

**LA BASE DE DONNEES DES RESSOURCES GENE-
TIQUES DES CAFEIERS AFRICAINS EN COLLECTION
EN COTE D'IVOIRE**

François Anthony
(Génétique)
(UR 507)

Centre ORSTOM de Montpellier
B.P.5 045
34032 Montpellier Cédex (France)

RESUME - Le Centre des ressources génétiques des caféiers africains, constitué en Côte d'Ivoire par l'ORSTOM et l'I.R.C.C. (Institut de Recherche du Café et du Cacao), est unique au monde par le nombre d'espèces représentées, l'importance des effectifs et la diversité des origines. Plus de 7500 géotypes sauvages appartenant à une quinzaine d'espèces de *Coffea* ont été collectés dans environ 250 populations naturelles réparties dans huit pays d'Afrique. Les collections sont installées dans deux stations de recherche présentant des conditions écologiques différentes. Les programmes de recherche basés sur ce matériel végétal sont effectués à l'aide de descripteurs nombreux et variés.

L'informatisation des données a commencé à la fin des années 70 par la définition d'un système de numérotation des géotypes applicables aux caféiers sauvages et à leurs hybrides des deux premières générations. En 1985, les deux stations de conservation ont été équipées de micro-ordinateurs Goupil 3PC fonctionnant sous MS-DOS. Le support physique des données est formé par des disquettes 5" (360 Koctets). Les logiciels utilisés pour la gestion et les analyses statistiques ont été développés par Noirot, Déjardin, Mullon et Savy (N.D.M.S.) en Sbasic (Publication ORSTOM, à paraître).

La base de données conçue a pour objectifs de :

- permettre la gestion des caféiers plantés en relation avec leur origine géographique,
- constituer un répertoire du matériel végétal disponible,
- standardiser les descripteurs utilisés dans les études génétiques et agronomiques,
- faciliter l'accès et l'exploitation des données de l'évaluation.

Les données codées concernent la numérotation des individus, leur provenance, leur plantation en collection et leur évaluation. Les données de l'évaluation, actuellement saisies, sont relatives à l'intensité des floraisons, aux productions par date de récolte et aux caractéristiques des fruits et des grains (fertilité, technologie...).

Il reste à coder les descripteurs de la morphologie des caféiers et des caractères enzymatiques révélés par électrophorèse.

Les données sont stockées dans deux types de fichiers :

- les *fichiers_des_individus* sont constitués uniquement de variables alpha-numériques représentant les codes des génotypes, de leur origine et de leur mise en collection (une ou deux positions). Ils sont découpés par espèce. L'ensemble de ces fichiers forme un répertoire permettant la gestion des collections.

- les *fichiers_de_l'évaluation* contiennent d'une part des variables alpha-numériques pour le code des caféiers et leur position en parcelle et d'autre part les variables numériques de leur évaluation. Pour chaque type de données (floraisons, récoltes...), des fichiers différents sont créés par parcelle, espèce et année.

La place mémoire totale occupée est d'environ 3,5 millions d'octets dont la moitié pour les fichiers des individus. Le logiciel N.D.M.S. réservant systématiquement huit octets par variable, il serait possible de diminuer des deux tiers la taille du répertoire si le programme de création de fichier tenait compte du format des champs.

Tous les fichiers sont triés sur le code des caféiers et exploités séquentiellement.

Les problèmes rencontrés proviennent du matériel et du logiciel. Il est évident que la capacité mémoire des disquettes est insuffisante. Pour le moment, les fichiers sont répartis sur 28 disquettes. Seul l'emploi d'un disque dur évitera les nombreuses manipulations de disquettes. Le logiciel N.D.M.S. n'a pas été conçu dans le cadre d'une base de données. Il ne permet d'exploiter qu'un seul fichier à la fois et les fonctions sont donc limitées. C'est pourquoi il est nécessaire d'utiliser un autre logiciel de gestion, plus performant, et de transformer ensuite les fichiers pour accéder à la bibliothèque de programmes statistiques N.D.M.S. qui est bien adaptée aux besoins des biologistes.

Les principales perspectives résident dans l'emploi d'un autre micro-ordinateur (Compaq 286) avec un disque dur (20 M. octets) à Montpellier. Les données seront prochainement transférées pour utiliser le logiciel dBASE III. Tous les fichiers pourront alors être indexés sur le code des caféiers. Cela permettra de constituer un système offrant de meilleures conditions d'exploitation.

