

## CHAPITRE VII

### "CORE COLLECTIONS" : SITUATION ET PERSPECTIVES

S. HAMON, M. NOIROT et F. ANTHONY

Laboratoire de Ressources Génétiques et Amélioration des Plantes Tropicales  
Centre ORSTOM, BP 3045, 34032 Montpellier Cedex.

L'utilisation des Ressources Phytogénétiques par les scientifiques peut se situer à différents niveaux : recherche de base, stratégique ou appliquée. Divers constats montrent que le niveau réel d'emploi est inférieur à 1 % du contenu des collections. De nombreux facteurs, plus ou moins complexes, se superposent pour aboutir à cette situation. A la lecture des premiers chapitres de cet ouvrage, consacré au Mil en Afrique, nous constatons que de nombreuses prospections ont été effectuées et que plusieurs équipes ont réalisé des travaux d'évaluation de la diversité génétique. C'est peut être le moment de réfléchir à une stratégie de Core Collection (CC) pour cette plante. Nous n'avons pas ici la prétention de fournir une recette miracle mais de suggérer des éléments de réflexion si un tel projet se met en place.

#### I - LES BESOINS ET LE CONCEPT

Les années 70 et 80 ont été des décennies où la plupart des pays, aussi bien du nord que du sud, ont pris conscience de l'intérêt des Ressources Phytogénétiques. L'ORSTOM pour sa part a fait de nombreuses prospections et évaluations (Charrier et Hamon 1991). Malheureusement, souvent devant l'urgence, la stratégie a été de favoriser la collecte du plus grand nombre d'échantillons possible au détriment de la qualité du lot ou de l'information associée. Ainsi, dans bien des cas, peu d'informations sont disponibles aussi bien au niveau des données du passeport que de l'évaluation (Peeters et Williams, 1984). Dans un premier temps, la grande majorité des collectes a concerné les formes cultivées des plantes économiquement importantes, principalement les céréales. Ces espèces, aux graines orthodoxes, conservent leur pouvoir germinatif sans problème pendant plusieurs décennies (IBPGR, 1985). Par la suite, les institutions ont élargi leur champ d'action en s'intéressant aux espèces apparentées sauvages et aux autres groupes d'espèces dites mineures (IBPGR, 1980). Les introductions ont été conservées soit dans des chambres froides soit dans des congélateurs ce qui était une bonne solution de moyen terme. Par contre, la conservation à long terme de ce patrimoine, surtout lorsqu'il s'agit de régénérer ou d'évaluer les échantillons, s'accompagne d'une importante augmentation des coûts. Parallèlement, il faut définir des stratégies de régénération, quelquefois délicates à mettre en oeuvre aussi bien du point de vue pratique que théorique.

Enfin un fossé d'incompréhension a longtemps existé entre les "curators", gestionnaires de collections, et les utilisateurs potentiels c'est à dire les scientifiques, les sélectionneurs, les Instituts de recherche ou de développement. En définitive, les grandes collections de Ressources Phytogénétiques demeurent peu performantes et sous utilisées car faiblement documentées. En particulier, elles ne peuvent que très rarement répondre à une demande précise.

La notion de collections contenant un minimum de redondance, une bonne représentation de la diversité génétique pour un ensemble, nommé complexe d'espèces au sens de Pernès et Lourd (1984), remonte au congrès de génétique de New Dehli (Frankel et Brown, 1984). Peu après apparaissait le terme de Core Collection (CC). En 1990 seule une dizaine de complexes d'espèces avait fait l'objet de réflexion dans ce sens (Hodkings, 1991). Les attendus principaux de la Core Collection sont de quatre ordres : permettre une meilleure caractérisation, faciliter la recherche de caractères désirés, obtenir une vision globale de la diversité génétique et donner des indications claires sur l'orientation des recherches d'échantillons dans la collection de base.

