragua) and the Andean countries (Colombia, Ecuador, Peru, Bolivia and Argentina) which are considered to be primary centres of diversity as per Tohme et al. (1995). Amongst the Andean countries, Colombia, Ecuador and northern Peru (the provinces of Amazonas, Cajamarca, La Libertad, Lambayeque, Piura and San Martin) are considered as north Andes. The definition of primary and secondary centres of diversity is, to some extent, arbitrary. Regions in which the wild common bean exists and domestication could have occurred constitute primary centers, while the regions of introduction in the post-Columbian age, such as Africa, are secondary centres. Accessions from the primary centres only were investigated in this study. Representation of accessions by countries was as follows: Argentina (19), Belize (4), Bolivia (20), Colombia (87), Costa Rica (22), Ecuador (62), El Salvador (14), Guatemala (86), Honduras (13), Mexico (393), Nicaragua (13) and Peru (338).

Experimental details

All accessions were planted (unreplicated) in Sep-

tember 1992 in the field in Tenerife, Department of Valle, Colombia which is at 3° 30′ 16″ N latitude and 76° 21′ 23″ W longitude. The elevation is 2200 m.a.s.l. and the average temperature is 15 °C. It has andosol (highly organic) soil and an average rainfall 1100 mm per year. Crops were managed under rainfed conditions. At this site, both highland and lowland adapted genotypes developed and flowered normally, providing comparable data on the entire set from a single site and planting season. Bush genotypes were planted in 1 m rows while climbing accessions were planted in two hill plots of four plants each. Climbers were staked to permit full development of climbing habit. One plot per accession was utilized. Data on morphological traits were taken as per Singh (1989), with seed analysed for biochemical (LKB GELSCAN 1986) data to provide categorical and quantitative attributes. More details on the seed and disease attribute measurements at CIAT are given by van Schoonhoven and Pastor-Corrales (1987).

Attributes used in the analysis

The attributes used in this study were protein types and concentrations, plant categorical (qualitative) characters and pod and seed quantitative characters, particularly because of their wide variation for the 1072 common bean accessions (Table 1). The protein types were phaseolin (C, Ca, H, CH, Ko, Pa, To, T, S, Sb, Sd, and M), lectin (K, T, V, 310, 312, B, Pr, M, and Po) and α-amylase inhibitor (A1–A7) types. Phaseolin type was characterized by comparison with standards defined by Ma and Bliss (1978), Toro et al. (1990), Brown et al. (1982). Brown et al. (1982) and Osborn et al. (1984) first defined and classified the

Table 1. Attributes recorded on 1072 accessions from the primary centres of origin in CIAT's common bean core collection.

Description	Category definition
Category attributes (protein types)	
Phaseolin types	
Lectin types	
α-amylase inhibitor types	
Categorical attributes (plant and seed)	
Hypocotyl color	green, pink and purple
Primary color of wings	white, white with tones, pink, and lilac, light purple and dark purple
Shape of the bracteoles	cordate, ovate, lanceolate and triangle
Leaf pubescence upper surface	no hairs, slight hairs, moderate and dense
Size of bracteoles	small, medium and large
Stripes on the neck of the flower	none, light and accentuated
Central vein	green, pink and purple
Color of raphe	white, cream, yellow or golden, brown, pink, red purple, black and others (gray and green)
Seed brilliance	opaque, semibrilliant and brilliant
Seed form	round, ovoid, elliptical, small squash, ovoid elongated, elongate, elongate (almost square), Kidney shaped and flat on the hilum side, kidney shaped and curved
Primary seed color	White, cream, yellow, brown, pink, purple, black and others
Quantitative attributes (protein concentrations)	
Percentage of protein concentrations	
Percentage of phaseolin concentrations	
Percentage of lectin concentrations	
Percentage of α-amylase inhibitor concentrations	
Quantitative attributes (pod and seed)	
Number of days to flower	
Average pod length (mm)	
Average pod width (mm)	
Seed weight (g/100 seed)	
Average seed length (mm)	
Average seed width (mm)	
Average seed thickness (mm)	

lectin types, while types of α-amylase inhibitor were classified by Ishimoto et al. (1995b).

The protein types in common bean seed were selected because of their previous application in gene pool classification. Some categorical morphological plant characters which are considered to reflect natural rather than human selection were included. Examples are colour of hypocotyl (the part of the plant beneath the cotyledons that exists at germination when the plantlet emerges from the soil), primary colour of wings and leaf pubescence (leaf hairiness of the upper surface). Some qualitative and quantitative characters (seed colour, seed brilliance and seed size) are assumed to be human selected or biased in selection by humans. Other quantitative characters measured were percentage of protein concentrations in the seed and some other pod and seed attributes. In selecting pod and seed quantitative characters, the correlation structure between them was taken into account. For example, with phenology data, only flowering data were used because of the high correlation between days to flowering and days to maturity ($r = 0.81$).

Statistical analysis

Complementary clustering and ordination techniques or pattern analysis (Williams 1976) was used to explore different aspects of the data by Islam (2001). The results from ordination and clustering were combined on the ordination display by using different symbols to depict the various clusters. The pattern analysis techniques chosen for the different common bean attribute types are given in Table 2.

An agglomerative hierarchical clustering technique with an incremental sum-of-squares sorting strategy (often called Ward's method (Ward 1963)) was used. This technique divided the accessions into distinct groups in such a way that the accessions within a group were more similar to each other than to the accessions in other groups.

Then ordination procedures were used to give a geometrical or spatial representation of the individuals in a low (usually two- or three-) dimensional space, such that the distance between points (individuals) represents their dissimilarity (Manly 1994). This reduced the original set of attributes into a smaller set of ordered uncorrelated components which accounted for the maximum variability as per Gabriel (1971).

