PN-ABM-7219. 241
79877

FINAL REPORT

Covering period: 9/01/89-8/31/92

Submitted to the Office of the Science Advisor
U.S. Agency for International Development

Establishment of a "core collection" for common bean (*Phaseolus vulgaris* L.) genetic resources conservation

Principal Investigator: Paul Gepts
Grantee Institution: University of California, Davis

Collaborators: S.P. Singh, D.G. Debouck, W.M. Roca
Institution: CIAT, Cali, Colombia

Project Number: 9.241
Grant Number: AID DHR-5542-G-SS-9020-00

A.I.D. Grant Project Officer:
Harvey Hortik

Project Duration: 9/01/89 - 8/31/92

REC'D IN R&D/R

NOV 5 1992

cc: CD/E+BSTID

TABLE OF CONTENTS

EXECUTIVE SUMMARY 1

FINAL REPORT 2
- **I. Research Objectives:** 2
- **II. Methods and Results** 3
  - *Objective 1. Identification of molecular markers (RFLPs) that allow us to determine patterns of diversity in common bean* 3
  - *Objective 2. Study of patterns of diversity using molecular and morph-agronomic traits and establish groups of related genotypes ("races" or "gene pools")* 4
  - *Objective 3. Germplasm exploration in southern Colombia and Ecuador* 9
  - *Objective 4. Test methodology to identify markers without the use of radioactively labeled probes* 10
  - *Objective 5. Develop a molecular linkage map for common bean* 10
  - *Objective 6: Mapping of the I gene for BCMV resistance and QTLs for Rhizobium nodule number and resistance to common bacterial blight* 13
- **III. Impact, Relevance, and Technology Transfer** 15
  - **A. Additional Knowledge** 15
  - **B. Training of Individuals from Third World Countries** 16
  - **C. Materials Produced** 17
- **IV. Project Activities and Output** 17
  - **A. Meetings attended and Invited Presentations** 17
  - **B. Publications** 19
  - **C. Training** 20
- **V. Project Productivity** 20
- **VI. Future Work** 21

EXECUTIVE SUMMARY

In the last 20 years, large collections of common bean (*Phaseolus vulgaris*) diversity have been assembled in gene banks, notably by the USDA and CIAT in Colombia. While the genetic diversity contained in these collections represent a long-term investment for bean improvement, their sheer size (up to 45,000 entries at CIAT) makes them less amenable to short-term use by breeders. The purpose of this study is to devise a strategy to identify a subsample of 5-10 % of the total number of entries (i.e. a "core" collection) that is representative of the diversity contained in the total collection. Our guiding principle has been that the core collection should include the various evolutionary lineages contained in the species; these lineages can best be identified using biotechnological tools such as molecular markers: only those common bean lines that exhibit different markers and, hence, have a different ancestor, should be included in the core collection.

Our results over the three years can be summarized as follows: 1) We have been able to identify very high frequencies of differences for molecular markers. These frequencies are higher than reported previously in common bean and will considerably facilitate subsequent work in genetic diversity conservation and breeding at CIAT and in U.S. programs; 2) We have identified suitable statistical tools to relate molecular and agronomic (i.e. disease resistance, adaptation, growth habit, etc.) variation; using these tools, we have been able to identify gene pools and races that can be distinguished at the molecular as well as agronomic levels; 3) We have identified wild *P. vulgaris* populations in Ecuador where they had never been reported before; analyses show these populations to be intermediate between the Mesoamerican and Andean subspecies of common bean and may help us understand the process of gene transfer between the two subspecies; 4) Analyses of cultivated materials show that within the same region, Andean cultivars are grown at higher altitudes and therefore appear to be better adapted to cooler conditions than their Middle American counterparts. 5) We have established a molecular linkage map based on genomic probes established during the first year of the project. The high levels of polymorphisms allowed us to establish this map in no more than a year. 6) We have started locating genes for agronomic traits on this molecular linkage map. The *I* gene for Bean Common Mosaic Virus resistance, which is currently introduced into most advanced cultivars, was located on the map as were 4 quantitative trait loci (QTLs) each for *Rhizobium* nodule number and for *Xanthomonas campestris* resistance. Transfer of technology from UCD to CIAT has included: 1) knowledge of the gene pool and racial organization of genetic diversity in common bean; 2) transfer of methodology for molecular marker analysis in common bean; 3) transfer of the *Pst*I genomic library used in mapping; 4) training of a member of the CIAT Biotechnology Unit in RAPD methodology; 5) training of an Ecuadoran scientist in germplasm exploration; and 6) identification of regions in Ecuador where wild bean populations subsist.

FINAL REPORT

I. Research Objectives:

The major thrust of this project was to clarify genetic relationships within common bean, *Phaseolus vulgaris*, in order to expedite the establishment of a "core collection." Core collections have been proposed by Frankel (1984) and Brown (1989) as a means towards improving the access to and utilization of germplasm collections, especially very large ones such as the bean collection at CIAT which contains over 40,000 entries. These collections would consist of some 5-10% of the main collection and would be characterized in priority over the main collection. The increased access and utilization would result primarily from this improved characterization.

The establishment of such a collection requires sound sampling procedures so that the diversity within the core collection can be maximized. This requires a knowledge of patterns of diversity within the species to determine which segments of the species carry the most diversity or unusual diversity. Molecular markers are particularly appropriate to characterize genetic diversity because they are genotypic traits and thus directly reflect changes at the DNA level. Genetic distances among entries of a bank established on the basis of molecular markers are more accurate than those based on phenotypic (morpho-agronomic) trait measurements.

Hence, our first first two objectives were as follows:

Objective 1. Identification of molecular markers (RFLPs) that allow us to determine patterns of diversity in common bean

Objective 2. Study of patterns of diversity using molecular and morph-agronomic traits and establish groups of related genotypes ("races" or "gene pools").

Preliminary observation prior to this proposal had indicated that wild beans from northern Peru could be potential intermediate forms between the Middle American and Andean gene pools of common bean. However, no wild common bean from neighboring Ecuador had ever been reported. Because of the potential importance of these intermediate forms as potential bridging forms and to elucidate the origin of common bean, we proposed an exploration in Ecuador and southern Colombia in search of wild common bean.

Objective 3. Germplasm exploration in Ecuador and southern Colombia to discover unreported wild common bean populations

Because RFLP technology is cumbersome and generally requires the use of radioisotopes, it is usually impractical in developing countries. Alternative strategies that

obviate the use of radioactivity should be pursued.

Objective 4. Test methodology to identify markers without the use of radioactively labeled probes.

Having completed our objectives, we turned our attention to the application of markers we had identified to the mapping of the common bean genome. We established a molecular linkage map and with this map were able to locate the first disease resistance genes and genes for *Rhizobium* nodulation. The availability of map information

Objective 5. Develop a molecular linkage map for common bean

Objective 6. Locate genes for resistance to Bean Common Mosaic Virus, common bacterial blight (Xanthomonas campestris pv. phaseoli), and nodulation by Rhizobium tropici.

II. Methods and Results

Objective 1. Identification of molecular markers (RFLPs) that allow us to determine patterns of diversity in common bean

Scientists involved: R. Nodari (UCD graduate student from Brazil), E. Koinange (UCD graduate student from Tanzania), V. Becerra (UCD graduate student from Chile), P. Gepts (UCD)

Two genomic libraries were obtained, one obtained by digestion of bean DNA with *Eco*RI and *Bam*HI and the other with *Pst*I. These two libraries showed contrasting characteristics with regard to single copy sequence frequency and levels of polymorphisms (Table 1). While the *Pst*I library, as expected, exhibited a much higher frequency of single copy sequences, it also, unexpectedly, exhibited a very high frequency of polymorphisms. That this is not some experimental fluke was demonstrated by the presence of high levels of polymorphisms in several populations. In addition to the BAT 93 x Jalo EEP558 and Rio Tibagi x ABA 71 populations, we obtained even higher levels (approx. 90 %) in a third population also involving a Mesoamerican and an Andean genotype. Polymorphisms within the Andean or Mesoamerican groups, while lower than those between these groups, were still quite high (Table 2). These high levels of DNA polymorphisms were one of the highlights of our research in this grant because they facilitated our characterization of genetic diversity in *Phaseolus vulgaris* germplasm and, in collaboration with other groups, speeded up the mapping of the bean genome.

Publications:

Nodari RO, Koinange EMK, Kelly JD, 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

TABLE 1: COMPARISON BETWEEN THE *Eco*RI-*Bam*HI AND *Pst*I GENOMIC LIBRARIES

	*Eco*RI-*Bam*HI	*Pst*I
PROPORTION OF SINGLE COPY SEQUENCES	102/484 (21%)	381/426 (89%)
PROPORTION OF SINGLE COPY POLYMORPHIC CLONES: POP. A[1]	18/57 (32%)	50/61 (82%)
POP. B[2]	0/7 (0%)	23/28 (82%)

[1] BAT 93 x JALO EEP-558 [2] RIO TIBAGI x ABA 71

Gepts P,Llaca V, Nodari RO, Panella L. 1992. Analysis of seed proteins, isozymes, and RFLPs for genetic and evolutionary studies in Phaseolus. *In*: Linskens HF,Jackson JF Ieds), Seed analysis, Modern Methods of Plant Analysis, Vol. 14. Springer, Berlin: pp. 63-93.

Objective 2. Study of patterns of diversity using molecular and morph-agronomic traits and establish groups of related genotypes ("races" or "gene pools").

Objective 2a: Characterization of common bean germplasm with seed protein and isozyme electrophoresis: Scientists involved: S.P. Singh (CIAT, on leave at UCD); P. Gepts (UCD)

Two-hundred and twenty seven landraces of common bean landraces were analyzed for phaseolin and allozyme diversity. Nine polymorphic loci were observed all of which were unlinked to the phaseolin locus. This total of ten loci covered, hence, some 500-600 cM of the total bean genome. Although the precise size of the common bean genome is not known, figures from other crop plants suggest that this represented some 30 % of the common bean genome. A cluster analysis based on Nei's genetic distance (D) was performed according to the unweighted paired group method of Sneath and Sokal. Our results confirmed the existence of two major groups in cultivated common bean, Mesoamerican versus Andean American, confirming the hypothesis of multiple domestications; provided indications of gene flow from wild to cultivated beans; and suggested at least five subgroups within Mesoamerican and four within Andean

TABLE 2 INTER vs. INTRA "SUBSPECIES" RFLP LEVELS

GENOTYPES: ANDEAN: JALO EEP 558, ABA 71, MIDAS
MESOAMERICAN: BAT 93, G12873, SIERRA, LEF-2RB, AC1028

ENZYMES: *Bam*HI, *Eco*RI, *Eco*RV, *Hind*III, *Pst*I, *Xba*I

COMPARISON	NO. OF PROBES	NO. OF ENZYMES	RFLP LEVEL (%)
	INTER		
B x J	57	6	86
	61	3	82
M x G	61	3	89
	INTRA		
MESOAMERICAN	53	6	62
ANDEAN	57	6	49

cultivar groups.

Publication: Singh, S.P., Nodari, R. Gepts, P. 1991. Genetic diversity in cultivated Phaseolus vulgaris. I. Allozymes. Crop Sci. 31:19-23

Objective 2b: RFLP characterization of genetic diversity in wild and cultivated common bean Scientists involved: V. Becerra and P. Gepts (UCD)

Using probes selected from the *Pst*I library, eighty-five accessions of common bean (*Phaseolus vulgaris* L.), representing a wide geographic area of distribution from Mexico to Argentina, were tested for restriction fragment length polymorphisms (RFLPs) to: 1) determine the organization of the genetic diversity of the species, both among and within genotypes, and 2) to compare the results with isozyme and phaseolin markers

previously studied. Accessions were classified into two major groups with a distinct geographical distribution; Middle American genotypes and Andean genotypes. Within each gene pool, cultivated accessions clustered together with wild forms from the same geographical area. Estimates of Nei's genetic distances, among the cultivated races from the two different gene pools varied from 0.13 to 0.39; however, genetic distances within races from the same gene pool were smaller, and ranged from 0.04 to 0.11. Genetic distances obtained by RFLP analysis tended to be higher than those obtained by isozyme analysis. A highly significant positive correlation was obtained between Nei's distances from RFLPs and isozyme data (0.82).

Table 3. Total (Ht), within (Hs) and between (Dst) genetic diversity in common bean (*Phaseolus vulgaris* L.) detected by RFLP and isozyme markers.

Marker	Ht	Hs	Dst
Isozyme[a]	0.19	0.06	0.13
RFLP	0.34	0.24	0.11

[a] Data of Koenig and Gepts (1989), Singh et al. (1991c)

The level of genetic diversity (Ht= 0.34) was twice the value obtained with isozyme analysis (Table 3). Genetic diversity within races (Hs= 0.24) was at least 3-4 times higher compared to isozymes. On the other hand, genetic diversity between races (Dst= 0.11) was similar for both categories of markers. Average heterozygosity showed some reduction in the levels of diversity in cultivated genotypes compared to wild genotypes within the two gene pools. In addition, average heterozygosity values obtained from RFLP data were 3-4 times higher than those obtained from isozyme data (Table 4).

These results corroborate parallel studies on the characterization of genetic diversity in common bean, based on phaseolin, isozyme, and morphological data, that clearly show the existence of two distinct gene pools, Middle American and Andean. Moreover, RFLP markers are superior to isozymes when they are used for genetic variability studies because they provided better coverage of the genome and revealed higher level of polymorphisms.

Table 4. Trends in genetic diversity in *Phaseolus vulgaris* based on isozyme and RFLPs (Koenig and Gepts 1989; Singh et al. 1991b; V. Becerra 1992

			Average heterozygosity	
		N	Isozymes[a]	RFLPS[b]
Middle American	Wild	11	0.13	0.33
	Cultivated	36	0.09	0.27
Andean	Wild	11	0.07	0.25
	Cultivated	26	0.03	0.27

[a] 19 loci; [b] 13 probes for nuclear sequences, 3 restriction enzymes

Publications:
Becerra Velasquez VL.1992. Characterization of the genetic diversity in common beans (Phaseolus vulgaris L.) using RFLP markers. MS thesis, University of California, Davis.

Objective 2c: Morphoagronomic characterization of common bean landraces: Scientists involved: S.P. Singh (CIAT), P. Gepts (UCD)

Three hundred and six landraces representative of the geographic, ecological and morphological diversity of common bean in Latin America were evaluated for morphological and agronomic traits. Data on pigmentation, growth habit, and leaflet, pod, seed, and phenology traits, and reaction to four important diseases and an insect pest (Table 5), obtained from field evaluations at three environmentally contrasted locations in Colombia during the 1987-1988 cropping season, were analyzed by multivariate statistical analyses. Results permitted separation of these landraces into Mesoamerican and Andean groups, confirming prior phaseolin and allozyme data. A marker-based multivariate analysis, using phaseolin or allozymes as an initial classification criterion (see next section), followed by a corroborating analysis of morpho-agronomic traits, suggested the existence of subgroups within each of the major Andean and Mesoamerican groups, with distinctive morphology, adaptation, and disease resistances. Based on this information, a classification system is being proposed for the cultivated materials of *P. vulgaris* which includes three races in both Mesoamerican and Andean materials.

TABLE 5. MORPHOLOGICAL AND AGRONOMIC TRAITS USED TO CHARACTERIZE GENETIC DIVERSITY IN COMMON BEAN

PIGMENTATION:	GROWTH HABIT:
HYPOCOTYL COLOR	CIAT CLASSIFICATION
FLOWER COLOR	LENGTH 5TH INTERNODE
	NODES TO FIRST FLOWER (STEM)
LEAFLET:	NODES TO FIRST FLOWER (BRANCH 1)
LENGTH	NODES TO FIRST FLOWER (BRANCH 2)
WIDTH	
ELONGATION (L/W)	SEED:
	LENGTH
POD:	HEIGHT
LENGTH	WIDTH
WIDTH	ELONGATION (L/W)
NUMBER/PLANT	FLATNESS (H/W)
	SIZE
PHENOLOGY:	NUMBER/PLANT
DAYS TO FLOWERING	YIELD/PLANT
DAYS TO MATURITY	

DISEASE AND PEST RESISTANCE:
BEAN COMMON MOSAIC VIRUS (*I*)
EMPOASCA LEAFHOPPER
ANTHRACNOSE: LEAF
ANTHRACNOSE: POD
COMMON BACTERIAL BLIGHT (1ST TRIFOLIOLATE)
COMMON BACTERIAL BLIGHT (ASPERSION)
ANGULAR LEAFSPOT

Publications:

Singh, S.P., Gutiérrez, A., Molina, A., Urrea, C., Gepts, P. 1991. Genetic diversity in cultivated *Phaseolus vulgaris*. II. Marker-based analysis of morpho-agronomic diversity. Crop Sci. 31:23-29

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

Gepts P. 1992. The use of molecular and biochemical markers in crop evolution studies. Evol. Biol., in press

Objective 3. Germplasm exploration in southern Colombia and Ecuador

Objective 3a: Germplasm exploration
Because of scheduling conflicts and an unusual drought, the exploration was divided into two parts. The first part took place from June 13 to July 2, 1990, in Ecuador; the team consisted of Dr. D. Debouck (IBPGR), N. Díaz (Ecuadoran national with INIAP, Ecuador), A. Maquet (*P. lunatus* specialist at CIAT), O. Toro (Colombian national and germplasm introduction officer at CIAT), and P. Gepts. The second part took place from August 23 to September 8, 1990, in Colombia; the team consisted of Daniel Debouck (IBPGR), L. López (Colombian national with IBPGR), O. Toro (CIAT), and Néstor Angulo (Colombian national with ICA, Colombia).

Forty-one samples were collected in Ecuador: *P. lunatus*: wild or escaped: 21, cultivated: 10; *P. vulgaris*: wild: 4, *P. coccineus-polyanthus*: 2; *P. "rosei"*: 1; *P. augusti*: 1; *Vigna hookeri*: 2. Eighty-nine samples were collected in Colombia: *P. lunatus*: cultivated: 26; *P. vulgaris*: wild: 3, cultivated: 38; *P. coccineus* subsp. *polyanthus*: 3; *Vigna candida*: 7; *V. venusta*: 6; *V. antillana*: 2; *V. adenantha*: 1; *V. lasiocarpa*: 1; *V. vexillata*: 1; and *Ciphomandra betacea*: wild: 1.

With the exception of two wild *P. lunatus* accessions, it was possible to secure mature seeds. In addition, for all the accessions complete passport data were obtained and herbarium specimens were obtained for all the wild materials. Half of the seeds harvested were left with INIAP (Ecuador) and ICA (Colombia). The remainder of the seed is currently at CIAT where it will be increased and sent to the USDA for addition to the U.S. collection.

One of the *P. vulgaris* sites (Huigra, Ecuador) was sampled extensively and seeds of the places in the site were maintained separate. This population will be analyzed collaboratively between the Genetic Resources Unit of CIAT and UCD to examine intra-population diversity and the concept of accession. One of the *P. coccineus* populations is actually a hybrid between *coccineus* and *polyanthus*. This population will be studied at Davis to further examine speciation and introgression in the *P. coccineus* complex.

We have been fortunate to receive full support from CIAT for this activity and particularly from Dr. M. Iwanaga who has been very cooperative and has greatly facilitated our work.

Objective 3b: Isozyme and phaseolin diversity of wild *Phaseolus vulgaris* of Colombia and Ecuador: Scientists involved: D. Debouck (IBPGR), O.Toro (CIAT), M. Paredes and P. Gepts (UCD)

Wild populations of this region were of interest because previous isozyme analyses had shown that the transition between the Middle American and Andean genotypes of *P. vulgaris* occurred somewhere between northern Colombia and Southern Peru. Until

recently, however, there was a dearth if not a complete absence of wild materials from this region. Our exploration demonstrated the existence of wild *P. vulgaris* in Ecuador where it had not been reported before suggesting that our knowledge of the biological diversity in that region could certainly be improved. Isozyme analyses demonstrated that these accessions were intermediate between the Andean and Middle American wild *P. vulgaris* accessions (Fig. 1).

The intermediate nature resulted from the presence of either Middle American or Andean alleles depending on the locus. It is not clear at this stage what the cause is of this intermediate nature. It could result from hybridization between Middle American and Andean genotypes. Alternatively, the southern Colombia-Ecuador-northern Peru region could reprsent the area of origin of the species from which the Middle American and Andean groups have radiated.

From a practical standpoint, the existence of intermediate genotypes suggests that it should be possible to obtain viable and fertile recombinants with sufficient fitness in crosses between Middle American and Andean genotypes. These crosses have represented a challenge for breeders so far but may in the long run represent a source of novel variability.

Publications: Debouck DG , Toro O, Paredes O, Johnson W, and Gepts P. Genetic Diversity and Ecological Distribution of *Phaseolus vulgaris* in Northwestern South America. In preparation.

Objective 4. Test methodology to identify markers without the use of radioactively labeled probes.

Originally this objective was to identify non-radioactive labeling methods for Southern hybridization. Our experience as well as the introduction of the PCR method has led us to revise this objective. Instead, we have attempted to use PCR to distinguish genotypes. Specifically, we have designed primers amplifying a portion of the phaseolin locus and were able to distinguish between Middle American and Andean materials, *e.g.*, between 'S' phaseolin materials on one hand and 'T' or 'C' materials on the other (Fig. 2). This technique was performed on DNA extracted directly from seed (instead of leaves) and required low amounts of tissue. We feel that the rapidity of the PCR analysis as well as the direct seed analyses offer advantages over other markers.

Objective 5. Develop a molecular linkage map for common bean: Scientists involved: P. Gepts (UCD), S. Singh (CIAT)

The high levels of polymorphism observed with the *Pst*I genomic clones developed during the first year of this project suggested that rapid progress could be made in developing a molecular linkage map for common bean. This was indeed the case as we were able in

about one year to develop a Restriction Fragment Length Polymorphism (RFLP)-based linkage map for common bean (*Phaseolus vulgaris* L.) covering 327 centiMorgans (cM) based on a F_2 mapping population derived from a cross between BAT93 and Jalo EEP558 (Fig. 3). The parental genotypes were chosen because they exhibited differences in evolutionary origin, allozymes, phaseolin type, and for several agronomic traits. Segregation of 152 markers was analyzed, including 115 RFLP loci, 7 isozyme loci, 8 Random Amplified Polymorphic DNA (RAPD) marker loci, 19 loci corresponding to 15 clones of known genes, 1 virus resistance gene, 1 flower color gene, and 1 seed color pattern gene. Using MAPMAKER and LINKAGE-1, 143 markers could be assigned to

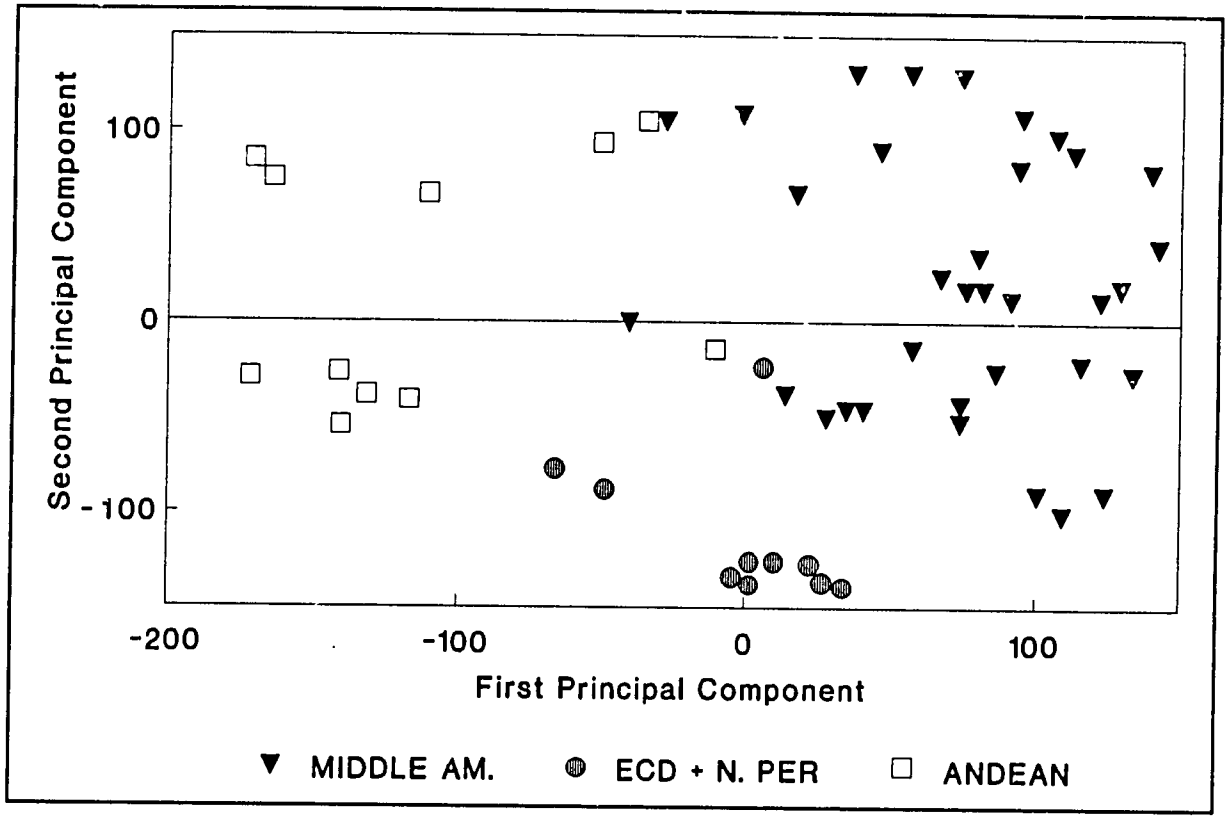


Figure 1. Principal component analysis of isozyme diversity in wild *Phaseolus vulgaris*. Middle Am. = Middle America; ECD = Ecuador; N. PER = Northern Peru

15 linkage groups, whereas 9 markers remained unassigned. The average interval between markers was 6.5 cM; only one interval was larger than 30 cM. A small fraction (9%) of the markers deviated significantly from the expected Mendelian ratios (1:2:1 or 3:1) and mapped into four clusters. Probes of known genes belonged to three categories: seed proteins, pathogen response genes, and *Rhizobium* response genes. Within each category, sequences homologous to the various probes were unlinked.

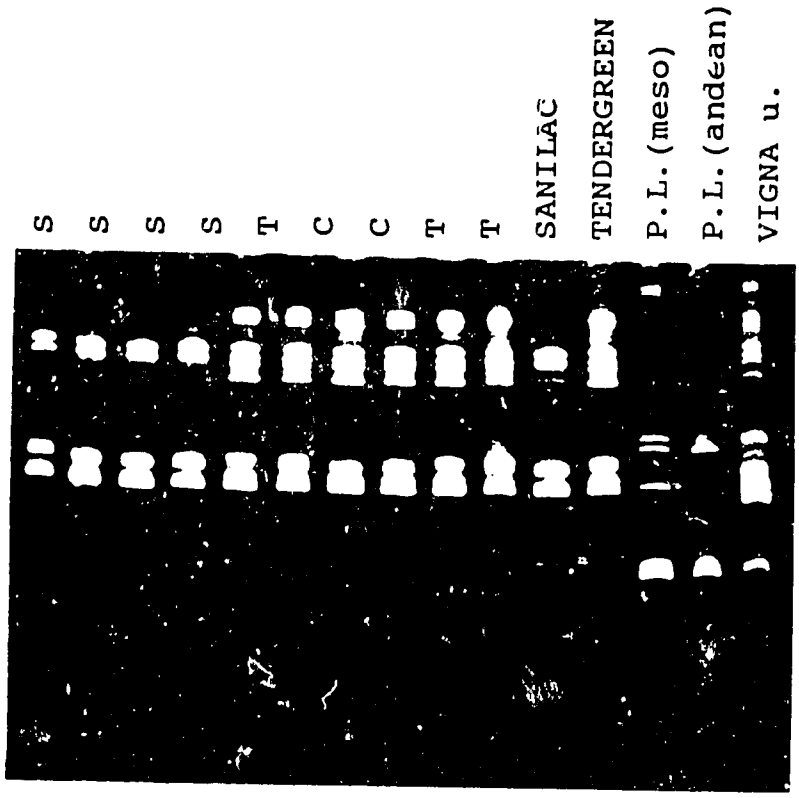


Figure 2. PCR amplification of a portion of the phaseolin locus in selected genotypes of Middle American ('S') or Andean ('T' or 'C') origin in *Phaseolus vulgaris* and related species

Publications:

Nodari RO, Tsai SM, Gilbertson RL, Gepts P. 1992. Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. Theor. Appl. Genet., in press.

Gepts P. 1992. Linkage map of common bean (Phaseolus vulgaris L.). In: O'Brien SJ (ed), Genetic Maps. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, in press

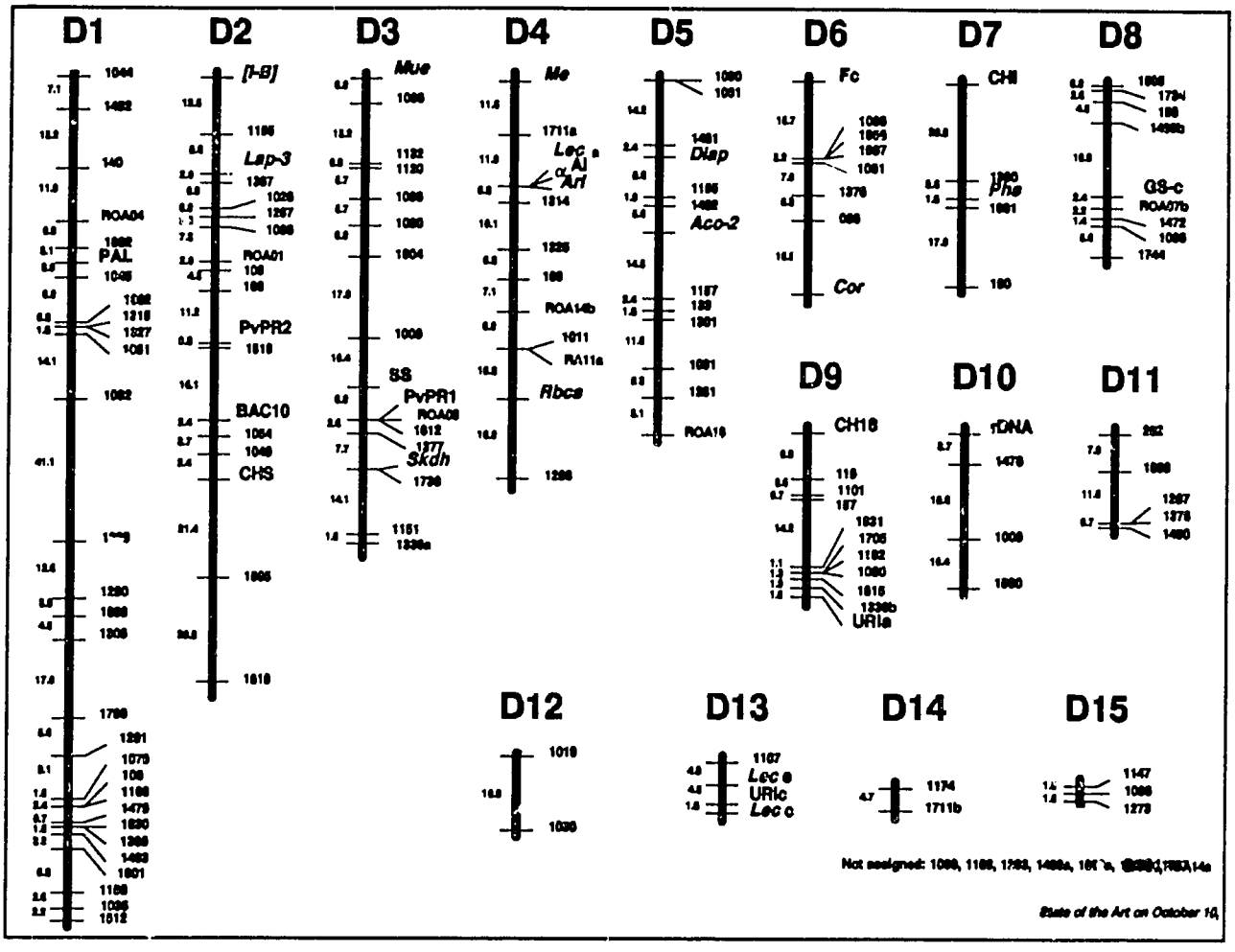


Figure 3. RFLP-based linkage map of common bean

Objective 6: Mapping of the I gene for BCMV resistance and QTLs for Rhizobium nodule number and resistance to common bacterial blight. Scientists involved: P. Gepts and R. Gilbertson (UCD), S.M. Tsai (CENA, Brasil)

The *I* gene for bean common mosaic virus resistance is a very important gene in bean improvement because it confers resistance to all known non-necrotic strains of the disease. The gene is now being incorporated in all improved bean genotypes. Linkage experiments showed that this gene is located at one end of linkage group D2 (Fig. 3). The closest marker is located at 12.5cM suggesting that additional efforts should be spent to identify more closely linked markers for indirect selection purposes. The *I* gene is the first disease resistance gene to be located on the common bean genetic linkage map.

RFLP-based genetic linkage maps allow us to dissect the genetic control of quantitative traits (QT) by locating individual quantitative trait loci (QTLs) on the linkage map and determing their type of gene action and the magnitude of their contribution to the

phenotype of the QT. We have performed such an analysis for the interaction between the plant host and the *Rhizobium* symbiont, specifically nodule number (NN), and between the host plant and *Xanthomonas campestris* pv. *phaseoli* (resistance to common bacterial blight). Analyses were conducted in the progeny of a cross between BAT93 (fewer nodules) and Jalo EEP558 (more nodules). An RFLP-based linkage map for common bean based on 152 markers had previously been derived in the F_2 of this cross (see above). Seventy F_2-derived F_3 families were inoculated in separate greenhouse experiments with *Rhizobium tropici* strain UMR1899 or *Xanthomonas campestris* pv. *phaseoli* isolate W18. Regression and interval mapping analyses were used to identify genomic regions involved in the genetic control of these traits. The two methods identified the same genomic regions for each trait, with a few exceptions. For each trait, at least 4 putative QTLs were identified, which accounted for approximately 50% and 75% of the phenotypic variation in NN and CBB resistance, respectively. A chromosome region on linkage group D7 carried factor(s) influencing both traits. In all other cases, the putative QTLs affecting NN and and CBB were located in different linkage groups or in the same linkage group, but far apart (more than 50 cM). Both BAT93 and Jalo EEP558 contributed alleles associated with higher NN, whereas CBB resistance was always associated with BAT93 alleles. The potential common factor influencing both NN and CBB on linkage group D7 could be involved in the synthesis of flavonoids, which have been shown that have either a stimulatory or deleterious effect on Rhizobia and pathogenic bacteria. In a separate experiment, we have mapped the *P* gene controlling the presence of flavonoids to the same region on linkage group D7. Identification of these QTLs raises the possibility of initiating map-based cloning and marker-assisted selection for these traits.

Publications:

Nodari RO, Tsai SM, Gilbertson RL, Gepts P. 1992. Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. Theor. Appl. Genet., in press.

Nodari RO, Tsai SM, Guzmán P, Gilbertson RL, Gepts P. Towards an Integrated Linkage Map of Common Bean. 3. Mapping Genetic Factors controlling Host-Bacterium Interactions. Submitted.

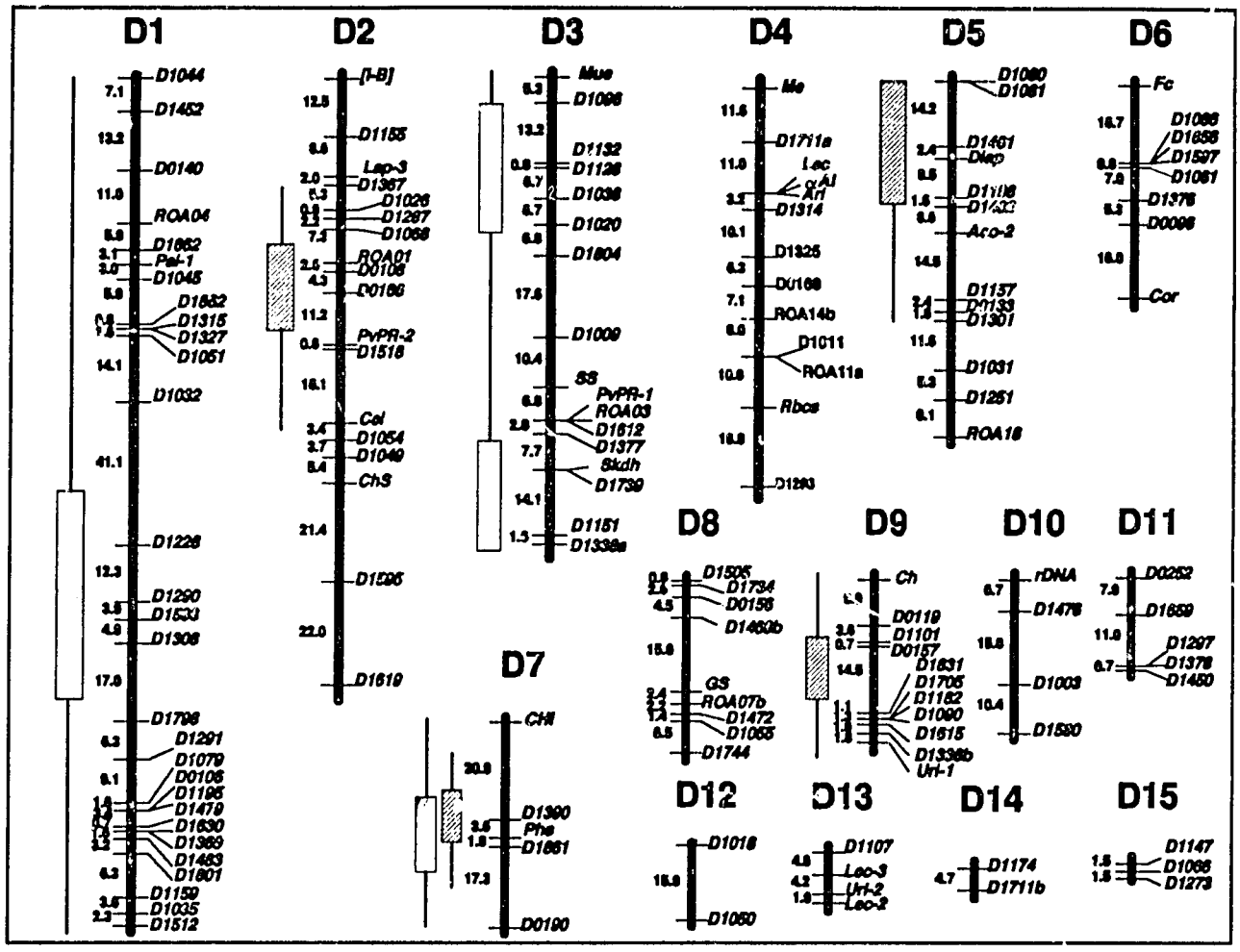


Figure 4. Location of quantitative trait loci influencing *Rhizobium* nodule number (stippled blocks) and resistance to common bacterial blight (hatched blocks) in the cross BAT93 x Jalo EEP558. The blocks indicate the most likely region

III. Impact, Relevance, and Technology Transfer

The project has had a threefold impact: 1) it has generated additonal knowledge about genetic diversity in common bean; 2) it has trained individuals from Third World countries; and 3) it has generated materials (DNA probes and a genetic stock population) for future genetic studies in common bean.

A. Additional Knowledge

1. Gene pool and racial organization of genetic diversity in common bean: This information is having an impact on the working of the *Phaseolus* collection at CIAT (Drs. J. Tohme and M. Iwanaga) and on the breeding program of Dr. S. Singh. Regarding the former, the Genetic Resources and Biotechnology Units of CIAT have

developed a core collection for common bean. Selection criteria for this collection included, among other criteria, information from molecular data generated in this project. This core collection has been presented at a recent meeting organized by IBPGR, CENARGEN, and CNG in Brasilia (August 23-28, 1992). Regarding the latter, membership in races is one of the criteria used by Dr. S. Singh to select particular parents for crosses. He has observed that generally crosses made among genotypes belonging to different races are more productive than those made within a race (*e.g.* Singh et al. 1989).

2. The availability of a genetic linkage map is expected to impact bean breeding in the near future because of enhanced knowledge of the genetic control of important traits. For example, our finding of a major gene for commmon bacterial blight resistance on linkage group D7 is consistent with previous observations that suggested the existence in certain circumstances of a major gene (e.g., Silva et al. 1989). It also paves the way for marker-assisted selection for certain traits difficult to select for, for example, when the screening is too cumbersome or when several genes for disease resistance need to be pyramided.

3. Until 1989, wild common bean had not been reported from Ecuador. Research in this proiect was instrumental in focusing on this country as a potential source of additional germplasm of wild common bean. Succesful explorations were conducted that revealed the existence of wild common bean populations in that country. Members of the Ecuadoran Genetic Resources Unit (Estación Experimental de Sanata Catarina, Quito) have been made aware of the precise locations and have received half of the seed collected. Herbarium specimen have been deposited at the Herbarium of the Catholic University in Quito and elsewhere.

B. Training of Individuals from Third World Countries

Four individuals from Third World countries have been trained in relationship to this project:

Name	Country	Support	Current position
Viviana Becerra	Chile	RA-ship from this project	Will return within 6 months to Chile upon completion of husband's PhD
Rubens Nodari	Brazil	PhD research partly supported by this project	Graduated and has returned to Brazil; currently on the faculty of the Crop Science Department, Universidade Federal de Santa Catarina

Epimaki Koinange	Tanzania	PhD research partly supported by this project	Graduated and is returning to Tanzania; will continue on the staff of the Tanzania Ministry of Agriculture
Dr. S.-M. Tsai	Brazil	Research partly supported by this project	Has returned to Brasil and continues as Head of the Microbiology Section of CENA, Universidade de São Paulo
Ing. Fernando Tenjo	Colombia	One month training supported by this project	Member of the Biotechnology Unit of CIAT and has returned there

C. Materials Produced

Three types of materials have been produced. A genomic library of *Pst*I clones which was transferred in December 1990 to Dr. W. Roca (head of the Biotechnology Unit at CIAT) and a recombinant inbred population (BAT93 x Jalo EEP558 F6) transferred to the Genetic Resources Unit of CIAT. Seeds collected during the germplasm exploration have been divdied equally between the national programs (Eciador and Colombia) and CIAT. After seed increase these have been transferred to the estern Plant Introduction Station in Pullman, WA.

IV. Project Activities and Output

A. Meetings attended and Invited Presentations

1. Bean Improvement Cooperative Meeting, Toronto, November 7-9, 1989: Presentation (with S. Singh, A. Gutiérrez, C. Urrea, R. Koenig, and R. Nodari): Marker-based analysis of genetic diversity in common bean.

2. UCLA Symposium, Molecular Strategies for Crop Improvement, April 16-22, 1990; Poster (with Shree Singh, Rubens Nodari, Belén Garrido, and Epimaki Koinange): Towards an integrated linkage map of common bean (*Phaseolus vulgaris* L.).

3. Washington State University (IAREC), Prosser, WA, May 17, 1990: Invited presentation: The use of molecular markers in genetic studies of common bean, *Phaseolus vulgaris*

4. Colegio de Ingenieros Agrónomos de Loja y Zamora-Chinchipe, Loja, Ecuador,

June 22, 1990: Invited presentation: Domesticación de cultivos y recursos fitogenéticos (Domestication of cultivated plants and plant genetic resources)

5. Estación Experimental Agrícola de Santa Catalina, INIAP, Quito, Ecuador, July 2, 1990: Invited presentation: La domesticación del frijol: implicaciones para los recursos fitogenéticos y el mejoramiento. (Domestication of common bean: implications for genetic resources and breeding).

6. Centro Internacional de Agricultura Tropical, Cali, Colombia, September 11-14, 1990: Invited presentation: Improved characterization of *Phaseolus vulgaris* genetic variability.

7. Ninth SUA/WSU Bean/Cowpea CRSP and 2nd SADCC/CIAT Regional Bean Research Workshop, Sokoine University, Morogoro, Tanzania, September 17-22, 1990: Invited presentation: Genetic resources available in *Phaseolus vulgaris*

8. American Society of Agronomy 1990 Annual Meeting, San Antonio, October 21-26, 1990: Presentation: with S.P. Singh, J.A. Gutiérrez, A. Molina, C. Urrea: Marker-based classification of genetic diversity in common bean

9. American Society of Agronomy 1990 Annual Meeting, San Antonio, October 21-26, 1990: Presentation: by R. Nodari (graduate student); S.P. Singh, P. Gepts : RFLPs and linkage mapping in common bean

10. University of California, Davis, Plant Biology Graduate Group Seminar presentation, December 7, 1990: Germplasm exploration for wild beans in Ecuador

11. Michigan State University, East Lansing, MI. Mini-Symposium on Molecular markers in plant genetics and crop evolution, Invited presentation, December 14, 1990: Molecules, phenotypes, and multivariate analyses in *Phaseolus* evolution

12. North Carolina State University, Symposium on Plant Breeding in the 1990s, Poster presentation, with R. Nodari: Linkage map of common bean (*Phaseolus vulgaris* L.)

13. North Carolina State University, Symposium on Plant Breeding in the 1990s, Poster presentation, with S.P. Singh, J.A. Gutiérrez, A. Molina, C.A. Urrea: Potential of wild common bean populations for improvement of cultivars in the tropics.

14. 13th North American Symbiotic Nitrogen Fixation Conference 1991, Banff, Canada, August 25-30, 1991: Poster presentation with S.M. Tsai, R.O. Nodari, P.A. Arraes, E. Koinange: RFLP analysis of bean (*Phaseolus vulgaris* L.) cultivars

selected for nitrogen fixation.

15. AID/PSTC networking meeting, Banff, Canada, September 1-6, 1991. Presentation, with S.P. Singh: Establishing a core collection for *Phaseolus vulgaris* genetic resources conservation.

16. AID/PSTC networking meeting, Banff, Canada, September 1-6, 1991. Invited plenary presentation: Usefulness of molecular identification of plant selections and microorganisms.

17. UC Davis, Dept. of Pomology:Variations in common bean: potential implications for breeding and disease resistance management

18. Ferry-Morse Seed Co., Hollister, CA: December 4, 1991: Invited presentation: The potential of RAPD markers in breeding

19. Miami Bio/Technology Winter Symposia: January 20-24, 1992: Poster, selected for short presentation: with R.O. Nodari and S.M. Tsai, Locating genetic factors for nodulation intensity on the RFLP map of common bean (*Phaseolus vulgaris*)

20. UCD-Pacific Rim Food and Agricultural Biotechnology Conference: Sacramento, June 20-24, 1992: Poster presentation, with R.L. Gilbertson, R.O. Nodari, P. Guzmán, and S.M. Tsai, Mapping genetic factors controlling host-bacteria interactions in common bean.

21. Third International Legume Conference, Kew, U.K., July 12-17, 1992: Invited Presentation: Origin and Evolution of Cultivated *Phaseolus* Species

22. Academia Brasileira de Ciênia, Rio de Janeiro, July 24, 1992: Presentation by J. Dobereiner: S.M. Tsai, R.O. Nodari, and P. Gepts: Analise quantitativa e mapeamento de genes da nodulaçao em feijoeiro (*Phaseolus vulgaris* L.) [Quantitative analysis and mapping of nodulation genes in common bean (*Phaseolus vulgaris* L.)]

23. International Board for Plant Genetic Resources-CENARGEN-CNG, Brasilia: August 23-29, 1992: Invited presentation: Genetic markers and core collections

B. Publications

1. 1991 Singh, S.P., Nodari, R. Gepts, P. Genetic diversity in cultivated Phaseolus vulgaris. I. Allozymes. Crop Sci. 31: 19-23.

2. 1991 Singh, S.P., Gutiérrez, J.A., Molina, A., Urrea, C., Gepts, P. Genetic diversity in cultivated Phaseolus vulgaris. II. Marker-based analysis of

morpho-agronomic diversity. Crop Sci. 31: 23-29.

3. 1991 Singh, S.P., Gepts, P, Debouck, D.G. Races of common bean, (Phaseolus vulgaris, Fabaceae). Econ. Bot. 45: 379-396.

4. 1992 Nodari, R.O., Koinange, E.M.K., Kelly, J.D., Gepts, P. Towards an integrated linkage map of common bean. I. Development of genomic DNA probes and levels of restriction fragment length polymorphism. Theor. Appl. Genet. 84: 186-192

5. 1992 Gepts P,Llaca V, Nodari RO, Panella L. Analysis of seed proteins, isozymes, and RFLPs for genetic and evolutionary studies in Phaseolus. In: Linskens HF,Jackson JF Ieds), Seed analysis, Modern Methods of Plant Analysis, Vol. 14. Springer, Berlin: pp. 63-93.

6. 1992 Nodari RO, Tsai SM, Gilbertson RL, Gepts P. Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. Theor. Appl. Genet., in press.

7. 1992 Gepts P. The use of molecular and biochemical markers in crop evolution studies. Evol. Biol., in press

8. 1992 Gepts P. Linkage map of common bean (Phaseolus vulgaris L.). In: O'Brien SJ (ed), Genetic Maps. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, in press

9. 1993 Gepts P Genetic markers and core collections. *In*: Hodgkin T, Brown AHD, Van Hintum T (eds) Proc. International Board for Plant Genetic Resources-CENARGEN-CNG Conference on Core Collections, Brasilia: August 23-29, 1992: in press.

C. Training: see above

V. Project Productivity

The project has accomplished all its original objectives (Objectives 1-4) and has gone well beyond them (Objectives 5-6). A possible exception is Objective 4. As mentioned earlier, this objective originally seeked to identify suitable non-radioisotopic labeling methods for Southern hybridization. The widespread introduction of the polymerase chain reaction (PCR) makes Southern hybridization less attractive because it is more cumbsersome and time-consuming compared to PCR. We have therefore focused on preliminary experiments to amplify a specific sequence of the phaseolin locus. These experiments have been successful and open the door for future research.

VI. Future Work

The core collection at CIAT that has been developed on the basis of agroecological data and molecular data needs to be validated to 1) examine whether the core collection represents an adequate cross-section of the diversity in the main collection; 2) deternine whether simpler methods (*e.g.* geographic or morphological criteria correlated with molecular markers) can also be adequate sampling criteria.

PCR-based markers offer the best opportunity to obtain a category of markers that is versatile (large numboer of loci, absence of environmental effects) and relatively easier to use so that they could be applied in developing countries. RFLPs need to be converted to PCR-based markers. Likewise, RAPDs should also be converted to full-fledged PCR markers (*e.g.*, SCARs) to avoid problems of reliability associated with RAPDs.

VII. Literature cited (in addition to citations in manuscripts)

Brown AHD (1989) The case for core collections. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds), The use of plant genetic resources. Cambridge Univ. Press, Cambridge, UK, pp. 136-156

Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Arber WK, Llimensee K, Peacock WJ, Starlinger P (eds), Genetic manipulation: impact on man and society. Cambridge Univ. Press, Cambridge, UK, pp. 161-170

Singh SP, Cajiao C, Gutierrez JA, Garcia J, Corrales MAP, Morales FJ (1989) Selection for seed yield in inter-gene pool crosses of common bean. Crop Sci 29:1126-31

Reprinted from *Crop Science*
Vol. 31, No. 1

Genetic Diversity in Cultivated Common Bean: I. Allozymes

Shree P. Singh, R. Nodari, and P. Gepts*

ABSTRACT

Previous studies using phaseolin seed protein as a marker have revealed that cultivated common bean (*Phaseolus vulgaris* L.) resulted from multiple domestications in Mesoamerica and in Andean South America. Because these studies were based on variation at a single locus, confirmation was sought by analyzing patterns of diversity at nine polymorphic allozyme loci, all unlinked to the phaseolin locus: ribulose bisphosphate carboxylase, shikimate dehydrogenase, cathodal peroxidase, malic enzyme, malate dehydrogenase (two loci), diaphorase (two loci), and leucine aminopeptidase. A total of 227 landraces representing geographical regions from Mexico to Argentina and Chile were analyzed for these enzyme systems. A crude homogenate of primary leaf or root tissue (depending on enzymes assayed) from five seedlings of each landrace grown in vermiculite was used for starch gel electrophoresis. A cluster analysis based on Nei's genetic distance (D) was performed according to the unweighted paired group method of Sneath and Sokal. Our results confirm the existence of two major groups in cultivated common bean, Mesoamerican vs. Andean American; provide indications of gene flow from wild to cultivated beans; and suggest at least five subgroups within Mesoamerican and four within Andean cultivar groups.

STUDIES IN CROP EVOLUTION have traditionally attempted to determine the crop's ancestry based on morphological similarities and the production of viable and fertile hybrids between wild ancestor and cultivated descendant. Recently, the use of molecular markers has helped identification of the actual ancestral populations in maize (*Zea mays* L.) (Doebley et al., 1987) and pea (*Pisum sativum* L.) (Palmer et al., 1985), and determination of the effect of domestication on genetic diversity in barley (*Hordeum vulgare* L.) (Clegg et al., 1984; Jana and Pietrzak, 1988) and pearl millet (*Pennisetum glaucum* [L.] R. Br.) (Gepts and Clegg, 1989).

The common bean is an annual, diploid ($2n = 2x = 22$) species that originated in the Americas and consists of wild and cultivated forms. The wild forms are distributed from northern Mexico to northern Argentina (Brücher, 1988; Delgado Salinas et al., 1988). They are morphologically similar to the cultivars and yield viable and fertile progenies when crossed with them (Brücher, 1988; Gentry, 1969). Because of the extensive distribution of the common bean's wild ancestor, the exact siting of its domestication has been subject to speculation. Phaseolin diversity data suggest that cultivated common bean arose from multiple domestications along this extended distribution (Gepts and Bliss, 1986; Gepts et al., 1986). In particular, two major domestications appear to have given rise to Mesoamerican and southern Andean cultivars, respectively. A third and minor domestication may have taken place in Colombia or Central America (Gepts and Bliss, 1986; Koenig et al., 1990).

Allozyme analyses of wild forms have confirmed the existence of these two major groups and clarified their geographic boundaries: the Mesoamerican forms include wild populations from northern Mexico to Colombia, whereas the Andean types include populations from Peru and Argentina (Koenig and Gepts, 1989b).

In this article, we report on allozyme analyses of 227 cultivated landraces representing a geographical distribution extending from Mexico to Argentina and Chile. The objectives were to determine whether cultivated common bean displays a geographic pattern of allozyme diversity similar to its wild ancestors and to identify subgroups within the large groups of Mesoamerican and Andean cultivars.

MATERIALS AND METHODS

The collection of 227 landraces analyzed in this study was obtained from the *Phaseolus* world collection at the International Center for Tropical Agriculture (CIAT), Cali, Colombia, and the Western Regional Plant Introduction Station, Pullman, WA. It is representative of major geographical and ecological regions in the area of origin of common bean, which extends from Mexico to Argentina. Among Mesoamerican and Caribbean materials, 66 originated in Mexico, 7 in Guatemala, 5 in El Salvador, 2 each in Nicaragua and Costa Rica, and 1 in the Dominican Republic. Among Andean South American materials, 54 landraces came from Colombia, 36 from Ecuador, 19 from Peru, 14 from Chile, 3 from Bolivia, and 1 from Argentina. From lowland South America, 17 Brazilian landraces were included. The wild *P. vulgaris* accessions included in the cluster analysis were those studied by Koenig and Gepts (1989b).

Allozyme analyses of a crude homogenate of primary leaf or root tissue (depending on the enzyme system assayed) from an average of five seedlings (10 d old) grown in vermiculite were performed as described by Koenig and Gepts (1989b). A total of nine enzyme systems showing polymorphism were assayed: ribulose bisphosphate carboxylase (small subunit; RBCS), shikimate dehydrogenase (SKDH), peroxidase (PRX), malic enzyme (ME), malate dehydrogenase (MDH), diaphorase (DIAP), and leucine aminopeptidase (LAP). The MDH and DIAP enzyme systems each had two independent loci. Seven additional enzyme systems did not reveal polymorphisms among cultivars: aspartate aminotransferase, fructokinase, glucose-6-phosphate dehydrogenase, glucose phosphate isomerase, peptidase, 6-phosphogluconate dehydrogenase, and triose phosphate isomerase. Allozyme loci and alleles were designated as in Koenig and Gepts (1989a). In the nine systems studied, the most common allele was designated 100 and all other allozymes were measured in millimeters from the standard.

In order to compare allozyme diversity in the wild ancestor and the cultivated descendant, data from cultivated landraces were analyzed jointly with data from 83 wild forms obtained earlier (Koenig and Gepts, 1989b). A dendrogram based on Nei's (1973) genetic distance (D) was constructed according to the unweighted paired group method of Sneath and Sokal (1973) with a statistical package provided by Dr. K. Ritland, Univ. of Toronto.

S.P. Singh, Centro Int. de Agric. Tropical (CIAT), Apartado Aereo 6713, Cali, Colombia; R. Nodari and P. Gepts, Dep. of Agron. and Range Sci., Univ of California, Davis, CA 95616. Received 15 Dec. 1989. *Corresponding author.

Published in Crop Sci. 31:19–23 (1991).

22

RESULTS AND DISCUSSION

The genetic control of the polymorphism for enzyme systems RBCS, SKDH, PRX, ME, MDH, DIAP, and LAP in common bean has been determined previously (Weeden, 1984, 1986; Koenig and Gepts, 1989a). In our study, polymorphism for five enzyme systems each originated at a single locus (the number of alleles is given in parentheses): *Rbcs* (2), *Skdh* (2), *Prx* (2), *Me* (3), and *Lap-3* (2); for the remaining two systems, two loci were involved: *Mdh-1* (2) and *Mdh-2* (2), and *Diap-1* (2) and *Diap-2* (2). The average number of alleles per polymorphic locus was therefore 2.1.

A total of 76 genotypes, each with its distinctive allele combination at the nine isozyme loci, were identified among the 83 wild forms (Koenig and Gepts, 1989b) and 227 cultivars studied (Fig. 1). Information on the composition of the allozyme genotypes (allelic combination, wild and cultivated accessions) can be

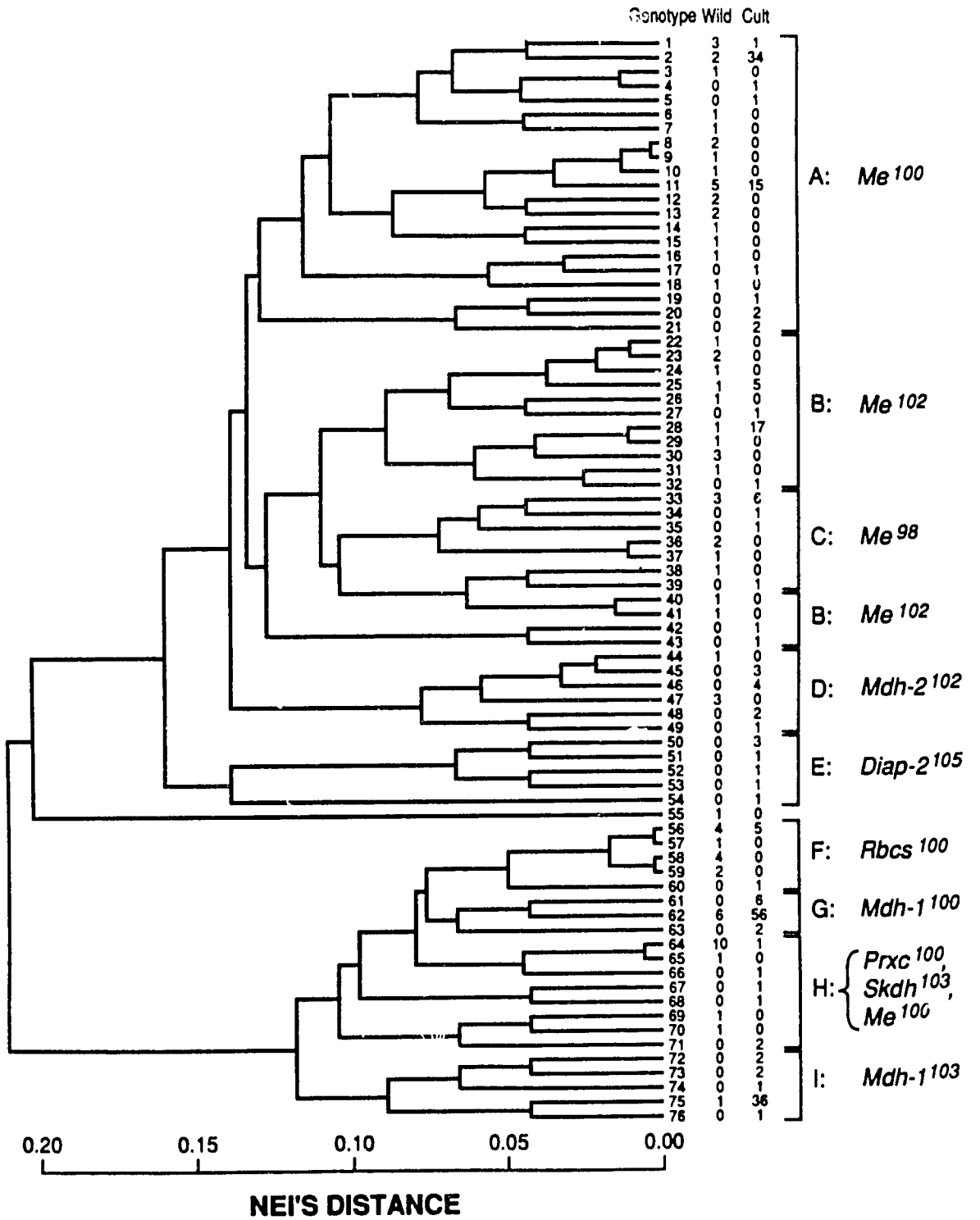


Fig. 1. Dendrogram of allozyme diversity in *Phaseolus vulgaris*. To the right of the consecutive genotype number is the number of wild and cultivated (cult) accessions for each genotype. Letters A to I indicate clusters of accessions sharing a common allozyme (as indicated to the right of the cluster letter).

obtained from the corresponding author. Because several allelic combinations were observed in more than one cultivar or wild population, only one cultivar and wild population per allelic combination were included in the cluster analysis (Fig. 1).

The cluster analysis based on Nei's (1973) genetic distance revealed the existence of two major groups of cultivars (Fig. 1). These two groups exhibited a differential geographic distribution. Most of the cultivars in the upper group (1 to 54) of Fig. 1 originated in Mesoamerica, while the cultivars of the lower group (56 to 76) originated in the Andes. Previous studies on the organization of genetic diversity at the molecular level in *P. vulgaris* had used phaseolin seed protein in wild populations and cultivars (Gepts and Bliss, 1985, 1986; Gepts et al., 1986; Koenig et al., 1990) and allozymes in wild populations (Koenig and Gepts, 1989b) and had provided evidence that the genetic diversity of *P. vulgaris* is organized into two major groups: Mesoamerican vs. Andean forms. As illustrated in Fig. 1, our allozyme data on cultivated landraces also distinguished two major groups: Mesoamerican vs. Andean. The allozyme profile of each of these two groups matched that of the corresponding wild populations in the same area (Koenig and Gepts, 1989b), providing additional evidence favoring independent origins for these two groups of cultivars. As had been observed in wild bean by Koenig and Gepts (1989b), cultivars with Type S, Sd, Sb, or B phaseolin had a Mesoamerican allozyme profile, whereas Type T, C, and H phaseolin cultivars had an Andean profile. The divergence between Mesoamerican and Andean landraces is therefore a consequence of the divergence between Mesoamerican and Andean wild ancestors, which predates the domestication process.

Table 1. Clusters of cultivated *Phaseolus vulgaris* from Latin America based on a characteristic common allozyme.

Cluster†	Genotype number†	Characteristic allozyme‡	Representative cultivars
		Mesoamerican	
A	1–21	Me^{100}	Puebla 152, Apetito, Flor de Mayo, Conejo, Cacahuate criollo, Frijola, Jalisco 20
B	22–32, 40–43	Me^{102}	G 2618, Pinto, Durango 222, Bayo Río Grande, Garrapato
C	33–39	Me^{90}	Rabia de Gato, Negro Argel, Pitouco
D	44–49	$Mdh\text{-}2^{102}$	Flor de Mayo, Rosa de Castilla, Naranja coral, Ecuador 299
E	50–54	$Diap\text{-}2^{105}$	Porillo Sintético, Rio Tibagi, H6 Mulatinho
		Andean	
F	56–60	$Rbcs^{100}$	Almidón, Balín d'Albenga, Bola 60 días
G	61–63	$Mdh\text{-}1^{100}$	Coscorrón, Frutilla, Burrito, Tórtola Los Angeles, Estrada Rosado, Cargamanto, Sangretoro, Algarrobo
H	64–66,	Prx^{100}	Blanco Sabanero
	67–68	$Skdh^{101}$	Pajaritos, Huevo de Pinche
	69–71	Me^{100}	Bagajo I, Huila 9
I	72–76	$Mdh\text{-}1^{101}$	Mortiño, Nuña, Bolón rojo, Bola roja, Bolón bayo

† Clustering and genotypes from Fig. 1.
‡ Diap = diaphorase, Me = malic enzyme; Mdh = malate dehydrogenase; Prx = peroxidase; Rbcs = ribulose bisphosphate carboxylase (small subunit); and Skdh = Shikimate dehydrogenase. Hyphenated numbers indicate locus; superscripted numbers indicate alleles.

At the molecular level, cultivated types of a major group (such as the Mesoamerican group) are more closely related to the wild ancestor of that same group than to cultivars of the other group (in this example, the Andean group).

Of 76 genotypes identified through analyses of allozyme data thus far in *P. vulgaris*, 34 were present only among wild populations and 32 were found among cultivars but not in wild forms; only 10 genotypes were found in both wild and cultivated germplasm (Fig. 1). Thus, the reduction in diversity observed upon domestication is not as pronounced for allozymes as for phaseolin (Gepts et al., 1986). This may be attributed to the lower level of total allozyme diversity present among wild bean at any given locus when compared to phaseolin diversity. For example, some 15 to 20 phaseolin patterns were observed among Mexican wild bean populations (Gepts et al., 1986), whereas only zero to four alleles per allozyme locus were observed among the same wild germplasm (Koenig and Gepts, 1989b).

Moreover, no definite trend could be established regarding increase or decrease in number of alleles per allozyme locus in wild vs. cultivated germplasm. For example, *Diap-1* displayed only two alleles ($Diap\text{-}1^{100}$ and $Diap\text{-}1^{95}$) in cultivars vs. four alleles ($Diap\text{-}1^{102}$, $Diap\text{-}1^{100}$, $Diap\text{-}1^{96}$, and $Diap\text{-}1^{95}$) in wild materials. On the other hand, *Diap-2* showed two alleles ($Diap\text{-}2^{105}$ and $Diap\text{-}2^{100}$) in cultivars and only one ($Diap\text{-}2^{100}$) in wild populations. The absence of the $Diap\text{-}2^{105}$ allele may be attributed to an insufficient sampling of the Mesoamerican wild populations, especially in Central America, from which most of the cultivated landraces carrying $Diap\text{-}2^{105}$ originated. If, however, the presence of $Diap\text{-}2^{105}$ among wild beans is confirmed, cultivars of Cluster E (see Fig. 1 and Table 1) would have evolved either as direct descendants or would represent introgressants from those wild beans. On the contrary, the $Diap\text{-}2^{105}$ allele could be the result of a mutation during or after domestication of this cultivar genotype. Allozymes of 34 genotypes found only in wild populations but absent in cultivars are likely due to inadequate sampling of the latter or else may suggest that all wild bean populations did not participate in the domestication and evolution of cultivars.

Landraces in Cluster D (Fig. 1 and Table 1) exhibited the $Mdh\text{-}2^{102}$ allele, which was observed previously only among Mesoamerican wild beans (Koenig and Gepts, 1989b). The absence of this allele among Andean cultivars and the most predominant clusters (A and B) of the Mesoamerican landraces suggests that it may have been introduced from Mesoamerican wild beans by occasional gene flow. Although common bean is a predominantly self-pollinated species, outcrossing rates of up to 70% have recently been reported (Wells et al., 1988). Outcrossing rates of much lower magnitude would, however, be sufficient to assure gene flow from wild to cultivated forms.

The Andean and Mesoamerican landraces exhibited contrasting alleles at the *Diap-1*, *Lap-3*, *Me*, *Prx*, *Rbcs*, and *Skdh* loci (Table 2). Alleles at the same or other loci distinguished subgroups within cultivars from the two major domestication centers. For example, approximately one-third of the Mesoamerican acces-

24

Table 2. Allozyme distribution in Mesoamerican and Andean cultivated *Phaseolus vulgaris*.

Geographic origin†	Allozymes and alleles‡																		
	Diap-1		Diap-2		Lap-3		Me			Mdh-1		Mdh-2		Prx		Rbcs		Skdh	
	100	95	105	100	103	100	102	100	98	103	100	103	100	100	98	100	98	103	100
	%																		
Mesoamerican (83)	14	82	4	96	5	95	29	60	11	1	99	8	92	68	32	96	4	87	13
Andean (127)	94	6	0	100	98	2	1	3	96	30	70	0	100	11	89	16	84	11	89
Brazil (17)	29	71	6	94	29	71	12	47	41	0	100	0	100	59	41	71	29	71	29

† Number of genotypes is given in parentheses.
‡ Diap = diaphorase, Lap = leucine aminopeptidase; Me = malic enzyme; Mdh = malate dehydrogenase; Prx = peroxidase; Rbcs = ribulose bisphosphate carboxylase (small subunit); and Skdh = Shikimate dehydrogenase. Hyphenated numbers identify the locus; numbers below the allozyme identify the alleles at that locus.

sions carried the Me^{102} allele, whereas the other two-thirds exhibited the Me^{100} allele, and 96% of the Andean landraces had the Me^{98} allele. Similarly, one-third of the Andean accessions showed the $Mdh\text{-}1^{103}$ allele, whereas the other two-thirds of the Andean accessions and most of the Mesoamerican accessions showed the $Mdh\text{-}1^{100}$ allele. Thus, closer inspection of Fig. 1 reveals that, within the Mesoamerican and Andean cultivated germplasm, clusters of landraces can be identified that share a common allozyme and, therefore, can presumably be traced to a common ancestor (with exceptions due to independent mutations or outcrossing). The common ancestry hypothesis accounts for the inclusion of genotypes 40 to 43 in Group B (Fig. 1). Table 1 provides the summary of five cluster groups in Mesoamerican cultivars and four cluster groups in Andean cultivars, each with their characteristic allozyme and representative cultivars. For example, in Mesoamerica, Cluster A, comprising genotypes 1 to 21, is characterized by the Me^{100} allele. Similarly, Cluster B, characterized by the Me^{102} allele, includes genotypes 22 to 32 and 40 to 43. Clusters F and H, characterized by the $Rbcs^{100}$ allele and the Prx^{100}, $Skdh^{103}$, and Me^{100} alleles, respectively, were considered to be distinct from the G and I clusters because their characteristic alleles are representative of Mesoamerican genotypes (Table 2). Landraces belonging to Clusters F and H may therefore represent hybrids between the Andean and Mesoamerican groups. The relationship between these clusters and morphological and agronomic variation is investigated in the companion paper (Singh et al., 1991).

One striking feature of allozyme diversity among cultivars is the existence of a limited number of genotypes represented by a large number of landraces and which differ from other genotypes at one or more allozyme loci: e.g., genotypes 2, 62, and 75, with 34, 56, and 36 cultivated representatives, respectively (Fig. 1). Whereas cultivars with the same allozyme genotype show similarities for certain morphological traits (see companion paper, this issue), they can exhibit considerable diversity for other morphological traits such as seed type (size, shape, color, and color pattern; data not shown). This observation suggests that most of the cultivars with the same allozyme genotype, following their origination from a common ancestor, have undergone further diversification for morphological traits but not for molecular markers. Exceptions to this evolutionary pattern could be due to independent mutations or occasional outcrosses. In plants, it has been suggested that certain morphological traits, notably those involving the presence or absence of certain organs, are also coded by a small number of genes or even a single gene (Knight, 1948; Hilu, 1983; Gottlieb, 1984). In common bean, traits that distinguish wild vs. cultivated beans and cultivated beans among them, are also coded by few genes (with the exception of seed size, which is a multigenic trait) (Gepts, 1990). Major phenotypic changes can therefore, have a simple genetic control and occur without major divergence at the molecular level.

