

# CORE COLLECTION: THEORETICAL AND APPLIED ASPECTS

S. HAMON<sup>1</sup>, T. HODGKIN<sup>2</sup>, S. DUSSERT<sup>1</sup>, F. ANTHONY<sup>3</sup>  
and M. NOIROT<sup>1</sup>

1: ORSTOM, Laboratoire de Ressources Génétiques et d'Amélioration des Plantes Tropicales, BP 5035, 34032, Montpellier, France

2. IPGRI, Via delle Sette Chiese, 142, 00146, Rome, Italy

3. CATIE, Aparto 125, 7170, Turrialba, Costa Rica.

## Abstract

A core collection has been described as a collection which contains, with a minimum of repetitiveness, the maximum possible genetic diversity of a crop species and its wild relatives. Such a collection is not intended to replace existing genebank collections but to make the variation contained within such collections more accessible to users. Core collections can provide a way of improving germplasm enhancement.

In this review paper we first examine theoretical approaches suggested for establishing a core collection: those based on the neutral allele theory and those taking into account morpho-agronomic traits. Possibilities for establishing a hierarchical strategy are then discussed. In the second part we describe current examples of core collections for crops such as barley, cassava, *Phaseolus* and coffee. In the third part we examine some of the main criticisms of the core strategy and discuss the future development of core collections.

A large part of this paper is based on the workshop on core collection held in Brasilia in 1992 for which proceedings will be available in 1994 (Wiley & Son, U.K.).

## INTRODUCTION

One of the major issues which genebank managers must face is the need to improve the accessibility of their collections to users. In practice, plant breeders (and most other users) are interested in having fairly small numbers of genotypes which possess, or are likely to possess, the characters needed in their breeding programmes. In contrast, genebank managers have a responsibility to conserve as much as possible of the total variation in a crop and its wild relatives. This leads to collections which are difficult for plant breeders and other research workers to use effectively and the sheer size of many collections has frequently been cited as a barrier to increased utilization of collections (Holden, 1984). Are these different needs incompatible? Breeders require rapid identification of desirable traits and immediate access to samples, germplasm specialists want samples for more in-depth studies, curators require knowledge of the latter to provide the former.

ORSTOM Fonds Documentaire

73

N° 41234 ep1

23 MARS 1995

Cote : B

Recognizing this, Frankel (1984) and Frankel and Brown (1984) proposed that one way of alleviating the problem lay through the development of core collections. These would represent "with minimum repetitiveness, the genetic diversity of a crop species and its relatives". These might include sets that represent the broad genetic variation available for a total crop genome. The concept later developed by Brown (1989a and b), has attracted considerable discussion and debate. Concerns have been expressed over the vulnerability of the accessions not included in the core, the difficulties involving in identifying accessions that represent total variation, the bias that may exist against usefulness in favour of total genetic diversity and the difficulty of modifying an established core collection. However, a number of national and international groups have now developed or are developing core collections of crops such as *Phaseolus*, barley, wheat, cassava, coffee, okra, sorghum and rice.

In this paper we intend to review the main features of a core collection, the processes involved in its establishment and the characteristics of some core collections under development. This will provide the framework for discussion of some of the concerns noted above.

## **I. Theoretical basis for core collections**

### **1. Models based on the neutral allele model**

Brown (1989a) on the basis of the neutral allele model developed by Kimura and Crow (1964) showed that a core of 10%, selected at random (R strategy), from the whole collection can be expected to contain over 70% of the variation in a species. In practice, the figure of 10% can be modified to take account of the known features of the crop, the needs of the users for larger or smaller cores or the methodology used to develop the core. Brown (1989b) has argued that the most effective strategies should involve a hierarchical stratification of the whole collection into groups of accessions which share common taxonomic, geographical, ecological or generic characteristics. These strategies involve the sampling of a constant fraction (C strategy); the sampling of accessions in proportion to the number of accessions available per group (P strategy); and the sampling of accessions in proportion to the logarithm of the number available per group (L strategy). Both these later approaches appear to be better than selecting a constant number from each group. The infinite neutral alleles model applied to large numbers of isolated (island) populations assumes that every allele that arises through mutation is unique and selectively neutral, and that populations are reproductively isolated from one another. This model lies between models that assume heterotic selection (many alleles with comparable frequency) and those that assume mutation-selection balance (fewer alleles, one common, the others rare).