## II - LES BASES DE L'ECHANTILLONNAGE

Dans toute procédure d'échantillonnage il est nécessaire de définir les critères de sélection et la méthodologie. Hormis la procédure totalement aléatoire, nous évoquons

De quels caractères disposons-nous pour effectuer un échantillonnage raisonné ? Ce que l'on appelle communément le passeport contient en principe des informations acquises pendant la prospection sur l'origine, le site de collecte et le statut des échantillons établis. La caractérisation comporte des informations sur la taxonomie, les marqueurs génétiques, morphologiques, moléculaires et agronomiques. On consultera utilement Peeters et Williams (1984) pour se faire une idée de la situation dramatique des collections dans ce domaine au milieu des années 80. Il est évident que le généticien ou le sélectionneur souhaite pouvoir disposer de données d'évaluation exprimées en termes génotypiques ou de fréquences alléliques. Au fil des années on a vu apparaître de nombreux articles scientifiques basés sur les données de type isoenzymatique (décennie 70 - Gottlieb, 1981) puis moléculaire (décennie 80 - REF). Ceux-ci traitent de divers aspects (diversité génétique, évolution, phylogénie,.....). Il n'y a toutefois pas de collection de Ressources Génétiques entièrement décrites de cette manière. Par contre, plus nombreuses sont celles qui ont fait l'objet de description, plus ou moins partielles, sur la base de caractères morphologiques soit qualitatifs soit quantitatifs.

### III - SELECTION DES INDIVIDUS

La première réflexion qui doit être menée concerne l'existence d'une évaluation et le niveau de fiabilité des données. Ensuite il faut pouvoir estimer la pertinence des regroupements effectués c.a.d. la valeur du degré de similarité (distance) afin de comparer des groupes comparables. Pour faire notre sélection nous disposons en fait de 4 grandes possibilités d'approche : l'aléatoire complet, le choix statistique sur les fréquences alléliques, l'utilisation de descripteurs particuliers, une analyse approfondie de la diversité. Le tout doit être associé à une petite dose de pragmatisme. Dès les premiers instants de réflexion les questions seront donc les suivantes : 1- De quelles connaissances botaniques, agronomiques, morphologiques, cytologiques, isozymiques, moléculaires, disposons-nous pour réaliser le choix ? 2- Assurer une couverture géographique ou écologique ? 3- doit-on conserver quelques échantillons de chaque espèce ?

Si on se réfère à la première définition d'une Core Collection deux points principaux sont évoqués : - maximum de diversité, - un nombre limité d'accessions. Quelle signification et comment interpréter ces termes ? Nous verrons ci-dessous, que lorsqu'il s'agit de définir la ligne à suivre plusieurs points de vue, apparemment antagonistes, se font jour. Brown (1989a, b), adepte d'une sélection neutraliste, base son raisonnement sur une sélection d'allèles à fréquence faible mais largement répandus. Galwey (1993) considère que l'on doit accorder une place plus importante à la diversité phénotypique, la diversité neutre n'ayant aucun intérêt en amélioration des plantes. Ces deux aspects sont à prendre en compte et la disponibilité d'informations sur les liaisons génétiques entre marqueurs moléculaires et caractères quantitatifs (QTL) seront évidemment d'un grand intérêt. L'expérience montre que la structure des populations peut être modulée en fonction de déterminants écologiques (Loveless et Hamrick, 1984). Tout récemment

Erskine et Muehlbauer (1991) proposent de prendre en compte le mode de reproduction, la diversité allozymique et la diversité morphologique pour constituer la CC de lentille. Enfin, pour certains auteurs, généralement proches des sélectionneurs, la réflexion va jusqu'à se demander s'il faut créer une ou plusieurs CC sur la base de caractéristiques différentes.

### 1) Choix sur la fréquence des allèles

Brown (1989a), auquel nous renvoyons pour plus de détails sur la procédure de calcul employée, propose une méthode qui conduit à retenir 10 % de l'effectif total d'une collection avec un maximum de 3000 accessions. La base théorique de la démarche se base sur la théorie du neutralisme de Kimura et Crow (1964). On considère que tout allèle peut se placer dans l'une des 4 catégories suivantes : Allèle rare ( $< 0,10$ ) : C1 = Commun ; C2 = Endémique / Allèle fréquent ( $> 0,10$ ) : C3 = Commun ; C4 = Endémique. Pour les allèles rares communs, l'auteur montre que pour 100 locus également polymorphes, un échantillonnage de 10 % permettra de retenir 70 % des allèles dans 95 % des cas. Un modèle basé sur des différences de polymorphisme et sur un échantillonnage sans remise montre qu'avec 3 000 individus dans la core collection on acheminera la conservation de 2,5 allèles par locus.