Homogeneity analysis was applied to nominal or unordered categorical (qualitative) attributes for interpreting the underlying structure of the data array (Gifi 1981). It is a weighted principal component analysis of a contingency table, decomposing the table into row and column coordinates that can be displayed graphically (SAS, Institute Inc. 1989).

Nonlinear principal component analysis which is an extension of ordinary principal component analysis was also applied to allow inclusion of quantitative attributes with categorical attributes (Jolliffe 1986). Transformation of the raw data was necessary for the use of nonlinear principal component analysis. Each of the quantitative attributes (Table 1) was grouped into six to eight categories based on their mean and standard deviation. The categorical values were assigned as follows $< \mu - 3\sigma = 1$, $\mu - 3\sigma$ to $\mu - 2\sigma = 2$,..., $> \mu + 3\sigma = 8$ as per Carbonel (1995). The new categories spanned equal intervals and no category had fewer than five accessions to prevent rare categories from unduly influencing the analysis (Gifi 1990).

Andean phaseolins A, Tca and To2 appear with frequencies less than or equal to two and these were

Table 2. Statistical analyses chosen for the different attribute types and combinations of attributes within those types.

Attributes	Statistical techniques
1. Categorical attributes	Homogeneity analysis and clustering
(a) Protein types	
(b) Plant categorical characters	
(c) Protein types and plant categorical characters	
2. Quantitative attributes	Clustering and principal component analysis
(a) Protein concentrations	
(b) Pod and seed quantitative characters	
(c) Protein concentrations and pod and seed quantitative characters	
3. Mixed attributes	Non-linear principal component analysis
(a) Protein types and concentrations	
(b) Plant categorical and pod and seed quantitative characters	
(c) Protein types and concentrations, plant categorical and pod and seed quantitative characters	

Table 3. Distribution of phaseolin types of 1072 accessions from the primary centres in CIAT's common bean core collection by their origin, ordered from Middle America (Mexico to Costa Rica) to Andean (Colombia to Argentina). The blanks are zeros.

Origin	Phaseolin types												
	C	Ca	H	CH	Ko	Pa	To	T	S	Sb	Sd	M	Total
Mexico								18	194	108	27	46	393
Belize									4				4
Guatemala								3	47	22	12	2	86
El Salvador									10		4		14
Honduras									6	5	2		13
Nicaragua									7	2	4		13
Costa Rica								2	10	5	5		22
Columbia		7	18	1				34	27				87
Ecuador	1	3	5	3			2	42	6				62
Peru	7	18	50	2	3	28	7	196	28				339
Bolivia		2			3			14					20
Argentina		2						16					19
Total	8	32	73	6	6	28	9	325	341	142	54	48	1072

grouped with Andean type phaseolin T. All these Andean phaseolins are associated with similar lectins and α-amylase inhibitor types and thus would not be expected to separate from T in any case. Similarly M1 and M15 types of Middle American origin with low frequency were grouped with M type phaseolin. These accessions likewise have similar lectins and α-amylase inhibitor as the M types. B type phaseolin was replaced by S type phaseolin, based on subsequent studies that suggested identity of these.

The analysis of categorical data by homogeneity analysis was achieved using the program HOMALS, while the analysis of mixed data by nonlinear principal component analysis was achieved using PRINCALS. Both were contained in SPSS (SPSS for Windows 1999). The statistical software package S-PLUS (S-PLUS Inc. 1995) was used for clustering and ordination of the quantitative attributes.

Results

Initially, the distribution of the origin of accessions by their protein types is described. Then, as can be seen from Table 2, the analysis of the germplasm data was approached in a systematic way. Within each data type, different combinations of attributes were analysed separately, i.e. (a) and (b), and then the combined data were analysed, i.e. (c). Full details can be found in Islam (2001), but only a selection of these, i.e. 1(a), 1(c) and 3(c), are presented here as they cover the range of summary descriptions obtained. The others are only remarked upon for completeness.

Distribution of the origin of accession by their protein types

The distribution of phaseolin, lectin and α-amylase inhibitor types by origin is given in Tables 3 to 5, respectively. Arrangement of phaseolin types (Table 3) shows that C, Ca, CH, H, Ko, Pa, T and To are phaseolins found mainly in the Andean region, while M, S, Sb and Sd predominantly occur in Middle America. Arrangement of lectin types (Table 4) shows that K, T, V, 312 and Pr are predominantly Andean, while M and Po are predominantly Middle American and lectin types 310 and B are widely distributed in both Andean and Middle American countries. The α-amylase inhibitor types (Table 5) are widely distributed as well, except for A1 type which is predominant in the Andean region and A7 and A3 types which are mainly in the Middle American region.

Analysis of categorical attributes

The first stage of the categorical data analysis, 1(a), involved the three protein types. The first component from the homogeneity analysis accounted for 61% of the total variability, while the second component accounted for 39%. The discrimination measures for the attributes for first and second components, component scores for phaseolin types (0.844 and 0.418), lectin types (0.774 and 0.516) and α-amylase inhibitor types (0.704 and 0.537) indicate that all protein types are important in understanding the variation among the accessions. This may not have been apparent from the analysis of each character separately.

Table 4. Distribution of lectin types of 1072 accessions from the primary centres in CIAT's common bean core collection by their origin, ordered from Middle America (Mexico to Costa Rica) to Andean (Colombia to Argentina). The blanks are zeros.