This contrast in diversity levels between morphological traits and biochemical markers raises the issue as to which trait should be used to assess genetic diversity. A higher plant contains several tens of thousands of genes (Kamalay and Goldberg, 1980). Most of these genes, including allozyme and phaseolin genes, do not have a strong effect on the phenotype and are most likely selectively neutral (Kimura, 1983); this distinguishes them from genes determining the morphology of cultivars (Gepts, 1990; Gepts and Debouck, 1991). Consequently, they are less likely to be subject to selection pressures; they could, however, be subject to genetic drift, which would ultimately lead to fixation. Because of their strong phenotypic effect and the high heritability of the traits they encode, genes for morphological traits constitute a sample of genes that is atypical of the large majority of genes in a genome. Their gene action, however, makes these traits highly responsive to selection during and after domestication, which accounts for their high levels of diversity in the primary cultivated gene pool as a whole as opposed to molecular markers. Allozymes and phaseolin offer the additional advantage that they can reveal genetic relationships among accessions. Based on these arguments, molecular markers may be more accurate descriptors of the overall levels of genetic diversity in a genome than morphological markers.

ACKNOWLEDGMENTS

We thank Dr. R. Kesseli for his assistance in the genetic diversity analysis and G. Arndt for useful comments. This research at the University of California, Davis, was supported by the International Center for Tropical Agriculture (CIAT), Cali, Colombia, and the U.S. Agency for International Development.

REFERENCES

Brücher, H. 1988. The wild ancestor of *Phaseolus vulgaris* in South America. p. 185–214. *In* P. Gepts (ed.) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, the Netherlands.
Clegg, M.T., A.H.D. Brown, and P.R. Whitfeld. 1984. Chloroplast

DNA diversity in wild and cultivated barley: Implications for genetic conservation. Genet. Res. 43:339–343.
Delgado Salinas, A., A. Bonet, and P. Gepts. 1988. The wild relative of *Phaseolus vulgaris* in Middle America. p 163–184. *In* P. Gepts (ed.) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, the Netherlands.
Doebley, J., W. Renfroe, and A. Blanton. 1987. Restriction site variation in the Zea chloroplast genome. Genetics 117:139–147.
Gentry, H.S. 1969. Origin of the common bean, *Phaseolus vulgaris*. Econ. Bot. 23:55–69.
Gepts, P. 1990. Biochemical evidence bearing on the domestication of *Phaseolus* beans. Econ. Bot. 44(3 Suppl.):28–38.
Gepts, P., and F.A. Bliss. 1985. F1 hybrid weakness in the common bean: Differential geographic origin suggests two gene pools in cultivated bean germplasm. J. Hered. 76:447–450.
Gepts, P., and F.A. Bliss. 1986. Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ. Bot. 40:469–478.
Gepts, P., and M.T. Clegg. 1989. Genetic diversity in pearl millet (*Pennisetum glaucum* [L.] R. Br.) at the DNA sequence level. J. Hered. 80:203–208.
Gepts, P.A., and D.G. Debouck. 1991. Origin, domestication, and evolution of the common bean (*Phaseolus vulgaris* L.). *In* A. van Schoonhoven and O. Voysest (ed.) Common beans: Research for crop improvement. CIAT, Cali, Colombia (in press).
Gepts, P.A., T.C. Osborn, K. Rashka, and F.A. Bliss. 1986. Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): Evidence for multiple centers of domestication. Econ. Bot. 40:451–468.
Gottlieb, L.D. 1984. Genetics and morphological evolution in plants. Am. Nat. 123:681–709.
Hilu, K. 1983. The role of single-gene mutation in the evolution of flowering plants. *In* M.K. Hecht, B. Wallace, and G.T. Prance (ed.) Evol. Biol. 16:97–128.
Jana, S., and L.N. Pietrzak. 1988. Comparative assessment of genetic diversity in wild and primitive cultivated barley in a center of diversity. Genetics 119:981–990.
Kamalay, J.C., and R.B. Goldberg. 1980. Regulation of structural gene expression in tobacco. Cell 19:935–946.
Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge Univ. Press, Cambridge, England.
Knight, R.L. 1948. The role of major genes in the evolution of economic characters. J. Genet. 48:370–387.
Koenig, R., and P. Gepts. 1989a. Segregation and linkage of genes for seed proteins, isozymes, and morphological traits in common bean (*Phaseolus vulgaris*). J. Hered. 80:455–459.
Koenig, R., and P. Gepts. 1989b. Allozyme diversity in wild *Phaseolus vulgaris*: Further evidence for two major centers for genetic diversity. Theor. Appl. Genet. 78:809–817.
Koenig, R., S.P. Singh, and P. Gepts. 1990. Novel phaseolin types in wild and cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ. Bot. 44:50–60.
Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci. USA 70:3321–3323.
Palmer, J.D., R.A. Jorgensen, and W.F. Thompson. 1985. Chloroplast DNA variation and evolution in *Pisum*: Patterns of change and phylogenetic analysis. Genetics 109:195–213.
Singh, S.P., J.A. Gutiérrez, A. Molina, C. Urrea, and P. Gepts. 1991. Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci. 31:23–29 (this issue).
Sneath, P.H.A., and R. Sokal. 1973. Numerical taxonomy. W.H. Freeman, San Francisco.
Weeden, N.F. 1984. Linkage between the gene coding the small subunit of ribulose biphosphate carboxylase and the gene coding malic enzyme in *Phaseolus vulgaris*. Annu. Rep. Bean Improv. Coop. 27:123–124.
Weeden, N.F. 1986. Genetic confirmation that the variation in the zymograms of 3 enzyme systems is produced by allelic polymorphism. Annu. Rep. Bean Improv. Coop. 29:117–118.
Wells, W.C., W.H. Isom, and J.G. Waines. 1988. Outcrossing rates of six common bean lines. Crop Sci. 28:177–178.

Genetic Diversity in Cultivated Common Bean: II. Marker-Based Analysis of Morphological and Agronomic Traits

Shree P. Singh, J. A. Gutiérrez, A. Molina, C. Urrea, and P. Gepts*

ABSTRACT

Knowledge of patterns of genetic diversity enhances the efficiency of germplasm conservation and improvement. This study examined the organization of diversity for morphological and agronomic characteristics in 306 landraces of cultivated common bean (*Phaseolus vulgaris* L.) from Latin America and its relationship with phaseolin seed protein and allozyme diversity of the landraces. Data on pigmentation, growth habit, and leaflet, pod, seed, and phenology traits, as well as reaction to four important diseases and an insect pest, obtained from field evaluations at three locations in Colombia during the 1987–1988 cropping season, were analyzed by multivariate statistical analyses. In addition, these same 306 landraces were characterized by electrophoresis for phaseolin seed protein and nine allozymes. Results permitted separation of these landraces into Mesoamerican and Andean groups, confirming prior phaseolin and allozyme data. A marker-based multivariate analysis, using phaseolin or allozymes as an initial classification criterion, followed by a corroborating analysis of morpho-agronomic traits, suggested the existence of subgroups within each of the major Andean and Mesoamerican groups, with distinctive morphology, adaptation, and disease resistances. Molecular analyses in conjunction with morphological and agronomic evaluations of gene bank accessions are recommended, because these provide complementary information and increase the resolving power of genetic diversity analyses.

CULTIVATED COMMON BEAN is a morphologically diverse crop (Hedrick, 1931). Strikingly large variations are found for growth habit, pigmentation, pod, seed, phenology, and other characters (Leakey, 1988; Singh, 1989; Vanderborght, 1986). This diversity reflects the wide range of ecological and human environments under which the crop has evolved over millennia. In spite of the importance of these traits, the organization of their genetic diversity is poorly understood.

Study of patterns of variation for phaseolin seed protein (Gepts and Bliss, 1985; Gepts et al., 1986) and allozymes (Koenig and Gepts, 1989; Singh et al., 1991) has revealed the existence of two major groups, a Mesoamerican and an Andean, within common bean germplasm. Attempts have been made to suggest subdivisions within the Mesoamerican and Andean germplasm. Evans (1976) and Vanderborght (1986) proposed classifications relying primarily on growth habit. Singh et al. (1989) identified three races within each of the Mesoamerican and Andean groups, and Singh (1988, 1989) proposed the existence of 12 gene pools in dry bean on the basis of growth habit and seed, morphological, and adaptational traits.

S.P. Singh, J.A. Gutiérrez, A. Molina, and C. Urrea, Bean Program, Int. Ctr. for Tropical Agric. (CIAT), Apartado aereo 6713, Cali, Colombia; and P. Gepts, Dep. of Agron. and Range Sci., Univ. of California, Davis, CA 95616. Received 15 Dec. 1989. *Corresponding author.

Published in Crop Sci. 31:23–29 (1991).

66

Multivariate statistical methods have been used previously to analyze patterns of genetic diversity. For example, Polignano and Spagnoletti-Zeuli (1985) examined variation of 10 morphological and agronomic traits in Mediterranean populations of *Vicia faba* L. with canonical discriminant and factor analyses. They were able to distinguish populations of different geographic origins within the Mediterranean basin, and they identified the specific traits responsible for the differences among them. In common bean, Vanderborght (1986) analyzed the morphological diversity of Mesoamerican and Andean germplasm using cluster analysis and principal component analysis.

Our objectives were to study patterns of diversity for morphological and agronomic traits in cultivated common bean and their relationship with patterns of diversity for phaseolin seed protein and allozymes.

MATERIALS AND METHODS

Plant Materials

Our sample consisted of 306 cultivated accessions from Latin America obtained from the International Center for Tropical Agriculture (CIAT), Cali, Colombia. To the best of our knowledge, all accessions were landraces as defined by Gepts et al. (1986). Seventy-two accessions originated in Mexico, 10 in Guatemala, 2 in Nicaragua, 5 in El Salvador, 1 in Honduras, 2 in Costa Rica, 2 in the Dominican Republic, 97 in Colombia, 44 in Ecuador, 31 in Peru, 17 in Chile, 1 each in Bolivia and Argentina, and 21 from Brazil. Seeds of all accessions used in this study originated from CIAT.

Morphological and Agronomic Traits

All 306 landraces were grown at two CIAT farms, Palmira (1000 m above sea level; mean growing temperature, 24 °C) and Popayán (1750 m above sea level; mean growing temperature, 18 °C), Colombia, in the 1987–1988 cropping season. Plots consisted of single 3-m rows spaced 60 cm apart. Within-row spacing was about 10 cm between plants. Plots were kept free from weeds, diseases, and insects throughout the growing season. Soil fertility and moisture conditions were adequate for normal crop growth at both locations. Five plants, surrounded by other plants in the row, under full competition were used for observations in each plot. Data were recorded on hypocotyl and flower color, days to flower and maturity, growth habit, length of fifth internode (cm), number of nodes to first flower, length and width of the central leaflet of the fifth trifoliolate leaf (cm), dry seed length and height (mm), seed shininess, 100-seed weight (g), and seed yield (g/plant).

Separate complementary nurseries were grown at Popayán for observation of the reaction to anthracnose (*Colletotrichum lindemuthianum* [Sacc. & Magn.] Scrib.), and at Santander de Quilichao (990 m above sea level; mean temperature, 24 °C) for reaction to common bacterial blight (*Xanthomonas campestris* pv. *phaseoli* [Smith] Dye) and angular leaf spot (*Phaeoisariopsis griseola* [Sacc.] Ferraris). Nurseries were inoculated with a mixture of local strains 35, 45, and 55 d after planting. Reaction to leafhoppers (*Empoasca kraemeri* Ross & Moore) was observed in a nursery established during the dry season in Palmira. The presence of the dominant *I* gene for resistance to bean common mosaic virus (BCMV) was verified in glasshouse inoculations of 7 to 10-d-old seedlings with the NL3 strain. Data on all these diseases and the insect damage were taken on a 1 to 9 scale (1 = immune and 9 = very susceptible) according to Van Schoonhoven and Pastor Corrales (1987).

Phaseolin Seed Protein

A small sample from cotyledons of a single seed of each landrace was taken for phaseolin determination using one-dimensional sodium dodecylsulfate–polyacrylamide gel electrophoresis (SDS–PAGE) as described by Brown et al. (1981). Details of these results have been reported elsewhere (Koenig et al., 1990).

Allozymes

Five individuals per accession were analyzed. The primary leaf or root apex tissue (depending on the enzymes assayed) of 10-d-old seedlings was utilized for preparation of crude homogenate. Paper wicks, dipped in homogenate, were then used for starch gel electrophoresis following the procedures of Weeden (1984). Nine polymorphic enzyme systems were assayed. The most common allele was designated as 100 and all other allozymes were measured in millimeters from the standard. Details of these allozyme diversity analyses are reported elsewhere (Singh et al., 1991).

Statistical Analysis

The SAS UNIVARIATE procedure (PLOT option) from SAS Institute (1985) was performed to test for normality. All variables approximate normality with the exception of hypocotyl and flower color, growth habit, seed shininess, number of nodes to first flower, and reaction to bean common mosaic virus, leafhoppers, common bacterial blight, anthracnose, and angular leaf spot. The first four represent categorical data, whereas the last five had highly skewed distributions. Analyses performed with the non-normal variables led, however, to results and conclusions similar to those of analyses performed without those variables. Results presented here come from analyses including all variables. In order to avoid effects due to scaling differences, variables were standardized to zero mean and unit variance using the SAS procedure STANDARD (SAS Institute, 1985).

Three types of multivariate analyses were performed, using SAS Institute (1985) programs. A principal component analysis was performed on the standardized variables listed in Table 1 using PROC PRINCOMP. Subsequently, a discriminant analysis was carried out with PROC DISCRIM. The initial classification criterion used was either the phaseolin type (Gepts and Bliss, 1985; Gepts et al., 1986; Koenig et al., 1990) or allozyme cluster membership (Singh et al., 1991). Posterior probabilities of membership were calculated using the same procedure. The contribution of each variable to the classification was estimated by calculating standardized discriminant coefficients (Afifi and Clark, 1984). Finally, a canonical discriminant analysis was performed using PROC CANDISC with phaseolin type or allozyme cluster membership as an initial classification criterion.

Univariate analyses of variance with PROC GLM (SAS Institute, 1985) were performed in order to identify those variables that distinguish the allozyme cluster groups within the Andean cultigens. Duncan's multiple range test was used to identify differences between means.

RESULTS

Morphological Differences between Mesoamerican and Andean Cultivated Genotypes

Principal component analysis showed that Mesoamerican and Andean cultigens (i.e., cultivated genotypes) had a distinct morphology and that the Mesoamerican group was morphologically more diverse than its Andean counterpart (Fig. 1). The first three principal components accounted for 24, 10, and 9%, respectively, of total variation. The traits respon-

sible for separation along the first principal component included (with loadings in parentheses) the fifth internode length (0.23), seed length (0.23), seed height (0.22), 100-seed weight (0.24), and number of seeds (−0.21). Traits affecting separation along the second principal component included growth habit (−0.23), days to flower (0.21), nodes to first flower (0.21), leaflet length (0.32), leaflet width (0.33), and number of pods per plant (−0.20). Along the third principal component, accessions were separated according to days to maturity (0.28) and number of nodes to first flower (0.33). In order to further examine the differences between Mesoamerican and Andean cultigens, a discriminant analysis was conducted using phaseolin type as a classification criterion. Landraces with Type S, Sd, Sb, and B phaseolin were classified as Mesoamerican, while landraces with Type T, C, and H phaseolin were classified as Andean (Gepts and Bliss, 1986; Gepts et al., 1986; Koenig et al., 1990; Koenig and Gepts, 1989). The variables that had the strongest effect on the discriminant function included length of fifth internode, node to first flower, leaflet length, leaflet width, seed length, seed height, and seed yield. Posterior probabilities of membership indicated that classification of landraces into the Mesoamerican or the Andean group corresponded in >96% of the cases with the classification into these same groups based on morpho-agronomic traits.

Landraces whose phaseolin and morphological classifications did not match included 'de Celaya' (Mexico), 'Mantequilla' (Mexico), Boyacá 17A (Colombia), and Cundinamarca 137 (Colombia) all of which had a Mesoamerican phaseolin but an Andean morphology. Conversely, landraces 'Burrito de Enrame' (PI 152313 from Ecuador), Antioquia 6 (Colombia), and Cauca 11 (Colombia) exhibited an Andean phaseolin and a Mesoamerican morphology. Most of these landraces originated from the Colombia–Ecuador area, which was hypothesized earlier to be a meeting place of the Mesoamerican and Andean germplasms (Gepts and Bliss, 1986).

A canonical discriminant analysis using phaseolin type as a classification criterion further confirmed the separation of the two groups of cultigens. The traits responsible for the separation of the two groups were similar to those identified by principal component analysis.

Identification of Subgroups within Mesoamerican and Andean Cultivated Genotypes

In a next stage, we examined whether the large Andean and Mesoamerican groups could be further divided into subgroups consisting of related landraces. Using principal component analysis, which attempts to identify distinct groups independently of any prior classification criterion, it was not possible to identify distinct subgroups within the Mesoamerican or Andean cultigens. The situation was improved when discriminant analysis or canonical discriminant analysis were used, both of which rely on prior classification criteria. However, we will focus only on the results obtained by canonical discriminant analysis, because this method provides a graphic output illustrating the existence of morphological groups. We used both phaseolin and allozymes as initial classification criteria.

Phaseolin

Using phaseolin as a classification criterion, canonical discriminant analysis allowed us to distinguish groups of cultigens with a common phaseolin type on the basis of their morphology. Because Andean cultivars were better adapted to the Popayán conditions, results from the Popayán nursery included fewer missing data compared to the Palmira nursery. Data from the former nursery will, therefore, be presented here. Among Mesoamerican cultigens, landraces with phaseolin Type B or an Sd could be distinguished from those with Type S or Sb phaseolin (Fig. 2). The canonical correlations corresponding to the first two axes ($r = 0.58$ and $r = 0.54$, respectively) were significantly different from zero ($P < 0.01$). Traits that separated groups along the first canonical variable included days to flowering and maturity, number of nodes to first flower, leaflet size, and seed yield (Table 1). Along the second canonical variable, groups were separated ac-

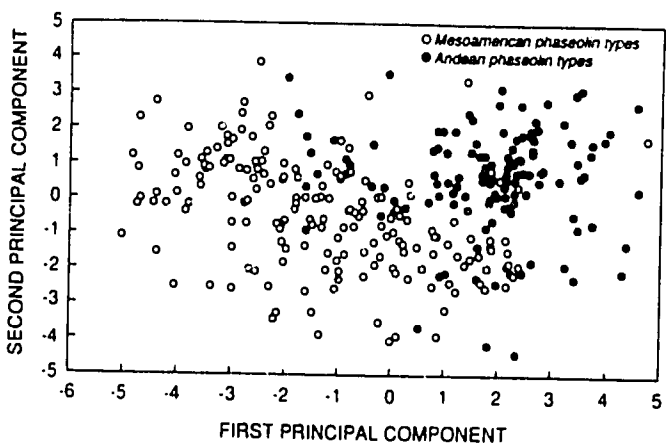


Fig. 1. Principal component analysis of diversity for morphological and agronomic traits in cultivated common bean.

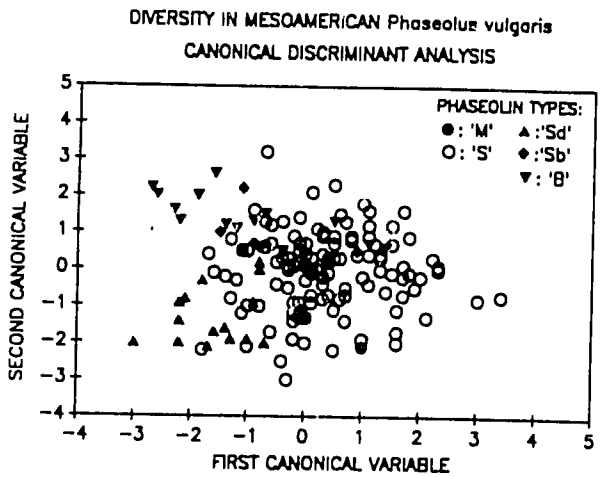


Fig. 2. Canonical discriminant analysis of diversity for morphological and agronomic traits of Mesoamerican cultivated genotypes of common bean using phaseolin type as an initial classification criterion.

Table 1. Correlation coefficients between original and canonical variables in a canonical discriminant analysis of cultivated common beans from Latin America using phaseolin as a classification criterion.

Trait	Mesoamerican landraces		Andean landraces	
	FCV†	SCV	FCV	SCV
	r			
Hypocotyl color	−0.62	0.52	0.78	0.63
Flower color	−0.08	0.83	−1.00	0.10
Days to flowering	0.87	0.39	1.00	0.06
Days to maturity	0.95	−0.05	0.89	−0.45
Fifth internode length	−0.51	−0.86	0.92	0.39
Number of nodes to first flower	0.90	−0.04	−0.91	−0.42
Leaflet length	0.85	0.44	1.00	0.07
Leaflet width	0.75	0.52	0.93	0.21
Seed length	0.03	−0.94	−0.70	0.71
Seed height	0.02	−0.98	0.58	0.05
Seed shininess	0.30	0.55	−0.15	0.99
100-seed weight	0.06	−0.96	0.93	0.38
Yield/plant	0.96	−0.20	0.60	0.80

† FCV and SCV = first and second canonical variables, respectively.

cording to flower color, fifth internode length, and seed size (Table 1).

Among Andean cultivars, landraces with 'T' phaseolin were morphologically distinct from those with 'C' or 'H' phaseolin type (Fig. 3). The canonical correlations of the first and second canonical variables ($r = 0.60$ and $r = 0.47$, respectively) were significantly at the $P < 0.01$ and $P < 0.05$ levels, respectively. The group of Type T phaseolin cultivars is characterized by a predominance of the *Mdh-1*^{100} allele (Table 2 and Singh et al., 1991). Compared to the next two groups, it has less strongly pigmented hypocotyls but more strongly pigmented flowers, it flowers and matures earlier, exhibits a more bushy growth habit (often Type I, as described by Singh, 1982), exhibits relatively smaller and more elongated leaflets, and has elongated seeds (Table 1). This group includes a high proportion of landraces from the northern Andes (Colombia and Ecuador), such as 'Algarrobo', 'Estrada Rosado', 'Sangretoro', 'Mortiñito', and 'Canario Alargado'. There are some notable exceptions to this general pattern, in particular with regard to growth habit. This group includes a proportion of cultigens with

Table 2. Distribution of phaseolin types among allozyme groups of cultivated common bean from Latin America.

Phaseolin type†	Allozyme groups‡								
	A	B	C	D	E	F	G	H	I
M	1	0	0	0	0	0	0	0	0
S	42	13	4	10	3	0	1	0	1
Sd	4	9	1	0	0	0	0	0	0
Sb	0	0	2	0	1	0	0	0	0
B	3	1	0	0	0	0	0	0	0
T	0	0	0	0	0	4	40	3	14
C	1	0	0	0	0	3	9	3	16
H	1	0	0	0	0	0	1	0	5

† After Koenig et al., 1990.
‡ After Singh et al., 1991.

either Growth Habit IV (e.g., 'Cargamanto', 'Calabozo', 'Radical'), often originating in the high altitude regions of the Andes, or Growth Habit III, originating mostly in the southern Andes (e.g., 'Coscorrón', 'Burrito', 'Frutilla Corriente', and 'Jalo'). The Growth Habit I cultivars of this group may correspond to the race 'Nueva Granada' of Singh et al. (1989). The Growth Habit III cultivars may fall within the race 'Chile' proposed by the same authors.

The Type C and H phaseolin cultivars are characterized by a high frequency of the *Mdh-1*^{103} allele (Table 2 and Singh et al., 1991). Morphologically, these cultivars exhibit many alternate traits from the preceding group, i.e., weaker pigmentation, late maturity, a stronger climbing tendency, larger and more obovate leaflets, and larger and more rounded seeds. Furthermore, landraces possessing Type C phaseolin can be distinguished from those with Type H phaseolin by their higher levels of seed shininess, their less elongated leaflets, longer seeds, and higher yield (Table 1). This group of Type C and H cultivars appears to correspond to the race 'Peru' of Singh et al. (1989).

Allozymes

Canonical correlation analysis was conducted using observations of the Popayán and Palmira nurseries. Because results of the two nurseries were similar, only the correlation coefficients between the original Popayán variables and the canonical variables are pre-

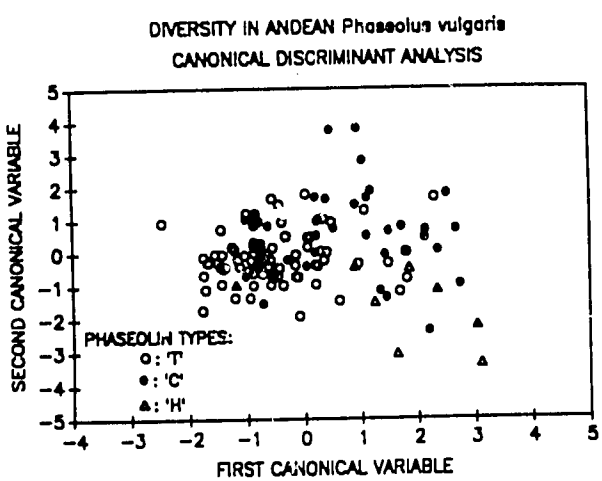


Fig. 3. Canonical discriminant analysis of diversity for morphological and agronomic traits of Andean cultivated genotypes of common bean using phaseolin type as an initial classification criterion.

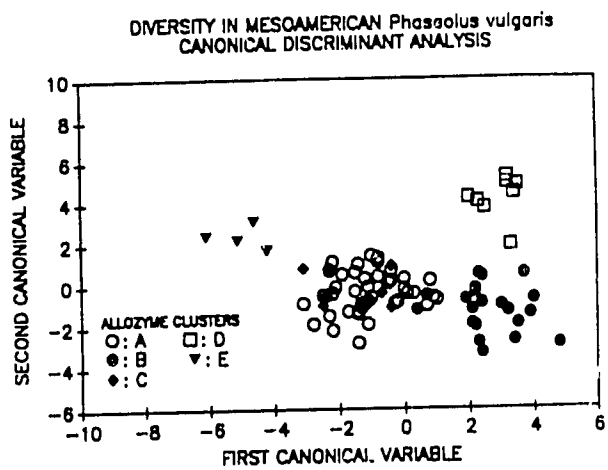


Fig. 4. Canonical discriminant analysis of diversity for morphological and agronomic traits of Mesoamerican accessions of cultivated common bean using allozyme groups as an initial classification criterion.

29

Table 3. Correlation coefficients between original and canonical variables in a canonical discriminant analysis of cultivated common bean from Mesoamerica using allozyme cluster membership as a classification criterion.

Trait	FCV†	SCV
	r	
Hypocotyl color	−0.97	0.00
Flower color	−0.93	−0.05
Days to flowering	−0.75	0.55
Days to maturity	0.13	0.95
Fifth internode length	0.79	0.10
Number of nodes to first flower	−0.77	0.24
Leaflet length	−0.27	0.91
Leaflet width	−0.24	0.89
Seed length	0.80	−0.08
Seed height	0.89	0.03
Seed shininess	0.65	0.32
100 seed weight	0.92	0.15
Yield/plant	0.33	0.86
Leafhoppers	0.48	−0.07
Bean common mosaic virus	0.84	−0.19
Common bacterial blight	0.40	0.70
Anthracnose	−0.79	−0.30
Angular leaf spot	−0.86	0.50

† FCV and SCV = first and second canonical variables, respectively.

sented. When canonical discriminant analysis was applied using individual isozyme loci as a classification criterion such as *Me* and *Mdh-1* that differentiate genotypes within the Mesoamerican and Andean gene pools, respectively (Singh et al., 1991), it was not possible to distinguish morphologically distinct groups. However, when the entire array of nine isozyme loci were used (Singh et al., 1991), a different picture emerged. The clusters of Mesoamerican cultigens with similar allozymes exhibited similar morphology, phenology, adaptation, and disease resistance (Fig. 4 and Table 3). The canonical correlation of the first axis ($r = 0.91$) was significantly different from zero ($P < 0.01$) and the canonical correlation of the second axis ($r = 0.85$) was nearly so ($P = 0.07$). Together, these two axes account for 73% of the total variation. The original variables that contributed most to the separation along the first canonical variable included hypocotyl and flower color, days to flowering, growth habit, fifth internode length, number of nodes to first flower, seed size, and reaction to BCMV, anthracnose, and angular leaf spot (Table 3). Along the second canonical variable, groups were separated according to days to maturity, leaflet length and width, and seed yield (Table 3).

The first canonical variable separated principally the E, C, A, and B clusters (in increasing order of their mean along this axis). A first major group, exhibiting the Me^{98} allele (Allozyme Group C) or allele $Diap\text{-}2^{105}$ (Allozyme Group E) and a high frequency of 'Sb' phaseolin (Table 2), contrasted with Allozyme Group B for many traits. It exhibited darker hypocotyl and flower pigmentation, later flowering, generally Growth Habit II, a shorter fifth internode, flowering at higher nodes, larger leaflets but smaller seeds, and increased incidence of resistance to BCMV and susceptibility to anthracnose. This group originates in the humid lowlands of Mexico, Central America, and Brazil and comprises tropical blacks and the mulatinhos, including landraces such as 'Rio Tibagi', 'Negro Argel', 'Pata de Zope', 'Rabia de Gato', and 'Porrillo Sintético'. These landraces correspond to race Mesoamerica of Singh et al. (1989). Compared to Cluster E, members of Cluster C flower earlier, have longer pods and seeds, have fewer seeds per plant, and exhibit a lower frequency of BCMV resistance.

A second major group is characterized by allele Me^{102} (Allozyme Group B) and a high frequency of Type Sd phaseolin (Table 3). It exhibits (compared to other Mesoamerican groups) reduced hypocotyl and flower pigmentation, Growth Habits III or IV, a longer fifth internode, flowering starting on lower nodes, earlier flowering, smaller leaflets, larger seeds,and a higher incidence of susceptibility to BCMV and resistance to anthracnose. Members of this group originate in the arid highlands of central and northern Mexico and include landraces such as 'Pinto', 'Bayo Río Grande', 'Durango 222', 'G 2618', 'Garbancillo Zarco', and 'Garrapato' (all from Mexico). This group corresponds to the race 'Durango' suggested by Singh et al. (1989).

The third group, characterized by the Me^{100} allele (Allozyme Groups A and D) and a high frequency of the Type S phaseolin (Table 2, with exceptions noted below), is a more heterogeneous group. Its core appears to consist of landraces with either Type III or Type IV growth habits and small- to medium-sized seeds. Most members originate in the humid highlands of Mexico, Central America, and the northern Andes. Landraces represented in this group include 'Flor de Mayo', 'Azufrado', 'Canario', 'Apetito', 'Conejo', and 'Cacahuate Criollo'. This group may correspond to the race Jalisco proposed by Singh et al. (1989) It could be further subdivided on the basis of additional markers such as Type B phaseolin (humid highlands of Central America and Colombia; Koenig et al., 1990) or $Mdh\text{-}2^{102}$ (Allozyme Group D; see below). Possible hybridizations with other groups is suggested by the presence of some landraces such as 'Pinto Texano' and 'Bayo' (Group B) and 'Carioca', 'Rim de Porco', and 'Mulatinho de Irece' (Groups C or E).

Allozyme Group D is separated from other groups along the second canonical variable. It consists of small- to medium-seeded landraces of growth habit IV originating in the humid highlands of Mexico, such as 'Charrito', 'Rosa de Castilla', 'Colorado de Teopisca', and 'Naranja Coral'. Compared to cultivars of other allozyme groups, cultivars in this group are later, have larger leaflets and a higher seed yield per plant. This group is characterized by the presence of the $Mdh\text{-}2^{102}$ allele (Singh et al., 1991). This allele was observed only in some wild *P. vulgaris* populations from Mexico (Koenig and Gepts, 1989), suggesting that Group D represents cases of introgression from wild common beans.

Canonical discriminant analysis of morphological and agronomic traits distinguishing Andean allozyme groups did not reveal a canonical correlation that was significantly different from zero. However, analyses of variance revealed some significant differences among the various groups. A comparison between Groups G and I, which contain a large number of landraces, reveals that those of Group G tend to have a bush (as opposed to climbing) growth habit, tend to have rel-

30

atively smaller and more elongated leaflets, and have more elongated seeds, when compared to landraces of Group I.

DISCUSSION

Results from the three methods of multivariate analyses consistently identified fifth internode length, number of nodes to first flower, leaflet size, and seed weight as major traits separating cultigens of Andean and Mesoamerican origin. The Andean germplasm possessed a higher number of nodes to first flower, and larger leaflets and seeds than Mesoamerican germplasm. Our results confirm those obtained previously by Gepts et al. (1986), Gepts and Bliss (1986), and Sprecher and Isleib (1989). In addition, further subdivisions within the Mesoamerican and Andean groups were possible. The identification of these groups of cultivars is made possible by the existence of strong associations between the traits considered in this study. Such associations can arise through multilocus genetic associations or developmental correlations. The former are favored by the predominantly autogamous reproductive system of common bean and by geographical (e.g., Mesoamerican vs. Andean) and ecological (e.g., arid highlands vs. humid lowlands in Mesoamerica) isolations. The latter are exemplified by the relationships between yield and its components.

The approach we have followed to identify subgroups or races within Mesoamerican and Andean germplasm marks a departure from the traditional way of analyzing genetic diversity in cultivated common bean, which relied exclusively on morphological traits (Evans, 1976; Hidalgo, 1988; Vanderborght, 1986). Our approach relied on a prior classification of cultigens using molecular markers (phaseolin and isozymes), followed by a corroborating analysis of morphological and agronomic traits. This strategy was first used to determine that common bean cultivars of Andean origin, on the average, had larger seeds than Mesoamerican common bean cultivars (Gepts et al., 1986). More recently, a preliminary report outlines morphological differences, as detected by principal component analysis, in Malawian common bean cultivars of Mesoamerican and Andean origin, as determined by allozyme analysis (Sprecher and Isleib, 1989). The importance of the prior (molecular classification is twofold. First, molecular markers are more useful than morpho-agronomic traits in determining the genetic relationships among accessions. Because morpho-agronomic traits are phenotypic traits, accessions may be similar morphologically, yet be distant genetically. For example, molecular markers show that Growth Habit I cultivars had two independent origins, one in Mesoamerica and the other in the Andes. Second, the complex genetic control of many of the morpho-agronomic traits often precludes the determination of the precise genotype underlying each phenotype. The complexity is caused by the number of genes involved, the presence of pleiotropy and epistasis, and environmental influence. Because molecular markers are mostly devoid of these disadvantages, several authors (e.g., Tanksley, 1983) have argued in favor of marker-based selection in breeding programs. By analogy, we show here that a marker-based classification of genetic diversity is an informative approach towards our understanding of the genetic structure of a species.

The amount of information provided by this marker-based approach will depend on the type and number of markers, and their linkage relationships. Phaseolin type may be more useful in this respect than any individual allozyme because of its molecular complexity (Gepts, 1990). Two accessions may show the same allozyme because of common ancestry, independent mutation, or recombination. On the other hand, the molecular complexity of phaseolin makes it unlikely that the same phaseolin type would have appeared repeatedly through independent mutation. With either type of marker, the availability of several markers, preferably unlinked, may increase the power of the analysis because it would detect recombination. In common bean grown in Latin America, the importance or recombination especially between the Mesoamerican and Andean gene pool is probably minor because of the self-pollinating nature of the species, the geographic isolation of the two gene pools, and their reproductive isolation (Gepts and Bliss, 1985).

Finally, our analysis also reveals that within the allozyme groups, considerable morphological variation can exist with respect to growth habit and seed type, in spite of the uniformity at the allozyme level. Crosses may be attempted between two genotypes which, based on morphological arguments, may appear to be very diverse in order to maximize the potential gain from selection in the progeny. Our phaseolin and allozyme data indicate that in some cases these crosses may involve materials that are actually genetically closely related. Conversely, genotypes with similar morphological traits may be evolutionarily distant (within the primary gene pool) as exemplified by the existence of Growth Habit I cultivars in both the Mesoamerican and Andean gene pools. The benefit of molecular markers is that they characterize accessions at the genotypic level and allow us to evaluate evolutionary relationships and levels of genetic divergence more accurately than morphological markers. It is suggested here that gene banks should integrate morphological, agronomical, and biochemical and molecular evaluations of their genetic resources because the different types of traits provide complementary information.

ACKNOWLEDGMENTS

This work was jointly supported by the International Center for Tropical Agriculture (Cali, Colombia) and the US AID. We thank Gisela Arndt for useful comments, Bill Hardy for editorial assistance, and Aracelly Fernández for typing assistance.

REFERENCES

Afifi, A.A., and V. Clark. 1984. Computer-aided multivariate analysis. Van Nostrand Reinhold, New York.

Brown, J.W.S., and Y. Ma, F.A. Bliss, and T.C. Hall. 1981. Genetic variation in the subunits of globulin-1 storage protein of French bean. Theor. Appl. Genet. 59:83–88.

Evans, A. 1976. Beans. p. 168–172. *In* N.W. Simmonds (ed.) Evolution of crop plants. Longman, London.

Gepts, P. 1990. Biochemical evidence bearing on the domestication

of *Phaseolus* (Fabaceae) beans. Econ. Bot. 44 (3 Suppl.):28–38.
Gepts, P., and F.A. Bliss. 1985. F_1 hybrid weakness in the common bean: Differential geographic origin suggests two gene pools in cultivated bean germplasm. J. Hered. 76:447–450.
Gepts, P., and F.A. Bliss. 1986. Phaseolin variability among wild and cultivated common beans from Colombia. Econ. Bot. 40:469–478.
Gepts, P., T.C. Osborn, K. Rashka, and F.A. Bliss. 1986. Electrophoretic analysis of phaseolin protein variability in wild forms and landraces of the common bean, *Phaseolus vulgaris*: Evidence for multiple centers of domestication. Econ. Bot. 40:451–468.
Hedrick, U.P. 1931. The vegetables of New York: The beans of New York. New York (Geneva) Agric. Exp. Stn. Rep. 1(2):1–110.
Hidalgo, R. 1988. The *Phaseolus* world collection. p. 67–90. *In* P. Gepts (ed.) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, the Netherlands.
Koenig, R., and P. Gepts. 1989 Allozyme diversity in wild *Phaseolus vulgaris*: Further evidence for two major centers of genetic diversity. Theor. Appl. Genet. 44:809–817.
Koenig, R., S.P. Singh, and P. Gepts. 1990. Novel phaseolin types in wild and cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ. Bot. 44:50–60.
Leakey, C.L.A. 1988. Genotypic and phenotypic markers in common bean. p. 245–327. *In* P. Gepts (ed.) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, the Netherlands.
Polignano, G.B., and P.L. Spagnoletti-Zeuli. 1985. Variation and covariation in *Vicia faba* L. populations of Mediterranean origins. Euphytica 34:659–668.
SAS Institute. 1985. SAS user's guide: Statistics. 5th ed. SAS Institute, Cary, NC.
Singh, S.P. 1982. A key for identification of different growth habits of *Phaseolus vulgaris* L. Annu. Rep. Bean Improv. Coop. 25:92–95.
Singh, S.P. 1988. Gene pools in cultivated dry bean. Annu. Rep. Bean Improv. Coop. 31:180–182.
Singh, S.P. 1989. Patterns of variation in cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ. Bot. 43:39–57.
Singh, S.P., D.G. Debouck, and P. Gepts. 1989. Races of common bean, *Phaseolus vulgaris* L. p. 75–89. *In* S. Beebe (ed.) Current topics in breeding of common bean. Working document 47. Bean Program, CIAT, Cali, Colombia.
Singh, S.P., R. Nodari, and P. Gepts. 1991. Genetic diversity in cultivated common bean: I. Allozymes. Crop Sci. 31:19–23 (this issue).
Sprecher, S.L., and T.G. Isleib. 1989. Morphological and phenological diversity between bean gene pools. Annu. Rep. Bean Improv. Coop. 32:54–55.
Tanksley, S.D. 1983. Molecular markers in plant breeding. Plant Mol. Biol. Rep. 1:3–8.
Vanderborght, T. 1986. L'étude de la variabilité chez le haricot commun (*Phaseolus vulgaris* L.) par l'utilisation de méthodes statistiques multivariées appliquées à une banque de données. Ph.D. diss. Fac. des Sci. Agron., Gembloux, Belgium (Diss. Abstr. 47:3190B–3191B).
Van Schoonhoven, A., and M.A. Pastor Corrales. 1987. Standard system for the evaluation of bean germplasm. CIAT, Cali, Colombia.
Weeden, N.F. 1984. Linkage between the gene coding the small subunit of ribulose bisphosphate carboxylase and the gene coding malic enzyme in *Phaseolus vulgaris*. Annu. Rep. Bean Improv. Coop. 27:123–124.

32

RACES OF COMMON BEAN (*PHASEOLUS VULGARIS*, FABACEAE)[1]

SHREE P. SINGH, PAUL GEPTS, AND DANIEL G. DEBOUCK

Shree P. Singh *(Bean Program, CIAT, AA 6713, Cali, Colombia),* **Paul Gepts** *(Department of Agronomy and Range Sciences, University of California, Davis, CA 95616-8515, USA), and* **Daniel G. Debouck** *(IBPGR, Via delle Sette Chiese 142, 00145 Rome, Italy).*

RACES OF COMMON BEAN (*PHASEOLUS VULGARIS*, FABACEAE). *Economic Botany 45(3):379–396. 1991. Evidence for genetic diversity in cultivated common bean* (Phaseolus vulgaris) *is reviewed. Multivariate statistical analyses of morphological, agronomic, and molecular data, as well as other available information on Latin American landraces representing various geographical and ecological regions of their primary centers of domestications in the Americas, reveal the existence of two major groups of germplasm: Middle American and Andean South American, which could be further divided into six races. Three races originated in Middle America (races Durango, Jalisco, and Mesoamerica) and three in Andean South America (races Chile, Nueva Granada, and Peru). Their distinctive characteristics and their relationships with previously reported gene pools are discussed.*

Razas de Fríjol Común (*Phaseolus vulgaris*, Fabaceae). *Se presenta una revisión sobre la evidencia de variabilidad genética en el fríjol cultivado* (Phaseolus vulgaris). *De acuerdo con los análisis estadísticos multivariados de datos morfológicos, agronómicos y moleculares y con información adicional disponible sobre variedades criollas de América Latina que representan varias regiones ecológicas y geográficas de sus centros primarios de domesticación en las Américas, se establece la existencia de los dos grupos principales de germoplasma: los de Mesoamérica y de los Andes suramericanos; los cuales pueden ser divididos en seis razas. Tres razas se originaron en Mesoamérica (razas Durango, Jalisco y Mesoamérica) y tres en los Andes suramericanos (razas Chile, Nueva Granada y Perú). Se discuten sus características distintivas y sus relaciones con otros acervos de genes reportados anteriormente.*

Key Words: *Phaseolus vulgaris*, domestication, crop plant evolution, genetic diversity, races of common bean, gene pools.

All available archaeological, morphological, and molecular (phaseolin seed protein and allozymes) evidence suggests that cultivated common bean (*Phaseolus vulgaris* L., Fabaceae) evolved from its closest relative, wild common bean (Brücher 1988; Delgado Salinas et al. 1988; Gentry 1969; Gepts 1988a,b, 1990; Gepts and Bliss 1986; Gepts and Debouck 1991; Gepts et al. 1986; Kaplan 1965, 1981; Kaplan and Kaplan 1988; Miranda Colín 1967). Although some exceptions occur (Brunner and Beaver 1988; Wells et al. 1988), both wild populations and cultivated forms are self-pollinating (Ortega V. 1974; Pereira Filho and Cavariani 1984; Rutger and Beckham 1970; Stoetzer 1984; Tucker and Harding 1975; Vieira 1960) and are diploid, with 2n = 2x = 22 chromosomes. They hybridize with each other easily, producing viable and fertile individuals (Harmsen et al. 1987; Koenig and Gepts 1989b; Motto et al. 1978; Weiseth 1954).

Wild common bean is distributed from northern Mexico to northwestern Argentina (Brücher 1988; Burkart and Brücher 1953; Debouck and Tohme 1989; Delgado Salinas et al. 1988; Gentry 1969; Harlan 1975; Koenig and Gepts 1989a; Miranda Colín 1967; Nabhan et al. 1986; Toro et al. 1990). Marked differences are found in morphological (Debouck and Tohme 1989; Delgado Salinas et al. 1988; Gepts and Debouck 1991) and molecular characteristics (Koenig and Gepts 1989a; Koenig et al. 1990) among the wild populations from the two extremes of their geographical distribution. Wild beans from Costa Rica, Panama, Venezuela, Colombia, Ecuador, and northern Peru possess intermediate traits

[1] Received 10 July 1990; accepted 8 April 1991.

Economic Botany 45(3) pp. 379–396. 1991
© 1991, by The New York Botanical Garden, Bronx, NY 10458 U.S.A.

(Brücher 1988; Debouck and Tohme 1989; Koenig and Gepts 1989a). Debouck and Tohme (1989) and Gepts and Debouck (1991) have summarized variations that occur in seed size, phaseolin seed protein patterns, hypocotyl texture, bracteole size and shape, and days to flowering among wild bean populations from different regions of Latin America. Based on differences in allelic frequencies for nine polymorphic allozyme loci, Koenig and Gepts (1989a) identified five major geographical groups, namely, Mexico, Central America, Colombia, southern Peru, and Argentina, in wild populations. Fewer samples of wild beans compared to cultivated types (< 1000 versus > 35 000) are currently available in gene banks (Hidalgo 1991; Toro et al. 1990). As new accessions of wild common bean are examined, especially from regions where collections have not yet been made, a more complete picture of the organization of genetic diversity in wild bean populations is likely to emerge.

Wild bean populations have undergone major changes in morphological, physiological, biochemical, and genetic characteristics under domestication (Brücher 1988; Evans 1980; Gepts and Debouck 1991; Kaplan 1965; Smartt 1988, 1990a,b). The most apparent changes include the appearance of indeterminate and determinate upright bush growth habits; gigantism of leaf, pod, and seed characteristics; suppression of explosive pod dehiscence; loss of seed dormancy; appearance of a vast variety of seed sizes, shapes, and colors; and selection for insensitivity to photoperiod.

Polymorphism for morphological, physiological, agronomic, and molecular characteristics also exists among cultivated forms. Their patterns of variation seem to parallel those found in wild populations along their geographical range of distribution in the Americas (Debouck and Tohme 1989; Gepts et al. 1986; Koenig et al. 1990; Singh et al. 1991a,b; Vanderborght 1987). This is probably because different wild bean populations participated in the initial domestications in different regions, thus supporting the earlier proposed hypothesis of multiple domestication centers in Middle America and in Andean South America (Gepts 1988a,b, 1990; Gepts and Debouck 1991, Gepts et al. 1986; Harlan 1975, Heiser 1979, Kaplan 1981, Vargas et al. 1990). Morphological characteristics and phaseolin seed protein and allozyme patterns that facilitate characterization of and separation between cultivated common beans of Middle American and Andean South American origins are summarized in Table 1.

Multiple domestications from diverged ancestral *P. vulgaris* populations, the self-pollinating nature of the species, and geographical or ecological separation over millennia must have permitted multitrait genetic and developmental associations to appear and, consequently, have led to the evolution of distinct groups of related cultivated populations of common bean (Singh et al. 1991a).

In this article, we review the evidence accumulated in recent years on patterns of genetic diversity within cultivated *P. vulgaris*. This evidence consists of data on molecular diversity (phaseolin: Gepts and Bliss 1986; Gepts et al. 1986; Koenig et al. 1990; allozymes: Koenig and Gepts 1989a; Singh et al. 1991b; Sprecher 1988; mitochondrial RFLPs: Khairallah et al. 1990), morphological diversity (Evans 1973, 1980; Singh 1989; Singh et al. 1990, 1991a; Urrea and Singh 1991; Vanderborght 1987), breeding behavior (Nienhuis and Singh 1988), and reproductive isolation (Gepts and Bliss 1985; Singh and Gutiérrez 1984; Singh and Molina 1991; Vieira et al. 1989). In particular, we identify specific associations of molecular and morphological markers, agronomic traits, reproductive isolation factors, ecological adaptation, and geographical distribution that characterize "races" of cultivated common bean.

The term race is used here to denote a group of related landraces (Gepts 1988a) within the Middle American and Andean components of the species. Members of each group share certain distinctive morphological, agronomic, physiological, and biochemical or molecular traits, and differ from other groups in allelic frequencies of the genes controlling differences in those traits. Each of the races defined in this article consists of one or more gene pools defined earlier by Singh (1988, 1989, 1991b) on the basis of morphological traits and breeding behavior. A tentative description of these races was presented in a bean breeders' workshop (Singh et al. 1989b).

Materials and Methods

Plant Materials

Of 306 cultivated landraces studied, 72 were from Mexico, 10 from Guatemala, 2 from Nicaragua, 5 from El Salvador, 1 from Honduras, 2 each from Costa Rica and the Dominican Republic, 97 from Colombia, 44 from Ecuador, 31 from Peru, 17 from Chile, 1 each from Bolivia and Argentina, and 21 from Brazil. Seed samples of all landraces were obtained from the *Phaseolus* germplasm bank of the Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.

TABLE 1. PRINCIPAL CHARACTERISTICS OF CULTIVATED COMMON BEAN (*PHASEOLUS VULGARIS* L.) FROM MIDDLE AMERICAN AND ANDEAN SOUTH AMERICAN DOMESTICATION CENTERS.

Characteristic	Domestication center	
	Middle America	Andean South America
Shape of terminal leaflet of trifoliolate leaf	Ovate, cordate	Hastate or lanceolate, rhombohedric
Straight leaf hairs	Sparse, short	Dense, long
Length of the fifth internode	Short	Long
Pod-bearing inflorescence	Multinoded	Single-noded
Shape of bracteole	Cordate, ovate	Lanceolate, triangular
Base of standard (banner petal)	Striped	Smooth
Pod beak position	Placental (dorsal suture)	Between placental and ventral sutures
Seed size	Small, medium	Large
Phaseolin seed protein patterns	S, Sb, Sd, B	T, C, H, A
Allozyme	*Diap-1*^{95}, *Lap-3*^{100}, *Rbcs*^{100}, *Skdh*^{103}	*Diap-1*^{100}, *Lap-3*^{103}, *Rbcs*^{98}, *Skdh*^{100}

Field Evaluations

All 306 landraces were evaluated at Palmira (1000 m altitude; mean growing temperature, 24°C; Mollisol soil) and Popayán (1750 m altitude; mean growing temperature, 18°C; Inceptisol soil), Colombia, in 1987–1988. Each plot consisted of a single row, 3 m in length. Spacing between rows was 60 cm and within a row 10 cm. Optimum conditions for crop growth and development were provided. Data on five competitive plants in each plot were recorded for hypocotyl, flower, and pod color; length and width of the central leaflet of the fully developed trifoliolate leaf; length of the fifth internode; growth habit; number of nodes to the first flower; days to flower and maturity; bracteole size and shape, absence or presence of stripes at the external base of flower standard; origin of pod beak; dry seed length and height; seed shape; seed shininess; 100-seed weight; and seed yield. Additionally, evaluations for reaction to bean common mosaic virus, leafhoppers, common bacterial blight, angular leaf spot, and anthracnose were made in separate complementary nurseries either at Palmira, Popayan, or Santander de Quilichao (990 m altitude; mean growing temperature, 24°C; Oxisol soil), Colombia, in 1987–1988. For details on materials and methods, readers should refer to Singh et al. (1991a,b).

Criteria Used for Race Identification

Shape and size of leaflet. The shape of the central or terminal leaflet of the fully developed trifoliolate leaves can be cordate, ovate, rhombohedric, and lanceolate or hastate (Fig. 1). These differences also occur in wild bean populations (Urrea and Singh 1991). Similarly, large differences in the width and length of the central leaflet are found.

Leaf hairiness. Straight leaf hairs can be short and sparse versus long and dense. These differences do not refer to the hooked hairs which are found in all *Phaseolus* species (Maréchal et al. 1978).

Length of internodes. This is the length of the fifth internode from the cotyledonary node measured on the main stem acropetally.

Number of nodes to flower. This refers to the number of nodes on the main stem (including the nodes of cotyledonary and primary simple leaves) at the axil where the first raceme to flower appears.

Shape and size of bracteoles. According to Singh et al. (1990), flower bracteoles can be small, medium, or large (i.e., exceeding the calyx length). The shape can be cordate, ovate, lanceolate, or slender and triangular (Fig. 2).

Inflorescence. The axillary and terminal fruit-

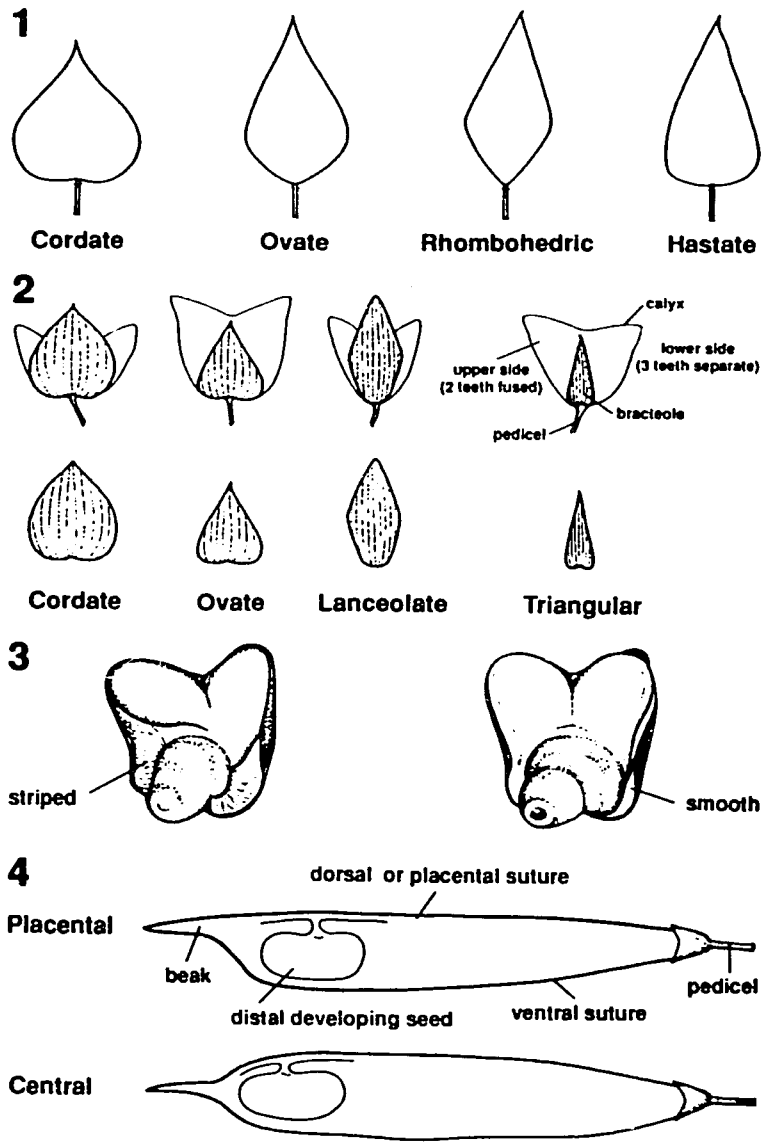


Fig. 1-4. Fig. 1. Variation in shape of the central leaflet of the trifoliolate leaves of common bean. **Fig. 2.** Principal bracteole types found in common bean. **Fig. 3.** Striped versus smooth basal outer surface of flower standard (banner petal) of common bean. **Fig. 4.** Placental versus central pod beak position in cultivated common bean.

bearing inflorescence can be multinoded (≥ 2) or single-noded. In Middle American germplasm, ontogenesis seems to be strong and flowering somewhat more synchronous through the raceme. This, together with more even distribution of photosynthates within the raceme, probably allows it to bear pods on several nodes. In the Andean South American accessions, instead, on-

togenesis is weaker, and the first two flowers of the multiflowered inflorescence will develop into pods at the first node before further development of pods at other nodes of the raceme, often forcing the terminal part of the raceme to senesce prematurely.

Standard. The outer base of the standard (banner petal) of the corolla can be striped and colored (often purple or pink) or smooth and greenish (Fig. 3). Similarly, color of the petals can be white, white with pink stripes, white with a pink or purple blotch, pink, or purple with varying intensities.

Pod beak origin. The beak or tip of the pods may extend straight from the placental (dorsal) suture or may have an intermediate or central position between the placental and ventral sutures (Fig. 4).

Days to maturity. This varies from about 50 days to more than 250 days and is highly affected by sensitivity to photoperiod, growing temperature, their interactions, and other factors.

Seed size and shape. Seed weight (hence, size) can vary from <15 to >90 g/100 seeds. Accessions were grouped into small- (<25 g), medium- (25–40 g), and large-seeded (>40 g/100 seeds) germplasm by Voysest (1983). Seed shape depends upon the length, height, and width of fully mature and dry seed. When the three dimensions are considered jointly, seed shape can be round, oval, rhombohedric, kidney, or cylindrical (Fig. 5). Also, large variations in seed colors and their spotting, striping, and speckling patterns are found (Leakey 1988; Voysest and Dessert 1991).

Growth habit. Singh (1982) used the type of terminal bud (vegetative versus reproductive), stem stiffness (strong versus weak), twining ability (absent, weak, or strong), and distribution of pod load or fruiting patterns (basal, along the entire length, or largely in the upper portion of the plant) to characterize the four major growth habits in common bean: determinate upright (I); indeterminate upright (II); indeterminate, weak-stemmed, prostrate nonclimbing or semiclimbing (III); and indeterminate (or determinate), weak-stemmed, with long guides or leaders (i.e., elongated terminal internodes, with twining ability) and strong climbing ability (IV). Growth habits of common bean are thoroughly discussed by Debouck (1991) and Debouck and Hidalgo (1986).

Phaseolin seed protein. Phaseolin seed protein was analyzed at the University of California, Davis, California, in 1986–1987 by sodium dodecyl sulfate polyacrylamide gel electrophoresis as described by Brown et al. (1981) and Ma and Bliss (1978). For details, readers should refer to Koenig et al. (1990) and Singh et al. (1991a). The cultivated landraces from Mexico, Central America, Colombia, Venezuela, and Brazil usually carry S, Sb, Sd, and B phaseolin seed protein types, and those from Andean South America (including some from Colombia) usually possess T, C, H, and A phaseolin types (Gepts and Bliss 1986; Gepts et al. 1986; Koenig et al. 1990; Vargas et al. 1990). In general, there is a much larger variability in phaseolin protein patterns in wild populations, only a small proportion of which is found in landraces (Gepts et al. 1986; Koenig et al. 1990; Vargas et al. 1990).

Allozymes. Allozymes were analyzed at the University of California, Davis, California, in 1986–1987 as described by Koenig and Gepts (1989a,b) and Singh et al. (1991b). For details, readers should refer to Singh et al. (1991b). Seven polymorphic allozymes, namely, ribulose biphosphate carboxylase (small subunit; RBCS), shikimate dehydrogenase (SKDH), peroxidase (PRX), malic enzyme (ME), malate dehydrogenase (MDH), diaphorase (DIAP), and leucine aminopeptidase (LAP) were surveyed. The MDH and DIAP enzyme systems each had two loci. Allozyme loci and alleles were designated as in Koenig and Gepts (1989a). The multivariate statistical analyses of these allozymes also supported the existence of the Middle American and Andean groups of common beans (Singh et al. 1991b; Sprecher 1988). Allelic frequencies at the allozyme loci distinguish five subgroups within Middle American and four subgroups within Andean landraces. For example, approximately two-thirds of the Middle American accessions showed the Me^{100} allele, whereas one-third of the Middle American accessions carried the Me^{102} allele, and 96% of the Andean landraces had the Me^{98} allele (Singh et al. 1991b).

Multivariate Statistical Analyses

Multivariate statistical analyses (principal component, discriminant, and canonical discriminant analyses) were applied to study the organization of genetic diversity among landraces (Singh et al. 1991a,b). Principal component analysis does not require prior classification of entries and was most useful in the identification of traits distinguishing Middle American and

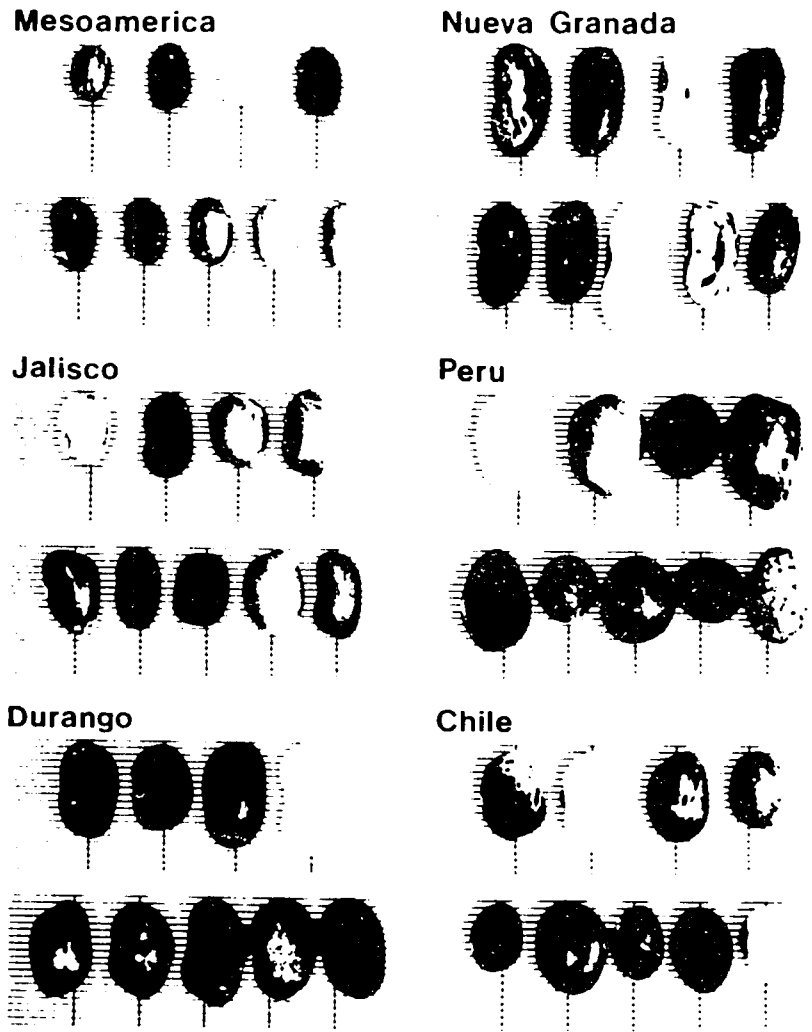


Fig. 5. Characteristics of dry seeds of different races of cultivated common bean.

TABLE 2. CORRELATION COEFFICIENTS† BETWEEN ORIGINAL AND CANONICAL VARIABLES IN A CANONICAL DISCRIMINANT ANALYSIS OF CULTIVATED COMMON BEAN FROM LATIN AMERICA USING PHASEOLIN AND ALLOZYME CLUSTER MEMBERSHIP AS CLASSIFICATION CRITERIA.

	Middle American landraces				Andean landraces	
	Phaseolin		Allozyme		Phaseolin	
Trait	FCV	SCV	FCV	SCV	FCV	SCV
Hypocotyl color	−0.62	0.52	−0.97	0.00	0.78	0.63
Flower color	−0.08	0.83	−0.93	−0.05	−1.00	0.10
Days to flowering	0.87	0.39	−0.75	0.55	1.00	0.06
Days to maturity	0.95	−0.05	0.13	0.95	0.89	−0.45
Fifth internode length	−0.51	−0.86	0.79	0.10	0.92	0.39
Number of nodes to first flower	0.90	−0.04	−0.77	0.24	−0.91	−0.42
Leaflet length	0.85	0.44	−0.27	0.91	1.00	0.07
Leaflet width	0.75	0.52	−0.24	0.89	0.98	0.21
Seed length	0.03	−0.94	0.80	−0.08	−0.70	0.71
Seed height	0.02	−0.98	0.89	0.03	0.58	0.05
Seed shininess	0.30	0.55	0.65	0.32	−0.15	0.99
Hundred-seed weight	0.06	−0.96	0.92	0.15	0.93	0.38
Yield/plant	0.96	−0.20	0.33	0.86	0.60	0.80
Leafhoppers	—	—	0.48	−0.07	—	—
Bean common mosaic virus	—	—	0.84	−0.19	—	—
Common bacterial blight	—	—	0.40	0.70	—	—
Anthracnose	—	—	−0.79	−0.30	—	—
Angular leaf spot	—	—	−0.86	0.50	—	—

† Adopted from Singh et al. 1991a. FCV and SCV = first and second canonical variables, respectively.

Andean genotypes (Table 1). Discriminant and canonical discriminant analyses require prior classification for which molecular markers were used as described in Singh et al. (1991a). The two analyses were useful mostly to distinguish races within the Middle American and Andean groups. Correlation coefficients between some original and canonical morphological, physiological, and agronomic variables in a canonical discriminant analysis using phaseolin and allozyme groups as prior classification criteria are given in Table 2 (modified from Singh et al. 1991a).

RESULTS

DIFFERENCES BETWEEN MIDDLE AMERICAN AND ANDEAN CULTIVATED LANDRACES

Through multiple domestications from an already diverged wild ancestor, two major groups of landraces of common bean have appeared: Middle American and Andean South American. These can be distinguished by molecular markers (phaseolin, allozymes) and vegetative and reproductive traits (Table 1). Molecular markers and/or other ancestral traits (i.e., traits of wild beans still found in cultivated landraces; Singh et al. 1989b) served as indicators of evolutionary origin. For example, it was possible to distinguish the two groups of landraces by the shape of the terminal leaflet of the trifoliolate leaf (Fig. 1), the density and length of the straight hairs, the shape and size of the flower bracteoles (Fig. 2), the presence or absence of stripes on the outer base of the standard petal (Fig. 3), the number of nodes of pod-bearing inflorescence, the pod beak position (Fig. 4), and the size and shape of dry seeds (Fig. 5).

Middle American Races

Mesoamerica. This race includes small-seeded (<25 g/100 seed) landraces of all seed colors and growth habits (Table 3 and Fig. 5). Leaf size and internode length are small, intermediate, or large. The group is often characterized by an ovate, cordate, or hastate terminal leaflet of the trifoliolate leaves and large, broad cordate or lanceolate bracteoles. Flower standard often possesses marked stripes at the outer base. Color of petals can be white, white with pink stripes, or purple. Inflorescences are multinoded. Pods are 8–15 cm long, slender, fibrous or parchmented, and easy

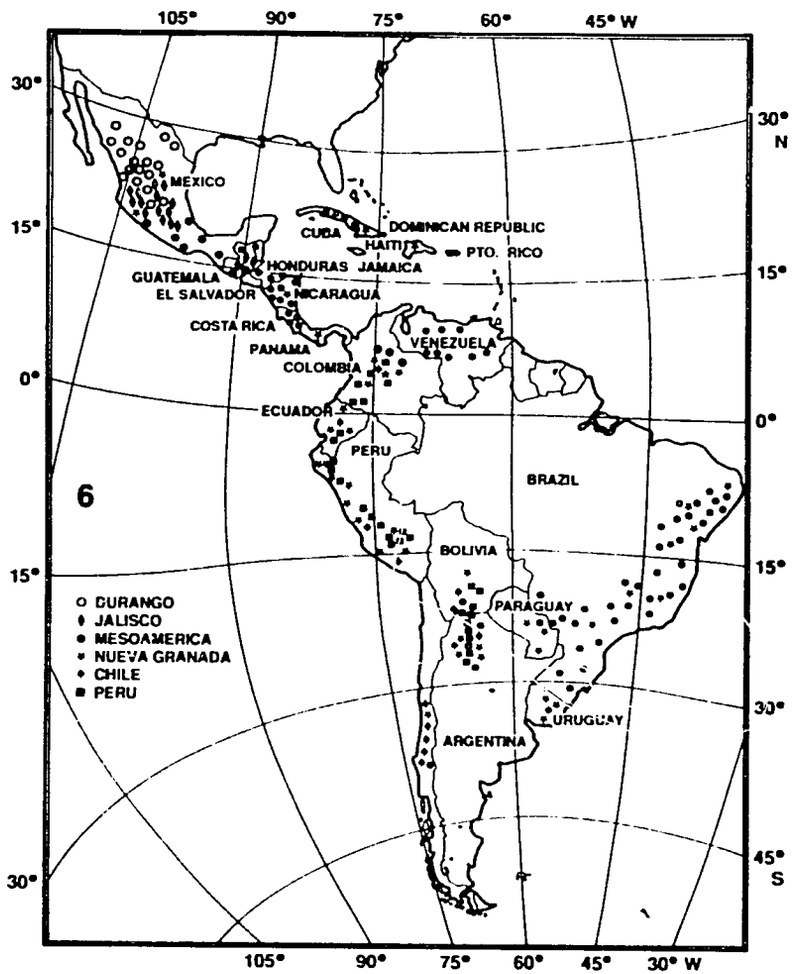


Fig. 6. Distribution of races of cultivated common bean in Latin America. Races Mesoamerica, Jalisco, and Durango are from Middle America and races Nueva Granada, Peru, and Chile are from the Andes.

to thresh; they contain six to eight seeds. Phaseolin types are predominantly S, but can also be Sb and B. The race is distributed throughout the tropical lowlands and intermediate altitudes of Mexico, Central America, Colombia, Venezuela, and Brazil (Fig. 6). Based on growth habit, number of nodes to flower, and days to flowering and maturity, phaseolin types, and/or differences in allelic frequencies at some allozyme loci, two subgroups could be identified within this race. One group was represented by indeterminate, erect, type II landraces such as 'Rio Tibagi' (G 4830), 'Porrillo Sintético' (G 4495), and 'H6 Mulatinho' (G 5059), all of which had the *Diap-2*[105] allozyme allele (cluster E of Singh et al. 1991b). The other subgroup was formed by indeterminate, prostrate, type III early maturing landraces including 'Rabia de Gato' (G 3184), 'Negro Argel' (G 3758), and 'Pitouco' (G 21972), characterized by the *Me*[98] allele (cluster C of Singh et al. 1991b). The latter group appears to be more primitive than the former, but both were con-

TABLE 3. SOME CHARACTERISTICS OF THE RACES OF CULTIVATED COMMON BEAN (*PHASEOLUS VULGARIS* L.), THEIR RELATIONSHIPS WITH GENE POOLS, AND DISTRIBUTION IN THE PRIMARY CENTERS OF DOMESTICATION IN THE AMERICAS.

