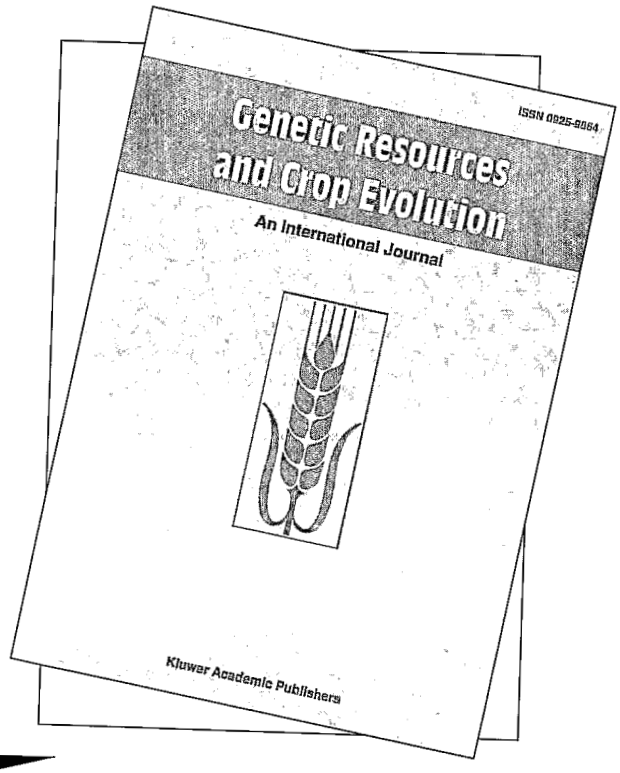
ISSN 0925-9864

OFFPRINT FROM



Genetic Resources and Crop Evolution

Genetic Resources and Crop Evolution is devoted to all aspects of plant genetic resources research. It publishes original articles in the fields of taxonomical, morphological, physiological, biochemical, genetical, cytological or ethnobotanical research of genetic resources and includes contributions to gene-bank management in a broad sense, that means to collecting, maintenance, evaluation, storage and documentation.
Areas of particular interest include:
- crop evolution
- domestication
- crop-weed relationships
- related wild species
- history of cultivated plants including palaeoethnobotany.

Genetic Resources and Crop Evolution also publishes short communications, e.g. newly described crop taxa, nomenclatural notes, reports of collecting missions, evaluation results of gene-bank material etc. as well as book reviews of important publications in the field of genetic resources. Every volume will contain some review articles on actual problems. The journal is the internationalized continuation of the German periodical *Die Kulturpflanze*, published formerly by the Institute of Plant Genetics and Crop Plant Research at Gatersleben, Germany. All contributions are in the English language and are subject to peer reviewing.

Genetic Resources and Crop Evolution is indexed/abstracted in *Current Awareness in Biological Sciences; CAB Abstracts; Biological Abstracts*

Kluwer Academic Publishers, P.O. Box 17, 3300 AA Dordrecht, The Netherlands

Genetic Resources and Crop Evolution **43:** 1–6, 1996.

The principal component scoring: A new method of constituting a core collection using quantitative data

M. Noirot[1], S. Hamon[1] & F. Anthony[2]
[1]*Laboratoire de Ressources Génétiques et d'Amélioration des Plantes Tropicales, ORSTOM, BP 5045, Montpellier Cedex 1, France;* [2] *Catié, Apartado 59, 5170 Turrialba, Costa Rica*

Received 3 August 1993; accepted 3 May 1994

Key words: genetic resources, core collection, quantitative traits, principal components

Abstract

The principal aims of a core collection are: 1) to include the maximum of diversity of the base collection in a sample of minimum size; and 2) to avoid redundancies. We present here a new method (the Principal Component Scoring) which fulfills these aims. The use of P.C.S. has consequences for sampling stratification and choice of sample size. P.C.S. requires quantitative data, but with small changes can be used for qualitative data.

Introduction

Vavilov (1935) was among the first to demonstrate the importance of gene banks in crop breeding. Some decades later, Harlan (1970), Frankel & Bennet (1970) and Pernès (1984) encouraged the evaluation of diversity of wild populations. Main crops, related wild species and minor crops, were collected on the initiative of IBPGR (International Board for Plant Genetic Resources). However, this routine collecting soon led to management and preservation problems. Many collections became very large, and therefore difficult to regenerate and maintain.

