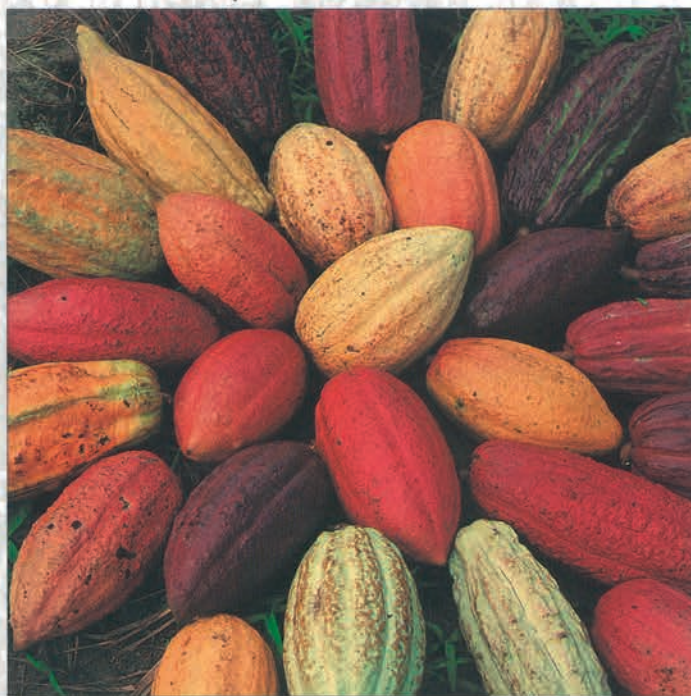


■ REPÈRES

# Diversité génétique des plantes tropicales cultivées

---

Perla Hamon, Marc Seguin, Xavier Perrier  
et Jean Christophe Glaszmann  
*Editeurs scientifiques*



---

CIRAD

## LES ÉDITEURS

---

Perla Hamon est maître de conférence à l'université Montpellier III. Depuis près de vingt ans, ses activités de recherche portent sur la diversité génétique des plantes tropicales. Elle a en particulier mené des travaux sur les ignames en Côte d'Ivoire, qui constituent le sujet de la thèse d'Etat qu'elle a soutenue en 1987. Ses travaux au Cirad, à Montpellier, ont porté sur les sorghos.

Marc Seguin a étudié pendant neuf ans les ressources génétiques et le génome de l'hévéa, au Cirad à Montpellier. Il a aussi travaillé sur la diversité génétique du palmier à huile, de la canne à sucre et du cocotier. Il est actuellement en poste en Thaïlande pour y développer le marquage moléculaire de l'hévéa dans le cadre de son amélioration génétique.

Xavier Perrier anime la délégation mathématiques et informatique appliquées au sein de la direction scientifique du Cirad. Une part importante de ses activités concerne les méthodes de gestion et d'analyse des données de diversité génétique. Il est l'auteur d'une thèse de mathématiques consacrée aux méthodes de construction d'arbres de diversité et à leur comparaison.

Jean Christophe Glaszmann est responsable du programme biotechnologies et ressources génétiques végétales du Cirad. Auparavant, il a étudié pendant huit ans la diversité génétique du riz, notamment à l'IRRI, International Rice Research Institute, et élaboré une classification isoenzymatique de l'espèce, aujourd'hui largement utilisée par les sélectionneurs. Il a ensuite abordé l'étude du génome de la canne à sucre à l'aide des marqueurs moléculaires.

## LE CIRAD

---

Le Cirad, Centre de coopération internationale en recherche agronomique pour le développement, est un organisme scientifique spécialisé en agriculture des régions tropicales et subtropicales. Il réalise, dans une cinquantaine de pays, des opérations de recherche, de développement agricole et de formation.

*Illustration de couverture*

Cabosses de cacaoyer (Claire Lanaud-Cirad).