Origin	Lectin types									
	K	T	V	310	312	B	Pr	M	Po	Total
Mexico	2	61		33	1	54	1	170	71	393
Belize								4		4
Guatemala	1	11		5		9		40	20	86
El Salvador	1	3		2				8		14
Honduras	2	1				3		4	3	13
Nicaragua	1	2				3		7		13
Costa Rica	2	4		3		2		10	1	22
Colombia	1	50	2	1	8	9	8	6	2	87
Ecuador		37	4	3	1	9	4	3	1	62
Peru	15	261	1	7	4	28	9	13	1	339
Bolivia		19						1		20
Argentina	3	14				2				19
Total	28	463	7	54	14	119	22	266	99	1072

The accessions can be presented graphically in the Euclidean space spanned by the first and second components. This can be done by constructing displays for the accession scores labeled by phaseolin type (Figure 1). Note that a single letter was used to denote phaseolin types in Figures 1, 3 and 4 to make it easier to distinguish the position of individual accessions. A large group of accessions having M, S and Sb type phaseolin (negative on vector 1 and to the lower left in Figure 1) are the Middle American accessions in contrast with the accessions with H, Pa and T type phaseolin (to the right in Figure 1), which are the Andean type phaseolin (Table 3). However, one small group of accessions appears as outlying points on the top-left hand corner of Figure 1, and these have CH type phaseolin (represented as D in the Figure 1). The accession points were also plotted using their lectin and α-amylase inhibitor types (figures are not shown). By observing the phaseolin, lectin and α-amylase inhibitor types associated with accession points, it was noted that the Middle American group is characterized by 310, B and Po lectin types and A7 α-amylase inhibitor types, while the Andean group has K and T lectin types and A1 α-amylase inhibitor types (Tables 4 and 5).

A cluster analysis was performed on the two-dimensional object scores found from homogeneity analysis as per Jongman et al. (1987), Hill (1991), Carey et al. (1995) to formalise the grouping pattern found in that analysis. The dendrogram was truncated at the four group level which explained 82% of the variation in the component scores (Figure 2). Among the four groups, group 1 was formed with 98% Middle American accessions; group 2 with mixed

Table 5. Distribution of α-amylase inhibitor types of 1072 accessions from the primary centres in CIAT's common bean core collection by their origin, ordered from Middle America (Mexico to Costa Rica) to Andean (Colombia to Argentina). The blanks are zeros.

Origin	α-amylase inhibitor types							
	A1	A2	A3	A4	A5	A6	A7	Total
Mexico	17	8	6	5	25	25	307	393
Belize					3	1		4
Guatemala	5	4	1		4	2	70	86
El Salvador		1	1			1	11	14
Honduras							13	13
Nicaragua	1						12	13
Costa Rica	2	3				1	16	22
Colombia	26	9		7	16	13	16	87
Ecuador	35	3		3	5	3	13	62
Peru	183	16		24	26	5	85	339
Bolivia	12	2		2			4	20
Argentina	8			2	1		8	19
Total	289	46	8	43	80	51	555	1072

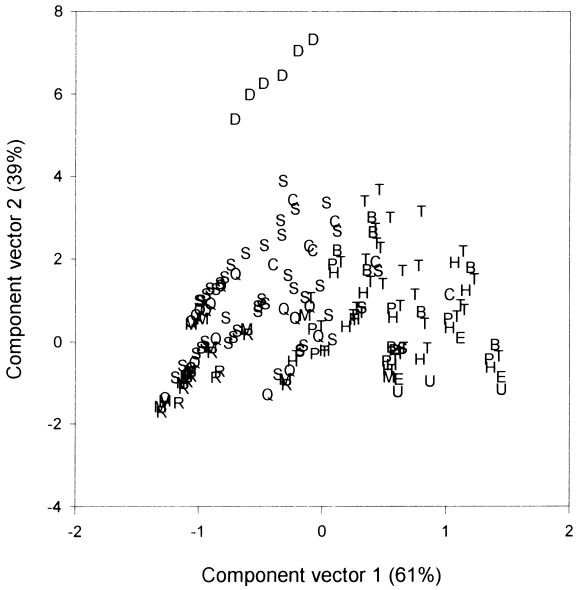


Figure 1. Plot of the accession scores for the two component vectors for the 1072 common bean accessions from the homogeneity analysis of the protein tyupe categorical attributes. Accessions are labelled by phaseolin type using a single letter notation for maximum clarity (C, B = Ca, D = CH, H, E = Ko, M, P = Pa, S, Q = Sb, R = Sd, T, U = To).

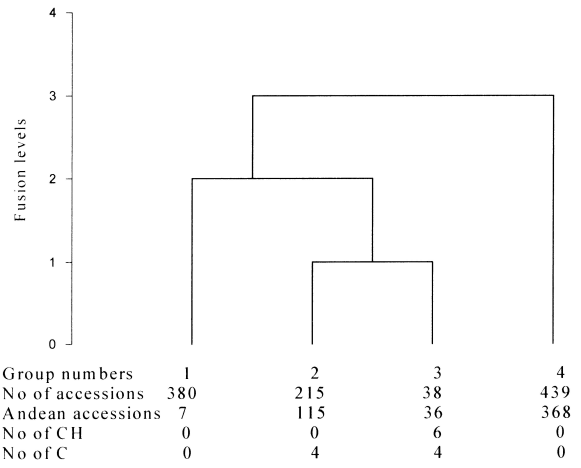


Figure 2. Partial dendrpgram from the cluster analysis of the two component scores for the 1072 common bean accessions from the homogeneity analysis of the protein type categorical attributes.

accessions (59% Andean); group 3 with mainly North Andean accessions (74% from North Andean countries which includes Colombia, Ecuador and northern Peru; 21% Andean and 5% Middle American); and group 4 with mainly Andean accessions (86% Andean and North Andean). The North Andean group (group 3) had 38 accessions including all six CH phaseolin types.

In an attempt to emphasize the non-human selected attributes, only plant categorical characters were included in the second stage analysis, 1(b). Regardless of the number or combinations considered, the analyses did not show any separation of accessions into two gene pools nor a group of outliers as found in Figure 1(Islam 2001).