Frankel & Brown (1984) were the first to highlight the necessity of a collection of small size and maximal diversity: the "core collection". Size reduction inevitably leads to decreased diversity, but Brown (1989) shows that 10% of the base collection, and a maximum of 2000 to 3000 individuals, allow preservation of about 80% of alleles.

Most researchers currently believe that sampling should first be stratified, according to the organization of variability in groups and sub-groups (Frankel & Brown, 1984; Hintum, 1995; Yonezawa *et al.*, 1995). Usually, random sampling is used either for the whole of the base collection (in the rare cases for which the organization in groups is unknown or absent), or for each previously defined group. Such sampling has the advantage of being statistically representative of the base collection. In fact, for the great majority of users, the main goal of core collection is to avoid redundancies of genotypes (doubles). Redundancy is often related to the mode of reproduction. Rare in allogames, redundancy is very frequent in autogames and becomes common in apomictic species or in plants with vegetative multiplication.

Here, we propose a new sampling methodology – the principal component scoring (P.C.S.) – which effectively maximises sample diversity. An application to simulated data is described. In our sampling method, diversity is measured using quantitative variables, but the choice of the type of traits is discussed. The effects of the use of P.C.S. on sampling stratification and sample size are discussed.

Principles and methods

Choice of the distance, colinearity and weighting

Within population diversity is determined by between-individual differences for one or more traits. The differences can be estimated by a distance that, in our case, must be metric. The choice of distance depends on the kind of traits – qualitative or quantitative. In our

case, the use of quantitative traits leads to the selection of the Euclidian distance.

Generally, quantitative traits are of heterogenous type. Some of them are lengths (plant height, stem diameter, etc...), others are leafs (leaf area, pollen area, etc...), or weights (aerial biomass, reproductive biomass, reproductive biomass, etc...), or durations (date of flowering, fructification duration, etc.). In addition, quantitative traits have different variability. In order to give the same contribution (the same weight) to each trait j, the Euclidian distance is weighted by the reciprocal of the standard deviation σ_j. The distance d_{ik} between two individuals i and k for the J quantitative traits is defined by the following formula:

$$d_{ik} = \sqrt{\sum_{j=1}^{J} [(x_{ij} - x_{kj}) \cdot \sigma_j^{-1}]^2},$$

where x_{ij} and x_{kj} are the observed value of the trait j on the individuals i and k respectively.

The between-individuals distance is directly related to the number of differences. If differences belong to highly correlated traits (positively or negatively), this greatly overestimates the distance between some individuals relative to others. For example, if we measure the stem diameter of a tree at different heights (1 m, 1.10 m, 1.20 m, etc.) from the ground, the Euclidean distance between two trees will be highly influenced by the diameter differences. This example is obvious, but such effect, named colinearity, exists in all correlated traits.

In order to avoid the effect of variable colinearity, principal component analysis was applied to standardized data, and gave J new statistically independent and centred variables: the factors. The distance between two individuals i and k for the J factors is computed using a similar formula:

$$d_{ik} = \sqrt{\sum_{j=1}^{J} [(z_{ij} - z_{kj}) \cdot \sqrt{\lambda_j^{-1}}]^2},$$

where the square root of the λ_j eigenvalue allows weighting, and where z_{ij} and z_{kj} are the coordinates of the individuals i and k, respectively, on the factor j.

Such a procedure takes into account all factors with the same weight, including residual components – the result of chance or notation errors – in distance estimation. Removal of factors for which the eigenvalue is below one is arbitrary applied to eliminate this disadvantage.

Choice of individuals maximizing diversity

The generalized sum of square (*GSS*) of a set of N individuals in the factorial space of K standardized (mean = 0; variance = 1) and independent (correlation coefficient = 0) variables is equal to the product $N.K$ (Lebart et al., 1977). The contribution P_i of the individual i to the *GSS*, is equal to the sum of the squares of its K new coordinates:

$$P_i = \sum_{j=1}^{K} x_{ij}^2$$

The relative contribution CR_i of the individual i to the *GSS* of the set is given by:

$$CR_i = P_i/(N.K)$$

Preserving the greatest variability is equivalent to maximizing the score of the sub-set of sampled individuals using an *GSS* estimator. The first step consists in keeping the farthest individual of the set centre as initial sub-set, i.e. the individual with highest relative contribution. Iterative selection of individuals that maximize sub-set variability increases sub-set size and provides a core collection. At every iteration, the cumulative *GSS* of the sub-set (expressed in percentage of the total *GSS*) is known. The procedure can be stopped according to either the sub-set size or the percentage *GSS*. The two criteria can be simultaneously taken into account; in this case, the first criterion to be reached defines the stopping of sampling.