La première formulation de la CC considérait la collection dans son intégralité. Très rapidement un consensus est apparu sur la nécessité de procéder par hiérarchisation et de définir des groupes. Plusieurs suggestions ont été faites par Brown (1989b). Deux stratégies extrêmes peuvent être proposées. Dans la première, dite Constante (C), on prend un nombre équivalent d'accessions quelle que soit la taille initiale du groupe. Dans la seconde, dite proportionnelle (P), on sélectionne une fraction identique de l'effectif initial pour constituer chaque groupe. Ces deux méthodes sont clairement biaisées : (C) introduit un biais pour les petits groupes, (P) un pour les grands. L'auteur propose un échantillonnage logarithmique (L) où un groupe est représenté par le log de son effectif.

### 2) Choix sur les caractères quantitatifs

Le travail sur la base de caractères quantitatifs peut se révéler très intéressant pour construire une Core collection à conditions de prendre quelques précautions élémentaires et d'effectuer l'analyse au sein d'un groupe de diversité bien identifié. On peut ainsi y appliquer le principe d'additivité généralisé pour l'ensemble des caractères et considérer qu'il n'y a pas de barrière reproductive entre les individus. En d'autres termes le croisement entre deux individus éloignés sur un plan factoriel ou une classification est non seulement possible mais est susceptible de redonner dans les descendance des individus de phénotype intermédiaire.

Hamon *et al.* (1993) proposent une méthode appliquée au complexe d'espèces des caféiers. Soit un ensemble d'individus (génotypes) obéissant aux conditions énumérées ci-dessus et qui a été caractérisé par un nombre  $j$  de variables quantitatives. On a ainsi un tableau  $N_i * j$ . La distance retenue est la distance euclidienne. La colinéarité des variables est éliminée en pratiquant une analyse multivariée en composantes principales. Si on

considère l'ensemble du nuage de points, chaque individu contribue à l'inertie totale par sa contribution relative (CRI). Le modèle de sélection consiste à retenir les individus qui maximisent la somme des CRI. On s'aperçoit ainsi que l'on peut obtenir 90 % de la variabilité avec 50 % des individus. Cette méthode est compatible avec la démarche neutraliste mais permet d'optimiser la représentation de la diversité morpho-physiologique.

### 3) Choix sur les caractères qualitatifs

La plupart des discussions sur les Core Collections ont été centrées sur la conservation des allèles. Ceci est utile d'un point de vue théorique, le seul inconvénient est que ce type d'information est rarement connu pour l'ensemble d'une collection. Par ailleurs, il y a au moins deux types de plantes difficiles à appréhender de ce point de vue : - les plantes à multiplication uniquement végétative (i.e. ignames) où la valeur d'un génotype ne s'exprime pas en fréquence allélique et les phénotypes rares peuvent être importants ; - les plantes allogames où les groupes de liaison ne correspondent pas simplement à des zones susceptibles de recombinaison. Il faut donc souvent considérer les descripteurs et leur état.

Galwey (1993) propose, pour estimer la diversité d'un caractère discret, l'index de Shannon (H) tel que :  $H = - \sum (i) \log p(i)$  où  $p(i)$  = proportion des accessions ayant l'état (i) du descripteur. L'échantillonnage peut se faire ensuite suivant les modes P, C ou L définis précédemment. Il suggère également que la core soit basée sur des groupes ayant une amplitude de diversité équivalente.

## IV - QUELQUES EXEMPLES DE CORE COLLECTIONS

L'adoption d'un mode de pensée, la référence à des modèles n'impliquent pas forcément, loin s'en faut, une démarche identique face à la réalité. A titre d'illustration nous donnons ci-dessous quatre exemples de Core Collections en cours de réalisation. Elles ont fait l'objet de présentation au cours du symposium international organisé sur ce thème à Brasilia (Brésil) en août 1992 par l'IBPGR et le CENARGEN.