Schoen and Brown (1994) introduced two other strategies H and M. The H (Heterozygosity) strategy refers to the Nei's gene diversity index defined as one minus the sum of squared of allelic frequencies at the  $i$ th locus in the  $j$ th group (Nei, 1973). The M strategy (Maximization) differs from all other procedures because it refers to individual accessions and the variance and covariance measured at different loci. It is assumed that variation at a selected number of marker loci is representative of the variation at loci of interest in genetic conservation. Schoen and Brown (1994) have tested all strategies mentioned above on real data sets with both estimated and target loci. The results showed that for the overall average, ranking of the six strategies, in order of highest (rank 1) to lowest (rank 6), expected allele retention is :  $M > H > P > L > C > R$ .

### **2. The choice of a hierarchical structure**

The first stage in the development of any core collection is to assemble the available data on the whole collection. This will certainly include the passport data available to the genebank and as much characterization data as has been collected. It may also include

evaluation data and, where information is available, data from biochemical studies of isozymes, seed proteins, molecular markers etc.

Genetic diversity is not randomly distributed among plant populations, but has a structure that can generally be represented by a hierarchical model, a tree. One approach could be based on passport data combined with knowledge about the structure of the gene pool. In that case, the assumption is that the identity and the origin of material allows predictions about the genetic diversity in that material. This also implies that the passport data is reliable. The second approach is based on a phenetic analysis of characterization data on accessions from which the core is to be selected. In that case the assumption is that the observed diversity for morphological, molecular or other marker represents the underlying total genetic diversity.

#### *Grouping according to major features*

Hintum (1994) suggested defining a hierarchical structure by grouping the accessions in a collection according to major features which are known or expected to influence the distribution of diversity.

These may be ecogeographic, involving characteristics such as country or area of origin. They may take account of the major different features of the crop such as 2 row and 6 row types of barley or spring and winter wheats; or they may involve a mixture of these features. This procedure will ensure that unevenly distributed diversity, which is bound to exist in a collection, can be adequately taken into account in developing the core. Grouping of accessions in this way, on the basis of known or expected similarities, lies at the heart of the successful development of core collections. There is considerable evidence to suggest that country of origin is a reliable unit of grouping and indicator of genetic diversity and this may provide a useful approach to those exploring the development of groups for a particular collection (Peeters and Martinelli, 1989). Where there is reliable agronomic data, such an approach may be modified to take account of phenotypic similarity of accessions from different countries (Spagnoletti Zeuli and Qualset, 1987, 1993). Practical experience suggests that there will always be problems at this stage as to the completeness and reliability of the data available. Passport data may be minimal for many accessions with only the country of origin recorded. Characterization data may also be lacking for a significant proportion of a genebank's holding depending on the facilities available for such work. There will also be decisions to be made concerning how much data should be used and how it may be combined. (Hintum, 1994) concluded that it is often possible to describe the structure of the genetic diversity to some extent by describing these groups and their relationships. In many cases this can be sufficiently be represented in a hierarchical model

#### *Grouping on the basis of markers*

Genetic markers have been used for different purposes such as taxonomic studies, the search for the centre of diversity of a species, the route of domestication, for the relation between environment and diversity, and studies on the complete genepool (Gepts 1994). At this level it is necessary to define what is genetic diversity vs genetic differentiation. Genetic diversity can be defined as the extent to which heritable material differs within a group of plants. Genetic differentiation is the extent to which heritable material differs between a group of plants. The heritable material comprises its genomic and cytoplasmic DNA. It can differ at the level of DNA - sequences (alleles) but also at the level of allele combinations (genotypes). In addition, two genotypes could be similar by homology or homoplasy. Homology is the resemblance due to inheritance from a common ancestry; Homoplasy is the resemblance due to parallelism and convergence. Quantitative markers are dependent on both genotype and environment but they include morphological and agronomically useful characters (earliness, drought resistance) not suited for the study of genetic diversity. Genetic diversity can best be quantified on the basis of data as close to DNA as possible. Expression of morphological and agronomic characters indicates adaptation to environmental factors rather than genetic diversity and differentiation. RFLP (Restriction Fragments Length

Polymorphism) have more polymorphic loci than isozymes and more variant per locus. Microsatellites which appear to be highly polymorphic could also be very useful in diversity studies but little is known of the molecular basis of hypervariable sequences (Messmer et al 1991).