Diversité génétique  
des plantes  
tropicales cultivées





# Diversité génétique des plantes tropicales cultivées

Perla Hamon, Marc Seguin, Xavier Perrier  
et Jean Christophe Glaszmann  
*Editeurs scientifiques*





# Sommaire

---

- 7 Avant-propos
- 11 Abstract
- 13 Les méthodes de marquage biochimique et moléculaire  
Laurent Grivet, Jean-Louis Noyer
- 43 Les méthodes d'analyse des données  
Xavier Perrier, Albert Flori, François Bonnot
- 77 Une méthode de constitution de *core collections*  
Michel Noirot, François Anthony, Stéphane Dussert, Serge Hamon
- 89 Les agrumes  
Patrick Ollitrault, Camille Jacquemond, Cécile Dubois,  
François Luro
- 113 Les bananiers  
Christophe Jenny, Françoise Carreel, Kodjo Tomekpe, Xavier Perrier,  
Cécile Dubois, Jean-Pierre Horry, Hugues Tézenas du Montcel
- 141 Le cacaoyer  
Claire Lanaud, Juan-Carlos Motamayor, Olivier Sounigo

- 175 Le caféier, *Coffea canephora*  
Stéphane Dussert, Philippe Lashermes, François Anthony,  
Christophe Montagnon, Pierre Trouslot, Marie-Christine Combes,  
Julien Berthaud, Michel Noirot, Serge Hamon
- 195 La canne à sucre  
Jean Christophe Glaszmann, Nazeema Jannoo, Laurent Grivet,  
Angélique D'Hont
- 219 Le cocotier  
Patricia Lebrun, Yavo-Pierre N'cho, Roland Bourdeix, Luc Baudouin
- 241 L'hévéa  
Marc Seguin, Albert Flori, Hyacinthe Legnaté,  
André Clément-Demange
- 271 Le manioc  
Gérard Second, Jean-Pierre Raffailac, Carlos Colombo
- 309 Le mil  
Gilles Bezançon
- 327 Le riz asiatique  
Jean Christophe Glaszmann, Laurent Grivet, Brigitte Courtois,  
Jean-Louis Noyer, Claude Luce, Michel Jacquot, Laurence Albar,  
Alain Ghesquière, Gérard Second
- 351 Le sorgho  
Monique Deu, Perla Hamon, François Bonnot, Jacques Chantereau
- 383 Annexes  
Liste des abréviations 385  
Adresses des auteurs 386



## Avant-propos

La naissance de l'agriculture marque les débuts de l'amélioration des plantes. Elle remonte au paléolithique, il y a dix à douze mille ans, lorsque des groupes humains, qui tiraient jusque là l'essentiel de leur subsistance de la cueillette, de la chasse et de la pêche, se sont sédentarisés. Depuis cette époque et jusqu'au milieu du XIX<sup>e</sup> siècle s'est déroulé le processus de domestication de la quasi-totalité des plantes aujourd'hui cultivées. Celles-ci passèrent du statut d'espèce sauvage à celui d'écotype ou de variété traditionnelle en acquérant progressivement un ensemble de caractéristiques plus ou moins éloignées de celles de leurs ancêtres, qui constitue le « syndrome de domestication ». Cette domestication a eu lieu soit dans des espaces très limités — comme pour le maïs et la pomme de terre, pour leur histoire précolombienne —, soit sur plusieurs continents comme c'est le cas du riz et du sorgho. L'évolution des espèces cultivées s'est donc effectuée sous l'effet de pressions de sélection environnementales et humaines liées à des processus à la fois cumulatifs et divergents, d'introggression permanente et de diversification. L'homme a exploité cette diversité en valorisant tout d'abord les ressources de son environnement immédiat puis, à partir du XVI<sup>e</sup> siècle, celles que l'essor des échanges et des transports rendait accessibles. Ainsi est née une infinité de systèmes de culture, qui constituent chacun une réponse originale aux besoins et aux contraintes propres à chaque communauté humaine.