### A - La core collection européenne d'orge (Knüpfner, 1993).

L'orge est une Poacée (ex graminée) économiquement très importante, bien connue au niveau génétique. Les formes cultivées dérivent d'une espèce diploïde (*Hordeum vulgare*) mais les pools secondaire et tertiaire renferment une quarantaine d'espèces plus ou moins spontanées. En 1989, le groupe de travail européen du réseau d'échange de ressources génétiques (ECP / GR) a décidé de réfléchir sur la construction d'une core collection d'orge. L'échéancier adopté est le suivant : 1- (1989-92) réflexion sur le projet ; 2 - (1992-94) établissement de la core ; 3- (1995) BCC opérationnelle.

Les responsables ont choisi de considérer la diversité de cette plante d'une manière hiérarchique. La Core Collection d'orge (BCC) contiendra : 500 cultivars et 800 variétés traditionnelles (landraces) de *H. vulgare* ; 150 à 200 de *H. spontaneum* dont 2/3 originaires du centre de la zone de distribution et le reste pris sur les marges ; 60 à 100 échantillons à raison de 2 à 3 génotypes pour les autres espèces.

**B - La core collection de caféiers à l'ORSTOM (France) (Hamon *et al.* 1993)**

Se basant sur la disponibilité des espèces et sur la connaissance actuelle de la diversité génétique du complexe des caféiers, l'ORSTOM a décidé de construire une CC de caféiers. La démarche retenue pour cette plante pérenne tropicale diffère sensiblement de celle évoquée pour l'orge dans la mesure où elle prend beaucoup plus en compte les espèces spontanées. L'espèce cultivée principale, *C. arabica*, est traitée séparément en raison de son mode de reproduction original (autogamie). L'espèce cultivée *C. canephora*, employée en culture depuis relativement peu de temps, est considérée au même niveau que les espèces spontanées.

Sur la base des données d'évaluation qui prennent aussi bien en compte les possibilités de croisement, les caractères phénologiques que les descripteurs isoenzymatiques 88 groupes de diversité correspondant à 64 espèces botaniques ont été retenus. Pour chaque groupe de diversité un échantillonnage en collection (bulk avec une à deux graines par génotype de la collection de base) a été effectué. Les graines ont été

Colombie, conserve environ 24 000 échantillons de haricot commun (*Phaseolus vulgaris*). Des introductions en provenance de zones sans intérêt sont sur-représentées et des échantillons qui pourraient se révéler nécessaires font cruellement défaut.

La stratégie suivante a été proposée pour constituer une core de 1400 accessions : 1200 du centre primaire de diversité, 200 des centres secondaires, et 25 variétés sélectionnées. La core collection des formes spontanées sera considérée ultérieurement. La sélection est basée sur la compréhension actuelle de l'organisation des pools majeurs avec trois niveaux de prise en compte : 1 - Poids arbitraire en fonction de l'appartenance à un site primaire ou secondaire ; 2 - Caractères morpho-physiologiques (habitat, couleur et forme des graines en prenant les caractéristiques de chaque site Mésoaméricain et Andin) ; 3 - Prise en compte des données sur l'écologie de la zone d'origine.

## V - CONCLUSIONS ET PERSPECTIVES

Les grandes collections de Ressources Phytogénétiques, dites collections de base, sont indispensables mais posent des problèmes de coût en regard à l'utilisation qui est faite du matériel conservé. Il est nécessaire pour mieux les valoriser de mettre en place parallèlement des collections de dimensions plus réduites représentatives de la diversité génétique appelées Core Collections. Le choix des entrants en CC peut se faire sur différents types de caractères tels que l'origine géographique des échantillons, la caractérisation du matériel végétal, les résultats d'une évaluation approfondie. Toutefois, pour un ensemble de plantes apparentées et (ou) faisant partie d'un même complexe d'espèces, la diversité génétique est organisée. Il faut en tenir compte dans la conception de la collection.

Il semble que pour encourager l'utilisation des ressources génétiques par les sélectionneurs, la core collection doit être établie par le "curator" et représenter la diversité en collection. Lors de l'introduction d'un nouveau matériel en collection, la question se pose dans les termes suivants : la nouvelle accession introduit-elle une nouvelle potentialité ou y a-t-il un risque de redondance ? Que faire d'une introduction : va-t-elle en CC ou en collection de base ? Il est bien entendu souhaitable que les échantillons utilisés pour la CC soient si possible de première main et qu'ils soient multipliés en fonction de leur structure génétique et du mode de reproduction propre à l'espèce.