INTRODUCTION

Le Centre des Ressources Génétiques des caféiers africains, progressivement mis en place en Côte d'Ivoire par l'ORSTOM et l'I.R.C.C. (Institut de Recherche du Café et du Cacao), est unique au monde par le nombre d'espèces représentées, l'importance des effectifs et la diversité des origines. Plus de 7500 géotypes sauvages, appartenant à une quinzaine d'espèces de *Coffea*, ont été collectés dans environ 250 populations naturelles réparties dans huit pays d'Afrique. Les collections vivantes sont installées dans deux stations de recherche présentant des conditions écologiques différentes : au Mont Tonkoui (1100 m.), près de Man, pour les caféiers d'altitude et à Divo pour les caféiers de basse altitude. La diversité morpho-physiologique de ces caféiers spontanés et leur valeur agronomique sont étudiées à l'aide de descripteurs nombreux et variés dont certains, utilisés pendant plusieurs années, sont communs à la plupart des espèces.

L'informatisation des données a commencé à la fin des années 70 par la définition d'un système de numérotation des caféiers. Depuis 1985, les deux stations de conservation des ressources génétiques sont équipées de micro-ordinateurs Goupil 3PC fonctionnant sous MS-DOS. Le support physique des données est constitué par des disquettes 5 1/4 pouces d'une capacité mémoire de 360 K.octets. Le logiciel de gestion utilisé appartient à la bibliothèque de programmes de calculs statistiques développés en Sbasic par Noiroto, Déjardin, Mullon et Savy (Publication ORSTOM, à paraître).

La base de données a pour objectifs de :

- constituer un répertoire du matériel végétal disponible,
- permettre la gestion des caféiers plantés en relation avec leur origine géographique,
- standardiser les descripteurs utilisés dans les études génétiques et agronomiques,
- faciliter l'accès et l'exploitation des données de l'évaluation.

Après avoir expliqué le codage des informations stockées, nous présenterons l'organisation générale de la base de données en dégageant les applications pratiques. Enfin, nous exposerons les problèmes rencontrés et les perspectives pour améliorer les performances du système.

1. LE CODAGE DES DONNEES

Les données codées concernent la numérotation des génotypes, leur origine géographique, leur plantation en collection et divers résultats de leur évaluation.

1.1. La numérotation des génotypes

Chaque génotype est repéré par :

- son niveau de ploïdie (format = 1 caractère),
- un identificateur de l'espèce (format = 4),
- un numéro d'ordre défini par identificateur d'espèce (format = 4).

Par exemple, l'individu 4.0001.1253 est un caféier tétraploïde (4), de l'espèce *C.arabica* (0001), numéroté 1253.

Pour les caféiers sauvages, le niveau de ploïdie et l'identificateur d'espèce sont liés : tous les *Coffea* connus sont diploïdes à l'exception de *C.arabica* qui est tétraploïde.

Néanmoins, ce codage a l'avantage de conserver l'identificateur d'espèce et le numéro d'ordre des génotypes ayant subi une duplication chromosomique. Les couples isogéniques diploïde-tétraploïde sont ainsi facilement distingués.

Ce système de numérotation est utilisable pour les hybrides des deux premières générations. L'identificateur d'espèce devient alors un identificateur de la combinaison hybride, formé à partir des deux identificateurs parentaux. Dans le cadre de cette communication, nous nous limiterons au fonctionnement de la base de données des caféiers sauvages.

1.2. L'origine géographique

L'origine des caféiers est expliquée par :

- le pays de collecte (format = 2),
- la population dans le pays (format = 3),
- la nature du matériel prélevé (graine, plantule ou axe végétatif) (format = 1),
- et éventuellement le code de l'arbre-mère.

1.3. La plantation en collection

La plantation en collection est décrite par :

- la station de recherche (format = 1),
- la parcelle dans la station (format = 2),
- la position dans la parcelle (numéro de ligne, format = 3, et rang sur la ligne, format = 2),
- l'année de plantation (format = 2),
- la nature du matériel planté (plant issu de semis, bouture ou greffe, format = 1),

- éventuellement la duplication en collection dans la même station,
- ou dans une autre station.

La plupart des génotypes sont plantés deux fois dans la même station à l'exception des descendances issues de graines (caféiers apparentés).

1.4. Les données de l'évaluation

Les données saisies de l'évaluation sont relatives à l'intensité des floraisons, aux poids de cerises fraîches récoltées et aux caractéristiques des fruits et des grains. Les champs des floraisons et des récoltes ont pour nom la date de la notation.

Les descripteurs des fruits et des grains sont :

- le poids de 100 grains à 12% d'humidité,
- le rendement en café marchand,
- le taux de remplissage des loges des fruits (deux loges potentielles par fruit),
- le taux de grains caracolis (un grain par fruit),
- le taux de loges vides,
- la teneur en caféine,
- les différentes catégories de cerises,
- la longueur du pédoncule des fruits,
- les dimensions (longueur, largeur, épaisseur).

Les quatre premiers descripteurs sont obtenus chaque année sur tous les arbres fructifères dont la récolte est pesée individuellement. La teneur en caféine est déterminée une fois pour chaque génotype. Les autres descripteurs sont utilisés selon les besoins de l'évaluation et leur liste n'est pas close.