Jusque dans les années 30, l'amélioration variétale moderne fondée sur les schémas de sélection reposait uniquement sur l'utilisation de variétés ou d'écotypes traditionnels et la notion de ressources génétiques était limitée aux plantes cultivées. A la suite des travaux de Vavilov, cette notion s'est étendue aux espèces sauvages apparentées puis aux genres de plus en plus éloignés. Néanmoins, l'importance des ressources génétiques n'est apparue réellement qu'avec la menace de leur disparition. Ce phénomène a été particulièrement marqué en Europe après la Seconde Guerre mondiale. En effet, les équilibres naturels des milieux ont été fortement perturbés du fait de l'extension considérable des agglomérations aux dépens des zones agricoles, de la mise en culture, avec ou sans déforestation, de nouveaux espaces, souvent peu adaptés et fragiles, et de la dégradation de vastes régions naturelles, jusque là préservées.

Consciente de ce risque, la communauté scientifique internationale s'est mobilisée dans les années 60-70 pour collecter et conserver un maximum de ressources génétiques, en donnant la priorité aux espèces alimentaires qui présentaient un intérêt économique majeur à l'échelle mondiale. C'est ainsi qu'une multitude de collections ont été rassemblées à travers le monde. La FAO a publié une synthèse sur ce sujet, le *Rapport sur l'état des ressources phytogénétiques dans le monde*, à l'occasion de la Conférence internationale de Leipzig de juin 1996. Actuellement, la plupart de ces collections ont atteint

une taille qui rend difficiles leur entretien et la caractérisation systématique de leurs accessions. La question de leur gestion se pose de manière aiguë. Il devient nécessaire de trouver un équilibre entre conservation, évaluation et utilisation des ressources génétiques. Sans évaluation, il n'y a pas d'utilisation raisonnée possible, sans utilisation, la conservation perd sa justification pour les institutions qui ne sont pas spécifiquement mandatées pour l'assurer.

Pour répondre à ces préoccupations, Frankel et Brown ont introduit dans les années 80 le concept de *core collection*, qu'ils définissent comme un échantillon limité d'accessions issues d'une collection plus vaste, dite de base, et choisies pour représenter au mieux le spectre de diversité existant. Une telle *core collection* peut servir de nombreux objectifs. Elle permet d'identifier le matériel dont la conservation est prioritaire, mais aussi d'accéder de manière rationnelle à la diversité génétique disponible dans la collection de base, ce qui facilite la recherche de nouvelles sources de caractères utiles pour la sélection.

En général, on dispose pour les collections de base d'informations « passeport », comme l'origine géographique et écologique des accessions. Ces informations sont complétées le cas échéant par des données morphotaxonomiques, agronomiques et génétiques issues des marqueurs biochimiques (isoenzymes, polyphénols...) ou moléculaires (RFLP, RAPD...).

Lors de la construction d'une *core collection*, les caractères utiles sont prioritaires pour le sélectionneur, mais leur évaluation directe est parfois lourde et peut cacher des déterminismes génétiques complexes. De plus, cette évaluation peut être obérée par la méconnaissance des contraintes qui pèseront sur la culture dans le futur. D'un autre côté, les marqueurs génétiques moléculaires ne sont d'aucune utilité directe, mais ils traduisent des structures de diversité générale, qui peuvent à leur tour servir de base pour construire la *core collection*. En effet, les marqueurs révèlent parfois des groupes d'accessions partiellement isolés les uns des autres, qui peuvent avoir fixé des allèles distincts pour des caractères utiles, par effet de fondation, par dérive génétique ou encore sous l'action de différentes pressions de sélection.

Les relations entre les deux niveaux de variabilité — celui des marqueurs génétiques moléculaires, probablement en majorité neutres, et celui des caractères agronomiques, généralement plus complexes et soumis à une sélection naturelle ou humaine — sont mal connues. Elles sont au cœur des préoccupations de la communauté scientifique impliquée dans la gestion des ressources génétiques, comme cela apparaît dans les conclusions du colloque sur les ressources génétiques végétales et animales et leurs méthodologies d'étude et de gestion, organisé par l'INRA et le BRG à Montpellier, en septembre 1993.