The rate of genetic diversity change under domestication is much higher as compared to the relatively slow process under natural evolution (Pickersgill 1984). Genetic drift and founder effect could lead to severe bottlenecks. Differentiation can also be marked such as for *Brassica oleracea* (Cabbage, Kale, Chinese Kale, cauliflower, broccoli, Brussels sprouts and kohlrab). Some other species like cultivated okra (*Abelmoschus esculentus* and *A. caillei*) show a good level of agronomic useful polymorphism with a very low level of genetic diversity at the molecular level (Hamon 1989). Plant breeding techniques can also influence the level of diversity. Several phenomena like induced mutations, hybridization between previously incompatible populations and introgressions can increase the diversity. Others, like inbred lines, cultivars for high input agriculture, lead to a reduction. The result is a complex multi-dimensional structure of variation within a crop. In practice, as noted by Gepts (1994), it appears that there may be considerable variation in the degree of concordance between the information obtained on patterns of diversity from agromorphological, biochemical and molecular studies.

### 3. The use of quantitative and (or) qualitative data

Hamon et al. (1994), Noirot et al. (1993) proposed and tested a global strategy with coffee. The procedure is as follows: the accessions are first hierarchically grouped according to taxonomic, ecological data, crossing fertility, genetic pattern of diversity. Then, within each group when data are available, a Principal Component Score procedure (PCS strategy) is used. This procedure makes use of characterization and evaluation data and permits accessions to be identified which maximize the variability in the core collection following appropriate principal component analysis. Each accession is characterized by its inertia in the factorial space. Selection is made on accessions which maximize the selected inertia. A good concordance is found between the test on real values (Hamon et al. 1994) and synthetic normally distributed variables and simulation (Noirot et al, submitted). The advantage of this procedure is that existing evaluation data may be used to identify the core hierarchy and on the other hand to maximize the within group variability. The choice of core accessions and the process used can be adapted to breeders' needs in respect of the numbers selected. The extent to which the procedure results in the inclusion in the core of the maximum genetic diversity in respect of qualitative traits is yet to be determined.

Spagnoletti and Qualset (1993) have tested five strategies for selecting a core collection of 3000 *Triticum durum* accessions using four qualitative and eight quantitative characters. Each of the following strategies generated about 500 accessions for the core sample: random, random systematic according to chronology of entries into the collection, stratified by country of origin, stratified by log frequency by country of origin, and stratified by canonical variables. The first three strategies produce samples representative to the whole collection but the remaining two produce the desired effects of increasing frequency from less represented countries of origin. The stratified canonical sample increased phenotypic variation. These authors conclude that multivariate approach is extremely useful but requires considerable data from the whole collection.

In conclusion, within a well defined group it is possible to improve the percentage of sampled diversity, without modifying the relative intensity of selection (i.e. 10%). To be really efficient, this procedure must involve only quantitative characters having a strong heritability.

## II. Some current examples of core collections

The core collection concept has merit in selecting a set of representative diversity from a larger assembly of germplasm. Different users will have different objectives in sampling

germplasm. The following examples, choiced among developping cores, show this difference but also the distance between sampling theory and the user choice.

### **1. The European barley Core collection (*Hordeum*)**

The Barley Core Collection Project is a collaborative international initiative (Knüpffer and Hintum, 1994). The accessions selected (approx. 2000) will cover the entire *Hordeum* genepool with defined numbers of landraces, improved cultivars and wild relatives from the primary, secondary and tertiary genepools. Genetic stocks will also be included. It is envisaged that the core collection will be held at a number of centres and, to ensure that it remains constant, it is intended that each accession should be a homozygous line. This raises an interesting new dimension in core collection work in that the accessions now become additional entities maintained separately from those from which they have been derived. The selection, management, maintenance and use of these accessions have more fully described by von Bothmer *et al.* (1990). The barley cultivated species is *Hordeum vulgare* L.. The species is diploid and shares the primary genepool with the *H. spontaneum* complex. The secondary gene pool comprises *H. bulbosum*. The tertiary gene pool includes about 30 species. For this complex, the core collection is not a selection from the germplasm collection of a single institution but from the entire genepool of a crop. It is a part of an existing gene bank but maintained separately.