An example of application to simulated data

A data table comprising 20 normally-distributed variables and 2000 individuals was obtained by simulation. Eighteen of the variables were distributed into three independent groups within which correlations were high: V1 to V10, V11 to V15, V16 to V18. Variables V19 and V20 were completely independent. Figure 1 represents the between-variable correlations that were defined for simulation.

The correlation matrix is given in Table 1. The first six factors of P.C.A. show an eigenvalue greater than 1 (Fig. 2), and represent 79.5% of set *GSS*. The simulated structure of the data table can be seen in the three first factors. The latter represent the three groups of variables V1 to V10, V11 to V15 and V16 to V18, respectively.

GSS increasing of sub-set is apparent in Fig. 3. *GSS* sampled by P.C.S. is always greater than that sampled

Correlations computed from simulated data. Significant correlations are italicised

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
V20	–0.03	–0.04	0.02	–0.04	–0.04	0.02	–0.01	–0.01	0.00	0.01	–0.02	–0.01	–0.01	–0.02	–0.02	–0.02	–0.01	–0.00	–0.02	*1.00*
V19	–0.04	–0.04	0.02	–0.03	–0.04	0.05	0.06	0.04	0.04	0.02	0.03	0.03	0.00	0.04	0.00	0.01	–0.00	–0.03	*1.00*	
V18	–0.05	–0.04	0.04	–0.01	–0.02	0.01	–0.00	–0.01	–0.03	–0.02	–0.01	–0.01	–0.00	0.00	–0.03	*0.73*	*0.85*	*1.00*		
V17	–0.05	–0.03	0.04	–0.02	–0.02	0.01	0.01	0.00	–0.02	–0.01	–0.01	–0.01	–0.01	0.01	–0.03	*0.86*	*1.00*			
V16	–0.03	–0.02	0.02	0.00	–0.01	0.01	0.00	–0.00	–0.01	–0.01	–0.01	–0.00	–0.01	–0.00	–0.03	*1.00*				
V15	–0.03	–0.03	0.02	–0.02	–0.03	0.02	0.01	0.01	0.03	0.01	*0.43*	*0.38*	*0.85*	*0.36*	*1.00*					
V14	–0.01	–0.01	0.01	0.03	–0.01	0.01	0.00	–0.00	0.01	0.00	*0.85*	*0.72*	*0.42*	*1.00*						
V13	–0.02	–0.01	0.03	0.00	–0.02	0.01	–0.00	–0.00	–0.01	0.00	*0.50*	*0.42*	*1.00*							
V12	–0.01	–0.01	0.01	0.00	–0.02	0.02	0.00	0.00	0.01	–0.01	*0.85*	*1.00*								
V11	–0.00	–0.01	0.01	0.01	–0.02	0.00	–0.01	–0.00	0.00	–0.00	*1.00*									
V10	*–0.30*	*–0.38*	*0.26*	*–0.14*	*–0.41*	*0.50*	*0.60*	*0.70*	*0.84*	*1.00*										
V9	*–0.38*	*–0.47*	*0.34*	*–0.19*	*–0.52*	*0.61*	*0.71*	*0.84*	*1.00*											
V8	*–0.46*	*0.56*	*0.39*	*–0.23*	*–0.62*	*0.73*	*0.85*	*1.00*												
V7	*–0.53*	*–0.65*	*0.45*	*–0.26*	*–0.72*	*0.85*	*1.00*													
V6	*–0.61*	*–0.76*	*0.51*	*–0.29*	*–0.84*	*1.00*														
V5	*0.72*	*0.89*	*–0.60*	*0.35*	*1.00*															
V4	*0.52*	*0.40*	*–0.44*	*1.00*																
V3	*–0.84*	*–0.66*	*1.00*																	
V2	*0.80*	*1.00*																		
V1	*1.00*																			