Ces relations varient probablement en fonction de la structure des populations. Elles seront généralement fortes s'il existe des déséquilibres gamétiques forts. Mais pour avancer dans cette réflexion, il convient d'aborder des questions précises : les différents types de marqueurs moléculaires sont-ils équivalents ? de fortes structurations à l'échelle moléculaire, qui témoignent de désé-



quilibres gamétiques généralisés sur l'ensemble du génome, sont-elles systématiquement associées à de fortes structurations pour les caractères agronomiques ? les structures à ces deux niveaux sont-elles alors concordantes ?

Plusieurs méthodes ont été proposées pour construire des *core collections*. En règle générale, elles reposent sur une stratification de la collection de base puis sur un prélèvement aléatoire au sein de chaque groupe ainsi défini selon différentes modalités. Lorsque des données d'évaluation morphoagronomiques sont disponibles, il est intéressant de considérer la structure de la diversité génétique établie à l'aide des marqueurs génétiques afin de maximiser la diversité agronomique et de minimiser la perte des allèles rares de manière globale mais ponctuellement importants.

La constitution d'une *core collection* doit donc s'appuyer sur une excellente description des populations et sur une bonne compréhension de leur structuration. Les outils statistiques sont ici indispensables. Ainsi, pour chaque type de marqueur, il est important d'utiliser une mesure de dissemblance entre les unités taxonomiques (individus ou populations) pertinente au regard de ses propriétés mathématiques et de son interprétation en termes génétiques. De même, le choix des algorithmes de représentation des dissimilarités doit reposer sur un équilibre entre une efficacité maximale et une complexité minimale, afin d'être en mesure de traiter des tableaux de grande taille, supérieure à 100 individus. La comparaison des structurations observées à l'aide des différents types de descripteurs permet de considérer de manière globale l'organisation de la diversité génétique chez une plante. Cette question peut être abordée de différentes manières et, entre autres, en recherchant les structures communes à deux ou plusieurs arbres. Dans ces conditions, comment construire des arbres consensus ou des arbres minimaux communs ? Quelle est leur fiabilité et leur signification biologique ?

Cet ouvrage fournit des éléments de réponse à ces questions en partant de l'étude de onze plantes, choisies de façon à couvrir une large gamme de caractéristiques biologiques (plantes pérennes, annuelles, autogames, allogames...) : les agrumes, le bananier, le cacaoyer, le caféier, la canne à sucre, le cocotier, l'hévéa, le manioc, le mil, le riz et le sorgho. Trois chapitres méthodologiques viennent compléter ces études : le premier est consacré à l'utilisation des marqueurs biochimiques et moléculaires pour analyser la diversité des collections, le deuxième traite de l'analyse des données et le troisième décrit une méthode de constitution de *core collections* fondée sur la maximisation de la variabilité.



# Abstract

## Genetic diversity of cultivated tropical plants

It was in the 1960s that the scientific community first became aware of the threat that disrupting natural environments posed to the germplasm of numerous cultivated species. It began working to collect that germplasm, and thus it was that a multitude of collections were set up worldwide. These collections are now so large that they are proving difficult to maintain and characterize, and their management has become a crucial issue. Germplasm conservation, evaluation and usage all need to be rethought.

In response to those concerns, Frankel and Brown introduced the idea of core collections in the 1980s: a sample of accessions from a larger collection is taken to provide as accurate a picture as possible of the existing range of diversity. However, according to which criteria and using which tools should the sample be chosen?

Agronomic criteria are primordial for breeders, but are sometimes difficult to evaluate, and their genetic determinism is often complex. Molecular genetic markers, which are of no direct use, reveal the structure of the existing diversity, which can be used as the basis for setting up a core collection. Little is known about the relations between these two levels of variability: are the different types of molecular markers equivalent? does a strong molecular structure systematically mean a strong structure based on agronomic criteria too? do the two types of structure necessarily tally?

Statistical tools, which can be used to analyse the resemblances between individuals or populations, are essential in identifying the structure of diversity, if indeed it is structured. But which method is most appropriate for each type of marker, how reliable is the picture it gives of diversity and what biological significance can it be assumed to have?