The third stage of the categorical data analysis, 1(c), was a homogeneity analysis of protein types and various combinations of different numbers of plant categorical characters. As before, seed attributes were not included as they were considered to be influenced by human selection. In every case, one small group of accessions with CH type phaseolin appears as outlying points, as was found in the analysis of protein types on their own. When one or more of the plant characters, hypocotyl colour, primary colour of wings and shape of bracteole, were added to protein types, homogeneity analysis identified another small group of outlying accessions in addition to those found from the analysis of the protein types. The first four components from the homogeneity analysis accounted for 87% of the variability in the original six attributes. The discrimination measures (Table 6) indicate that the first, third and fourth components mainly reflect phaseolin types, lectin types and α-amylase inhibitor types, while the second component mainly reflects hypocotyl colour and primary colour of wings.

The object scores for each accession are plotted in the Euclidean space spanned by the first two components (Figure 3(a)) and the third and fourth components (Figure 3(b)). Component vectors 1 and 2 give a separation consistent with two gene pools (left and right-hand sides of the display), while component vectors 3 and 4 do not show any clear separation except for two minor groups appearing on the top left-hand corner and bottom part of the plot. The group located on the top left-hand corner corresponds to the CH type phaseolin found in the homogeneity analysis of the protein types. However, a separate group of accessions with C type phaseolin is also distinguished in this analysis.

Analysis of quantitative attributes

No clear separation was obtained from the first stage of quantitative data analysis, 2(a), involving only protein concentrations in bean seeds (Islam 2001). In contrast, a wide separation of accessions into the two

Table 6. Discrimination measures for each attribute (protein type, hypocotyl colour, shape of bracteole and primary colour of wings) per component from the homogeneity analysis of the 1072 common bean accessions.

Attributes	Components			
	1	2	3	4
Phaseolin type	0.812	0.135	0.430	0.454
Lectin type	0.698	0.040	0.488	0.449
α-amylase inhibitor type	0.630	0.023	0.522	0.205
Hypocotyl colour	0.207	0.801	0.007	0.004
Primary colour of wings	0.254	0.827	0.006	0.083
Shape of bracteole	0.197	0.044	0.054	0.117
Variation accounted for	0.32	0.21	0.17	0.15

major gene pools of an Andean one and a Middle American one was found when either an analysis of the quantitative pod and seed characters, 2(b), or the combination of the percentages of protein concentrations and pod and seed characters, 2(c), was undertaken (Islam 2001).

Analysis of mixed attributes

The analysis of the protein types and concentrations, 3(a), gave a similar result as for the combined categorical data analysis. The first two component vectors from the nonlinear principal component analysis gave two distinct gene pools while the third and fourth component vectors gave two additional small groups of points, one group with CH type phaseolin and the other with C type phaseolin (Islam 2001). The analysis of plant categorical and pod and seed quantitative characters, 3(b), separated the Andean and the Middle American gene pools (Islam 2001).

The third stage of the analysis, 3(c), involved all the mixed attributes, i.e., protein types and concentrations, plant categorical and pod and seed quantitative characters. The first four components from the nonlinear principal component analysis accounted for 43% of the variability in the original 25 attributes of the common bean germplasm data (Table 7). This analysis was also performed without the attribute of primary seed colour, and the results were quite similar.

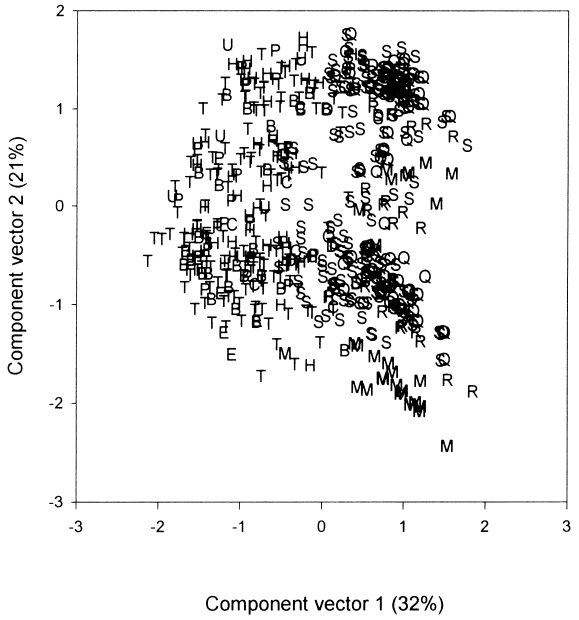
The discrimination measures (Table 7) indicate that component vector 1 places importance on attributes such as phaseolin types, lectin types, α-amylase inhibitor types, length of pod, length of seed, thickness of seed and width of seed in distinguishing between accessions, while component vector 2 shows that hypocotyl colour, primary colour of wings and primary seed colour are important in distinguishing between accessions.

The object scores for each accession are plotted in the Euclidean space spanned by the first two component vectors (Figure 4(a)) and third and fourth component vectors (Figure 4(b)). It appears that the human selected attributes strongly influenced the separation of accessions (Table 7). From the first two components, the accessions can be roughly divided into two groups (left and right sides of the display) consistent with the two major gene pools. Perhaps the only inference that can be made from components 3 and 4 is that four out of the six accessions with CH type phaseolin were in the bottom right-hand corner and hence were in an outlying region and somewhat distinct from other accessions.