The objectives are to facilitate the coordination of efforts and sharing responsibilities, to increase the knowledge about barley genepool, to use the existing germplasm, to provide standards for studies of genetic diversity. Consequently this implies an accumulation of large amount of data for a limited standard set of accessions. The research starting point is a small sample covering a considerable part of the whole collection which avoid expensive screening of large collections with duplicated material. The size should not exceed 2000 accessions. It is as follows : Category 1, cultivars 500 (phylogenetic group, including oriental and occidental); Category 2, landraces 800 (ecogeographical data, agricultural system practised, type of use) ; Category 3, *H. spontaneum* (including *agriocrithon* and introgression products with cultivated barley) 150 - 200 (ecogeographical data 2/3 from central, 1/3 from marginal areas); Category 4, Other wild species 60-100 accessions (2 per species - ecogeographical and morphological data); Category 5 : genetic stocks, and reference material (selected by barley genetic experts) 200. To ensure continued integrity accessions will be homozygous and homogeneous lines as far as possible. Heterogeneous samples are only accepted for the 2 outcrossing species. The advantages are identical multiplication over generations and locations and correspondance between information and material. The disadvantages are that variation within landraces is reflected with considerable reduction of the number of alleles. Concerning the wild relatives it is presumed that a single line contains the genetic background common to the material it represents.

### **2. Core collection of Brazilian cassava accessions (*Manihot*)**

In developing a core collection of Brazilian cassava accessions, Cordeiro *et al.* (1994) have shown that the data assembling phase can be extremely valuable in its own right to improve genebank management. Brazilian cassava accessions are distributed among a high number of centres involved in maintenance or breeding and the data collection phase established the origin of the existing accessions and allowed putative duplicates to be identified. Some 4132 accessions were identified from different collections in Brazil, of which 1200 were considered to be duplicates. This phase also led to the development of an effective hierarchical classification in which the first criterion was the category of the material (landraces - 2035 accessions, improved selections - 339 accessions and of unknown nature - 558 accessions) and the second was the agroecological zone of origin, of which nine were defined. A final criterion for grouping was based on characterization and evaluation data.

### 3. The CIAT bean core collection (*Phaseolus*)

In the CIAT *Phaseolus* core collection the same principles were applied, but in a much more sophisticated manner (Tohme *et al.*, 1994). Following preliminary grouping, according to known history of bean cultivation (in order to give more weight to centres of high diversity) an agroecological classification of Latin America was constructed, in which, each 10 minute grid was classified into one of 54 classes, on the basis of soil type, altitude, available water and photoperiod. The origin of the Latin American landraces for which passport data were available was determined and they were grouped into the 54 identified agroecological groups. A second step involved a weighting process based to ensure that variation in growth habit, seed colour and seed size was fully represented in the core collection.

### 4. The coffee core collection (*Coffea*)

The genetic organisation of the coffee gene pool was examined at 3 different levels: biogeography, genetic resources and available data. A core collection for coffee should consist of 88 genetic diversity groups of 3 types, according to their genetic history, the available genetic knowledge and the germplasm available: a *Coffea arabica* group, groups with well studied species like *C. liberica* and *C. canephora* and groups with a large number of neglected species.

For Type 1, *C. arabica*, 256 genotypes are kept in the core (128 *in vitro* and 128 in greenhouse). Most of them are issued from the original collection made in Ethiopia in 1966. Groups of type 2, presenting obvious agronomic interest were more evaluated for agromorphological traits than wild species. The PCS strategy (Principal Component Score), described above, could be applied. *C. liberica* was cultivated in Africa before the rise of *Fusarium* during the forties. We have found that about half the inertia is obtained when 10% of the 338 genotypes are selected, and 90% of the total inertia is obtained with a sample of 50% of these genotypes. The advantage of this procedure, which has also been tested with okra, is that existing evaluation data may be used to identify the core accessions and that the process can be adapted to breeders' needs in respect of the numbers selected. In our case, this is a "virtual core" because it can be selected and modified at any time from the base collection. For the real core, *C. liberica* was split in 3 groups according their geographical origin (guinean, congolese and Koto - Cameroon) and sampling was made as for the wild species. For group 3, constituted by wild species, it was possible to adopt different strategies. We have seen above for barley that the core collection strategy may lead to considerable reduction in the available diversity of wild forms. Users only want to keep the global coadapted complex and so only 2 or 3 wild genotypes are included. Crossa *et al.* (1993) demonstrate that about 200 plants are needed to capture alleles at frequency of 0.05 in 150 loci with a 90-95% probability in a composite. We have decided to adopt an intermediate strategy. For each diversity group, a bulk was established by collecting one seed per genotype. Then seeds were mixed and 100 seeds selected. In order to conserve a total of 40, fifty were sown in a greenhouse, fifty introduced *in vitro*. Today the *Coffea* core collection comprises 1348 genotypes corresponding to 30 diversity groups plus 11 species with several genotypes.