This work goes some way towards answering those questions, based on a study of the genetic diversity of eleven tropical plants. Three methodological chapters—on biochemical and molecular marking, data analysis and setting up core collections—complement the study.





# Les méthodes de marquage biochimique et moléculaire

---

Laurent Grivet, Jean-Louis Noyer

Les marqueurs biochimiques et moléculaires ont de nombreuses applications en génétique des plantes. Ils permettent d'observer, de façon plus ou moins fine, le polymorphisme de séquences de l'ADN d'un certain nombre de sites ou de locus répartis sur le génome. Plus précisément, les marqueurs biochimiques révèlent le polymorphisme des séquences de certaines protéines et donc, de façon indirecte, le polymorphisme des séquences d'ADN à partir desquelles elles sont traduites. Les marqueurs moléculaires révèlent directement le polymorphisme de l'ADN, les séquences ciblées correspondant ou non à des séquences codantes.

De par leurs propriétés, les marqueurs biochimiques et moléculaires constituent un outil puissant pour étudier la structuration de la variabilité génétique au sein d'une espèce et retracer son histoire évolutive. Ils sont très pénétrants, c'est-à-dire que leur niveau d'expression est peu affecté par l'environnement ou le fond génétique. Ils permettent donc de comparer des individus ne figurant pas dans les mêmes essais ou établis en collections dans des lieux différents. On admet généralement qu'ils révèlent un polymorphisme neutre, c'est-à-dire non soumis à la sélection. Ils sont relativement peu sensibles à l'homoplasie : le risque est faible d'observer deux allèles identiques mais résultant d'histoires mutationnelles différentes.

Ce chapitre vise à préciser les caractéristiques des différentes techniques de marquage biochimique et moléculaire disponibles pour étudier la diversité des collections de matériel génétique chez les plantes. Il existe maintenant plus d'une dizaine de techniques de marquage génétique (voir par exemple, WEISING *et al.*, 1995 ; KARP *et al.*, 1996 ; SANTONI, 1996 ; KARP *et al.*, 1997 ; KARP *et al.*, 1998 ; DE VIENNE et SANTONI, 1998). Nous en aborderons cinq, largement répandues ou prometteuses : les isoenzymes, les RFLP (*restriction fragment length polymorphism*), les RAPD (*random amplified polymorphic DNA*), les AFLP (*amplified fragment length polymorphism*) et les microsatellites. Le principe de chaque technique est brièvement décrit en annexe. Après avoir rappelé la structure du génome des plantes, nous décrivons les aspects méthodologiques — séquences ciblées, nature et niveau du polymorphisme détecté, similarité génétique — et pratiques — coût, rapidité de mise en œuvre, infrastructures nécessaires — qui différencient ces cinq techniques.

## Rappels sur l'organisation et la variabilité du génome des plantes

### L'organisation du génome

Le génome des plantes est distribué dans trois compartiments cellulaires : les mitochondries, les chloroplastes et le noyau.

Le génome mitochondrial est composé d'une molécule d'ADN maître circulaire de 200 à 2 500 kilobases (1 kilobase = 1 kb =  $10^3$  paires de bases) suivant les espèces, comportant 100 à 120 gènes. Le génome chloroplastique est aussi circulaire et sa taille est d'environ 150 kilobases. Il porte une centaine de gènes. Les génomes chloroplastique et mitochondrial sont, le plus souvent, hérités par la voie maternelle.

Le génome nucléaire est composé d'autant de molécules linéaires qu'il y a de chromosomes. Les gènes sont dispersés dans une matrice d'ADN non codant, essentiellement constituée d'ADN répété. La taille du génome nucléaire varie de façon considérable d'une espèce à l'autre : elle est de l'ordre de 400 mégabases (1 mégabase = 1 Mb =  $10^6$  paires de bases) chez le riz et de 16 000 mégabases chez le blé tendre. Le plus gros génome nucléaire connu chez les angiospermes, celui de *Fritillaria assyriaca*, contient 600 fois plus d'ADN que le plus petit, celui d'*Arabidopsis thaliana* (BENNETT et SMITH, 1991). Ces variations sont dues à des différences de niveau de ploïdie et surtout à des différences de quantité d'ADN répété dispersé non codant. On estime que le nombre de gènes portés par un génome nucléaire de plante est compris entre 20 000 et 50 000. Bien que diluée, l'information contenue dans le noyau reste donc net-

tement plus importante que celle qui est contenue dans les organites cytoplasmiques. Le génome nucléaire a une hérédité biparentale via la méiose et la fécondation.