Bien que la description de la diversité soit souvent loin d'être idéale, un choix raisonné est préférable au pur hasard. Des considérations statistiques sur les isoenzymes suggèrent que 10 % de l'effectif avec un maximum de 3.000 échantillons sont un bon compromis. Les considérations statistiques sur les données quantitatives recommandent une proportion plus importante. Enfin, la CC est bien entendu évolutive et sa composition peut changer en fonction de nouvelles introductions ou de nouvelles données. La hiérarchisation de la structure pourra ainsi être modifiée pour tenir compte de la révision des catégories affines ou d'une modification des priorités et des besoins.

## BIBLIOGRAPHIE

- BERNATSKY et TANKSLEY, 1989. PEACOCK, 1989 *In the use of plant genetic resources*. Brown, Frankel, Marshall, Williams eds.
- BROWN (A. H. D.), 1989a. The case for core collections. *In The use of plant genetic resources.*, A. H. D. Brown, O. H. Frankel, D. R. Marsahll et J. T. Williams Edts. Cambridge University Press : 136-156.
- BROWN (A. H. D.), 1989b. Core collections : a practical approach to genetic resources management. *Genome* 31 : 818-824.
- CHARRIER (A.) et HAMON (S.), 1991. Les activités de collecte, de conservation et d'utilisation du germplasm à l'ORSTOM). *In Crop Genetic Ressources of Africa*, vol II ; Ng, P. Perrino, F. Attere & H. Zedan Edts : 41 à 53.
- ERSKINE (W.) and MUEHLBAUER (F.J.), 1991. Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. *Theor. Appl. Genet.* 83 : 119-125.
- FRANKEL (O.H.) and BROWN (A.H.D.), 1984. Plant genetic resources today : a critical appraisal. *In Genetics : new frontiers*, Vol 4. Oxford and IBH Publishing Co, New Dehli : 1-11.
- GALWEY (N.W.), 1993. Verifying and validating the representativeness of a core collection. *In Proc. Int. IBPGR / CENARGEN Workshop on Core Collection*. Brazilia, Brésil, 22-27 août 1992 (sous presse).
- GOTTLIEB (L.D.), 1981. Electrophoretic evidence and plant populations. *Progress Phytochem.* 7 : 1-46.
- HAMON (S.), NOIROT (M.) and ANTHONY (F.), 1993. A statistical procedure for selecting a core collection with quantitative data. Application to coffee. *In Proc. Int.*

- LOVELESS (M.D.) and HAMRICK (J.L.), 1984. Ecological determinants of genetic structure in plant populations. *Ann. Rev. Ecol. Syst.* 15 : 65-95.
- NATALI (L.), CAVALLINI (A.), CIONINI (G.), SASSOLI (O.), CIONNINI (P.G.) and DURANTE (M.), 1993. Nuclear DNA changes within *Helianthus annuus* L. within single progenies and their relationships with plant development. *Theor. Appl. Genet.* 85 : 506-512.
- NOIROT (M.) et HAMON (S.), 1992. Contraintes imposées à l'amélioration des plantes par la diversité des modes de reproduction au sein des complexes d'espèces. *In Actes du colloque international en Hommage à J. Pernès. Paris 8-10 janvier 1992* : 301-312.
- PEETERS (J.P.) and WILLIAMS (J.T.), 1984. Towards better use of genebanks with reference to information. *IBPGR Plant Genet. Resources Newsl.* 60 : 22-32.
- PERNES (J.) et LOURD (M.), 1984. Organisation des complexes d'espèces *In Gestion des ressources génétiques des plantes, Tome 2 : manuel, J. Pernès éd., ACCT (Paris)* : 5-106.
- RAO (K.E.P.) and RAO (V.R.), 1993. The use of characterization data in developing a core collection. *In Proc. Int. IBPGR / CENARGEN Workshop on Core Collection. Brazilia, Brésil, 22-27 aout 1992 (sous presse).*
- TOHME (J.), JONES (P.), BEEBE (S.) and IWANAGA (M.), 1993. The CIAT *Phaseolus vulgaris* core collection established by the combined use of evolutionary, characterization and agroecological data. *In Proc. Int. IBPGR / CENARGEN Workshop on Core Collection. Brazilia, Brésil, 22-27 aout 1992 (sous presse).*