

Ann. Inst. Nat. Agron. El-Harrach, 1989,
Vol. 13, N°2, 704 - 711.

**L'INFORMATIQUE AU SERVICE DES RESSOURCES PHYTOGENETIQUES:
BASE DE DONNEES ET EXPLOITATION STATISTIQUE SUR MICRO-
ORDINATEUR.**

Par **BEN KHALIFA A. et BRAC DE LA PERRIERE R.A.**
U.R.Z.A. Station d'EL-GOLEA BP 13 EL-GOLEA
ALGERIE

R E S U M E

Pour faciliter la gestion et réaliser des traitements de données de ressources génétiques, sur un micro-ordinateur, les auteurs ont essayé différents logiciels, depuis trois années, en utilisant des données recueillies sur Palmier Dattier.

I N T R O D U C T I O N

Les premières taches préoccupantes dans l'étude des ressources génétiques, sont les inventaires et les évaluations. Le problème d'accumulation de données se pose quelque soit la méthode suivie. Il s'agit, toujours de recenser espèces, cultivars, lignées ou populations, qu'il faut caractériser par un grand nombre de variables observées, mesurées ou analysées. Les descripteurs peuvent être des caractéristiques phénotypiques, biochimiques, cytogénétiques ou agronomiques, auxquelles s'ajoutent les données relatives aux conditions d'observations. Quel que soit leur nature, les informations recueillies doivent être stockées dans des fichiers, numérotées correctement de façon que chacune puisse être recherchée, modifiée et rendue accessible dans les délais favorisant des utilisations

fréquentes et diverses. Il peut s'agir de milliers ou de millions de données ou informations à enregistrer avec fiabilité de façon pratique et aisément communicable. Pour ces raisons l'étude des données de ressources génétiques renvoie à tous les aspects de la gestion des bases de données et à l'utilisation de l'informatique.

Face aux multiplications répétitives et fastidieuses, l'ordinateur est un moyen indispensable à acquérir et à accepter. Si les grandes machines sont coûteuses et nécessitent généralement des équipes spécialisées, les micro-ordinateurs sont moins coûteux et peuvent être placés plus facilement dans les laboratoires et les établissements de recherche ou de développement.

Nous travaillons dans le cadre des recherches sur les ressources génétiques du Palmier dattier sur un micro-ordinateur IBM-PC avec double lecteur de disquettes (5,25) et une imprimante IBM 4201. Nous ne voulons ni présenter un document d'informatique ni faire de la publicité sur les ordinateurs et les accessoires électroniques. L'intérêt est de montrer au titre d'échange d'expériences avec nos collègues, combien un micro-ordinateur de puissance modérée avec la maîtrise de quelques logiciels on peut rendre d'importants services dans le cadre des recherches sur les ressources génétiques. Nous montrons l'utilisation d'un micro-ordinateur pour réaliser:

1. Un stockage de fichiers peu encombrant avec une consultation rapide;
2. Une exploitation statistique à des niveaux approfondis

1. LE TRAITEMENT DE FICHIERS

Il n'existe pas un système parfaitement au point et directement transposable d'une espèce végétale à l'autre pour constituer et gérer une base de données des ressources génétiques. Pour réaliser un stockage et permettre une consultation rapide de nos données, nous avons essayé deux logiciels de gestion de fichiers; Filing Assistant et DBASE III+. Ces logiciels fonctionnent sur les micro IBM et compatibles (Olivetti M24 par exemple). Ce matériel est déjà disponible dans certains de nos laboratoires ou même acquis par nos collègues.

Le principe est toujours le même. Il s'agit de constituer une fiche type "maquette" et enregistrer les données une à une. Ce qui demande un peu de temps, mais une fois que le fichier est enregistré on a l'avantage de sélectionner les données, les classer, les dupliquer, les corriger et rajouter de nouvelles données.

Le premier logiciel Filing Assistant est installé sur une seule disquette (5,25 double face double densité). Le menu est très simple. Les fonctions d'aide peuvent être demander à tous les niveau. Le programme permet de créer une fiche type d'une ou plusieurs pages, de saisir des fiches et de les corriger de rechercher et de trier des fiches selon une ou plusieurs rubriques, de copier ou imprimer les fiches triées ou recherchées. Ce logiciel n'autorise qu'une seule rubrique par ligne et on ne peut pas conserver les informations saisies lors d'une modification des intitulés existants.

Par ailleurs sa multiplication est très simple et aucune connaissance ou formation spécialisée en informatique n'est exigée.

Exemple d'une fiche enregistrée:

FICHE D'ENQUETE OCT, NOV, 1985. -----

CODE

CULTIVAR: TAQUERBUCH
LOCALITE: TIMOUDI
CARACTERISTIQUES DU CULTIVAR

Synonymes locaux.....: ?
Sens du nom.....: ?
Origine présumée.....: ?

Importance dans la localité.....: RARE
Date de récolte.....: NOVEMBRE
Utilisation de la datte.....: CONSERVEE
Commercialisation.....: FAIBLE
Appréciation.....: EXCELLENTE
Digestibilité.....: FROIDE
Conservation.....: ECRASEE
Sensibilité.....: INCONNUE
Prix du kilo.....: 10

INFORMATION SUR LE PIED MERE -----

Age du pied mère.....: JEUNE
Nombre de régimes portés.....: 5....10
Irrigation.....: BONNE
Nombre de rejets.....: AUCUN
Maladies.....: Sain

REFERENCES COMPLEMENTAIRES -----

N° de l'échantillon.....: 70
Date de l'observation.....: ?
Nom de l'observateur.....: ?