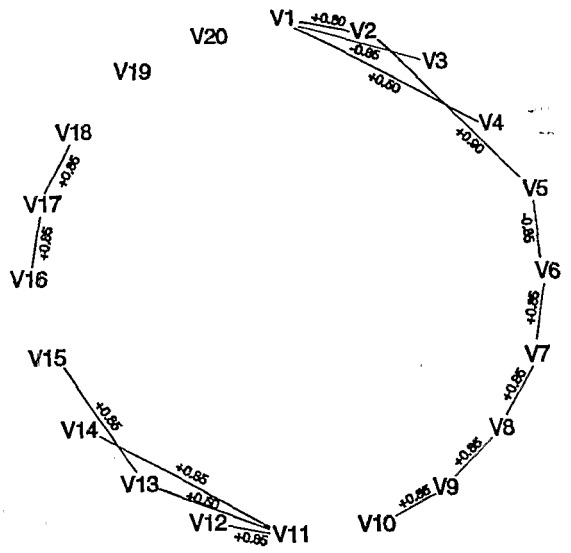


Fig. 1. Theoretical correlations between variables (before simulation).

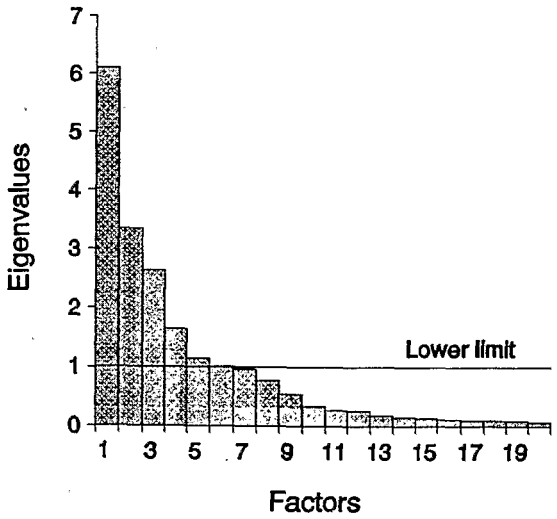


Fig. 2. Eigenvalue distribution.

randomly. Hence, 10% of individuals represent 22.4% of *GSS*, whereas 50% of *GSS* includes less than 30% of individuals. The slope of the curve decreases progressively, and the addition of new elements brings less and less diversity. Thus, the 500 later individuals (25% of the set size) contribute less than 10% of *GSS*.

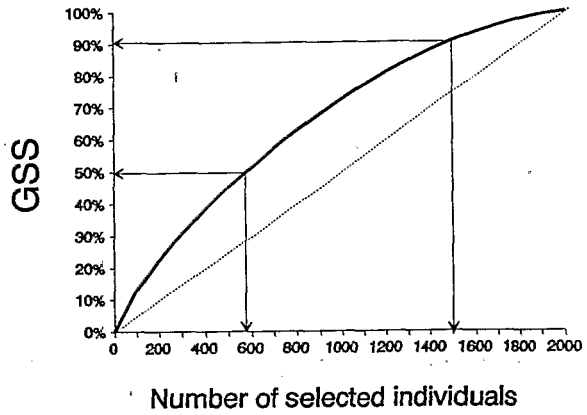


Fig. 3. Sampled *GSS* as a function of sample size. The dotted line represents the expected trend with a random sampling.

Discussion

Aims of the core collection

The main aim of core collections (Frankel & Brown, 1984) is to include the maximum genetic diversity, with a minimum of redundancies, of the crop and its related species in a sample of minimal size (Brown & Clegg, 1983). The simplest method for creating a core collection is random sampling of the whole of base collection. When the genetic structure of the base collection is unknown, such sampling represents the best solution (Brown, 1989). However, its efficiency is low for alleles that are locally common but rare in the whole base collection. The stratification of sampling improves efficiency (Frankel & Brown, 1984). Nevertheless, random sampling within groups does not achieve the first aim of a core collection, i.e. to sample the maximal diversity. In addition, random sampling does not avoid doubles in autogames, apomicts and species with vegetative multiplication. Our method modifies principally the mode of sampling, which is no longer random, but which is designed to maximize diversity and avoid doubles. The method meets the expected aims of a core collection.

Implied conditions of our mode of sampling. Consequences on stratification

By preferential sampling of the farthest individuals, the method requires two conditions to be functional and effective. The more important condition assumes that all the individuals of the sub-set can breed together to give intermediate types. This requires previous

knowledge of the genetic structure of the complex of species (Pernès, 1970) and a good estimation of the level of reproductive barriers between pools. It therefore implies the use of sampling stratification, as in the case of random sampling.