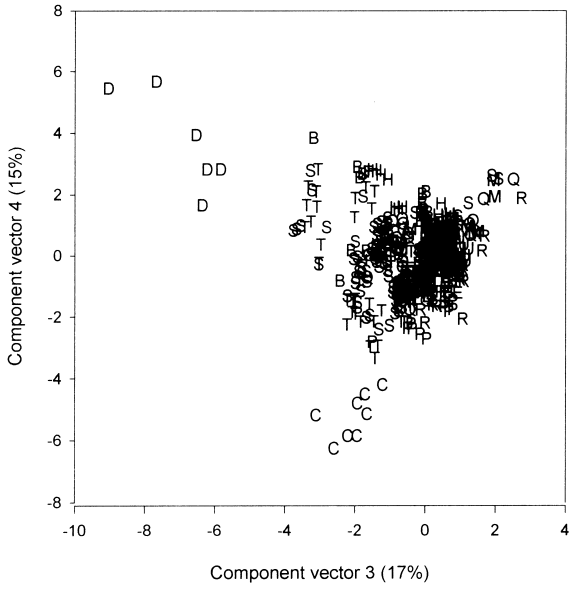
Discussion

Results of the different analyses presented led to somewhat different conclusions. The homogeneity analysis based on protein types indicated the existence of two gene pools of cultivated common bean (as is widely accepted) and also a minor gene pool including CH type phaseolin. Inclusion of some other qualitative characters also suggested the existence of two major gene pools and some other germplasm with CH and C type phaseolin not fitting into the two major pools.

Using quantitative attributes, the analysis of pod and seed indicated two major gene pools. Inclusion of protein concentrations with pod and seed quantitative characters produced a similar result of two major gene pools, but the separation of the two groups became weaker in the presence of protein concentrations. In fact, the quantitative data of pod and seed characters were heavily weighted with attributes that reflect human selection for seed size (seed weight, pod and seed attributes). Apparently, seed size and pod characters were critical to a clear definition of two gene



(a)



(b)

Figure 3. Plot of the accession scores for the (a) first and second component vectors and (b) third and fourth component vectors for the 1072 common bean accessions from the homogeneity analysis of the protein type and plant categorical attributes. Accessions are labelled by phaseolin type using a single letter notation for maximum clarity (C, B = Ca, D = CH, H, E = Ko, M, P = Pa, S, Q = Sb, R = Sd, T, U = To).

pools in the quantitative analysis. Therefore, it is possible that the evolutionary origins of the cultivated common bean independent of human selection could be better reflected by the categorical data alone. The clustering pattern based on the homogeneity analysis of the categorical attributes (either protein types or a combination of protein types and plant characters) did reveal Middle American and Andean groups with distinct patterns, as well as other groups with an intermediate pattern. Thus the present results are consistent with the distinction of an Andean and a Middle American gene pool, but suggest additional variability which does not fit in either one of these.

The intermediate groups found in the protein type analysis are of particular interest. Amongst the intermediate groups, one group comprised mixed accessions (59% Andean) and another group of 38 accessions consisted mainly of North Andean accessions. The North Andean group presented phaseolin types CH (all six found in the study), S (7) and T (16); lectin types 312, Pr, B and K; and mostly A5, A6 and A4 α-amylase inhibitor types. Many of these protein morphotypes are uncommon across the primary centres, and a disproportionate number of rare types occurred within this group. The mixed group with Andean and Middle American accessions separated from the North Andean group, in part due to lectin (B and Pr types) and α-amylase inhibitors types (A4, A5 and A6). Thus the protein types appear to have contributed significantly to the definition of the North Andean and mixed clusters.

In the past, phaseolin has been used extensively as an evolutionary marker (Gepts 1988), and in the present analyses this has been quantified by the discrimination measures. Phaseolin consistently gave amongst the highest values when it was included in the analysis. Possibly this is the first time that lectin and α-amylase inhibitor types have been used to discriminate common bean gene pools. Although genes for lectin and α-amylase inhibitor are controlled by linked loci (Nodari et al. 1992), no lectin was associated with a unique α-amylase inhibitor type, nor *vice versa*.

When the cluster analysis was performed using the combined categorical data, results were similar to those found by the analysis with only three protein types alone. However, in the analysis of combined categorical data, the accessions with C type phaseolin fused with a mixed cluster while they formed a small group in the homogeneity analysis. The North Andean group with CH type phaseolin still appeared as dis-

Table 7. Discrimination measures for the mixed attributes (protein types and concentrations, plant categorical and pod and seed quantitative characters) per component from the nonlinear principal component analysis of the 1072 common bean accessions.

Attributes	Components			
	1	2	3	4
Phaseolin type	0.710	0.148	0.096	0.233
Lectin type	0.499	0.069	0.117	0.119
α-amylase inhibitor type	0.398	0.041	0.040	0.129
Hypocotyl color	0.240	0.530	0.117	0.136
Primary color of wings	0.164	0.668	0.068	0.032
Shape of bracteoles	0.039	0.031	0.036	0.138
Leaf pubescence upper surface	0.313	0.184	0.008	0.060
Size of bracteoles	0.057	0.010	0.054	0.042
Strips on the neck of the flower	0.384	0.138	0.011	0.015
Central vein	0.279	0.248	0.011	0.095
Color of raphe	0.307	0.202	0.155	0.292
Seed brilliance	0.086	0.009	0.087	0.065
Seed form	0.239	0.184	0.467	0.194
Primary seed color	0.288	0.413	0.354	0.296
Percentage of protein concentration	0.163	0.006	0.155	0.058
Percentage of phaseolin concentration	0.097	0.061	0.009	0.020
Percentage of lectin concentration	0.051	0.019	0.030	0.029
Percentage of α-amylase inhibitor concentration	0.009	0.008	0.011	0.022
Number of days to flower	0.071	0.078	0.060	0.012
Average pod length (mm)	0.244	0.161	0.151	0.015
Average pod width (mm)	0.426	0.010	0.007	0.142
Seed weight (g/100 seed)	0.023	0.013	0.064	0.016
Average seed length (mm)	0.451	0.078	0.160	0.064
Average seed width (mm)	0.479	0.010	0.080	0.129
Average seed thickness (mm)	0.556	0.010	0.212	0.038
Variance accounted for	0.21	0.11	0.07	0.06

tinct within a group of mainly North Andean accessions when homogeneity analysis was performed with protein types and some other categorical data. The plant categorical data alone did not indicate outliers which could be explored as new gene pools.