## III. Remaining problems and issues in the development of core collections

### *Criticisms on the core collection concept*

The fact that entries in the core are chosen to cover the genetic spectrum of the collection, the collections, the genepool, ... has led to the claim that the core forms a sub-optimal sample. We try to summarise in the following paragraphs answers given to most of the problems identified based on comments by Brown (1994).

One major criticism concerns the bias to representing diversity ignores usefulness. It is true that in many instances the core is unlikely to contain the single most useful source of



character. It provides a logical general strategy to identify the best source. One advantage for the breeder is the chance to become acquainted with the diversity of phenotypes in a crop and its related wild species. Some critics have also commented that the core collection is inflexible. There is a difficulty in estimating the rate at which changes should be made in the composition of the core. The base collection changes so the core should change. Other criticisms concern the variation within accessions which is often ignored and the insufficient attention given to very rare variants. For inbreeding species, like barley, where the project coordinators have chosen homogenous samples this is true. For outbreeding species an optimal sampling size has been proposed earlier (Marshall and Brown 1975). It has always been recognized that core collections will not necessarily include extremely rare genes. A proportion will be included, but this will be due to chance and many will not be present.

It has been also argued that the creation of a reasonably representative core collection would require so much information that it is impracticable, and that, once sufficient information has been collected to create such a collection, it would be unnecessary. The examples noted above have shown that core collections can be constructed on the basis of what is already known and that limited passport data and basic characterization data for major morphological characters can provide an effective grouping strategy for the development of a core collection. As in the case of cassava noted above, the actual process can be informative and lead to a better germplasm collection by assisting in the identification of duplicates and creating a well structured collection based on existing data.

#### *The future of base collections*

Another criticism by those who have expressed doubts about the development of core collections has been the fate of the accessions that are not included in the core. It has been argued that these accessions will be neglected and that administrators will be less prepared to provide resources for the maintenance of large collections if core collections are established.

There is no reason why this should be so. 1) the core is not an entity on its own, it is a guide and an entry point for the whole collection; 2) core collections are designed to stimulate and improve the use of genetic resources and hence, are likely to increase the value of the whole collection rather than decrease the value of the non-core element. However, studies on the links between the core collection and the whole collection are still needed. It is assumed that the core collection provides a simple and effective first step in a two stage process of identifying the most desirable accession. This needs to be studied in practice now that sufficient core collections exist to make tests possible.

#### *One core or many*

The difficulties in handling large and growing numbers of accessions in gene banks during decade before 1984 was largely responsible for the core collection proposal. (Mackay, 1994) suggest that this constraint to utilisation is largely negated by modern database technology. One basis for considering many core collections is that germplasm users often request a set of accessions that is likely to contain a characteristic the require which has not been previously described. He also considered that the core collection concept has a merit in selecting a set of representative diversity from a larger assembly of germplasm but different germplasm users will have different objectives in sampling germplasm. Three reasons are forwarded to avoid single core collections: 1- it is unlikely that the breeder is totally without information in respect to the information being sought; 2- the probability of identifying a useful accession within a single core is lower than in a specially selected one, 3- the genetic diversity in a single core, say representing the general variation of a single genebank, could be far from representative of the whole range of genetic diversity.