According to Frankel & Brown (1984), Hintum (1993) and Yonezawa *et al.* (1993), the construction of groups includes in the following order: the geographical origin, the systematics, to isozymes (i.e. genetic markers), qualitative and quantitative traits, and finally biogeographical and bioclimatic information. In our case, the stratification of sampling must depend on the genetic structure of populations and on recombination limits. When such data are lost or absent, systematics should be taken into account. The bioclimatic and biogeographical information should then improve the structure by establishing sub-groups corresponding to genetic differentiation (sub-species, ecotypes). Principal component scoring is applied within the sub-groups and groups.

The second condition affects the efficiency of P.C.S. In fact, efficiency assumes generalized additivity and strong heritability *s.l.* for quantitative traits. In the cases for which this assumption is not valid (dominance, superdominance, high plasticity, etc.), the selection of phenotypic diversity does not necessarily lead to selection of genetic diversity. The sampling is then considered as random for the hidden genetic variability.

Traits used for evaluating group diversity

For estimating group diversity, Frankel (1974) distinguished two types of characters from the complexity of their inheritance. The first type includes characters with known inheritance, and for which the identification of genotype from phenotype is relatively simple (resistance to diseases, colouration, etc.). Isozymes also belongs to the group, and are often used in diversity estimation, but their use is debatable: 1) enzymatic diversity does not necessarily correlate with morphological diversity (Davis & Gilmatin, 1985); 2) its economic value is often minor; 3) morphological divergence may precede enzymatic divergence (Crawford, 1985), as in the case of *Coffea arabica*. Finally, enzymatic diversity measures phylogeny more than adaptation.

The second type includes traits of unknown, and often polygenic, inheritance. Symbolized by quantitative data, the traits have the main disadvantage of measuring factors other than genetic diversity, such as its interaction with the environment (phenotype) (Frankel, 1974). Nevertheless, quantitative polymorphism reflects the diversity of selective conditions rather than phylogeny. Quantitative traits are therefore of major interest for the user of genetic resources. In addition, in numerous situations, and especially in tropical plants, quantitative traits (height, flowering day, production, fruit size, etc.) are the only ones for which we have notations.

The size of the core collection and its sub-groups

Core collection size has long been under discussion (Frankel & Brown, 1984). Using neutral alleles theory (Kimura & Crow, 1964) and sampling theory, Brown (1989) showed that 10% of the base collection includes at least 80% of alleles, with a statistical error risk of 5%. According to the author, the results are robust towards the type of frequency distribution of alleles at each locus. This value of 10% is not modified by our mode of sampling.

The size of each sub-set within the core collection has also been studied (Brown, 1989). Three methods were proposed: 1) the same number of accessions per group; 2) a number proportional to the group size, and 3) a number proportional to the logarithm of the group size. The author showed that the third solution constitutes a good compromise. Nevertheless, the choice of the size of sub-sample from the size of group assumes that a relation exists between group diversity and size. This is far from always the case, and the diversity/size ratio of the group depends on its mode of reproduction (highly probable presence of doubles in the autogames and apomicts) and on its economic importance (crop or related species). This is particularly striking in the species complex (Pernès, 1984) of the genus *Coffea*, for example, where autogamous crop *C. arabica* with low variability (historical bottleneck) is over-represented in base collections compared with the allogamous and more variable species from East Africa *C. sessiliflora* (Noirot *et al.*, 1993).

The sampling method we propose here allows monitoring the trend in sub-set diversity as sampling increases (Fig. 3). The size choice truly depends on diversity. The method is particularly efficient in groups with high redundancy (apomictic populations, for example).

Conclusions

The main advantage of P.C.S. is to improve the percentage of sampled diversity without modifying the relative intensity of the selection (10%) proposed by Brown (1989). This is particularly true for autogamous plants, apomicts and species with vegetative reproduction. This should allow fulfilment of the expected qualities of a core collection: maximum diversity and minimum redundancy. Note also that the mode of sampling we propose affects the choice of the stratification characteristics and the sample size in each group.