Race	Seed: Size[a]	Seed: Shape	Bracteole	Phaseolin[b]	Allozyme[c]	Growth habit[d]	Gene pool[e]	Distribution[f]	Landraces
Middle American									
Mesoamerica	Small	Cylindrical, kidney, oval	Large cordate	S, Sb, B	*Me*[98] or *Diap-2*[105]	Determinate I; indeterminate II, III, and IV	1, 2, 3, 4	Lowlands of Latin America	Brazil 2, Jamapa, Porrillo, Rojo de Seda, Mulatinho, Magdalena 3
Durango	Medium	Rhombohedric	Small ovate	S, Sd	*Me*[102]	Indeterminate III	5	Semiarid highlands of Mexico	Pinto, Morado de Agua, Bayo, Ojo de Cabra
Jalisco	Medium	Oval, round, cylindrical	Medium lanceolate	S	*Me*[100] or *Mdh-2*[102]	Indeterminate IV	6	Humid highlands of Mexico	Garbancillo Zarco, Frijola, Flor de Mayo, Apetito, Conejo
Andean South American									
Nueva Granada	Medium and large	Cylindrical, kidney	Small lanceolate or triangular	T	*Rbsc*[100] or *Me*[100]	Determinate I, indeterminate II and III	7, 8, 9	Intermediate altitudes of Andes	Cargabello, Antioquia 8, Radical, Jalo, Bagajo
Chile	Medium	Oval, round	Small triangular	C, H	*Mdh-1*[100]	Indeterminate III	10	Southern Andes	Frutilla, Tórtolas, Coscorrón, Bolita Cristal
Peru	Medium and large	Oval, round	Large lanceolate or cordate	T, C, H	*Mdh-1*[103]	Indeterminate or determinate IV	11, 12	Andean highlands (>2000 m)	Overitos, Caballeros, Nuñas, Bolón Bayo, Bola Roja, Cargamanto

[a] Weight of 100 seeds (hence, size) of small is < 25 g, medium varies between 25 and 40 g, and large is >40 g. [b] According to Gepts and Bliss 1986, Gepts et al. 1986, Koenig et al. 1990. [c] According to Singh de al. 1991b. [d] According to Singh 1982. [e] According to Singh 1989, 1991b. [f] See also Fig. 6.

sidered subgroups within race Mesoamerica and did not warrant independent race status, at least for the time being. Landraces of this race can carry the *Dl-1* gene, leading to F_1 hybrid dwarfism or lethality in the presence of the *Dl-2* gene of Andean origin (Gepts and Bliss 1985; Shii et al. 1980; Singh and Gutiérrez 1984; Vieira et al. 1989), or give rise to deformed leaflets (virus-like symptoms) in segregating generations of interracial crosses (Singh and Molina 1991).

Insensitivity to photoperiod and resistance to bean common mosaic virus (*I1* gene) and tolerance to angular leaf spot, bean golden mosaic virus, high temperatures, moisture stress, and low soil fertility can be found in this race.

Durango. Landraces are predominantly of indeterminate, prostrate growth habit III, which is characterized by relatively small to medium ovate or cordate leaflets, thin stems and branches, short internodes, and fruiting commencing from and concentrated in the basal nodes. These landraces often possess small ovate bracteoles with a pointed tip. Germplasm in this group possesses medium-sized (5–8 cm) flattened pods with four to five flattened rhombohedric seeds of medium size (25–40 g/100 seeds). Seed colors are often tan-like ('bayo'), but may also be yellow, cream, gray, black, white, red, or pink, with or without spots or stripes. Phaseolin types are predominantly S, but some accessions can carry the Sd type. The characteristic allozyme allele is Me^{100} (cluster B of Singh et al. 1991b). The race is distributed in the semiarid central and northern highlands of Mexico and the southwestern USA. The race Durango may include groups 2 and 3 of Bukasov (1930).

This race could be a source of early maturity, drought tolerance, high harvest index, positive general combining ability (GCA) for seed yield (Nienhuis and Singh 1988), and tolerance to some viral diseases (Morales and Singh 1991) and anthracnose. Members of this race are often non-carriers of the *Dl-1* allele (Singh 1990).

Jalisco. This race is often characterized by indeterminate growth habit IV. Plant height can be over 3 m in its natural habitat. The terminal leaflet of trifoliolate leaves is hastate, ovate, or rhombohedric and sometimes relatively large. Stems and branches are weak and have medium-sized or long internodes. Most germplasm from this race possesses medium-sized, cordate, ovate, or lanceolate bracteoles. Fruiting is distributed either along the entire length of the plant or mostly in its upper part. Pods are 8–15 cm long and have five to eight medium-sized seeds, whose shape is round, oval, or slightly elongated and cylindrical or kidney-shaped (Fig. 5). They carry an S phaseolin type and the characteristic allozyme allele is Me^{100} (cluster A of Singh et al. 1991b). Their natural habitat is the humid highlands of central Mexico and Guatemala, where maximum diversity is found. Some small-seeded landraces of growth habit III (e.g., 'Carioca' (G 4017) and 'San Cristóbal 83' (G 17722)) fell into this group, as determined by multivariate statistical analyses. Thus, some heterogeneity was found, and its small-seeded members were included in race Mesoamerica.

High seed yield, positive GCA for yield, high levels of resistance to *Apion* spp. and anthracnose, and tolerance to angular leaf spot and low soil fertility can be found in this race.

South American Races

Nueva Granada. Germplasm is mostly of growth habits I, II, and III with medium (25–40 g/100 seeds) and large seeds (>40 g/100 seeds) of often kidney or cylindrical shapes which vary greatly in color. Leaves are often large with hastate, ovate, or rhombohedric central trifoliolate leaflets and long, dense, straight hairs. Stem internodes are intermediate to long. Bracteoles are small or medium, and ovate, lanceolate, or triangular. Dry pods are fibrous, hard, medium to long (10–20 cm), and leathery, and possess four to six seeds. The pod beak often originates between the placental and ventral sutures. The predominant phaseolin pattern is the T type, combined with $Mdh\text{-}1^{100}$ (cluster G of Singh et al. 1991b) as a characteristic allozyme allele. This race is distributed mostly at intermediate altitudes (<2000 m) of the northern Andes in Colombia, Ecuador, and Peru, but it is also found in Argentina, Belize, Bolivia, Brazil, Chile, Panama, and some Caribbean countries, including the Dominican Republic, Haiti, and Cuba (Lioi et al. 1990). Some landraces of this group may carry the *Dl-2* allele (Gepts and Bliss 1985; Shii et al. 1980; Singh and Gutiérrez 1984; Vieira et al. 1989) or produce deformed leaflets in the segregating generations upon crossing with races of Middle American origin (Singh and Molina 1991).

Insensitivity to photoperiod, early maturity, and resistance to bean common mosaic virus, halo blight, anthracnose, and angular leaf spot can be found in this race.

Chile. Landraces are predominantly of indeterminate growth habit III. These are characterized by relatively small or medium hastate, rhombohedric, or ovate leaves; short internodes; small or medium, and narrowly triangular, spatulate, or ovate bracteoles; light pinkish or white flower; medium-sized (5–8 cm) pods, often with reduced fiber content; and round to oval seeds (three to five per pod). The most common phaseolin patterns are C and H types, and the characteristic allozyme allele is $Mdh\text{-}1^{100}$ (cluster G of Singh et al. 1991b). Morphologically, these landraces largely resemble germplasm from race Durango, except that seeds of race Chile are round or oval, and fruiting is more sparse. In some of the landraces (e.g., 'Coscorrón' (G 4474) and 'Frutilla' (G 5852)), pods exhibit an attractive anthocyanin striping, and in many countries these are harvested for green seeds (green shelled or "granados") before physiological maturity. Some members of this race carry the *Dl-2* allele. This race is distributed in relatively drier regions at lower altitudes in the southern Andes (southern Peru, Bolivia, Chile, and Argentina).

Peru. Key morphological characteristics of germplasm belonging to this race are the large hastate or lanceolate leaves (often basal) and long and weak internodes with either indeterminate or determinate type IV climbing growth habit (Debouck et al. 1988). In its natural habitat, it is always grown in association with maize and other crops. Pods are often long (10–20 cm) and leathery. Fruiting is distributed either along the entire stem length or only in the upper part of the plants. Seeds are large and often round or oval but can also be elongated. Predominant phaseolin protein patterns are C, H, and T types (as one moves from the southern to northern Andes), and the characteristic allozyme allele is $Mdh\text{-}1^{103}$ (cluster I of Singh et al. 1991b). Flower bracteoles are large lanceolate in the 'nuña', 'numia', 'apa', or popping beans from the southern Andes (Debouck 1989) and are large, broad cordate in landraces from northern Peru, Ecuador, and Colombia. The latter group often has an ovate, large terminal leaflet and pod beak originating from the placental suture, resembling Middle American germplasm. This group is highly photoperiod-sensitive and is adapted to moderately wet and cool temperatures often requiring more than 250 days to maturity. The race is distributed from the northern Colombian highlands (>2000 m altitude) to Argentina. Its members in the southern Andes (e.g., 'overitos', 'nuñas', 'tiachos') occur at relatively lower altitudes, are earlier maturing, and possess comparatively smaller seeds with distinctive speckling and spottings.

Discussion

The combination of an autogamous reproductive system, geographical and ecological isolation, and human selection over a long period will increase multilocus associations. This will occur even among genes for apparently unrelated traits such as, for example, morphological characteristics, phaseolin seed proteins, and allozymes, as illustrated by the results for common bean summarized in this article and by Singh et al. (1991a). Although common bean is generally considered to be a self-pollinating species (Ortega V. 1974; Pereira Filho and Cavariani 1984; Rutger and Beckham 1970; Stoetzer 1984; Tucker and Harding 1975; Vieira 1960), the recent studies of Brunner and Beaver (1988) and Wells et al. (1988) indicate that under specific genotype × environment interactions, common bean can exhibit high levels of outcrossing. Whereas outcrossing can account for some of the exceptions to our racial classification (see below) and the occurrence of potential weedy complexes (Debouck et al. 1989), its existence does not necessarily preclude the formation of races as illustrated by the case of outcrossing species such as maize. However, other features such as geographical or ecological isolation and human selection over millennia can also promote multitrait associations.

These associations of traits evolved and preserved over thousands of years translate into a hierarchy of groups of related landraces within this species (Singh et al. 1991a). At the highest level, a major separation between Middle American and Andean landraces results from the divergence prior to domestication of the wild ancestral populations. This is revealed by a wide array of traits, including primarily molecular markers (Gepts and Bliss 1985; Gepts et al. 1986; Khairallah et al. 1990; Koenig et al. 1990; Sprecher 1988), but also morphological traits (Evans 1973, 1980; Singh 1989; Singh et al. 1990, 1991a; Sprecher and Isleib 1989; Vanderborght 1987), reproductive isolation (Evans 1970; Gepts and Bliss 1985; Koenig and Gepts 1989b; Singh and Gutiérrez 1984; Singh and Molina 1991;

Sprecher and Khairallah 1989; Vieira et al. 1989), and geographical and ecological adaptation (Singh 1989; Voysest and Dessert 1991). Because of the level of divergence and the presence of reproductive isolation, one may be tempted to give the status of subspecies to the Middle American and Andean South American groups, although we are not suggesting that here. At the next hierarchical level, molecular, morphological, and physiological markers also help identify a set of subdivisions we proposed to call races, three each in the Middle American and Andean South American groups. These races exhibit a specific geographical and ecological adaptation and distribution (Fig. 6) to the extent that members of all races never grow and develop normally in any single environment simply because they evolved and were selected in different environments over thousands of years of evolution under domestication. In the next step, some races (e.g., races Mesoamerica, Nueva Granada, and Peru) can be further subdivided into gene pools according to their growth habit, crop duration, and yield potential, as described by Singh (1988, 1989, 1991b), primarily to facilitate germplasm improvement suitable for different cropping systems and growing environments.

The identification and use of molecular markers (because of their genotypic nature) and other ancestral traits increase the probability that the races we have identified represent evolutionarily distinct lineages. Similarities identified in this study are, therefore, not merely phenotypic but are based on common ancestry. This advantage of molecular markers and other ancestral traits is particularly useful in common bean germplasm characterization. In both Middle American and Andean groups, races with similar morphology and adaptation have been identified. For example, races Durango and Chile have similar growth habits, have medium-sized seeds, and are adapted to more arid conditions at higher latitudes. Races Jalisco and Peru have predominantly climbing growth habits and are adapted to humid highlands. Yet, molecular markers and other ancestral traits indicate that, in spite of their phenotypic similarities, the races in each of these pairs have different evolutionary origins (i.e., Middle American versus Andean). The apparently independent evolution of a similar phenotype (i.e., growth habit, seed size) adapted to similar environmental conditions in geographically distant areas raises the question whether these constitute ideal plant types providing maximum productivity in these environmental conditions. Further experimentation is needed to distinguish this possibility from others, including coincidence.

Molecular markers were instrumental in identifying the major Middle American and Andean groups in *P. vulgaris* (Gepts 1988a,b, 1990; Gepts and Debouck 1991; Khairallah et al. 1990; Singh et al. 1991b; Sprecher 1988), as well as their direct subdivisions into races (Tables 1, 3). They did not assist definitively in further subdividing races into gene pools due in large part to the lack of variability among landraces within a race for phaseolin and allozymes. More polymorphic markers with high resolution power, at the DNA level (RFLP), need to be identified and utilized for that purpose. Among morphological traits, size and shape of the terminal leaflet of trifoliolate leaves, flower bracteoles, and fully mature dry seed; presence or absence of stripes at the outer base of the standard or banner petal; and pod beak position were among the most useful traits to identify the region of domestication of individual landraces and their assignment to specific races. The resolution power of morphological, agronomic, and physiological traits improved considerably when molecular markers (either phaseolin seed protein patterns or allozymes) were used as a prior classification criterion for multivariate statistical analyses (Singh et al. 1991a).

Because of the strong multilocus associations mentioned above, a good correspondence between classification based on morphological and molecular traits was generally observed (Table 2). Nonetheless, in a few cases (<5%), classifications based on morphological and molecular traits did not match (data not shown). For example, landraces 'de Celaya' (G 13614) and 'Mantequilla' (G 13673) from Mexico and 'Boyacá 17A' (G 8174) and 'Cundinamarca 137' (G 8190) from Colombia all had a Middle American phaseolin type but an Andean morphology. On the other hand, 'Burrito de Enrame' (G 17182) from Ecuador and 'Antioquia 6' (G 3668) and 'Cauca 11' (G 14978) (both from Colombia) possessed Andean phaseolin patterns and a Middle American morphology. Similarly, several landraces belonging to the race Peru from northern Peru, Ecuador, and Colombia (e.g., 'Caballeros' (G 12597), 'Bolón Bayo' (G 12407), 'Mortiño' (G 12709), 'Liborino' (G 11819), 'Sangretoro' (G 5708), 'Blanco Sabanero' (G 12667), and 'Bola Azuay' (G 12229)), possessed an ovate terminal leaflet and large, broad cordate bracteoles with pod beak originating from the placental (dorsal) suture, all characteristics of Middle American germplasm, although they possessed Andean phaseolin, allozyme, and seed characteristics.

Similarly, when landraces were classified based only on their phaseolin seed protein and allozyme patterns, in some cases morphologically distinct landraces belonging to different geographical and ecological regions were grouped together. For example, landraces 'Brazil 2' (G 3807), 'Favinha' (G 5019), 'Rim de Porco' (G 6508), and 'Siempre Asim' (G 21977) from northeastern Brazil; 'Boyacá 22' (G 2511) and 'Magdalena 3' (G 2525) from Colombia; and 'Guatemala 1240' (G 10813), 'Orgulloso' (G 14027), and 'Rabia de Gato' (G 3184) from Central America, all of which possess small seeds and are adapted to lowland tropics, fell into the same group as landraces 'Conejo' (G 22029), 'Flor de Mayo' (G 10945), and 'Frijola' (G 2793), all accessions with medium-sized seeds from the humid highlands of Mexico. Conversely, races Nueva Granada and Mesoamerica each comprise landraces which show contrasting differences for growth habit and maturity. Although Singh (1988, 1989, 1991b) separated them into different gene pools to facilitate breeding, they were grouped together (Table 3) principally because they often possess similar marker traits, are sympatric, and differences in growth habit and maturity are controlled by only a few major genes (Singh 1991a).

The discrepancy between molecular, on the one hand, and morphoagronomic, ecological, and geographical data, on the other, can be attributed to hybridization and recombination or, alternatively, to independent mutations in evolutionarily distinct lineages that lead to materials with identical allozyme profiles. Hybridization is likely to occur more often among races within the Middle American and Andean groups than among races between the two major groups because of reduced geographical isolation among the former. In such circumstances, relatively more weight was given to their classifications based on the morphoagronomic and physiological traits for the following reasons. First, none of the allozyme or phaseolin loci (probably neutral alleles) is known to be associated with adaptive traits, whereas there appears to be strong association of traits such as internode length, leaflet size and shape, seed size and shape, and phenological traits with ecological adaptations (Singh 1989; Voysest and Dessert 1991). These latter groups of traits would have responded positively or negatively to different selection pressures throughout domestication, with severe consequences on the organization of genetic diversity within the species. Second, if the discrepancy results from an outcross, the phenotypic resemblance of the progeny with one of the parents implies that through linkage with morphological traits a substantial part of the genome of the parent may be represented in the progeny; hence, the accession should be included in the parental race it resembles most. An independent mutation affecting an allozyme or phaseolin locus in an accession is unlikely to affect the remainder of the genome; therefore, the genome of this accession should be primarily like that of the original material, and it should not be reclassified.

Some landraces belonging to race Peru and distributed in the northern Andes possessed bracteole, leaflet, pod, and flower characteristics typical of small-seeded race Mesoamerica, but they had Andean seed characteristics and allozyme and phaseolin protein patterns. This might reflect the fact that the wild beans from this region (Colombia, Ecuador, northern Peru) also show a morphology that is intermediate between the northernmost Middle American and southernmost Andean wild bean populations (Brücher 1988; Debouck and Tohme 1989). Moreover, small-seeded cultivated landraces possessing B phaseolin and morphology typical of race Mesoamerica are distributed from northeastern Brazil to Central America, including Colombia where their wild counterpart is also distributed (Gepts and Bliss 1986; Koenig et al. 1990). Moreover, there are strong indications that beans were exchanged in both directions between Middle American and Andean regions by pre-Columbian natives after domestication had begun. For example, in northern Cajamarca in Peru (e.g., Cutervo, Chota, Celendin), beans were brought in frequently from different places (including the northern Andes and perhaps Middle America) and became more diversified. These sites contrast with places such as Junín and Apurimac in southern Peru, where beans actually seem to have been domesticated. Individual landraces from the latter sites are more uniform for ancestral traits and phaseolin types typical for the Andes, where-

as landraces from northern Cajamarca show a mixture of traits, including those found in accessions typically from Middle America.

In regions where geographical and ecological separation is largely a function of differences in altitudes (e.g., in Colombia, Ecuador, and Peru in the Andes; and Guatemala, Jalisco, Oaxaca, Puebla, etc., in Middle America), movement of germplasm from one microregion to another by pre-Columbian natives could have occurred rather easily and frequently. Gene exchange through outcrossing in sympatric regions would have occurred among wild forms and cultivated landraces. For example, in some parts of Peru and Mexico, it is not uncommon even today to find wild, weedy, and cultivated forms growing in close proximity or in the same field, thus increasing and facilitating occasional gene exchange through outcrossing (Debouck et al. 1989; Delgado Salinas et al. 1988; Vanderborght 1983). Outcrossing will very likely blur the limits of races and will also be responsible for the occasional lack of correspondence in morphological and molecular traits and some heterogeneity observed among members of races Jalisco, Mesoamerica, and Peru. Thus, for identification of races we would need several characters including molecular, morphological, physiological, agronomic, and adaptive attributes, along with breeding behavior.

In the opinion of one author (D. Debouck), both Andean and Middle American beans were probably initially harvested for their green, young pods. That use continued in Middle America until ceramics was found, but in the southern Andes, in addition to use as a vegetable, the aesthetic value of dry seed was highlighted and progressively selected for popped beans or nuñas. Occurrence of polymorphism for phaseolin types (C, H, and T types) in nuñas may indicate that the selection pressure for popping ability was widely distributed in the region and looked for in several wild populations of common bean. But once ceramics was discovered and maize introduced into the Andes, other ways of cooking beans were developed. A unique selective pressure was disrupted, and with it, probably a possibility of more phenotypic uniformity in the race Peru. Additionally, the southern Andes are not uniform climatically, and 3 to 5 microregions may be recognized. Since environment is thought to be one of the most important sources of selective pressure, that could have helped evolution and maintenance of different ecotypes. Moreover, several major phaseolin types (T, C, H, A, TO) are found in cultivated landraces from the Andes, suggesting that several wild bean populations were domesticated there, at different places and times. Moreover, race Peru may still be evolving, partly due to the presence of the wild-weed–crop complex (Debouck et al. 1989).

The proposed classification of cultivated landraces into races (this article and Singh et al. 1989b) and gene pools (Singh 1988, 1989, 1991b) should by no means be considered final. Domestication efforts were multiple in space and time (and probably are still taking place even today in some Andean regions [Debouck 1990; Debouck et al. 1989]) and relatively small samples of primitive landraces have been carefully studied for molecular markers. As more accessions of wild and cultivated bean populations are examined, especially by using molecular markers at the DNA level and by gathering information on crossing and breeding behavior, further refinements of the current classification are likely as are more convincing explanations of the occasional discrepancies recorded among molecular and morphological data.

The genetic distance between Middle American and Andean races seems to be considerable. As a consequence, although hybridization among these races is easily effected, various degrees and kinds of hybrid problems (Coyne 1965; Evans 1970; Gepts and Bliss 1985; Koenig and Gepts 1989b; Shii et al. 1980; Singh and Gutiérrez 1984; Singh and Molina 1991; Sprecher and Khairallah 1989; Vieira et al. 1989) are observed beginning in the F_1 and subsequent generations, thus interfering with effective recombination of genes among races. These phenomena are more pronounced in crosses between parents belonging to the race Mesoamerica (probably the most primitive) and those of Andean origins (races Nueva Granada and Chile are probably most highly evolved, in that order), possibly because they are farthest apart genetically. However, as noted earlier in this article and by Singh (1989), some members of each race possess certain desirable traits that are either not found or adequately expressed in other races. In addition, differences for most agronomic traits, including responses to varying photoperiod and temperature, time to maturity, plant type, seed yield, tolerance to drought and low soil fertility, and resistance to diseases and insects, are much greater among

races than within them. It is therefore imperative that for any tangible and long-lasting progress, all available genetic variation across races and gene pools must be used in breeding programs for the benefit of mankind. For example, the large-seeded landraces of Andean America, particularly the bush determinate and indeterminate members of the races Chile and Nueva Granada, are often characterized by reduced overall growth (crop dry weight), relative growth rate, harvest index, and seed yield compared to their small-seeded counterparts (race Mesoamerica) from Middle America in tropical environments (White and González 1990; White and Izquierdo 1991; White et al. 1991). Moreover, from crosses within races and gene pools, little or no progress could be achieved for yield due to insufficient genetic variation (Singh and Gutiérrez 1990; Singh et al. 1989c). On the other hand, substantial yield gains could be made from interracial populations involving parents possessing positive general combining ability for seed yield (Singh et al. 1989a, 1991c; Singh and Gutiérrez 1990). Similarly, for diseases such as anthracnose caused by *Colletotrichum lindemuthianum* (Sacc. & Magn.) Scrib., angular leaf spot caused by *Phaeoisariopsis griseola* (Sacc.) Ferraris, and rust caused by *Uromyces appendiculatus* (Pers.) Unger var. *appendiculatus*, all characterized by a large variation in their pathogenic populations, the sources of resistance identified in the Middle American races of common bean seem to offer protection against a much wider range of pathogenic populations occurring in the Andean South American bean-growing environments and vice versa (M. A. Pastor-Corrales, pers. comm.). The implications are that different groups of pathogen populations have coevolved with each common bean race, and that each probably carries a different mechanism and gene for pathogenicity in the fungus and genes for resistance in common bean races. Thus, higher levels and more stable resistance to a much broader range of pathogenic populations could be achieved by combining sources of resistance from different common bean races from both centers of domestication. Preliminary evidence supporting this hypothesis could be obtained from bean rust, where race-nonspecific (Shaik and Steadman 1988) and adult plant (Mmbaga and Steadman 1990) resistance appear to be associated with leaf pubescence found in large-seeded landraces (e.g., Pompadour Checa) belonging to the race Nueva Granada. This could be combined with race-specific resistance found in small-seeded germplasm (e.g. Ecuador 299, Compuesto Negro Chimaltenango, Stavely 1990) of the race Mesoamerica, in order to obtain much broader and stable resistance. A similar case could be made for bean golden mosaic virus (Morales and Singh 1991), anthracnose (Schwartz et al. 1982), angular leaf spot (Correa-Victoria 1987), drought, and soil fertility problems. However, more systematic studies on combining ability and genetic recombinations among and within races are required for their efficient use in improvement programs. Some strategies for gene transfer among races and gene pools were discussed earlier by Singh (1989, 1991b).

Characterization and classification of germplasm bank accessions into races (this article and Singh et al. 1989b) and gene pools (Singh 1988, 1989, 1991b) should also facilitate (1) elimination of duplications, (2) formation of a core collection, (3) acquisition of new germplasm, (4) genetic and evolutionary studies, and (5) efficient management and conservation of genetic resources. Correlating genetic distance between parents (as indicated by group membership) and their breeding behavior may help improve the predictiveness of their progeny performance for agronomic traits.

Acknowledgments

We are very grateful to Bill Hardy for editorial assistance and Aracelly Fernandez for typing of the manuscript. Seed samples supplied by the Genetic Resources Unit and illustrations prepared by the Graphic Arts Section of CIAT are also gratefully acknowledged.

Literature Cited

Brown, J. W. S., Y. Ma, F. A. Bliss, and T. C. Hall. 1981. Genetic variation in the subunits of globulin-1 storage protein of French bean. Theor. Appl. Genet. 59:83–88.

Brücher, H. 1988. The wild ancestor of *Phaseolus vulgaris* in South America. Pages 185–214 *in* P. Gepts, ed., Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, Netherlands.

Brunner, B. R., and J. S. Beaver. 1988. Estimation of outcrossing of dry beans in Puerto Rico. Ann. Rep. Bean Improv. Coop. 31:42–43.

Bukasov, S. M. 1930. The cultivated plants of Mexico, Guatemala and Colombia. Bull. Appl. Bot. Genet. Pl. Breed., Leningrad, Supplem. 459–553.

Burkart, A., and H. Brücher. 1953. *Phaseolus aborigineus* Burkart, die mutmassliche andine Stammform der Kulturbohne. Züchter 23:65–72.

Correa-Victoria, F. J. 1987. Pathogenic variation

production of toxic metabolites, and isoenzyme analysis in *Phaeoisariopsis griseola* (Sacc.) Ferr. Ph.D. dissertation, Michigan State University, East Lansing.

Coyne, D. P. 1965. A genetic study of "crippled" morphology resembling virus symptoms in *Phaseolus vulgaris* L. J. Hered. 56:162–163.

Debouck, D. G. 1989. Early beans (*Phaseolus vulgaris* L. and *P. lunatus* L.) domesticated for their aesthetic value? Ann. Rep. Bean Improv. Coop. 32: 62–63.

———. 1990. Wild beans as a food resource in the Andes. Ann. Rep. Bean Improv. Coop. 33:102–103.

———. 1991. Systematics and morphology. Pages 55–118 *in* A. van Schoonhoven and O. Voysest, eds., Common beans: research for crop improvement. C.A.B. Int., Wallingford, UK and CIAT, Cali, Colombia.

———, **M. Gamarra F., V. Ortiz A., and J. Tohme.** 1989. Presence of a wild-weed–crop complex in *Phaseolus vulgaris* L. in Peru? Ann. Rep. Bean Improv. Coop. 32:64–65.

———, **and R. Hidalgo.** 1986. Morphology of the common bean plant. Study guide. CIAT, Cali, Colombia:16–25.

———, **H. Rubiano, and M. del Carmen Menéndez S.** 1988. Determinate climbers among Argentinean materials of *Phaseolus vulgaris* L. Ann. Rep. Bean Improv. Coop. 31:189–190.

———, **and J. Tohme.** 1989. Implications for bean breeders of studies on the origins of common beans, *Phaseolus vulgaris* L. Pages 3–42 *in* S. Beebe, ed., Current topics in breeding of common bean. Working document no. 47. CIAT, Cali, Colombia.

Delgado Salinas, A., A. Bonet, and P. Gepts. 1988. The wild relative of *Phaseolus vulgaris* in Middle America. Pages 163–184 *in* P. Gepts, ed., Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, Netherlands.

Evans, A. 1970. Heterosis for yield in *Phaseolus vulgaris* crosses. Ann. Rep. Bean Improv. Coop. 13: 52–54.

———. 1973. Exploitation of the variability in plant architecture in *Phaseolus vulgaris*. Pages 279–286 *in* D. Wall, ed., Potentials of field beans and other food legumes in Latin America. Series Seminar 2E. CIAT, Cali, Colombia.

———. 1980. Structure, variation, evolution, and classification in *Phaseolus*. Pages 337–347 *in* R. J. Summerfield and A. H. Bunting, eds., Advances in legume science. HMSO, Royal Botanic Gardens, Kew, London.

Gentry, H. S. 1969. Origin of the common bean, *Phaseolus vulgaris*. Econ. Bot. 23:55–69.

Gepts, P. 1988a. Phaseolin as an evolutionary marker. Pages 215–241 *in* P. Gepts, ed., Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, Netherlands.

———. 1988b. A Middle American and an Andean common bean gene pool. Pages 375–390 *in* P. Gepts, ed., Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, Netherlands.

———. 1990. Biochemical evidence bearing on the domestication of *Phaseolus* (Fabaceae) beans. Econ. Bot. 44(supplement):28–38.

———, **and F. A. Bliss.** 1985. F_1 hybrid weakness in the common bean: differential geographic origin suggests two gene pools in cultivated bean germplasm. J. Hered. 76:447–450.

———, **and** ———. 1986. Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ. Bot. 40:469–478.

———, **and D. Debouck.** 1991. Origin, domestication, and evolution of the common bean (*Phaseolus vulgaris* L.). Pages 7–53 *in* A. van Schoonhoven and O. Voysest, eds., Common beans: research for crop improvement. C.A.B. Int., Wallingford, UK and CIAT, Cali, Colombia.

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

Harlan, J. R. 1975. Geographic patterns of variation in some cultivated plants. J. Hered. 66:184–191.

Harmsen, R., F. A. Bliss, and T. C. Osborn. 1987. Breeding beans resistant to bruchids. Ann. Rep. Bean Improv. Coop. 30:44–45.

Heiser, C. B. 1979. Origin of some cultivated New World plants. Ann. Rev. Ecol. Syst. 10:309–326.

Hidalgo, R. 1991. CIAT's world *Phaseolus* collection. Pages 163–197 *in* A. van Schoonhoven and O. Voysest, eds., Common beans: research for crop improvement. C.A.B. Int., Wallingford, UK and CIAT, Cali, Colombia.

Kaplan, L. 1965. Archaeology and domestication in American *Phaseolus* (beans). Econ. Bot. 19:358–368.

———. 1981. What is the origin of the common bean? Econ. Bot. 35:240–254.

———, **and L. N. Kaplan.** 1988. *Phaseolus* in archaeology. Pages 125–142 *in* P. Gepts, ed., Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, Netherlands.

Khairallah, M. M., M. W. Adams, and B. B. Sears. 1990. Mitochondrial DNA polymorphisms of Malawian bean lines: further evidence for two major gene pools. Theor. Appl. Genet. 80:753–761.

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

———, **and** ———. 1989b. Segregation and linkage

of genes for seed proteins, isozymes, and morphological traits in common bean (*Phaseolus vulgaris*). J. Hered. 80:455–459.

———, **S. P. Singh, and P. Gepts.** 1990. Novel phaseolin types in wild and cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ. Bot. 44:50–60.

Leakey, C. L. A. 1988. Genotypic and phenotypic markers in common bean. Pages 245–327 *in* P. Gepts, ed., Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, Netherlands.

Lioi, L., M. Esquivel, L. Castiñeiras, and K. Hammer. 1990. Phaseolin variation among common bean landraces from Cuba. Biol. Zentr. Bl. 109:231–233.

Ma, Y., and F. A. Bliss. 1978. Seed proteins of common bean. Crop Sci. 18:431–437.

Maréchal, R., J. M. Mascherpa, and F. Stainier. 1978. Etude taxonomique d'un groupe complexe d'espèces des genres *Phaseolus* et *Vigna* (Papilionaceae) sur la base de données morphologiques et polliniques, traitées par l'analyse informatique. Boissiera 28:1–273.

Miranda Colín, S. 1967. Origen de *Phaseolus vulgaris* L. (frijol común). Agrociencia 1:99–109.

Mmbaga, M. T., and J. R. Steadman. 1990. Adult plant rust resistance and leaf pubescence on dry beans. Ann. Rep. Bean Improv. Coop. 33:61–62.

Morales, F. J., and S. P. Singh. 1991. Genetics of resistance to bean golden mosaic virus in *Phaseolus vulgaris* L. Euphytica 52:113–117.

Motto, M., G. P. Sorresi, and F. Salamini. 1978. Seed size inheritance in a cross between wild and cultivated common beans (*Phaseolus vulgaris* L.). Genetica 49:31–36.

Nabhan, G. P., J. S. Munzuga M., F. Cárdenas R., and B. T. Burns. 1986. Wild bean explorations in northwest Mexico and southwest USA. Plant Genet. Resour. Newsl. 65:23–25.

Nienhuis, J., and S. P. Singh. 1988. Genetics of seed yield and its components in common bean (*Phaseolus vulgaris* L.) of Middle-American origin. I. General combining ability. Plant Breed. 101:143–154.

Ortega V., S. 1974. Polinización cruzada natural de la caraota (*P. vulgaris* L.) en Venezuela. Agronomía Tropical (Venezuela) 24:27–32.

Pereira Filho, I. A., and C. Cavariani. 1984. Taxa de hibridação natural do feijoeiro comum em Patos de Minas, Minas Gerais. Pes. Agrop. Bras. 19:1181–1183.

Rutger, J. N., and L. S. Beckham. 1970. Natural hybridization of *Phaseolus vulgaris* L. × *Phaseolus coccineus* L. J. Am. Soc. Hort. Sci. 95:659–661.

Schwartz, H. F., M. A. Pastor-Corrales, and S. P. Singh. 1982. New sources of resistance to anthracnose and angular leaf spot of beans (*Phaseolus vulgaris* L.). Euphytica 31:741–754.

Shaik, M., and J. R. Steadman. 1988. Nonspecific resistance to bean rust and its association with leaf pubescence. Ann. Rep. Bean Improv. Coop. 31:62–63.

Shii, C. T., M. C. Mok, S. R. Temple, and D. W. S. Mok. 1980. Expression of developmental abnormalities in hybrids of *Phaseolus vulgaris*. J. Hered. 71:218–222.

Singh, S. P. 1982. A key for identification of different growth habits of *Phaseolus vulgaris* L. Ann. Rep. Bean Improv. Coop. 25:92–95.

———. 1988. Gene pools in cultivated dry bean. Ann. Rep. Bean Improv. Coop. 31:180–182.

———. 1989. Patterns of variation in cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ. Bot. 43:39–57.

———. 1990. Bridging-parents for incompatible crosses between Mesoamerican and Andean common beans. Ann. Rep. Bean Improv. Coop. 33:111.

———. 1991a. Bean genetics. Pages 199–286 *in* A. van Schoonhoven and O. Voysest, eds., Common beans: research for crop improvement. C.A.B. Int., Wallingford, UK and CIAT, Cali, Colombia.

———. 1991b. Breeding for seed yield. Pages 383–443 *in* A. van Schoonhoven and O. Voysest, eds., Common beans: research for crop improvement. C.A.B. Int., Wallingford, UK and CIAT, Cali, Colombia.

———, **C. Cajiao, J. A. Gutiérrez, J. García, M. A. Pastor-Corrales, and F. J. Morales.** 1989a. Selection for seed yield in inter-gene pool crosses of common bean. Crop Sci. 29:1126–1131.

———, **D. G. Debouck, and P. Gepts.** 1989b. Races of common bean, *Phaseolus vulgaris* L. Pages 75–89 *in* S. Beebe, ed., Current topics in breeding of common bean. Working document no. 47. CIAT, Cali, Colombia.

———, ———, **and C. A. Urrea.** 1990. Variation for bracteoles and its association with races of common bean. Ann. Rep. Bean Improv. Coop. 33:112.

———, **and J. A. Gutiérrez.** 1984. Geographical distribution of the Dl_1 and Dl_2 genes causing hybrid dwarfism in *Phaseolus vulgaris* L., their association with seed size, and their significance to breeding. Euphytica 33:337–345.

———, **and** ———. 1990. Effect of plant density on selection for seed yield in two population types of *Phaseolus vulgaris* L. Euphytica 51:173–178.

———, ———, **A. Molina, C. Urrea, and P. Gepts.** 1991a. Genetic diversity in cultivated *Phaseolus vulgaris*. II. Marker-based analysis of morphological and agronomic traits. Crop Sci. 31:23–29.

———, **and A. Molina.** 1991. Occurrence of deformed leaflets (virus-like symptoms) in segregating generations of interracial populations of common bean. Ann. Rep. Bean Improv. Coop. 34:134–135.

———, **R. Nodari, and P. Gepts.** 1991b. Genetic diversity in cultivated *Phaseolus vulgaris* I. Allozymes. Crop Sci. 31:19–23.

———, **M. A. Pastor-Corrales, A. Molina, C. Urrea, and C. Cajiao.** 1991c. Independent, alternate, and simultaneous selection for resistance to anthracnose and angular leaf spot and effects on seed yield in common bean (*Phaseolus vulgaris* L.). Plant Breed. 104:in press.

———, **C. A. Urrea, J. A. Gutiérrez, and J. Garcia.** 1989c. Selection for yield at two fertilizer levels in small-seeded common bean. Can. J. Plant Sci. 69: 1011–1017.

Smartt, J. 1988. Morphological, physiological and biochemical changes in *Phaseolus* beans under domestication. Pages 143–161 *in* P. Gepts, ed., Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, Netherlands.

———. 1990a. The evolution of agriculturally significant legumes. Plant Breed. Abstr. 60:725–731.

———. 1990b. Grain legumes—evolution and genetic resources. Cambridge Univ. Press, Cambridge, UK.

Sprecher, S. L. 1988. New isozyme variants at two NADH diaphorase loci in dry beans: correlations to gene pools and commercial classes. Ann. Rep. Bean Improv. Coop. 31:92.

———, **and T. G. Isleib.** 1989. Morphological and phenological diversity between bean gene pools. Ann. Rep. Bean Improv. Coop. 32:54–55.

———, **and M. Khairallah.** 1989. Association of male sterility with gene pool recombinants in bean. Ann. Rep. Bean Improv. Coop. 32:56–57.

Stavely, J. R. 1990. Recent progress toward obtaining rust resistant beans. Ann. Rep. Bean Improv. Coop. 33:63–64.

Stoetzer, H. A. I. 1984. Natural cross-pollination in bean in Ethiopia. Ann. Rep. Bean Improv. Coop. 27:99–100.

Toro, O., J. Tohme, and D. G. Debouck. 1990. Wild bean (*Phaseolus vulgaris* L.): description and distribution. IBPGR and CIAT, Cali, Colombia.

Tucker, C. L., and J. Harding. 1975. Outcrossing in common bean, *Phaseolus vulgaris* L. J. Am. Soc. Hort. Sci. 100:283–285.

Urrea, C. A., and S. P. Singh. 1991. Variation for leaflet shape in wild and cultivated landraces of common bean. Ann. Rep. Bean Improv. Coop. 34: 133.

Vanderborght, T. 1983. Evaluation of *P. vulgaris* wild types and weedy forms. Plant Genet. Resour. Newsl. 54:18–25.

———. 1987. The study of common bean (*Phaseolus vulgaris* L.) variability by the use of multivariate statistical methods applied to a database. Diss. Abstr. 47:3190B–3191B.

Vargas, J., J. Tohme, and D. G. Debouck. 1990. Common bean domestication in the southern Andes. Ann. Rep. Bean Improv. Coop. 33:104–105.

Vieira, A. L., M. A. Patto Ramalho, and J. B. dos Santos. 1989. Crossing incompatibility in some bean cultivars utilized in Brazil. Rev. Brasil. Genet. 12:169–171.

Vieira, C. 1960. Sobre a hibridação natural em *Phaseolus vulgaris* L. Revista Ceres. 11:103–107.

Voysest, O. 1983. Variedades de frijol en América Latina y su origen. CIAT, Cali, Colombia.

———, **and M. Dessert.** 1991. Bean cultivars: classes and commercial seed types. Pages 119–162 *in* A. van Schoonhoven and O. Voysest, eds., Common beans: research for crop improvement. C.A.B. Int., Wallingford, UK and CIAT, Cali, Colombia.

Weiseth, G. 1954. Una variedad silvestre del poroto común (*Phaseolus vulgaris*), autóctona del Noroeste Argentino y su relación genética con variedades cultivadas. Rev. Agron. Noroeste Argentino 1:71–81.

Wells, W. C., W. H. Isom, and J. G. Waines. 1988. Outcrossing rates of six common bean lines. Crop Sci. 28:177–178.

White, J. W., and A. González. 1990. Characterization of the negative association between seed yield and seed size among genotypes of common bean. Field Crops Res. 23:159–175.

———, **and J. Izquierdo.** 1991. Physiology of yield potential and stress tolerance. Pages 287–382 *in* A. van Schoonhoven and O. Voysest, eds., Common beans: research for crop improvement. C.A.B. Int., Wallingford, UK and CIAT, Cali, Colombia.

———, **S. P. Singh, C. Pino, M. J. Rios B., and I. Buddenhagen.** 1991. Effects of seed size and photoperiod response on crop growth and yield of common bean. Field Crops Res. 28:in press.

Theor Appl Genet (1992) 84: 186-192

© Springer-Verlag 1992

Towards an integrated linkage map of common bean

1. Development of genomic DNA probes and levels of restriction fragment length polymorphism

R. O. Nodari *, E. M. K. Koinange **, J. D. Kelly *, and P. Gepts ******

Department of Agronomy and Range Science, University of California, Davis, CA 95616-8515, USA

Received April 29, 1991; Accepted October 17, 1991
Communicated by A. L. Kahler

Summary. Two genomic libraries were established to provide markers to develop an integrated map combining molecular markers and genes for qualitative and quantitative morpho-agronomic traits in common bean. Contrasting characteristics were observed for the two libraries. While 89% of the *Pst*I clones were classified as single-copy sequences, only 21% of the *Eco*RI-*Bam*HI clones belonged in that category. Clones of these two libraries were hybridized against genomic DNA of nine genotypes chosen according to their divergent evolutionary origin and contrasting agronomic traits. Eight restriction enzymes were used in this study. *Pst*I clones revealed 80-90% polymorphism between the Andean and Middle American gene pools and 50-60% polymorphism within these gene pools. However, under the same conditions only 30% of the *Eco*RI-*Bam*HI clones showed polymorphism between the Middle American and Andean gene pools. Hybridization with *Pst*I clones to *Eco*RI-, *Eco*RV-, or *Hin*dIII-digested genomic DNA resulted in a cumulative frequency of polymorphism of approximately 80%. Hybridizations to *Bam*HI-, *Hae*III-, *Hin*fI-, *Pst*I-, and *Xba*I-digested genomic DNA detected no additional polymorphisms not revealed by the former three enzymes. In the *Pst*I library, a positive correlation was observed between the average size of hybridizing restriction fragments and the frequency of polymorphism detected by each restriction enzyme. This relationship is consistent with the higher proportion of insertion/deletion events compared with the frequency of nucleotide substitutions observed in that library.

Key words: *Phaseolus vulgaris* - RFLP - Genetic diversity

* *Permanent address:* Departamento de Fitotecnica, Universidade Federal de Santa Catarina, Florianpolis, SC, Brazil
** *Permanent address:* Ministry of Agriculture, Lyamungu Research Institute, P.O. Box 3004, Moshi, Tanzania
*** *Permanent address:* Department of Crop and Soil Sciences, Michigan State University, East Lansing, MI 48824, USA
**** Correspondence to: P. Gepts

Introduction

Common bean (*Phaseolus vulgaris* L., $2n = 2 \times = 22$) is grown worldwide on more than 12 million hectares (Singh 1989) and constitutes a major source of dietary protein in Third World countries (Pachico 1989). Most of the common bean production and consumption consists of dry bean, although green pods can also be important on a regional basis (Singh 1989). Recently, its health benefits (e.g., source of fibers, hypocholestesterolemic effect) have also received considerable attention (Andersen et al. 1984).

Although common bean has been used to identify or confirm important genetic concepts (e.g., segregation, Mendel 1865; genotype versus phenotype, Johannsen 1909; quantitative trait locus, Sax 1923), there have been fewer genetic studies conducted with common bean than with other crop plants such as maize, pea, or tomato. This has been due in part to the lack of available markers. This situation can be remedied using restriction fragment length polymorphism (RFLP) markers. Compared to traditional morphological and biochemical markers, the advantages of RFLPs include their high number, their higher levels of polymorphism, co-dominance, freedom from environmental, epistatic, or deleterious effects, and detectability in all tissues and developmental stages (Beckmann and Soller 1986; Tanksley 1983).

Our long-term goal is to establish an integrated linkage map of common bean, i.e., a linkage map including

43

both molecular markers (particularly RFLPs) and genetic factors controlling agronomic traits. Because most of common bean breeding takes place within the primary gene pool *sensu* Harlan and de Wet (1971), polymorphisms at the molecular level should be identified within *P. vulgaris*. This, in turn, should increase the opportunities of applying the information form the linkage map to the development of improved bean cultivars by marker-assisted introduction of recessive or quantitative traits. Hence, a first objective of our mapping effort, detailed in this article, has been to identify bean genotype-DNA probe combinations that reveal high levels of intraspecific polymorphism. We describe here the establishment of two common bean genomic libraries and compare levels of polymorphisms between and within gene pools of *P. vulgaris*.

Material and methods

Plant material

In order to increase levels of polymorphism both at the molecular and agronomic levels, nine *P. vulgaris* genotypes were chosen according to two criteria: (1) divergent evolutionary origin; and (2) contrasting agronomic traits. BAT93 is a Middle American breeding line that is resistant to bean common mosaic virus (BCMV), rust (*Uromyces phaseoli*), common bacterial blight (*Xanthomonas phaseoli*), and anthracnose (*Colletotrichum lindemuthianum*), but is susceptible to angular leafspot (*Phaeoisariopsis angularis*). Jalo EEP558, an Andean landrace, exhibits opposite reactions to the same pathogens. 'Rio Tibagi', a small black-seeded Middle American cultivar, and ABA71, a large white-seeded Andean breeding stock, have distinct response to abiotic stress such as drought and low soil phosphorus levels. Midas, an Andean wax snap bean, and G12873, a Middle American wild accession, exhibit contrasting growth habits, pigmentation, pod fiber content, and photoperiod sensitivity.

Sierra, LEF2RB, and AC1028, all of Middle American origin, are breeding lines developed at Michigan State University and represent introgression between germ plasm from the arid upland Durango race with the humid lowland Mesoamerica race (Kelly et al. 1990; Singh et al. 1991 a).

DNA isolation, restriction enzyme digestion, and electrophoresis

Genomic DNA was isolated as described in Gepts and Clegg (1989) and Gepts et al. (1992). Seven to ten micrograms of DNA was digested at 37 °C for 4–5 h with different restriction enzymes (4–5 U μg DNA) according to the recommendations of the manufacturers (BRL and New England Biolabs). Digested DNA was separated by electrophoresis on 0.8% agarose gels in TAE running buffer for 16–20 h at 1 V/cm of gel.

Southern blotting and hybridization

The Southern hybridization followed the Zetabind membrane protocol of AMF-CUNO (Meriden, Conn). Hybridizations with ^{32}P-labelled probes ($1-15 \times 10^8$ dpm) were performed at 42 °C in 50% formamide. Sequential washes were carried out for 60 min each in 2 × SSC, 0.1% SDS at room temperature, 0.1 × SSC, 0.1% SDS at room temperature, and 0.1 × SSC, 0.1% SDS at 60 °C. The washed nylon membranes were exposed to X-Omat X-ray film for 1 h to 5 days. Inspection of the autoradiograms revealed that under these experimental conditions most samples had been digested to completion as indicated by the detection of single bands by almost 90% of the probes. In addition, the banding patterns observed in these screening experiments were heritable (R. Nodari and P. Gepts unpublished observations).

Development of genomic libraries

Total plant DNA from *P. vulgaris* cv 'California Dark Red Kidney' was purified in a cesium chloride gradient (Sambrook et al. 1989). After digestion with either *Eco*RI and *Bam*HI or *Pst*I, fragments smaller than 5.0 kb were size selected on a 0.8% agarose gel for the *Eco*RI-*Bam*HI library or on a 10%–40% sucrose gradient for the *Pst*I library (Weis and Quertermous 1987). Subsequently, DNA was precipitated with 95% ethanol at −20 °C (overnight), rinsed with cold 70% ethanol, and resuspended in TE. The size-selected fragments were ligated into the pUC18 plasmid (Sambrook et al. 1989), and competent bacterial cells (DH5-α) were transformed with the ligated products. After selection on LB medium with ampicillin, X-gal, and IPTG (Sambrook et al. 1989), plasmids from white colonies were isolated through a minipreparation (Holmes and Quigley 1981).

Identification of single-copy genomic probes

Approximately 1 μg of plasmid DNA from clones of each library (around 500) was digested with either *Eco*RI and *Bam*HI or *Pst*I, size separated on a 0.8% agarose gel to verify the presence of an insert, and subsequently transferred to a Zetabind membrane, as described above. These blots were probed with 1 μg of total genomic DNA radio-labelled with α-^{32}PdCTP (Amersham) by nick translation (Rigby et al. 1977). Clones showing the absence or near-absence of signal on the X-ray film were selected as single-copy clones.

Probes

Single-copy sequence inserts were separated from the plasmid after restriction digestion and electrophoresis in low-melting-temperature agarose. Approximately 30 ng of DNA from each insert was radio-labeled with α-^{32}PdCTP (Amersham) by the random primer method (Feinberg and Vogelstein 1984).

Results

Comparison of the two genomic libraries

The two genomic libraries, one resulting from the digestion of genomic DNA with *Pst*I and the other with *Eco*RI-*Bam*HI, exhibited contrasting characteristics. Whereas 89% of 426 *Pst*I clones were classified as single-copy sequences, only 21% of 484 clones from the *Eco*RI-*Bam*HI library belonged to the same category.

The two libraries also exhibited different levels of polymorphism detected by the single-copy sequences. Eighty-two percent of 61 randomly chosen *Pst*I clones identified polymorphisms with at least one of eight restriction enzymes (*Bam*HI, *Eco*RI, *Eco*RV, *Hae*III, *Hin*dIII, *Hin*fI, *Pst*I, and *Xba*I) in the BAT93 versus Jalo EEP558 comparison. A similar level was observed in a second comparison, 'Rio Tibagi' versus ABA71. This high level of polymorphism detected by the *Pst*I library was in marked contrast with lower levels of polymor-

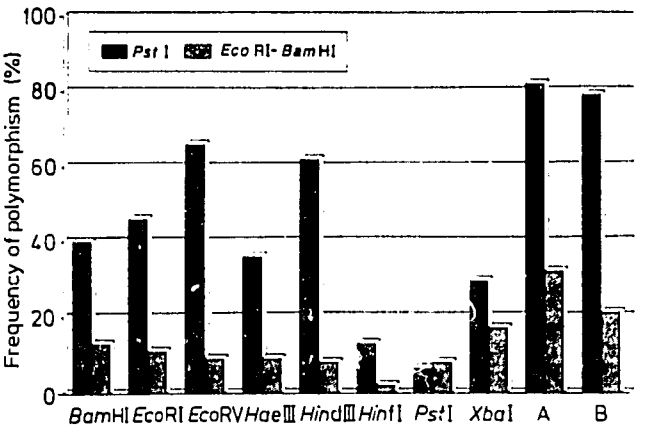


Fig. 1. Level of polymorphism detected by various restriction enzymes. *A* Cumulative polymorphism frequency detected by all eight restriction enzymes; *B* cumulative polymorphism frequency detected by *Eco*RI, *Eco*RV, and *Hin*dIII

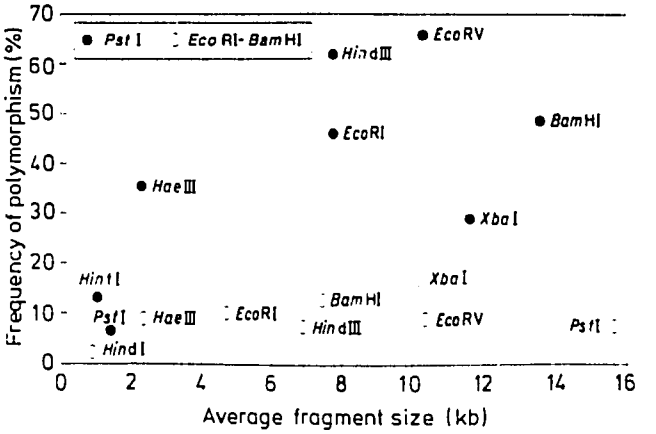


Fig. 2. Relationship between average fragment size and polymorphism level revealed by *Pst*I or *Eco*RI-*Bam*HI clones after restriction digestion of genomic DNA

phism (32%) detected by 57 *Eco*RI-*Bam*HI clones in the BAT93 versus Jalo EEP558 comparison using the same eight restriction enzymes.

When *Pst*I clones were used, we observed that the restriction enzymes used to digest parental DNA detected different levels of polymorphism. In the BAT93 versus Jalo EEP558 comparison, levels of polymorphism ranged between 6.5% (*Pst*I) and 66% (*Eco*RV) (Fig. 1). In the comparison involving 'Rio Tibagi' and ABA71, levels of polymorphism varied between 10% (*Hin*fI) and 67% (*Hin*dIII) (not shown). In both comparisons, high levels of polymorphisms (60–70%) were detected by *Eco*RV and *Hin*dIII, and intermediate levels were detected by *Eco*RI, *Hae*III, *Bam*HI, and *Xba*I. Any polymorphism seen with *Bam*HI, *Hae*III, *Hin*fI, *Pst*I, or *Xba*I, was also seen with *Eco*RI, *Eco*RV, or *Hin*dIII. Hence, the polymorphism observed with the latter three enzymes was as high as that observed with all eight restriction enzymes, i.e., approximately 80% (Fig. 1).

A different pattern was observed for the *Eco*RI-*Bam*HI clones. All of the enzymes detected similar low levels of polymorphism in the comparison BAT93 versus Jalo EEP558 (2–17%). Combined polymorphisms detected by these clones with the eight restriction enzymes amounted to 32%, whereas polymorphisms detected in *Eco*RI-, *Eco*RV-, or *Hin*dIII-digested genomic DNA amounted to 21% (Fig. 1).

What is the probability that an additional clone will reveal a polymorphism between two bean genotypes and, hence, can be mapped? Data from single comparisons – BAT93 versus Jalo EEP558 or Midas versus G12873 – indicated that approximately four-fifths of the *Pst*I clones were capable of detecting a polymorphism with the restriction enzymes *Eco*RI, *Eco*RV, or *Hin*dIII. When the data from two different comparisons were combined, 46 out of 61 clones (75%) detected polymorphisms in both comparisons; 12 out of 61 randomly chosen *Pst*I clones (20%) detected differences in one or the other comparison, and 3 out of 61 clones (5%) detected no polymorphism in either comparison. Hence, using two populations, it would be possible to map 95% of the *Pst*I clones.

Fragment size and frequency of polymorphism

An analysis involving 61 *Pst*I probes in the comparison BAT 93 versus Jalo EEP558 showed significant differences in the average size of the genomic fragments generated by different restriction enzymes (Table 1). Whereas *Hin*fI, *Pst*I, and *Hae*III produced small fragments of sizes 1.0, 1.4, and 2.3 kb, respectively, *Eco*RV, *Xba*I, and *Bam*HI generated fragments with an average size of 10.4, 11.6 and 13.5 kb, respectively. Between these extremes, two restriction enzymes, *Eco*RI and *Hin*dIII, generated fragments with average size of 7.7 and 7.8 kb, respectively. In general, the average fragment size generated by each restriction enzyme was positively correlated with the level of polymorphism revealed by that enzyme (Fig. 2). This association between fragment size and level of polymorphism was not observed for the *Eco*RI-*Bam*HI clones (Fig. 2).

Interversus Intra-gene pool polymorphism

The level of polymorphism was also affected by the degree of genetic similarity among genotypes chosen to identify polymorphisms. Results on levels of polymorphism presented in the previous two paragraphs involved comparisons between Mesoamerican and Andean genotypes. Comparisons within the Andean or Mesoamerican group yielded the following results. Of 57 *Pst*I clones 49% detected polymorphisms between the three cultivated genotypes from the Andean region (ABA71, Jalo EEP558, and Midas) (Table 2). Within the Middle American germ plasm (BAT93, G12873, AC1028, LEF2RB,

Table 1. Average fragment size generated by various restriction enzymes

Restriction enzyme	*Pst*I library			*Eco*RI-*Bam*HI library		
	Number of bands	Average size (kb)	Standard error	Number of bands	Average size (kb)	Standard error
*Hin*fI	50	1.0 a[a]	0.06	35	0.8 a	0.08
*Pst*I	49	1.4 b	0.15	33	15.6 f	1.24
*Hae*III	65	2.3 c	0.31	52	2.3 b	0.33
*Eco*RI	57	7.7 d	0.94	37	4.7 c	0.84
*Hin*dIII	65	7.8 d	0.82	40	6.8 cd	0.88
*Eco*RV	58	10.4 de	0.89	35	10.3 d	1.30
*Xba*I	48	11.6 e	1.21	38	10.2 d	1.26
*Bam*HI	52	13.5 e	1.25	32	7.4 cd	1.19

[a] Averages followed by the same letter within a column are significantly different according to pairwise *t*-tests (Steel and Torrie 1980)

Table 2. RFLP levels between and within the Mesoamerican and Andean cultivar groups

Comparison	Number of probes	Number of enzymes	RFLP level (%)
	Between Mesoamerican and Andean cultivar groups		
BAT93, Jalo EEP558	57	6[b]	86
	61	3[c]	82
Midas, G12873	61	3	89
	Within Mesoamerican or Andean cultivar groups		
Mesoamerican[a]	53	6	62
Andean[a]	57	6	49

[a] Mesoamerican: BAT93, G12873, Sierra, LEF2RB, AC1028; Andean: Jalo EEP558, ABA71, Midas
[b] *Bam*HI, *Eco*RI, *Eco*RV, *Hin*dIII, *Pst*I, *Xba*I
[c] *Eco*RI, *Eco*RV, *Hin*dIII

Table 3. Frequency (in percentage) of site mutations and chromosomal rearrangements

Comparison	Number of probes	Frequency (%)		
		Site mutations	Insertions/deletions	Ambiguous
	*Pst*I library			
BAT93 versus Jalo EEP558	50	20	32	48
Rio Tibagi versus ABA71	23	13	35	52
	*Eco*RI-*Bam*HI library			
BAT93 versus Jalo EEP558	17	59	0	41

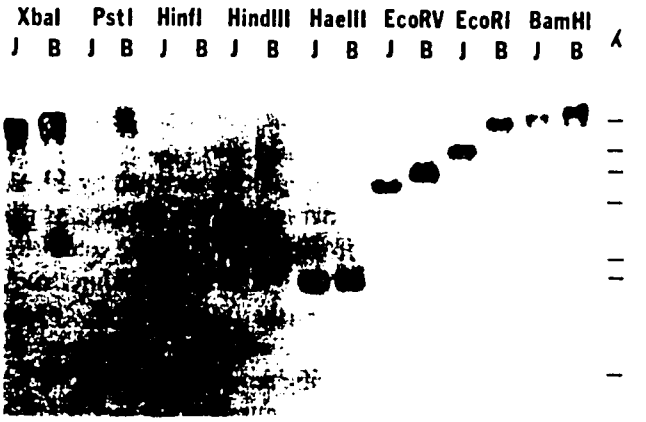


Fig. 3. Polymorphism detected by *Pst*I genomic clone GUC1049 between BAT93 (*B*) and Jalo EEP558 (*J*)

and Sierra), 62% of 53 *Pst*I clones revealed a polymorphism. As expected, therefore, comparisons within the Mesoamerican or Andean groups showed lower levels of polymorphism than between these two groups.

Type of polymorphisms

Polymorphisms can be generated either through nucleotide substitution or through insertions/deletions. If digestion with one or two of the eight restriction enzymes showed a polymorphism, then a nucleotide substitution at the specific recognition site was inferred. If a consistency in the polymorphism was observed among five or more restriction enzymes (e.g., one parent consistently exhibited a larger fragment, as illustrated in Fig. 3), the polymorphism was assumed to be due to an insertion deletion (or chromosomal rearrangement). Intermediate cases were classified as ambiguous. Using these criteria, a majority of the polymorphisms revealed by *Pst*I clones appeared to be due to an insertion/deletion event, whereas *Eco*RI-*Bam*HI clones recognized mostly nucleotide substitutions (Table 3).

46

Discussion

Libraries enriched with single-copy sequences have been obtained in common bean (95% single-copy sequences; Chase et al. 1991) and in other species such as rice (85%; McCouch et al. 1988), maize (50.6%; Burr et al. 1988), tomato (92%; Miller and Tanksley 1990b), and barley (89%; Graner et al. 1990) when DNA was digested with *Pst*I, a methylation-sensitive enzyme (as opposed to *Bam*HI and *Eco*RI, two methylation-insensitive enzymes). Lack of methylation has been associated with gene activity in vertebrates and plants (e.g., Antequera and Bird 1988; Schwartz 1989; Monk 1990).

The relatively high level of polymorphism revealed by the *Pst*I library compared to that detected by the *Eco*-RI-*Bam*HI library (80% versus 32%) was unexpected. *Pst*I clones also exhibited high levels of polymorphisms in other species. In tomato, higher levels of polymorphism were observed with *Pst*I clones than with *Eco*RI clones (Miller and Tanksley 1990a). Of the *Pst*I probes 75% were polymorphic with three enzymes when tested in 38 2n potato lines (Gebhardt et al. 1989).

The wide range of polymorphisms revealed by different libraries illustrate the paramount importance of the restriction enzymes used to digest DNA for the establishment of a genomic library. The cloning procedure, such as the restriction enzyme and the fragment size, apparently induces a sampling bias which favors certain regions of the genome or specific sequences, depending on their base composition, copy number, or methylation level. *Pst*I will not cleave DNA at the recognition sequence (5'-CTGCAG-3') if the cytosine at the 5' end is methylated (Nelson and McClelland 1987; Gruenbaum et al. 1981). Our data suggest that DNA sequences homologous to the *Pst*I clones (or DNA sequences adjacent to them) are more variable than those homologous to the *Eco*RI-*Bam*HI clones. This variability revealed by *Pst*I clones appears to be due more often to small rearrangements than to nucleotide substitutions, although the latter could also play an important role. Variability revealed by *Eco*RI-*Bam*HI clones appeared to be due mostly to nucleotide substitutions. We are not aware of an explanation that could account for this difference in polymorphism levels detected with the two libraries.

Recently, two other groups have obtained estimates of RFLP levels in common bean. Chase et al. (1991) found that in a group of 14 Middle American and Andean accessions, 5 out of 18 (28%) random *Pst*I clones revealed polymorphisms in *Eco*RI digests of genomic DNA. For the *Eco*RV and *Hin*dIII digests, polymorphisms were observed with 5 out of 18 (28%) and 7 out of 18 (39%) of the clones, respectively. Within the Middle American group (excluding accession XR-235-1-1, which includes *P. coccineus* in its background), 5 out of 18 (28%) revealed polymorphisms whereas in the Andean gene pool 4 out of 18 (22%) revealed polymorphisms. The lower levels of polymorphisms observed by these authors can be attributed to differences in the plant material analyzed. For example, all Middle American accessions analyzed by Chase et al. (1991) belong to race Mesoamerica, whereas the accessions analyzed in our study belong to either race Mesoamerica or race Durango of the Middle American gene pool (Singh et al. 1991a).

Guo et al. (1991) observed only 22–26% polymorphisms within *P. vulgaris* but 70–80% between *P. vulgaris* and *P. coccineus*. The lower level of polymorphisms may be due to the use of a single restriction enzyme (*Eco*RI), to the expression of their results as the proportion of polymorphic bands to the total number of bands and to different sources of probes (mostly anonymous cDNA probes and clones of genes of known function).

Although experimental conditions are not strictly comparable, the RFLP levels we have identified in common bean (around 80–90%) compare favorably with those observed in other species (Table 4), especially self-pollinated plants such as soybean (40–50%) and tomato (14%) (Keim et al. 1989; Helentjaris et al. 1985). Table 2 shows that while between gene pools we were able to detect 80–90% polymorphism, within gene pools we still detected 50–60% polymorphism. Hence, RFLP analysis confirmed the divergence between Mesoamerican and Andean gene pools detected previously with morphological, biochemical, and reproductive isolation data (Vanderborght 1987; Gepts et al. 1986; Gepts 1990; Gepts and Debouck 1991; Sprecher 1988; Khairallah et al. 1990; Koenig and Gepts 1989; Shii et al. 1980; Singh et al. 1991b, c). In order to increase levels of polymorphisms at the molecular level, parents should belong to contrasting gene pools, i.e., Mesoamerican versus Andean. However, the level of polymorphism within gene pools appears to be still high enough to develop efficient applications with these markers, such as marker-assisted selection or evolutionary studies.

Strong correlation have been found between the average size of the hybridizing restriction fragments and the frequency of polymorphisms generated by different restriction enzymes (McCouch et al. 1988; Miller and Tanksley 1990a). By testing 61 probes against BAT93 and Jalo EEP558 genomic DNA digested with eight restriction enzymes, we found similar results. This positive correlation is consistent with our finding that a majority of the RFLPs can be attributed to insertions/deletions.

From a practical point of view, mapping efficiency increases when the number of restriction enzymes used to digest DNA of parents and of segregating generations can be decreased without a significant reduction in polymorphism levels. Our results indicate that the level of polymorphism detected by the combined use of *Eco*RV, *Eco*RI, and *Hin*dIII is the same as the one detected by all

Table 4. Comparison of RFLP levels in common bean and other plants

Crop	Genotypes	Type of clone	Number of enzymes	RFLP (%)	Source
Arabidopsis	3 strains	Genomic: *Eco*RI	3	46	Chang et al. (1988)
Barley	48 cultivars	Genomic: *Pst*I	3	43	Graner et al. (1990)
	4 cultivars	Genomic: *Pst*I	5	54	Heun et al. (1991)
Brassica	3 species	Genomic: *Pst*I	1	95 between species	Figdore et al. (1988)
	(37 cultivars or genetic stocks)			80 between subspecies	Figdore et al. (1988)
				70 within subspecies	Figdore et al. (1988)
Common Bean	7 cultivars + 1 wild	Genomic: *Pst*I	3	89 between subspecies	Nodari et al.
				62 within Mesoamerican	(present results)
				49 within Andean	
Lentil	2 cultivars +	Genomic: *Pst*I, *Eco*RI	4	38	Havey and Muehlbauer
	2 wild	cDNA	4	64	(1989)
Lettuce	4 cultivars	Genomic: *Mbo*I	9	13	Landry et al. (1987)
		cDNA	9	25	Landry et al. (1987)
Maize	3 cultivars	cDNA	3	83	Helentjaris et al. (1985)
Potato	*phureja* × (*tuberosum*	Genomic: *Pst*I, *Eco*RI	11	60	Bonierbale et al. (1988)
	× *chacoense*)	cDNA	11	51	Bonierbale et al. (1988)
Rice	*indica* × *javanica*	Genomic: *Pst*I	11	78	McCouch et al. (1988)
Soybean	5 cultivars	Genomic: *Sau*3AI, *Eco*RI	13	20	Apuya et al. (1980)
	58 wild + cultivated	Genomic: *Sau*3Ai	4	40–50	Keim et al. (1989)
Tomato	2 cultivars	cDNA	3	14	Helentjaris et al. (1985)
	esculentum-cheesmannii	cDNA	3	68	Helentjaris et al. (1985)
	esculentum-pennellii	cDNA	3	100	Helentjaris et al. (1985)

eight restriction enzymes together (around 80%). These three enzymes were also very efficient in detecting polymorphisms in lettuce (Landry et al. 1987) and potato (Bonierbale et al. 1988). When the data from two different pairwise genotypic comparisons were combined, the level of polymorphism was even higher than in a single comparison. This observation suggests that mapping in more than one population will lead to a denser map, in general, and will increase the probability of identifying polymorphic markers in a chromosome region of interest.

Acknowledgements. This work was supported by the United States Agency for International Development and the International Board for Plant Genetic Resources. R. Nodari and E. Koinange are recipients of a fellowship from CNPq and SAD-CC.CIAT, respectively. We thank Gisela Arndt, Belén Garrido, and Nino Polignano for assistance, and T. Wilkins for useful discussions.