The foregoing analyses present more evidence that wild bean populations in the North Andes have been domesticated. This would offer the possibility that unique alleles exist in North Andean landraces that are absent in the other two gene pools. This possibility was first posed by Gepts and Bliss (1986), based on the identification of a 'B' phaseolin in wild and cultivated common bean from Colombia. We have had difficulty finding consistent differences between 'B' and 'S' phaseolin, although essentially the same conclusion resulted from the discovery of 'CH' and 'L' phaseolin types in both wild and cultivated common bean from Colombia (Chacón et al. 1996; Beebe et al. 1997). This conclusion was supported by AFLP analysis of nuclear DNA, showing a relationship between Colombian wild and cultivated common bean and their uniqueness compared to wild and cultivated common bean of the two major gene pools (Chacón et al. 1996). In the present study, this conclusion is further strengthened by the presence of lectin and α-amaylase inhibitor types in North Andean landraces that are atypical of the major gene pools, together with unusual combinations of morphological traits. The North Andean group was also observed to be later to flower than other groups.

Phaseolin is not the only marker that indicates the presence of a North Andean pool, as 'CH' and 'L' phaseolins also appear to be indicators. In particular, CH phaseolin is the most common type recovered from collections of wild bean in Colombia, representing about 85% of accessions (unpublished data, CIAT). Weedy types present about 50% CH type. However, the proportion of 'CH' and 'L' types in cultivated common bean is very low. In the core collection, no 'L' types and only six 'CH' types were identified out of 1441 total accessions. All six 'CH' accessions are from the North Andean region. The core collection contains 283 accessions from this same region, thus CH accessions represent 2% of this subset. Colombian accessions in the core number 87, including one CH, or 1.5%. Chacón et al. (1996)

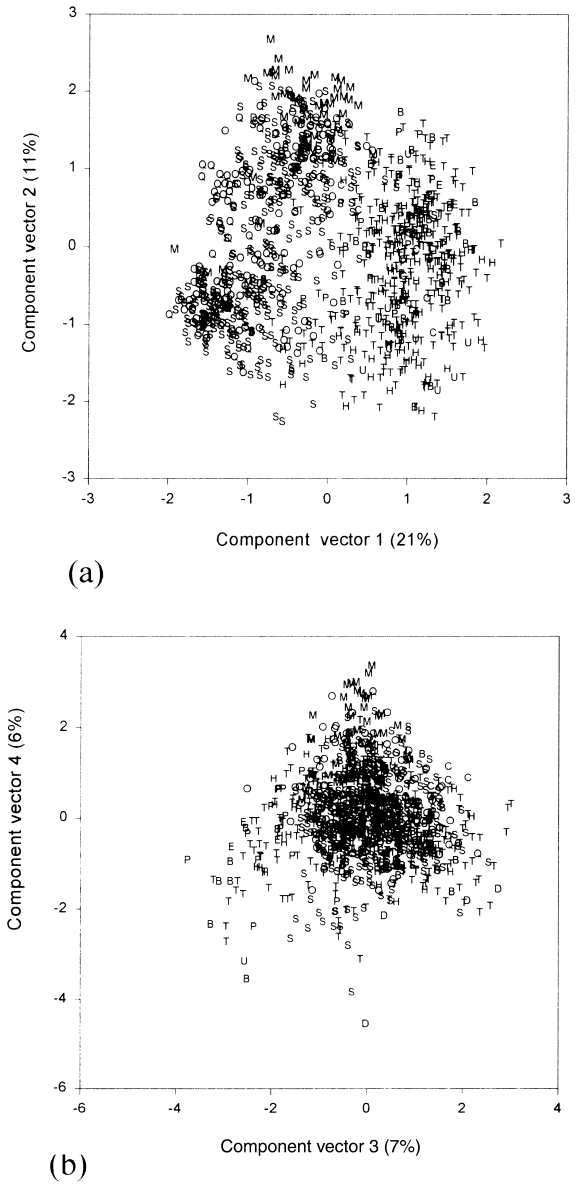


Figure 4. Plot of the accession scores for the (a) first and second component vectors and (b) third and fourth component vectors for the 1072 common bean accessions from the nonlinear PCA for mixed attributes. The accessions are labelled by phaseolin type using a single letter notation for maximum clarity (C, B = Ca, D = CH, H, E = Ko, M, P = Pa, S, Q = Sb, R = Sd, T, U = To).

reported 3% each of 'CH' and 'L' types in Colombian cultivars. By any criterion these proportions suggest that this pool is of limited size.

The North Andean group of cultivated common bean must be viewed in light of the structure of wild bean from the same region. Although wild bean has been found in a rather small area of the North Andes, in North East Colombia and in Ecuador-North Peru (Toro et al. 1990), useful comparisons with cultivated common bean can be drawn. It was found by using AFLP markers (Tohme et al. 1996) that some wild bean from Colombia is genetically quite unique, while others tend to cluster with germplasm from Middle America or the Andean zone. This has been interpreted as reflecting substantial introgression into Colombian wild bean. There is evidence that some North Andean landraces are intermediate for several traits between Middle American and Andean gene pools, suggesting considerable introgression among these major gene pools. Middle American and Andean common bean are both cultivated in the North Andean region, and evidence exists that these hybridize with each other and with local bean populations (Beebe et al. 1997). Thus, North Andean landraces can reflect the influence of both introduced germplasm and that derived from local domestication events, and in varying degrees. The North Andean region has probably been a crossroads of both wild and cultivated common bean from the north and south. It may also be that conditions that contribute to cross pollination have contributed to introgression in both wild and cultivated common bean. Such introgression seems to be a characteristic of both wild and the cultivated common bean North Andean gene pool, and has contributed significantly to their evolution. It is possible that introgression with the major gene pools may have compromised the integrity of the North Andean gene pool.