REMARQUÈS:

Le deuxième logiciel est le D base III+. Bien que ce type de logiciel soit beaucoup plus consistant pour construire une base de données, sous forme d'enregistrements, il peut être aussi utilisé comme une calculatrice en mode interactif.

Aux avantages du DBASE III+ lui même, il s'ajoute la possibilité d'utiliser les données d'un fichier de DBASE III+ avec d'autres logiciels de traitements de statistiques (cas du STAT-ITCF).

Il occupe deux disquettes (5,25 DS DD), et possède un mode de fonctionnement automatique. Avec un utilitaire ASSIST qui va plus loin que l'aide-mémoire, on peut manipuler DBASE III+ même si on ne connaît pas les commandes.

Chaque enregistrement prend une ligne qui se divise en plusieurs rubriques appelées champs. Les enregistrements de DBASE III+ peuvent contenir jusqu'à 128 champs et un fichier peut comporter jusqu'à un milliard d'enregistrements.

Les commandes permettent d'ajouter et supprimer des informations, de modifier des enregistrements, d'effectuer un survol rapide du fichier entier et de sélectionner uniquement les informations désirées.

Un apprentissage est fortement conseillé pour pratiquer le DBASE III+.

2. LE TRAITEMENT STATISTIQUE DES DONNEES

Les informations recueillies pour l'étude des ressources génétiques sont des variables statistiques. Pour mieux les utiliser il faut les décrire, les représenter lisiblement (sous forme de tableaux et de graphiques) et les analyser pour mieux les exploiter. La pratique de l'analyse de données s'avère nécessaire dans les analyse de distance, l'étude de la variabilité des échantillons ou l'établissement des classifications.

Nous utilisons pour le traitement de données des prospections des palmeraies algériennes le STAT-ITCF (ver.3 du Fév. 1987), rédigé par le service informatique de l'Institut Technique des Céréales et des Fourrages (France).

Ce logiciel est installé sur 14 disquettes et composé de 25 modules de traitements statistiques (voir sommaire). Les programmes sont limités en général à 60 variables initiales mais le nombre d'individus n'est limité que par la capacité des disquette(s). Il existe certaines limites spécifiques à des programmes précis (100 individus et 115 modalités issues de 60 variables pour le cas d'une analyse factorielle de correspondances multiples).

Le temps d'exécution des programmes dépend de la taille des tableaux à analyser mais il n'est considérable que pour les programmes d'analyses multivariées (près d'une heure pour une analyse de correspondances multiples et quelques heures pour une analyse de classification automatique sur un grand tableau).

L'utilisation de ce logiciel est facilitée par les représentations pédagogiques du manuel ou l'aide écran accessible à tous les niveaux (deux disquettes sont réservées). La maîtrise des applications exige une détermination parfaite du but à atteindre, une bonne préparation des données et savoir interpréter des résultats statistiques.

-----S T A T I T C F -----

version 3 - copyright-1987

OPTIONS:

- | | |
|--|---|
| A. GESTION DES DONNEES | N. TESTS NON PARAMETRIQUES |
| B. HISTOGRAMMES | O. TABLEAUX CROISES ET TEST KHI-2 |
| C. GRAPHIQUES | P. ANALYSE COMBINATOIRE |
| D. PARAMETRES STATISTIQUES | Q. DISTRIBUTIONS DE PROBABILITES |
| E. COMPARAISON DE 2 MOYENNES | R. COMPARAISON DE 2 PROPORTIONS |
| F. PUISSANCE-Nb DE BLOCS
(ou d'ESSAIS) | S. ETUDE D'ECHANTILLONNAGE |
| G. GESTION DES DONNEES POUR
UNE ANALYSE DE VARIANCE | T. ANALYSE EN COMPOSANTES PRINCIPALES |
| H. ANALYSE DE VARIANCE | U. ANALYSE FACTORIELLE DISCRIMINANTE |
| I. REGROUPEMENT D'ESSAIS | V. ANALYSE FACTORIELLE DES
CORRESPONDANCES |
| J. MATRICE DE CORRELATIONS | W. CLASSIFICATION AUTOMATIQUE |
| K. REGRESSION LINEAIRE | |
| M. SERIE CHRONOLOGIQUES | |
| | Z. (fin) |

CHOISISSEZ UNE OPTION

F1 = aide

F2 = Présentation

C O N C L U S I O N

Il n'existe pas de système parfaitement au point et directement transposable de n'importe quelle espèce pour constituer et gérer une banque de données dans le cadre des ressources génétiques. Les grands ordinateurs ne sont pas disponibles et nécessitent des équipes spécialisées. Les micro-ordinateurs et leurs logiciels ont énormément évolué et permettent de réaliser sur un petit matériel informatique les calculs les plus compliqués dans un temps négligeable.

DBASE III+ ou FILING ASSISTANT sont deux logiciels de traitement de fichier sur micro-ordinateur qui peuvent rendre un très grand service dans la gestion de données de ressources génétiques.

Le STATITCF est un logiciel de traitement statistique qui permet la pratique de l'analyse de données jusqu'au modèle multivarié.

Pour permettre un échange d'expérience et faciliter le transfert des informations et du matériel entre établissements chercheurs ou collègues, il est nécessaire de concorder les efforts pour homogénéiser les méthodes de stockage de données notamment dans le choix du système de gestion de base de données, le choix des descripteurs et de leurs états entre les personnes travaillant sur le même matériel ou avec les mêmes types de traitement.