References

Andersen JW, Story L, Sieling B, Chen W-JL, Petro MS, Story J (1984) Hypocholesterolemic effects of oat-bran or bean intake for hypercholesterolemic men. Am J Clin Nutr 40:1146–1155

Antequera F, Bird AP (1988) Unmethylated CpG islands associated with genes in higher plant DNA. EMBO J 7:2295–2299

Apuya NR, Frazier BL, Keim P, Roth EJ, Lark KG (1988) Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. Theor Appl Genet 75:889–901

Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms and genetic improvement of agriculatural species. Euphytica 35:111–124

Bonierbale M, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095–1103

Burr B, Burr FA, Thompson KH, Albertsen MC, Stuber CW (1988) Gene mapping with recombinant inbreds of maize. Genetics 118:519

Chang C, Bowman JL, DeJohn AW, Lander ES, Meyerowitz EM (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc Natl Acad Sci USA 85:6856–6860

Chase CD, Ortega VM, Vallejos CE (1991) DNA restriction fragment length polymorphisms correlate with isozyme diversity in *Phaseolus vulgaris* L. Theor Appl Genet 81:806–811

Feinberg AP, Vogelstein B (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267

Figdore S, Kennard WC, Song KM, Slocum MK, Osborn TC (1988) Assessment of the degree of restriction fragment length polymorphism in *Brassica*. Theor Appl Genet 75:833–840

Gebhardt C, Ritter E, Debener T, Schachtschabel V, Wakemeier B, Uhrig H, Salamini F (1989) RFLP analysis and linkage mapping in *Solanum tuberosum*. Theor Appl Genet 78:65–75

Gepts P (1990) Biochemical evidence bearing on the domestication of *Phaseolus* beans. Econ Bot 44S:28–38

Gepts P, Clegg MT (1989) Genetic diversity in pearl millet (*Pennisetum glaucum* [L.] R.Br.) at the DNA sequence level. J Hered 80:203–208

Gepts P, Debouck DG (1991) Origin, domestication, and evolution of the common bean, *Phaseolus vulgaris*. In: Voysest O, Van Schoonhoven A (eds) Common beans: research for crop improvement. CAB, Oxon, UK, pp 7–53

48

Gepts P, Osborn TC, Rashka K, Bliss FA (1986) Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. Econ Bot 40:451–468

Gepts P, Llaca V, Nodari RO, Panella L (1992) Analysis of seed proteins, isozymes, and RFLPs for genetic and evolutionary studies in *Phaseolus*. In: Linskens HF, Jackson JF (eds) Modern methods of plant analysis. New series, Vol. 14. Springer, Berlin Heidelberg New York (in press)

Graner A, Siedler H, Jahoor A, Herrmann RG, Wenzel G (1990) Assessment of the degree and the type of restriction fragment polymorphism in barley (*Hordeum vulgare*). Theor Appl Genet 80:826–832

Gruenbaum Y, Maveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292:260-262

Guo M, Lightfoot DA, Mok MC, Mok DWS (1991) Analyses of *Phaseolus vulgaris* L. and *P. coccineus* Lam. hybrids by RFLP: preferential transmission of *P. vulgaris* alleles. Theor Appl Genet 81:703–709

Harlan JR, de Wet JMJ (1971) Towards a rational classification of cultivated plants. Taxon 20:509–517

Havey MJ, Muehlbauer FJ (1989) Linkages between restriction fragment length, isozyme, and morphological markers in lentil. Theor Appl Genet 77:395–401

Helentjaris T, King G, Slocum M, Siedenstrang C, Wegmann S (1985) Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol Biol 5:109–118

Heun M, Kennedy AE, Anderson JA, Lapitan NLV, Sorrells ME, Tanksley SD (1991) Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*). Genome 34:437–447

Holmes DS, Quigley M (1981) A rapid boiling method for the preparation of bacterial plasmids. Anal Biochem 114:193–197

Johannsen W (1909) Elementen der exakten Erblichkeitslehre. Fisher, Jena, FRG

Keim P, Shoemaker RC, Palmer RG (1989) Restriction fragment polymorphism diversity in soybean. Theor Appl Genet 77:786–792

Kelly JD, Adams MW, Saettler AW, Hosfield GL, Varner GV, Uebersax MA, Taylor J (1990) Registration of 'Sierra' pinto bean. Crop Sci 30:745–746

Khairallah MM, Adams MW, Sears BB (1990) Mitochondrial DNA polymorphins of Malawian bean lines: further evidence for two major gene pools. Theor Appl Genet 80:753–761

Koenig R, Gepts P (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. Theor Appl Genet 78:809–817

Landry BS, Kesseli RV, Farrara B, Michelmore RW (1987) A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331–337

McCouch SR, Kochert G, Yu ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815–829

Mendel G (1865) Experiments on plant hybrids. In: Stern C, Sherwood ER (eds) The origin of genetics. W. H. Freeman, San Francisco, pp 1–48

Miller JC, Tanksley SD (1990a) Effect of different restriction enzymes, probe source, and probe length on detecting restriction fragment length polymorphism in tomato. Theor Appl Genet 80:385–389

Miller JC, Tanksley SD (1990b) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet 80:437–448

Monk M (1990) Variation in epigenetic inheritance. Trends Genet 6:110–114

Nelson M, McClelland M (1987) The effect of site-specific methylation on restriction-modification enzymes. Nucleic Acids Res 15:219–230

Pachico D (1989) Trends in world common bean production. In: Schwartz HF, Pastor-Corrales MA (eds) Bean production problems in the tropics. CIAT, Cali, Colombia, pp 1–8

Rigby PWJ, Dieckman M, Rhodes C, Berg O (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick-translation with DNA polymerase. J Mol Biol 113:237–251

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sax K (1923) The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552–560

Schwartz D (1989) Gene-controlled cytosine demethylation in the promoter region of the *Ac* transposable element in maize. Proc Natl Acad Sci USA 86:2789–2793

Shii CT, Mok MC, Temple SR, Mok DWS (1980) Expression of developmental abnormalities in hybrids of *Phaseolus vulgaris* L. J Hered 71:218–222

Singh SP (1989) Patterns of variation in cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ Bot 43:39–57

Singh SP, Gepts P, Debouck DG (1991a) Races of common bean (*Phaseolus vulgaris* L., Fabaceae). Econ Bot 45:379–396

Singh SP, Gutiérrez JA, Molina A, Urrea C, Gepts P (1991b) Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci 31:23–29

Singh SP, Nodari R, Gepts P (1991c) Genetic diversity in cultivated common bean. I. Allozymes. Crop Sci 31:19–23

Sprecher SL (1988) Allozyme differentiation between gene pools in common bean (*Phaseolus vulgaris* L.), with special reference to Malawian germplasm. Ph.D thesis, Michigan State University, East Lansing, Mich

Steel RGD, Torrie JH (1980) Principles and procedures of statistics. McGraw Hill, New York

Tanksley SD (1983) Molecular markers in plant breeding. Plant Mol Biol Rep 1:3–8

Vanderborght T (1987) The study of common bean (*Phaseolus vulgaris* L.) variability by the use of multivariate statistical methods applied to a data base. Diss Abstr 47:3190B-1B

Weis JH, Quertermous T (1987) Size fractionation using sucrose gradients. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds) Current protocols in molecular biology. Wiley, New York, pp 5.3.2-5.3.8

49

Theor Appl Genet (1992) 84: 186–192

© Springer-Verlag 1992

Towards an integrated linkage map of common bean

1. Development of genomic DNA probes and levels of restriction fragment length polymorphism

R. O. Nodari *, E. M. K. Koinange **, J. D. Kelly *, and P. Gepts ******

Department of Agronomy and Range Science, University of California, Davis, CA 95616-8515, USA

Received April 29, 1991; Accepted October 17, 1991
Communicated by A. L. Kahler

Summary. Two genomic libraries were established to provide markers to develop an integrated map combining molecular markers and genes for qualitative and quantitative morpho-agronomic traits in common bean. Contrasting characteristics were observed for the two libraries. While 89% of the *Pst*I clones were classified as single-copy sequences, only 21% of the *Eco*RI-*Bam*HI clones belonged in that category. Clones of these two libraries were hybridized against genomic DNA of nine genotypes chosen according to their divergent evolutionary origin and contrasting agronomic traits. Eight restriction enzymes were used in this study. *Pst*I clones revealed 80–90% polymorphism between the Andean and Middle American gene pools and 50–60% polymorphism within these gene pools. However, under the same conditions only 30% of the *Eco*RI-*Bam*HI clones showed polymorphism between the Middle American and Andean gene pools. Hybridization with *Pst*I clones to *Eco*RI-, *Eco*RV-, or *Hin*dIII-digested genomic DNA resulted in a cumulative frequency of polymorphism of approximately 80%. Hybridizations to *Bam*HI-, *Hae*III-, *Hin*fI-, *Pst*I-, and *Xba*I-digested genomic DNA detected no additional polymorphisms not revealed by the former three enzymes. In the *Pst*I library, a positive correlation was observed between the average size of hybridizing restriction fragments and the frequency of polymorphism detected by each restriction enzyme. This relationship is consistent with the higher proportion of insertion/deletion events compared with the frequency of nucleotide substitutions observed in that library.

Key words: *Phaseolus vulgaris* – RFLP – Genetic diversity

Introduction

Common bean (*Phaseolus vulgaris* L., $2n = 2 \times = 22$) is grown worldwide on more than 12 million hectares (Singh 1989) and constitutes a major source of dietary protein in Third World countries (Pachico 1989). Most of the common bean production and consumption consists of dry bean, although green pods can also be important on a regional basis (Singh 1989). Recently, its health benefits (e.g., source of fibers, hypocholestesterolemic effect) have also received considerable attention (Andersen et al. 1984).

Although common bean has been used to identify or confirm important genetic concepts (e.g., segregation, Mendel 1865; genotype versus phenotype, Johannsen 1909; quantitative trait locus, Sax 1923), there have been fewer genetic studies conducted with common bean than with other crop plants such as maize, pea, or tomato. This has been due in part to the lack of available markers. This situation can be remedied using restriction fragment length polymorphism (RFLP) markers. Compared to traditional morphological and biochemical markers, the advantages of RFLPs include their high number, their higher levels of polymorphism, co-dominance, freedom from environmental, epistatic, or deleterious effects, and detectability in all tissues and developmental stages (Beckmann and Soller 1986; Tanksley 1983).

Our long-term goal is to establish an integrated linkage map of common bean, i.e., a linkage map including

* *Permanent address:* Departamento de Fitotecnica, Universidade Federal de Santa Catarina, Florianpolis, SC, Brazil
** *Permanent address:* Ministry of Agriculture, Lyamungu Research Institute, P.O. Box 3004, Moshi, Tanzania
*** *Permanent address:* Department of Crop and Soil Sciences, Michigan State University, East Lansing, MI 48824, USA
**** Correspondence to: P. Gepts

50

both molecular markers (particularly RFLPs) and genetic factors controlling agronomic traits. Because most of common bean breeding takes place within the primary gene pool *sensu* Harlan and de Wet (1971), polymorphisms at the molecular level should be identified within *P. vulgaris*. This, in turn, should increase the opportunities of applying the information form the linkage map to the development of improved bean cultivars by marker-assisted introduction of recessive or quantitative traits. Hence, a first objective of our mapping effort, detailed in this article, has been to identify bean genotype-DNA probe combinations that reveal high levels of intraspecific polymorphism. We describe here the establishment of two common bean genomic libraries and compare levels of polymorphisms between and within gene pools of *P. vulgaris*.

Material and methods

Plant material

In order to increase levels of polymorphism both at the molecular and agronomic levels, nine *P. vulgaris* genotypes were chosen according to two criteria: (1) divergent evolutionary origin; and (2) contrasting agronomic traits. BAT93 is a Middle American breeding line that is resistant to bean common mosaic virus (BCMV), rust (*Uromyces phaseoli*), common bacterial blight (*Xanthomonas phaseoli*), and anthracnose (*Colletotrichum lindemuthianum*), but is susceptible to angular leafspot (*Phaeoisariopsis angularis*). Jalo EEP558, an Andean landrace, exhibits opposite reactions to the same pathogens. 'Rio Tibagi', a small black-seeded Middle American cultivar, and ABA71, a large white-seeded Andean breeding stock, have distinct response to abiotic stress such as drought and low soil phosphorus levels. Midas, an Andean wax snap bean, and G12873, a Middle American wild accession, exhibit contrasting growth habits, pigmentation, pod fiber content, and photoperiod sensitivity.

Sierra, LEF2RB, and AC1028, all of Middle American origin, are breeding lines developed at Michigan State University and represent introgression between germ plasm from the arid upland Durango race with the humid lowland Mesoamerica race (Kelly et al. 1990; Singh et al. 1991 a).

DNA isolation, restriction enzyme digestion, and electrophoresis

Genomic DNA was isolated as described in Gepts and Clegg (1989) and Gepts et al. (1992). Seven to ten micrograms of DNA was digested at 37 °C for 4–5 h with different restriction enzymes (4–5 U μg DNA) according to the recommendations of the manufacturers (BRL and New England Biolabs). Digested DNA was separated by electrophoresis on 0.8% agarose gels in TAE running buffer for 16–20 h at 1 V/cm of gel.

Southern blotting and hybridization

The Southern hybridization followed the Zetabind membrane protocol of AMF-CUNO (Meriden, Conn). Hybridizations with ^{32}P-labelled probes ($1-15 \times 10^8$ dpm) were performed at 42 °C in 50% formamide. Sequential washes were carried out for 60 min each in 2 × SSC, 0.1% SDS at room temperature, 0.1 × SSC, 0.1% SDS at room temperature, and 0.1 × SSC, 0.1% SDS at 60 °C. The washed nylon membranes were exposed to X-Omat X-ray film for 1 h to 5 days. Inspection of the autoradiograms revealed that under these experimental conditions most samples had been digested to completion as indicated by the detection of single bands by almost 90% of the probes. In addition, the banding patterns observed in these screening experiments were heritable (R. Nodari and P. Gepts unpublished observations).

Development of genomic libraries

Total plant DNA from *P. vulgaris* cv 'California Dark Red Kidney' was purified in a cesium chloride gradient (Sambrook et al. 1989). After digestion with either *Eco*RI and *Bam*HI or *Pst*I, fragments smaller than 4.0 kb were size selected on a 0.8% agarose gel for the *Eco*RI-*Bam*HI library or on a 10%–40% sucrose gradient for the *Pst*I library (Weis and Quertermous 1987). Subsequently, DNA was precipitated with 95% ethanol at −20 °C (overnight), rinsed with cold 70% ethanol, and resuspended in TE. The size-selected fragments were ligated into the pUC18 plasmid (Sambrook et al. 1989), and competent bacterial cells (DH5–α) were transformed with the ligated products. After selection on LB medium with ampicillin, X-gal, and IPTG (Sambrook et al. 1989), plasmids from white colonies were isolated through a minipreparation (Holmes and Quigley 1981).

Identification of single-copy genomic probes

Approximately 1 μg of plasmid DNA from clones of each library (around 500) was digested with either *Eco*RI and *Bam*HI or *Pst*I, size separated on a 0.8% agarose gel to verify the presence of an insert, and subsequently transferred to a Zetabind membrane, as described above. These blots were probed with 1 μg of total genomic DNA radio-labelled with α-^{32}PdCTP (Amersham) by nick translation (Rigby et al. 1977). Clones showing the absence or near-absence of signal on the X-ray film were selected as single-copy clones.

Probes

Single-copy sequence inserts were separated from the plasmid after restriction digestion and electrophoresis in low-melting-temperature agarose. Approximately 30 ng of DNA from each insert was radio-labeled with α-^{32}PdCTP (Amersham) by the random primer method (Feinberg and Vogelstein 1984).

Results

Comparison of the two genomic libraries

The two genomic libraries, one resulting from the digestion of genomic DNA with *Pst*I and the other with *Eco*RI-*Bam*HI, exhibited contrasting characteristics. Whereas 89% of 426 *Pst*I clones were classified as single-copy sequences, only 21% of 484 clones from the *Eco*RI-*Bam*HI library belonged to the same category.

The two libraries also exhibited different levels of polymorphism detected by the single-copy sequences. Eighty-two percent of 61 randomly chosen *Pst*I clones identified polymorphisms with at least one of eight restriction enzymes (*Bam*HI, *Eco*RI, *Eco*RV, *Hae*III, *Hin*dIII, *Hin*fI, *Pst*I, and *Xba*I) in the BAT93 versus Jalo EEP558 comparison. A similar level was observed in a second comparison, 'Rio Tibagi' versus ABA71. This high level of polymorphism detected by the *Pst*I library was in marked contrast with lower levels of polymor-

51

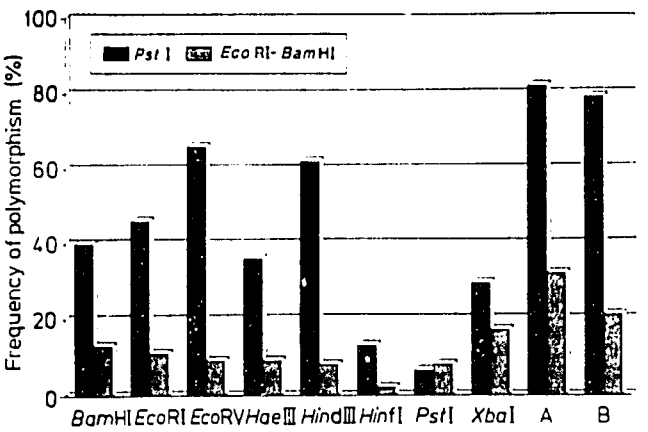


Fig. 1. Level of polymorphism detected by various restriction enzymes. *A* Cumulative polymorphism frequency detected by all eight restriction enzymes; *B* cumulative polymorphism frequency detected by *Eco*RI, *Eco*RV, and *Hin*dIII

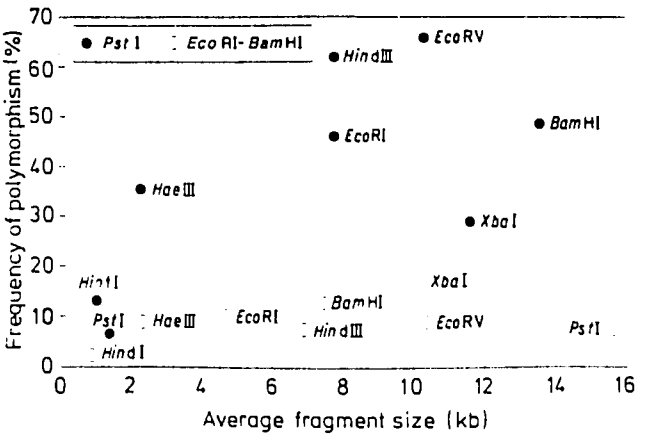


Fig. 2. Relationship between average fragment size and polymorphism level revealed by *Pst*I or *Eco*RI-*Bam*HI clones after restriction digestion of genomic DNA

phism (32%) detected by 57 *Eco*RI-*Bam*HI clones in the BAT93 versus Jalo EEP558 comparison using the same eight restriction enzymes.

When *Pst*I clones were used, we observed that the restriction enzymes used to digest parental DNA detected different levels of polymorphism. In the BAT93 versus Jalo EEP558 comparison, levels of polymorphism ranged between 6.5% (*Pst*I) and 66% (*Eco*RV) (Fig. 1). In the comparison involving 'Rio Tibagi' and ABA71, levels of polymorphism varied between 10% (*Hin*fI) and 67% (*Hin*dIII) (not shown). In both comparisons, high levels of polymorphisms (60–70%) were detected by *Eco*RV and *Hin*dIII, and intermediate levels were detected by *Eco*RI, *Hae*III, *Bam*HI, and *Xba*I. Any polymorphism seen with *Bam*HI, *Hae*III, *Hin*fI, *Pst*I, or *Xba*I, was also seen with *Eco*RI, *Eco*RV, or *Hin*dIII. Hence, the polymorphism observed with the latter three enzymes was as high as that observed with all eight restriction enzymes, i.e., approximately 80% (Fig. 1).

A different pattern was observed for the *Eco*RI-*Bam*HI clones. All of the enzymes detected similar low levels of polymorphism in the comparison BAT93 versus Jalo EEP558 (2–17%). Combined polymorphisms detected by these clones with the eight restriction enzymes amounted to 32%, whereas polymorphisms detected in *Eco*RI-, *Eco*RV-, or *Hin*dIII-digested genomic DNA amounted to 21% (Fig. 1).

What is the probability that an additional clone will reveal a polymorphism between two bean genotypes and, hence, can be mapped? Data from single comparisons – BAT93 versus Jalo EEP558 or Midas versus G12873 – indicated that approximately four-fifths of the *Pst*I clones were capable of detecting a polymorphism with the restriction enzymes *Eco*RI, *Eco*RV, or *Hin*dIII. When the data from two different comparisons were combined, 46 out of 61 clones (75%) detected polymorphisms in both comparisons; 12 out of 61 randomly chosen *Pst*I clones (20%) detected differences in one or the other comparison, and 3 out of 61 clones (5%) detected no polymorphism in either comparison. Hence, using two populations, it would be possible to map 95% of the *Pst*I clones.

Fragment size and frequency of polymorphism

An analysis involving 61 *Pst*I probes in the comparison BAT 93 versus Jalo EEP558 showed significant differences in the average size of the genomic fragments generated by different restriction enzymes (Table 1). Whereas *Hin*fI, *Pst*I, and *Hae*III produced small fragments of sizes 1.0, 1.4, and 2.3 kb, respectively, *Eco*RV, *Xba*I, and *Bam*HI generated fragments with an average size of 10.4, 11.6 and 13.5 kb, respectively. Between these extremes, two restriction enzymes, *Eco*RI and *Hin*dIII, generated fragments with average size of 7.7 and 7.8 kb, respectively. In general, the average fragment size generated by each restriction enzyme was positively correlated with the level of polymorphism revealed by that enzyme (Fig. 2). This association between fragment size and level of polymorphism was not observed for the *Eco*RI-*Bam*HI clones (Fig. 2).

Interversus Intra-gene pool polymorphism

The level of polymorphism was also affected by the degree of genetic similarity among genotypes chosen to identify polymorphisms. Results on levels of polymorphism presented in the previous two paragraphs involved comparisons between Mesoamerican and Andean genotypes. Comparisons within the Andean or Mesoamerican group yielded the following results. Of 57 *Pst*I clones 49% detected polymorphisms between the three cultivated genotypes from the Andean region (ABA71, Jalo EEP558, and Midas) (Table 2). Within the Middle American germ plasm (BAT93, G12873, AC1028, LEF2RB,

52

Table 1. Average fragment size generated by various restriction enzymes

Restriction enzyme	*Pst*I library			*Eco*RI-*Bam*HI library		
	Number of bands	Average size (kb)	Standard error	Number of bands	Average size (kb)	Standard error
*Hin*fI	50	1.0 a[a]	0.06	35	0.8 a	0.08
*Pst*I	49	1.4 b	0.15	33	15.6 f	1.24
*Hae*III	65	2.3 c	0.31	52	2.3 b	0.33
*Eco*RI	57	7.7 d	0.94	37	4.7 c	0.84
*Hin*dIII	65	7.8 d	0.82	40	6.8 cd	0.88
*Eco*RV	58	10.4 de	0.89	35	10.3 d	1.30
*Xba*I	48	11.6 e	1.21	38	10.2 d	1.26
*Bam*HI	52	13.5 e	1.25	32	7.4 cd	1.19

[a] Averages followed by the same letter within a column are significantly different according to pairwise *t*-tests (Steel and Torrie 1980)

Table 2. RFLP levels between and within the Mesoamerican and Andean cultivar groups

Comparison	Number of probes	Number of enzymes	RFLP level (%)
	Between Mesoamerican and Andean cultivar groups		
BAT93, Jalo EEP558	57	6[b]	86
	61	3[c]	82
Midas, G12873	61	3	89
	Within Mesoamerican or Andean cultivar groups		
Mesoamerican[a]	53	6	62
Andean[a]	57	6	49

[a] Mesoamerican: BAT93, G12873, Sierra, LEF2RB, AC1028; Andean: Jalo EEP558, ABA71, Midas
[b] *Bam*HI, *Eco*RI, *Eco*RV, *Hin*dIII, *Pst*I, *Xba*I
[c] *Eco*RI, *Eco*RV, *Hin*dIII

Table 3. Frequency (in percentage) of site mutations and chromosomal rearrangements

Comparison	Number of probes	Frequency (%)		
		Site mutations	Insertions/deletions	Ambiguous
	*Pst*I library			
BAT93 versus Jalo EEP558	50	20	32	48
Rio Tibagi versus ABA71	23	13	35	52
	*Eco*RI-*Bam*HI library			
BAT93 versus Jalo EEP558	17	59	0	41

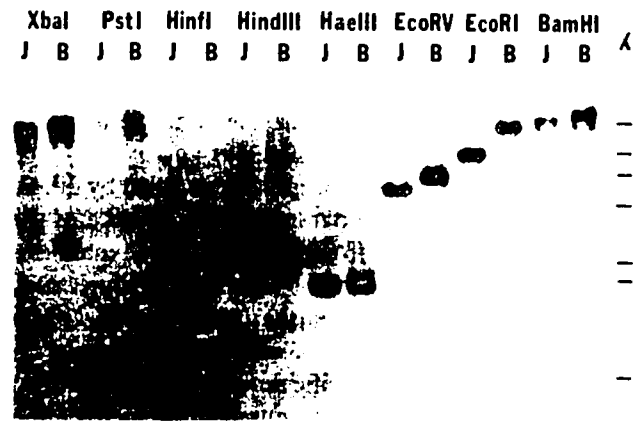


Fig. 3. Polymorphism detected by *Pst*I genomic clone GUC1049 between BAT93 (*B*) and Jalo EEP558 (*J*)

and Sierra), 62% of 53 *Pst*I clones revealed a polymorphism. As expected, therefore, comparisons within the Mesoamerican or Andean groups showed lower levels of polymorphism than between these two groups.

Type of polymorphisms

Polymorphisms can be generated either through nucleotide substitution or through insertions/deletions. If digestion with one or two of the eight restriction enzymes showed a polymorphism, then a nucleotide substitution at the specific recognition site was inferred. If a consistency in the polymorphism was observed among five or more restriction enzymes (e.g., one parent consistently exhibited a larger fragment, as illustrated in Fig. 3), the polymorphism was assumed to be due to an insertion deletion (or chromosomal rearrangement). Intermediate cases were classified as ambiguous. Using these criteria, a majority of the polymorphisms revealed by *Pst*I clones appeared to be due to an insertion/deletion event, whereas *Eco*RI-*Bam*HI clones recognized mostly nucleotide substitutions (Table 3).

Discussion

Libraries enriched with single-copy sequences have been obtained in common bean (95% single-copy sequences; Chase et al. 1991) and in other species such as rice (85%; McCouch et al. 1988), maize (50.6%; Burr et al. 1988), tomato (92%; Miller and Tanksley 1990b), and barley (89%; Graner et al. 1990) when DNA was digested with *Pst*I, a methylation-sensitive enzyme (as opposed to *Bam*HI and *Eco*RI, two methylation-insensitive enzymes). Lack of methylation has been associated with gene activity in vertebrates and plants (e.g., Antequera and Bird 1988; Schwartz 1989; Monk 1990).

The relatively high level of polymorphism revealed by the *Pst*I library compared to that detected by the *Eco*RI-*Bam*HI library (80% versus 32%) was unexpected. *Pst*I clones also exhibited high levels of polymorphisms in other species. In tomato, higher levels of polymorphism were observed with *Pst*I clones than with *Eco*RI clones (Miller and Tanksley 1990a). Of the *Pst*I probes 75% were polymorphic with three enzymes when tested in 38 2n potato lines (Gebhardt et al. 1989).

The wide range of polymorphisms revealed by different libraries illustrate the paramount importance of the restriction enzymes used to digest DNA for the establishment of a genomic library. The cloning procedure, such as the restriction enzyme and the fragment size, apparently induces a sampling bias which favors certain regions of the genome or specific sequences, depending on their base composition, copy number, or methylation level. *Pst*I will not cleave DNA at the recognition sequence (5'-CTGCAG-3') if the cytosine at the 5' end is methylated (Nelson and McClelland 1987; Gruenbaum et al. 1981). Our data suggest that DNA sequences homologous to the *Pst*I clones (or DNA sequences adjacent to them) are more variable than those homologous to the *Eco*RI-*Bam*HI clones. This variability revealed by *Pst*I clones appears to be due more often to small rearrangements than to nucleotide substitutions, although the latter could also play an important role. Variability revealed by *Eco*RI-*Bam*HI clones appeared to be due mostly to nucleotide substitutions. We are not aware of an explanation that could account for this difference in polymorphism levels detected with the two libraries.

Recently, two other groups have obtained estimates of RFLP levels in common bean. Chase et al. (1991) found that in a group of 14 Middle American and Andean accessions, 5 out of 18 (28%) random *Pst*I clones revealed polymorphisms in *Eco*RI digests of genomic DNA. For the *Eco*RV and *Hin*dIII digests, polymorphisms were observed with 5 out of 18 (28%) and 7 out of 18 (39%) of the clones, respectively. Within the Middle American group (excluding accession XR-235-1-1, which includes *P. coccineus* in its background), 5 out of 18 (28%) revealed polymorphisms whereas in the Andean gene pool 4 out of 18 (22%) revealed polymorphisms. The lower levels of polymorphisms observed by these authors can be attributed to differences in the plant material analyzed. For example, all Middle American accessions analyzed by Chase et al. (1991) belong to race Mesoamerica, whereas the accessions analyzed in our study belong to either race Mesoamerica or race Durango of the Middle American gene pool (Singh et al. 1991a).

Guo et al. (1991) observed only 22–26% polymorphisms within *P. vulgaris* but 70–80% between *P. vulgaris* and *P. coccineus*. The lower level of polymorphisms may be due to the use of a single restriction enzyme (*Eco*RI), to the expression of their results as the proportion of polymorphic bands to the total number of bands, and to different sources of probes (mostly anonymous cDNA probes and clones of genes of known function).

Although experimental conditions are not strictly comparable, the RFLP levels we have identified in common bean (around 80–90%) compare favorably with those observed in other species (Table 4), especially self-pollinated plants such as soybean (40–50%) and tomato (14%) (Keim et al. 1989; Helentjaris et al. 1985). Table 2 shows that while between gene pools we were able to detect 80–90% polymorphism, within gene pools we still detected 50–60% polymorphism. Hence, RFLP analysis confirmed the divergence between Mesoamerican and Andean gene pools detected previously with morphological, biochemical, and reproductive isolation data (Vanderborght 1987; Gepts et al. 1986; Gepts 1990; Gepts and Debouck 1991; Sprecher 1988; Khairallah et al. 1990; Koenig and Gepts 1989; Shii et al. 1980; Singh et al. 1991b, c). In order to increase levels of polymorphisms at the molecular level, parents should belong to contrasting gene pools, i.e., Mesoamerican versus Andean. However, the level of polymorphism within gene pools appears to be still high enough to develop efficient applications with these markers, such as marker-assisted selection or evolutionary studies.

Strong correlation have been found between the average size of the hybridizing restriction fragments and the frequency of polymorphisms generated by different restriction enzymes (McCouch et al. 1988; Miller and Tanksley 1990a). By testing 61 probes against BAT93 and Jalo EEP558 genomic DNA digested with eight restriction enzymes, we found similar results. This positive correlation is consistent with our finding that a majority of the RFLPs can be attributed to insertions/deletions.

From a practical point of view, mapping efficiency increases when the number of restriction enzymes used to digest DNA of parents and of segregating generations can be decreased without a significant reduction in polymorphism levels. Our results indicate that the level of polymorphism detected by the combined use of *Eco*RV, *Eco*RI, and *Hin*dIII is the same as the one detected by all

Table 4. Comparison of RFLP levels in common bean and other plants

Crop	Genotypes	Type of clone	Number of enzymes	RFLP (%)	Source
Arabidopsis	3 strains	Genomic: *Eco*RI	3	46	Chang et al. (1988)
Barley	48 cultivars	Genomic: *Pst*I	3	43	Graner et al. (1990)
	4 cultivars	Genomic: *Pst*I	5	54	Heun et al. (1991)
Brassica	3 species	Genomic: *Pst*I	1	95 between species	Figdore et al. (1988)
	(37 cultivars or genetic stocks)			80 between subspecies	Figdore et al. (1988)
				70 within subspecies	Figdore et al. (1988)
Common Bean	7 cultivars + 1 wild	Genomic: *Pst*I	3	89 between subspecies	Nodari et al.
				62 within Mesoamerican	(present results)
				49 within Andean	
Lentil	2 cultivars +	Genomic: *Pst*I, *Eco*RI	4	38	Havey and Muehlbauer
	2 wild	cDNA	4	64	(1989)
Lettuce	4 cultivars	Genomic: *Mbo*I	9	13	Landry et al. (1987)
		cDNA	9	25	Landry et al. (1987)
Maize	3 cultivars	cDNA	3	83	Helentjaris et al. (1985)
Potato	*phureja* × (*tuberosum*	Genomic: *Pst*I, *Eco*RI	11	60	Bonierbale et al. (1988)
	× *chacoense*)	cDNA	11	51	Bonierbale et al. (1988)
Rice	*indica* × *javanica*	Genomic: *Pst*I	11	78	McCouch et al. (1988)
Soybean	5 cultivars	Genomic: *Sau*3AI, *Eco*RI	13	20	Apuya et al. (1980)
	58 wild + cultivated	Genomic: *Sau*3Ai	4	40–50	Keim et al. (1989)
Tomato	2 cultivars	cDNA	3	14	Helentjaris et al. (1985)
	esculentum-cheesmannii	cDNA	3	68	Helentjaris et al. (1985)
	esculentum-pennellii	cDNA	3	100	Helentjaris et al. (1985)

eight restriction enzymes together (around 80%). These three enzymes were also very efficient in detecting polymorphisms in lettuce (Landry et al. 1987) and potato (Bonierbale et al. 1988). When the data from two different pairwise genotypic comparisons were combined, the level of polymorphism was even higher than in a single comparison. This observation suggests that mapping in more than one population will lead to a denser map, in general, and will increase the probability of identifying polymorphic markers in a chromosome region of interest.

Acknowledgements. This work was supported by the United States Agency for International Development and the International Board for Plant Genetic Resources. R. Nodari and E. Koinange are recipients of a fellowship from CNPq and SAD-CC CIAT, respectively. We thank Gisela Arndt, Belén Garrido, and Nino Polignano for assistance, and T. Wilkins for useful discussions.

References

Andersen JW, Story L, Sieling B, Chen W-JL, Petro MS, Story J (1984) Hypocholesterolemic effects of oat-bran or bean intake for hypercholesterolemic men. Am J Clin Nutr 40:1146–1155

Antequera F, Bird AP (1988) Unmethylated CpG islands associated with genes in higher plant DNA. EMBO J 7:2295–2299

Apuya NR, Frazier BL, Keim P, Roth EJ, Lark KG (1988) Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. Theor Appl Genet 75:889–901

Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms and genetic improvement of agriculatural species. Euphytica 35:111–124

Bonierbale M, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095–1103

Burr B, Burr FA, Thompson KH, Albertsen MC, Stuber CW (1988) Gene mapping with recombinant inbreds of maize. Genetics 118:519

Chang C, Bowman JL, DeJohn AW, Lander ES, Meyerowitz EM (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc Natl Acad Sci USA 85:6856–6860

Chase CD, Ortega VM, Vallejos CE (1991) DNA restriction fragment length polymorphisms correlate with isozyme diversity in *Phaseolus vulgaris* L. Theor Appl Genet 81:806–811

Feinberg AP, Vogelstein B (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267

Figdore S, Kennard WC, Song KM, Slocum MK, Osborn TC (1988) Assessment of the degree of restriction fragment length polymorphism in *Brassica*. Theor Appl Genet 75:833–840

Gebhardt C, Ritter E, Debener T, Schachtschabel V, Wakemeier B, Uhrig H, Salamini F (1989) RFLP analysis and linkage mapping in *Solanum tuberosum*. Theor Appl Genet 78:65–75

Gepts P (1990) Biochemical evidence bearing on the domestication of *Phaseolus* beans. Econ Bot 44S:28–38

Gepts P, Clegg MT (1989) Genetic diversity in pearl millet (*Pennisetum glaucum* [L.] R.Br.) at the DNA sequence level. J Hered 80:203–208

Gepts P, Debouck DG (1991) Origin, domestication, and evolution of the common bean, *Phaseolus vulgaris*. In: Voysest O, Van Schoonhoven A (eds) Common beans: research for crop improvement. CAB, Oxon, UK, pp 7–53

Gepts P, Osborn TC, Rashka K, Bliss FA (1986) Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. Econ Bot 40:451–468

Gepts P, Llaca V, Nodari RO, Panella L (1992) Analysis of seed proteins, isozymes, and RFLPs for genetic and evolutionary studies in *Phaseolus*. In: Linskens HF, Jackson JF (eds) Modern methods of plant analysis. New series, Vol. 14. Springer, Berlin Heidelberg New York (in press)

Graner A, Siedler H, Jahoor A, Herrmann RG, Wenzel G (1990) Assessment of the degree and the type of restriction fragment polymorphism in barley (*Hordeum vulgare*). Theor Appl Genet 80:826–832

Gruenbaum Y, Maveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292:260–62

Guo M, Lightfoot DA, Mok MC, Mok DWS (1991) Analyses of *Phaseolus vulgaris* L. and *P. coccineus* Lam. hybrids by RFLP: preferential transmission of *P. vulgaris* alleles. Theor Appl Genet 81:703–709

Harlan JR, de Wet JMJ (1971) Towards a rational classification of cultivated plants. Taxon 20:509–517

Havey MJ, Muehlbauer FJ (1989) Linkages between restriction fragment length, isozyme, and morphological markers in lentil. Theor Appl Genet 77:395–401

Helentjaris T, King G, Slocum M, Siedenstrang C, Wegmann S (1985) Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol Biol 5:109–118

Heun M, Kennedy AE, Anderson JA, Lapitan NLV, Sorrells ME, Tanksley SD (1991) Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*). Genome 34:437–447

Holmes DS, Quigley M (1981) A rapid boiling method for the preparation of bacterial plasmids. Anal Biochem 114:193–197

Johannsen W (1909) Elementen der exakten Erblichkeitslehre. Fisher, Jena, FRG

Keim P, Shoemaker RC, Palmer RG (1989) Restriction fragment polymorphism diversity in soybean. Theor Appl Genet 77:786–792

Kelly JD, Adams MW, Saettler AW, Hosfield GL, Varner GV, Uebersax MA, Taylor J (1990) Registration of 'Sierra' pinto bean. Crop Sci 30:745–746

Khairallah MM, Adams MW, Sears BB (1990) Mitochondrial DNA polymorphins of Malawian bean lines: further evidence for two major gene pools. Theor Appl Genet 80:753–761

Koenig R, Gepts P (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. Theor Appl Genet 78:809–817

Landry BS, Kesseli RV, Farrara B, Michelmore RW (1987) A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331–337

McCouch SR, Kochert G, Yu ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815–829

Mendel G (1865) Experiments on plant hybrids. In: Stern C, Sherwood ER (eds) The origin of genetics. W. H. Freeman, San Francisco, pp 1–48

Miller JC, Tanksley SD (1990a) Effect of different restriction enzymes, probe source, and probe length on detecting restriction fragment length polymorphism in tomato. Theor Appl Genet 80:385–389

Miller JC, Tanksley SD (1990b) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet 80:437–448

Monk M (1990) Variation in epigenetic inheritance. Trends Genet 6:110–114

Nelson M, McClelland M (1987) The effect of site-specific methylation on restriction-modification enzymes. Nucleic Acids Res 15:219–230

Pachico D (1989) Trends in world common bean production. In: Schwartz HF, Pastor-Corrales MA (eds) Bean production problems in the tropics. CIAT, Cali, Colombia, pp 1–8

Rigby PWJ, Dieckman M, Rhodes C, Berg O (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick-translation with DNA polymerase. J Mol Biol 113:237–251

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sax K (1923) The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552–560

Schwartz D (1989) Gene-controlled cytosine demethylation in the promoter region of the *Ac* transposable element in maize. Proc Natl Acad Sci USA 86:2789–2793

Shii CT, Mok MC, Temple SR, Mok DWS (1980) Expression of developmental abnormalities in hybrids of *Phaseolus vulgaris* L. J Hered 71:218–222

Singh SP (1989) Patterns of variation in cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ Bot 43:39–57

Singh SP, Gepts P, Debouck DG (1991a) Races of common bean (*Phaseolus vulgaris* L., Fabaceae). Econ Bot 45:379–396

Singh SP, Gutiérrez JA, Molina A, Urrea C, Gepts P (1991b) Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci 31:23–29

Singh SP, Nodari R, Gepts P (1991c) Genetic diversity in cultivated common bean. I. Allozymes. Crop Sci 31:19–23

Sprecher SL (1988) Allozyme differentiation between gene pools in common bean (*Phaseolus vulgaris* L.), with special reference to Malawian germplasm. Ph.D thesis, Michigan State University, East Lansing, Mich

Steel RGD, Torrie JH (1980) Principles and procedures of statistics. McGraw Hill, New York

Tanksley SD (1983) Molecular markers in plant breeding. Plant Mol Biol Rep 1:3–8

Vanderborght T (1987) The study of common bean (*Phaseolus vulgaris* L.) variability by the use of multivariate statistical methods applied to a data base. Diss Abstr 47:3190B-1B

Weis JH, Quertermous T (1987) Size fractionation using sucrose gradients. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds) Current protocols in molecular biology. Wiley, New York, pp 5.3.2-5.3.8

Analysis of Seed Proteins, Isozymes, and RFLPs for Genetic and Evolutionary Studies in *Phaseolus*

P. GEPTS, V. LLACA, R.O. NODARI, and L. PANELLA

1 General Introduction

Common bean (*Phaseolus vulgaris* L.) has been used repeatedly as an experimental organism to derive important concepts in genetics. After discovering the rules of segregation and independent assortment governing the inheritance of heritable traits in pea (*Pisum sativum* L.), Mendel (1865) confirmed his observations on peas in crosses involving *P. nanus* (nowadays bush *P. vulgaris*) and *P. vulgaris* (nowadays climbing *P. vulgaris*). Johannsen (1909) was instrumental in introducing the distinction between phenotype and genotype following his investigations of seed size in pure lines of *P. vulgaris*. The first example to our knowledge of a linkage between a marker locus and a quantitative trait locus (QTL) was provided by Sax (1923) who established an association between a locus for seed pigmentation and a seed size factor in common bean. More recently, one of the first plant genes to be cloned was a gene for phaseolin seed protein (Sun et al. 1981). A comparison of the nucleotide sequence of cDNA and genomic clones of phaseolin established the presence of intervening sequences in higher plants in addition to other eukaryotic organisms. Our understanding of how plants defend themselves against pathogens has been considerably enhanced by the studies of Lamb and collaborators (reviewed in Dixon and Lamb 1990) on the relationship between common bean and *Colletotrichum lindemuthianum*, causal agent of anthracnose.

Common bean also constitutes an excellent model for evolutionary studies. The species consists of two major subdivisions, which are geographically separated (i.e., Mesoamerican vs. Andean) and show divergence for a wide range of characters, including morphological, agronomic, molecular, and adaptation traits. These two subdivisions are in the process of speciation as suggested by the simple genetic control of reproductive isolation between Mesoamerican and Andean genotypes (Shii et al. 1980; Gepts and Bliss 1985; Gepts 1988b, 1990a; Gepts and Debouck 1991). Hence, common bean can be used to investigate such speciation issues as the presence and configuration of hybrid zones and the molecular basis of reproductive isolation mechanisms. Furthermore, there is increasing evidence that domestication has induced a strong reduction in diversity at the molecular level in *Phaseolus* beans (e.g., Gepts et al. 1986; Schinkel and Gepts 1988). This reduction contrasts with the increase in diversity observed for morphological traits during and after domestication. Hence,

common bean can also be used to study the apparent uncoupling between molecular and phenotypic evolution.

The study of several of these issues has been and will continue to be substantially enhanced through the use of molecular markers. We include among these molecular markers seed proteins, isozymes, and RFLPs. The qualitative analysis of variation for these markers by various electrophoretic techniques provides an assessment of the degree of genetic relatedness (i.e., common ancestry) and levels of genetic diversity free of environmental influence in most cases. In addition, electrophoretic techniques allow us to analyze the larger number of samples necessary in evolution and population studies.

In this review, we describe the procedures and protocols adopted or developed in our laboratory to analyze seed proteins, isozyme, and RFLP variation primarily in *Phaseolus*. We have worked primarily with *P. vulgaris* but our methodology has also been applied in other *Phaseolus* species and in cowpea (*Vigna unguiculata*). For each of these marker types, we briefly illustrate how they have been applied to investigate evolutionary concepts. We end our review by a comparison of the relative merits of the three types of markers as evolutionary markers.

2 Seed Protein Analysis

2.1 Introduction

Seed proteins of common bean contain predominantly globulins, which are salt-soluble (Osborne 1907). Among globulins, phaseolin is the major constituent followed by lectins or phytohemagglutinins, which account for 35%–50% and 5%–10% of total bean seed protein, respectively (Ma and Bliss 1978; Osborn et al. 1985). Phaseolin, also called globulin-1 (G-1: McLeester et al. 1973) or glycoprotein II (Pusztai and Watt 1970), is a 7S vicilin-like protein with homologies to vicilin of *Pisum sativum* and *Vicia faba*, β-conglycinin in *Glycine max*, and β-conglutin in *Lupinus angustifolius* (reviewed in Gepts 1990b). It undergoes a pH-dependent, reversible association between polypeptide subunits of approximately 44 kD (3S) at pH 12, protomers of 163 kD (7.1S) at pH 7, and tetramers of protomers of 653 kD (18.2S) at pH 3.6. The polypeptide subunits range in molecular weight between 43 and 54 kD and in isoelectric point between pH 5.6 and 5.8. Qualitative variation as assessed by electrophoresis is controlled by a single locus (*Phs*) exhibiting codominant expression in heterozygotes (Brown et al. 1981a, b). The *Phs* locus is actually a complex locus constituted by six to eight tightly linked genes belonging to the α and β subfamilies of the phaseolin multigene family (Talbot et al. 1984; Slightom et al. 1985).

Seed lectin, also known as phytohemagglutinin (e.g., Staswick et al. 1986) or glycoprotein I (Pusztai 1966), has polypeptide subunits with molecular

weights of 29 to 36 kD and isoelectric points between pH 4.9 and 7.9. In its native conformation, lectin is a tetramer with molecular weight ranging between 85 and 150 kD. Lectin qualitative electrophoretic variation is coded by a single locus (*Lec*) consisting of two tightly linked genes (Brown et al. 1981a; Hoffman and Donaldson 1985; Osborn and Bliss 1985). The *Lec* locus is unlinked to the *Phs* locus (Brown et al. 1981a).

Arcelin, a seed protein first described by Romero-Andreas et al. (1986), is found in both the globulin and albumin fractions (Osborn 1988). It has polypeptide subunits of molecular weight ranging between 35 and 42 kD and isoelectric points that are more basic than those of phaseolin and lectin. It exhibits 80% nucleotide sequence similarity with lectin (Osborn et al. 1988). The arcelin locus (*Arl*) is tightly linked to the lectin locus (*Lec*: Osborn et al. 1986). For further information on phaseolin, lectin, and arcelin the reader is referred to reviews by Bliss and Brown (1983) and Osborn (1988).

Numerous analytical methods have been used to characterize seed proteins. Of these, electrophoretic techniques have been used most frequently to characterize seed protein variation because of their rapidity, low, cost, and capacity to handle a large number of samples. Stegemann and Pietsch (1983) have outlined the major electrophoretic separation methods, including polyacrylamide gel electrophoresis in one and two dimensions. In the next paragraphs, we will describe these methods of polyacrylamide gel electrophoresis as they are applied specifically to bean seed proteins in our lab. A more general overview of gel electrophoresis of proteins can be found in Hames and Rickwood (1981) and manufacturers of gel apparatuses also provide protocols similar to the ones presented here but with quantities adjusted for the size of their equipment.

2.2 One-Dimensional Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Most electrophoretic analyses of common bean and cowpea seed proteins have used a discontinuous sodium dodecyl sulfate (SDS) polyacrylamide slab gel electrophoresis system as described by Laemmli (1970) because of its high resolving power and the possibility of measuring the molecular weight of individual polypeptides. Polyacrylamide is the preferred separation matrix because it offers fewer problems of reproducibility, is chemically inert and stable over a wide range of pH, temperature, and ionic strength, and it offers a wider range of pore sizes. The discontinuous system is a system in which proteins migrate first through a large-pore "stacking" gel and second through a small-pore "running" gel. When moving through the stacking gel (3% polyacrylamide; pH 6.8), proteins form thin bands (or "stacks") one above the other in order of decreasing mobility; when they reach the running gel (12%–15% polyacrylamide; pH 8.8), the proteins encounter a higher pH and much smaller pore sizes, hence, they will migrate according to their intrinsic charge and size. The discontinuous system results in much greater resolution of individual protein bands in the running gel and has, therefore, become the preferred system.

SDS is an ionic detergent which will bind to proteins in a constant ratio of 1.4 g SDS/g protein and confers a negative charge. As a consequence, migration will be proportional to polypeptide size and not charge density. The added advantage of SDS is that minute amounts of protein (of the order of a µg) are necessary. An alternative dissociating agent is urea at high concentrations (approx. 8 M); unlike SDS, however, it does not affect charge.

2.2.1 Protein Sample Preparation

A flour sample is prepared by removing a portion of the raphe end of the seed, which is the nonembryo end of the seed. If needed for further progeny studies, the seed can thus be saved and germinated. Removal of the raphe end can take place in several ways. It can either be scraped with the edge of a razor blade directly into a 1.5-ml microfuge tube or it can be cut with a razor blade and crushed between glassine paper with a pestle. In either case the presence of the seed coat does not interfere with electrophoretic separation. When seeds are very small, such as seeds of wild relatives, the entire seeds may have to be analyzed destructively by crushing them with a pestle between glassine paper.

The flour sample is suspended for at least 30 min, but usually overnight, in a mixture of equal volumes of 0.5 M NaCl pH 2.4 and cracking buffer (0.0625 M Tris-HCl, pH 6.8; 2 mM EDTA; 2% w/v SDS; 40% w/v sucrose; 1% v/v 2-mercaptoethanol, 0.01% w/v bromophenol blue; Brown et al. 1981 b), to dissolve the seed proteins. Approximately, 25 to 30 mg of flour are suspended in 150 µl of the NaCl – cracking buffer mixture. Protein samples, extracted with cracking buffer only, provide satisfactory analyses as well. The capped tube with the mixture is heated at 100 °C for 10 min and then centrifuged at 15000 g for 5 min at room temperature. Five to 8 µl of the sample supernatant is loaded on the gel.

2.2.2 Gel Assembly

Discontinuous gel separation can only be done with a vertical electrophoresis apparatus, of which several designs exist. Usually slab gels are used because they can accommodate a larger number of samples. Besides commercially available slab gel apparatuses such as the Bio-Rad Protean II (Bio-Rad, Richmond, CA), a popular design has been the Studier (1973) model which can be easily and cheaply made in a laboratory workshop. For actual dimensions, the reader is referred to Hames (1981) or Schuler and Zielinski (1989).

Two glass plates of appropriate dimensions to fit the gel apparatus are used to mold the gel. One plate has a rectangular shape, whereas the other plate of the same overall dimensions has a 2.5-cm notch along one of the long sides leaving two 1-cm-wide "rabbit ears" at the edge of the plate (e.g., Hames 1981). Both plates are cleaned with 95% ethanol to remove any trace of protein and after drying are assembled to form a "sandwich". The plates are held at

Table 1. Composition of running and stacking gel solutions for one-dimensional SDS polyacrylamide gel electrophoresis

Running gel: 30 ml of solution for a 16 cm × 11 cm × 0.75 mm gel		
Stock solution	13% w/v Acrylamide [ml]	15% Acrylamide [m
Deionized H_2O	13.44	11.44
Acrylamide: bisacrylamide, 30:0.15	13.00	15.00
Tris-HCl 3.5 M, pH 6.8	4.00	4.00
SDS 20%	0.20	0.20
EDTA 0.5 M	0.16	0.16
Mix well and add		
Ammonium persulfate 7.5% (*prepared freshly*)	0.32	0.32
N,N,N',N'-tetraethylenediamine (TEMED)	0.04	0.04

Stacking gel Stock solution	Aliquc
	[ml]
Deionized H_2O	4.40
Acrylamide: bisacrylamide, 30: 1.5	0.50
Tris-HCl 1.25 M, pH 6.8	0.50
	[µl]
SDS 20%	[illegible]
EDTA 0.5 M	25
Ammonium persulfate 7.5%	80
N,N,N',N'-tetraethylenediamine (TEMED)	5
Optional: trace of bromophenol blue	

a correct distance with spacers of uniform thickness (between 0.75 anc 1.5 mm) and clamped together with binder clips on each side. The clamps are positioned in such a way that they will press on top of the spacers. To avoic any leakage of acrylamide during gel polymerization between the spacer anc the glass plates, both sides of the spacers are coated with stopcock grease.

One ml of running gel solution (Table 1) is injected with a syringe into the bottom of the gel mold in order to establish an impermeable bottom border. An additional 4 µl of ammonium persulfate 7.5% is added along the spacers at the bottom of the gel to help polymerization of this bottom seal of the running gel. (Care should be exercised not to add too much stopcock grease on the spacers because the excess grease may interfere with polymerization.)

After polymerization of the bottom seal, the remainder of the running gel is poured on top of the bottom border, leaving a space of about 3.5 cm, half of which is filled with H_2O-saturated n-butanol to obtain a straight to gel edge. The gel is allowed to set for at least 45 min and may be left overnight.

Following polymerization of the running gel as shown by the presence of an interface, the butanol layer is removed and the space on top of the running gel is thoroughly rinsed with distilled water (three or four times) and dried with

paper towels. Care should be exercised not to disturb the top edge of the running gel in this process as this will affect subsequent migration. The stacking gel (Table 1) is then poured on top of the running gel and a comb is inserted to form the wells. Air bubbles under the comb teeth should be removed. The upper corners of the comb are clamped with the uppermost glass plate. The stacking gel takes 45 to 60 min to polymerize.

The combs are carefully removed under running distilled water and the wells are rinsed to eliminate any excess acrylamide, after which excess water is also removed. Any well separations that have been disturbed can be repositioned using a needle or a spatula. The glass plates are placed in the gel apparatus and clamped with binder clips. To avoid leakage of the upper chamber buffer, the area around the apparatus notch is coated with stopcock grease. The position of the wells is marked with a felt-tip pen on the glass plates before adding the upper chamber buffer. Alternatively, a small amount of dye can be added to the stacking gel mixture before polymerization (Table 1). A small spatula can be used to clear any air bubbles in the wells.

2.2.3 Sample Loading and Running

Running buffer (0.025 M Tris; 0.192 glycine; 0.1% SDS; 0.002 M EDTA) is added in sufficient quantity to cover the wells in the top chamber by approximately 1 cm; at this time, one should check for any leaks between the apparatus and the gel plates. If none are found, running buffer is also added to cover the bottom of the gel in the lower chamber. The electrophoretic apparatus is then connected to a power supply, with the cathode connected to the top and anode to the bottom chamber. Initially, 7.5 mA per gel is used during gel loading. After all samples have been loaded, the current is increased to 12.5 mA per gel for 1 h or until the continuous blue dye line across the gel, formed by the loaded protein samples, reaches the running gel. For the remainder of the running time (approximately 4 h), a constant current of 30 mA per gel is used until the bromophenol blue dye line reaches the bottom of the running gel. If there is a provision for cooling the gel during operation, a current of up to 50 mA per gel may be used. The cooling of the gel will provide a better resolution of the bands.

2.2.4 Gel Staining and Drying

The glass plates are carefully pried apart and the gel is lifted from the plates and transferred to a glass dish containing Coomassie Brilliant Blue stain (0.15% w/v Coomassie Brilliant Blue R250; 45% v/v methanol; 9% v/v glacial acetic acid). The dish is placed on a shaker for overnight staining. The next day, the stain is replaced with destain solution (45% v/v methanol; 9% v/v glacial acetic acid) and after 5 to 8 h additional shaking (the destain solution can be changed if needed), the gel can be evaluated and photographed.

For permanent storage the gel can be dried according to two methods. In the first one, the gel is air dried by placing it between two sheets of cellophane and tightly clamping it in a plastic frame. Alternatively, the gel can be vacuum dried. In this case, both the gel and the cellophane are soaked overnight in a solution of 2% glycerol and 2% acetic acid on a shaker. If the soaking causes the gel to expand too much, it need not be soaked (but the cellophane must be). The next day, the gel is allowed to dry between the cellophane sheets during 2 h at 80 °C under vacuum. To avoid any acid damage to the pipes, the gel drier exhaust should be connected to a cold trap containing dry ice or liquid nitrogen.

Either before or after drying, molecular weights of the polypeptides can be determined according to the method of Weber and Osborn (1969). Molecular weight marker proteins can be purchased commercially and include, for example, phosphorylase B (94000 daltons), bovine serum albumin (68000 daltons), catalase (58000 daltons), fumarase (49000 daltons), aldolase (40000 daltons), malate dehydrogenase (34000 daltons), and soybean trypsin inhibitor (21000 daltons).

2.3 Two-Dimensional Polyacrylamide Gel Electrophoresis

Proteins can be separated either by mobility determined by size and net charge (as outlined in the previous section) or, alternatively, by equilibrium determined primarily by charge (electrofocusing). With both methods of separation, a given band can consist of different co-migrating proteins. It order to decrease the probability of this coccurrence, two-dimensional electrophoretic methods have been devised that will separate proteins at right angles according to different criteria. The two-dimensional method most used in bean seed protein analysis has been that of O'Farrell (1975) in which proteins are separated by isoelectric focusing in the first dimension and by SDS polyacrylamide gel electrophoresis (SDS/PAGE) in the second dimension. Separation in the first dimension provides an estimate of the isoelectric point (pI), whereas separation in the second dimension provides an estimate of the molecular weight. The specific methodology used to analyze common bean seed protein was first established by Ma and Bliss (1978) and Brown et al. (1981b) and will be described in the next sections.

2.3.1 Protein Sample Preparation

A sample of flour from the raphe end of the seed is suspended in 0.5 M NaCl pH 2.4 for at least 30 min to dissolve seed proteins. After centrifugation at 14000 g for 5 min, the supernatant is mixed 1:1 with sample or lysis buffer (9.5 M urea, 2% w/v NP-40, 5% β-mercaptoethanol, 1.6% pH 5–7 ampholines, and 0.4% pH 3.5–10 ampholines.) Protein samples can be stored at −20 or −70 °C.

2.3.2 First Dimension

The first dimension is run in tube gels of 1–2 mm diameter or less. The tubes are usually made of glass and should be extremely clean. We have used a 1:1 mixture of xylene and Micro detergent (Cole-Parmer), followed by rinses with distilled H_2O and a 1:600 Photoflo solution. The bottom end of each tubes is tightly wrapped in parafilm to avoid leakage of the gel solution and the tubes are maintained in a vertical position during the entire polymerization procedure. The gel solution (Table 2) is injected into the tubes using a syringe to which narrow plastic tubing has been attached. Care is exercised not to trap air bubbles. The top surface of the gel solution in each tube is evened out to obtain first dimension gels of similar length. Each tube gel is overlayed with a few drops of distilled H_2O and left to polymerize for 45–60 min.

The distilled H_2O and the parafilm are discarded. The bottom end of each tube is covered with dialysis tubing secured with rubber bands while care is taken not to trap air bubbles. The tubes are placed in the first dimension gel apparatus with the bottom end of the tubes extending into the acidic buffer of the lower chamber (0.01 M H_3PO_4). The top end of the gels is overlayed with 15 µl 8 M urea and the tubes are filled with the basic buffer of the upper chamber (0.02 M NaOH), after which the upper chamber is then filled with basic buffer and checked for leaks. The pH gradient in the tube gels is created by a pre-electrophoresis consisting of 30 min each at 100, 150, and 200 V (constant voltage).

After the pre-electrophoresis, the upper buffer chamber is discarded and the top of the tubes is rinsed with distilled H_2O. The samples are loaded and overlayed with 15 µl of 15% sucrose and 1% ampholines (7:3 v/v ratio of pH 5–7 and pH 3.5–10 ampholines). The tubes are filled with basic buffer solution, after which the upper chamber is also filled with basic buffer. Electrophoresis takes place – usually overnight – for 16 h at 200 V (constant voltage) followed by 1 h at 400 V (constant voltage).

After electrophoresis, the gels can be stored in their tubes at −70 °C until further analysis in the second dimension or analysis can proceed immediately

Table 2. Composition of gel solution for the isoelectric focusing dimension of two-dimensional IEF/SDS polyacrylamide gel electrophoresis

Urea	2.91 g
Acrylamide: bisacrylamide, 30:2.3	0.8 ml
NP-40 10%	1.2 ml
Deionized H_2O	0.8 ml
Ampholines pH 5–7	0.2 ml
Ampholines pH 3.5–10	0.1 ml
Ammonium persulfate 7.5%	8 µl
N,N,N',N'-tetraethylenediamine (TEMED)	6 µl

as follows. Gels are removed from their tubes by injecting distilled H_2O with a needle between the glass and the gel and rotating the tube. Subsequently, gels are placed in equilibration solution (0.0625 M Tris-HCl pH 6.8, 2.3% w/v SDS, 5% v/v β-mercaptoethanol, 10% w/v glycerol) for at least 20 min.

2.3.3 Second Dimension

The second dimension is run in polyacrylamide slab gels, cast as described in Section 2.2, with two major exceptions. The first exception regards the glass plates used to cast the gel. While one plate is a normal, thin (2.5 mm) gel plate, the other gel plate is thicker and beveled at the top so that the first dimension gel can be positioned and sealed into place. The second exception concerns the stacking gel, which is cast to the top of the plates and without well comb in order to assure continuity with the positioned first-dimension gel. Before polymerization of the stacking gel, individual wells can be formed at the side of the stacking gel for reference to one-dimensional gels, in which case samples should be loaded when the blue front of the second dimension reaches the bottom of the wells.

After equilibration of the first-dimension gels, agarose solution (1% agarose, 0.125 M Tris-HCl pH 6.8, 0.01% SDS, trace of bromophenol blue) is poured to fill the gap between the top of the plate and the first-dimension gel. The tube gel is positioned in the bevel and a note is made of the position of the acid end of the gel. After all air bubbles are removed, the top side of the gel is coated with agarose solution. The second dimension is run similarly to the one-dimension gel (Sect. 2.2) and stained with Coomassie Blue. Gels can be dried and stored as described for one-dimensional gels.

2.4 Applications

Phaseolin electrophoretic type has been the first molecular marker used to identify patterns of genetic diversity and domestication in *Phaseolus vulgaris*, *P. lunatus* (lima bean), and *P. acutifolius* (tepary bean) (Brown et al. 1981b, 1982a; Manen and Otoul 1981; Gepts and Bliss 1986; Gepts et al. 1986; Gepts 1988a; Schinkel and Gepts 1988, Debouck et al. 1989; Koenig et al. 1990; Lioi 1989a, b). Phaseolin diversity data provided the initial evidence that domestication had induced a sometimes marked bottleneck in genetic diversity and, hence, that the current relative lack of diversity in crop plants traced back several thousands of years ago (Gepts et al. 1986; Schinkel and Gepts 1988; Koenig et al. 1990).

In *P. vulgaris*, phaseolin also suggested that the species consisted of two major gene pools, one from Mesoamerica and the other from the Andes. Figure 1 displays phaseolin diversity among some Mesoamerican and Andean cultivars as revealed by one-dimensional polyaycrylamide gel electrophoresis. Through multiple domestications in the two areas, two major groups of cultivars were

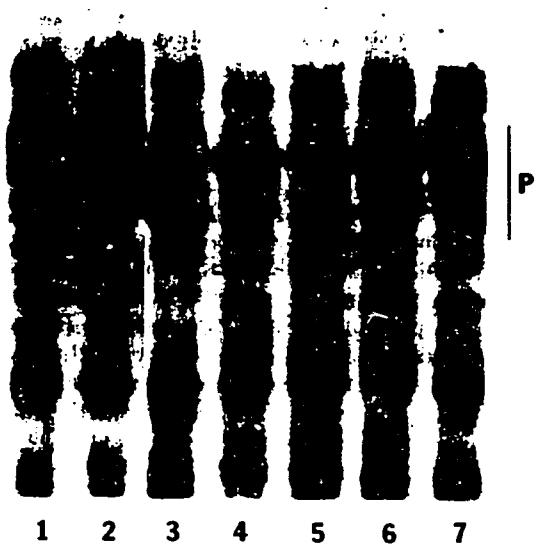


Fig. 1. Genetic diversity for phaseolin in common bean cultivars as determined by one-dimensional SDS polyacrylamide gel electrophoresis. *Lane 1*: "T" phaseolin (cultivar Jalo); *lane 2*: "H" (Cocacho); *lane 3*: "C" (Burrito); *lane 4*: "S" (BAT 93); *lane 5*: "B" (Orgulloso), *lane 6*: "Sb" (Black Turtle Soup); *lane 7*: "Sd" (Mexico 222). *Vertical bar (P)*: molecular weight range of phaseolin polypeptides. T, H, and C are Andean phaseolin types and S, B, Sb, and Sd are Mesoamerican phaseolin types as previously defined on the basis of one- and two-dimensional electrophoresis (Brown et al. 1981b; Gepts et al. 1986; Gepts and Bliss 1986; Koenig et al. 1990)

generated with distinctive agronomic traits (Gepts 1988b, 1990a) and some reproductive isolation (Singh and Guitérrez 1984; Gepts and Bliss 1985).

Lectin variation has also been analyzed by PAGE: 12 electrophoretic variants have been identified (Brown et al. 1981c, 1982b). In addition, a genetic association was identified between qualitative and quantitative lectin variation (Osborn et al. 1985). Arcelin was first identified by PAGE (Osborn et al. 1986; Romero-Andreas et al. 1986) and subsequently was found to possess insecticidal properties (Osborn et al. 1988).

3 Isozyme Analysis

3.1 Introduction

Isozymes have been defined as different variants of the same enzymes, having identical or similar functions, and present in the same individual (Markert and Moller 1959). Since their discovery by Hunter and Markert (1957), they have played an important role in numerous aspects of biological research (Stebbins 1989). In plants, in particular, isozymes have been used principally in population genetic and evolutionary studies (reviewed in Brown 1979; Gottlieb 1981; Tanksley and Orton 1983, Hamrick and Godt 1990). In recent years, several volumes have been published that include general descriptions of isozyme methodology and applications (Tanksley and Orton 1983; Pasteur et al. 1988; Soltis and Soltis 1989; Hillis and Moritz 1990). In this chapter, we will describe our isozyme methodology as applied to *Phaseolus*.

3.2 Preparation

3.2.1 Choice of Plant Tissue

When choosing the plant sample it is important that the enzyme of interest be present in sufficient quantity. Visualization of the enzyme is due to a successfully catalyzed reaction which results either in the deposition of a dye resulting in a colored band or the metabolization of a dye resulting in an achromatic area within the colored gel. In *Phaseolus*, young (2- to 4-week-old) seedlings provide tissue which is rapidly growing and has a large amount of enzymes present. Young tissues (before flowering) also contain less pigmentation than more mature plants and pigmentation can interfere with enzyme function. Only a small amount of tissue is needed (1 to 3 g) and the choice of tissue is generally a matter of trial and error for each species. Some enzymes are present in larger quantities in the roots, leaves, or cotyledons. In general, leaf tissue is easier to use, especially from field-grown plants. If the root tissue is needed, the plants, which may need to grow a little larger to provide enough tissue, can be grown in vermiculite or sand which is easily washed away without harming the plant. This is especially true of the small seeded wild taxa. The seed itself may also be used. It is imbibed overnight and the cotyledons are removed and used. This procedure necessitates the sacrifice of the seed of small seeded taxa but can be useful when screening homogeneous material, especially when time is an important consideration. Using root and leaf tissue allows us to assay a plant and still grow it to maturity. Table 3 lists the source of tissue that has given the best results for *P. vulgaris* under our conditions. Readers should be aware that under their specific conditions other tissues may also give satisfactory results (e.g., Sprecher and Vallejos 1989 for *P. vulgaris*). In addition, different species may require different sources of tissues or buffer systems (e.g., Schinkel and Gepts 1989; Gärvin et al. 1989 in *P. acutifolius*).

62

Table 3. Isozyme systems, loci and alleles in *Phaseolus vulgaris* (Weeden 1984a; Sprecher 1988; Koenig and Gepts 1989a, b; Sprecher and Vallejos 1989)

Isozyme	Enzyme system			Genetic data		
	Buffer[a]	Tissue	Stain recipe	Loci	Alleles	Source
Aconitase (EC 4.2.1.3)	H	Leaves	Morden et al. (1987)	*Aco*	100, 102	R. Nodari (unpubl. data)
Aspartate aminotransferase (EC 2.6.1.1)	L	Root	Vallejos (1983)	*Aat-1*	100	Koenig and Gepts (1989b)
				Aat-2	100	
Diaphorase (EC 1.6.4.3)	H	Root	Vallejos (1983)	*Diap-1*	95, 96, 100, 102	Sprecher (1988), Koenig and Gepts (1989b)
				Diap-2	100, 105	
Fructokinase (EC 2.7.1.4)	H	Leaves	Koenig and Gepts (1989b)	*Fk*	100	Koenig and Gepts (1989b)
Glucose-6-phosphate dehydrogenase (EC 1.1.1.49)	H	Root	Shaw and Prasad (1970)	*G6pd*	100	Koenig and Gepts (1989b)
Glucose and phosphate isomerase (EC 5.3.1.9)	L	Leaves	Vallejos (1983)	*Gpi-c1*	100	Weeden and Liang (1985); Weeden (1986a, b)
				Gpi-c2	96, 100	
Leucine aminopeptidase (EC 3.4.11.1)	L	Leaves	Shaw and Prasad (1970)	*Lap-1*	100	Koenig and Gepts (1989a)
				Lap-2	100	
				Lap-3	100, 103	
Malate dehydrogenase (EC 1.1.1.37)	L	Leaves	Vallejos (1983)	*Mdh-1*	100, 103	Koenig and Gepts (1989a)
				Mdh-2	100, 102	Koenig and Gepts (1989b)
Malic enzyme (EC 1.1.1.40)	L	Roots	(Brewer 1979)	*Me*	98, 100, 102	Weeden (1984a); Koenig and Gepts (1989a)
Methylumbelliferyl esterase	H	Leaves	Pasteur et al. (1988)	*Mue*	100, 103	Garrido et al. (1991)
Peptidase (EC 3.4.--)	L	Leaves	Weeden (1984a)	*Pep*	100	Koenig and Gepts (1989b)
Peroxidase (EC 1.11.1.7)	L	Root	Vallejos (1983)	*Prx*	98, 100	Weeden (1986a, b)
6-Phosphogluconate dehydrogenase (EC 1.1.1.44)	H	Root	Vallejos (1983)	*6Pgd*	100	Koenig and Gepts (1989b)
Ribulose bisphosphate carboxylase (small subunit) (EC 4.1.1.39)	L	Leaves	Weeden (1984a)	*Rbcs*	98, 100	Weeden (1984a)
Shikimate dehydrogenase (EC 1.1.1.25)	L	Leaves	Vallejos (1983)	*Skdh*	100, 103	Weeden (1984a)
Triosephosphate isomerase (EC 5.3.1.1)	H	Leaves	Vallejos (1983)	*Tpi-1*	100	Koenig and Gepts (1989b)
				Tpi-2	100	

[a] L, lithium hydroxide tris borate pH 8.1 (Selander et al. 1971); H, histidine citrate pH 6.5 (Cardy et al. 1980) (see Table 4).

3.2.2 Gel Preparation

Adequate separation of enzyme variants will depend on the buffer system utilized. In order to streamline our analyses, we have attempted to limit the number of buffer systems to two: either lithium hydroxide tris borate pH 8.1 (Selander et al. 1971) or histidine citrate pH 6.5 (Cardy et al. 1980; Table 4). Other buffers have been suggested by Weeden (1984a) and Sprecher and Vallejos (1989). The gels are poured into a horizontal form similar to that described by Johnson and Shaffer (1974). The "legs" are immersed in the buffer reservoirs. The wicking action of the gel is utilized to draw the buffer from the reservoir into the gel.

Table 4. Recipes for isozyme buffers and gels

A. Buffers: *Lithium hydroxide – boric acid buffer*
Lithium hydroxide (0.03 M) 1.2 g
Boric acid (0.19 M) 11.9 g
Add to 800 ml dH_2O and stir until dissolved. Bring volume to 1 l and adjust pH to 8.1 with LiOH or boric acid.

Histidine buffer
L-histidine (0.065 M) 10.1 g
Add to 800 ml dH_2O and stir until dissolved. Bring volume to 1 l and adjust pH to 6.5 with citric acid.

Tris-citrate buffer
Tris 6.2 g
Add to 800 ml dH_2O and stir until dissolved. Bring volume to 1 l and adjust pH to 8.4 with citric acid monohydrate (about 1.2 g/l).

B. Grinding buffer
Glutathione (reduced) 1.5 g
Add to 30 ml dH_2O and stir until dissolved. Bring volume to 40 ml and adjust pH to 7.6 with 1 M unacidified Tris.

C. Gels: *LiOH-boric acid – Tris-citrate starch gel*
Tris-citrate buffer 320 ml
LiOH-boric acid 30 ml
Mix well and separate 150-ml aliquot. The remainder is brought to boiling in the microwave.
Hydrolyzed starch 35 g
Sucrose 10 g
The starch and sucrose are mixed in a 1-l Erlenmeyer flask with side arm. The 200-ml buffer remaining is added and the mixture gently swirled until the starch and sucrose are dissolved. Follow the protocol described in Section 3.2.2.

Histidine starch gel
Histidine buffer 87.5 ml
dH_2O 262.5 ml
Mix well and separate 150-ml aliquot. The remainder is brought to boiling in the microwave.
Hydrolyzed starch 35 g
Sucrose 10 g
The starch and sucrose are mixed in a 1-l Erlenmeyer flask with side arm. The 200-ml buffer remaining is added and the mixture gently swirled until the starch and sucrose are dissolved. Follow the protocol described in Section 3.2.2.

The gel forms are carefully washed and then rinsed with Photoflo [1 : 600 deionized water (dH_2O] dilution] to prevent the starch from sticking to the gel form. Masking tape is used to prevent the heated liquid starch from leaking out of the form before the starch has solidified.

The starch solution can be heated using a microwave. This allows enough starch solution for two gels to be preparated at once. The starch and sucrose (see Table 4 for recipes) are well mixed in an Erlenmeyer flask with side arm (to allow degassing). Half of the buffer (either lithium or histidine) is added. The solution is gently shaken to dissolve the starch and sucrose. The other half of the buffer is heated in the microwave until boiling. The boiling liquid is added to the dissolved starch solution and vigorously shaken. The mixture is then returned to the microwave and heated until it begins to boil. The mixture should be removed from the microwave at 60-s intervals and swirled to insure uniform heating and to prevent the starch from burning onto the bottom of the flask.

Upon boiling the flask is removed (use an insulated glove) and immediately degassed for 30 to 60 s, the point at which the rising air bubbles become larger. The degassed liquid starch is poured carefully into the gel form to avoid forming air bubbles. If a few bubbles form the may be quickly removed with a plastic disposable pipette. If the starch has been degassed too long or burned on the bottom of the flask, pieces of resolidified starch can be seen in the clear liquid. These pieces and air bubbles can interfere with the protein migration and must be avoided. This is the easiest place at which to start over again.