A fourth wild bean gene pool in Ecuador and Northern Peru was detected by Tohme et al. (1996) with DNA markers, confirming that a unique group exists in this region as suggested by Debouck et al. (1993). This wild bean was characterized by an unusual phaseolin type designated I type. Kami et al. (1986) analyzed the DNA sequence for the I phaseolin allele and concluded that it could be ancestral to the phaseolin alleles of Middle American and Andean types. However, no I phaseolin has ever been found in a domesticated common bean. This fact, and the genetic distance that this wild bean presents in relation to other genetic groups, suggests that these wild beans were not involved in domestication.

The results suggest that accessions with C phaseolin merit more study. Analyses of protein types with other categorical data and mixed analyses of protein types with protein concentrations indicated a distinct group centered on C phaseolin. All the accessions of C type phaseolin are Peruvian. The C

phaseolin has been observed to be segregating in populations of a wild-weed-crop complex in Peru (Beebe et al. 1997). It is significant that C phaseolin is relatively uncommon in the Andean gene pool, being detected in less than 2% of Andean landraces in the core collection, while it appeared in all populations of the wild-weed-crop complex in Peru. Thus it is possible that it has been introgressed from wild populations together with alleles of wild bean in this region. It is also interesting that C phaseolin was detected at a higher frequency in secondary centre germplasm in the core collection (11 out of 228 with Andean phaseolin types, or 4.8%). Brown et al. (1982) reported about 6% C phaseolin in cultivars in the United States, while Gepts et al. (1986) found 21% C type in a sample of South American varieties, although largely in Chilean landraces. If C phaseolin has its origin close to Cuzco, where the wild-weed-crop complex was observed, this might have been a centre of distribution of germplasm to other parts of the world during the colonial period, leading to a higher proportion of C phaseolin in those secondary centres. However, even if C phaseolin has entered the cultivated common bean gene pool through the wild-weed-crop complex, the wild populations of southern Peru still fall in the greater Andean gene pool of wild bean (Tohme et al. 1996). Thus, the C phaseolin landraces deserve more study but should not be considered a separate gene pool at this stage.

The analysis of different numbers and combinations of attributes allowed an understanding of the effects of different traits and different kinds of data on the final results. Among qualitative traits that are believed to reflect ancestral origin, protein type consistently contributed the most to the discrimination measures. The inclusion of quantitative data for pod and seed attributes altered the results significantly in the sense that no North Andean pool could be discerned. The situation described in this paper regarding the effect of human selected traits on the results of analysis is likely to apply to other species as well. Germplasm experts who are analyzing diversity should therefore consider carefully what sorts of questions they wish to answer by the statistical analysis of phenotypic data, and choose their data accordingly. The statistical procedures utilized in this study could be of utility in the analysis of other species.

In conclusion, qualitative data on the common bean core collection support the hypothesis that at least one additional gene pool (in addition to the two major ones) exists in the North Andes of South America. The North Andean gene pool is visualized from the analysis of protein types alone. A third gene pool was again observed when plant categorical data were analyzed together with protein types. When pod and seed characters were analyzed in the quantitative analysis, the major gene pools dominated the results, although CH phaseolin accessions continued to appear as outliers in the mixed data analysis. In a companion paper (Islam et al. 2001) the practical implications of a North Andean gene pool are examined in terms of useful traits that it might possess.

Acknowledgements

This research was financially supported by a grant from the Australian Centre for International Agricultural Research (ACIAR), Canberra, ACT. We thank the editor and the reviewers for their constructive comments which allowed us to improve this paper considerably.

References

Becerra Velásquez L. and Gepts P. 1994. RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin. Genome 37: 256–263.

Beebe S., Toro O., Gonzalez A.V., Chacón M.I. and Debouck D.G. 1997. Wild weedy-crop complexes of common bean (*Phaseolus vulgaris* L., Fabaceae) in the Andes of Peru and Colombia, and their implications for conservation and breeding. Genet. Resour. Crop Evol. 44: 73–91.

Brown J.W.S., McFerson J.R., Bliss F.A. and Hall T.C. 1982. Genetic divergence among commercial classes of *Phaseolus vulgaris* in relation to phaseolin pattern. Hort. Sci. 17: 752–754.

Carbonel E.A. 1995. Statistical Methods in Plant Breeding: Quantification of Genetic Variability for Germplasm Conservation. Simposio Internacional de Estadistica, Estadistica en Agricultura y Medio Ambiente. Santa Marta, Colombia, 11–15 de junio de 1995.

Carey P.D., Preston C.D., Hill M.O., Usher M.B. and Wright S.M. 1995. An environmentally defined biogeographical zonation of Scotland designed to reflect species distributions. J. Ecol. 83: 833–845.

Chacón S.M.I., González A.V., Gutiérrez J.P., Beebe S.E. and Debouck D.G. 1996. Increased evidence for common bean (*Phaseolus vulgaris* L.) domesticated in Colombia. Ann. Rep. Bean Improv. Coop. 39: 201–202.

Debouck D.G., Toro O., Paredes O.M., Johnson W.C. and Gepts P. 1993. Genetic diversity and ecological distribution of *Phaseolus vulgaris* (Fabaceae) in northwestern South America. Econ. Bot. 47: 408–423.

Gabriel K.R. 1971. The biplot-graphical display of matrices with application to principal components analysis. Biometrika 58: 453–467.

Gatehouse A.M.R., Dewey F.M., Dove J., Fenton H.A. and Pusztai A. 1984. Effects of seed lectins from *Phaseolus vulgaris* on the development of larvae of *Callosobruchus maculatus*; mechanism of toxicity. J. Sci. Food Agric. 35: 373–380.

Gepts P. 1988. A middle American and an Andean common bean gene pool. In: Gepts P. (ed.), Genetic Resources of Phaseolus beans. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 375–390.

Gepts P. 1993. The use of molecular and biochemical markers in crop evolution studies. Evol. Biol. 27 M. K. Hecht (Ed.): 51–94.