Another advantage results from the use of quantitative traits. By contrast with the use of neutral qualitative traits which symbolize principally the phylogenetic diversity, quantitative traits should represent adaptative traits. However, this reflexion must not reject all qualitative traits. Indeed, the current progress in molecular biology should lead to the increasing availability of non-neutral qualitative markers, such as the quantitative traits loci (Q.T.L.) (Lander & Botstein, 1989). Use of these markers in the estimation of genetic diversity will then increase. In this case, P.C.S. may be modified for application to qualitative variables. Changes will affect 1) the type of distance, which must still be a metric (simple matching, Gower's distance (Gower, 1966) or χ^2 distance (Benzécri, 1972); 2) the method of data reduction (correspondence analysis (Benzécri, 1972) or principal coordinates (Gower, 1966)), and the weighting of factors in the *GSS* calculation.

Nevertheless, P.C.S. introduces an obvious statistical bias in estimating and infering from core collections the statistical and genetic parameters of the population comprising the core collection. However such biases very often also exists in the collecting method, implicitly for wild populations and obviously for crops.

Acknowledgment

We thank anonymous referees for their judicious comments.

References

Benzécri, J. P., 1972. Pratique de l'analyse des données: Analyse des correspondances. Dunod, Paris, 424 pp.

Brown, A. D. H., 1989. Core collections: a practical approach to genetic resources management. Genome 31: 818–824.

Brown, A. D. H., 1989. Size and structure of collection: the case for core collection. In: A. D. H. Brown *et al.* (Eds), The use of plant genetic resources pp. 136–156, Camb. Univ. Press, Cambridge.

Brown, A. D. H. & M. T. Clegg, 1983. Isozyme assessment of plant genetic resources. In: Rattazzi *et al.* (Eds), Isozymes: Current topics in biological and medical research (vol. 11) pp. 285–295, Alan R. Liss, New York.

Crawford, J. L., 1985. Electrophoretic data and plant speciation. Syst. Bot. 10(4): 405–415.

Davis, J. I. & A. J. Gilmartin, 1985 Morphological variation and speciation. Syst. Bot. 10(4): 416–425.

Frankel, O.H., 1974. Genetic conservation: our evolutionary responsibility. Genetics 70(1): 53–65.

Frankel, O.H. & E. Bennett, 1970. Genetic Resources in Plants. Their exploration and conservation. Blackwell Scientific Pub. Oxford.

Frankel, O.H. & A. D. H. Brown, 1984. Current plant genetic resources – A critical appraisal. In: Genetics New Frontiers. Proc. of XV Int. Congress of Genetics. Oxford & IBH Publishing Co (4): 3–13.

Gower, J. C., 1966. Some distance properties of latent and vector methods used in multivariate analysis. Biometrika 53: 325–328.

Harlan, J. R., 1970. The evolution of cultivated plants. In: O. H. Frankel & E. Bennet (Eds), Genetic Resources in Plants. Their exploration and conservation, pp. 19–32, Blackwell Scientific Publ. Oxford.

Hintum, T. J. L. van, 1995. Hierarchical approaches to the analysis of the genetic diversity of crop plants. In: T. Hodgkin, A. H. Brown, J. L. H. van Hintum and E. A. V. Morales (Eds), Chichester, UK, John Wiley & Sons, Core collections of plant genetic resources pp 23–34.

Kimura, M. & J. F. Crow, 1964. The number of alleles that can be maintained in a finite population. Genetics, 49: 725–738.

Lander, E. S. & D. Botstein, 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics, 121: 185–199.

Lebart, L., Morineau, A. & N. Tabard, 1977. Techniques de la description statistique. Méthodes et logiciels pour l'analyse des grands tableaux. Dunod, Paris, 352 pages.

Noirot, M., Hamon, S. & F. Anthony, 1993. L'obtention d'une Core collection de caféiers. Définition des groupes d'échantillonnage et méthodologie. XVIth Congress of ASIC, Montpellier 183–190.

Pernès, J., 1984. Gestion des ressources génétiques des plantes. A.C.C.T. Paris, 2 tomes.

Vavilov, N. I., 1935. The origin, variation immunity and breeding of cultivated plants. Chron. Bot. 13, 6 volumes.

Yonezawa, K., Nomura, T. & H. Morishima, 1993. Sampling strategies for use in stratified germplasm collections. T. Hodgkin, A. H. Brown, J. L. H. van Hintum and E. A. V. Morales (Eds), Chichester, UK, John Wiley & Sons, Core collections of plant genetic resources pp 35–53.