The starch should be allowed to cool on the bench until it begins to turn an opaque white (about 30–45 min). It should then be gently covered with plastic wrap and allowed to set up overnight.

3.3 Loading and Running the Starch Gels

3.3.1 Refrigerate the Gels and Buffer

The gels and running buffer should cool for at least 1 h before loading and running the gel. The gel, wrapped in the plastic wrap, is put into the refrigerator first thing in the morning. The running buffer needed should be measured out and put in a labeled graduated cylinder which is also set into the refrigerator to cool.

3.3.2 Preparing the Plant Tissue

A Plexiglass well board (with 2.5 cm diameter, numbered wells) is placed in a tub of ice at a slight angle and allowed to cool. One hundred fifty to 200 ml of grinding buffer (Table 4) is pipetted into each well to be used (grinding buffer should not be more than 2 weeks old). The plant tissue (leaf, root, or imbibed cotyledon) is put in the numbered wells and macerated in the grinding

buffer using a plastic pestle. The plant tissue should be as fresh as possible. We bring the growing plants in flats into the lab to remove the tissue immediately before grinding it. Once all the samples have been well ground a small spatula is used to push the solid plant material to the uphill side of the well to allow the tissue extract with the freed plant protein to gather at the bottom of the well. Small filter paper wicks (1.5×0.2 cm) are carefully placed into the tissue extract and allowed to absorb the solubilized protein.

3.3.3 Inserting the Wicks in the Gel or Freezing Them

The gel is cut lengthwise 4 cm from the cathodal end (unless an cathodal system is to be run, then it is cut 5 cm from the cathodal end) and wiped dry on either side of the cut (to make it easier to spread the gel at the cut). The wicks are removed from the wells with forceps, dabbed on paper towels to remove adhering plant tissue and excess tissue extract, and placed in the cut through the gel about 2–3 mm apart. The wicks containing the protein can be frozen at this time to be used at a latter date or as a backup in case of problems with the gels. We have kept wicks frozen in microtiter well plates wrapped in cellophane for up to 2 weeks at −20 °C.

Wicks soaked in bromophenol blue dye are placed on the outside of the samples. They are used to mark both where the samples are (when the gel is cut to be removed from the form) and how far the protein has migrated. A plastic spacer is placed between the frame and the short (4 cm) piece to hold the wicks firmly in the gel.

3.3.4 Running the Gel

The two buffer reservoirs are filled with the cooled running buffer. The gels run from the negative (black) to the positive (red) and the forms are placed on the buffer reservoirs with the legs in the buffer. Histidine-citrate gels are run at 10 mA until the proteins (i.e., the dye) has migrated about 1 cm from the wick (30 min). Lithium hydroxyide Tris borate gels are started at 20 mA and run for 40 min. Once the protein has migrated for that time, the wicks are carfully removed using forceps and the gels are covered with plastic wrap. It is always best to turn the power off when handling the gels. Histidine citrate gels are then run at 20 mA and the lithium hydroxide Tris borate gels at 40 mA. The gels run 5 to 6 h; for good reproducibility all gels should be run at the same amperage and for the same length of time each run.

3.3.5 Preparing and Staining Gel Slices

The gel is removed from the reservoirs and cut with a spatula. The front and back area which were in contact with the reservoir are removed. A small piece

is cut from the side to make it easier to remove the gel from the form. The gel is placed on the "cutting board" which consists of two pieces of Plexiglass (1 cm wide × 0.15 cm thick, which are about 10 cm longer than the gel) glued to a glass plate a little wider apart than the width of the gel which is laid between them. Four more equally sized spacers are laid on the glued spacers on either side of the gel to serve as guides. The 4-cm anodal piece of the gel is kept and used to press against the portion of the gel containing the migrated protein. This helps assure an even cut. A "cheese cutter" which has been made using a hack saw frame and a fine guitar string (metal) is used to cut the gel horizontally in 0.15-cm slices. It is slowly and continuously pulled through the gel guided by the spacers. One spacer from each side is removed and the gel is sliced again. This is repeated until there are five cuts, leaving a top and bottom piece (which will be discarded) and four 0.15-cm thick slices. The gel slices are stained in plastic boxes (19 × 11 × 3.5 cm). The stain boxes are lined with plastic wrap and the gel slices are carefully placed in them. The stain is mixed and placed over the gels (or vice versa) which are then set on the shaker to incubate (shake gently or they may break apart). The bands appear in anywhere from 0.5 h to overnight, depending on the enzyme-staining system, with the exception of methylumbelliferyl esterase (MUE), which has to be read on a UV transilluminator within 5 – 10 min after addition of the stain.

3.4 Stains

3.4.1 Staining Preparation

While the gels are running, the stain recipes should be prepared. It is best to start 1 or 2 h before the gel is done, depending on the number of different systems to be stained. Always measure the chemicals and solutions kept at room temperature first, followed by the refrigerated chemicals, and those stored in the freezer last. Whenever possible use stock solutions because it is both faster and more accurate to pipet small amounts. We have found that NAD and NADP solutions (10 mg/1 ml) can be aliquoted and frozen at −70 °C, PMS at −20 °C, and MTT refrigerated, with no loss of activity. When needed, they can be thawed and kept on ice until mixed into the stain.

It is best to measure the reagents beforehand, keeping the temperature-sensitive ones at the proper temperature until the stain solution is mixed. We line the boxes with plastic wrap because it makes the gel slices easier to handle after they have been stained. The staining solution should be mixed and added immediately to the gel slice in the box. It is then incubated on the shaker the requisite time needed for the staining reaction to occur. It is important to check each gel at regular intervals because some will reach an optimum point rather quickly (within 0.5 h) and after that point the background increases, making scoring more difficult.

3.4.2 Stain Recipes

For the sake of conciseness, stain recipes will not be repeated here. They have been detailed by several authors among which Shaw and Prasad (1970), Vallejos (1983), Pasteur et al. (1988), Wendel and Weeden (1989), and Murphy et al. (1990). Table 3 lists the particular recipe we have used in our *Phaseolus* analyses.

3.5 Interpretation of Gels

3.5.1 Naming Loci and Alleles

Nomenclature for isozyme loci and alleles in *Phaseolus* follows the proposal of Myers and Weeden (1988). Loci are designated using the appropriate two to four italicized letter abbreviation of the biochemical name of the enzyme, the first letter of the abbreviation being capitalized (e.g. *Me* for malic enzyme, *Skdh* for shikimate dehydrogenase). The accepted names for most enzymes are given in the International Union of Biochemistry (1984) and appropriate abbreviations have been established for most enzymes (Wendel and Weeden 1989). When more than one isozyme locus is identified, they are distinguished by a numeric suffix, the most anodal isozyme being given the suffix 1 (e.g., *Diap-1, Diap-2*).

The most common allele at each locus is designated 100 and all other alleles are designated by their migration distance measured in millimeters from the standard (e.g., slower migrating: 98; faster migrating: 102) (Koenig and Gepts 1989a, b). Figure 2 shows polymorphisms for various enzymes analyzed as described in Table 3. The two *P. vulgaris* reference cultivars included in all our gels have the following genotype:

ICA-Pijao: $Rbcs^{100}$, $Skdh^{103}$, Prx^{100}, Me^{100}, $Mdh\text{-}1^{100}$, $Mdh\text{-}2^{100}$, $Diap\text{-}1^{95}$, $Diap\text{-}2^{105}$, Lap^{100}, $Aco\text{-}2^{100}$, Mue^{100}

California Dark Red Kidney: $Rbcs^{98}$, $Skdh^{100}$, Prx^{98}, Me^{98}, $Mdh\text{-}1^{103}$, $Mdh\text{-}2^{100}$, $Diap\text{-}1^{100}$, $Diap\text{-}2^{100}$, $Lap\text{-}3^{103}$, $Aco\text{-}2^{102}$, Mue^{103}

3.5.2 Genetic Control

Table 3 lists the number of loci and their respective alleles as observed in our analyses of *P. vulgaris* genetic diversity. Additional polymorphic loci have been identified by Weeden (1984a, b; 1986a, b) and Weeden and Liang (1985). They include adenylate kinase (*Adk*; two alleles), esterase (*Est-1* and *Est-2*; each two alleles), glucose phosphate isomerase (*Gpi-c1*; two alleles), and N-acetyl glucoseaminidase (*Nag*). Most loci are unlinked with the possible exception of *Rbcs* and *Me* (Weeden 1984b; Koenig and Gepts 1989a), which in certain populations form a linkage group with the lectin-arcelin locus: *Rbcs*-[*Lec-Arl*]-*Me*.

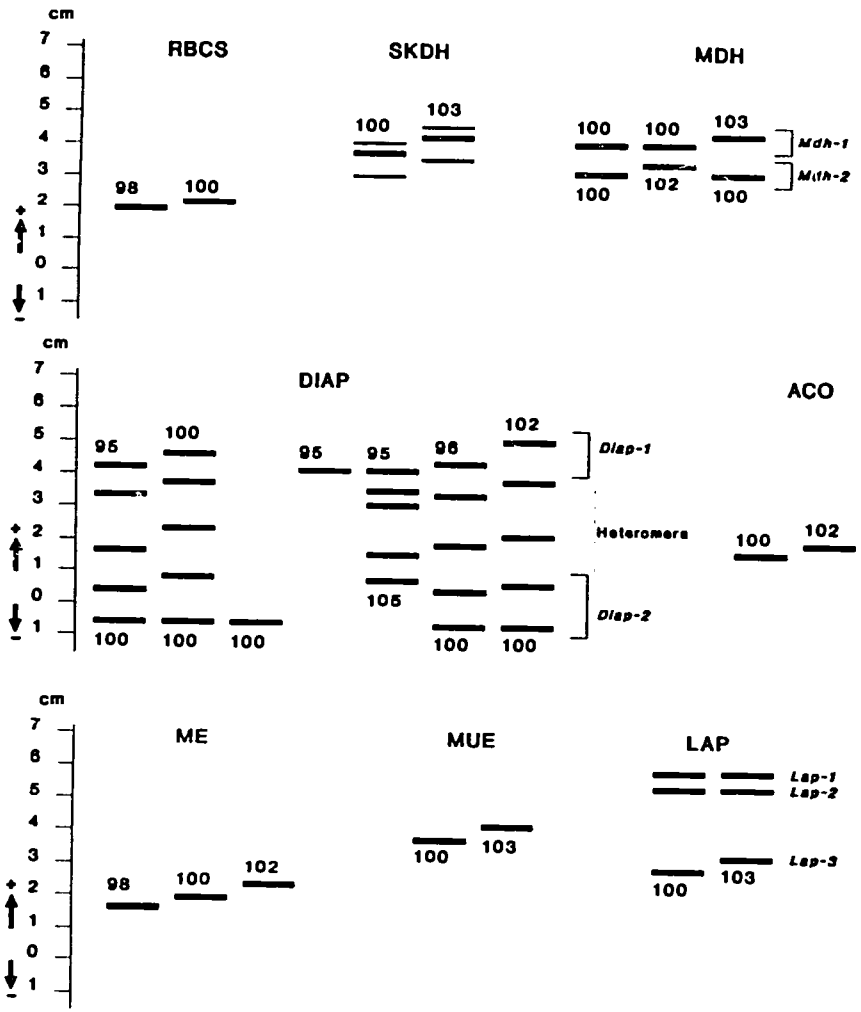


Fig. 2. Allozyme diversity for eight enzyme systems in wild and cultivated common bean. *RBCS* ribulose bisphosphate carboxylase (small subunit); *SKDH* shikimate dehydrogenase; *MDH* malate dehydrogenase; *DIAP* diaphorase; *ACO* aconitase; *ME* malic enzyme; *MUE* methylumbelliferyl esterase; *LAP* leucine aminopeptidase. For locus and allele designations, see Section 3.5.1

This general absence of linkage indicates that the polymorphic isozyme loci characterize different regions of the genome, an important feature in evolutionary studies.

3.6 Preservation of Gels

It is important that gels be preserved in order to compare results across experiments. Gels can be placed overnight in 50% ethanol and then stored indefinitely at 4 °C. It is important that the gels be scored before this is done since some resolution is lost through this processing. Photographs can also be taken at this time. A 35-mm camera with 100 ASA Ektachrome film is adequate to photograph the gel on a light box. These slides provide a more easily stored copy of the gel and are especially useful when a centimeter scale is included in the photograph. The lab has also used a Bio-Rad gel dryer at 70 °C for 1 h. The gels are dried between cellophane in the same manner as polyacrylamide gels (and on the same drying cycle), however, the dyes stain the backing and cover of the gel dryer quite severely.

3.7 Applications

The earliest isozyme analyses in *Phaseolus* were performed by Wall and Wall (1975) and Bassiri and Adams (1978a, b). More recently, several more detailed studies have been conducted on the patterns of genetic diversity in *P. vulgaris* and *P. acutifolius* (Weeden 1984a; Sprecher 1988, Koenig and Gepts 1989b; Singh et al. 1991b).

In *P. vulgaris*, these studies have revealed or confirmed that: (1) the species consists of two diverged geographical subdivisions – Mesoamerican vs. Andean – which, based on independent data, appear to represent two subspecies (Sprecher 1988; Koenig and Gepts 1989b; Singh et al. 1991b); (2) the species is characterized by extensive multilocus associations even among unlinked loci (Gepts, unpubl. observ.); (3) even in presumed secondary centers of diversity where both Mesoamerican and Andean genotypes are grown together, there is a dearth of recombinants between the two groups for reasons unknown so far (Sprecher 1988); (4) snap beans, which as a group are very homogeneous morphologically, appear to be relatively more diverse than other commercial bean classes (Weeden 1984a); and (5) certain allozyme variants correlate with morphological and geographical variation and have allowed us to identify races of related cultivars in both Andean and Mesoamerican cultivars (Singh et al. 1989, 1991a).

In *P. acutifolius*, isozyme analyses have confirmed the reduction of diversity during domestication revealed by phaseolin (Schinkel and Gepts 1989). They also identified geographic patterns of variation, such as the divergence between populations west and east of the Sierra Madre Occidental in Mexico.

4 RFLP Analysis

4.1 Introduction

The methodology for restriction fragment length polymorphism (RFLP) analysis is based on the properties of Type II restriction enzymes. Restriction enzymes are sequence-specific bacterial endonucleases. Of the three types of restriction enzymes that have been described (Yuan 1981), the Type II enzymes comprise ATP-independent endonucleases cleaving DNA at specific sites within or very close to short (usually 4 or 6 bp) recognition sequences. Hence, this cleavage activity will produce equimolar amounts of the same discrete fragments for a given DNA molecule. The fragments obtained by restriction digestion can then be separated by size using agarose gel electrophoresis and visualized by ethidium bromide staining. Defined fragments can be easily detected from the bulk of DNA fragments by hybridization procedures using radiolabeled fragment-specific sequences as probes and exposure to X-ray films.

Differences in the array of restriction fragments between two related sequences is indicative of modification in the DNA primary structure. These modifications can be due to gain/loss of recognition sites, rearrangements, or insertion/deletion of sequences between sites. RFLPs are thus molecular markers providing data about quantitative and qualitative differences between two sequences. Currently, they constitute one of the most efficient methods for detection of variation at the DNA level, being surpassed only by direct sequencing. Although RFLP analysis is less informative than direct sequencing, it is faster and less expensive. Moreover, in cases where many different samples must be analyzed, it is the only feasible procedure. As molecular markers, plant unclear RFLPs have already been broadly used in applied and basic genetics, systematics, and evolution (reviewed in Helentjaris and Burr 1989).

Variation at the DNA sequence level as determined by RFLP analysis has been examined for various seed proteins of common bean. These include phaseolin (e.g., Talbot et al. 1984), lectin (e.g. Staswick et al. 1986), and arcelin (Osborn et al. 1988). For example, phaseolin types can be recognized by specific restriction fragment patterns (Fig. 3). Each of these patterns is produced by the hybridization of a radiolabeled phaseolin gene fragment to the 7–8 genes in the phaseolin gene family. Nevertheless, because of tight linkage at the *Phs* locus, all the bands corresponding to the various phaseolin genes tend to segregate as a single Mendelian unit and represent a simple co-dominant marker (Brown et al. 1981a).

4.2 Methodology

Our methodology for RFLP analysis of bean DNA sequences, and, in particular, bean seed protein genes can be divided as follows:

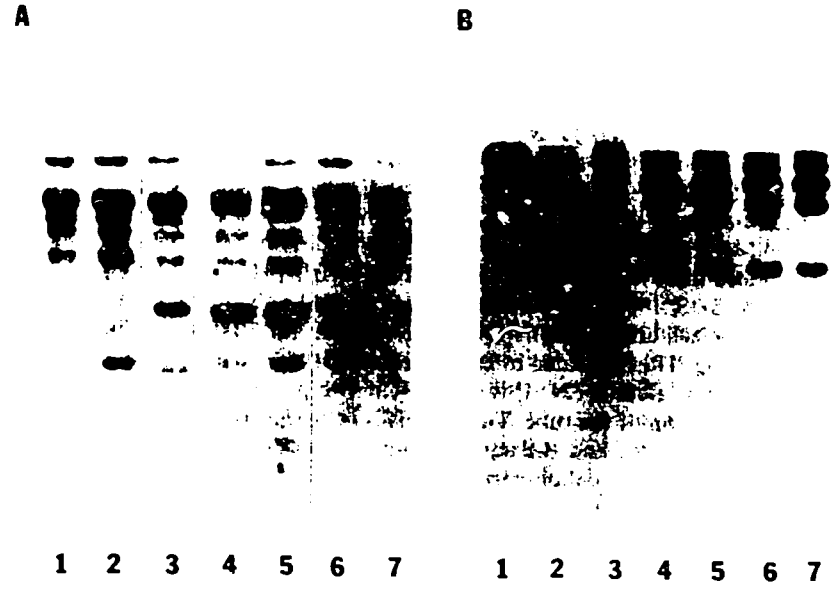


Fig. 3. Genetic diversity for phaseolin in common bean cultivars as determined by RFLP analysis. Genomic DNA of lines with various phaseolin types was digested with *Eco*RI (A) or *Hind*III (B), electrophoresed in a 0.8% agarose gel, Southern blotted, and hybridized to phaseolin probe MC3 (provided by JL Slightom). *Lane 1*: "T" phaseolin (cultivar Jalo); *lane 2*: "H" (Cocacho); *lane 3*: "C" (Burrito); *lane 4*: "S" (BAT 93); *lane 5*: "B" (Orgulloso); *lane 6*: "Sb" (Black Turtle Soup); *lane 7*: "Sd" (Mexico 222). T, H, and C are Andean phaseolin types and S, B, Sb, and Sd are Mesoamerican phaseolin types as previously defined on the basis of one- and two-dimensional electrophoresis (Brown et al. 1981b; Gepts and Bliss 1986; Gepts et al. 1986; Koenig et al. 1990)

1) Total or nuclear DNA is obtained from leaf tissue.
2) The DNA is digested with restriction enzymes.
3) The restriction fragments are separated by agarose gel electrophoresis.
4) Fragments in the gel are denatured and transferred to a nylon or nitrocellulose membrane.
5) Probes (genomic or cDNA) are radioactively labeled.
6) Hybridization of DNA on the membranes with the labeled probes.
7) The membrane is exposed to an X-ray film in order to visualize the restriction patterns for the specific DNA region.

4.2.1 DNA Extraction

Currently, a number of protocols for genomic DNA extraction from plants have been developed (Bendich et al. 1980; Murray and Thompson 1980; Saghai-Maroof et al. 1984; Rogers and Bendich 1988; Doyle and Doyle 1990). We have found that a modification of the method by Murray and Thompson

(1980; as modified by Saghai-Maroof et al. 1984 and Gepts and Clegg 1989) and Doyle and Doyle (1990) in leaf tissue is particularly adequate for *Phaseolus*. This method is relatively fast and inexpensive, and allows us to handle a large number of samples at the same time. Yields are generally good (approximately 50–100 µg/g fresh tissue) and the average size of the DNA molecules (approximately 40 kb) is sufficient for RFLP analysis.

Young leaves (4–6 cm) are harvested and used immediately (alternatively, they can be stored at −70 °C for several months). Three to 5 g of leaf tissue is ground to a powder in a mortar containing liquid nitrogen. The ground tissue is placed in 10 ml extraction buffer (50 mM Tris pH 9, 0.7 M NaCl, 1 mM EDTA, 1% SDS, 1% β-mercaptoethanol, 50 mg/ml polyvinyl polypyrrolidone or PVPP), mixed and promptly incubated 60 °C for 30–60 min, with occasional swirling. The incubated mixture is cooled off for 10 min. Ten ml of chloroform: isoamyl alcohol (24:1) is added. The two phases are mixed repeatedly by gentle inversion of the tube. The sample is centrifuged at 1600 g for 15 min at room temperature, and the aqueous (upper) phase is recovered. A second chloroform: isoamyl alcohol extraction is made. The aqueous phase is transferred and two-thirds of its volume in isopropanol is added. The tube is inverted gently several times until DNA strands appear. DNA strands are recovered by hooking them out with a bent Pasteur pipet (if the strands are too sheared to be hooked out, DNA is pelleted by centrifuging at 1600 g for a few seconds and the supernatant is discarded). DNA is removed to a sterile tube containing washing solution (76% ethanol, 10 mM ammonium acetate). At this stage, the DNA strands may have a greenish or brownish tint, which is usually eliminated after incubating 2–12 h at room temperature in washing solution with occasional swirling. If needed, the washing solution can be replaced with fresh solution. After washing, the solution is discarded, the DNA is air-dried and resuspended in 300–500 µl TE (10 mM Tris·HCl pH 8, 1 mM EDTA). After the DNA is dissolved the solution is centrifuged at 3500 g for 10 min and the supernatant is saved.

In some cases, the DNA appears brownish even after a long washing period, probably due to the oxidation of polyphenolics (Rogers and Bendich 1988). In this case, we suggest a digestibility test for the specific sample prior to further work. Extracted DNA may be further purified by a cesium chloride gradient (Sambrook et al. 1989). This procedure yields highly purified DNA, eliminating RNA, polysaccharides, and other contaminants, but it is time-consuming and expensive. In practice, we have seen that most of the samples can be efficiently digested without cesium chloride gradients.

The DNA concentration of samples purified by cesium chloride gradient can be estimated by UV spectrophotometry at 260 nm (1 unit, equivalent to 50 µg/ml DNA; Sambrook et al. 1989). However, if the sample is going to be used without purification, UV estimates are not reliable. In that case, DNA can be more precisely quantified using a fluorometer (Brunk et al. 1979; Cesarone et al. 1979). The final concentration for the extracted DNA in the TE solution should be around 0.4 to 1 µg/µl. For short periods, DNA solutions can be stored at 4 °C; long-term storage takes place at −70 °C.

4.2.2 Digestion

Restriction enzymes have different optimal assay conditions. All require Mg^{2+} as cofactor, and most of them have a similar pH range (7.4–8) and incubation temperature (37 °C). The most critical difference lies in the ionic strength (Fuchs and Blakesley 1983). Most of the manufacturers provide a buffer of the appropriate concentration and instructions to achieve optimal activity for each enzyme. Alternatively, a standard buffer (50 mM Tris·HCl pH 8, 10 mM $MgCl_2$, 1 mM dithiothreitol, 100 µg/ml bovine serum albumin) can be used, in which low, medium or high concentrations (0, 50, and 100 mM, respectively) of NaCl are used according to the enzyme (Fuchs and Blakesley 1983).

In a typical assay, 5–10 µg DNA is digested in a total volume of 30–50 µl. Generally, the restriction enzyme is added in five- to six-fold excess (25–30 units for a 5 µg digestion), and the reaction mix is incubated for 4–7 h, in order to achieve total digestion. If the sample to be used has not been purified by cesium chloride gradient, 1–4 µg DNAse-free RNAse (Sambrook et al. 1989) is added.

4.2.3 Gel Electrophoresis

Electrophoresis is performed according to Sambrook et al. (1989). Agarose at a concentration of 0.8% is dissolved in TAE (40 mM Tris-Acetate pH 7.4, 1 mM EDTA) to which 0.5 µg/ml ethidium bromide is added. For a large sample number, we use large horizontal gels (25×20 cm, available from several companies (e.g., Bethesda Research Laboratories, Gaithersburg, MD), allowing us to run up to 60 samples at the same time. The digested samples are mixed (6:1) with a solution containing glycerol (50%) and bromophenol blue (0.5%), and loaded into the wells. One µg lambda phage DNA digested with *Hind*III is used as size marker (Sambrook et al. 1989). The optimal voltage to be used is 1 V/cm.

After the overnight electrophoresis run, DNA is visualized by ethidium bromide fluorescence using a UV transilluminator to check for digestion. Completely digested DNA should be seen as a smear along the lane; some bands (e.g., cpDNA, highly repetitive DNA) may be brighter in relation to the smeared background. Incompletely digested or undigested DNA samples will show a strong band above the 23000 base pairs molecular weight marker band. The positions of the lambda *Hind*III fragments should be marked by puncturing the gel with an 18-gauge needle containing India ink in order to establish the correct position of the restriction fragments in the next steps.

4.2.4 Transfer

Restriction fragments can be transferred to either nylon- or nitrocellulose-based membranes. Nylon membranes are more desirable than nitrocellulose paper

because they have a much higher mechanical strength than the former and, hence, can be reused several times. In our hands, Zetabind membranes (AMF-CUNO, Meriden, CT) can be reused 10 to 15 times. Protocols can change slightly according to the properties of the membranes and it is therefore recommended to follow the protocol of the manufacturers. We describe here the Zetabind protocol.

The gel is washed for 30 min at 40–60 rpm in a tray with 0.2 N NaOH, 0.6 M NaCl to denature the DNA. The denaturing solution is replaced then by 0.5 M Tris pH 7.5, 1.5 M NaCl and again shaken gently (40–60 rpm) for 30 min, to neutralize the gel. Capillary transfer is made overnight according to Southern (1975; cited in Sambrook et al. 1989).

Membranes onto which DNA has been transferred are prehybridized in sealed plastic bags for 2–24 h at 42 °C in 5×SSC (20×SSC = 3 M NaCl, 0.3 M sodium citrate, pH 7), 10×Denhardt's solution (100×Denhardt's = 2% Ficoll, 2% polyvinyl pyrrolidone, 2% bovine serum albumin), 50 mM $NaPO_4$ pH 6.7, 5% dextran sulfate, 50% formamide, and 500 µg/ml sonicated or sheared salmon sperm DNA. Sonicated or sheared salmon sperm should be heated in a water bath at 100 °C for 10 min and immediately cooled off on ice for 10 min prior to addition to the prehybridization solution in order to maintain denaturation. Five ml of prehybridization solution is used for each 100 cm^2 of membrane. Several membranes can be place in the same bag. If not used immediately, they can be stored in their plastic bag at 4 °C.

4.2.5 Probes

Probes include either genomic or cDNA cloned fragments. They can be either anonymous sequences or sequences of specific genes, some of which are expressed in seeds such as phaseolin (Slightom et al. 1985), lectin (Hoffman and Donaldson 1985), or arcelin (Osborn et al. 1988). Plasmids containing clone sequences are amplified in *E. coli* liquid cultures and isolated by the quick-boiling minipreparation (Holmes and Quigley 1981). The vector is then cut with the appropriate restriction enzymes. Once digested, the fragment is isolated by low melting point agarose gel electrophoresis or by DEAE-NA45 paper affinity (Selden and Chory 1987).

The cloned sequence probe is radioactively labeled by the random primer method (Feinberg and Vogelstein 1984); kits for random priming are available from several manufacturers, who also give instructions for use. Alternatively, the component solutions such as the nucleotide solutions or the reaction buffer can be prepared in the lab. After labeling, unincorporated nucleotides are removed with a Sephadex G50 spun column as described by Sambrook et al. (1989). Specific activity for the probe is determined by measuring a small aliquot (1–2 µl) in a scintillation counter and should be at least $1\times10^8-10^9$ dpm/µg of probe to minimize background and exposure time. A total of $0.6-1\times10^7$ dpm of probe is used for each 100 cm^2 of membrane to be hybridized.

4.2.6 Hybridization

The prehybridization solution is totally removed and the following hybridization solution is added: 5×SSC, 1×Denhardt's, 20 mM $NaPO_4$ pH 6.7, 10% dextran sulfate ml, 50% formamide, and 100 mg/ml boiled sonicated or sheared salmon sperm DNA. This solution also contains the probe, previously denatured by boiling and cooled on ice.

Hybridization is carried on in a temperature-controlled orbital shaker (42 °C, 150 rpm) for 20 h according to Sambrook et al. (1989). After hybridization, the membrane is removed from the bag and briefly washed at low stringency in 2×SSC, 0.1% SDS at room temperature. Membranes are then washed twice in 0.1×SSC, 0.1% SDS at 60 °C. Membranes are lightly blotted to remove excess washing solution, wrapped in plastic wrap, and exposed to a Kodak X-Omat X-ray film. Exposure times can vary from 1 to several days, depending on the amount of DNA sample used and/or probe-specific activity.

4.2.7 Reuse of Solutions and Membranes

Nylon membranes can be reused by washing for 30 min in 0.4 M NaOH, neutralizing with 0.1×SSC, 0.5% SDS, 0.2 M Tris pH 7.5 for 30 min, and prehybridizing as described.

Prehybridization and hybridization solutions can be recovered after use and stored at −20 °C. Before reusing, the solutions are heated at 75–80 °C for 15 min in a water bath, and immediately added to the incubation bag. Prehybridization solution can be reused 3–4 times and stored for weeks. Hybridization solution should not be stored for more than 1 week.

4.3. Applications

This methodology has been applied successfully to several areas of study in our program that are still under way: (1) RFLP levels in wild and cultivated *Phaseolus vulgaris*. Figure 3 shows restriction fragment length polymorphisms detected with a phaseolin probe (R. Nodari, E. Koinange, V. Becerra, J. Kelly, and P. Gepts, in prep.); (2) chloroplast DNA evolution in the *P. vulgaris* – *P. coccineus* complex (V. Llaca and P. Gepts, in prep.); and (3) mapping of the common bean genome (Gepts et al. 1990).

5. Conclusion: Comparison of the Three Classes of Evolutionary Markers in Evolutionary Studies

Each of the three types of markers discussed in this chapter possess advantages and disadvantages as evolutionary markers. In general, an ideal marker should

exhibit a sufficiently high level of polymorphism to identify genetic diversity among the entries of the studies; yet, this level of polymorphism should not be so high as to include repeated or parallel mutations that would obscure patterns of common ancestry. The specific marker(s) most suitable for a given study will depend on the inherent variability of the marker and the genetic distance among entries of the study. For example, chloroplast DNA is generally highly conserved and is therefore most useful at higher taxonomic levels (Palmer 1987); on the other hand, hypervariable markers such M13 or human minisatellite sequences are usually sufficiently variable to detect differences among individuals of the same species, whether animal (Jeffries et al. 1985) or plant species (Dallas 1988; Rogstad et al. 1988). On the other hand, rare variants may be useful in detecting introgression and gene flow, for example between wild ancestral beans and cultivated descendant beans (e.g., the *Mdh-2*^{102} allele in *P. vulgaris;* Singh et al. 1991b). An empirical preliminary study will have to be conducted to identify the most suitable markers.

The second desirable attribute of evolutionary markers is environmental stability. Ideally, the expression of evolutionary markers should be independent of any confounding environmental influence. If this is the case, differences can be attributed exclusively to genetic causes. The third attribute is the number of loci detected by any category of markers. Categories that detect larger numbers of markers, preferably unlinked, will afford a more complete coverage of the genome; this will permit the detection of recombinants, translocations, inversions, linkage drag, etc.

A fourth attribute is the molecular basis of the variation. The more complex the molecular basis, the less likely that a particular variant will have appeared more than one. Hence, individuals that share a variant with a complex molecular basis, will most probably have a common ancestor (Gepts 1990a, b). Finally, preference should be given whenever possible to markers that are easy and cheap to analyze. In order to increase confidence levels, evolutionary studies require fairly large sample numbers that can only be achieved with markers that can be analyzed in a routine fashion.

How do these three categories of markers – phaseolin, isozmyes, and RFLPs – match these five attributes (Table 5)? As argued by Gepts (1990b), the advantages of phaseolin as a marker include a high level of polymorphism, a high degree of environmental stability, a complex molecular basis of the electrophoretic banding pattern, along with the simplicity and low cost of the

Table 5. Comparison of molecular electrophoretic markers in evolutionary studies

	Seed protein (phaseolin)	Isozymes	RFLPs
Polymorphism	High	Low	High
Environmental stability	High	Moderate	High
Number of loci	Single locus	Low (15 – 20 loci)	High
Molecular basis	Complex	Simple	Intermediate
Practicality	Quick, cheap	Quick, cheap	Slow, expensive

analysis. Its major disadvantage is the simple genetic control since all variation in the banding pattern is controlled by a single locus (*Phs*; Brown et al. 1981a; Bassett 1989).

The advantages of isozymes are the larger number (15 – 20) of (unlinked) loci and the simplicity and low cost of analysis. Their disadvantages include the low levels of polymorphism, their environmental sensitivity (which can be alleviated by working under standardized conditions), and their simple molecular basis (which can be alleviated by considering several isozmye loci simultaneously).

The advantages of RFLPs are their high levels of polymorphism, not only between the Mesoamerican and Andean subspecies, but also within these subspecies (Nodari, Koinange, Kelly, and Gepts, in prep.), their high level of environmental stability, and the high (potentially infinite) number of loci. Their main disadvantages is the high cost and cumbersomeness of the analysis and to a lesser extent the simple molecular basis (in many cases, single nucleotide substitutions), which can also be overcome by considering several RFLPs simultaneously.

Because each of these markers have advantages and disadvantages, none represents a straightforward choice when initiating a study. The particular marker(s) chosen will depend principally on the resources available (personnell, equipment, funds) and prior informtaion about molecular variation in the plant materials to be studied. Our own approach has been to start with the easiest and cheapest marker (phaseolin), continue with isozymes on a subset of the entries analyzed for phaseolin, and, finally to analyze an even smaller subset for RFLPs. Both isozymes and RFLPs have provided new information (e.g., strong multilocus associations shown by isozyme data) or confirmed previous findings (e.g., reduction of diversity during domestication by RFLPs). Hence, our repeated analyses with different markers have not been futile exercises but are providing a more in-depth and dependable picture of the evolutionary history of *Phaseolus* beans.

Acknowledgments. Our research on diversity at the molecular level in *Phaseolus* beans is funded by the U.S. Agency for International Development, the International Board for Plant Genetic resources, and the Lindbergh Fund. V. Llaca holds a fellowship of DGAPA, UNAM, Mexico. R.O. Nodari held, while at Davis, a fellowship of CNPq, Brazil.

References

Bassett MJ (1989) List of genes. Annu Rep Bean Improv Coop 32:1 – 15

Bassiri A, Adams MW (1978a) An electrophoretic survey of seedling isozymes in several *Phaseolus* species. Euphytica 27:447 – 459

Bassiri A, Adams MW (1978b) Evaluation of bean cultivar relationships by means of isozyme electrophoretic patterns. Euphytica 27:707 – 720

Bendich AJ, Anderson RS, Ward BL (1980) Plant DNA: long, pure, and simple. in: Leaver CJ (ed) Genome organization and expression. Plenum New York, pp 31 – 33

Bliss FA, Brown JWS (1983) Breeding common bean for improved quantity and quality of seed protein. Plant Breed Rev 1:59–102
Brown AHD (1979) Enzyme polymorphism in plant populations. Theor Popul Biol 15:1–42
Brown JWS, Bliss FA, Hall TC (1981a) Linkage relationships between genes controlling seed proteins in French beans. Theor Appl Genet 60:251–259
Brown JWS, Ma Y, Bliss FA, Hall TC (1981b) Genetic variation in the subunits of globulin-1 storage protein of French bean. Theor Appl Genet 59:83–88
Brown JWS, McFerson JR, Bliss FA, Hall TC (1982a) Genetic divergence among commercial classes of *Phaseolus vulgaris* in relation to phaseolin patterns. HortScience 17:752–754
Brown JWS, Osborn TC, Bliss FA, Hall TC (1982b) Bean lectin 2: relationship between qualitative lectin variation in *Phaseolus vulgaris* L. and previous observations on purified bean lectins. Theor Appl Genet 62:361–367
Brown JWS, Osborn TC, Bliss FA, Hall TC (1981c) Genetic variation in the subunits of globulin-2 and albumin seed proteins of French bean. Theor Appl Genet 60:245–250
Brunk C, Jones K, James T (1979) Assay for nanogram quantities of DNA in cellular homogenates. Anal Biochem 92:497–500
Cardy BJ, Stuber CW, Goodman MM (1980) Techniques for starch gel electrophoresis of enzymes from maize (*Zea mays* L.). Department of Statistics Mimeo Series No. 1317, North Carolina State University, Raleigh
Cesarone C, Bolognesi C, Santi L (1979) Improved fluorometric DNA determination in biological material using 3358 Hoechst. Anal Biochem 100:188–197
Dallas JF (1988) Detection of DNA "fingerprints" of cultivated rice by hybridization with a human minisatellite DNA probe. Proc Natl Acad Sci USA 85:6831–6835
Debouck DG, Maquet A, Posso CE (1989) Biochemical evidence for two different gene pools in lima beans. Annu Rep Bean Improv Coop 32:58–59
Dixon RA, Lamb CJ (1990) Molecular communication in interactions between plants and microbial pathogens. Annu Rev Plant Phys Plant Mol Biol 41:339–367
Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15
Feinberg AP, Vogelstein B (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267
Fuchs R, Blakesley R (1983) Guide to the use of type II restriction endonucleases. Methods Enzymol 100:3–38
Garrido B, Nodari R, Debouck DG, Gepts P (1991) *Uni-2*, a dominant mutation affecting leaf development in *Phaseolus vulgaris*. J Hered 82:181–183
Garvin DF, Roose ML, Waines JG (1989) Isozyme genetics and linkage in tepary bean, *Phaseolus acutifolius* A. Gray. J Hered 80:373–376
Gepts P (1988a) Phaseolin as an evolutionary marker. In: Gepts P (ed) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, pp 215–241
Gepts P (1988b) A Middle American and an Andean gene pool. In: Gepts P (ed) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, pp 375–390
Gepts P (1990a) Biochemical evidence bearing on the domestication of *Phaseolus* beans. Econ Bot 44(3):S28–S38
Gepts P (1990b) Genetic diversity of seed storage proteins in plants. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding, and genetic resources. Sinauer, Sunderland, MA, pp 64–82
Gepts P, Bliss FA (1985) F_1 hybrid weakness in the common bean: differential geographic origin suggests two gene pools in cultivated bean germplasm. J Hered 76:447–450
Gepts P, Bliss FA (1986) Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ Bot 40:469–478
Gepts P, Clegg MT (1989) Genetic diversity in pearl millet (*Pennisetum glaucum* [L.] R.Br.) at the DNA sequence level. J Hered 80:203–208
Gepts P, Debouck DG (1991) Origin, domestication, and evolution of the common bean, *Phaseolus vulgaris*. In: Voysest O, Van Schoonhoven A (eds) Common beans: research for crop improvement. CIAT, Cali, Colombia, pp 7–53

Gepts P, Osborn TC, Rashka K, Bliss FA (1986) Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. Econ Bot 40:451–468
Gepts P, Singh S, Nodari R, Garrido B, Koinange E (1990) Towards an integrated linkage map of common bean (*Phaseolus vulgaris* L.). J Cell Biochem (Suppl) 14E:281
Gottlieb LD (1981) Electrophoretic evidence and plant populations. Progr Phytochem 7:1–46
Hames BD (1981) An introduction to polyacrylamide gel electrophoresis. In: Hames BD, Rickwood D (eds) Gel electrophoresis of proteins. IRL, Oxford, pp 1–91
Hames BD, Rickwood D (1981) Gel electrophoresis of proteins: a practical approach. IRL, Oxford
Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding, and genetic resources. Sinauer, Sunderland, MA, pp 43–63
Helentjaris T, Burr B (1989) Development and application of molecular markers to problems in plant genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Hillis DM, Moritz C (1990) Molecular systematics. Sinauer, Sunderland, MA
Hoffman LM, Donaldson DD (1985) Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. EMBO J 4:883–889
Holmes DS, Quigley M (1981) A rapid boiling method for the preparation of bacterial plasmids. Anal Biochem 114:193–197
Hunter RL, Markert CL (1957) Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. Science 125:1294–1295
International Union of Biochemistry (1984) Enzyme nomenclature. Acad Press, New York
Jeffreys AJ, Wilson V, Thein SL (1985) Hpyervariable 'minisatellite' regions in human DNA. Nature 314:67–73
Johannsen W (1909) Elemente der exakten Erblichkeitslehre. Fischer, Jena, Germany
Johnson FM, Schaffer HE (1974) An inexpensive apparatus for horizontal gel electrophoresis. Isozyme Bull 7:4–6
Koenig R, Gepts P (1989a) Segregation and linkage of genes for seed proteins, isozymes, and morphological traits in common bean (*Phaseolus vulgaris*). J Hered 80:455–459
Koenig R, Gepts P (1989b) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. Theor Appl Genet 78:809–817
Koenig R, Singh SP, Gepts P (1990) Novel phaseolin types in wild and cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ Bot 44:50–60
Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685
Lioi L (1989a) Geographical variation of phaseolin patterns in an old world collection of *Phaseolus vulgaris*. Seed Sci Technol 17:317–324
Lioi L (1989b) Variation of the storage protein phaseolin in common bean (*Phaseolin vulgaris* L.) from the Mediterranean area. Euphytica 44:151–155
Ma Y, Bliss FA (1978) Seed proteins of common bean. Crop Sci 17:431–437
Manen JF, Otoul E (1981) Etudes électrophorétiques et détermination des fractions protéiques principales chez quelques cultivars élites de *Phaseolus lunatus* L. et de *Phaseolus vulgaris* L. Bull Rech Agron Gembloux 16:309–326
Markert CL, Moller F (1959) Multiple forms of enzymes: tissue, ontogenetic and species specific patterns. Proc Natl Acad Sci USA 45:753–763
McLeester RC, Hall TC, Sun SM, Bliss FA (1973) Comparison of globulin proteins from *Phaseolus vulgaris* with those of *Vicia faba*. Phytochemistry 2:85–93
Mendel G (1865) Experiments on plant hybrids. In: Stern C, Sherwood ER (eds) The origin of genetics. Freeman, San Francisco, pp 1–48
Murphy RW, Sites JW Jr, Buth DG, Haufler CH (1990) Proteins I: isozyme electrophoresis. In: Hillis DM, Moritz C (eds) Molecular systematics. Sinauer, Sunderland, MA, pp 45–126
Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucl Acid Res 8:4322–4325
Myers JR, Weeden NF (1988) A proposed revision of guidelines for genetic analysis in *Phaseolus vulgaris* L. Annu Rep Bean Improv Coop 31:16–19

O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007–4021
Osborn TC (1988) Genetic control of bean seed protein. CRC Crit Rev Plant Sci 7:93–116
Osborn TC, Bliss FA (1985) Effects of genetically removing lectin seed protein on horticultural and seed characteristics of common bean. J Am Soc Hortic Sci 110:484–488
Osborn TC, Brown JWS, Bliss FA (1985) Bean lectin 5: quantitative genetic variation in seed lectins of *Phaseolus vulgaris* and its relationship to qualitative genetic variation. Theor Appl Genet 70:22–31
Osborn TC, Blake T, Gepts P, Bliss FA (1986) Bean arcelin. 2. Genetic variation, inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris*. Theor Appl Genet 71:847–855
Osborn TC, Alexander DC, Sun SSM, Cardona C, Bliss FA (1988) Insecticidal activity and lectin homology of arcelin seed protein. Science 240:207–210
Osborne TB (1907) The vegetable proteins. Longmans and Green, London
Palmer JD (1987) Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Am Nat 130:S6–S29
Pasteur N, Pasteur G, Bonhomme F, Catalan J, Britton-Davidian J (1988) Practical isozyme genetics. Ellis Horwood, Chichester
Pusztai A (1966) The isolation of two proteins, glycoprotein I and a trypsin inhibitor, from seed of kidney bean (*Phaseolus vulgaris*). Biochem J 101:379–384
Pusztai A, Watt WB (1970) Glycoprotein. II. The isolation and characterization of a major antigenic and non-hemagglutinating glycoprotein from *Phaseolus vulgaris*. Biochim Biophys Acta 207:413–431
Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. In: Gelvin SB, Schilperoort RA (eds) Plant molecular biology manual. Kluwer, Dordrecht, pp A6:1–10
Rogstad SH, Patton JC II, Schaal BA (1988) M13 repeat probe detects DNA minisatellite-like sequences in gymnosperms and angiosperms. Proc Natl Acad Sci USA 85:9176–9178
Romero-Andreas J, Yandell BS, Bliss FA (1986) Bean arcelin. 1. Inheritance of a novel seed protein of *Phaseolus vulgaris* L. and its effect on seed composition. Theor Appl Genet 72:123–128
Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018
Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
Sax K (1923) The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552–560
Schinkel C, Gepts P (1988) Phaseolin diversity in the tepary bean, *Phaseolus acutifolius* A. Gray. Plant Breed 101:292–301
Schinkel C, Gepts P (1989) Allozyme variability in the tepary bean, *Phaseolus acutifolius* A. Gray. Plant Breed 102:182–195
Schuler MA, Zielinski RE (1989) Methods in plant molecular biology. Acad Press, San Diego
Selander RK, Smith MH, Yang SY, Johnson WE, Gentry JR (1971) Biochemical polymorphism and systematics in the genus *Peromyscus*. I. Variation in the old-field mouse (*Peromyscus polionotus*). Stud Genet VI Univ Texas Publ 7103:49–90
Selden RF, Chory J (1987) Isolation and purification of large DNA restriction fragments from agarose gels. In: Ausubel FM, Brent R, Kingston R, Moore DD, Seidman JG, Smith JA, Struhl K (eds) Current protocols in molecular biology. Wiley, New York, pp 2.6.1–2.6.8
Shaw CR, Prasad R (1970) Starch gel electrophoresis: a compilation of recipes. Biochem Genet 4:297–320
Shii CT, Mok MC, Temple SR, Mok DWS (1980) Expression of developmental abnormalities in hybrids of *Phaseolus vulgaris* L. J Hered 71:218–222
Singh SP, Gutiérrez AJ (1984) Geographical distribution of the DL_1 and DL_2 genes causing hybrid dwarfism in *Phaseolus vulgaris* L., their association with seed size, and their significance to breeding. Euphytica 33:337–345

Singh SP, Debouck DG, Gepts P (1989) Races of common bean, *Phaseolus vulgaris* L. In: Beebe S (ed) Current topics in breeding of common bean, Proc Int Bean Breed Workshop. CIAT, Cali, Colombia, pp 75–89
Singh SP, Gutiérrez JA, Molina A, Urrea C, Gepts P (1991a) Genetic diversity in cultivated common bean II. Marker-based analysis of morphological and agronomic traits. Crop Sci 31:23–29
Singh SP, Nodari R, Gepts P (1991b) Genetic diversity in cultivated common bean. I. Allozymes. Crop Sci 31:19–23
Slightom JL, Drong RF, Klassy RC, Hoffman LM (1985) Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. Nucl Acid Res 13:6483–6498
Soltis DE, Soltis PS (1989) Isozymes in plant biology. Dioscorides, Portland, OR
Sprecher SL (1988) Allozyme differentiation between gene pools in common bean (*Phaseolus vulgaris* L.), with special reference to Malawian germplasm. PhD Thesis, Michigan State University, East Lansing, MI
Sprecher SL, Vallejos CE (1989) System/tissue/enzyme combinations for starch gel electrophoresis of bean. Annu Rep Bean Improv Coop 32:32–33
Staswick P, Chapell J, Voelker T, Vitale A, Chrispeels M (1986) Molecular biology of seed storage proteins and lectins. In: Shannon LM, Chrispeels M (eds) Proc 9th Annu Symp Plant Phys Univ of California, Riverside, 9–11 January 1986. Am Soc Plant Phys, Rockville, MD, pp 107–115
Stebbins GL (1989) Introduction. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides, Portland, OR, pp 1–14
Stegemann H, Pietsch G (1983) Methods for quantitative and qualitative characterization of seed proteins of cereals and legumes. In: Gottschalk W, Müller HP (eds) Seed proteins: biochemistry, genetics, nutritive value. Martinus Nijhoff/Dr. W. Junk, The Hague, pp 45–75
Studier FW (1973) Analysis of bacteriophage T7 early RNAs and proteins on slab gels. J Mol Biol 79:237–248
Sun SM, Slightom JL, Hall TC (1981) Intervening sequences in a plant gene – comparison of the partial sequence of cDNA and genomic DNA of French bean phaseolin. Nature 289:37–41
Talbot DR, Adang MJ, Slightom JL, Hall TC (1984) Size and organization of a multigene family encoding phaseolin, the major seed storage protein of *Phaseolus vulgaris* L. Mol Gen Genet 198:42–49
Tanksley SD, Orton TJ (1983) Isozymes in plant genetics and breeding. Elsevier, Amsterdam
Vallejos CE (1983) Enzyme activity staining. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding. Elsevier, Amsterdam, pp 469–516
Wall JR, Wall SW (1975) Isozyme polymorphisms in the study of evolution in the *Phaseolus vulgaris-P. coccineus* complex of Mexico. In: Markert CL (ed) Isozymes IV. Acad Press, New York, pp 287–305
Weber K, Osborn M (1969) The reliability of molecular weight determinations by sodium dodecyl sulphate-polyacrylamide gel electrophoresis. J Biol Chem 244:4406–4412
Weeden NF (1984a) Distinguishing among white-seeded bean cultivars by means of allozyme genotypes. Euphytica 33:199–208
Weeden NF (1984b) Linkage between the gene coding the small subunit of ribulose biphosphate carboxylase and the gene coding malic enzyme in *Phaseolus vulgaris*. Annu Rep Bean Improv Coop 27:123–124
Weeden NF (1986a) Enzyme loci defined in *Phaseolus vulgaris*. Annu Rep Bean Improv Coop 29–53
Weeden NF (1986b) Genetic confirmation that the variation in the zymograms of 3 enzyme systems is produced by allelic polymorphism. Annu Rep Bean Improv Coop 29:117–118
Weeden NF, Liang CY (1985) Detection of a linkage between white flower color and *Est-2* in common bean. Annu Rep Bean Improv Coop 28:87–88
Wendel JF, Weeden NF (1989) Visualization and interpretation of plant isozymes. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides, Portland, OR, pp 5–45
Yuan R (1981) Structure and mechanisms of multifunctional restriction endonucleases. Annu Rev Biochem 50:285–316

In: O'Brien SJ (ed.), Genetic Maps 1992. Cold Spirng Harbor Laboratory Press, Cold Spring Harbor, NY

Linkage Map of Common Bean (*Phaseolus vulgaris* L.) 2N = 22

July 1, 1992

Paul Gepts
Department of Agronomy and Range Science
University of California
Davis, CA 95616-8515
Tel: (916) 752-7743 Fax: (916) 752-4361
Bitnet: plgepts@ucdavis Internet: plgepts@ucdavis.edu

The Genetics committee of the Bean Improvement Cooperative is a central clearinghouse for genetic data on common bean and allied species. Its functions are: 1) to establish guidelines and nomenclature rules for assigning new gene symbols; 2) to review manuscripts involving the interpretation of genetic control and assignment of new gene symbols; 3) to maintain a list of published gene symbols for mapped and unmapped loci (latest version: see 5). and 4) to establish a Genetic Stock Collection. Authors wishing additional information should contact the author.

In this report, a classical and a molecular map are presented, as well as a table of miscellaneous reported linkages. The classical map was established primarily by Lamprecht (36) with additional work by several authors as reviewed by Bassett (6). It involves mainly morphological traits and consists of 12 linkage groups (Fig. 1 and Table 1). Distances represent recombination fractions (in %). Three markers were linked to the [*I-B*] locus on linkage group III; however, their respective has not been determined yet. Two versions, established independently, exist for linkage group VII.

The molecular map was established by Nodari et al. (46) and involves primarily molecular markers (RFLPs, RAPDs, and isozymes) and additionally a virus (Bean Common Mosaic Virus) resistance gene (*I*), a flower color gene (Fc; presumably *V*), and a seedcoat color pattern gene (*Cor*) (Fig. 2 and Table 2). It currently consists of 14 linkage groups (linkage groups D6 and D12 have been joined) covering 960 cM. All intervals have a LOD score > 3.0 and map distance < 30 cM (Kosambi distance) except two intervals (on D1, D6) where the map distance > 30cM but the LOD score is > 3.0.

The correspondence between classical and molecular linkage groups is as follows: D1 - IV, D2 - III and VI, D6 - I and VII, D7 - X and XII. Additional miscellaneous linkages are reported in Table 3.

Note added in proof: An additional common bean molecular linkage map was recently published (65)

Figure 1. Common Bean Classical Linkage Map

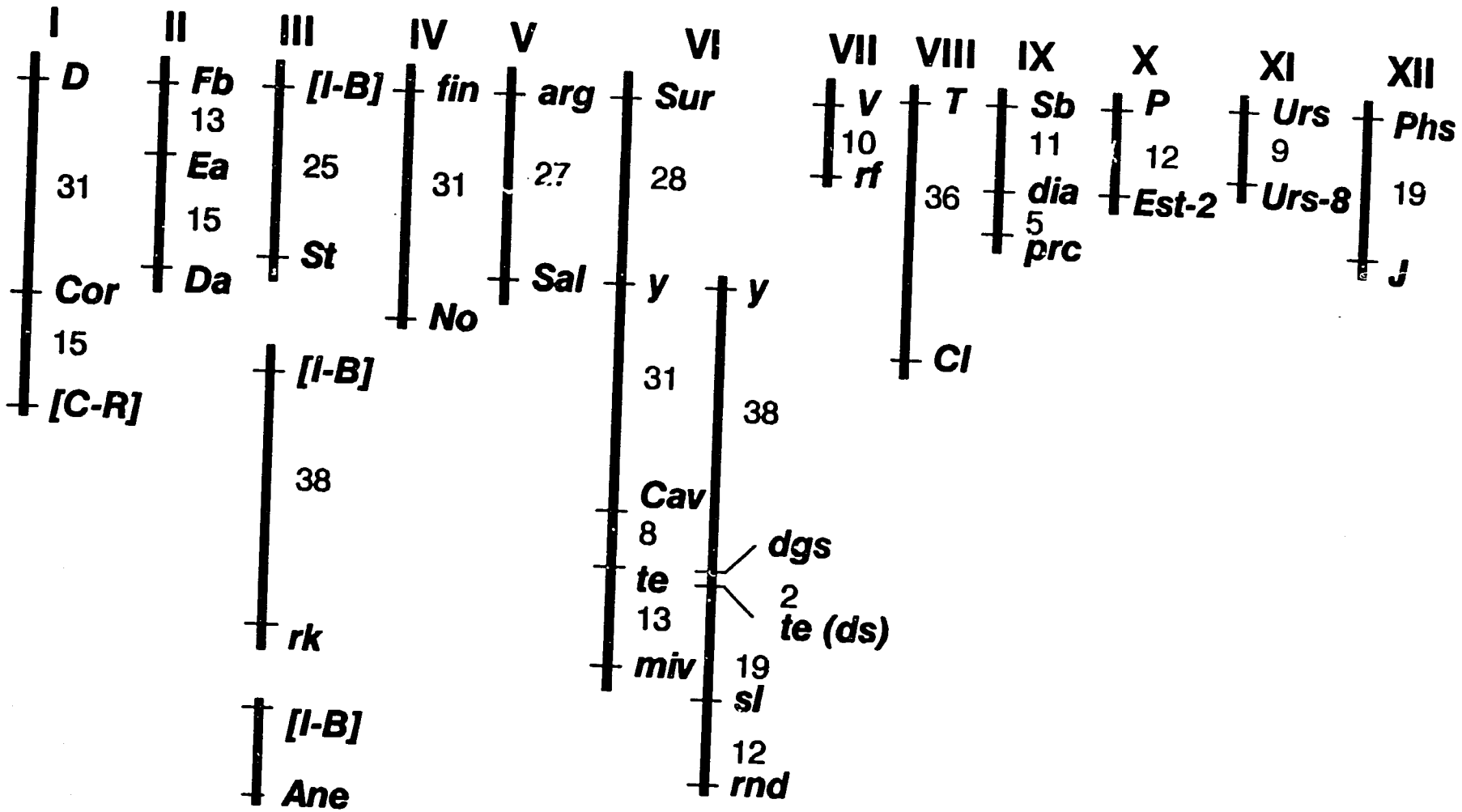

I
D
31
Cor
15
[C-R]
II
Fb
13
Ea
15
Da
III
[I-B]
25
St
[I-B]
38
rk
[I-B]
Ane
IV
fin
31
No
V
arg
27
Sal
VI
Sur
28
y
31
Cav
8
te
13
miv
y
38
dgs
2
te (ds)
19
sl
12
rnd
VII
V
10
rf
VIII
T
36
Cl
IX
Sb
11
dia
5
prc
X
P
12
Est-2
XI
Urs
9
Urs-8
XII
Phs
19
J

Figure 2. Common Bean Molecular Linkage Map

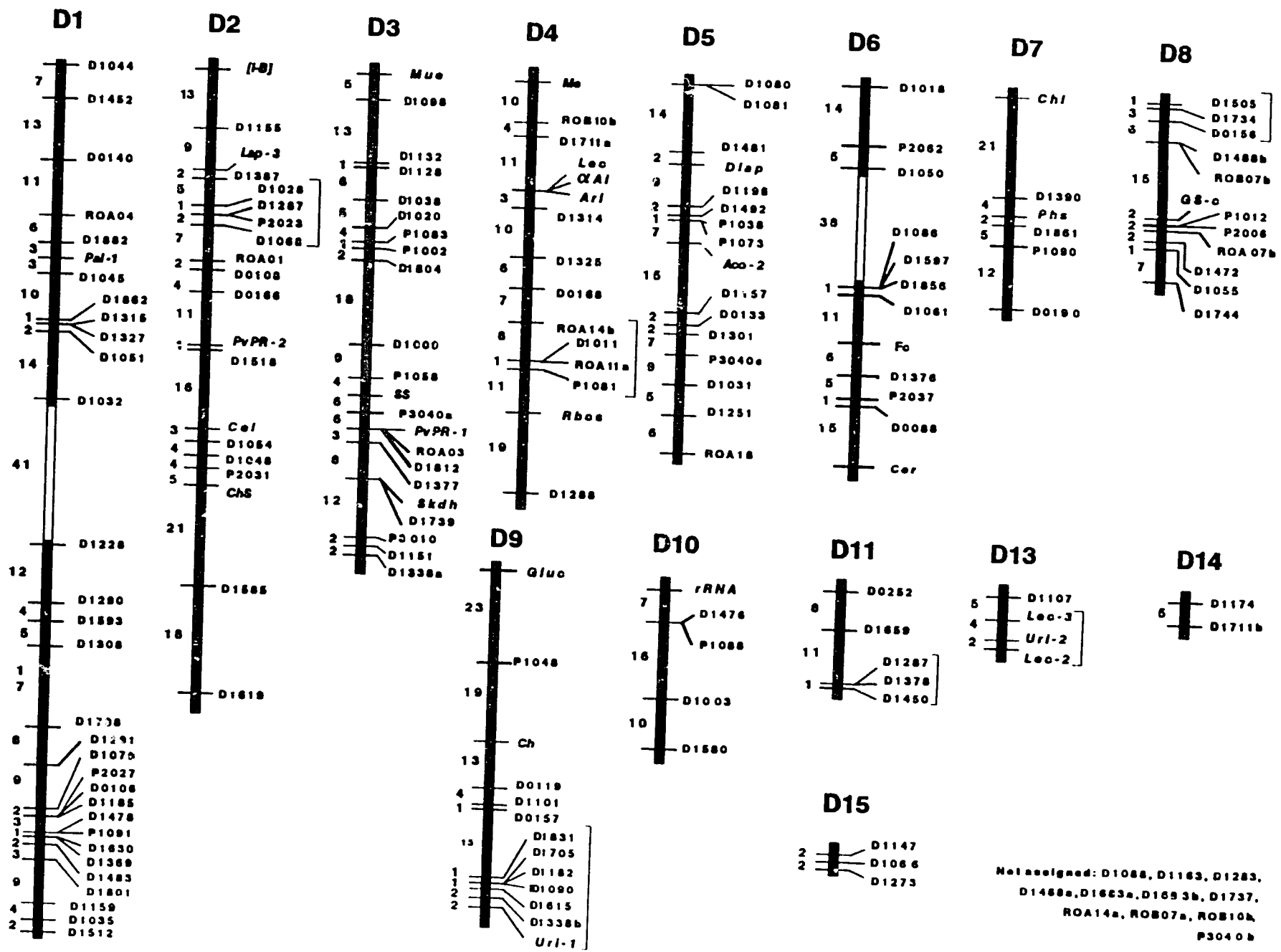

D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
D11
D13
D14
D15
Not assigned: D1088, D1163, D1283, D1468a, D1663a, D1693b, D1737, ROA14a, ROB07a, ROB10b, P3040b

Table 1. Common Bean Classical Map Gene List

Locus Symbol	Locus Name	Linkage Group	Phenotype	References
Ane	Anebulosus	III	Seedcoat color: Grayish brown mottling on buff background (e.g., cv. Contender)	27, 53
arg	argenteus	V	Greenish gray (silvery) pod color; with *y* gives white pod color	31
B	Bronze	III	Bronze-brown seedcoat color	19,26
C-R	Complex *C* locus	I	Several tightly linked genes or multiple alleles of the same locus influencing color and color pattern in flower, pod, and seed, including *R* (red seed), *Mf* (mancha na flor or flower blotch), *Mt* (seedcoat mottling), etc.	26,38,52-53,66
Cav	Caruncula verruca	VI	Wrinkling of the seedcoat starting from the caruncula	35
cl	circumlineatus	VIII	In partly colored seeds, colored areas are delineated with a strongly pigmented line	51
Cor	Corona	I	Colored ring outside the seed hilum ring	27
D	Hilum ring factor	I	Brown seed hilum ring	6,36,50
Da	Straight pod	II	In conjunction with *Db*, straight pod	25
dgs	dark green savoy	VI	Dark green, glossy, leathery and wrinkled leaves	42
dia	diamond	IX	Leaflets angular, slightly chlorotic, thick, and smaller	42
ds	dwarf seed	VI	Small seeds and short pods (xenia)	4
Ea	Flat pod	II	In conjunction with *Eb*, flat pod cross-section	25
Est-2	Esterase-2	X	Isozyme	70
Fb	Pod parchment	II	In conjunction with *Fa*, *Fc*, and *Fd*, fully parchmented pod walls	25
fin	finitis	IV	Determinate stem and branch growth	27
I	Inhibitor	III	Resistance to systemic infection by non-necrotic strains of Bean Common Mosaic Virus. Tight linkage or pleiotropy for resistance to Blackeye cowpea mosaic virus (*Bcm*), Blackeye cowpea mosaic virus (*Bcm*), Cowpea aphid-borne mosaic virus (*Cam*), Soybean mosaic virus (*Smv*), and Watermelon mosaic virus-2 (*Hsw*). Tight linkage to *B* .	1,24
J	Joker	XII	Seed shininess and postharvest seedcoat darkening; brown hilumring	30,50
miv	minor intervallis	VI	Short distance between successive funicula leading to seeds with flattened ends	33
No	Nopal	IV	Red (light salmon with brownish tinge) flower color	32

Locus Symbol	Locus Name	Linkage Group	Phenotype	References
P	Primary	X	Required for pigment production in all plant parts	30
Phs	Phaseolin	XII	Principal seed protein	5,8
prc	progressive chlorosis	IX	Chlorosis of fully expanded leaves; emerging and expanding leaves are green	2,42
rf	reclining foliage	VII	Downward slanting petiole	3
rk	red kidney	III	Red kidney seedcoat color	57
rnd	round	VI	Primary leaf and lateral leaflet tips rounded	42
Sal	Salmon	V	Salmon to geranium red flower color	32
sb	spindly branch	IX	Thin, drooping, short branches and petioles	2
sl	stipelless lanceolate	VI	Lanceolate leaflet form with absence of stipels from terminal leaflet	42
St	Stringless	III	Absence of "strings" (pod suture fibers)	49
Sur	Sursum versus	VI	Leaves and petals point downwards with pulvinus rotated 180°	29
T	Total	VIII	Colored flowers and self-colored seeds	27
te	tenuis	VI	Narrow and short pod	34
Urs	*Uromyces* resistance	XI	Nine linked loci (*Urs* to *Urs-9*) for resistance to *Uromyces phaseoli*	6,58
V	Violet	VII	Violet flower color and pale violet to black seed color	30
y	yellow	VI	Yellow (wax) pod	31

Table 2. Common Bean Molecular Map Gene List

Locus Symbol	Locus Name	Linkage Group	Phenotype or Comments	Reference
αAI	α-amylase inhibitor	D4	*αAI*, *Arl*, and *Lec* form a tightly linked small multigene family with around 80% nucleotide sequence similarity	41,48
Aco-2	Aconitase - 2	D5	Isozyme	64
Arl	Arcelin	D4	See *αAI*. The *Arl* locus consists of at least 4 structural genes.	20,47,48

Locus Symbol	Locus Name	Linkage Group	Phenotype or Comments	Reference
B	Bronze	D2	Bronze-brown seedcoat color (see classical map)	19,25
Cel	Cellulase	D2	RFLP to homologous common bean probe	62
Ch	Chitinase	D9	RFLP to homologous common bean probe	7
ChI	Chalcone Isomerase	D7	RFLP to homologous common bean probe	39
ChS	Chalcone Synthase	D2	RFLP to homologous common bean probe	54
Cor	Corona	D6	Colored ring outside the seed hilum ring (see classical marker map table)	27
Diap	Diaphorase	D5	Isozyme; consists of two tightly linked loci, *Diap-1* and *Diap-2*	58
Fc	Flower color	D6	Temporary symbol pending verification of allelism with *V* (see classical marker map table)	30,46
GS-c	Glutamine synthetase (cytosolic)	D8	RFLP to heterologous soybean probe	40
D0096-D0252			*Eco*RI-*Bam*HI random genomic probes developed at the University of California, Davis	45
D1009-D1862			*Pst*I random genomic probes developed at the University of California, Davis	45
Gluc	Glucanase	D9	RFLP to homologous common bean probe	13
I	Inhibitor	D2	Resistance to systemic infection by non-necrotic strains of Bean Common Mosaic Virus. Tight linkage or pleiotropy for resistance to Blackeye cowpea mosaic virus (*Bcm*), Blackeye cowpea mosaic virus (*Bcm*), Cowpea aphid-borne mosaic virus (*Cam*), Soyneam mosaic virus (*Smv*), and Watermelon mosaic virus-2 (*Hsw*). Tight linkage to *B* (see classical map)	1,24
Lap-3	Leucine aminopeptidase-3	D2	Isozyme	22
Lec	Seed lectin	D4	RFLP to homologous common bean probe. The *Lec* locus contains two structural genes, *dlec1* and *dlec2*, corresponding to erythro- and leucoagglutinating polypeptides ; co-segregation between RFLP and SDS-PAGE protein polymorphism at this locus. See also *αAI*.	8,18,46

Locus Symbol	Locus Name	Linkage Group	Phenotype or Comments	Reference
Lec-2 and *Lec-3*		D13	DNA sequence homology with *Lec*; however, no co-segregation with seed lectin protein polymorphism	46
Me	Malic enzyme	D4	Isozyme	22,68,69
Mue	Methylumbelliferyl esterase	D3	Isozyme	16
P1012 - P3040			*Mbo*I random genomic probes developed at the University of Paris	11,44
Pal	Phenylalanine ammonialyase	D1	RFLP to homologous common bean probe. Presumably *Pal-1*	10,14
Phs	Phaseolin	D7	RFLP for principal seed protein; small multigene family of 7-8 tightly linked genes. Co-segregation between RFLP and SDS-PAGE protein polymorphism	5,8,46,56,60
PvPR-1	*Phaseolus vulgaris* pathogenesis-related protein 2	D3	RFLP to homologous common bean probe; belongs to the same multigene family of about 12 genes as *PvPR-2*	67
PvPR-2	*Phaseolus vulgaris* pathogenesis-related protein 2	D2	RFLP to homologous common bean probe. See *PvPR-1*.	67
Rbcs	Ribulose bisphosphate carboxylase (small subunit)	D4	Isozyme	68,69
rRNA	Ribosomal RNA	D10	RFLP to heterologous pea probe for one of several rRNA loci identified by *in situ* hybridization	12,21,46
ROA01 - ROB10			RAPD markers (Operon primers)	44,46
Skdh	Shikimate dehydrogenase	D3	Isozyme	68,69
SS	Sucrose synthetase	D3	RFLP to heterologous soybean probe	15,61

Locus Symbol	Locus Name	Linkage Group	Phenotype or Comments	Reference
Uri and *Uri-2*	Uricase	D9 and D13	RFLP to heterologous soybean probe	43

Table 3. Additional linkages and correspondance between classical and molecular linkage maps

Traits (linkage distance in cM unless otherwise noted)	Molecular Linkage Group if known	Source
Est-2 - 12% - *P*	D7	70
Adh-1 - 2 - *Aat-2*		64
Est-2 - 11 - *Phs* - 8- *Spe/Spg* - 24 - *P* - 9 - *Spa/Spb* - 16 - *Spba* - 2[illegible] - *Mdh-1*	D7	65
Spd/Spf/Sph - 5 - *Spca*		65
fin - 8% - early flowering	D1	9,23

REFERENCES CITED

1. Ali, M. (1950) Phytopathology 40: 69-79
2. Awuma, K., Bassett, M.J. (1988) J. Amer. Soc. Hort. Sci. 113: 464-467
3. Bassett, M.J. (1976) HortScience 11: 238-240
4. Bassett, M.J. (1982) J. Amer. Soc. Hort. Sci. 107: 1058-1061
5. Bassett, M.J. (1989) Annu. Rept. Bean Improv. Coop. 32: 1-15
6. Bassett, M.J. (1991) HortScience 26: 834-836
7. Broglie, K.E., Gaynor, J.J., Broglie, R.M. (1986) Proc. Natl. Acad. Sci. USA 83: 6820-6824
8. Brown, J.W.S., Bliss, F.A., Hall, T.C. (1981) Theor. Appl. Genet. 60: 251-259
9. Coyne, D.P., Schuster, M.L. (1974) Euphytica 23: 195-204
10. Cramer, C.L., Edwards, K., Dron, M., Liang, X., Dildine, S.L., Bolwell, G.P., Dixon, R.A., Lamb, C.J., Schuch, W. (1989) Plant Molec. Biol. 12: 367-383
11. Dron, M., Grisvard, J., Adam, A.-F., pers. comm
12. Durante, M., Cionini, P.G., Avanzi, S., Cremonini, R., D'Amato, F. (1977) Chromosoma 60: 269-282
13. Edington, B.V., Lamb, C.J., Dixon, R.A. (1991) Plant Molec. Biol. 16: 81-94
14. Edwards, K.E., Cramer, C.L., Bolwell, G.P., Dixon, R.A., Schuch, W., Lamb, C.J. (1985) Proc. Natl. Aca Sci. USA 82: 6731-6735
15. Fuller, F., Kunstner, P.W., Nguyen, T., Verma, D.P.S. (1983) Proc. Natl. Acad. Sci. USA 80: 2594-2598
16. Garrido, B., Nodari, R., Debouck, D.G., Gepts, P. (1991) J. Hered. 82: 181-183
17. Gepts, P., Llaca, V., Nodari, R.O., Panella, L. (1992) *In* Linskens, H.-F., Jackson, J.F. (eds.), Modern methods of plant analysis (New Series): Seed analysis. Springer, Berlin: p. in press
18. Hoffman, L.M., Donaldson, D.D. (1985) EMBO J. 4: 883-889
19. Johannsen, W. (1909) Elementen der exakten Erblichkeitslehre. Fisher, Jena, Germany
20. John, M.E., Long, C.M. (1990) Gene 86: 171-176
21. Jorgensen, R.A., Cuellar, R.E., Thompson, W.P., Kavanagh, T.A. (1987) Plant Molec. Biol. 8: 3-12
22. Koenig, R., Gepts, P. (1989) J. Hered. 80: 455-459
23. Koinange, E., Gepts, P., unpubl. results
24. Kyle, M.M., Dickson, M.H. (1988) J. Hered. 79: 308-311
25. Lamprecht, H. (1932) Hereditas 16: 295-340
26. Lamprecht, H. (1932) Hereditas 16: 169-211
27. Lamprecht, H. (1934) Hereditas 19: 163-176
28. Lamprecht, H. (1935) Hereditas 20: 71-93
29. Lamprecht, H. (1937) Bot. Notiser 1937: 341-354
30. Lamprecht, H. (1939) Hereditas 25: 255-288
31. Lamprecht, H. (1947) Agri Hort. Genet. 5: 72-84
32. Lamprecht, H. (1948) Agri Hort. Genet. 6: 64-81
33. Lamprecht, H. (1952) Agri Hort. Genet. 10: 105-112
34. Lamprecht, H. (1952) Agri Hort. Genet. 10: 141-151
35. Lamprecht, H. (1955) Agri Hort. Genet. 13: 143-153
36. Lamprecht, H. (1961) Agri Hort. Genet. 19: 319-332
37. Lamprecht, H. (1964) Agri Hort. Genet. 22: 256-271
38. Leakey, C.L.A. (1988) *In* Gepts, P. (ed.), Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, the Netherlands: pp. 245-327
39. Mehdy, M.C., Lamb, C.J. (1987) EMBO J. 6: 1527-1533
40. Miao, G.-H., Hirel, B., Marsolier, M.C., Ridge, R.W., Verma, D.P.S. (1991) Plant Cell 3: 11-22
41. Moreno, J., Chrispeels, M.J. (1989) Proc. Natl. Acad. Sci. USA 86: 7885-7889
42. Nagata, R.T., Bassett, M.J. (1984) J. Amer. Soc. Hort. Sci. 109: 513-516
43. Nguyen, T., Zelechowska, M., Foster, V., Bergmann, H., Verma, D.P.S. (1985) Proc. Natl. Acad. Sci. USA 82: 5040-5044
44. Nodari, R., Tsai, S.-M., Gepts, P., unpubl. results
45. Nodari, R.O., Koinange, E.M.K., Kelly, J.D., Gepts, P. (1992) Theor. Appl. Genet. : in press
46. Nodari, R.O., Tsai, S.M., Gilbertson, R.L., Gepts, P. (1992) Theor. Appl. Genet. : in press
47. Osborn, T.C., Blake, T., Gepts, P., Bliss, F.A. (1986) Theor. Appl. Genet. 71: 847-855
48. Osborn, T.C., Alexander, D.C., Sun, S.S.M., Cardona, C., Bliss, F.A. (1988) Science 240: 207-210
49. Prakken, R. (1934) Genetica 16:: 177-294

50. Prakken, R. (1970) Meded. Landbouwhogesch. Wageningen 70-23: 1-38
51. Prakken, R. (1972) Meded. Landbouwhogesch. Wageningen 72-29: 1-82
52. Prakken, R. (1974) Meded. Landbouwhogesch. Wageningen 74-24: 1-36
53. Prakken, R. (1977) Annu. Rept. Bean Improv. Coop. 20: 32-35
54. Ryder, T.B., Cremer, C.L., Bell, J.N., Robbins, M.P., Dixon, R.A., Lamb, C.J. (1984) Proc. Natl. Acad. Sci. USA 81: 5724-5728
55. Ryder, T.B., Hedrick, S.A., Bell, J.N., Liang, X., Clouse, S.D., Lamb, C.J. (1987) Mol. Gen. Genet. 210: 219-233
56. Slightom, J.L., Sun, S.M., Hall, T.C. (1983) Proc. Natl. Acad. Sci. USA 80: 1897-1901
57. Smith, F.L. (1939) Hilgardia 12: 553-621
58. Sprecher, S.L. (1988) Ph.D. thesis, Michigan State University, East Lansing, MI
59. Stavely, J.R. (1984) Phytopathology 74: 339-344
60. Talbot, D.R., Adang, M.J., Slightom, J.L., Hall, T.C. (1984) Mol. Gen. Genet. 198: 42-49
61. Thummler, F., Verma, D.P.S. (1987) J. Biol. Chem. 262: 14730-14736
62. Tucker, M.L., Milligan, S.B. (1991) Plant Physiol. 95: 928-933
63. Vallejos, C.E., Chase, C.D. (1991) Theor. Appl. Genet. 82: 353-357
64. Vallejos, C.E., Chase, C.D. (1991) Theor. Appl. Genet. 81: 413-419
65. Vallejos, E.C., Sakiyama, N.S., Chase, C.D. (1992) Genetics 131: 733-740
66. Vieira, C., Shands, H. (1969) Crop Sci. 5: 371
67. Walter, M.H., Liu, J., Grand, C., Lamb, C.J., Hess, D. (1990) Mol. Gen. Genet. 222: 353-360
68. Weeden, N.F. (1984) Annu. Rept. Bean Improv. Coop. 27: 123-124
69. Weeden, N.F. (1984) Euphytica 33: 199-208
70. Weeden, N.F., Liang, C.Y. (1985) Annu. Rept. Bean Improv. Coop. 28: 87-88

The Use of Molecular and Biochemical Markers in Crop Evolution Studies

Paul Gepts

Department of Agronomy and Range Science,
University of California, Davis, CA 95616-8515

[... No doubt man selects varying individuals, sows their seeds, and again selects their varying offspring ... Man therefore may be said to have been trying an experiment on a gigantic scale; and it is an experiment which nature during the long lapse of time has incessantly tried ...]