2. L'ORGANISATION GENERALE

La base de données est formée par deux types de fichiers répondant à des contraintes et à des objectifs différents : les fichiers des individus et les fichiers de l'évaluation.

2.1. Les fichiers des individus

Les fichiers des individus sont constitués uniquement par des variables alpha-numériques représentant les codes des génotypes, de leur origine et de leur plantation en collection. Ils sont découpés par niveau de ploïdie et par identificateur d'espèce. Puis ils sont triés sur le numéro d'ordre des caféiers. L'ensemble de ces fichiers forme un répertoire du matériel végétal planté avec mention de la provenance géographique.

Ces fichiers sont créés au fur et à mesure de la plantation en collection de nouveaux génotypes. Ils sont mis à jour au moins une fois par an. Au total, ils occupent plus d'un million et demi d'octets.

Le logiciel utilisé disposant de fonctions limitées, les principales applications pratiques se résument à :

- recenser les génotypes par espèce, pays et population,
- trouver les positions en collection des caféiers d'une origine donnée,
- établir la liste des caféiers morts à remplacer chaque année,
- reconstituer les familles pour des analyses de transmission de caractères,
- rechercher les génotypes qui étaient adultes au moment de leur collecte pour étudier l'influence des pressions de sélection dans les populations naturelles.

2.2. Les fichiers de l'évaluation

Les fichiers de l'évaluation contiennent d'une part les variables alpha-numériques du code des caféiers et de leur position en parcelle et d'autre part les variables numériques des données de l'évaluation. Pour chaque type de données (floraisons, récoltes...), des fichiers différents sont créés par station, par parcelle, par espèce et par année. L'année représente un cycle annuel de reproduction. Ainsi, les données de fructification sont enregistrées avec la même année que celles des floraisons correspondantes.

Les fichiers disponibles des floraisons, des récoltes et des caractéristiques des fruits et des grains occupent environ deux millions d'octets.

3. PROBLEMES ET PERSPECTIVES

Les problèmes rencontrés proviennent du matériel et du logiciel. Il est évident que la capacité mémoire des disquettes est insuffisante. Pour le moment, les fichiers sont répartis sur 28 disquettes. Seul l'emploi d'un disque dur évitera les nombreuses manipulations de disquettes pendant les travaux de gestion.

Le logiciel utilisé n'a pas été conçu dans le cadre d'une base de données. Les fichiers sont exploités séquentiellement et séparément. Les fonctions sont donc limitées. Par ailleurs, le logiciel réserve systématiquement 8 octets par variable. Avec un programme de création de fichier tenant compte du format des variables, il sera possible de diminuer des deux tiers la taille du répertoire.

Les meilleures perspectives d'amélioration résident dans l'emploi d'un autre micro-ordinateur (Compaq 286) avec un disque dur (20 M.octets) à Montpellier. Les fichiers seront prochainement

transformés pour être exploités avec un logiciel de gestion plus performant et plus répandu tel que dBASE III. L'ouverture simultanée de plusieurs fichiers permettra d'effectuer des opérations à la fois dans le répertoire et dans les fichiers de l'évaluation. Il sera alors possible de sélectionner les caféiers d'une espèce, provenant d'une origine choisie et possédant des caractéristiques précises de floraison, récolte, etc.. Pour cela, il faudra constituer un système relationnel. Les fichiers devront être découpés autrement car certains géotypes sont plantés plusieurs fois dans la même parcelle. Le répertoire sera sans doute éclaté en un fichier des origines, indexé sur le code des caféiers, et un fichier des plantations, indexé sur la position en collection. Les fichiers de l'évaluation seront aussi indexés sur la position en parcelle. Les analyses statistiques des données seront toujours réalisées à l'aide du logiciel N.D.M.S. après transformation des fichiers.

CONCLUSION

Les objectifs fixés sont partiellement atteints. Nous avons saisi la plupart des données accumulées depuis plus de dix ans. Le répertoire du matériel végétal vivant en collection est disponible et peut être diffusé sous forme de listing. Cependant, il est indispensable d'améliorer l'accès aux données et leur traitement. L'organisation générale de la base de données devra être modifiée pour parvenir à un système performant.

La saisie des informations continue. De nouveaux fichiers sont à l'étude pour décrire :

- les populations d'origine (coordonnées géographiques, description de la forêt, ...),
- les hybridations réalisées en station (nombre de fleurs fécondées, taux de nouaison et de fructification),
- les caractères enzymatiques révélés par électrophorèse,
- les caractères morphologiques, l'architecture et la croissance des caféiers.

Enfin, il est utile de rappeler que la base de données constituée pour les caféiers sauvages est applicable aux hybrides des deux premières générations. Seules les variables de l'origine sont modifiées. Le sélectionneur peut ainsi traiter ensemble les données des hybrides et de leurs parents.