For information about current subscription rates and prices for back volumes for
Genetic Resources and Crop Evolution, ISSN 0925-9864
please contact one of the customer service departments of Kluwer Academic Publishers or return the form overleaf to:

Kluwer Academic Publishers, Customer Service, P.O. Box 322, 3300 AH Dordrecht, the Netherlands, Telephone (+31) 78 524 400, Fax (+31) 78 183 273, Email: services@wkap.nl

or

Kluwer Academic Publishers, Customer Service, P.O. Box 358, Accord Station, Hingham MA 02018-0358, USA, Telephone (1) 617 871 6600, Fax (1) 617 871 6528, Email: kluwer@world.std.com

Call for papers

Authors wishing to submit papers related to any of the themes or topics covered by *Genetic Resources and Crop Evolution* are cordially invited to prepare their manuscript following the 'Instructions for Authors'. Please request these instructions using the card below.

Author response card

Genetic Resources and Crop Evolution

I intend to submit an article on the following topic:

_____

_____

Please send me detailed 'Instructions for Authors'.

NAME : _____

INSTITUTE : _____

DEPARTMENT : _____

ADDRESS : _____

_____

_____

Telephone : _____

Telefax : _____

Email : _____

Library Recommendation Form

Route via Interdepartmental Mail

To the Serials Librarian at: _____

From: _____ Dept./Faculty of: _____

Dear Librarian,

I would like to recommend our library to carry a subscription to
Genetic Resources and Crop Evolution, ISSN 0925-9864
published by Kluwer Academic Publishers.

Signed: _____ Date: _____

Request for information about current subscription rates and prices for back volumes of

Genetic Resources and Crop Evolution, ISSN 0925-9864

Please fill in and return to:

Kluwer Academic Publishers, Customer Service, P.O. Box 322, 3300 AH Dordrecht, the Netherlands

Kluwer Academic Publishers, Customer Service, P.O. Box 358, Accord Station, Hingham MA 02018-0358, USA

☐ Please send information about current program and prices
☐ Please send a free sample copy

NAME : _____
INSTITUTE : _____
DEPARTMENT : _____
ADDRESS : _____
Telephone : _____
Telefax : _____
Email : _____

REF. OPC

STAMP

Genetic Resources and Crop Evolution
Kluwer Academic Publishers,
P.O. Box 17
3300 AA Dordrecht
The Netherlands

REF. OPC

TO : The Library
FROM: _____

VIA INTERDEPARTMENTAL MAIL

Reçu le 20/3/96

LISTE DE DIFFUSION et FICHE DESCRIPTIVE

ORSTOM	ORSTOM	ORSTOM
70-74 Rte d'Aulnay	**214 rue Lafayette**	**2051 av. du Val de Montferrand**
93143 BONDY	**75480 PARIS cedex 10**	**34032 MONTPELLIER cedex**
Cellule de collecte FDO	Département MAA	UR MAA 3A
Base Horizon	Commission scient. 4	
3 exemplaires	1 ex. + 1 ex. expédiés séparément	1 exemplaire

Auteur(s) : **M. NOIROT, S. HAMON, F. ANTHONY**

Titre original : The principal component scoring : A new method of constituting a core collection using quantitative data

Titre en français : Méthodologie d'échantillonnage pour la constitution d'une Core collection
Un nouveau modèle pour les données quantitatives : la P.C.S.

Mots clés : Ressources génétiques, core collection, caractère quantitatif, analyse de composant principal

Résumé en français :

Une nouvelle méthode, la P.C.S. (Principal Component Score), a été élaborée pour répondre aux objectifs des core collection, lorsqu'il existe une évaluation de la collection de base avec des descripteurs quantitatifs. Elle consiste dans un premier temps à effectuer une analyse en composantes principales sur données centrées réduites, à éliminer les composantes d'inertie inférieure à 1 et à donner le même poids à chacune des composantes retenues dans le calcul des distances entre individus. Dans un deuxième temps, nous retenons par ordre décroissant les individus dont la contribution à la variabilité générale est la plus importante. L'utilisation de la P.C.S. a des conséquences sur la stratification d'échantillonnage et le choix de la taille de l'échantillon. La P.C.S. a été élaborée pour des données quantitatives et des modifications sont envisagées afin d'utiliser également des données qualitatives.

Les titres, mots-clés matières et résumés en Anglais sont indispensables pour les documents destinés à entrer dans les Bases AGRIS et ASFA (Aquatic Sciences and Fisheries Abstracts).