C. DARWIN (1868), The Variation of Animals and Plants under Domestication.

TABLE OF CONTENTS

I. INTRODUCTION 4

II. METHODOLOGICAL ASPECTS 6

III. PATTERNS OF GENETIC DIVERSITY AND DOMESTICATION 11
A. *Phaseolus* 11
1. Identification of the wild ancestor 11
2. Divergence in the cultivated gene pool 14
3. Gene flow between wild and cultivated forms 15
B. Other crops 17
1. Identification of the wild ancestor 17
Avocado 17
Barley 17
Brassica spp. 17
Cotton 18
Lentil 19
Lettuce 19
Maize 19
Pea 20
Peanut 20
Pearl millet 20
Peppers 20
Potato 21
Rice 22
Soybean 22
Tomato 22
Wheat 23
Wild rice 24
2. Divergence in the cultivated gene pool 24
Brassica spp. 24
Maize 25
Potato 25
Rice 26
3. Gene flow between wild and cultivated forms 26
Brassica spp. 26
Maize 27
Pearl millet 27
Rice 27
Tomato 27

IV. DIFFERENTIATION BETWEEN WILD PROGENITOR AND

CULTIVATED DESCENDANT 29
A. Fate of genetic diversity during and after domestication 29
1. *Phaseolus* spp. 29
2. Other crops 31
B. Genetics of morphological and physiological differentiation between wild ancestor and cultivated descendant 32

V. DISCUSSION 36

VI. CONCLUSIONS 42

VII. REFERENCES 44

I. INTRODUCTION

When de Candolle (1882) initiated the study of crop evolution, his major concern was to identify the geographic origin, *i.e.* the domestication center, of individual crops. In his opinion, the following four types of data would shed light on this question: 1) archeology; 2) botany; 3) philology or linguistics; and 4) history. Of these four types of data, the first two are the most reliable (Harlan and de Wet 1973). Archaeological remains attested to the antiquity of the cultivation of a crop in a region compared to other regions devoid of archeological remains. Botanical arguments referred to the existence in a defined region of a wild-growing form that was sufficiently similar morphologically to the crop that it could represent its ancestral form (or at least the immediate descendant of the ancestral form). The existence of words designating a particular crop, particularly in native languages, was considered a testimony to the relative antiquity of cultivation of the crop. Finally, historical description such as the treatises of Dioscorides and Theophrastus, the Herbals of the sixteenth and seventeenth centuries, or descriptions of the New World short after the Spanish Conquista by, for example, Cieza de León (between 1541 and 1550) or Acosta (1590), also would provide some evidence towards identifying the area of origin of crops.

While these types of evidence are still relied upon in the study of crop evolution, the range of available techniques of analysis has expanded since de Candolle's time. Consequently, crop evolutionists have been able not only to ask a broader range of questions but also in certain cases to obtain more specific answers to their questions. In this review, I will focus on the use of molecular and biochemical markers to strengthen the botanical argument in the study of crop evolution. Molecular markers include any markers reflecting directly changes at the DNA sequence level, principally restriction fragment length polymorphisms (RFLPs), but also random amplified polymorphic DNA (RAPDs), minisatellite markers, etc. Because crop evolution studies require a sufficient sampling of the diversity contained in the species, direct DNA sequencing has rarely been used, if at all, because of its inherent cumbersomeness. Biochemical markers have included principally isozymes and seed proteins. The major advantage of these markers is their evolutionary neutrality which allows us to distinguish those similarities that are due

to common ancestry from similarities due to convergence. Over the last years, an increasing body of data on crop evolution has been accumulated based on molecular and biochemical markers (reviewed in Doebley 1989, 1992; Gepts 1990; Clegg 1990).

This review will be divided in three parts. The first part will address some methodological aspects, including the need for adequate sampling and the relative merits of various classes of molecular or biochemical markers in crop evolution studies. The second part will deal primarily with the traditional endeavor of crop evolution studies, i.e. the identification of the wild ancestor of crop plants and related questions such as the organization of genetic diversity in crop species and possible introgression between wild and cultivated forms of crop species. The second part will be devoted to examining differences between wild and cultivated forms at the genetical level, an area that is perhaps more recent because the identity of the wild ancestor should be known at the onset of these studies. Throughout this review, I will use the information that has been accumulated on the evolution of *Phaseolus* cultivated species as a starting point because this genus has proven to be an interesting experimental material for crop evolution studies; I then compare these results with those obtained for other crop species. Because of the extensive literature in this area, I will not attempt to do an exhaustive review but rather I will emphasize some of the more recent results. Additional references are available in the reviews mentioned above.

II. METHODOLOGICAL ASPECTS

Crop evolution studies in a given species require an adequate sampling according to two criteria. A first criterion is a geographic one. Both wild and cultivated genotypes from the entire area of origin should be adequately represented in the initial sample in order to increase the confidence level of the study. Genotypes of the presumed wild progenitor provide an essential geographic framework of reference that allows us to trace certain variants observed among the cultivated genotypes to a specific geographic area. Unfortunately, most of the emphasis in germplasm conservation was concentrated until recently on the cultivated materials. Hence, with few exceptions, it was difficult to assemble a representative sample of wild progenitor accessions. This emphasis may be shifting now in part because of studies involving molecular and biochemical markers that show higher levels of diversity among the wild progenitors compared to the cultivated descendant (see below).

Among cultivated materials, preference should be given to so-called landraces. Landraces are defined here as cultivated genotypes that have not resulted from scientific breeding programs. Hybridizations in these programs will tend to obscure associations of traits observed in either the wild ancestor or cultivated races descended from the ancestor. In addition, these programs may also introduce successful varieties from other countries. Until recently, most of the cultivated genotypes found in the domestication areas were landraces because developing countries in the tropical and subtropical areas where most crops were domesticated, did not have the resources to establish strong breeding programs. In addition, these new breeding programs have often emphasized for practical reasons a limited number of the commercially most desired types. Hence, most of the cultivated materials collected until recently in these countries are landraces and, in general, it should not be difficult, at least for the major crops, to constitute an adequate sample of cultivated genotypes.

This geographic sampling can also be an iterative process as illustrated by our data on allozyme diversity in common bean. As explained in more detail below, allozyme frequency data indicated that the transition between the Middle American and Andean gene pools in this species was located somewhere between northeastern

Colombia and southern Peru (Koenig and Gepts 1989b). With the exception of a few wild common bean accessions collected by D.G. Debouck in northern Peru, no other wild materials were available from this region, particularly in southern Colombia and Ecuador. On the basis of these results, germplasm explorations were organized in 1989 and 1990, which found several wild *P. vulgaris* in Ecuador where this species had previously not been reported. Allozyme and phaseolin seed protein data of these newly collected accessions confirmed that they represent a group that is distinct, not only at the biochemical level, but also ecologically from the Middle American and Andean gene pools (O.M. Paredes, D.G. Debouck, and P. Gepts, unpubl. results). It is striking that at this stage we are still defining the boundaries of the natural distribution of the wild progenitor of a crop that is a mainstay in the nutrition of Latin American countries. In the 1980s alone, previously undescribed wild bean populations have been identified in northern Mexico (Chihuahua: Nabhan 1985), eastern Mexico (Querétaro) and Costa Rica (Gepts and Debouck 1991), and Panama (Brücher 1988) in addition to the populations of Ecuador and northern Peru mentioned above. This observation has some bearing upon other crops, especially those for which the progenitor has not yet been identified. While extinction of the ancestor is a possible explanation, absence in germplasm collections is an equally plausible one in light of our experience in *Phaseolus*.

The second sampling criterion is the genome. Plants harbor one nuclear and two cytoplasmic genomes (in the chloroplasts and the mitochondria). Each of these can provide molecular or biochemical markers for crop evolution studies although the type of information provided by each of them may not be the same. Cytoplasmic genome markers may shed light on the origin of one of the parents especially in polyploid plants for those taxa where cytoplasmic inheritance is uniparental, whether maternal or paternal (Soltis et al. 1992a). There are differences, however, between mitochondrial and chloroplast DNA markers. The peculiar evolutionary dynamics of plant mtDNA, which is characterized by high levels of rearrangements, low rates of point mutations, and the presence of foreign sequences, makes it difficult to establish phylogenies based on mtDNA sequence variation although by the same token the rapid changes can make this genome useful to carry out rapid surveys of genome types (Palmer 1992). The size and

gene order of chloroplast DNA, on the other hand, is highly conserved but its nucleotide substitution rate is higher than in mtDNA (Palmer 1990). Intraspecific diversity for cpDNA has been observed and used in phylogenetic and crop evolution studies (Soltis et al. 1992b).

Nuclear markers, whether single-copy or repeated sequences, will provide information on both progenitor genotypes because of the biparental inheritance of the nuclear genome. Depending on the particular sequence, a wide range of diversity exists in regard to level and molecular basis of the polymorphisms (Clegg 1990). Whenever possible, it is preferable to obtain information about linkage relationships of the markers. This information will help ascertain that different regions of the nuclear genome are being characterized and may help interpret, for example, cases of introgression or discrepancies in trends affecting genetic diversity when measured by different markers.

A second methodological consideration is the choice of molecular markers. A survey of the literature reveals that isozymes, seed proteins, RFLPs (for nuclear and cytoplasmic sequences), minisatellite sequences, and RAPDs have been used in crop evolution studies. There are several attributes by which one can assess the potential usefulness of a particular category of markers. These include level of polymorphism, environmental stability, the number of loci, molecular basis of the polymorphism, and the ease and cost of analysis (Table 1). Depending on the objective of the study, a certain level of polymorphism is required: at higher taxonomic level (species or above), more conserved markers are needed, whereas at the population level more variable markers are desirable. Electrophoretic patterns should be free of environmental influence to confirm that observed differences are genotypic differences. The number of loci should be as high as possible to ensure adequate genome coverage. The molecular basis of the polymorphism should be complex, thus decreasing the probability of homoplasy. Finally, the methodology employed should be as cheap and simple as possible to permit the analysis of samples of adequate size.

The main advantages of isozymes include the simplicity and low cost of the analysis and a reasonable genome coverage (10-20 loci per species: Hamrick and Godt 1990). In addition, standardized experimental conditions allow us to detect genotypic

differences although isozymes are phenotypic markers. Disadvantages include a generally low level of polymorphism with only a few alleles per locus (Hamrick and Godt 1990) and a simple molecular basis: some nucleotide changes lead to charged amino acid substitutions and, hence, changes in electrophoretic migration.

Advantages of seed proteins as electrophoretic markers include their high level of polymorphism, their high level of environmental stability (although a few exceptions have been reported: e.g., Gayler and Sykes 1985), and the complex molecular basis of the electrophoretic patterns that includes nucleotide substitutions, insertions/deletions, and co- and posttranslational modifications. This complexity at the molecular level, however, makes it difficult to relate phenotypic changes in electrophoretic banding pattern with changes at the molecular level. Hence, one is usually limited to phenetic analyses with this category of markers. A disadvantage is the low number of loci involved (usually less than 10) (reviewed in Gepts 1989).

RFLP markers can display a wide range of levels of polymorphism depending on the species (Table 2; Nodari et al. 1992), the genome (e.g., cytoplasmic vs. nuclear: Curtis and Clegg 1984; Palmer 1987; Zurawski and Clegg 1987; Wolfe et al. 1987), or the particular sequence. In general, however, RFLPs are be more polymorphic than isozymes. For example, in a direct comparison between the two classes of markers in the same set of maize genotypes, Messmer et al. (1991) found that RFLPs were more polymorphic than isozymes both in terms of the number of polymorphic loci (94 vs. 68 %) and the average number of variants per polymorphic locus (3.4 vs. 2.5). Additional advantages of RFLPs include their better genome coverage and environmental stability. The molecular basis of RFLPs can be quite simple as a single nucleotide change can lead to a restriction site loss, or less likely a site gain. The likelihood of a repeat mutation involving a nucleotide change may be higher than that involving a rearrangement characterized by its size as well as location. Through restriction mapping, it is possible to identify the molecular basis of the polymorphism although this limits the number of sequences that can be analyzed (Gepts and Clegg 1989). Finally, the RFLP technology is cumbersome and costly, which effectively limits the sample size.

Minisatellite markers, principally those revealed through cross-hybridization with

human mini-satellites or M13 (e.g., Dallas 1988; Rogstad et al.1989; Stockton et al. 1992), reveal very high levels of polymorphism. The fingerprinting pattern is not influenced by environmental conditions (G. Sonnante, T. Stockton, and P. Gepts, unpubl. results). Little is known about the actual molecular basis of hypervariable sequences in plants. The complexity of the fingerprinting pattern and actual mapping data (T. Stockton, R. Nodari, and P. Gepts, unpubl. results) suggest, however, that several loci are involved although the genome coverage may not be as extensive as that of RFLPs or RAPDs. The technology is similar to RFLP technology with the exception that several loci can be sampled at once.

The advantages of RAPD markers are that they can be more polymorphic than RFLPs (Williams et al. 1990), they offer genome coverage equivalent to that of RFLPs, and their methodology can be quite simple provided a rigorously standardized methodology with the necessary controls is adhered to. Disadvantages include limited information about the molecular basis of RAPDs and the environmental stability of the polymorphism.

In summary, each class of molecular or biochemical markers possesses advantages and disadvantages. Depending on the goals of the study, one or the other or a combination of markers can be used.

III. PATTERNS OF GENETIC DIVERSITY AND DOMESTICATION

A. *Phaseolus*

The genus *Phaseolus* consists of some 55 species (Debouck 1991), all of which originated on the American continent (Maréchal et al. 1978; Delgado Salinas 1985). Four *Phaseolus* species contain cultivated forms: *P. vulgaris* (common bean), *P. lunatus* (lima bean), *P. coccineus* (runner bean and year bean), and *P. acutifolius* (tepary bean). Depending on the taxonomic treatment, the year bean is either included in *P. coccineus* as a subspecies (*P. coccineus* subsp. *darwinianus*: Hernandez Xolocotzi et al. 1959; or *P. coccineus* subsp. *polyanthus*: Maréchal et al. 1978) or considered a separate species (*P. polyanthus*: Schmit and Debouck 1991). Each of these species has a characteristic life history and reproductive system. Common bean is an annual species with predominantly self-pollinated reproductive system (< 5% outcrossing) although higher levels of outcrossing have been reported by Wells et al. (1988). Lima bean is similar to common bean but with a longer life cycle and higher levels of outcrossing (Webster et al. 1979; Lyman et al. 1985; Baudoin 1988). Runner bean is a perennial, predominantly allogamous species pollinated by insects (honey bees and bumblebees) (Webster et al. 1980; Delgado Salinas 1988). Finally, tepary bean is a short-lived annual with a markedly autogamous (cleistogamous) reproductive system (Pratt and Nabhan 1988). Of these four species, *P. vulgaris* is by far the most important because it is grown on a worldwide scale and, hence, has been the subject of most studies. The other three species can be important locally either as a subsistence crop (all three species) or as a commercial crop (e.g., lima bean in California); some more specific studies have been conducted on these species.

1. Identification of the wild ancestor

The use of molecular markers has confirmed the identity of the wild ancestor of cultivated *Phaseolus* species previously determined on the basis of morphological similarities and crossability studies. In addition, molecular markers have provided additional information on the pattern of domestication, specifically whether the cultivars in the various species resulted from a single or multiple domestications.

The considerable distribution of wild *P. vulgaris*, which extends from northern

Mexico to Argentina (reviewed in Delgado Salinas 1988; Brücher 1988), the distinctness of seed remains among the various archeological sites in Latin America, and the differential geographic distribution between large-seeded and small-seeded cultivars (see below), led Kaplan (1965), Harlan (1971), Heiser (1965), and Evans (1976) to suggest that common bean had two or more domestication centers.

Various biochemical and molecular markers (phaseolin seed protein, isozymes, RFLPs for nuclear and mitochondrial DNA, and minisatellite sequences), correlated with morphological markers and F_1 hybrid lethality, have now provided solid evidence for this initial suggestion (Table 2). Analyses with molecular markers reveal a divergence between Middle American and Andean wild-growing bean populations as illustrated by Fig. 1, which shows the result of a principal component analysis of banding pattern frequencies for RFLPs of nuclear single copy sequences. Similar results have been obtained with phaseolin seed protein (Gepts et al. 1986; Gepts and Bliss 1986; Koenig et al. 1990), isozymes (Koenig and Gepts 1989b), and sequences hybridizing with M13 or human minisatellite sequences (G. Sonnante, T. Stockton, and P. Gepts, unpubl. results). Cultivars of Middle American and Andean origin display the same divergence and show the highest similarity with the wild beans of their respective region as illustrated by RFLP data (Fig. 1) or data from other molecular or biochemical markers (Table 2). This divergence between Middle American and Andean genotypes is responsible in part for the high levels of RFLP present in *P. vulgaris* compared to other species, especially autogamous species (Table 3).

This divergence at the molecular level is correlated with divergence for morphological traits. Multivariate analyses, such as principal component and canonical discriminant analyses, of traits including seed size and shape, leaf size and shape, internode length, reveal that Andean cultivars, as determined on the basis of molecular markers, have larger leaves, longer internodes, and larger seeds (Singh et al. 1991a).

Prior to analyses with molecular markers, various cases of F_1 hybrid lethality or weakness had been described (e.g., Shii et al. 1980; Singh and Gutiérrez 1984). Markers allowed us to determine that, so far without exceptions, each of these case arose in

crosses between a Middle American and an Andean cultivar (Gepts and Bliss 1985). This F_1 lethality is controlled by two complementary genes (*Dl-1* and *Dl-2*: Shii et al.1980) present not only in cultivated beans but also in wild beans (Koinange and Gepts 1992). The simple genetic control of this reproductive isolation mechanism and the overall divergence at the molecular, biochemical, and morphological levels between the Middle American and Andean genotypes suggests that *P. vulgaris* could have entered a speciation process leading to two new species if the *Dl* genes were to become fixed over an evolutionary time scale in their respective gene pools or "subspecies". Additional forms of reproductive isolation include chlorophyll variegation (e.g., Coyne 1966) and segregation distortion (Koenig and Gepts 1989a).

The domestication scenario suggested by these results is the following. The wild progenitor appears to have diverged prior to domestication, which took place some 7,000 to 10,000 years ago into two entities showing incipient reproductive isolation and often contrasting alleles at molecular and biochemical marker loci. Independent domestications in these two regions has led to two groups of cultivars whose divergence mirrors that of their respective wild ancestors. While this multiple domestication scenario appears to be the most likely possibility, other scenarios involving a single domestication in one area followed by migration and escape from cultivation in the other area, or involving gene flow between wild and cultivated beans cannot entirely be excluded.

Recently, we have identified a group of wild *P. vulgaris* that appears to bridge the gap, both geographically and at the molecular level, between the Middle American and Andean gene pools. Isozyme analyses had shown that the wild beans from the Middle American gene pool extend from northern Mexico to northeastern Colombia and wild beans from the Andean gene pool are distributed from southern Peru to Argentina. It was only in the mid 1980s that wild *P. vulgaris* in northern Peru were first described and in 1989-1990 in Ecuador (Debouck et al. 1989a; Gepts and Debouck 1991). Wild beans in northern Peru and Ecuador are distributed on the western slope of the Andes in Ecuador and northern Peru whereas wild beans elsewhere in the Andes are distributed on the eastern slope of the Andes. Isozyme analyses show that this group carries a combination of Middle American and Andean isozyme alleles (O.M. Paredes and P.

97

Gepts, unpubl. results). It has apparently not been involved in bean domestication because its phaseolin type ('I') is not represented among bean cultivars (Koenig et al. 1990).

Phaseolin studies in other cultivated *Phaseolus* species have revealed that tepary bean resulted probably from a single domestication although it was not possible to determine where this occurred in the distribution area of wild *P. acutifolius*, which extends from the southwestern U.S. to Central America (Schinkel and Gepts 1988). In *P. coccineus* subsp. *polyanthus*, the highest diversity was observed among wild accessions, all originating in Guatemala. Cultivars displayed a divergence between the northern half of their distribution (predominance of the 'b' phaseolin type) and the southern half (predominance of the 'k' phaseolin type) (Schmit and Debouck 1991). It should be noted that the 'k' phaseolin type has yet to be discovered among wild accessions. In lima bean, phaseolin studies showed that the wild progenitor is also diverged in two major groups as in common bean (Debouck et al. 1989b). The two groups, however, have a different geographic distribution. One group, which gave rise to the small-seeded cultivars found mostly in Mexico, Central America, the Caribbean, and Brazil) (the so-called Sieva and Potato types of Mackie 1943, is distributed from Mexico to Argentina. In South America, in particular, it is found on the Eastern slope of the Andes. The other group, which gave rise to the large seeded group (the so-called Big Lima group of Mackie 1943, has a much more restricted distribution on the western slope of the Andes in Ecuador and Peru (Debouck et al. 1989b; D.G. Debouck, A. Maquet, and P. Gepts, unpubl. results).

2. <u>Divergence in the cultivated gene pool</u>

Information derived from phaseolin and isozyme analyses has provided evidence for further divergence within the Middle American and Andean gene pools. When phaseolin or isozymes are used as a priori classification criteria, multivariate analyses such as canonical discriminant analysis reveal correlations with phenotypic traits such as growth habit, internode length, leaf and seed size, phenology, disease resistances, and general ecological adaptations (Singh et al. 1991a, b). These data have led to a proposal for 6 races or groups of related cultivars, three in each major gene pool (Singh et al.

1991c). In the Middle American gene pool, race Jalisco represents the predominantly climbing cultivars of the southern, humid highlands of Mexico and Central America. Race Durango includes cultivars with a prostrate growth habit from the northern, arid highlands of Mexico and race Mesoamerica includes bush indeterminate cultivars from the humid, hotter lowlands of Mexico, Central America, and South America.

In the Andean gene pool, race Nueva Granada includes cultivars with determinate bush or indeterminate climbing growth habits adapted to moderate altitudes. Race Peru consists of cultivars with a climbing growth habit adapted to higher altitudes, whereas race Chile includes cultivars with a prostrate growth habit. In it is interesting to note that races Durango and Chile display a similar phenotype, which includes medium-sized seeds and light pigmentation, in addition to the prostrate growth habit. These two races are diverged, however, at the molecular and biochemical levels (Gepts et al. 1986; Singh et al. 1991b; O.M. Paredes, V. Becerra and P. Gepts, unpubl. results). Their similar phenotype may have resulted from convergence due to selection for adaptation to arid environments prevalent in northern Mexico and Chile.

3. Gene flow between wild and cultivated forms

In spite of their morphological distinctness, the wild progenitor and cultivated descendant in *P. vulgaris* and in other cultivated *Phaseolus* species belong to the same biological species. The two forms can be crossed and the progeny is viable and fertile (Burkart and Brücher 1953; Miranda Colín 1979; Evans 1980; Koenig and Gepts 1989a). Although common bean is generally considered to be a predominantly self-pollinated species, cross-pollination of various intensities has been reported (Bliss 1980; Wells et al. 1988). In addition, it may be that in the areas of origin there is a higher frequency of cross-pollinating insects than in the other bean-growing areas where most of the estimates of cross-pollination have been established. During dissemination of beans from their domestication area to other regions, these insects would likely not have been dispersed. The absence of reproductive isolation as well as the sympatry of wild and cultivated forms, favor, therefore, the existence of gene flow between them.

On the other hand, the natural and cultivated growing environments are quite

distinct and hybrid forms may be adapted to neither of these environments. In addition, humans display a marked preference for certain qualities of the harvested product, for example, seed color or cooking time that can be lost in the hybrids. Selection in these contrasting environments will then lead to a return to the respective parental phenotype. Although the genetic control of the differences conferring adaptation to the two environments and the linkage relationships among the genes involved are only imperfectly known (see below), it is likely that several genes are implicated. Linkage drag around these genes during the selection process subsequent to hybridization may further reduce the possibility of recombination between wild and cultivated forms. Hence, disruptive selection and linkage drag may reduce the apparent frequency of gene flow.

The detection of gene flow poses a methodological problem. Gene flow between wild and cultivated forms will result in parallel geographic patterns of variation between the two forms. However, other processes such as convergent evolution and common ancestry can also lead to such patterns. Traditionally, gene flow has been inferred from morphological traits. Because of their phenotypic nature, however, these traits can be subject to selection and convergence. The selective neutrality of molecular and biochemical markers allows us to distinguish similarity through convergence from the other two possibilities. In our experience, gene flow can then be inferred -- with caution -- from the existence of rare molecular or biochemical variants with a limited geographic distribution as illustrated below.

In *P. vulgaris*, most wild and cultivated genotypes display the *Mdh-2*^{S} allele. Exceptions include a limited number of wild-growing accessions and race Jalisco cultivars from the southern highlands of Mexico, which show the *Mdh-2*^{F} allele. The rarity and the narrow geographic localization argue in favor of gene flow although it is not possible to determine the direction of the gene flow (Singh et al. 1991b). It is interesting to note, however, that the phenotype of the cultivars involved does not display any hint of past hybridization with wild beans (S. Singh, pers. comm.). Phaseolin data also provide evidence for occasional outcrosses between wild and cultivated beans. All cultivated accessions from Mexico analyzed so far show the 'S' phaseolin type with the exception of

one accession with an 'M' phaseolin type characteristic of wild-growing accessions of the same region. The latter accession also displays morphological signs of hybridization with wild beans such as smaller seeds and the striping and spotting pattern characteristic of wild bean seeds (Koenig et al. 1990). Some wild accessions from Colombia, Ecuador, and northern Peru also show signs of introgression from cultivars based on phaseolin data (Gepts and Bliss 1986; Gepts et al. 1986; O.M. Paredes and P. Gepts, unpublished results).

In summary, evidence for gene flow between wild and cultivated *P. vulgaris* has been identified although it is difficult to quantify the magnitude of this process because of our reliance on markers that are rare in the wild or cultivated gene pools.

B. Other crops

1. Identification of the wild ancestor

Avocado: The use of molecular markers has helped clarify the relationships between the three varieties of *Persea americana* currently cultivated (var. *americana*, var. *drymifolia*, and var. *guatemalensis*) and other *Persea* species. Restriction site analyses of cpDNA, rDNA, and cellulase confirmed the distinction between the three varieties made on morphological grounds (Furnier et al. 1990). *P. americana* var. *guatemalensis* appeared to be a hybrid between *P. steyermarkii* as a female parent and *P. nubigena* as a male parent, both species being native to the forested mountains of western Guatemala. The combination of nuclear and cytoplasmic markers appeared to be a more powerful tool to assess phylogenetic relationships than either type alone.

Barley: Barley is one of the most intensively studied species for variation at the molecular and biochemical levels, principally, through the activities of the Davis (e.g., Saghai-Maroof et al. 1984; Neale et al. 1986) and Israel groups (e.g., Nevo et al. 1979). Isozyme variation studies by Jørgensen (1976), Kahler and Allard (1981), and Nevo et al. (1979; 1986) and cpDNA or mtDNA analyses by Clegg et al. (1984), Neale et al. (1986), and Holwerda et al. (1986) confirmed that *Hordeum vulgare* subsp. *spontaneum* is the progenitor of cultivated barley.

Brassica spp.: The relationships among the nuclear genomes of cultivated *Brassica*

species are summarized by the triangle of U (1935). In this triangle, the diploid species (*B. rapa* (syn. *campestris*), n=10; *B. nigra*, n=8; *B. oleracea*, n=9) occupy the apices and the three amphidiploid species (*B. carinata*, n=17; *B. juncea*, n=18; *B. napus*, n=19) the sides between their respective progenitors.

Because of maternal inheritance of cpDNA, it was possible to identify the maternal parent of the amphidiploids (Erickson et al. 1983; Palmer et al. 1983). *B. carinata* and *B. juncea* derived their cytoplasm from *B. nigra* and *B. campestris*, respectively. Part of *B. napus* derived its cytoplasm from *B. oleracea*, whereas the other part may have derived its cytoplasm through introgressive hybridization from another *Brassica* species (Palmer et al. 1983). Identity of the parents of the amphidiploids was confirmed by Song et al. (1988a) and Hosaka et al. (1990) on the basis of nuclear genome-specific RFLP markers and by Delseny et al. (1990) on the basis of nuclear rRNA gene polymorphism. The molecular data also confirmed previous results obtained by various approaches such as cytogenetics, isozymes, and artificial resynthesis (see Hosaka et al. 1990 for references).

Cotton: There are four species of cultivated cotton. Two are diploid, Old World species with an A genome -- *Gossypium arboreum* from Asia and *G. herbaceum* from Africa -- and the other two are allotetraploid, New World species combining the A and D genomes -- *G. barbadense and G. hirsutum*. Wendel et al. (1989) observed that the two Old World species, which overlap considerably in their morphology, could readily be distinguished by certain isozyme alleles. These data were consistent with earlier cytogenetic and F_2 hybrid breakdown data and suggested that the two species resulted from independent domestication although the actual ancestors of these two species are not known with reasonable certainty and may have become extinct. Isozyme analyses of *G. barbadense* by Percy and Wendel (1990) showed that the area of northwestern South America contains the highest diversity and that advanced cultivated genotypes are mostly derived from western Andean materials. In addition, the diversity of advanced cultivars appears to have been enriched through introgression with *G. hirsutum*. Analyses of cpDNA showed that an A genome was maternal parent of both New World allotetraploids (Wendel 1989). In addition, the degree of divergence of between cpDNA

of the New World species and that of Old World A genome species suggested that the allotetraploid species arose 1-2 million years ago most likely after natural dispersal of an A genome species to the New World (Wendel 1989).

Lentil: Morphological, ecological, cytogenetic, and archaeological data support *Lens culinaris* subsp. *orientalis* (nomenclature of Ladizinsky et al. 1984) as the ancestor of lentil, *L. culinaris* subsp. *culinaris*. Allozyme and nuclear RFLP analyses confirmed the close relationship between the two taxa (Pinkas et al. 1984; Hoffman et al. 1986; Havey and Muehlbauer 1989).

Lettuce: Kesseli et al. (1991) analyzed diversity for nuclear RFLP loci detected by cDNA or genomic *Pst*I clones in *Lactuca* spp. The probes were chosen on the basis of two criteria: 1) the polymorphism, preferably a simple banding pattern, had a known location on the lettuce genetic map; and 2) the loci identified were distributed at regular intervals along linkage groups, at the end of linkage groups, or near disease resistance genes. Based on these analyses, the closest wild relative of lettuce -- *Lactuca sativa* -- is *L. serriola*, whereas other species -- *L. saligna* and *L. virosa* -- were more distantly related. The different cultivar groups of lettuce (butterhead, crisphead, cos, and looseleaf) could be distinguished at the RFLP level. Some 30 % of the alleles of *L. sativa*, however, were not found in *L. serriola*. Some of these alleles were found either in the butterhead or crisphead cultivar groups but not both. Overall, these data suggest that cultivated lettuce resulted from several domestications each leading to a separate cultivar group, although it was not possible to identify all the actual wild ancestors.

Maize: Several theories have been proposed to account for the origin of maize (Galinat 1971; Mangelsdorf 1974; Beadle 1978; Iltis 1983). One of the issues distinguishing these theories is the role of teosinte, a wild relative of maize from Mexico and Central America. Several lines of evidence, among them molecular and biochemical markers, now favor teosinte as the actual and immediate progenitor of maize. Multivariate analyses of allozyme frequencies in the genus *Zea* by Doebley et al. (1984) showed that the wild taxon most similar isoenzymatically to the cultivars is *Z. mays* subsp. *parviglumis*, a short-spikeleted annual teosinte adapted to mesic intermediate altitudes. Other teosintes, such as *Z. mays* subsp. *mexicana*, a large-spikeleted annual

teosinte adapted to arid high altitudes, *Z. luxurians*, an annual teosinte from southeastern Guatemala, and perennial teosintes *Z. perennis* and *Z. diploperennis*, were more distantly related to the cultivars. Chloroplast DNA restriction site analyses confirmed that annual teosintes were closely related to maize; however, these analyses could not distinguish between *Z. mays* subsp. *parviglumis* and *mexicana* as the most likely ancestral form (Doebley et al. 1987). Further confirmation that teosinte is the ancestor of maize was provided by rDNA restriction site analysis (Zimmer et al. 1988).

Isozyme data agreed with morphological and ecological information in that they were able to distinguish between *Z. mays* subsp. *parviglumis* and subsp. *mexicana*, which also showed differences in morphology and adaptation. Isozyme data did not agree with morphological data in that they showed *Z. mays* subsp. *parviglumis* to be the most closely related to maize, whereas morphologically *Z. mays* subsp. *mexicana* is most similar (maizoid) to maize (Doebley 1990).

Pea: An analysis of cpDNA diversity in *Pisum* spp., including cultivated pea, *Pisum sativum*, and wild pea species *P. humile*, *P. elatior*, and *P. fulvum*, showed that the most common cpDNA type among cultivars was also found in an accession of *P. humile* of northern Israel confirming morphological and isozyme data (Palmer et al. 1985).

Peanut: Peanut (*Arachis hypogea*) is an allotetraploid native to South America as is a closely related wild species, *A. monticola*. RFLP analyses using random *Pst*I genomic probes and RAPD markers, revealed high level of similarity between the two species, confirming earlier observations resulting from morphology, crossability, and seed protein studies (Halward et al. 1991; Kochert et al. 1991). This strongly suggested that *A. monticola* is the wild ancestor of peanut.

Pearl millet: Chloroplast DNA data showed no differences between *Pennisetum glaucum* subsp. *monodii* and *P. glaucum* subsp. *glaucum*, confirming that the former could be the ancestor of this African cereal (Gepts and Clegg 1989).

Peppers: The genus *Capsicum* is native to the Americas and contains five domesticated taxa. Isozyme analyses have shown that these five taxa did not result from a single domestication, but rather that at least three domestications can be suggested

(McLeod et al. 1982, 1983). Indeed, dendrograms based on isozyme genetic distances showed three clusters, each containing at least one domesticated and two wild taxa. In one group, *C. baccatum* var. *baccatum* is the ancestor of *C. baccatum* var. *pendulum*. In the second group, it is not possible to suggest a progenitor for *C. pubescens* but the closest wild species is *C. eximium*. Finally, in the third group, the cultivated forms of three species -- *C. annuum*, *C. frutescens*, and *C. chinense* -- were very similar among themselves and to a wild taxon, *C. annuum* var. *aviculare*. No allozyme analyses were conducted on the wild taxa within *C. frutescens* and *C. chinense* and, therefore, it was not possible to determine whether the three taxa in this group resulted from a single domestication followed by divergence or from three domestications involving an already diverged ancestor. Numerical taxonomy study by Pickersgill et al. (1979), however, suggested that the wild forms of these three species form a complex that shows a weak geographic differentiation and in which multiple domestications could have taken place.

Potato: Cultivated potato includes several species at various ploidy levels. At the diploid level, there are three cultivated species: *Solanum goniocalyx*, *S. phureja*, and *S. stenotomum*. At the tetraploid level, *S. tuberosum* is subdivided into two subspecies: subsp. *andigena* and subsp. *tuberosum*. Restriction digests of cpDNA showed that the three cultivated diploid species are closely related, confirming morphological, genetic, and biochemical traits. In addition, *S. goniocalyx and S. phureja* did not originate independently but appeared to have been derived from *S. stenotomum* (Hosaka and Hanneman 1988b). At the tetraploid level, *S. tuberosum* subsp. *andigena* contained several cpDNA types some of which (A, S, and C types) were also found among the diploid cultivated species, suggesting that this taxon was derived from the cultivated diploid species probably by repeated bilateral sexual polyploidizations involving the fusion of 2n gametes. This polyploidization process may also have involved wild diploid species as shown by the presence of the W type cpDNA (Hosaka and Hanneman 1988b). *S. tuberosum* subsp. *tuberosum* display a characteristic T cpDNA type, which is not found in any wild materials but can also be observed in *S. tuberosum* subsp. *andigena*. In South America, this T type is found at high frequencies in coastal Chile in both subsp. *tuberosum* and subsp. *andigena*. These data suggest that subsp. *tuberosum* was derived

from subsp. *andigena* from Chile. Additional data show that subsp. *tuberosum* supplanted subsp. *andigena* in Europe after the late blight epidemic in the 1840s.

Rice: Cultivated rice consists of two species, each with its corresponding wild species: Asian rice (*Oryza sativa*) and the associated wild species *O. rufipogon*, African rice (*O. glaberrima*) and the associated wild species *O. breviligulata* (syn. *O. barthii*). Relationships among and within these species have been extensively studied by Second and collaborators (Second 1982; Cordesse et al. 1990; Dally and Second 1990) and Japanese workers (e.g., Endo and Morishima 1983) on the basis of a combination of molecular, biochemical, and hybridization data. Isozyme analyses showed that Asian rice was derived from *O. rufipogon* via at least two domestications, one in South or Southeast Asia leading to the Indica types and the other in China leading to the Japonica types. Isozyme data further showed that African rice was distinct from the Asian species and arose from *O. breviligulata*. (Second 1982). Chloroplast DNA and ribosomal gene spacer length data confirmed independent domestication of *O. glaberrima* and the origin of the differentiation within *O. sativa* (Indica or Hsien and Japonica or Keng) in a divergence within *O. rufipogon* predating domestication (Cordesse et al. 1990; Dally and Second 1990).

Soybean: Cultivated soybean (*Glycine max*) and its presumed wild ancestor (*G. soja*), both annual species, showed identical ribosomal gene repeats and no variability was found within these two taxa in contrast with the wild perennial species, which displayed extensive repeat and restriction site variation (Doyle and Beachy 1985).

Tomato: The cultivated tomato (*Lycopersicon esculentum*) is part of a group of three *Lycopersicon* species synthesizing colored carotenoid pigments in contrast with other, green-fruited *Lycopersicon* species. Rick and Fobes (1975) examined isozyme variation in these three species to determine the wild ancestor of the cultivated tomato. The taxon with the closest relationship appeared to be *L. esculentum* var. *cerasiforme*, the cherry tomato, an aggressive weed from the Andean region. *L. cheesmannii* is native to the Galapagos Islands and it crosses freely with *L. esculentum*. The majority of its accessions deviated from the *L. esculentum* isozyme profile by one or two alleles. *L.*

cheesmannii did not apparently play a role in tomato domestication but was derived from *L. pimpinellifolium* populations in northwestern Peru as suggested by the close relationship in isozyme profile with accessions of the latter from that area. The heterogeneity at the allozyme level was much higher in *L. pimpinellifolium* than in *L. esculentum* and the number of allelic differences between the two species ranged from 2 to 7. Based on these data, Rick and Fobes (1975) propose that the ancestor of the cultivated tomato is *L. esculentum* var. *cerasiforme*. Mitochondrial DNA studies of McClean and Hanson (1986) and single-copy nuclear RFLP studies of Miller and Tanksley (1990) confirmed the ancestral nature of *L. esculentum* var. *cerasiforme*.

Wheat: Wild and cultivated wheats form a polyploid series at three ploidy levels. The diploid level includes *Triticum monococcum* and *T. urartu* (AA genome), *T. speltoides* (genome related to the B genome of tetraploid and hexaploid wheats), and *T. tauschii* (genome related to the D genome of tetraploid and hexaploid wheats). The tetraploid level includes *T. timopheevii* (AAB^tB^t or AAGG genome) and *T. turgidum* (AAB^cB^c genome), whereas the hexaploid level consists of *T. aestivum* (AAB^cB^cDD genome) (Dvořák et al. 1988). Several studies have attempted to identify the actual wild ancestor for each of the cultivated taxa and the donors of the various genomes. Studies based on chromosome pairing have provided a substantial amount of information on these questions although these results should be interpreted with caution because of the genetic control over chromosome pairing in amphidiploids (Dvořák 1988). Molecular and biochemical markers have been used to provide additional evidence in support of cytogenetic analyses or in those case where cytogenetic analyses had not been conclusive.

Isozyme analyses had shown that the wild ancestor of *T. monococcum* subsp. *monococcum* (einkorn wheat) is *T. monococcum* subsp. *aegilopoides* (syn. var. *boeoticum*) and that the donor of the D genome is *T. tauschii* subsp. *strangulata* (Asins and Carbonell 1986; Jaaska 1980, 1981; Nakai 1981). Analyses of cpDNA and mt DNA show that *T. turgidum* var. *dicoccoides* was the maternal parent of *T. turgidum* var. *turgidum* and that the latter, in turn was the progenitor of *T. aestivum* (Tsunewaki and Ogihara 1988; Graur et al. 1989). More recently, Dvořák et al. (1988) showed that

polymorphisms in the abundance and length of restriction fragments of repetitive sequences (RSAPs and RFLPs, respectively) were a very useful tool to determine the origin of genomes in polyploid plants and, particularly, in wheat. Specifically, they were able to confirm that while the donor of the A genome of *T. monococcum* subsp. *monococcum* is *T. monococcum* subsp. *aegilopoides*, the donor of the A genome in the tetraploid and hexaploid species is *T. urartu* (Dvořák et al. 1988).

The origin of the B genome has been a long-standing controversy. Based on morphological and isozyme analyses, the most likely donor appeared to be *T. speltoides* although other species of the Sitopsis section (*T. longissimum*, *T. searsii*, *T. bicorne*, and *T. sharonense*) could not be excluded (Sarkar and Stebbins 1956; Jaaska 1980, 1981; Asins and Carbonell 1986). Analyses of cytoplasmic genomes showed that a B genome species was the maternal parent of the polyploid species but could not identify specifically this species (Tsunewaki and Ogihara 1988; Graur et al. 1989). However, nucleotide sequences of 26S and 18S rRNA genes (Dvořák and Appels 1982) and RSAPs and RFLPs of repetitive sequences (Dvořák and Zhang 1990) revealed that *T. speltoides* is the closest extant species of the Sitopsis section to both tetraploid species. The G genome of *T. timopheevii* was contributed by *T. speltoides* whereas the B genome of *T. turgidum* was contributed by a species in the evolutionary lineage of *T. speltoides*. This species may not have been discovered yet or is extinct.

Wild rice: Isozyme data of Warwick and Aiken (1986) confirmed a taxonomic treatment that considered two species, *Zizania aquatica* with three varieties, and *Z. palustris* with two varieties, var. *palustris* and var. *interior*. These analyses further showed that the cultivars were derived from *Z. palustris* var. *interior*.

2. Divergence in the cultivated gene pool

In addition to the information mentioned for several crops in the previous section, some more detailed studies have been performed on the following crops:

Brassica spp.: Nuclear RFLPs showed that *B. rapa* consists of two major groups, the European group, including turnip, turnip rape, spring broccoli raab, and sarson, and

the East Asian group, including Chinese cabbage, pak choi, and narinosa. This separation confirms previously obtained morphological and isozyme data. Initial domestication may have taken place in Europe from a wild *B. rapa*. Wild populations or primitive cultivars may have been disseminated to Asia where additional domestication or diversification took place. Within the European group, turnip, turnip rape, and spring broccoli raab, can be distinguished from sarson, the former having originated in Europe and the latter in India. In the Eastern Asian group, Chinese cabbage appears to have been derived from pak choi, the more primitive form (Song et al. 1988b, 1990).

B. oleracea appears to have a monophyletic origin based on nuclear RFLP data (Song et al. 1988b, 1990). Diversification following the initial domestication has yielded a wide variety of morphotypes. Portuguese tree kale and Chinese kale seemed to be the most closely related forms to wild *B. oleracea*. Cabbage was closely related to these kale forms, whereas broccoli and cauliflower were clustered together.

Maize: Wellhausen et al. (1952) established a racial classification of Mexican maize on morphological grounds. Doebley et al. (1985) provided evidence that these maize races can be grouped into three complexes of morphologically and ecologically similar races: 1) the high elevation, Mexican pyramidal complex, including races Palomero Toluqueño, Chalqueño, and Cónico; 2) the northern complex, which includes races Azul and Apachito; 3) the remaining bulk of races, including the southern and western lowland dents and flours such as Tuxpeño and Tabloncillo. Frequencies of 22 isozyme alleles were correlated with altitude, which, in Mexico, is associated with precipitation and temperature, i.e. the lowlands are hotter and moister whereas the highlands tend to be cooler and drier.

In the United States, the Northern Flint and Southern Dent, the two landraces that gave rise by hybridization to the Midwestern dent, the basis for many of the contemporary elite hybrid maize cultivars, exhibit a marked divergence at the isozyme level (Doebley et al. 1988). This differentiation is correlated with pronounced differences in morphology and genome size and with a strong heterotic response.

Potato: Chloroplast DNA in *S. tuberosum* subsp. *andigena*, displayed a continuous gradient between Mexico and Chile. At the northern end of this gradient, this subspecies

was monomorphic for the A cpDNA type. Further south, the frequency of the A type decreased whereas the frequency of the C, S, and W cpDNA types increased. At the southern end of the distribution in Chile, a high frequency of T cpDNA was observed (Hosaka and Hanneman 1988a).

Rice: Second (1982) and Glaszmann (1987) were able to identify two major groups of *O. sativa* varieties based on isozyme diversity. These two groups correspond to the Indica and Japonica types previously identified on the basis of morphological, hybridization, and ecogeographical arguments, among others. The Japonica group also included the Javanica types - tropical variants of the temperate Japonica types, confirming previous results of Oka (1958). Second (1982) also suggested that some of the allozyme diversity of *O. sativa* could have originated from crosses between Indica and Japonica types. Studies of the rDNA spacer polymorphism showed that the Japonica genotypes have generally shorter spacers than Indica genotypes. No variation was observed for spacer length in *O. glaberrima* (Cordesse et al. 1990). RFLPs for nuclear sequences confirmed the divergence between the Japonica and Indica types, showed that the Indica group contains a higher amount of variation compared to the Japonica group, and revealed that the process Indica-Japonica differentiation affected at least 11 of the 12 chromosomes of Asian rice (Zhang et al. 1992; Wang et al 1992). This organization of genetic diversity in Asian rice resembles in many ways that of *P. vulgaris*.

3. Gene flow between wild and cultivated forms

Brassica spp.: *B. napus* contains at least two cpDNA types, one of which is similar to the cpDNA of *B. campestris*, the maternal parent of the amphidiploid. The other cpDNA appears to have been derived from an ancestral form prior to the divergence of *B. campestris* and *B. oleracea*. Irrespective of their cpDNA type, the *B. napus* accessions tested showed equal contributions of the *B. campestris* and *B. oleracea* genomes. This contradiction between cpDNA-based and nuclear DNA-based phylogenies suggested to Palmer et al. (1983) that introgression may have played a role in the origin of *B. napus*. The latter cpDNA type may have been introduced by hybridization and backcrossing via the maternal parent into *B. napus*. Additional evidence for a polyphyletic origin of *B.*

napus - and *B. juncea* - was obtained by Song et al. (1988a) and Hosaka et al. (1990) using nuclear RFLP markers.

Maize: As mentioned earlier, the progenitor of cultivated maize is an annual teosinte, *Z. mays* subsp. *parviglumis*. A distinct annual teosinte, *Z. mays* subsp. *mexicana*, is not the direct progenitor of maize but may have contributed genetic diversity to the cultivated gene pool: in the region where maize and subsp. *mexicana* are sympatric, isozyme alleles characteristic of the wild taxon were observed in cultivars. These alleles were absent from the cultivars elsewhere (Doebley et al. 1987). Evidence for introgression in the other direction is provided by the perennial teosinte *Z. diploperennis* in which a plant showed isozyme alleles at two linked loci characteristic of the cultivated gene pool, suggesting that the chromosome segment marked by the two loci was introduced from maize (Doebley et al. 1984).

Pearl millet: Cultivated pearl millet is sympatric with its wild ancestor in the Sahel. Viable and fertile hybrids can be formed, although male sterility in these crosses has been described (Marchais and Pernès 1985). Isozyme analyses revealed that pollen from wild forms was preferentially involved in fertilization on wild pistils in pollination experiments with mixtures of pollen from wild and cultivated forms. Conversely, pollen from cultivated forms preferentially effected fertilization on cultivated pistils after mixed pollinations with wild pollen (Robert et al. 1991). This type of reproductive isolation may explain at least in part why wild and cultivated pearl millet maintain their phenotypic integrity in spite of their sympatry and predominant allogamy.

Rice: Some accessions of *O. breviligulata*, the ancestor of African rice, contained isozyme alleles that were absent in most of the other *O. breviligulata* accessions, but were found in Asian rice, *O. sativa*. This suggests that some *O. breviligulata* accessions arose from hybridizations between wild *O. breviligulata* and *O. sativa*, after the introduction of cultivars of the latter in Africa (Second 1982).

Tomato: *L. pimpinellifolium*, the currant tomato, is native to the coastal areas of Ecuador and Peru with maximum isozyme diversity observed in northern Peru. In coastal areas of Ecuador and Peru, *L. pimpinellifolium* and *L. esculentum*, including var. *cerasiforme*, share certain isozyme alleles that are not found in *L. esculentum* of other

regions. This suggests introgression from *L. pimpinellifolium* into *L. esculentum* in that region and confirms similar observations based on shared traits such as the *cm* gene (severe leaf distortions at low temperatures) and the Ge^{p} gene (gamete eliminator) (Rick and Fobes 1975).

IV. DIFFERENTIATION BETWEEN WILD PROGENITOR AND CULTIVATED DESCENDANT

Having identified the likely wild ancestor, it is possible then to examine the evolutionary processes that have led from this wild ancestor to the modern cultivated descendant. The study of crop evolution is a potentially useful experimental approach to the study of evolution in general because both the starting material and the end product are available, the time frame (approximately 10,000 years) is known, and the genetics of crop plants is generally more developed than that of wild plants.

A. Fate of genetic diversity during and after domestication

The domestication process has involved selection -- whether conscious or unconscious (Harlan 1975; Rindos 1984; Heiser 1988) -- for particular characteristics that make the crop more amenable to cultivation and more attractive to consumers. Like any other selection process, domestication is expected to lead to a reduction in diversity (Ladizinsky 1985). Isozyme and, to a lesser extent, seed protein and RFLP studies have confirmed this expectation, although exceptions to this trend have been described.

1. *Phaseolus* spp.

Comparison of phaseolin diversity in wild and cultivated *P. vulgaris* shows a reduction in diversity during the process of domestication. For example, in the Middle American gene pools, wild beans from Mexico show some 20 'M' phaseolin variants whereas Mexican cultivars show one major phaseolin type ('S') and two variants ('Sb' and 'Sd') (Gepts et al. 1986; Koenig et al. 1990). In the Andean gene pool, the reduction is not as marked partly because the diversity in the wild ancestor is not as high and perhaps also because multiple domestications may have taken place leading to the existence of several phaseolin types among cultivars of that region ('C', 'T', 'H', and 'A'). Various phaseolin types have, however, been identified among wild beans in the Andean gene pool that are absent from the cultivated gene pool.

Further evidence for reduction in diversity during domestication has come from the identification of arcelin seed protein and its role in resistance against seed weevils

(Bruchidae). Resistance against these insects has only been found in wild *P. vulgaris* but not in cultivated *P. vulgaris* in spite of surveys of several thousands of cultivated accessions (Van Schoonhoven and Cardona 1982). Resistance is due to the presence of arcelin seed protein (Osborn et al. 1988), which has only been found in wild *P. vulgaris* accessions from Mexico (Romero et al. 1986; Osborn et al. 1986). Further analyses have shown that this protein is part of a tightly linked family of genes coding for seed proteins including also a lectin and an α-amylase inhibitor (Hoffman and Donaldson 1985; Osborn et al. 1986, 1988; Moreno and Chrispeels 1989).

Isozyme analyses have not shown as pronounced a reduction in diversity during domestication mainly because levels of diversity, i.e. the number of alleles per isozyme locus, are much lower than those for phaseolin in wild beans (Koenig and Gepts 1989b; Singh et al. 1991b). These same analyses show, however, that among cultivars there are large groups of genotypes that cannot be distinguished by isozymes (although they often can be distinguished at the morphological level) whereas wild accessions that cannot be distinguished by isozymes are much rarer. It appears that domestication has led to an increase in the frequency of certain multilocus associations. It is not known, however, whether these specific associations have some type of adaptive advantage or if they result from random processes.

Additional evidence for this reduction in diversity has been provided by a study on photosynthetic parameters by Lynch et al. (1992). They reported that wild beans sampled along the distribution area from Mexico to Argentina displayed considerably more diversity for photosynthetic parameters than landraces in the same areas.

After the initial domestication process, dissemination of common bean cultivars from their domestication area resulted in additional reductions in diversity. Phaseolin diversity data of landraces and heirloom varieties in Europe, Africa, and the U.S.A., show a reduction in the number of phaseolin types present compared to those observed in landraces of the domestication areas (Gepts and Bliss 1988; Gepts et al. 1988). Minisatellite sequence fingerprinting data of California common bean cultivars show high levels of uniformity within commercial classes such as the kidney and pink beans. This near-uniformity contrasts with the high levels of polymorphism observed among landrace

cultivars of the corresponding races in Latin America, races Nueva Granada and Durango, respectively (G. Sonnante, T. Stockton, and P. Gepts, unpubl. data). Such uniformity is brought about by the stringent agronomic and product quality requirements imposed on any new cultivar, which in turn favor crosses between closely related, elite lines.

This reduction in diversity at the molecular and biochemical levels is in contrast with the marked increase in diversity observed for morphological and phenological traits, of which seed color is perhaps the most illustrative example. Wild beans usually show seed of agouti color resulting from a dark striping and spotting pattern superposed on a clear background. There is little variation for this seed color pattern, which probably plays a role in protection from predators. In contrast, cultivated beans are known for the wide variety of colors and color patterns. This apparent contradiction between genotypic (molecular) and phenotypic (morphological) data on the evolution of diversity during domestication will be addressed in the discussion.

2. Other crops

In a sample of approximately 20 wild ancestor-cultivated descendant comparisons of isozyme diversity, Doebley (1989) calculated that, on the average, total heterozygosity (H_t) and expected heterozygosity per population (H_e) decreased around 25%, and the proportion of polymorphic loci and number of alleles per locus around 10% in the domesticate compared to the wild progenitor. Some 7 wild ancestor-cultivated descendant comparisons of cpDNA diversity showed a reduction in RFLP diversity (reviewed in Doebley 1992). Nuclear DNA markers such as genes for ribosomal RNA and single-copy RFLPs were also characterized by a reduction in diversity (Doyle 1988; Allard 1988; Gepts and Clegg 1989; Havey and Muehlbauer 1989; Cordesse et al. 1990). Recently, Kochert et al. (1991) and Halward et al. (1991) analyzed diversity for RFLPs and RAPD markers in peanut, *Arachis hypogea*, its wild ancestor *A. monticola*, both tetraploid species, and related wild diploid *Arachis* species. They observed considerable levels of diversity at the diploid level but not at the tetraploid level. The limited number or single allotetraploidization event that led relatively recently to the *A. hypogea* - *A.*

115

monticola complex, may have induced a strong bottleneck in genetic diversity. Similar bottlenecks have taken place in other allopolyploid species of recent origin such as hexaploid wheat (Chao et al. 1989). Allopolyploidy is not, however, always associated with low diversity levels as shown by Wang et al. (1992) for *Oryza latifolia*, a wild allotetraploid species.

A few exceptions to this trend have been reported. In barley, isozymes and cpDNA showed equivalent levels of diversity in wild and cultivated accessions whereas mtDNA diversity was higher in the cultivars than in the wild ancestor(Holwerda et al. 1986; Jana and Pietrzack). These results contrast with those of Brown and Munday (1982) for isozymes and Clegg et al. (1984) and Neale et al. (1986) for cpDNA. It is not clear what the cause is of these discrepancies. One possibility is the sampling of the plant material which differed among these studies for the total number of accessions, the relative number of wild and cultivated accessions, and their respective geographic origins. In pearl millet (*Pennisetum glaucum*), RFLPs of genes for ribosomal RNA and the *Adh-1* locus analyzed on the same set of wild and cultivated accessions showed contrasting trends in diversity: whereas the former showed a clear reduction in diversity between the wild ancestor and the cultivated descendant, the latter showed comparable levels of diversity (Gepts and Clegg 1989). It is possible that the diversity at the *Adh-1* locus in the cultivars is maintained through introgression with the sympatric wild ancestor. For rRNA genes, this mechanism would not operate if these genes are tightly linked to an essential feature of the cultivated phenotype; hence, after outcrossing with a wild population, selection for the cultivated phenotype in subsequent generations would eliminate introgressed rRNA genes from wild populations. Information on the actual linkage relationships between these molecular markers and genes controlling the cultivated phenotype is needed to verify this hypothesis.

B. <u>Genetics of morphological and physiological differentiation between wild ancestor and cultivated descendant</u>

There have been surprisingly few comprehensive studies on the genetics of the morphological differences distinguishing wild progenitor and cultivated descendant in

crops. An important exception is the study of Doebley et al. (1990) and Doebley and Stec (1991) in maize. They crossed a cultivated accession of the primitive maize race Chapalote with a wild teosinte belonging to *Z. mays* subsp. *mexicana*. The F_2 progeny was then analyzed for both vegetative and reproductive traits distinguishing teosinte and maize and for RFLP markers distributed at regular intervals on the maize RFLP map. The difference for most traits was controlled by one or two major loci and various loci with smaller effects. The loci with major effects tended to be grouped in five chromosome regions on chromosomes 1 to 4 of the maize genome. This genome distribution may explain why maize-like and teosinte-like plants can be recovered at relatively high frequency in segregating populations. These data also provided some insights on the essential steps involved in the transformation of teosinte into maize. These included 1) changes in the lateral branches, including shortening of the internodes and transformation of the lateral inflorescence into female rather than male structures; 2) changes that led to development of male secondary traits in female structures such as soft glumes and paired spikelets; 3) transformation from two-ranked to four-ranked cupules; and 4) evolution to limit disarticulation of the ear, a character shared with other cereals.

In pearl millet, some efforts have been made to map traits distinguishing wild and cultivated forms (Pernès et al. 1984 and references therein; Marchais and Tostain 1985). Their studies indicated linkage of certain spikelet traits including the length of the spikelet pedicel, the dehiscence of the spikelet, and length of the spikelet bristles. As in maize, this linkage of traits would explain why parental types, whether wild or cultivated, can be recovered relatively easily in the progeny of these crosses. Their studies also showed, however, that other mechanisms may act to limit the gene flow and recombination between the wild and cultivated gene pools such as male sterility (Marchais and Pernès 1985) or pollen competition (Marchais and Tostain 1985; Robert et al. 1991).

During the domestication process, selection has significantly altered the morphology and physiology of the bean plant. The principal traits affected include seed dormancy, dispersal, and size, growth habit, photoperiod sensitivity, and harvest index

117

(reviewed by Smartt 1988; Debouck 1991; Gepts and Debouck 1991). Seed of cultivated genotypes are non-dormant, i.e. they will start imbibing water and germinating immediately unlike seeds of the wild ancestor. The reduced fiber content in pods of cultivated beans prevents the explosive dehiscence of the fruit at maturity and dispersal of the seed, characteristic of the wild progenitor. In cultivars, seed as well as other organs such as leaves and pods show "gigantism" compared to wild materials.

The growth habit of the wild ancestor is a climbing habit characterized by numerous internodes and branches on the main stem. Compared to wild beans, cultivars show fewer internodes and branches. The end result of this evolution of the growth habit is the determinate bush cultivars found not only in certain Latin American landraces but also in snap bean and kidney bean cultivars in temperate countries. Concurrently with this trend towards a more compact growth habit is an increased earliness. In addition, the more compact growth habit, the larger seeds, and a shift in the partitioning of photosynthates towards the seeds, leads to a higher harvest index.

Wild beans are photoperiod sensitive, i.e. they will flower only under the short days characteristic of their native environments in the tropics. In these environments, photoperiod sensitivity fulfills an important adaptive function as it allows the plant to time its flowering so that seed maturation will coincide with the onset of the dry season. Maturation in moist conditions will lead to seed germination in pods. The loss of photoperiod sensitivity allowed cultivated beans to be grown under different latitudes particularly at the higher latitudes of temperate countries.

What is the genetic control of these differences between wild and cultivated beans? There are few published reports on segregation analyses of crosses between wild and cultivated beans. An important exception is the contribution of Motto et al. (1978), which showed that seed weight, an essential difference between wild and cultivated beans, is a quantitative trait of high heritability (h^2_{NS} = 0.86). Two independent studies performed under different locations and involving cultivated x cultivated crosses obtained comparable heritability estimates (Chung and Stevenson 1973: h^2_{NS} = 0.58; Nienhuis and Singh 1988b: h^2_{NS} = 0.74)

Photoperiod sensitivity in a wild x cultivated bean cross was conditioned by two

complementary genes (Rüdorf 1959). This genetic control was also observed by Coyne (1967), although other segregation ratios have been reported (reviewed by Singh 1991).

For the other traits distinguishing wild and cultivated beans, only information from cultivated x cultivated crosses is available. This information, reviewed by Gepts (1990) and Gepts and Debouck (1991), shows that genetic control of these traits is simple: it involves only a few genes with a major phenotypic effect. Experiments are under way to obtain segregation and linkage information for these traits from a wild bean x cultivated bean cross. Preliminary results show that the genetic control of these traits involves the same number of genes as in cultivated x cultivated crosses (E. Koinange and P. Gepts, unpubl. results).

V. DISCUSSION

This review leads to a number of issues, which I will discuss in the next paragraphs.

Firstly, to what extent can we develop a general model of the domestication process? This model would describe the general features in genetic terms of the process, including, for example, the domestication pattern (one vs. several domestications), the fate of genetic diversity (reduction or increase and of what magnitude), the role of gene flow between wild ancestor and cultivated descendant after domestication (direction, magnitude, and effect on overall diversity in the gene pool), and the genetic and physiological differentiation between wild ancestor and cultivated descendant. From the examples discussed in this review, it is apparent that such a model can only be couched in general terms because each crop appears to have undergone a unique domestication process. This should not come as a surprise given the large number of biological, environmental, and cultural factors that may have affected domestication.

Among the biological factors that can influence domestication, are the geographic and ecological distribution of the wild ancestor, the reproductive system (including the possibility of vegetative reproduction), and the life cycle (particularly the generation time). The larger the distribution, the more likely multiple domestications become although this need not be so as shown by results in cowpea (*Vigna unguiculata*) for which the wild ancestor is distributed across the Sahel and into southern Africa. In spite of this very extensive distribution, cowpea appears to have been domesticated only once (L. Panella and P. Gepts, unpubl. results). The ecological distribution and, particularly, the ecological amplitude can influence the distribution of the crop and the extent to which the crop will be disseminated to areas outside of its original environment. The reproductive system will influence the recombination system of the crop (*sensu* Grant 1975) and this may be important especially in the initial phases of the domestication when various pre-existing genes or newly-arisen mutations had to be recombined into the same progeny in order to generate a cultivated genotype. In addition, the reproductive system plays a role in the level of gene flow between wild and cultivated forms. The

generation time will also affect the recombination system and particularly the speed at which new recombinants can be formed and spread through the early domesticated populations. These three factors alone, and notwithstanding environmental and cultural differences, can, therefore, cause major differences in the domestication pattern among various crops. Extrapolation of results from one crop to the other should generally not be attempted. One possible exception is the reduction in diversity at the molecular level during domestication that has been observed for several crops.

A second issue is the apparent uncoupling between morphological variation, on one side, and molecular and biochemical variation, on the other. The reduction in diversity at the molecular level during domestication is accompanied by an increase in diversity at the phenotypic level. Cultivars exhibit great diversity in phenology, growth habit, seed type, etc., that is usually unmatched at the molecular and biochemical levels. How can this apparent paradox be resolved? Population genetics tells us that the usual fate of a mutation will be extinction (Fisher 1958, p. 81; Crow and Kimura 1970, p. 421; Kimura 1983, p. 196). Among mutations, there are, however, differences in selective value. Other factors being equal, mutations that confer a selective advantage have a higher chance of survival than neutral or deleterious mutations (Table 4). Mutations or novel gene combinations responsible for the cultivated phenotype (or so-called domestication syndrome, Hammer 1984) had a strong selective advantage because they conferred adaptation to the cultivated environment that was markedly different from the natural environment or because they struck the fancy of the early agriculturalists. Examples of such mutations are the lack of seed dispersal mechanism in cereals or legumes or white-seededness in beans. Under certain harvest systems, lack of rachis shattering will be selected for, whether consciously or unconsciously. Likewise, white seededness in a cultivated environment may have a selective advantage because of its novelty (although its effect on plant vigor and fitness is somewhat deleterious according to Dickson and Petzoldt 1988. Selective values of 2.50 have been obtained for non-brittle rachises compared to brittle rachises when einkorn wheat plants were harvested by uprooting or reaping with sickles (Hillman and Davies 1990). (It should be clear also that these two traits, while they have a selective advantage in the cultivated environment,

can be deleterious in a natural environment because of reduced seed dispersal and increased predation, respectively.) Molecular and biochemical markers, on the other hand, are usually neutral, i.e. most of the changes at the DNA sequence level will not have a selective advantage. The divergent trends can, therefore, be reconciled by the following scenario. Domestication traces back to a limited number of populations leading to a bottleneck in genetic diversity at the molecular and biochemical levels. During domestication, strong selection within this limited array of populations for morphological or physiological variants increased phenotypic diversity; however, the time span between initial domestication and the present (some 8,000-10,000 yr) has been insufficient to replenish the diversity at the molecular level of the cultivated gene pool. Clearly, more information is needed on domestication as a selection process from actual selection and simulation experiments such as those described by Hillman and Davies (1990).

A corollary of this scenario is that, for selection for traits of the domestication syndrome to be effective within the small populations and short time frame of domestication, these traits had to have a relatively high heritability. In this respect, high heritability may be as important as the number and magnitude of the genes involved. There are very few data on the heritability of phenotypic traits distinguishing wild ancestor and cultivated descendant. More data are available from cultivated x cultivated crosses. In *Phaseolus vulgaris*, these data (reviewed by Gepts 1990, Gepts and Debouck 1991) show that most of these traits are controlled by a limited number of genes with a high heritability including seed size for which the narrow-sense heritability can reach values above 0.70 (Motto et al. 1978; Singh and Nienhuis 1988). Additional information stemming from actual wild x cultivated crosses on inheritance (such as those of Doebley et al. 1990 and Doebley and Stec 1991) and heritability in a range of crops is required to test this hypothesis.