#### *Management and future issues of the cores*

If we try to find a general consensus for establishing a core, it could be suggested that the most flexible approach is to use passport, characterization and other data to first develop

a hierarchical classification of accessions into groups. Then, accessions will be assigned to progressively smaller groups based on shared characteristics with respect to taxonomy, geographic origin, ecological origin, genetic markers and agronomic data. Model and existing sampling strategies give a general scheme which must be always in mind even if the user needs very specific traits. Once a set of core accessions has been identified, a number of management issues need to be resolved.

The origin of the cultivated species, the number of wild relatives in the gene pool, their reproductive strategies,... are parameters of importance for the determination of an optimal management of accessions within each group. For some crops in some genebanks, it may be sufficient to mark the accessions in a database as belonging to the core and to identify the different groups of which they are the representatives. In most cases, this would be combined with further multiplication to ensure adequate populations of the core accessions were maintained for the additional research envisaged and for distribution. For other crops such as clonally propagated or recalcitrant seeded ones, the core may be maintained in a separate plantation or *in vitro*.

Most of the available information on the management and use of core collections is given by the international barley core collection but several other works exist. A core collection of about 60 populations of French raygrass (*Lolium perenne*) was evaluated for agronomic characters in France at seven locations. Charmet et al. (1993) showed for most characters there was significant environment interaction. In this case, mapping the regression coefficients allows plant breeders to identify the populations most tolerant to specific limiting factors of the environment. Genetics traits could change during cycles of regeneration or multiplication in different locations.

A considerable amount of work remains to be done to improve methods of selecting accessions for inclusion in core collections and to gain experience in their management and use. This will necessarily come from the experience of those who have undertaken the practical work of developing and testing such collections. There is a need to gain experience with a much wider range of crops including self and cross pollinated crops and clonally propagated ones. Developing core collections for this last group of crops raises specific questions of both a theoretical and practical nature since core collection theory is largely based on the application of population genetics to seed propagated crops. The extent to which data on quantitative traits can be used needs further study, as do the ways in which such data can be combined with that from biochemical or molecular studies. Ways of testing the extent to which the objective of representing diversity has been achieved also need to be improved.

In conclusion, the core collection provides a relatively simple and effective way of improving the accessibility of collections and the quality of information on the variation held in them. This is a prerequisite to improving the use of the plant genetic resources held in collections.



## REFERENCES

- Brown, AHD, 1989a. The case for core collections. In: Brown, AHD, Frankel, OH, Marshall RD and Williams, JT (eds.), *The use of Plant Genetic Resources*, pp. 136-156. CUP, Cambridge, UK.
- Brown, AHD, 1989b. Core collections: A practical approach to genetic resources management. *Genome*, 31: 818-824.
- Brown, AHD, 1994. The core collection at the crossroads. Proceedings of an IBPGR/CENARGEN/CGN Workshop on Core Collections, Brasilia, August, 1992 (in press).
- Bothmer, R von, Fischbeck, G, van Hintum, Th Hodgkin, T and Knupffer, H, 1990. The Barley Core Collection, 26 Sept., 1990.
- Charmet, G., Balfourier, F. Ravel, C. and Denis, JB. 1993. Genotype x environment interactions in a core collection of French perennial ryegrass populations. *Theor. Appl. Genet.* 86 : 731-736.
- Cordeiro, CMT, Morales, E, Ferreira, P, Rocha, DMS, Costa, I, Valois, A and Silva, S, 1994. Towards a Brazilian core collection for cassava. Proceedings of an IBPGR/CENARGEN/CGN Workshop on Core Collections, Brasilia, August, 1992 (in press).
- Crossa, J., Hernandez, CM., Bretting, P. Eberhart, SA. and Taba, S. 1993. Statistical genetic considerations for maintaining germ plasm collections. *Theor. Appl. Genet.* 86 : 673-678.
- Frankel, OH, 1984. Genetic perspectives in germplasm conservation. In: Arber, W, Llimensee, K, Peacock, WJ and Starlinger, P (eds.), *Genetic Manipulation: Impact on Man and Society*, pp. 161-170. CUP, Cambridge, UK.
- Frankel, OH and Brown, AHD, 1984. Current plant genetic resources - a critical appraisal. In *Genetics: New Frontiers*, Vol IV, pp 1-11. Oxford and IBH Publ. Co., New Delhi.
- Gepts, P, 1994. Genetic markers and core collections. Proceedings of an IBPGR/CENARGEN/CGN Workshop on Core Collections, Brasilia, August, 1992 (in press).
- Hamon, S. 1989. Characterization and evaluation of okra. In: Brown, AHD, Frankel, OH, Marshall RD and Williams, JT (eds.), *The use of Plant Genetic Resources*, pp. 173-196. CUP, Cambridge, UK.
- Hamon, S, Noirot, M, and Anthony, F, 1994. Suggested procedures for selecting a coffee core collection. Proceedings of an IBPGR/CENARGEN/CGN Workshop on Core Collections, Brasilia, August, 1992 (in press).
- Hintum (van), Th 1994. Hierarchical approaches to the analysis of genetic diversity in crop plants. Proceedings of an IBPGR/CENARGEN/CGN Workshop on Core Collections, Brasilia, August, 1992 (in press).
- Holden, JHW, 1984. The second ten years. In Holden, JHW and Williams, JT (eds.). *Crop Genetic Resources: Conservation and Evaluation*, pp. 277-285. George Allen & Unwin, London, UK.