Gepts P. 1998. Origin and evolution of common bean: Past events and recent trends. Hor. Sci. 33: 1124–1130.

Gepts P. and Bliss F.A. 1986. Phaseolin variability among wild and cultivated common Beans (*Phaseolus vulgaris*) from Colombia. Econ. Bot. 40: 469–478.

Gepts P., Osborn T.C., Rashka K. and Bliss F.A. 1986. Phaseolin protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris* L.): evidence for multiple centres of domestication. Econ. Bot. 40: 451–468.

Gifi A. 1981. Non Linear Multivariate Analysis. Department of data Theory, University of Leiden.

Gifi A. 1990. Non Linear Multivariate Analysis. John Wiley and Sons, Chichester, pp. 151–192.

Hill M.O. 1991. Patterns of species distribution in Britain elucidated by canonical correspondence analysis. J. Biogeogr. 18: 247–255.

Ishimoto M., Suzuki K., Iwanaga M. and Kikuchi F. 1995b. Variation of seed α-amylase inhibitor in the common bean. Theor. Appl. Genet. 90: 425–429.

Ishimoto M., Yamada T. and Kaga A. 1995a. Insecticidal activity of an alpha-amylase inhibitor-like protein resembling a putative precursor of alpha-amylase inhibitor in the common bean, *Phaseolus vulgaris* L. Biochim. Biophys. Acta 1432: 104–112.

Islam F.M.A. 2001. Evaluation and Utilization of Common Bean Germplasm Collections, PhD, The University of Queensland, Australia.

Islam F.M.A., Basford K.E., Jara C., Redden R.J. and Beebe S. 2001. Seed compositional and disease resistance differences among gene pools in cultivated common bean. Genet. Resour. Crop. Evol. (in press).

Jolliffe I.T. 1986. Principal Component Analysis. Springer-Verlag, New York.

Jongman R.H.G., ter Braak C.J.F. and van Tongeren O.F.R. 1987. Data Analysis in Community and Landscape Ecology. Pudoc, Wageningen.

Kami J., Becerra Velásquez L., Debouck D.G. and Gepts P. 1986. Identification of present ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. Proceedings of the National Academy of Science. USA 92: 1101–1104.

Koenig R. and Gepts P. 1989. Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centres of genetic diversity. Theor. Appl. Genet. 78: 809–817.

Lin J.J. and Kuo J. 1995. AFLP™: A Novel PCR-based assay for plant and bacterial DNA fingerprinting. FOCUS 17: 66–70.

LKB GELSCAN 1986. LKB 2400 GELSCAN XL™ software. LKB produkter AB, Box 305, S-161 26Bromma, Sweden.

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

Ma Y. and Bliss F.A. 1978. Seed proteins of common bean. Crop Sci. 17: 431.

Manly B.F.J. 1994. Multivariate Statistical Methods. 2nd edn. Chapman and Hall, New York, pp. 215.

Nodari R.O., Koinange E.M.K., Kelly J.D. and Gepts P. 1992. Towards an integrated linkage map of common bean. 1. Development of genomic DNA probes and levels of restriction fragment length polymorphism. Theor. Appl. Genet. 84: 186–192.

Osborn T.C., Manen J.F., Brown J.W.S. and Bliss F.A. 1984. Bean lectins 4: genetic variation in the non-denatured structure of lectins from from different *Phaseolus vulgaris* L. cultivars. Theor. Appl. Genet. 67: 547–552.

Osborn T.C., Alexander D.C., Sun S.S.M., Cardona C. and Bliss F.A. 1988. Insecticidal activity and lectin homology of arcelin seed protein. Science 240: 207–210.

SAS, Institute Inc. 1989. SAS/STAT* User's Guide. Version 6. 4th edn. SAS Institute Inc., Cary, North Carolina. 1.

Singh S.P. 1989. Patterns of variation in cultivated common bean. Econ. Bot. 43: 39–57.

Singh S.P., Gepts P. and Debouck D.G. 1991a. Races of common bean (*Phaseolus vulgaris*, Fabaceae). Econ. Bot. 45: 379–396.

Singh S.P., Nodari R. and Gepts P. 1991b. Genetic diversity in cultivated common bean: I. Allozymes. Crop Sci. 31: 19–23.

SPSS for Windows (1999). Copyright©SPSS Inc. 1989–1999.

S-PLUS Inc. 1995. StatSci Division MathSoft, Inc. Seattle, Washington.

Tohme J., Jones P., Beebe S. and Iwanaga M. 1995. The combined use of agroecological and characterisation data to establish the CIAT *Phaseolus vulgaris* core collection. In: Hodgkin T., Brown A.H.D., van Hintum Th.J.L. and Morales E.A.V. (eds), Core Collections of Plant Genetic Resources. International Plant Genetic Resources Institute (IPGRI) A Wiley-Sayce Publication, pp. 95–107.

Tohme J., Gonzalez D.O., Beebe S. and Duque M.C. 1996. AFLP analysis of gene pools of a wild bean core collection. Crop Sci. 36: 1375–1384.

Toro O., Tohme J. and Debouck D.G. 1990. Wild bean (*Phaseolus vulgaris* L.): description and distribution. Centro Internacional de Agricultura Tropical, Cali, Colombia.

van Schoonhoven A. and Pastor-Corrales M.A. 1987. Standard system for the evaluation of bean germplasm. Centro Internacional de Agricultura Tropical, Cali, Colombia.

van Schoonhoven A. and Voysest O. 1991. Common Beans. Research for Crop Improvement. C. A. B International, Wallingford, Oxon., OX10 8DE, UK.

Williams W.T. 1976. Pattern Analysis in Agricultural Science. Elsevier, New York.

Ward J.H. 1963. Hierarchical grouping to optimize an objective function. J. Amer. Stat. Assoc. 58: 236–244.