In addition, such experiments would shed light on another question, namely that of the linkage relationships of genes of the domestication syndrome. Both Doebley et al. (1990) in maize and Pernès et al. (1984) in pearl millet observed linkage among certain traits of the syndrome. Harlan (1975) suggested that linkage should be expected because linked genes will respond more easily to selection. One could argue that linkage may be

122

favorable when the desired alleles are in coupling, but may be disadvantageous when they are in repulsion. Maize and pearl millet are cross-pollinated species in which linkage may be important to maintain favorable associations of alleles when cultivated and wild forms are repeatedly subjected to gene influx from the other form in their respective environments. In predominantly self-pollinated species such as common bean, linkage maybe less important because inbreeding will strengthen multilocus associations even among unlinked loci.

There is limited information on the molecular basis of the changes that led to the domestication syndrome. If indeed many of these changes involved genes with major effect, as suggested by Harlan (1975), Doebley et al. (1990), and Doebley and Stec (1991), then it remains to be determined whether the major effect of a gene is due to a single mutation or a series of mutations with incremental effects at the locus of the gene. This information would also bear on the debate taking place currently as to the basis of major morphological changes in plants. Knight (1948), Hilu (1983), and Gottlieb (1984) have argued in favor of mutations with major effects, whereas Lande (1983) and Coyne and Lande (1985) have argued in favor of mutations with small effects.

The bottleneck in diversity that apparently distinguishes wild and cultivated forms of crop species may have occurred at any time from the earliest stages of domestication up to the present. Modern plant breeding, responding to the need for a highly uniformized cropping systems, distribution systems and consumer preferences, has tended to emphasize crosses within a narrow gene pool of elite materials (e.g., Silbernagel and Hannan 1988) with until recently little contribution of exotic germplasm (Duvick 1984). Dissemination of cultivated materials from the domestication areas to other regions of the world has also led to reductions in diversity because only small samples of the original diversity were introduced in the new regions. Perhaps the most significant reduction may have taken place in the domestication areas themselves, somewhat paradoxically because of the direct availability of the wild ancestor. In common bean, for example, cultivars from Mexico appear to have originated from a single domestication, possibly in the west-central part of the country (Gepts 1988). Wild common bean has, however, an extensive distribution in this country but populations from outside the

putative domestication area seem not to have contributed to the cultivated gene pool except for occasional cases of introgression. It is as if, once the first domesticated genotypes were obtained, they or the first domesticators somehow preempted additional domestications in other areas from local wild bean populations. It may have been easier to merely adopt the first domesticates than to repeat the selection procedures that led to domestication of the crop. An alternative and not mutually exclusive hypothesis is that the people who first domesticated crops obtained a selective advantage compared to people still relying on hunting and gathering perhaps because of a more abundant or more regular food supply. Following domestication, they spread to other regions and introduced their new technology. Some evidence for the latter hypothesis is provided by the studies of human gene frequencies on the European continent. Agriculture in Europe was introduced from the Near East during the Neolithic. The spread of agriculture could have taken place either via cultural diffusion without genetic consequences for human populations or via demic diffusion involving migration of the early agriculturalists from the Near East into Europe. The data of Ammerman and Cavalli-Sforza (1984), Sokal (1991) and Sokal and Rohlf (1991) provide evidence favoring demic diffusion. This may explain why, for example, cultivated cereals were introduced into Europe and no local domestications apparently took place, although wild relatives of these cereals are distributed in Europe. Analysis of these temporal and spatial patterns of variation during domestication may benefit from ancient DNA studies performed on archaeological seeds (Pääbo 1989; Pääbo et al. 1989).

Finally, crop evolution studies have important consequences for genetic conservation and breeding. They allow us to identify segments of the germplasm that are not represented or insufficiently represented in genetic resources collections such as the wild ancestral populations, which until recently had been neglected as a source of genetic diversity for breeding programs. In common bean, the discovery of arcelin genes in Mexican wild bean populations and the evidence for a genetic bottleneck between wild and cultivated common bean has led to an increased focus on wild ancestral materials in germplasm explorations and evaluations for disease resistance. These studies also provide us with a better understanding of the genetic relationships between various segments of

the cultivated gene pool. In common bean, our understanding of the divergence between the Middle American and Andean gene pools and the existence of races is leading to different breeding strategies. For example, combining ability studies have shown that in the Middle American gene pool, races Jalisco and Durango have high general combining ability for yield but race Mesoamerica has low general combining ability for yield (Nienhuis and Singh 1988a; Singh et al. 1989, 1992).

VI. CONCLUSIONS

A wide array of molecular and biochemical markers has been used to address issues in crop evolution. Each of these markers have their respective advantages and disadvantages. Depending on the objective of a particular study, one or the other will be more appropriate.

Molecular and biochemical markers have been very useful tools to provide information that either confirmed previous evidence based on morphology, cytogenetics, and other areas, or provided new evidence not available previously. Among the most significant contribution has been the more precise identification of specific populations or groups of populations as the presumptive ancestor of crops. Such identifications set the stage for comparisons between wild ancestor and cultivated descendant involving overall levels of diversity and the genetics and linkage relationships of traits distinguishing wild and cultivated forms. Data available suggest a reduction in diversity at the molecular level during domestication which contrasts with an increase in diversity at the phenotypic level.

Future studies will likely include: 1) a continuation of the current lines of study aimed at identifying the wild ancestor and the organization of genetic diversity in those crops where this has not been accomplished as yet; 2) characterization of the genetic and physiological differences between wild ancestor and cultivated descendant; 3) identification of spatial and temporal patterns of variation using ancient DNA from archaeological remains; and 4) study of the molecular basis of morphological and physiological evolution under domestication.

Acknowledgments: Research leading to this review on crop evolution was funded by the Research Programme of the International Board for Plant Genetic Resources, Rome, Italy, and the PSTC and Bean/Cowpea CRSP programs of the Agency for International

1 Development, Washington, DC.

VII. REFERENCES

ACOSTA, J. DE, 1590 *Historia natural y moral de las Indias*. Juan de León, Sevilla. Modern ed. Valencia Cultural, Valencia.

AMMERMAN, A. J., and L. L. CAVALLI-SFORZA, 1984 *The neolithic transition and the genetics of populations in Europe*. Princeton Univ. Press, Princeton, NJ.

APUYA, N. R., B. L. FRAZIER, P. KEIM, E. J. ROTH and K. G. LARK, 1988 Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. Theor. Appl. Genet. **75:** 889-901.

ASINS, M. J., and E. A. CARBONELL, 1986 A comparative study on variability and phylogeny of *Triticum* species. 2. Interspecific relationships. Theor. Appl. Genet. **72:** 551-558.

BAUDOIN, J. P., 1988 Genetic resources, domestication and evolution of lima bean, *Phaseolus lunatus*, pp. 393-407 in *Genetic resources of* Phaseolus *beans*, edited by P. GEPTS. Kluwer, Dordrecht, the Netherlands.

BEADLE, G. W., 1972 The mystery of maize. Field Mus. Nat. Hist. Bull. **43:** 2-11.

BLISS, F. A., 1980 Common bean, pp. 273-284 in *Hybridization of crop plants*, edited by W. R. FEHR and H. H. HADLEY. Crop Science Society of America, Madison, WI.

BONIERBALE, M., R. L. PLAISTED and S. D. TANKSLEY, 1988 RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics **120:** 1095-1103.

BROWN, A. H. D., and J. MUNDAY, 1982 Population genetic structure and optimal sampling of landraces of barley from Iran. Genetica **58:** 85-96.

BRÜCHER, H., 1988 The wild ancestor of *Phaseolus vulgaris* in South America, pp. 185-214 in *Genetic resources of* Phaseolus *beans*, edited by P. GEPTS. Kluwer, Dordrecht, the Netherlands.

BURKART, A., and H. BRÜCHER, 1953 *Phaseolus aborigineus* Burkart, die mutmassliche andine Stammform der Kulturbohne. Züchter **23:** 65-72.

DE CANDOLLE, A., 1882 *L'origine des plantes cultivées. English translution: The origin of cultivated plants*. Appleton, New York.

CHANG, C., J. L. BOWMAN, A. W. DEJOHN, E. S. LANDER and E. M. MEYEROWITZ, 1988 Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **85:** 6856-6860.

CHAO, S., P. J. SHARP, A. J. WORLAND, E. J. WARHAM, R. M. D. KOEBNER and M. D. GALE, 1989 RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor. Appl. Genet. **78:** 495-504.

CHASE, C. D., V. M. ORTEGA and C. E. VALLEJOS, 1991 DNA restriction fragment length polymorphisms correlate with isozyme diversity in *Phaseolus vulgaris* L. Theor. Appl. Genet. **81:** 806-811.

CHUNG, J. H., and E. STEVENSON, 1973 Diallel analysis of the genetic variation in some quantitative traits in dry beans. N.Z. J. Agric. Res. **16:** 223-231.

128

CIEZA DE LEON, P. DE, between 1541 and 1550 *La crónica del Perú*. Espasa-Calpe, Madrid.

CLEGG, M. T., 1990 Molecular diversity in plant populations, pp. 98-115 in *Plant population genetics, breeding, and genetic resources*, edited by A. H. D. BROWN, M. T. CLEGG, A. L. KAHLER and B. S. WEIR. Sinauer, Sunderland, MA.

CLEGG, M. T., A. H. D. BROWN and P. R. WHITFELD, 1984 Chloroplast DNA diversity in wild and cultivated barley: implications for genetic conservation. Genet. Res. **43:** 339-343.

CORDESSE, F., G. SECOND and M. DELSENY, 1990 Ribosomal gene spacer length variability in cultivated and wild rice species. Theor. Appl. Genet. **79:** 81-88.

COYNE, D. P., 1966 A mutable gene system in *Phaseolus vulgaris* L. Crop Sci. **6:** 307-310.

COYNE, D. P., 1967 Photoperiodism: inheritance and linkage studies in *Phaseolus vulgaris*. J. Hered. **58:** 313-314.

COYNE, J., and R. LANDE, 1985 The genetic basis of species differences in plants. Amer. Nat. **126:** 141-145.

CROW, J. F., and M. KIMURA, 1970 *An introduction to population genetics theory*. Harper and Row, New York.

CURTIS, S. E., and M. T. CLEGG, 1984 Molecular evolution of chloroplast DNA sequences. Mol. Biol. Evol. **1:** 291-301.

DALLAS, J. F., 1988 Detection of DNA "fingerprints" of cultivated rice by hybridization with a human minisatellite DNA probe. Proc. Natl. Acad. Sci. USA **85:** 6831-6835.

DALLY, A. M., and G. SECOND, 1990 Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, section Oryza). Cladistic-mutation and genetic-distance analysis. Theor. Appl. Genet. **80:** 209-222.

DEBOUCK, D. G., 1991 Systematics and morphology, pp. 55-118 in *Common beans; research for crop improvement*, edited by A. VAN SCHOONHOVEN and O. VOYSEST. CAB, Wallingford, Oxon, UK.

DEBOUCK, D. G., R. CASTILLO T. and J. M. TOHME, 1989a Observations of little-known *Phaseolus* germplasm of Ecuador. Plant Genet. Res. Newsl. **80:** 15-21.

DEBOUCK, D. G., A. MAQUET and C. E. POSSO, 1989b Biochemical evidence for two different gene pools in lima beans. Annu. Rept. Bean Improv. Coop. **32:** 58-59.

DELGADO SALINAS, A., 1985 Systematics of the genus *Phaseolus* (Leguminosae) in North and Central America. Ph.D., University of Texas, Austin.

DELGADO SALINAS, A., A. BONET and P. GEPTS, 1988 The wild relative of *Phaseolus vulgaris* in Middle America, pp. 163-184 in *Genetic resources of* Phaseolus *beans*, edited by P. GEPTS. Kluwer, Dordrecht, the Netherlands.

DELSENY, M., J. M. MCGRATH, P. THIS, A. M. CHEVRE and C. F. QUIROS, 1990 Ribosomal RNA

genes in diploid and amphidiploid *Brassica* and related species: organization, polymorphism, and evolution. Genome **33:** 733-744.

DICKSON, M. H., and R. PETZOLDT, 1988 Deleterious effects of white seed due to *p* gene in beans. J. Amer. Soc. Hort. Sci. **111-114:** 111-114.

DOEBLEY, J., 1989 Isozymic evidence and the evolution of crop plants, pp. 165-191 in *Isozymes in plant biology*, edited by D. E. SOLTIS and P. S. SOLTIS. Dioscorides, Portland, OR.

DOEBLEY, J., 1990 Molecular systematics of *Zea* (Gramineae). Maydica **35:** 143-150.

DOEBLEY, J., 1992 Molecular systematics and crop evolution, pp. 202-222 in *Molecular systematics of plants*, edited by P. S. SOLTIS, D. E. SOLTIS and J. J. DOYLE. Chapman Hall, New York.

DOEBLEY, J., W. RENFROE and A. BLANTON, 1987 Restriction site variation in the *Zea* chloroplast genome. Genetics **117:** 139-147.

DOEBLEY, J., and A. STEC, 1991 Genetic analysis of the morphological differences between maize and teosinte. Genetics **129:** 285-295.

DOEBLEY, J., A. STEC, J. WENDEL and M. EDWARDS, 1990 Genetic and morphological analysis of a maize-teosinte F_2 population:implications for the origin of maize. Proc. Natl. Acad. Sci. USA **87:** 9888-9892.

DOEBLEY, J. F., M. M. GOODMAN and C. W. STUBER, 1985 Isozyme variation in the races of maize from Mexico. Amer. J. Bot. **72:** 629-639.

DOEBLEY, J. F., J. D. WENDEL, J. S. C. SMITH, C. W. STUBER and M. M. GOODMAN, 1988 The origin of cornbelt maize: the isozyme evidence. Econ. Bot. **42:** 120-131.

DOYLE, J. J., and R. N. BEACHY, 1985 Ribosomal gene variation in soybean (*Glycine*) and its relatives. Theor. Appl. Genet. **70:** 369-376.

DUVICK, D. N., 1984 Genetic diversity in major farm crops on the farm and in reserve. Econ. Bot. **38:** 161-178.

DVORAK, J., 1988 Cytogenetical and molecular inferences about the evolution of wheat, pp. 187-192 in *7th Internat. Wheat Genetics Symp*, Vol. 1, edited by T. E. MILLER and R. M. D. KOEBNER. Institute Plant Sci. Res., Cambridge, UK.

DVORAK, J., and R. APPELS, 1982 Genome and nucleotide differentiation in genomes of polyploid *Triticum* species. Theor. Appl. Genet. **63:** 349-360.

DVORAK, J., P. E. MCGUIRE and B. CASSIDY, 1988 Apparent sources of the A genomes of wheat inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome **30:** 680-689.

DVORAK, J., and H.-B. ZHANG, 1990 Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. Proc. Natl. Acad. Sci. USA **87:** 9640-9644.

ENDO, T., and H. MORISHIMA, 1983 Rice, pp. 129-146 in *Isozymes in plant genetics and breeding*, Vol. B,

edited by S. D. TANKSLEY and T. J. ORTON. Elsevier, Amsterdam.

EVANS, A. M., 1976 Beans, pp. 168-172 in *Evolution of crop plants*, edited by N. W. SIMMONDS. Longman, London, UK.

FIGDORE, S., W. C. KENNARD, K. M. SONG, M. K. SLOCUM and T. C. OSBORN, 1988 Assessment of the degree of restriction fragment length polymorphism in *Brassica*. Theor. Appl. Genet. **75:** 833-840.

FURNIER, G., M. P. CUMMINGS and M. T. CLEGG, 1990 Evolution of the avocados as revealed by DNA restriction fragment variation. J. Hered. **81:** 183-188.

GALINAT, W. C., 1971 The origin of maize. Annu. Rev. Genet. **5:** 447-478.

GAYLER, K. R., and G. E. SYKES, 1985 Effect of nutritional stress on the storage proteins of soybeans. Plant Physiol. **78:** 582-585.

GEPTS, P., 1988 Phaseolin as an evolutionary marker, pp. 215-241 in *Genetic resources of* Phaseolus *beans*, edited by P. GEPTS. Kluwer, Dordrecht, the Netherlands.

GEPTS, P., 1990 Genetic diversity of seed storage proteins in plants, pp. 64-82 in *Plant population genetics, breeding, and genetic resources*, edited by A. H. D. BROWN, M. T. CLEGG, A. L. KAHLER and B. S. WEIR. Sinauer, Sunderland, MA.

GEPTS, P., and F. A. BLISS, 1985 F_1 hybrid weakness in the common bean: differential geographic origin suggests two gene pools in cultivated bean germplasm. J. Hered. **76:** 447-450.

GEPTS, P., and F. A. BLISS, 1986 Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ. Bot. **40:** 469-478.

GEPTS, P., and F. A. BLISS, 1988 Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability. II. Europe and Africa. Econ. Bot. **42:** 86-104.

GEPTS, P., and M. T. CLEGG, 1989 Genetic diversity in pearl millet (*Pennisetum glaucum* [L.] R.Br.) at the DNA sequence level. J. Hered. **80:** 203-208.

GEPTS, P., and D. G. DEBOUCK, 1991 Origin, domestication, and evolution of the common bean, *Phaseolus vulgaris*, pp. 7-53 in *Common beans: research for crop improvement*, edited by O. VOYSEST and A. VAN SCHOONHOVEN. CAB, Oxon, UK.

GEPTS, P., T. C. OSBORN, K. RASHKA and F. A. BLISS, 1986 Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. Econ. Bot. **40:** 451-468.

GEPTS, P., K. KMIECIK, P. PEREIRA and F. A. BLISS, 1988 Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability. I. The Americas. Econ. Bot. **42:** 73-85.

GHADERI, A., M. W. ADAMS and A. W. SAETTLER, 1982 Environmental repsonse patterns in commercial classes of common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. **63:** 17-22.

GLASZMANN, J. C., 1987 Isozymes and classification of Asian rice varieties. Theor. Appl. Genet. **74:** 21-30.

GOTTLIEB, L. D., 1984 Genetics and morphological evolution in plants. Amer. Nat. **123:** 681-709.

GRANER, A., H. SIEDLER, A. JAHOOR, R. G. HERRMANN and G. WENZEL, 1990 Assessment of the degree and the type of restriction fragment polymorphism in barley (*Hordeum vulgare*). Theor. Appl. Genet. **80:** 826-832.

GRANT, V., 1975 *Genetics of flowering plants*. Columbia Univ. Press, New York.

GRAUR, D., M. BOGHER and A. BRIEMAN, 1989 Restriction endonuclease profiles of mitochondrial DNA and the origin of the B genome of wheat, *Triticum aestivum*. Heredity **62:** 335-342.

GUO, M., D. A. LIGHTFOOT, M. C. MOK and D. W. S. MOK, 1991 Analyses of *Phaseolus vulgaris* L. and *P. coccineus* Lam. hybrids by RFLP: preferential transmission of *P. vulgaris* alleles. Theor. Appl. Genet. **81:** 703-709.

HALWARD, T. M., H. T. STALKER, E. A. LARUE and G. KOCHERT, 1991 Genetic variation detectable with molecular markers among unadapted germ-plasm resources of cultivated peanut and related wild species. Genome **34:** 1013-1020.

HAMMER, K., 1984 Das Domestikationssyndrom. Kulturpflanze **32:** 11-34.

HAMRICK, J. L., and M. J. W. GODT, 1990 Allozyme diversity in plant species, pp. 43-63 in *Plant population genetics, breeding, and genetic resources*, edited by A. H. D. BROWN, M. T. CLEGG, A. L. KAHLER and B. S. WEIR. Sinauer, Sunderland, MA.

HARLAN, J. R., 1971 Agricultural origins: cenetrs and non-centers. Science **174:** 468-474.

HARLAN, J. R., 1975 *Crops and man*. Amer. Soc. Agron., Madison, WI.

HARLAN, J. R., and J. M. J. DE WET, 1973 On the quality of evidence for origin and dispersal of cultivated plants. Curr. Anthrop. **14:** 51-62.

HAVEY, M. J., and F. J. MUEHLBAUER, 1989 Variability for restriction fragment lengths and phylogenies in lentil. Theor. Appl. Genet. **77:** 839-843.

HEISER, C. B., 1965 Cultivated plants and cultural diffusion in nuclear America. Amer. Anthropol. **67:** 930-949.

HELENTJARIS, T., G. KING, M. SLOCUM, C. SIEDENSTRANG and S. WEGMAN, 1985 Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Molec. Biol. **5:** 109-118.

HERNANDEZ XOLOCOTZI, E., S. MIRANDA COLIN and C. PRWYER, 1959 El origen de *Phaseolus coccineus* L. *darwinianus* Hernandez X. & Miranda C. subspecies nova. Rev. Soc. Mex. Hist. Nat. **20:** 99-121.

HEUN, M., A. E. KENNEDY, J. A. ANDERSON, N. L. V. LAPITAN, M. E. SORRELLS and S. D. TANKSLEY, 1991 Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*). Genome **34:** 437-447.

B2

HILLMAN, G. C., and M. S. DAVIES, 1990 Domestication rates in wild-type wheats and barley under primitive cultivation. Biol. J. Linn. Soc. **39:** 39-78.

HILU, K., 1983 The role of single-gene mutation in the evolution of flowering plants, pp. 97-128 in *Evolutionary Biology*, Vol. 16, edited by M. K. HECHT, B. WALLACE and G. T. PRANCE., .

HOFFMAN, D. L., D. E. SOLTIS, F. J. MUEHLBAUER and G. LADIZINSKY, 1986 Isozyme polymorphism in *Lens* (Leguminosae). Syst. Bot. **11:** 392-402.

HOFFMAN, L. M., and D. D. DONALDSON, 1985 Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. EMBO J. **4:** 883-889.

HOLWERDA, B. C., S. JANA and W. L. CROSBY, 1986 Chloroplast and mitochondrial DNA variation in *Hordeum vulgare* and *Hordeum spontaneum*. Genetics **114:** 1271-1291.

HOSAKA, K., and R. E. HANNEMAN, 1988a Origin of chloroplast DNA diversity in the Andean potatoes. Theor. Appl. Genet. **76:** 333-340.

HOSAKA, K., and R. E. HANNEMAN, 1988b The origin of the cultivated tetraploid potato based on chloroplast DNA. Theor. Appl. Genet. **76:** 172-176.

HOSAKA, K., S. F. KIANIAN, J. M. MCGRATH and C. F. QUIROS, 1990 Development and chromosomal localization of genome-specific DNA markers of *Brassica* and the evolution of amphidiploids and n = 9 diploid species. Genome **33:** 131-142.

ILTIS, H. H., 1983 From teosinte to maize: the catastrophic sexual transmutation. Science **222:** 886-894.

JAASKA, V., 1980 Electrophoretic survey of seedling esterases in wheats in relation to their phylogeny. Theor. Appl. Genet. **56:** 273-284.

JAASKA, V., 1981 Aspartate aminotransferase and alcohol dehydrogenase isoenzymes: intraspecific differentiation in *Aegilops tauschii* and the origin of the D genome polyploids in the wheat group. Plant Syst. Evol **137:** 259-273.

JANA, S., and L. PIETRZAK, 1988 Comparative assessment of genetic diversity in wild and primitive cultivated barley in a center of diversity. Genetics **119:** 981-990.

KAHLER, A. L., and R. W. ALLARD, 1981 Worldwide patterns of genetic variation among four esterase loci in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. **59:** 101-111.

KAPLAN, L., 1965 Archaeology and domestication in American *Phaseolus*. Econ. Bot. **19:** 358-368.

KEIM, P., R. C. SHOEMAKER and R. G. PALMER, 1989 Restriction fragment polymorphism diversity in soybean. Theor. Appl. Genet. **77:** 786-792.

KESSELI, R., O. OCHOA and R. W. MICHELMORE, 1991 Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. sativa*). Genome **34:** 430-436.

KHAIRALLAH, M. M., M. W. ADAMS and B. B. SEARS, 1990 Mitochondrial DNA polymorphisms of Malawian bean lines: further evidence for two major gene pools. Theor. Appl. Genet. **80:** 753-761.

KIMURA, M., 1983 *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, UK.

KNIGHT, R. L., 1948 The role of major genes in the evolution of economic characters. J. Genet. **48:** 370-387.

KOCHERT, G., T. HALWARD, W. D. BRANCH and C. E. SIMPSON, 1991 RFLP variability in peanut (*Arachis hypogea* L.) cultivars and wild species. Theor. Appl. Genet. **81:** 565-570.

KOENIG, R., and P. GEPTS, 1989a Segregation and linkage of genes for seed proteins, isozymes, and morphological traits in common bean (*Phaseolus vulgaris*). J. Hered. **80:** 455-459.

KOENIG, R., and P. GEPTS, 1989b Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. Theor. Appl. Genet. **78:** 809-817.

KOENIG, R., S. P. SINGH and P. GEPTS, 1990 Novel phaseolin types in wild and cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ. Bot. **44:** 50-60.

KOINANGE, E. M. K., and P. GEPTS, 1992 Hybrid weakness in wild *Phaseolus vulgaris*. J. Hered. : in press.

LADIZINSKY, G., 1985 Founder effect in crop-plant evolution. Econ. Bot. **39:** 191-198.

LADIZINSKY, G., D. BRAUN, D. GOSHEN and F. J. MUEHLBAUER, 1984 The biological species of the genus *Lens*. Bot. Gaz. **145:** 253-261.

LANDE, R., 1981 The minimum number of genes contributing to quantitative variation between and within populations. Genetics **99:** 541-553.

LANDE, R., 1983 The response to selection on major and minor mutations affecting a metrical trait. Heredity **50:** 47-65.

LANDRY, B. S., R. KESSELI, H. LEUNG and R. W. MICHELMORE, 1987 Comparison of restriction endonucleases and sources of probes for thier efficiecy in dtecting restriction fragment length polymorphisms in lettuce (*Lactuca sativa* L.). Theor. Appl. Genet. **74:** 646-653.

LYMAN, J. M., J. P. BAUDOIN and R. HIDALGO, 1985 Lima bean, pp. 477-519 in *Grain legume crops*, edited by R. H. SUMMERFIELD and E. H. ROBERTS. Collins, London.

LYNCH, T., A. GONZALEZ, J. TOHME and J. GARCIA, 1992 Variation in characters related to leaf photosynthesis in wild bean populations. Crop Sci. **32:** in press.

MACKIE, 1943 Origin, dispersal and variability of the lima bean, *Phaseolus lunatus*. Hilgardia **15:** 1-29.

MANGELSDORF, P. C., 1974 *Corn: its origin, evolution and improvement*. Harvard Univ. Press, Cambridge, MA.

MARCHAIS, L., and J. PERNES, 1985 Genetic divergence between wild and cultivated pearl millets (*Pennisetum typhoides*). I. Male sterility. Z. Pflanzenzüchtg. **95:** 103-112.

MARCHAIS, L., and S. TOSTAIN, 1985 Genetic divergence between wild and cultivated pearl millets (*Pennisetum typhoides*). II. Characters of domestication. Z. Pflanzenzüchtg. **95:** 245-261.

MARECHAL, R., J.-M. MASCHERPA and F. STAINIER, 1978 Etude taxonomique d'un groupe complexe d'espèces des genres *Phaseolus* et *Vigna* (Papilionaceae) sur la base de données morphologiques et polliniques, traitées par l'analyse informatique. Boissiera **28:** 1-273.

McCLEAN, P. E., and M. R. HANSON, 1986 Mitochondrial DNA sequence divergence among *Lycopersicon* and related *Solanum* species. Genetics **112:** 649-667.

McCOUCH, S. R., G. KOCHERT, Z. Y. YU, G. S. KHUSH, W. R. COFFMAN and S. D. TANKSLEY, 1988 Molecular mapping of rice chromosomes. Theor. Appl. Genet. **76:** 815-829.

McLEOD, M. J., S. I. GUTTMAN and W. H. ESHBAUGH, 1982 Early evoution of chili peppers (*Capsicum*). Econ. Bot. **36:** 361-368.

McLEOD, M. J., S. I. GUTTMAN, W. H. ESHBAUGH and R. E. RAYLE, 1983 An electrophoretic study of evolution in *Capsicum* (Solanaceae). Evolution **37:** 562-574.

MESSMER, M. M., A. E. MELCHINGER, M. LEE, W. L. WOODMAN, E. A. LEE and K. R. LAMKEY, 1991 Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. Theor. Appl. Genet. **83:** 97-107.

MILLER, J. C., and S. D. TANKSLEY, 1990 RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. **80:** 437-448.

MORENO, J., and M. J. CHRISPEELS, 1989 A lectin gene encodes the α-amylase inhibitor of the common bean. Proc. Natl. Acad. Sci. USA **86:** 7885-7889.

MOTTO, M., G. P. SORESSI and F. SALAMINI, 1978 Seed size inheritance in a cross between wild and cultivated common beans (*Phaseolus vulgaris* L.). Genetica **49:** 31-36.

NABHAN, G. P., 1985 Native crop diversity in Aridoamerica: conservation of regional gene pools. Econ. Bot. **39:** 387-399.

NEALE, D. B., M. A. SAGHAI-MAROOF, R. W. ALLARD, Q. ZHANG and R. A. JORGENSEN, 1986 Chloroplast DNA diversity in populations of wild and cultivated barley. Genetics **120:** 1105-1110.

NEVO, E., A. BEILES and D. ZOHARY, 1986 Genetic resources of wild barley in the Near East: structure, evolution and application in breeding. Biol. J. Linn. Soc. **27:** 355-380.

NEVO, E., D. ZOHARY, A. H. D. BROWN and M. HABER, 1979 Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum*, in Israel. Evolution **33:** 815-833.

NIENHUIS, J., and S. P. SINGH, 1988a Genetics of seed yield and its components in common bean (*Phaseolus vulgaris* L.) of Middle-American origin. I. General combining ability. Plant Breed. **101:** 143-154.

NIENHUIS, J., and S. P. SINGH, 1988b Genetics of seed yield and its components in common bean (*Phaseolus vulgaris* L.) of Middle American origin. II. Genetic variance, heritability, and expected response from selection. Plant Breed. **101:** 155-163.

NODARI, R. O., E. M. K. KOINANGE, J. D. KELLY and P. GEPTS, 1992 Towards an integrated linkage

map of common bean. I. Development of genomic DNA probes and levels of restriction fragment length polymorphism. Theor. Appl. Genet. : in press.

OGIHARA, Y., and K. TSUNEWAKI, 1988 Diversity and evolution of chloroplast DNA in *Triticum* and *Aegilops* as revealed by restriction fragment analysis. Theor. Appl. Genet. **76:** 321-332.

OKA, H. I., 1958 Intervarietal variation and classification of cultivated rice. Ind. J. Genet. Plant Breed. **18:** 79-89.

OSBORN, T. C., T. BLAKE, P. GEPTS and F. A. BLISS, 1986 Bean arcelin. 2. Genetic variation, inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris*. Theor. Appl. Genet. **71:** 847-855.

OSBORN, T. C., D. C. ALEXANDER, S. S. M. SUN, C. CARDONA and F. A. BLISS, 1988 Insecticidal activity and lectin homology of arcelin seed protein. Science **240:** 207-210.

PÄÄBO, S., 1989 Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc. Natl. Acad. Sci. USA **86:** 1939-1943.

PÄÄBO, S., R. HIGUCHI and A. C. WILSON, 1989 Ancient DNA and the polymerase chain reaction. J. Biol. Chem. **264:** 9709-9712.

PALMER, J. D., 1987 Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Amer. Nat. **130:** S6-S29.

PALMER, J. D., 1990 Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet. **6:** 115-120.

PALMER, J. D., 1992 Mitochondrial DNA in plant systematics: applications and limitations, pp. 36-49 in *Molecular systematics of plants*, edited by P. S. SOLTIS, D. E. SOLTIS and J. J. DOYLE. Chapman Hall, New York.

PALMER, J. D., R. A. JORGENSEN and W. F. THOMPSON, 1985 Chloroplast DNA variation and evolution in *Pisum*: patterns of change and phylogenetic analysis. Genetics **109:** 195-213.

PALMER, J. D., C. R. SHIELDS, D. B. COHEN and T. J. ORTON, 1983 Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theor. Appl. Genet. **65:** 181-189.

PERCY, R. G., and J. F. WENDEL, 1990 Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. Theor. Appl. Genet. **79:** 529-542.

PERNES, J., D. COMBES and J. M. LEBLANC, 1984 Le mil, pp. 159-197 in *Gestion des ressources génétiques des plantes*, Vol. 1: Monographies, edited by J. PERNES. Techniques et Documentation - Lavoisier, Paris.

PICKERSGILL, B., C. B. HEISER and J. MCNEILL, 1979 Numerical taxonomy studies on variation and domestication in some species of *Capsicum*, pp. 679-700 in *The biology and taxonomy of the Solanaceae*, edited by J. G. HAWKES, R. N. LESTER and A. D. SKELDING. Academic Press, New York.

PINKAS, R., D. ZAMIR and G. LADIZINSKY, 1985 Allozyme divergence and evolution in the genus *Lens*. Plant Syst. Evol **151:** 131-140.

PRATT, R. C., and G. P. NABHAN, 1988 Evolution and diversity of *Phaseolus acutifolius* genetic resources, pp. 409-440 in *Genetic resources of* Phaseolus *beans*, edited by P. GEPTS. Kluwer, Dordrecht, the Netherlands.

RICK, C. M., and J. F. FOBES, 1975 Allozyme variation in the cultivated tomato and closely related species. Bull. Torrey Bot. Club **102:** 376-384.

RINDOS, D., 1984 *The origins of agriculture*. Academic Press, Orlando, FL.

ROBERT, T., R. LESPINASSE, J. PERNES and A. SARR, 1991 Gametophytic competition as influencing gene flow between wild and cultivated forms of pearl millet (*Pennisetum typhoides*). Genome **34:** 195-200.

ROGSTAD, S. H., J. C. PATTON II and B. A. SCHAAL, 1988 M13 repeat probe detects DNA minisatellite-like sequences in gymnosperms and angiosperms. Proc. Natl. Acad. Sci. USA **85:** 9176-9178.

ROMERO-ANDREAS, J., B. S. YANDELL and F. A. BLISS, 1986 Bean arcelin. 1. Inheritance of a novel seed protein of *Phaseolus vulgaris* L. and its effect on seed composition. Theor. Appl. Genet. **72:** 123-128.

RÜDORF, W., 1959 Genetics of *Phaseolus aborigineus*. Proc. Xth Internat. Genet. Congr., Vol. 2: p. 243 (Abstract)

SAGHAI-MAROOF, M. A., K. M. SOLIMAN, R. A. JORGENSEN and R. W. ALLARD, 1984 Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. USA **81:** 8014-8018.

SARKAR, P., and G. L. STEBBINS, 1956 Morphological evidence concerning the origin of the B genome in wheat. Amer. J. Bot. **43:** 297-304.

SCHINKEL, C., and P. GEPTS, 1988 Phaseolin diversity in the tepary bean, *Phaseolus acutifolius* A. Gray. Plant Breed. **101:** 292-301.

SCHINKEL, C., and P. GEPTS, 1989 Allozyme variability in the tepary bean, *Phaseolus acutifolius* A. Gray. Plant Breed. **102:** 182-195.

SCHMIT, V., and D. G. DEBOUCK, 1991 Observations on the origin of *Phaseolus polyanthus* Greenman. Econ. Bot. **45:** 345-364.

SECOND, G., 1982 Origin of the genetic diversity of cultivated rice (*Oryza spp.*): study of the polymorphism scored at 40 enzyme loci. Jap. J. Genet. **57:** 25-57.

SHII, C. T., M. C. MOK, S. R. TEMPLE and D. W. S. MOK, 1980 Expression of developmental abnormalities in hybrids of *Phaseolus vulgaris* L. J. Hered. **71:** 218-222.

SILBERNAGEL, M. J., and R. M. HANNAN, 1988 Utilization of genetic resources in the development of commercial bean cultivars in the U.S.A, pp. 561-596 in *Genetic Resources of* Phaseolus *Beans*, edited by P. GEPTS. Kluwer, Dordrecht, the Netherlands.

SINGH, S. P., 1991 Bean genetics, pp. 199-286 in *Common beans: research for crop improvement*, edited by A. VAN SCHOONHOVEN and O. VOYSEST. C.A.B., Wallingford, Oxon, UK.

SINGH, S. P., and A. J. GUTIÉRREZ, 1984 Geographical distribution of the DL_1 and DL_2 genes causing hybrid dwarfism in *Phaseolus vulgaris* L., their association with seed size, and their significance to breeding. Euphytica **33:** 337-345.

SINGH, S. P., C. CAJIAO, J. A. GUTIÉRREZ, J. GARCIA, M. A. PASTOR-CORRALES and F. J. MORALES, 1989 Selection for seed yield in inter-gene pool crosses of common bean. Crop Sci. **29:** 1126-1131.

SINGH, S. P., J. A. GUTIÉRREZ, A. MOLINA, C. URREA and P. GEPTS, 1991a Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci. **31:** 23-29.

SINGH, S. P., R. NODARI and P. GEPTS, 1991b Genetic diversity in cultivated common bean. I. Allozymes. Crop Sci. **31:** 19-23.

SINGH, S.P., P. GEPTS, and D.G. DEBOUCK, 1991c Races of common bean (*Phaseolus vulgaris*, Fabaceae). Econ. Bot. 45:379-396.

SINGH, S. P., J. TERAN, A. MOLINA and J. A. GUTIÉRREZ, 1992 Combining ability for seed yield and its components in common bean of Andean origin. Crop Sci. **32:** 81-84.

SOKAL, R. R., 1991 The continental population structure of Europe. Annu. Rev. Anthrop. **20:** 119-140.

SOKAL, R. R., N. L. ODEN and C. WILSON, 1991 Genetic evidence for the spread of agriculture in Europe by demic diffusion. Nature **351:** 143-145.

SOLTIS, P. S., J. J. DOYLE and D. E. SOLTIS, 1992a Molecular data and polyploid evolution in plants, pp. 177-201 in *Molecular systematics of plants*, edited by P. S. SOLTIS, D. E. SOLTIS and J. J. DOYLE. Chapman Hall, New York.

SOLTIS, D. E., P. S. SOLTIS and B. G. MILLIGAN, 1992b Intraspecific chloroplast DNA variation: systematic and phylogenetic implications, pp. 117-150 in *Molecular systematic of plants*, edited by P. S. SOLTIS, D. E. SOLTIS and J. J. DOYLE. Chapman Hall, New York.

SONG, K. M., T. C. OSBORN and P. H. WILLIAMS, 1988a *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 1. Genome evolution of diploid and amphidiploid species. Theor. Appl. Genet. **75:** 784-794.

SONG, K. M., T. C. OSBORN and P. H. WILLIAMS, 1988b *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 2. Preliminary analysis of subspecies within *B. rapa* (syn. *campestris*) and *B. oleracea*. Theor. Appl. Genet. **76:** 593-600.

STOCKTON, T., G. SONNANTE and P. GEPTS, 1992 Detection of minisatellite sequences in *Phaseolus vulgaris*. Plant Molecular Biology Reporter **10:** 47-59.

U, N., 1935 Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Japan. J. Bot. **7:** 389-452.

VAN SCHOONHOVEN, A., and C. CARDONA, 1982 Low levels of resistance to the Mexican bean weevil in dry beans. J. Econ. Entom. **75:** 567-569.

WANG, Z. Y., G. SECOND and S. D. TANKSLEY, 1992 Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor. Appl. Genet. **83:** 565-581.

WEBSTER, B. D., S. P. LYNCH and C. L. TUCKER, 1979 A morphological study of the development of reproductive structures of *Phaseolus lunatus* L. J. Amer. Soc. Hort. Sci. **104:** 240-243.

WEBSTER, B. D., R. M. ROSS and M. C. SIGOURNEY, 1980 A morphological study of the development of reproductive structures of *Phaseolus coccineus*. J. Amer. Soc. Hort. Sci. **105:** 825-833.

WELLHAUSEN, E. J., L. M. ROBERTS and E. HERNANDEZ X., 1952 *Races of maize in Mexico*. Bussey Inst., Harvard Univ., Cambridge, MA.

WELLS, W. C., W. H. ISOM and J. G. WAINES, 1988 Outcrossing rates of six common bean lines. Crop Sci. **28:** 177-178.

WENDEL, J. F., 1989 New World tetraploid cottons contain Old World cytoplasm. Proc. Natl. Acad. Sci. USA **86:** 4132-4236.

WENDEL, J. F., P. D. OLSON and J. MCD. STEWART, 1989 Genetic diversity, introgression, and independent domestication of Old World cultivated cottons. Amer. J. Bot. **76:** 1795-1806.

WHITE, J. W., C. MONTES and L. Y. MENDOZA, 1992 Use of grafting to characterize and alleviate hybrid dwarfness in common bean. Euphytica : in press.

WILLIAMS, J. G. K., A. R. KUBELIK, K. J. LIVAK, J. A. RAFALSKI and S. V. TINGEY, 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Ac. Res. **18:** 6531-6535.

WOLFE, K. H., W.-H. LI and P. M. SHARP, 1987 Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84:** 9054-9058.

ZHANG, Q., M. A. SAGHAI MAROOF, T. Y. LU and B. Z. SHEN, 1992 Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. Theor. Appl. Genet. **83:** 495-499.

ZIMMER, E. A., E. R. JUPE and V. WALBOT, 1988 Ribosomal gene structure, variation, and inheritance in maize and its ancestors. Genetics **120:** 1125-1136.

ZURAWSKI, G., and M. T. CLEGG, 1987 Evolution of higher plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Annu. Rev. Plant Physiol. **38:** 391-418.

1 Table 1. Comparison of molecular electrophoretic markers in evolutionary, genetic, and breeding studies
2

	Allozymes	Seed Proteins	RFLPs	Minisatellite Sequences	RAPDs
Polymorphism	low	high	low-high	very high	low-high
Environmental stability	moderate	high	high	high	high
Number of loci	moderate (<20 loci)	low (<10 loci)	high	moderate	high
Molecular basis	simple	complex	intermediate	complex	complex
Practicality	quick, cheap	quick, cheap	slow, expensive	intermediate	quick, expensive

3
4
5
6
7
8
9
10

Table 2. Evidence supporting the existence of the Middle American and Andean gene pools in *Phaseolus vulgaris*

Evidence		Source
Biochemical markers	Phaseolin	Gepts et al. 1986; Gepts and Bliss 1986; Koenig et al. 1990
	Isozymes	Koenig and Gepts 1989b; Singh et al. 1991b
Molecular markers	Nuclear DNA: random *Pst*I genomic probes	Chase et al., 1991; Guo et al. 1991; Nodari et al. 1992
	Mitochondrial DNA	Khairallah et al. 1990
Morphological and agronomic traits		Kaplan 1965; Evans 1976, 1980; Gepts et al. 1986; Singh et al. 1991a
Reproductive isolation	F_1 hybrid lethality	Shii et al. 1980; Singh and Gutiérrez 1984; Gepts and Bliss 1985; White et al. 1992; Koinange and Gepts 1992
	Chlorophyll variegation	Coyne 1966
	Segregation distortion	Koenig and Gepts 1989a
Yield	Genotype x environment interactions	Ghaderi et al. 1982

141

1 General combining ability Singh et al. 1992

2

3

142

Table 3. Comparison of RFLP levels in common bean and other plants (from Nodari et al. 1992)

Crop	Genotypes	Type of clone	No. of enzymes	RFLP (%)	Source
Arabidopsis	3 strains	Genomic: *Eco*RI	3	46	Chang et al. 1988
Barley	48 cultivars	Genomic: *Pst*I	3	43	Graner et al. 1990
	4 cultivars	Genomic: *Pst*I	5	54	Heun et al. 1991
Brassica	3 species (37 cultivars or genetic stocks)	Genomic: *Pst*I	1	95 between species 80 between subsp. 70 within subsp.	Figdore et al. 1988
Common Bean	7 cultivars + 1 wild	Genomic: *Pst*I	3	89 between subsp. 62 within Middle American 49 within Andean	Nodari et al., 1992
Lentil	2 cultivars +	Genomic: *Pst*I, *Eco*RI	4	38	Havey and Muehlbauer 1989
	2 wild	cDNA	4	64	
Lettuce	4 cultivars	Genomic: *Mbo*I	9	13	Landry et al. 1987
		cDNA	9	25	
Maize	3 cultivars	cDNA	3	83	Helentjaris et al. 1985
Potato	*phureja* x (*tuberosum* x *chacoense*)	Genomic: *Pst*I, *Eco*RI	11	60	Bonierbale et al. 1988
			11	51	
Rice	indica x javanica	Genomic: *Pst*I	11	78	McCouch et al. 1988
Soybean	5 cultivars	Genomic: *Sau*3AI,*Eco*RI	13	20	Apuya et al. 1988
	58 wild + cultivated	Genomic: *Sau*3Ai	4	40-50	Keim et al. 1989
Tomato	2 cultivars	cDNA	3	14	Helentjaris et al. 1985

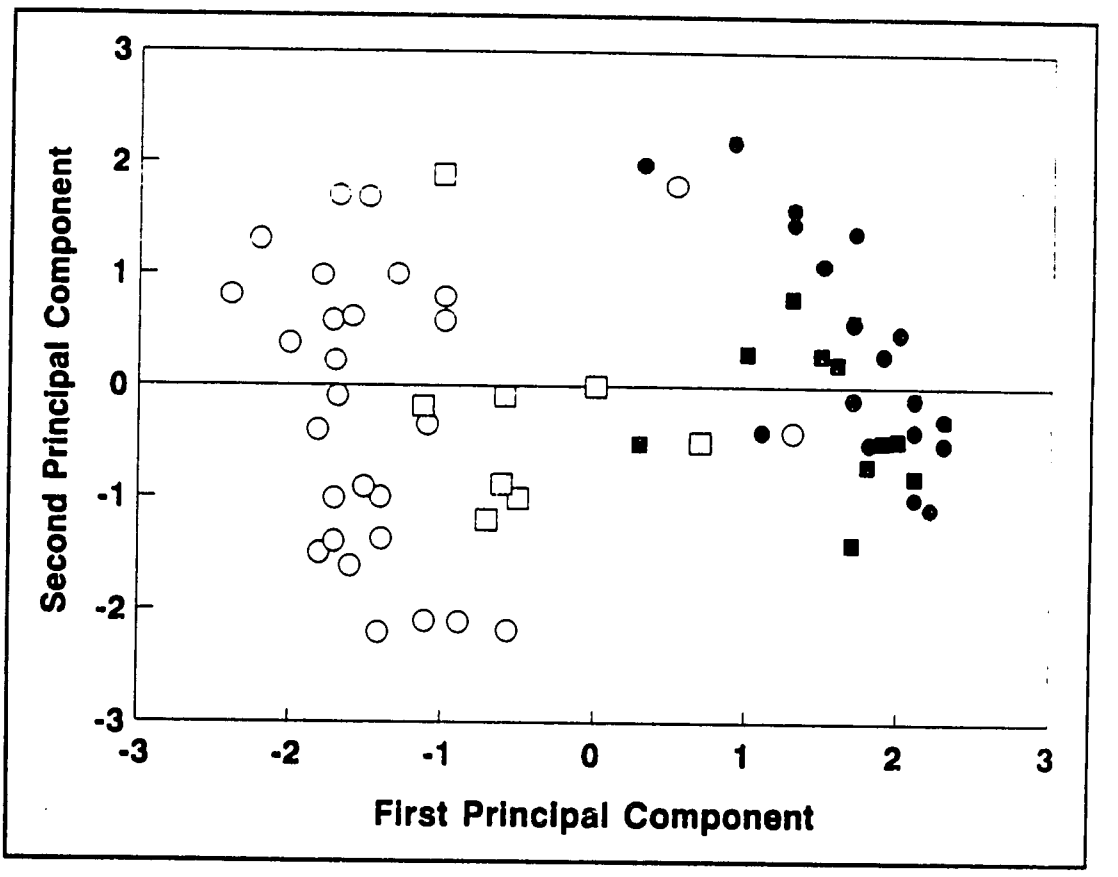
esculentum-cheesmannii	cDNA	3	68	"
esculentum-pennellii	cDNA	3	100	"

Table 4. Probability of survival of mutant genes determined according to the branching process method (modified from Crow and Kimura 1970)

	Selective Advantage			
Generation	-10%	0 %	10 %	50 %
0	1.00	1.00	1.00	1.00
1	0.59	0.63	0.67	0.78
10	0.08	0.16	0.25	0.61
100	<0.001	0.02	0.18	0.58
Infinite	0.00	0.00	0.18	0.58

Figure caption

Figure 1. Principal component analysis of RFLP banding pattern frequencies in *Phaseolus vulgaris*. Genomic DNA of 23 wild and 62 cultivated accessions was digested with *Eco*RV and, after Southern transfer, hybridized with 11 *P. vulgaris* nuclear DNA clones. Open symbols: Andean accessions; closed symbols: Middle American accessions. Squares: wild-growing; circles: cultivated (V. Becerra and P. Gepts, unpubl. results).


Second Principal Component
First Principal Component
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3

Theor Appl Genet (1992) ■: ■ - ■

© Springer-Verlag 1992

Towards an integrated linkage map of common bean 2. Development of an RFLP-based linkage map

R. O. Nodari [1,*], S. M. Tsai [1,**], R. L. Gilbertson [2], and P. Gepts [1]

[1] Department of Agronomy and Range Science, University of California, Davis, CA 95616-8515, USA
[2] Department of Plant Pathology, University of California, Davis, CA 95616-8680, USA

Received December 1, 1991; Accepted May 5, 1992
Communicated by A. L. Kahler

Summary. A restriction fragment length polymorphism (RFLP)-based linkage map for common bean (*Phaseolus vulgaris* L.) covering 827 centiMorgans (cM) was developed based on a F_2 mapping population derived from a cross between BAT93 and Jalo EEP558. The parental genotypes were chosen because they exhibited differences in evolutionary origin, allozymes, phaseolin type, and for several agronomic traits. The segregation of 152 markers was analyzed, including 115 RFLP loci, 7 isozyme loci, 8 random amplified polymorphic DNA (RAPD) marker loci, and 19 loci corresponding to 15 clones of known genes, 1 virus resistance gene, 1 flower color gene, and 1 seed color pattern gene. Using MAPMAKER and LINKAGE-1, we were able to assign 143 markers to 15 linkage groups, whereas 9 markers remained unassigned. The average interval between markers was 6.5 cM; only one interval was larger than 30 cM. A small fraction (9%) of the markers deviated significantly from the expected Mendelian ratios (1:2:1 or 3:1) and mapped into four clusters. Probes of known genes belonged to three categories: seed proteins, pathogen response genes, and *Rhizobium* response genes. Within each category, sequences homologous to the various probes were unlinked. The *I* gene for bean common mosaic virus resistance is the first disease resistance gene to be located on the common bean genetic linkage map.

Key words: RFLP – RAPD – Linkage map – Bean Common Mosaic Virus resistance – Segregation distortion

* *Permanent address:* Department of Plant Technology, Universidade Federal de Santa Catarina, Caixa Postal 476, 88049-Florianopolis, S.C., Brazil
** *Permanent address:* CENA, Universidade de São Paulo, Caixa Postal 96, 13400-Piracicaba, SP, Brazil
Correspondence to: P. Gepts

Introduction

The current linkage map of common bean (*Phaseolus vulgaris*; 2n = 2x = 22) is poorly developed. It consists of a small number of linkage groups that include genes controlling mostly morphological and pigmentation traits such as flower and seed color or color patterning (reviewed by Bassett 1991; Gepts 1988). Many of these traits are subject to epistatic, pleiotropic, and environmental effects, and with the exception of the locus for phaseolin seed protein and loci for ribosomal RNA, none has been assigned to a particular chromosome (Schumann et al. 1990; Durante et al. 1977).

Restriction fragment length polymorphisms (RFLPs) can be used as genetic markers to construct a saturated linkage map (Botstein et al. 1980). In the past 5 years, RFLP linkage maps have been developed for *Arabidopsis thaliana* (Reiter et al. 1992) and several crops such as barley (Heun et al. 1991), *Brassica napus* (Landry et al. 1991), *Brassica rapa* (Song et al. 1991), lettuce (Kesseli et al. 1990), maize (Coe et al. 1990), potato (Bonierbale et al. 1988; Gebhardt et al. 1989), rice (McCouch et al. 1988), soybean (Diers et al. 1992), and tomato (Paterson et al. 1991). The advantages of RFLP markers over morphological and biochemical markers include their relatively high number, absence of pleiotropic and epistatic effects, and higher frequency of polymorphism.

Preliminary studies with a set of approximately 60 random *Pst*I clones hybridized to genomic DNA digested with three restriction enzymes revealed high levels of RFLP in common bean. These levels reach 80–90% between the Middle American and Andean gene pools and 50–60% within these gene pools (Nodari et al. 1992). In the work presented here we have used RFLPs to develop a linkage map of common bean for use in evolutionary, genetic, and breeding studies. We also identified linkages

TAG
Ms.: 480
K. Triltsch, Würzburg
(provisorische Seitenzahlen)

Springer-Verlag, Berlin
S.: 1–8
1. Korr.: 21.9.92

between RFLP loci and other molecular markers including RAPDs, isozymes, and seed proteins. We mapped sequences involved in responses to pathogens and *Rhizobium*. Furthermore, we mapped the *I* gene for bean common mosaic virus resistance.

Material and methods

Plant material

The two parents of the F_2 mapping population were selected because they had divergent evolutionary origins and exhibited contrasting disease responses. BAT93 is a breeding line developed at the Centro Internacional de Agricultura Tropical (CIAT, Cali, Colombia) and was derived from a double cross involving four Middle American genotypes (Veranic 2, PI 207262, Jamapa, and Great Northern Tara). Jalo EEP558 was a selection from the Andean landrace Jalo obtained at the Estação Experimental de Pato de Minas (Minas Gerais, Brazil). Phaseolin and isozyme profiles of BAT93 and Jalo EEP558 confirmed their Middle American and Andean evolutionary origins, respectively (Koenig et al. 1989 b; Singh et al. 1991 b). BAT93 is resistant to bean common mosaic virus (BCMV), rust (*Uromyces phaseoli*), common bacterial blight (*Xanthomonas phaseoli*), and anthracnose (*Colletotrichum lindemuthianum*), but is susceptible to angular leaf spot (*Phaeoisariopsis angularis*). Jalo EEP558 exhibits opposite reactions to the same pathogens (Schwartz et al. 1982; S. Singh, personal communication; R. Stavely, personal communication). Jalo EEP558 develops a large number of nodules after inoculation with *Rhizobium tropici*, whereas BAT93 develops few nodules (Tsai et al. unpublished results). The two parents also differ in morpho-agronomic traits such as seed size, flower color, presence of the seed corona, and length of the fifth internode (Singh et al. 1991 a). An F_2 mapping population of 75 greenhouse-grown individuals was obtained after selfing of the F_1 generation under insect-free conditions.

Markers: source and nomenclature

Genomic clones and probes of known genes

Segregations for RFLPs detected by genomic clones and probes of known genes were determined in the F_2 population of 75 individuals. Two genomic libraries, *Eco*RI-*Bam*HI (clones numbered from GUC001 to GUC999) and *Pst*I (clones numbered from GUC1001 to GUC1862), were constructed as described in Nodari et al. (1992). Single-copy clones were selected after hybridization of the insert with total radiolabeled bean genomic DNA as described in Nodari et al. (1992). The following sequences were also used for RFLP mapping: nuclear ribosomal RNA gene (pHA2; Jorgensen et al. 1987), α-amylase inhibitor (pHU20; Moreno and Chrispeels 1989), seed lectin or phytohemagglutinin (pMcPHA; M. Chrispeels, personal communication), phaseolin (pPhasMC31; Slightom et al. 1983), cellulase (pBAC10; Tucker and Milligan 1991), chitinase (pCH18, Broglie et al. 1986), glucanase (pG101; Edington et al. 1991), chalcone isomerase (pCHI1; Mehdy and Lamb 1987), chalcone synthase (pCHS1; Ryder et al. 1984), phenylalanine ammonia-lyase (pPAL5; Edwards et al. 1985), pathogenesis-related proteins 1 and 2 (pEIT23 and pEIT14a; Walter et al. 1990), glutamine synthetase (pCPGS2; DPS Verma, personal communication), sucrose synthetase (nodulin-100; Fuller et al. 1983; Thummler and Verma 1987), and uricase (pNOD35; Nguyen et al. 1985). When the same probe detected ~~one~~ more than one polymorphic band, fragments of decreasing size were labeled alphabetically with an additional lower case letter. The same rule was followed for RAPD markers.

more

Random amplified polymorphic DNA (RAPD)

DNA sequences from the parents and 54 randomly chosen plants of the F_2 mapping population were amplified with randomn 10-mer primers (Kit A, Operon, Alameda, CA) in a thermal cycler (Ericomp). The cycling parameters (40 cycles) for denaturation, annealing and extension were: 1 min at 94 °C, 1 min at 35 °C, and 2 min at 72 °C, respectively. One last cycle of 10 min at 72 °C was performed to complete the reaction. Amplification was carried out with 25 ng of total genomic DNA in a 25 µl solution containing 1 × buffer (10 × = 100 m*M* TRIS-HCl pH 8.3, 500 m*M* KCl, 15 m*M* $MgCl_2$, 0.01% gelatin), 0.1 m*M* of each dNTP, 0.45 m*M* $MgCl_2$, 0.4 µ*M* primer, and 1 U *Taq* DNA polymerase (Promega). One-half of the volume was electrophoresed on a 1.5% agarose gel (containing 0.5 µg/ml ethidium bromide), and the RAPD polymorphisms were scored as the presence or absence of a specific band. Each amplification run included two negative controls (without DNA, without polymerase) and the two parents to verify the repeatability of the experiment. The scored bands were labeled with an R (for RAPD), followed by an O (Operon), an A (Kit A from Operon) and by the number of the primer in the kit. When the same probe detected more than 1 locus, fragments of decreasing size were labeled alphabetically with an additional lower case letter.

Phaseolin and isozymes

Polyacrylamide and starch gel electrophoresis assays were performed as described by Koenig and Gepts (1989 a, b). Seeds from 70 individuals of the mapping population were analyzed for seed-protein phaseolin type. Leaf and root tissues from the same individuals were scored for seven polymorphic isozyme systems: aconitase (ACO, E.C. 4.2.1.3), diaphorase (DIAP, E.C. 1.6.4.3), leucine aminopeptidase (LAP, E.C. 3.4.11.1), malic enzyme (ME, E.C. 1.1.1.40), methyl-umbelliferyl esterase (MUE), the small subunit of ribulose bisphosphate carboxylase (RBCS, E.C. 4.1.1.39), and shikimate dehydrogenase (SKDH, E.C. 1.1.1.25). The stain recipes and genetic control for all enzymes except aconitase have been reported previously (Koenig and Gepts 1989 a, b; Garrido et al. 1991). Aconitase was stained as described by Gepts et al. (1992) and showed two bands of activity. The most cathodal of these was polymorphic, with BAT93 and Jalo EEP558 carrying the *Aco*-2^{100} and *Aco*-2^{102} alleles, respectively.

Morphological and agronomic traits

Three morpho-agronomic traits were evaluated: flower color, corona, and reaction to bean common mosaic virus (BCMV). Flower color (~~pink~~ pink versus white) was evaluated in 75 F_2 plants, whereas the presence or absence of corona (*Cor:* Lamprecht 1961) was evaluated in F_3 seeds of the same 75 F_2 plants. The presence (hypersensitivity as revealed by "black root" necrosis symptoms in leaves and or whole plants) or absence (typical common mosaic symptoms) of the dominant *I* allele, which confers resistance to non-necrosis-inducing strains of bean common mosaic potyvirus (BCMV; Drijfhout et al. 1978), was tested in 10 individuals of 70 F_3 families by mechanically inoculating primary leaves with NL-3, a necrosis-inducing strain of BCMV (Morales and Castaño 1987). The inoculation of 10 individuals per F_3 family allowed us to distinguish at the $P = 0.95$ level between F_2 plants homozygous or heterozygous for the dominant allele at the *I* locus, ~~which confers resistance to BCMV~~.

pink

149

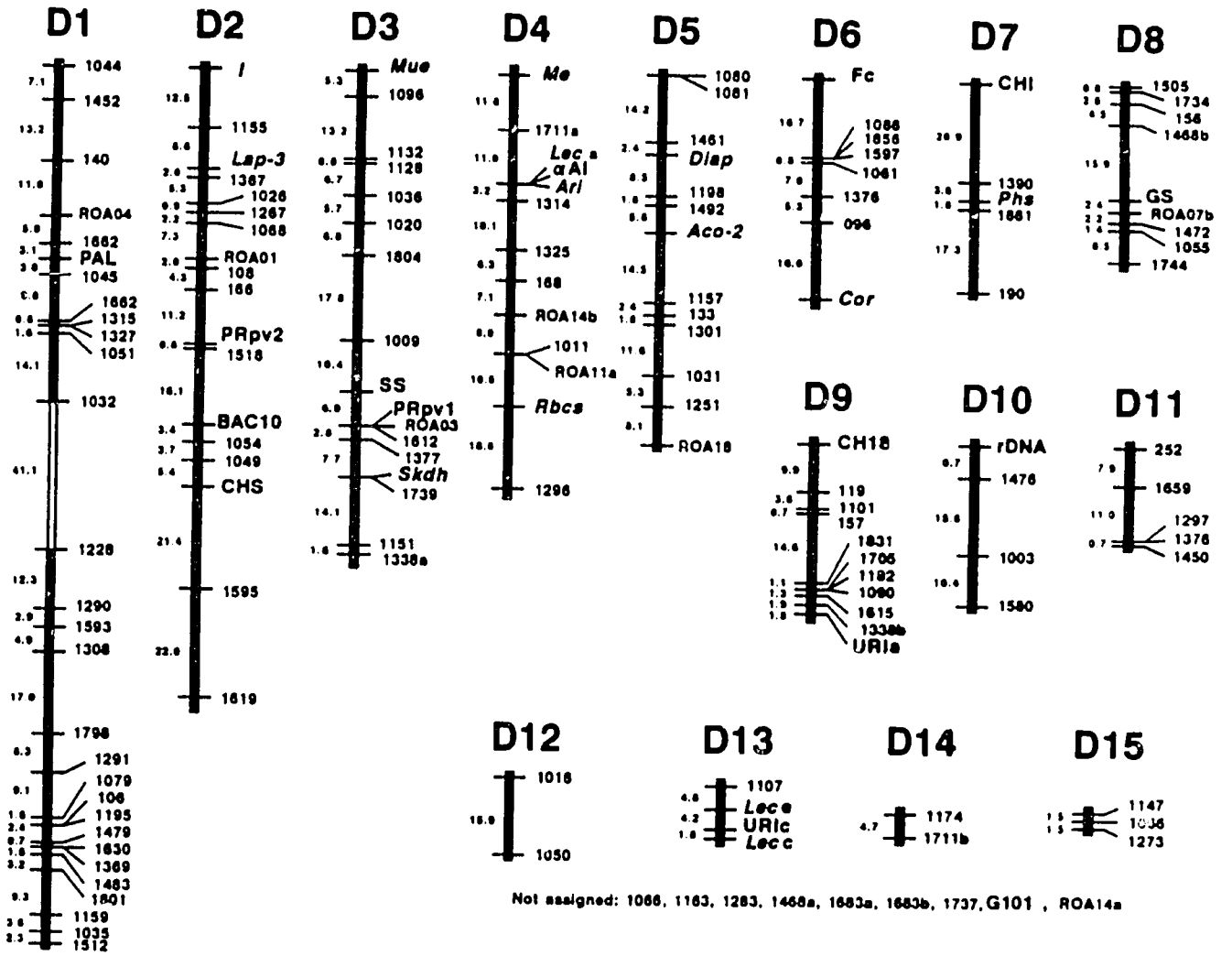


Fig. 1. RFLP-based linkage map of common bean. Linkage group numbers are indicated on *top*. Kosambi map distance are indicated at the *left side* of each interval between two markers. Genomic markers are numbered from *001* to *999* (from *Eco*RI-*Bam*HI library) and from *1001* to *1862* (from *Pst*I library) (the GUC prefix was removed for the sake of simplicity). *Aco-2*, *Diap*, *Lap-3*, *Me*, *Mue*, *Rbcs*, and *Skdh* are isozyme loci. RAPD marker loci are designated by *ROA* followed by a *number* (see text for explanations). Morpho-agronomic traits are *C*, *Cor*, and *I* (see text for explanations). The following markers represent sequences coding for products of known function: *Arl*, *αAl*, BAC10, CHI, CHS, CH18, GS, *Lec*, PAL, *Phs*, PRPv1, PRPv2, rDNA, SS, and URI (see text for explanations). The *open bar* in linkage group D1 represents an interval with a LOD score above 3.0 but with a recombination frequency above 30%

DNA extraction, digestion, and electrophoresis

genomic

Total genomic DNA from frozen leaf tissue (5–7 g) that had been harvested prior to flowering was extracted as described by Nodari et al. (1992). Genomic DNA samples (5–10 μg) from parents and F_1 and F_2 plants were digested with either *Eco*RI, *Eco*RV, *Hae*III, or *Hind*III restriction enzymes because these enzymes revealed the highest levels of polymorphism between BAT93 and Jalo EEP558 (Nodari et al. 1992). Restriction digestions (4–6 U enzyme/μg of genomic DNA) were performed according to the manufacturers' recommendations (Bethesda Research Laboratories and New England Biolabs) for 4–6 h at 37 °C. The digested DNA was electrophoresed in 0.8% agarose gels (prepared with TAE: 40 m*M* TRIS-acetate pH 7.4, 1 m*M* EDTA) for 16–20 h (1 V/cm of gel) in TAE running buffer.

Southern blotting, hybridization, and autoradiography

Southern hybridization was performed according to the Zetabind protocol (AMF-CUNO, Meriden, Conn.). Approximately 30–40 ng of the insert of each clone, except pHA2, were radiolabeled with α-[^{32}P]dCTP (Amersham) by the random primer method (Feinberg and Vogelstein 1984). The pHA2 plasmid was radiolabeled with the same isotope by nick translation (Rigby et al. 1977). Hybridizations were performed at 42 °C in 50% formamide. Washes after hybridization were performed according to recommendations of the manufacturer (2 × SSC, 0.1% SDS at room temperature; 0.1 × SSC, 0.1% SDS at room temperature and twice at 60 °C for 30 min each). Kodak X-Omat X-ray film was exposed to the membranes for a few hours to a week depending on the intensity of the signal.

Linkage analysis

Linkage analysis was performed with MAPMAKER (Lander et al. 1987) and LINKAGE-1 (Suiter et al. 1983). To identify the linkage groups, pairwise comparisons and grouping of markers were performed with MAPMAKER under the following conditions: (1) recombination frequency below 30% and (2) LOD

3.0 score equal to or above 6. To establish the most likely order of markers within each group, the "orders" command was used with the above-mentioned linkage criteria and with a three-point exclusion threshold of −2.0 LOD score units. Additional markers were located in these groups using the "try" command with the aforementioned linkage criteria. Kosambi map distances among linked markers were used to generate the map (Fig. 1). Pairwise comparisons to confirm detected linkage relationships between markers and goodness-of-fit tests to detect distorted segregation ratios were performed with LINKAGE-1.

Results and discussion

Segregation of markers

The majority of the random *Pst*I and *Eco*RI-*Bam*HI genomic clones (75%) hybridized to a single band and, with few exceptions, segregated in a co-dominant manner. Four clones – GUC1338, GUC1468, GUC1683, and GUC1711 – detected two bands, which, in all four cases, segregated independently and represent duplications. The remaining clones displayed a hybridization pattern with more than two bands. However, in all cases just one segregating locus could be scored. Probes representing cloned genes in general gave more complex hybridization patterns than random genomic clones. With the exception of BAC10 (cellulase), G101 (glucanase), and CHI (chalcone isomerase), which hybridized to a single band, all other gene clones hybridized to multiple bands. For example, seven and six segregating bands were observed after hybridization with CHS (chalcone synthase) and *Lec* (seed lectin) sequences, respectively. In the case of chalcone synthase (CHS), all segregating bands mapped to the same locus. For seed lectin (*Lec*), four bands (*Leca*, *Lecb*, *Lecd*, and *Lecf*) co-segregated at one locus in linkage group D4 and the other two (*Lecc* and *Lece*) mapped in linkage group D13 (see below). The presence of these multiple bands points to the existence of multigene families. Evidence of Ryder et al. (1987) suggests that CHS is represented by a family of six to eight genes, some of which are tightly linked. Hoffman and Donaldson (1985) identified two lectin genes 4 kb apart and tightly to the the loci coding for the homologous seed proteins arcelin and α-amylase inhibitor (Osborn et al. 1986; Moreno and Chrispeels 1989). RAPD markers were scored for the presence of absence of specific bands of high intensity that distinguished the two parents; hence, these markers segregated as dominant markers. Isozymes and phaseolin seed protein segregated as co-dominant markers as reported previously (Koenig and Gepts 1989a; Garrido et al. 1991; Gepts et al. 1992).

The inheritance pattern of the 152 markers analyzed in the segregating population of 75 individuals followed the expected F_2 co-dominant (1:2:1) or dominant (3:1) Mendelian ratios in 91% of the cases. A small fraction (9%) of the markers deviated significantly ($P<0.05$) from the ratios mentioned above. A similar proportion of distorted segregation (10%) was recently reported in lettuce (Kesseli et al. 1990) and in barley (Heun et al. 1991). Higher proportions of distorted markers have been detected in rice (18.8%; McCouch et al. 1988) and potato (25.5%; Gebhardt et al. 1988). The proportions of BAT93 and Jalo EEP558 alleles were almost identical (0.492:0.508). The 14 markers with distorted segregation were clustered in four regions of the bean genome. Seven of them mapped adjacent to the PAL locus in linkage group D1 (GUC140, ROA04, GUC1662, GUC1045, GUC1862, GUC1315, and GUC1327); 1 marker was located in linkage group D2 (BAC 10); 2 markers were placed 2.4 cM apart in linkage group D8 (GS and ROA07b), and the other 4 constituted linkage group D13 (*Lecc*, *Lece*, URIc, and GUC1107). The clustering of markers with distorted segregation has been observed previously in barley (Heun et al. 1991) and potato (Bonierbale et al. 1988). Possible causes for these segregation distortions include small population size and or, genetic factors affecting the transmission of the genes. Of the 7 markers in linkage group D16 showed an excess of the Jalo EEP558 alleles. The 7th marker showed an excess of heterozygotes. The 4 markers in linkage group D13 showed a distortion favoring the BAT93 alleles. The number of markers and the consistency in the direction of the deviation among the markers in these two regions suggest a possible genetic origin for the deviation on the linkage groups D1 and D13.

Linkage mapping of markers

Of the 152 markers 143 (94%) could be assigned into 15 linkage groups, comprising a total of 827 cM (Fig. 1 and Table 1). These linkage groups are temporarily labeled D (as in Davis) followed by a number pending further genetic and cytogenetic confirmation. The number (and in parentheses the percentage) of markers per linkage group ranged from 2 (1.4%) to 29 (20.6%). Almost the same order of magnitude was found for variation in length of the linkage groups (3.0–191.5 cM). Among linkage groups the average distance between 2 markers varied from 1.5 cM to 15.9 cM. The average distance between pairs of markers among all linkage groups was 6.5 cM. Only one interval between 2 markers was larger than 30 cM, and it was located in linkage group D1; however, the LOD score for this interval was above the threshold of 3.0. The majority of the intervals (74%) were smaller than 10 cM.