- Kay, MC, 1994. One or many core collections ? Proceedings of an IBPGR/CENARGEN/CGN Workshop on Core Collections, Brasilia, August, 1992 (in press).
- Kimura, M. and Crow, JF 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49 : 725-738.
- Knüpffer, H and van Hintum, ThJL, 1994. The barley core collection - An international effort. Proceedings of an IBPGR/CENARGEN/CGN Workshop on Core Collections, Brasilia, August, 1992 (in press).
- Marshall, DR and Brown, ADH 1975. Optimum sampling in genetic conservation. In *Crop genetic Resources for today and tomorrow*, pp; 53-80, Frankel, O. and Hawkes (eds.). Cambridge University Press, Cambridge.
- Messmer, MM, Melchinger, AE, Lee, M, Woodman, WL Lee, EA and Lamkey, KR 1991. Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population : comparison of allozymes and RFLP data. *Theor. Appl. Genet.* 83 : 97-107.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *P.N.A.S., USA* 70 : 3321-3323.
- Noirot, M, Hamon, S. et F. Anthony 1993. L'obtention d'une core collection de caféiers. Définition des groupes d'échantillonnage et méthodologie. *Proc. 15ème Colloque ASIC, Montpellier, France*, pp 183-190.
- Peeters JP and Martinelli, JA, 1989. Hierarchical cluster analysis as a tool to manage variation in germplasm collections. *Theor. Appl. Genet.* 78: 42-48.
- Perry, MC, McIntosh, MS and Stoner, AK, 1991. Geographical patterns of variation in the USDA soybean germplasm collection. II. Allozyme frequencies. *Crop Sci.* 31: 1356-1360.
- Pickersgill, B. 1984. Evolution of hierarchical variation patterns under domestication and their taxonomic treatment. In Styles, BT (ed.) *Infraspecific classification of wild and cultivated plants*. Oxford, UK : Clarenton pp 191-209.
- Schoen DJ and Brown, AHD, 1994. Maximizing allelic diversity in core collections of wild crop relatives: the role of genetic markers. Proceedings of an IBPGR/CENARGEN/CGN Workshop on Core Collections, Brasilia, August, 1992 (in press).
- Spagnoletti-Zeuli PL and Qualset, CO, 1987. Geographical diversity for quantitative spike characters in a world collection of durum wheat. *Crop Sci.* 27: 235-241.
- Spagnoletti-Zeuli PL and Qualset, CO, 1993. Evaluation of five strategies for obtaining a core subset from large genetic resources collection of *Triticum durum*. *Theor. Appl. Genet.* 87 : 295-304.
- Tohme J, Jones, P, Beebe, S and Iwanaga, M, 1993. The combined use of agroecological and characterization data to establish the CIAT *Phaseolus vulgaris* core collection. Proceedings of an IBPGR/CENARGEN/CGN Workshop on Core Collections, Brasilia, August, 1992 (in press).

# EVALUATION AND EXPLOITATION OF GENETIC RESOURCES PRE-BREEDING



Photo INRA M. Pichon

PROCEEDINGS OF THE GENETIC RESOURCES SECTION

MEETING OF EUCARPIA



15 - 18 March 1994

CLERMONT-FERRAND FRANCE