The 7 segregating isozyme loci were distributed in 4 linkage groups. *Me* (malic enzyme) and *Rbcs* (locus for the small subunit of ribulose bisphosphate carboxylase) had been found previously to be linked (Weeden 1984; Koenig and Gepts 1989a). We confirmed the linkage

fraction

151

Table 1. Size distribution of linkage groups in the common bean RFLP map

Linkage group	Markers		Length		Average distance (cM)
	Number	% of total number or probes	cM	% of total	
D1	29	20.3	191.5	23.6	6.6
D2	18	12.6	129.1	15.6	7.6
D3	17	11.9	99.9	12.1	6.2
D4	13	9.1	87.0	10.5	7.2
D5	13	9.1	76.3	9.2	6.4
D6	8	5.6	45.9	5.6	6.6
D7	5	3.5	43.5	5.3	10.9
D8	9	6.3	36.4	4.4	4.6
D9	11	7.7	35.0	4.2	3.5
D10	4	2.8	32.9	4.0	11.0
D11	5	3.5	19.7	2.4	4.9
D12	2	1.4	15.9	1.9	15.9
D13	4	2.8	10.8	1.3	3.6
D14	2	1.4	4.7	0.6	4.7
D15	3	2.1	3.0	0.4	1.5
Total	143	100	826.6	100	
Average					6.5

between *Me* and *Rbcs* by identifying several intermediate markers in linkage group D4, including the tightly linked seed lectin (*Lec*) and the α-amylase inhibitor (*αAI*) loci. Although the arcelin locus was not mapped here, previous studies by Osborn et al. (1986) showed that *Lec* and *Arl* were also tightly linked. Two other isozyme loci, *Aco-2* and *Diap-1*, were located 16 cM apart in linkage group D5, confirming the observations of Vallejos and Chase (1991). The remaining two allozyme loci, *Skdh* and *Muel* mapped to the same linkage group D3, although they were separated by a distance of more than 50 cM.

Mue,

Cloned gene sequences used for RFLP mapping belonged to three broad categories. Disease response genes included chitinase (CH18), glucanase (G101), chalcone synthase (CHS), chalcone isomerase (CHI), phenylalanine ammonialyase (PAL), and pathogenesis-related proteins 1 and 2 (PRPv1 and PRPv2). *Rhizobium* nodulation response genes consisted of glutamine synthetase (GS), sucrose synthetase (SS), and uricase (URI). Seed proteins included phaseolin (*Phs*), seed lectin or phytohemagglutinin (*Lec*), and α-amylase inhibitor (*αAI*). Within each of these three categories, sequences homologous to the various probes were unlinked (Fig. 1).

We were able to map one locus for the ribosomal RNA genes located in linkage group D10. However, *in situ* hybridizations revealed at least 3 loci for rRNA genes (Durante et al. 1977). This discrepancy can be attributed to the lower level of polymorphism associated with rDNA sequences in population BAT93 × Jalo EEP558 compared to that of other sequences. Of the four restriction enzymes (*Eco*RI, *Eco*RV, *Hae*III, and *Hin*dIII) used in this study, only *Eco*RV revealed a polymorphism for rRNA genes between the two parents. The use of additional restriction enzymes may identify additional polymorphisms for the rDNA sequences and should allow us to map additional rRNA loci.

Eight genomic clones and 3 RAPD marker remained unassigned to linkage groups. A small proportion of loci have been reported as unassigned for other RFLP maps [e.g., 10% in potato (Gebhardt et al. 1988) and 13% in *Brassica napus* (Landry et al. 1991)]. This observation, together with the discrepancy between the haploid chromosome number (n = 11) and the current number of linkage groups, suggest that additional markers should be mapped to fill the gaps between some linkage groups and between linkage groups and currently unassigned markers. Genetic walking experiments based on bulked segregant analysis of RAPD markers as described by Michelmore et al. (1991) should facilitate this objective.

Segregation and mapping of morpho-agronomic traits

The *Cor* (corona) gene induces a colored ring on the outside of the hilum ring, which surrounds the seed hilum of the seed (von Lamprecht 1934; Leakey 1988). Jalo EEP558 and BAT93 carry the dominant and recessive *Cor* alleles, respectively, of this gene. Presence or absence of the corona segregated according to a 3:1 ratio in the F_2 generation ($\chi^2 = 3.74$; $P = 0.053$). Flower color (provisionally labeled Fc) also segregated according to a 3:1 ratio ($\chi^2 = 1.69$; $P = 0.19$) and appeared to be linked to the *Cor* locus in linkage group D6 (Fig. 1). The gene actually involved in flower color determination in this cross is not known. The *P*, *T*, and *V* loci (Urakken 1970; Leakey 1988) have been shown to control flower color. The *P* and *T* genes can be excluded because they also induce a complete or partial absence, respectively, of pigments in the seeds, which was not observed in this population. The *V* locus has, so far, not been shown to be linked to the *Cor* locus. On the other hand, the *C* locus has been shown to be linked to *Cor* (Lamprecht 1961). It is possible that flower color in this cross is controlled by the *C* pigmentation locus because the mapping distance of 45 cM between *Cor* and the flower color locus identified here corresponds to a recombination value of 25% that is very close to the recombination value of 23% detected previously by Lamprecht (1961). Test crosses are needed to identify the gene responsible for flower color segregation.

The *I* locus conditioning dominant resistance to the non-necrosis-inducing strains of BCMV and hypersensitivity to the necrosis-inducing strains (Drijfhout et al.

158

Table 2. Comparison of RFLP maps of common bean and other species

Species	C (pg haploid genome)[a]	Number of markers	Number of linkage groups haploid chromosome number	Length (cM)	Average distance (cM)	Amount of DNA map unit (kbp/cM)[b]	Reference
Arabidopsis thaliana	0.15	320	5/5	630	2	230	Reiter et al. (1992)
Barley	5.05	157	7/7	1,096	7.3	4,843	Heun et al. (1991)
Brassica napus	1.20	120	19/19	1,413	14.0	820	Landry et al. (1991)
Brassica rapa	0.53	280	10/10	1,850	6.9	276	Song et al. (1991)
Common bean	0.65	143	15/11	827	6.5	758	Present results
Lettuce	2.73	147	16/9	1,404	10.7	1,876	Kesseli et al. (1990)
Maize	~2.8	261	10/10	1,776	7.1	1,521	Coe et al. (1990)
Potato	~1.8	141	12/12	690	5.3	2,517	Gebhardt et al. (1989)
Potato	~1.8	134	12/12	606	5.0	2,866	Bonierbale et al. (1988)
Rice	~0.45	135	12/12	1,389	11.3	313	McCouch et al. (1988)
Soybean	1.15	252	31/20	2,147	8.5	517	Diers et al. (1992)
Tomato	0.95	800	12/20	1,480	1.9	619	Paterson et al. (1991)

[a] Arumuganathan and Earle (1991)
[b] 1 pg = 965 mbp (Arumuganathan and Earle 1991)

1978) segregated according to a 3:1 ratio ($\chi^2 = 0.60$; $P = 0.44$) as reported previously (Ali 1950) and mapped to linkage group D2.

Comparison with RFLP maps in other crops

Table 2 provides a comparison of mapping parameters for various published RFLP maps. Caution should be exercised in comparing these various maps because recombination rates will vary depending on the biological material (e.g., inter- versus intraspecific crosses) and the environmental conditions. In addition, recombination rates, and, hence, the amount of DNA per map unit, vary widely within a genome. Of particular interest is the average interval between 2 adjacent markers, which ranged from approximately 2.0 cM (*A. thaliana* and tomato) to 14.0 cM (*Brassica napus*). The average interval in common bean (6.5 cM) is sufficiently small to warrant mapping of quantitative trait loci. The average physical distance per unit of genetic distance ranged from 230 kbp/cM in *A. thaliana* to 4,843 kbp/cM in barley. The current average physical distance (expressed as amount of DNA per map unit) in common bean is only 3.2 times that of *A. thaliana*, which raises the possibility of map-based cloning of selected genes provided additional tightly linked markers can be identified that lie within a relatively short physical distance around genes of interest. Genes controlling agronomic traits of interest segregating in this mapping population include genes for disease resistance and genes for *Rhizobium* nodulation intensity.

In conclusion, we have developed a low-density genetic linkage map for the common bean genome. Markers used to develop this map included RFLPs, RAPDs, isozymes, and morphoagronomic traits, including a virus resistance gene. The development of this map was facilitated by the high levels of RFLP observed in *P. vulgaris*. Additional markers need to be mapped in order to increase its density and to reduce the number of linkage groups to 11, the haploid chromosome number of common bean. We will use this map to study the genetic control of morpho-agronomic traits in order to improve our understanding of the genetics of this species. Furthermore, some of the markers located on this map may serve as starting points for map-based cloning.

Acknowledgements. This research was funded by the Agency for International Development. R.O.N. holds a fellowship from CNPq, Brazil. S.M.T. holds fellowships from CNPq and FAPESP, Brazil. We thank the following individuals: S. Singh for advice on the choice of parents for our mapping population; J. Wendel, and E. Lander and S. Lincoln for the current versions of LINKAGE-1 and MAPMAKER, respectively; those listed in Materials and methods for providing clones; G. Arndt Shonnard, B. Garrido, P. Guzmán, N. Polignano, and N. Tuschla for technical assistance; and R. Kesseli and R. Michelmore for advice on the use of MAPMAKER.

References

Ali M (1950) Genetics of resistance to the common bean mosaic virus (bean virus 1) in the bean (*Phaseolus vulgaris* L.) Phytopathology 40:69–79

Aramuganathan K, Earle DE (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Bassett MJ (1991) A revised linkage map of common bean HortScience 26:834–836

Bonierbale M, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095–1103

Botstein D, White RL, Skolnick M, Davis RV (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

1 one)

153

Broglie KE, Gaynore JJ, Broglie RM (1986) Ethylene-regulated gene expression: molecular cloning of the genes encoding an endochitinase from *Phaseolus vulgaris*. Proc Natl Acad Sci USA 83:6820–6824

Coe EH, Hoisington DA, Neuffer MG (1990) Linkage map of corn (maize) (*Zea mays* L.) (2n = 20). In: O'Brien SJ (ed) Genetic maps. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp 639–667

Diers BW, Keim P, Fehr WR, Shoemaker RC (1992) RFLP analysis of soybean seed protein and oil content. Theor Appl Genet 83:608–612

Drijfhout E, Silbernagel MJ, Burke DW (1978) Differentiation of strains of bean common mosaic virus. Neth J Plant Pathol 84:13–26

Durante M, Cionini PG, Avanzi S, Cremonini R, D'Amato F (1977) Cytological localization of the genes for the four classes of ribosomal RNA (25S, 18S, 5.8S, and 5S) in polytene chromosomes of *Phaseolus coccineus*. Chromosoma 60:269–282

Edington BV, Lamb CJ, Dixon RA (1991) cDNA cloning and characterization of a putative 1,3-β-D-glucanase transcript induced by fungal elicitor in bean cell suspension cultures. Plant Mol Biol 16:81–94

Edwards KE, Cramer CL, Bolwell GP, Dixon RA, Schuch W, Lamb CJ (1985) Rapid transient induction of phenylalanine ammonia-lyase mRNA in elicitor-treated bean cells. Proc Natl Acad Sci USA 82:6731–6735

Feinberg AP, Vogelstein B (1984) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267

Fuller F, Kunstner PW, Nguyen T, Verma DPS (1983) Soybean nodulin genes: analysis of cDNA clones reveals several major tissue-specific sequences in nitrogen-fixing root nodules. Proc Natl Acad Sci USA 80:2594–2598

Garrido B, Nodari R, Debouck DG, Gepts P (1991) *Uni-2*, a dominant mutation affecting leaf development in *Phaseolus vulgaris*. J Hered 82:181–183

Gebhardt C, Ritter E, Debener T, Schachtschabel V, Wakemeier B, Uhrig H, Salamini F (1989) RFLP analysis and linkage mapping in *Solanum tuberosum*. Theor Appl Genet 78: 65–75

Gepts P (1988) Provisional linkage map of common bean. Annu Rep Bean Improv Coop 31:20–25

Gepts P, Llaca V, Nodari RO, Panella L (1992) Analysis of seed proteins, isozymes, and RFLPs for genetic and evolutionary studies in *Phaseolus*. In: Linskens HF, Jackson JF (eds) Modern methods of plant analysis (new series): seed analysis. Springer, Berlin Heidelberg New York ~~in press~~ pp. 63–93

Heun M, Kennedy AE, Anderson JA, Lapitan NLV, Sorrells ME, Tanksley SD (1991) Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*). Genome 34:437–447

Hoffman LM, Donaldson DD (1985) Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. EMBO J 4:883–889

Jorgensen RA, Cuellar RE, Thompson WP, Kavanagh TA (1987) Structure and variation in ribosomal RNA genes of pea. Plant Mol Biol 8:3–12

Kesseli RV, Paran I, Michelmore RW (1990) Genetic linkage map of lettuce (*Lactuca sativa*, 2n = 18). In: O'Brien SJ (ed) Genetic maps. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp 6100–6102

Koenig R, Gepts P (1989a) Segregation and linkage of genes for seed proteins, isozymes, and morphological traits in common bean (*Phaseolus vulgaris*). J Hered 80:455–459

Koenig R, Gepts P (1989b) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. Theor Appl Genet 78:809–817

Lamprecht H (1961) Weitere Kopplungsstudien an *Phaseolus vulgaris* mit einer Übersicht über die Koppelungsgruppen. Agric Hortic Genet 19:319–332

Lamprecht H von (1934) Zur Genetik von *Phaseolus vulgaris*. VII. Zwei weitere Gene für die Sameneigenschaft, *Cor* und *Fast*. Hereditas 19:163–176

Lander ES, Green P, Abrahamson J, Barlow A, Daly M, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Landry BS, Hubert N, Etoh T, Harada JJ, Lincoln SE (1991) A genetic map for *Brassica napus* based on restriction fragment length polymorphisms detected with expressed DNA sequences. Genome 34:543–552

Leakey CLA (1988) Genotypic and phenotypic variation in common bean. In: Gepts P (ed) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, the Netherlands, pp 245–327

McCouch SR, Kochert G, Yu ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 75:815–829

Mehdy MC, Lamb CJ (1987) Chalcone isomerase cDNA cloning and mRNA induction by fungal elicitor, wounding and infection. EMBO J 6:1527–1533

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Morales FJ, Castaño M (1987) Seed transmission characteristics of selected bean common mosaic virus strains in differential bean cultivars. Plant Dis 71:51–53

Moreno J, Chrispeels MJ (1989) A lectin gene encodes the α-amylase inhibitor of the common bean. Proc Natl Acad Sci USA 86:7885–7889

Nguyen T, Zelechowska M, Foster V, Bergmann H, Verma DPS (1985) Primary structure of the soybean nodulin-35 gene encoding uricase II localized in the peroxisomes of uninfected cells of nodules. Proc Natl Acad Sci USA 82:5040–5044

Nodari RO, Koinange EMK, Kelly JD, Gepts P (1992) Towards an integrated linkage map of common bean. I. Development of genomic DNA probes and levels of restriction fragment length polymorphism. Theor Appl Genet ~~in press~~ 84:136–142

Osborn TC, Blake T, Gepts P, Bliss FA (1986) Bean arcelin. 2. Genetic variation, inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris*. Theor Appl Genet 71:847–855

Paterson AH, Tanksley SD, Sorrells ME (1991) DNA markers in plant improvement. Adv Agron 46:39–90

Prakken R (1970) Inheritance of colour in *Phaseolus vulgaris* L. II. Critical review. Meded Landbouwhogesch Wageningen 72-23:1–38

Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV, Scolnik PA (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc Natl Acad Sci USA 89:1477–1481

Rigby PWJ, Dieckman M, Rhodes C, Berg O (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick-translation with DNA polymerase. J Mol Biol 113:237–251

Ryder TB, Cremer CL, Bell JN, Robbins MP, Dixon RA, Lamb CJ (1984) Elicitor rapidly induces chalcone synthase mRNA in *Phaseolus vulgaris* cells at the onset of the phytoalexin defense response. Proc Natl Acad Sci USA 81:5724–5728

Ryder TB, Hedrick SA, Bell JB, Liang X, Clouse SD, Lamb CJ (1987) Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. Mol Gen Genet 210:219–233

Schumann K, Baumann A, Nagl W (1990) Localization of phaseolin genes in the polytene chromosomes of *Phaseolus coccineus* (Leguminosae). Genetica 83:73–76

Schwartz HF, Pastor Corrales MA, Singh SP (1982) New sources of resistance to anthracnose and angular leafspot of beans (*Phaseolus vulgaris* L.). Euphytica 31:741–754

Singh SP, Gutiérrez JA, Molina A, Urrea C, Gepts P (1991a) Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci 31:23–29

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

Slighton JL, Sun SM, Hall TC (1983) Complete nucleotide sequence of a French bean storage protein gene: phaseolin. Proc Natl Acad Sci USA 80:1897–1901

Song KM, Suzuki JY, Slocum MK, Williams PH, Osborn TC (1991) A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci. Theor Appl Genet 82:296–304

Suiter KA, Wendel JF, Case JS (1983) LINKAGE-1, a PASCAL computer program for the detection and analysis of genetic linkage. J Hered 74:203–204

Thummler F, Verma DPS (1987) Nodulin-100 of soybean is the subunit of sucrose synthase regulated by availability of free heme in nodules. J Biol Chem 262:14730–14736

Tucker ML, Milligan SB (1991) Sequence analysis and comparison of avocado fruit and bean abscission cellulases. Plant Physiol 95:928–933

Vallejos CE, Chase CD (1991) Linkage between isozyme markers and a locus affecting seed size in *Phaseolus vulgaris* L. Theor Appl Genet 81:413–419

Walter MH, Liu J, Grand C, Lamb CJ, Hess D (1990) Bean-pathogenesis-related (PR) proteins deduced from elicitor-induced transcripts are members of a ubiquitous new class of conserved PR proteins including pollen allergens. Mol Gen Genet 222:353–360

Weeden NF (1984) Distinguishing among white-seeded bean cultivars by means of allozyme genotypes. Euphytica 33:199–208

155

Genetic Markers and Core Collections

Paul Gepts
Department of Agronomy and Range Science
University of California
Davis, CA 95616-8515

Presented at the IBPGR/CGN/CENARGEN workshop on:
Core Collections: improving the management and use of plant germplasm collections
Brasilia, 23-28 August, 1992

Genetic Markers and Core Collections

Paul Gepts
Department of Agronomy and Range Science
University of California
Davis, CA 95616-8515

Abstract: Various types of molecular and biochemical markers have been used to characterize genetic diversity in germplasm collections among which principally isozymes, RFLPs, and seed proteins. These various categories of markers can be distinguished by several attributes: level of polymorphism, degree of environmental stability, number of loci, molecular basis of the polymorphism, and practicality. One or the other of these markers can be chosen depending on the objective of the study. Using these markers it has been possible to identify or confirm centers of domestication. Markers in particular have been able to identify in some cases more specifically than other traits 1) which area was the most likely center of dometication; 2) multiple areas of doemstication; and 3) in polyploid crops, the female or male parent. Markers have been used to establish further subdivisions in the cultivated gene pool and, thus, determine in some cases whether phenotypic similarity was due to common ancestry or alternative causes such hybridization, polyphyly, etc. Markers constitute tools to investigate the occurrence of gene flow between wild and cultivated forms because in some cases they provide rare alleles that will reveal gene flow where morphological traits will not. Molecular and biochemical marker studies generally show a reduction of diversity during domestication although exceptions exist. The principal advantage of molecular and biochemical markers is that they are genotypic markers and, hence, will reflect the actual genetic distance between accessions and their common ancestry more accurately than phenotypic markers. Their main disadvantage is their cumbersomeness. It is suggested that this disadvantage could be circumvented by establishing 1) correlations between molecular and morphological markers; and 2) discriminant functions that would allow us to use morpho-agronomic traits to assign accessions to groups of evolutionarily related groups

("gene pools") as determined by molecular and biochemical markers.

Keywords: Seed proteins - isozymes - RFLPs - RAPDs - minisatellite sequences - M13 - genetic diversity - domestication - crop evolution - multivariate analysis

Introduction

According to the definition of Frankel (1984), a core collection is a subset of the general (or bulk or reserve) collection that represents the genetic diversity of a crop species and its relatives with a minimum of repetitiveness. There are therefore two essential steps in the establishment of core collection: a first one involves the characterization of the genetic diversity of of species and their wild relatives and the second one involves the choice of accessions to be included in the core collection based on genetic diversity data.

Both of these steps involve a sampling strategy. In particular, sampling in the first step involves a choice of criteria by which genetic diversity is characterized. Plant genomes may contain some 100,000 genes (Kamalay and Goldberg 1980). Therefore, when one choses a particular class of traits or markers to characterize diversity, one performs in effect a sampling of the plant genome.

Various criteria have been used to analyze genetic diversity in crop plants among which morphological, agronomic, ecogeographical, and biochemical or molecular traits or markers. Each of these have their advantages and disadvantages as markers of genetic diversity. Molecular markers include any markers reflecting direct changes at the DNA sequence level, principally restriction fragment length polymorphisms (RFLPs), but also random amplified polymorphic DNA (RAPDs), minisatellite markers, etc. Because crop evolution studies require a sufficient sampling of the diversity contained in the species, direct DNA sequencing has rarely been used, if at all, because of its inherent cumbersomeness.

Biochemical markers have included principally isozymes and seed proteins. The major advantage of molecular and biochemical markers is their presumed selective neutrality, although cases of non-neutrality have been reported such as *Adh* in *Drosophila* (Anderson and McDonald 1983). This general neutrality allows us to distinguish those similarities that are due to common ancestry from similarities due to convergence. Over the last few years, an increasing body of data on crop evolution has accumulated based on molecular and biochemical markers (reviewed in Doebley 1989, 1992; Gepts 1990, 1992; Clegg 1990).

In this presentation, molecular and biochemical markers will be discussed as indicators of patterns and levels of diversity. After a discussion of the attributes of various types of markers, some of their contributions to our knowledge of the organization of genetic diversity in crop plants will be presented.

Attributes of molecular and biochemical markers as genetic diversity indicators

Molecular and biochemical markers have been used in a substantial number of studies of genetic diversity in crop plants. A survey of the literature shows that isozymes, RFLPs (restriction fragment length polymorphism), and seed proteins have been used most often in these studies (reviewed in Doebley 1989, 1992; Gepts 1990; Clegg 1990). More recently, RAPDs (random amplified polymorphic DNA) and sequences hybridizing to minisatellite markers have also been used.

There are several attributes by which one can assess the potential usefulness of a particular category of markers. These include level of polymorphism, environmental stability, the number of loci, molecular basis of the polymorphism, and the ease and cost of analysis (Table 1). Depending on the objective of the study, a certain level of polymorphism is required: at higher taxonomic levels (species or above), more conserved markers are needed, whereas at the population level more variable markers are desirable. Electrophoretic patterns should be free of environmental influence to confirm that observed differences are genotypic differences. The number of loci should be as high as possible to ensure adequate genome coverage. The molecular basis of the polymorphism should be known so that genetic distance and diversity parameters can be determined. Markers for which the polymorphism results from simple changes at the molecular level such as nucleotide substitutions (*e.g.*, isozymes, RFLPs) are therefore more amenable to such quantitative analysis than markers for which the polymorphism involves a more complex series of events (*e.g.*, seed proteins). On the other hand, the probability of homoplasy is higher for the former class of markers than for the latter. Finally, the methodology employed should be as inexpensive and simple as possible to permit the analysis of samples of adequate size.

The main attributes of **isozymes** include the simplicity and low cost of the analysis, the simple molecular basis of their polymorphism, and a reasonable genome coverage (10-50 loci per species). In addition, standardized experimental conditions allow us to detect genotypic differences although isozymes are phenotypic markers. A disadvantage is the general but not universal lower level of polymorphism (Wendel and Weeden 1989; Weeden and Wendel 1989; Doebley 1989; Hamrick and Godt 1990).

RFLP markers can display a wide range of levels of polymorphism depending on the species (Table 4 of Nodari et al. 1992), the genome (e.g., cytoplasmic vs. nuclear: Curtis and Clegg 1984; Palmer 1987; Zurawski and Clegg 1987; Wolfe et al. 1987), or the particular sequence. In general, however, RFLPs are more polymorphic than isozymes. For example, in a direct comparison between the two classes of markers in the same set of maize genotypes, Messmer et al. (1991) found that RFLPs were more polymorphic than isozymes both in terms of the number of polymorphic loci (94 vs. 68 %) and the average number of variants per polymorphic locus (3.4 vs. 2.5). Additional advantages of RFLPs include their better genome coverage and environmental stability. The molecular basis of RFLPs can be as simple as a single nucleotide change which can lead to a restriction site loss, or less likely a site gain. The likelihood of a repeat mutation involving a nucleotide change may be higher than that involving a rearrangement characterized by its size as well as location. Through restriction mapping, it is possible to identify the molecular basis of the polymorphism although this limits the number of sequences that can be analyzed (Gepts and Clegg 1989). Finally, the RFLP technology is cumbersome and costly, which effectively limits the sample size.

Attributes of **seed proteins** as electrophoretic markers include their high level of polymorphism, their high level of environmental stability (although a few exceptions have been reported: e.g., Gayler and Sykes 1985), and the complex molecular basis of the electrophoretic patterns that includes nucleotide substitutions, insertions/deletions, and co- and posttranslational modifications. This complexity at the molecular level, however, makes it difficult to relate phenotypic changes in electrophoretic banding pattern with changes at the molecular level. Hence, one is usually limited to phenetic analyses with this category of markers. A disadvantage is the low number of loci involved (usually less than 10) (reviewed in Gepts 1990).

Minisatellite markers, principally those revealed through cross-hybridization with human mini-satellites or M13-derived sequences (e.g., Dallas 1988; Rogstad et al. 1989; Stockton et al. 1992), often reveal very high levels of polymorphism. The fingerprinting pattern does not appear to be influenced by environmental conditions (G. Sonnante, T. Stockton, and P. Gepts, unpubl. results). Little is known about the actual molecular basis of hypervariable sequences in plants. The complexity of the fingerprinting pattern and actual mapping data (T. Stockton, R. Nodari, and P. Gepts, unpubl. results) suggest, however, that several loci are involved, although the genome coverage may not be as extensive as that of RFLPs or RAPDs. The technology is similar to RFLP technology with the exception that several loci can be sampled at once.

The advantages of **RAPD markers** are that they can be more polymorphic than RFLPs (Williams et al. 1990), they offer genome coverage equivalent to that of RFLPs, and their methodology can be quite simple provided that a rigorously standardized methodology with the necessary controls is adhered to. Disadvantages include limited information about the environmental stability of the polymorphism and the molecular basis of RAPDs.

In summary, each class of molecular or biochemical markers possesses advantages and disadvantages. Depending on the goals of the study, one or the other or a combination of markers can be used. In general, however, isozymes and RFLPs have been preferred because they allow us to recognize homologies and determine genetic distance and diversity parameters.

Information on genetic diversity provided by molecular and biochemical markers

Studies using molecular and biochemical markers have addresses the following issues affecting organization and distribution of genetic diversity: identification of the wild ancestor and the pattern of domestication, divergence in the cultivated gene pool, gene flow between wild ancestor and cultivated descendant, and fate of genetic diversity and partitioning of genetic diversity during domestication

Identification of the wild ancestor and the pattern of domestication

Identification of the wild ancestor is important in crop genetic resources conservation because it tells us which wild taxon contributed genetic material to the cultivated gene pool. Conversely, it also tells us which wild taxa did not contribute genetic material to the cultivated gene pool. Both of these aspects are obviously important when identifying accessions for a core collection that is supposed the span the spectrum of genetic diversity for the improvement of the crop.

Traditionally, the wild ancestor has been identified on the basis of the overall morphological similarity between the crop and a wild taxon. More recently, other arguments have been used, including cytogenetic, crossability, and chemotaxonomic data (*e.g.*, Simmonds 1976). It is clear from these study that the wild ancestor, although sometimes very different morphologically from its cultivated descendant, is conspecific with it and belongs to its primary gene pool (Harlan and de Wet 1972).

Molecular and biochemical markers have usually been used to confirm the identity of the ancestral taxon and in some cases identify more specifically which populations were involved in domestication (Tables 2 and 3). Three cases - *Zea mays*, *Phaseolus vulgaris*, and *Brassica* spp. - provide examples of the power of molecular markers.

Case 1: Molecular and biochemical markers have confirmed teosinte as the actual and immediate progenitor of **maize**. Multivariate analyses of allozyme frequencies in the genus *Zea* by Doebley et al. (1984) showed that the wild taxon most similar isoenzymatically to the cultivars is *Z. mays* subsp. *parviglumis*, a short-spikeleted annual teosinte adapted to mesic intermediate altitudes. Other teosintes, such as *Z. mays* subsp. *mexicana*, a large-spikeleted annual teosinte adapted to arid high altitudes, *Z. luxurians*, an annual teosinte from southeastern Guatemala, and perennial teosintes *Z. perennis* and *Z. diploperennis*, were more distantly related to the cultivars.

Chloroplast DNA restriction site analyses also confirmed that annual teosintes were closely related to maize; however, these analyses could not distinguish between *Z. mays* subsp. *parviglumis* and *mexicana* as the most likely ancestral form (Doebley et al. 1987). Further confirmation that teosinte is the ancestor of maize was provided

by rDNA restriction site analysis (Zimmer et al. 1988).

Isozyme analyses by Doebley further suggested that the *Z. mays* subsp. *parviglumis* populations of the Balsas and Jalisco regions were the most probable progenitors of cultivated maize. Isozyme data did not agree with morphological data in that they showed *Z. mays* subsp. *parviglumis* to be the most closely related to maize, whereas morphologically *Z. mays* subsp. *mexicana* is most similar (maizoid) to maize (Doebley 1990).

Case 2: Wild ***Phaseolus vulgaris*** has a very large distribution, extending from northern Mexico to northern Argentina (Brücher 1988; Delgado Salinas et al. 1988). In addition, the archaeological record included finds of approximately equal ages in Mexico and Peru (Kaplan and Kaplan 1988). A survey of morphological diversity among cultivars suggested the existence of two groups, one distributed in Mexico and Central America and the other in the Andes (Evans 1976). These data, however, did not permit a more conclusive picture with regard to the domestication pattern of common bean. Molecular and biochemical markers, on the other hand, have revealed that the conspecific wild ancestor has diverged into two major groups, the Middle American and Andean groups, and that separate domestications in the Middle American and Andean areas have led to two groups of distinct cultivars (reviewed in Gepts 1990, 1992).

Case 3: The relationships among the nuclear genomes of cultivated ***Brassica*** species are summarized by the triangle of U (1935). In this triangle, the diploid species (*B. rapa* (syn. *campestris*), n=10; *B. nigra*, n=8; *B. oleracea*, n=9) occupy the apices and the three amphidiploid species (*B. carinata*, n=17; *B. juncea*, n=18; *B. napus*, n=19) the sides between their respective progenitors.

Because of maternal inheritance of cpDNA, it was possible to identify the maternal parent of the amphidiploids (Erickson et al. 1983; Palmer et al. 1983). *B. carinata* and *B. juncea* derived their cytoplasm from *B. nigra* and *B. campestris*, respectively. Part of *B. napus* derived its cytoplasm from *B. oleracea*, whereas the other part may have derived its cytoplasm through introgressive hybridization from another *Brassica* species (Palmer et al. 1983). Identity of the parents of the amphidiploids was confirmed by Song et al. (1988a) and Hosaka et al. (1990) on the basis of

nuclear genome-specific RFLP markers and by Delseny et al. (1990) on the basis of nuclear rRNA gene polymorphism. The molecular data also confirmed previous results obtained by various approaches such as cytogenetics, isozymes, and artificial resynthesis (see Hosaka et al. 1990 for references).

The following lessons can be drawn from these studies: 1) Markers can confirm or identify the wild ancestral taxon of crop plants. 2) This capability allows us to identify more specifically than with most other traits which populations more specifically were involved in the domestication process, especially in the case of crops with multiple domestications. For polyploid crops with uniparental maternal cytoplasmic inheritance, molecular and biochemical markers allow us to identify the cytoplasm inheritance. 3) Results from molecular or biochemical marker and morphological trait studies are not always correlated. Discrepancies may be attributed to possible selective effects associated more likely with morphological traits than with molecular markers. This underscores the fact that more than one category of markers or traits should be used to characterize genetic diversity.

Divergence within the cultivated gene pool

Divergence within the cultivated gene pool according to morphological, agronomic, ecological, etc.factors, will provide additional sampling criteria to set up a core collection. In addition to the information mentioned for several crops in the previous section, some more detailed studies have been performed on a number of crops (Table 4).

Case 1: Wellhausen et al. (1952) established a racial classification of Mexican **maize** on morphological grounds. Doebley et al. (1985) provided evidence that these maize races can be grouped into three complexes of morphologically and ecologically similar races: 1) the high elevation, Mexican pyramidal complex, including races Palomero Toluqueño, Chalqueño, and Cónico; 2) the northern complex, which includes races Azul and Apachito; 3) the remaining bulk of races, including the southern and western lowland dents and flours such as Tuxpeño and Tabloncillo. Frequencies of 22 isozyme alleles were correlated with altitude, which, in Mexico, is associated with precipitation and temperature, i.e. the lowlands are

hotter and moister whereas the highlands tend to be cooler and drier.

Case 2 : In **common bean**, when phaseolin or isozymes are used as a priori classification criteria, multivariate analyses such as canonical discriminant analysis reveal correlations with phenotypic traits such as growth habit, internode length, leaf and seed size, phenology, disease resistances, and general ecological adaptations (Fig. 1; Singh et al. 1991a, b). These data have led to a proposal for 6 races or groups of related cultivars, three in each major gene pool (Singh et al. 1991c). In the Middle American gene pool, race Jalisco represents the predominantly climbing cultivars of the southern, humid highlands of Mexico and Central America. Race Durango includes cultivars with a prostrate growth habit from the northern, arid highlands of Mexico and race Mesoamerica includes bush indeterminate cultivars from the humid, hotter lowlands of Mexico, Central America, and South America.

In the Andean gene pool, race Nueva Granada includes cultivars with determinate bush or indeterminate climbing growth habits adapted to moderate altitudes. Race Peru consists of cultivars with a climbing growth habit adapted to higher altitudes, whereas race Chile includes cultivars with a prostrate growth habit. It is interesting to note that races Durango and Chile display a similar phenotype, which includes medium-sized seeds and light pigmentation, in addition to the prostrate growth habit. These two races are diverged, however, at the molecular and biochemical levels (Gepts et al. 1986; Singh et al. 1991b; O.M. Paredes, V. Becerra and P. Gepts, unpubl. results). Their similar phenotype may have resulted from convergence due to selection for adaptation to arid environments prevalent in northern Mexico and Chile.

In summary, it is possible to identify examples that show or do not show concordance between patterns of diversity identified either by molecular and biochemical markers or morpho-agronomic traits. Discrepancies can be attributed to outcrosses between wild ancestor and cultivated descendants or within the cultivated gene pools. Alternatively, polyphyletic origins for the morpho-agronomic traits could also account for dissimilar results.

Gene flow between wild ancestor and cultivated descendant

Gene flow between wild ancestor and cultivated descendant will introduce additional genetic diversity in the cultivated gene pool. Groups of cultivars that show some of evidence of introgression from wild ancestors may need to be included in a core collection because they may carry alleles not usually represented in the cultivated gene pool. Methodologically, however, the identification of these introgressants is somewhat difficult (discussed in Doebley 1989, Gepts 1992).

Case 1: As mentioned earlier, the progenitor of cultivated **maize** is an annual teosinte, *Z. mays* subsp. *parviglumis*. A distinct annual teosinte, *Z. mays* subsp. *mexicana*, is not the direct progenitor of maize but may have contributed genetic diversity to the cultivated gene pool: in the region where maize and subsp. *mexicana* are sympatric, isozyme alleles characteristic of the wild taxon were observed in cultivars. These alleles were absent from the cultivars elsewhere (Doebley et al. 1984). Evidence for introgression in the other direction is provided by the perennial teosinte *Z. diploperennis* in which a plant showed isozyme alleles at two linked loci characteristic of the cultivated gene pool, suggesting that the chromosome segment marked by the two loci was introduced from maize (Doebley et al. 1984).

Case 2: Cultivated **pearl millet** is sympatric with its wild ancestor in the Sahel. Viable and fertile hybrids can be formed, although male sterility in these crosses has been described (Marchais and Pernès 1985). Isozyme analyses revealed that pollen from wild forms was preferentially involved in fertilization on wild pistils in pollination experiments with mixtures of pollen from wild and cultivated forms. Conversely, pollen from cultivated forms preferentially effected fertilization on cultivated pistils after mixed pollinations with wild pollen (Robert et al. 1991). This type of reproductive isolation may explain at least in part why wild and cultivated pearl millet maintain their phenotypic integrity in spite of their sympatry and predominant allogamy.

Case 3: In *P. vulgaris*, most wild and cultivated genotypes display the *Mdh-2*^{100} allele. Exceptions consists of a limited number of wild-growing accessions and race Jalisco cultivars from the southern highlands of Mexico, which show the *Mdh-2*^{102}

allele. The rarity and the narrow geographic localization of the *Mdh-2*^{102} allele argue in favor of gene flow although it is not possible to determine the direction of the gene flow (Singh et al. 1991b). It is interesting to note, however, that the phenotype of the cultivars involved does not display any hint of past hybridization with wild beans (S. Singh, pers. comm.). Phaseolin data also provide evidence for occasional outcrosses between wild and cultivated beans. All cultivated accessions from Mexico analyzed so far show the 'S' phaseolin type with the exception of one accession with an 'M' phaseolin type characteristic of wild-growing accessions of the same region. The latter accession also displays morphological signs of hybridization with wild beans such as smaller seeds and the striping and spotting pattern characteristic of wild bean seeds (Koenig et al. 1990). Some wild accessions from Colombia, Ecuador, and northern Peru also show signs of introgression from cultivars based on phaseolin data (Gepts and Bliss 1986; Gepts et al. 1986; O.M. Paredes and P. Gepts, unpublished results).

Fate of genetic diversity during domestication

Doebley (1989) surveyed isozyme studies examining differences indiversity between wild ancestor and cultivated descendants. Total heterozygosity (H_t) decreased from 0.24 in wild forms to 0.19 in cultivated forms, the proportion of polymorphic loci per population from 0.32 to 0.25, the proportion of polymorphic loci per taxon from 0.58 to 0.49, and the number of alleles per locus per taxon from 2.47 to 2.15. Chloroplast DNA data surveyed by Doebley (1992) also showed a reduction in diversity (calculated as the average probability per variable site that two accessions will differ: Clegg et al. 1984) and the proportion of polymorphic restriction sites (Table 5).

Nuclear DNA markers such as genes for ribosomal RNA and single-copy RFLPs were also characterized by a reduction in diversity (Doyle 1988; Allard 1988; Gepts and Clegg 1989; Havey and Muehlbauer 1989; Cordesse et al. 1990).

A few exceptions to this trend have been reported. In barley, isozymes and cpDNA showed equivalent levels of diversity in wild and cultivated accessions whereas mtDNA diversity was higher in the cultivars than in the wild ancestor (Holwerda et al. 1986; Jana and Pietrzack 1988). These results contrast with those of Brown and Munday (1982) for

isozymes and Clegg et al. (1984) and Neale et al. (1986) for cpDNA. It is not clear what the cause is of these discrepancies. One possibility is the sampling of the plant material which differed among these studies for the total number of accessions, the relative number of wild and cultivated accessions, and their respective geographic origins.

In pearl millet (*Pennisetum glaucum*), RFLPs of genes for ribosomal RNA and the *Adh-1* locus analyzed on the same set of wild and cultivated accessions showed contrasting trends in diversity: whereas the former showed a clear reduction in diversity between the wild ancestor and the cultivated descendant, the latter showed comparable levels of diversity (Gepts and Clegg 1989). It is possible that the diversity at the *Adh-1* locus in the cultivars is maintained through introgression with the sympatric wild ancestor. For rRNA genes, this mechanism would not operate if these genes are tightly linked to an essential feature of the cultivated phenotype; hence, after outcrossing with a wild population, selection for the cultivated phenotype in subsequent generations would eliminate introgressed rRNA genes from wild populations. Information on the actual linkage relationships between these molecular markers and genes controlling the cultivated phenotype is needed to verify this hypothesis.

In common bean, a comparison of trends in genetic diversity revealed by isozymes, low-copy nuclear RFLPs, and sequences hybridizing to M13-derived sequences, also showed a reduction in diversity in the cultivars compared with wild *P. vulgaris* (Table 6). An exception are the RFLP data for the Andean gene pool for which a slight increase in diversity was observed. Possible explanations include a sampling effect and the lack of representation of wild populations from certain areas in the Andes.

In summary, available data suggest a reduction in diversity in the cultivated gene pool compared to the wild ancestral gene pool. Wild forms represent therefore a source of additional diversity and should be represented in some way in the core collection of a crop.

Discussion

The major advantage of molecular and biochemical markers is their genotypic nature, *i.e.* they reflect directly changes at the DNA level. Morphological traits, on the other hand,

are phenotypic traits. Because the same phenotype can be achieved by different genotypes, it is sometimes difficult to equate phenotype and genotype. Accessions with a similar phenotype may sometimes be evoutionarily unrelated. For example, in common bean, races Durango and Chile show similar phenotypes yet biochemical markers show they belong to different gene pools, Middle America and Andes, respectively, of common bean (Singh et al. 1991c). An additional advantage of markers is that they are generally, although not always, selectively neutral and, hence, will show patterns of diversity independent of selection.

A disadvantage of molecular and biochemical markers can be their cumbersomeness. It is obviously unrealistic to subject an entire collection or even a large fraction of it to molecular and biochemical analyses. Simplifications to the procedures such as the introduction of the polymerase chain reaction may help alleviate this problem. Another possible soultion to this problem is to correlate molecular and biochemical markers with morphological markers. For example, in common bean, certain alleles of morphological markers are found at high frequency in certain gene pools or races. Races can be distinguished by their leaflet and bracteole shape and size, seed size and shape, pod beak insertion, and pigmentation patterns in flower and seed. The *Mf* gene causes the appearance of purple veination at the bottom of the flower and is found at high frequency in race Mesoamerica. The *Punc* (colored dots on seeds), *Mt* (seed mottling), *Str* (seed striping), and *Cor* (seed corona) are Andean alleles. Andean materials generally show larger leaves, longer internodes, and larger seeds (Gepts et al. 1986; Singh et al. 1991a).

Multivariate methods could be used to assign individual accessions on the basis of phenotypic data to specific groups (gene pools, races, etc.) that have been recognized previously using marker-based analysis of diversity (see above). Posterior probabilities of membership in the Andean or Middle American group obtained by discriminant analysis of morpho-agronomic traits were above 95% (when the prior classification criterion was Middle American vs. Andean phaseolin type). Posterior probabilities of membership ranged between 65 and 85% when the prior classification was the phaseolin type itself (P. Gepts, unpubl. results).

This approach would involve a dynamic analysis consisting of multiple rounds in each of

which diversity is analyzed at the molecular and morphological levels until a satisfactory grouping of genotypes is achieved such that most accessions can be assigned to evolutionarily distinct groups. Further experiments are needed to corroborate this approach.

References

Allard, R.W. 1988. Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild relatives. J. Hered. 79: 225-238

Anderson, S.M. and McDonald, J.F. 1983. Biochemical and molecular analysis of naturally occurring *Adh* variants in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 80: 4798-4802

Asins, M.J. and Carbonell, E.A. 1986. A comparative study on variability and phylogeny of *Triticum* species. 2. Interspecific relationships. Theor. Appl. Genet. 72: 551-558

Brown, A.H.D. and Munday, J. 1982. Population genetic structure and optimal sampling of landraces of barley from Iran. Genetica 58: 85-96

Brücher, H. 1988. The wild ancestor of *Phaseolus vulgaris* in South America. *In*: Gepts, P. (ed.), Genetic resources of *Phaseolus* beans, pp. 185-214. Kluwer, Dordrecht, the Netherlands

Clegg, M.T. 1990. Molecular diversity in plant populations. *In*: Brown, A.H.D., Clegg, M.T., Kahler, A.L. and Weir, B.S. (eds.), Plant population genetics, breeding, and genetic resources, pp. 98-115. Sinauer, Sunderland, MA

Clegg, M.T., Brown, A.H.D. and Whitfeld, P.R. 1984. Chloroplast DNA diversity in wild and cultivated barley: implications for genetic conservation. Genet. Res. 43: 339-343

Cordesse, F., Second, G. and Delseny, M. 1990. Ribosomal gene spacer length variability in cultivated and wild rice species. Theor. Appl. Genet. 79: 81-88

Curtis, S.E. and Clegg, M.T. 1984. Molecular evolution of chloroplast DNA sequences. Mol. Biol. Evol. 1: 291-301

Dallas, J.F. 1988. Detection of DNA "fingerprints" of cultivated rice by hybridization with a human minisatellite DNA probe. Proc. Natl. Acad. Sci. USA 85: 6831-6835

Dally, A.M. and Second, G. 1990. Chloroplast DNA diversity in wild and cultivated

species of rice (Genus *Oryza*, section Oryza). Cladistic-mutation and genetic-distance analysis. Theor. Appl. Genet. 80: 209-222

Delgado Salinas, A., Bonet, A. and Gepts, P. 1988. The wild relative of *Phaseolus vulgaris* in Middle America. *In*: Gepts, P. (ed.), Genetic resources of *Phaseolus* beans, pp. 163-184. Kluwer, Dordrecht, the Netherlands

Delseny, M., McGrath, J.M., This, P., Chevre, A.M. and Quiros, C.F. 1990. Ribosomal RNA genes in diploid and amphidiploid *Brassica* and related species: organization, polymorphism, and evolution. Genome 33: 733-744

Doebley, J. 1989. Isozymic evidence and the evolution of crop plants. *In*: Soltis, D.E. and Soltis, P.S. (eds.), Isozymes in plant biology, pp. 165-191. Dioscorides, Portland, OR

Doebley, J. 1990. Molecular systematics of *Zea* (Gramineae). Maydica 35: 143-150

Doebley, J. 1992. Molecular systematics and crop evolution. *In*: Soltis, P.S., Soltis, D.E. and Doyle, J.J. (eds.), Molecular systematics of plants, pp. 202-222. Chapman Hall, New York

Doebley, J., Renfroe, W. and Blanton, A. 1987. Restriction site variation in the *Zea* chloroplast genome. Genetics 117: 139-147

Doebley, J.F., Goodman, M.M. and Stuber, C.W. 1984. Isoenzymatic variation in *Zea* (Gramineae). Syst. Bot. 9: 203-218

Doebley, J.F., Goodman, M.M. and Stuber, C.W. 1985. Isozyme variation in the races of maize from Mexico. Amer. J. Bot. 72: 629-639

Doyle, J.J. and Beachy, R.N. 1985. Ribosomal gene variation in soybean (*Glycine*) and its relatives. Theor. Appl. Genet. 70: 369-376

Duvall, M.R. and Doebley, J.F. 1990. Restriction site variation in the chloroplast genome of *Sorghum* (Poaceae). Syst. Bot. 15: 472-480

Dvorak, J. and Appels, R. 1982. Genome and nucleotide differentiation in genomes of polyploid *Triticum* species. Theor. Appl. Genet. 63: 349-360

Dvorak, J. and Zhang, H.-B. 1990. Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. Proc. Natl. Acad. Sci. USA 87: 9640-9644

Dvorak, J., McGuire, P.E. and Cassidy, B. 1988. Apparent sources of the A genomes of

wheat inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome 30: 680-689

Erickson, L.R., Straus, N.A. and Beversdorf, W.D. 1983. Restriction patterns reveal origins of chloroplast genomes in *Brassica* amphiploids. Theor. Appl. Genet. 65: 201-206

Evans, A.M. 1976. Beans. *In*: Simmonds, N.W. (ed.), Evolution of crop plants, pp. 168-172. Longman, London, UK

Frankel, O.H. 1984. Genetic perspectives of germplasm conservation. *In*: Arber, W.K., Llimensee, K., Peacock, W.J. and Starlinger, P. (eds.), Genetic manipulation: impact on man and society, pp. 161-170. Cambridge Univ. Press, Cambridge, UK

Furnier, G., Cummings, M.P. and Clegg, M.T. 1990. Evolution of the avocados as revealed by DNA restriction fragment variation. J. Hered. 81: 183-188

Gayler, K.R. and Sykes, G.E. 1985. Effect of nutritional stress on the storage proteins of soybeans. Plant Physiol. 78: 582-585

Gepts, P. 1990. Genetic diversity of seed storage proteins in plants. *In*: Brown, A.H.D., Clegg, M.T., Kahler, A.L. and Weir, B.S. (eds.), Plant population genetics, breeding, and genetic resources, pp. 64-82. Sinauer, Sunderland, MA

Gepts, P. 1992. The use of molecular and biochemical markers in crop evolution studies. Evol. Biol. : in press

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. and Clegg, M.T. 1989. Genetic diversity in pearl millet (*Pennisetum glaucum* [L.] R.Br.) at the DNA sequence level. J. Hered. 80: 203-208

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*): evidence for multiple centers of domestication. Econ. Bot. 40: 451-468

Graur, D., Bogher, M. and Brieman, A. 1989. Restriction endonuclease profiles of mitochondrial DNA and the origin of the B genome of wheat, *Triticum aestivum*. Heredity 62: 335-342

Halward, T.M., Stalker, H.T., LaRue, E.A and Kochert, G. 1991. Genetic variation detectable with molecular markers among unadapted germ-plasm resources of

cultivated peanut and related wild species. Genome 34: 1013-1020

Hamrick, J.L. and Godt, M.J.W. 1990. Allozyme diversity in plant species. *In*: Brown, A.H.D., Clegg, M.T., Kahler, A.L. and Weir, B.S. (eds.), Plant population genetics, breeding, and genetic resources, pp. 43-63. Sinauer, Sunderland, MA

Harlan, J.R. and de Wet, J.M.J. 1972. A simple classification of cultivated sorghum. Crop Sci. 12: 172-176

Havey, M.J. and Muehlbauer, F.J. 1989. Variability for restriction fragment lengths and phylogenies in lentil. Theor. Appl. Genet. 77: 839-843

Hoffman, D.L., Soltis, D.E., Muehlbauer, F.J. and Ladizinsky, G. 1986. Isozyme polymorphism in *Lens* (Leguminosae). Syst. Bot. 11: 392-402

Holwerda, B.C., Jana, S. and Crosby, W.L. 1986. Chloroplast and mitochondrial DNA variation in *Hordeum vulgare* and *Hordeum spontaneum*. Genetics 114: 1271-1291

Hosaka, K. and Hanneman, R.E. 1988a. Origin of chloroplast DNA diversity in the Andean potatoes. Theor. Appl. Genet. 76: 333-340

Hosaka, K. and Hanneman, R.E. 1988b. The origin of the cultivated tetraploid potato based on chloroplast DNA. Theor. Appl. Genet. 76: 172-176

Hosaka, K., Kianian, S.F., McGrath, J.M. and Quiros, C.F. 1990. Development and chromosomal localization of genome-specific DNA markers of *Brassica* and the evolution of amphidiploids and n = 9 diploid species. Genome 33: 131-142

Hutchinson, J.B. 1951. Intra-specific differentiation in *Gossypium hirsutum*. Heredity 5: 161-193

Jaaska, V. 1980. Electrophoretic survey of seedling esterases in wheats in relation to their phylogeny. Theor. Appl. Genet. 5o: 273-284

Jaaska, V. 1981. Aspartate aminotransferase and alcohol dehydrogenase isoenzymes: intraspecific differentiation in *Aegilops tauschii* and the origin of the D genome polyploids in the wheat group. Plant Syst. Evol 137: 259-273

Jana, S. and Pietrzak, L. 1988. Comparative assessment of genetic diversity in wild and primitive cultivated barley in a center of diversity. Genetics 119: 981-990

Jorgensen, R.B. 1976. Relationships in the barley genus (*Hordeum*): an electrophoretic examination of proteins. Hereditas 104: 273-291

Kamalay, J.C. and Goldberg, R.B. 1980. Regulation of structural gene expression in

tobacco. Cell 19: 935-946

Kaplan, L. and Kaplan, L.N. 1988. *Phaseolus* in archaeology. *In*: Gepts, P. (ed.), Genetic resources of *Phaseolus* beans, pp. 125-142. Kluwer, Dordrecht, the Netherlands

Kesseli, R., Ochoa, O. and Michelmore, R.W. 1991. Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. sativa*). Genome 34: 430-436

Khairallah, M.M., Adams, M.W. and Sears, B.B. 1990. Mitochondrial DNA polymorphisms of Malawian bean lines: further evidence for two major gene pools. Theor. Appl. Genet. 80: 753-761

Kochert, G., Halward, T., Branch, W.D. and Simpson, C.E. 1991. RFLP variability in peanut (*Arachis hypogea* L.) cultivars and wild species. Theor. Appl. Genet. 81: 565-570

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

Marchais, L. and Pernès, J. 1985. Genetic divergence between wild and cultivated pearl millets (*Pennisetum typhoides*). I. Male sterility. Z. Pflanzenzüchtg. 95: 103-112

McClean, P.E. and Hanson, M.R. 1986. Mitochondrial DNA sequence divergence among *Lycopersicon* and related *Solanum* species. Genetics 112: 649-667

McLeod, M.J., Guttman, S.I. and Eshbaugh, W.H. 1982. Early evoution of chili peppers (*Capsicum*). Econ. Bot. 36: 361-368

McLeod, M.J., Guttman, S.I., Eshbaugh, W.H. and Rayle, R.E. 1983. An electrophoretic study of evolution in *Capsicum* (Solanaceae). Evolution 37: 562-574

Messmer, M.M., Melchinger, A.E., Lee, M., Woodman, W.L., Lee, E.A. and Lamkey, K.R. 1991. Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. Theor. Appl. Genet. 83: 97-107

Miller, J.C. and Tanksley, S.D. 1990. RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. 80: 437-448

Morden, C.W., Doebley, J. and Schertz, K.F. 1990. Allozyme variation among the spontaneous species of *Sorghum* section Sorghum (Poaceae). Theor. Appl. Genet. 80: 296-304

Neale, D.B., Saghai-Maroof, M.A., Allard, R.W., Zhang, Q. and Jorgensen, R.A. 1986.

Chloroplast DNA diversity in populations of wild and cultivated barley. Genetics 120: 1105-1110

Nevo, E., Zohary, D., Brown, A.H.D. and Haber, M. 1979. Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum*, in Israel. Evolution 33: 815-833

Nevo, E., Beiles, A. and Zohary, D. 1986. Genetic resources of wild barley in the Near East: structure, evolution and application in breeding. Biol. J. Linn. Soc. 27: 355-380

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

Nybom, H., Schaal, B.A. and Rogstad, S.H. 1989. DNA "fingerprints" can distinguish cultivars of blackberries and raspberries. Acta Hort. 262: 305-310

Ogihara, Y. and Tsunewaki, K. 1988. Diversity and evolution of chloroplast DNA in *Triticum* and *Aegilops* as revealed by restriction fragment analysis. Theor. Appl. Genet. 76: 321-332

Palmer, J.D. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Amer. Nat. 130: S6-S29

Palmer, J.D., Shields, C.R., Cohen, D.B. and Orton, T.J. 1983. Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theor. Appl. Genet. 65: 181-189

Palmer, J.D., Jorgensen, R.A. and Thompson, W.F. 1985. Chloroplast DNA variation and evolution in *Pisum*: patterns of change and phylogenetic analysis. Genetics 109: 195-213

Pinkas, R., Zamir, D. and Ladizinsky, G. 1985. Allozyme divergence and evolution in the genus *Lens*. Plant Syst. Evol 151: 131-140

Rick, C.M. and Fobes, J.F. 1975. Allozyme variation in the cultivated tomato and closely related species. Bull. Torrey Bot. Club 102: 376-384

Rieseberg, L.H. and Seiler, G. 1990. Molecular evidence and the origin and development of the domesticated sunflower (*Helianthus annuus* L.). Econ. Bot. 44 (3S): 79-91

Robert, T., Lespinasse, R., Pernès, J. and Sarr, A. 1991. Gametophytic competition as

influencing gene flow between wild and cultivated forms of pearl millet (*Pennisetum typhoides*). Genome 34: 195-200

Second, G. 1982. Origin of the genetic diversity of cultivated rice (*Oryza spp.*): study of the polymorphism scored at 40 enzyme loci. Jap. J. Genet. 57: 25-57

Simmonds, N.W. 1976. Evolution of crop plants. Longman, London.

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

Singh, S.P., Gutiérrez, J.A., Molina, A., Urrea, C. and Gepts, P. 1991b. Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci. 31: 23-29

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

Song, K., Osborn, T.C. and Williams, P.H. 1990. *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 3. Genome relationships in *Brassica* and related genera and the origin of *B. oleracea* and *B. rapa* (syn. *campestris*). Theor. Appl. Genet. 79: 497-506

Song, K.M., Osborn, T.C. and Williams, P.H. 1988a. *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 2. Preliminary analysis of subspecies within *B. rapa* (syn. *campestris*) and (*B. oleracea*). Theor. Appl. Genet. 76: 593-600

Song, K.M., Osborn, T.C. and Williams, P.H. 1988b. *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 1. Genome evolution of diploid and amphidiploid species. Theor. Appl. Genet. 75: 784-794

Stockton, T., Sonnante, G. and Gepts, P. 1992. Detection of minisatellite sequences in *Phaseolus vulgaris*. Plant Molecular Biology Reporter 10: 47-59

U, N. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Japan. J. Bot. 7: 389-452

Weeden, N.F. and Wendel, J.F. 1989. Genetics of plant isozymes. *In*: Soltis, D.E. and Soltis, P.S. (eds.), Isozymes in plant biology, pp. 46-72. Dioscorides, Portland, OR

Wellhausen, E.J., Roberts, L.M. and Hernández X., E. 1952. Races of maize in Mexico. Bussey Inst., Harvard Univ., Cambridge, MA.

Wendel, J.F. 1989. New World tetraploid cottons contain Old World cytoplasm. Proc. Natl. Acad. Sci. USA 86: 4132-4236

Wendel, J.F. and Weeden, N.F. 1989. Visualization and interpretation of plant isozymes. *In*: Soltis, D.E. and Soltis, P.S. (eds.), Isozymes in plant biology, pp. 5-45. Dioscorides, Portland, OR

Wendel, J.F., Brubaker, C.L. and Percival, A.E. 1992. Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton. : in press

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Ac. Res. 18: 6531-6535

Wolfe, K.H., Li, W.-H. and Sharp, P.M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84: 9054-9058

Zimmer, E.A., Jupe, E.R. and Walbot, V. 1988. Ribosomal gene structure, variation, and inheritance in maize and its ancestors. Genetics 120: 1125-1136

Zurawski, G. and Clegg, M.T. 1987. Evolution of higher plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Annu. Rev. Plant Physiol. 38: 391-418

Table 1. Comparison of molecular electrophoretic markers in evolutionary, genetic, and breeding studies: for further explanations, see text (from Gepts 1992)

	Allozymes	RFLPs	Seed Proteins	Minisatellite Sequences	RAPDs
Polymorphism	low	low-high	high	very high	low-high
Environmental stability	moderate	high	high	high	high
Number of loci	moderate (<50 loci)	high	low (<10 loci)	moderate	high
Molecular basis of polymorphism	simple	intermediate	complex	complex	complex
Practicality	quick, cheap	slow, expensive	quick, cheap	intermediate	quick, expensive

Table 2. Examples of closest wild relative of diploid species identified or confirmed using molecular and biochemical markers

Crop	Closest wild relative	Markers used	Source
Avocado (*Persea americana* var. *guatemalensis*: one of three cultivated avocado races)	*P. steyermarkii* (♀) x *P. nubigena* (♂)	cpDNA, rDNA, cellulase	Furnier et al. 1990
Barley (*Hordeum vulgare* subsp. *vulgare*)	*H. vulgare* subsp. *spontaneum*	Isozymes, cpDNA, mtDNA	Jørgensen 1976, Nevo et al. 1979, 1986; Clegg et al. 1984; Neal et al. 1986; Holwerda et al. 1986
Common bean (*Phaseolus vulgaris* var. *vulgaris*	*P. vulgaris* var. *aborigineus* and *mexicanus*	Isozymes, seed proteins, mtDNA, nuclear single-copy RFLPs	Gepts et al. 1986; Koenig and Gepts 1989; Khairallah et al. 1990; Singh et al. 1991a
Lentil (*Lens culinaris* subsp. *culinaris*)	*L. culinaris* subsp. *orientalis*	Isozymes, single-copy nuclear RFLPs	Pinkas et al. 1985; Hoffman et al. 1986; Havey and Muehlbauer 1989
Lettuce (*Lactuca sativa*)	*L. serriola* and other unidentified wild taxa	Nuclear single-copy RFLPs	Kesseli et al. 1991
Maize (*Zea mays*)	*Z. mays* subsp. *parviglumis* (Balsas and Jalisco)	Isozymes	Doebley et al. 1984
Pea (*Pisum sativum*)	*P. humile*	cpDNA	Palmer et al. 1985

Table 2. Examples of closest wild relative of diploid species identified or confirmed using molecular and biochemical markers

	Crop	Closest wild relative	Markers used	Source
1	Peanut (*Arachis hypogea*)	*A. monticola*	Nuclear single-copy RFLPs, RAPDs	Halward et al. 1991; Kochert et al. 1991
2 3	Pearl millet (*Pennisetum glaucum*)	*P. glaucum* subsp. *monodii*	cpDNA	Gepts and Clegg 1989
4	Peppers:		Isozymes	McLeod et al. 1982, 1983
5 6	1) *Capsicum baccatum* var. *pendulum*	1) *C. baccatum* var. *baccatum*		
7 8 9 10	2) *C. pubescens* 3) *C. annuum*, *C. frutescens*, *C. chinense*	2) *C. eximium* (presumably not the wild ancestor but a closely related species) 3) *C. annuum* var. *aviculare*		
11	Rice:		Isozymes, cpDNA	Second 1982; Cordesse et al. 1990; Dally and Second 1990
12 13	1) *Oryza sativa* 2) *O. glaberrima*	1) *O. rufipogon* 2) *O. breviligulata*		
14	Sorghum (*Sorghum bicolor*)	*S. bicolor* subsp. *arundinaceum*	cpDNA	Duvall and Doebley 1990
15	Soybean (*Glycine max*)	*G. soja*	rDNA	Doyle and Beachy 1985

Table 2. Examples of closest wild relative of diploid species identified or confirmed using molecular and biochemical markers

Crop	Closest wild relative	Markers used	Source
Tomato (*Lycopersicon esculentum*)	*L. esculentum* var. *cerasiforme*	Isozymes, mtDNA, nuclear single-copy RFLPs	Rick and Fobes 1975; McClean and Hanson 1986; Miller and Tanksley 1990

Table 3. Origin of certain polyploid crops as determined by molecular and biochemical markers

Crop	Parental taxa	Markers used	Source
Brassica spp.: 1) *Brassica carinata* (n=17); 2) *B. juncea* (n=18) 3) *B. napus* (n=19)	1) *B. nigra* (n=8) x *B. oleracea* (n=9); 2) *B. campestris* (n=10) x *B. nigra* (n=8) 3) *B. oleracea* (n=9) x *B. campestris* (n=10)	cpDNA, nuclear single-copy RFLPs, rDNA	Palmer et al. 1983; Erickson et al. 1983; Song et al. 1988; Hosaka et al. 1990; Delseny et al. 1990
Cotton: *Gossypium barbadense*, *G. hirsutum*	A genome species (Old World; ♀) x D genome species (New World; ♂)	cpDNA	Wendel 1989
Potato (tetraploid): 1) *Solanum tuberosum* var. *andigena*	Cultivated diploid species: *S. goniocalyx*, *S. phureja*, *S. stenotomum*; and a wild diploid *Solanum* species	cpDNA	Hosaka and Hanneman 1988a,b

183

Wheat:			
1) Diploid: *Triticum monococcum* (AA)	1) *T. monococcum* subsp. *aegilopoides*	1) Isozymes	1) Asins and Carbonell 1986; Jaaska 1980, 1981
2) Tetraploid: *T. turgidum* var. *turgidum* (AB)	2) *T. turgidum* var. *dicoccoides*	2) cpDNA, mtDNA	2) Tsunewaki and Ogihara 1988; Graur et al. 1989
A genome donor	*T. urartu*	nuclear repetitive RFLPs and RSAPs; rDNA	3) Dvořák and Appels 1982; Dvořák et al. 1988; Dvořák and Zhang 1990
B genome donor	*T. speltoides* or related species		
3) Hexaploid: *T. aestivum* (ABD)			
AB genome donor	3) *T. turgidum* var. *turgidum*	3) cpDNA, mtDNA	3) Tsunewaki and Ogihara 1988; Graur et al. 1989
D genome donor	*T. tauschii* subsp. *strangulata*	Isozymes	Asins and Carbonell 1986; Jaaska 1980, 1981; Nakai 1981)

Table 4. Examples of studies involving molecular and biochemical markers on divergence within cultivated gene pools

Crop	Divergence within the cultivated gene pools	Markers used	Source
Brassica spp.:			
B. rapa	2 groups: Indo-European (turnip, turnip rape, spring broccoli raab, sarson) vs. East Asian (pak choi, Chinese cabbage)	Nuclear single-copy RFLPs	Song et al. 1988b, 1990

1 2 3	Common bean (*Phaseolus vulgaris* L.)	3 Middle American (Durango, Jalisco, and Peru) and Andean (Nueva Granada, Peru, and Chile) differentiated by morphological, agronomic, and ecological characteristics	Isozymes, phaseolin seed proteins	Singh et al. 1991a,b,c
4 5	Cotton (*Gossypium hirsutum*)	Geographic differentiation that does not support a previous subdivision in 7 races on morphological grounds	Isozymes	Wendel et al. (1992); Hutchinson 1951
6	Maize (*Zea mays*)	3 complexes of morphologicall and ecologically similar races: 1) high elevation Mexican pyramidal complex: *e.g.*, Palomero Toluqueño, Chalqueño; 2) northern complex: *e.g.*, Azul, Apachito; 3) remainder: southern and western lowland dents and flours: *e.g.*, Tuxpeño, Tabloncillo	Isozymes	Doebley et al. 1985
7 8	Sorghum (*Sorghum bicolor*)	Geographic differentiation that does not support a previous classification based on morphological grounds	Isozymes	Morden et al. 1989; Harlan and de Wet 1972

Table 5. Chloroplast DNA diversity in wild ancestors and cultivated descendants in crop species (from Doebley 1992)

Progenitor-Crop	N	TS	P	D	Source
Glycine soja-	8	2	1.0	0.47	Close et al. 1990
Glycine max	46	2	1.0	0.20	
Helianthus annuus wild/weedy-	11	4	1.0	0.25	Rieseberg and
Helianthus annuus cultivated	23	4	0.0	0.00	Seiler 1990
Hordeum vulgare subsp. *spontaneum*-	11	5	1.00	0.43	Clegg et al.
Hordeum vulgare subsp. *vulgare*	9	5	0.20	0.08	1984
Hordeum vulgare subsp. *spontaneum*-	30	3	1.00	0.45	Neale et al.
Hordeum vulgare subsp. *vulgare*	51	3	0.33	0.03	1988
Pisum humilis-	4	6	0.67	0.43	Palmer et al.
Pisum sativum	13	6	0.50	0.10	1985
Sorghum bicolor subsp. *arundinaceum*-	6	8	1.00	0.35	Duvall and
Sorghum bicolor subsp. *bicolor*	3	8	0.13	0.06	Doebley 1990
Zea mays subsp. *parviglumis* var. *parviglumis*-	31	3	1.00	0.38	Doebley 1990
Zea mays subsp. *mays*	80	3	0.67	0.22	

1 [a] N: number of accessions analyzed; TS: total number of polymorphic sites; P: proportion of total sites that are polymorphic; D: diversity,

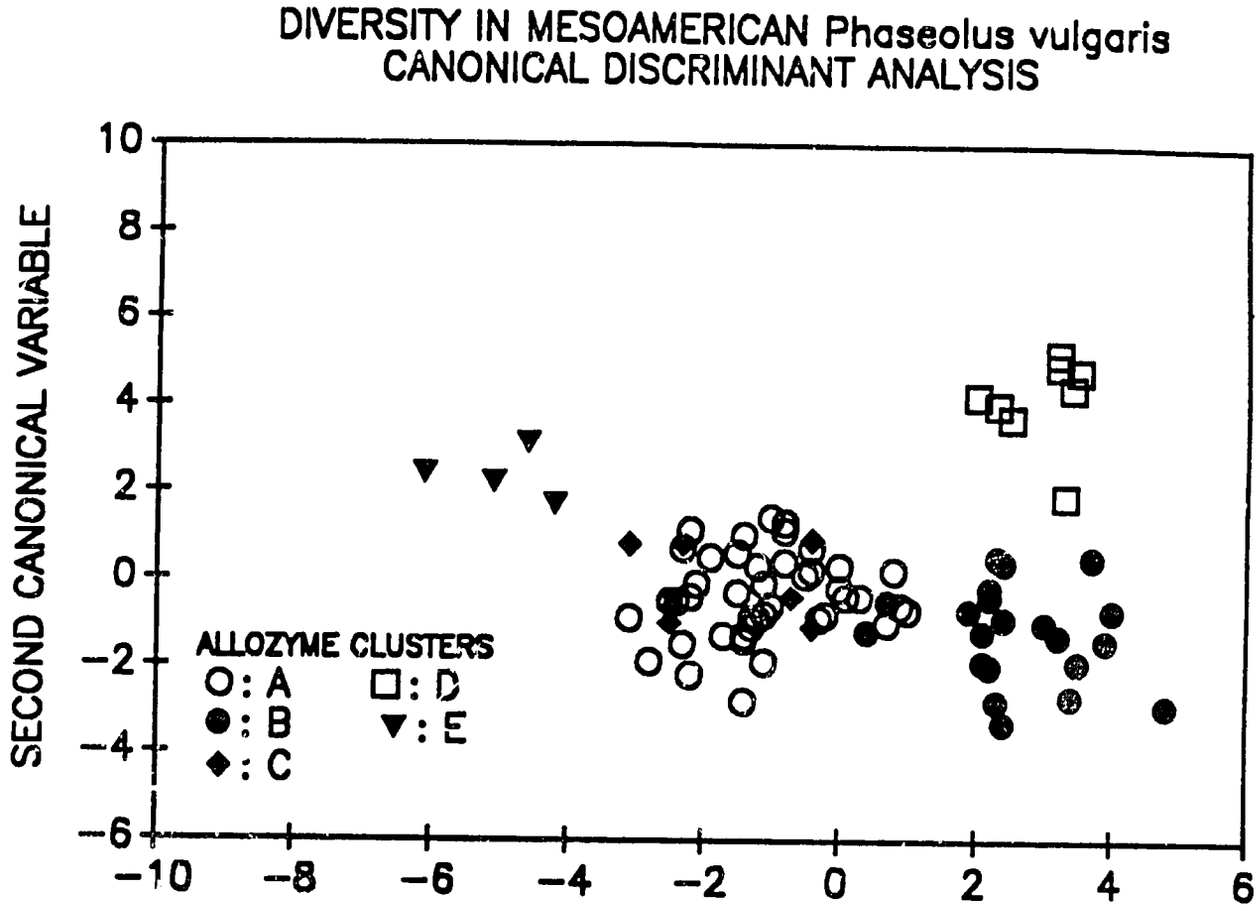
2 *i.e.* average probability that two cpDNAs will differ at a polymorphic site (Clegg et al. 1984)

Table 6. Trends in genetic diversity in *Phaseolus vulgaris* based on isozyme, RFLP, and M13-homologous sequences (Koenig and Gepts 1989; Singh et al. 1991b; V. Becerra, G. Sonnante, T. Stockton, and P. Gepts, unpubl. results)

			Average heterozygosity		
		N	Isozymes[a]	RFLPS[b]	M13[c]
Middle American	Wild	11	0.13	0.33	0.24
	Cultivated	36	0.09	0.27	0.20
Andean	Wild	11	0.07	0.25	0.20
	Cultivated	26	0.03	0.27	0.16

[a] 19 loci; [b] 13 probes for nuclear sequences, 3 restriction enzymes; [c] 1 probe, 1 restriction enzyme

1 **Fig. 1. Canonical discriminant analysis of morpho-agronomic diversity in the Middle**
2 **American gene pool of *Phaseolus vulgaris* (from Singh et al. 1991a)**



-181