## Le polymorphisme de l'ADN

Le polymorphisme de l'ADN résulte de l'accumulation de mutations, c'est-à-dire de modifications de séquences sous l'action de facteurs endogènes ou exogènes. Les mutations peuvent se présenter sous la forme de macroréarrangements visibles à l'échelle cytogénétique (délétion, translocation, inversion) ou sous la forme de modifications ponctuelles de séquences.

Les marqueurs biochimiques et moléculaires permettent essentiellement de détecter les variations ponctuelles de séquences. On en distingue généralement deux types : les mutations correspondant à la substitution d'une base par une autre et les mutations par insertion ou délétion d'un court fragment d'ADN.

Le polymorphisme de l'ADN engendré par les mutations ponctuelles est un fait d'observation très courant. Il affecte l'ensemble du génome, bien qu'à des niveaux différents en fonction du type de séquence et du compartiment génomique. Il est, par exemple, élevé dans les séquences répétées nucléaires de type micro ou minisatellite, mais faible dans l'ADN chloroplastique. A titre d'exemple, le séquençage, sur 1 933 paires de bases, du gène nucléaire *Opa-que-2* du maïs chez 21 lignées représentant la diversité du germoplasme tempéré a montré que les 21 séquences étaient toutes différentes. Au total, 14 % des bases présentaient un polymorphisme résultant de mutations ponctuelles et 26 phénomènes d'insertion-délétion ont été détectés (HENRY et DAMERVAL, 1997).

Un type particulier de polymorphisme de l'ADN est exploité par les marqueurs microsatellites. Un microsatellite est généralement composé d'une répétition en tandem d'un motif formé d'un petit nombre de paires de bases, en général deux ou trois. Les microsatellites sont très nombreux dans les génomes eucaryotes et sont dispersés sur l'ensemble des chromosomes. Chacun est encadré de séquences qui lui sont spécifiques. Pour chaque site microsatellite, des mutations se produisant avec une fréquence élevée provoquent une variation du nombre de répétitions du motif de base, ce qui engendre une forte diversité allélique. Ces mutations pourraient être dues à un « glissement » de la polymérase lors de la réplication ou à des crossing-over inégaux (JARNE et LAGODA, 1996, pour revue).

Il existe parfois des différences significatives de quantité d'ADN nucléaire total entre espèces proches interfertiles ou même entre groupes génétiques au sein d'une même espèce. Ces différences sont vraisemblablement à mettre sur le compte de variations dans la quantité d'ADN répété. Elles peuvent être mises

en évidence à l'aide de la cytométrie en flux (DOLEZEL, 1991 ; BENNETT et LEITCH, 1995). Elles peuvent parfois apporter des éléments utiles pour différencier des groupes au sein d'une espèce ou d'un complexe d'espèces.

## La nature de l'information révélée par les différentes techniques de marquage génétique

Nous abordons ici les aspects méthodologiques qui différencient les cinq techniques de marquage génétique envisagées, puis nous examinons ces différences pour voir comment elles affectent l'estimation de la diversité génétique et les similarités entre individus.

### La nature et l'interprétation génétique du polymorphisme

Chaque type de marqueur repose sur une méthodologie qui va conditionner la nature de l'information obtenue. Il est possible de distinguer deux grandes catégories de marqueurs : les marqueurs qui permettent de révéler une série de plusieurs allèles pour chaque locus étudié, ce sont les marqueurs dits multialléliques, et les marqueurs qui permettent de révéler la présence ou l'absence d'un seul allèle pour chaque locus, et cela simultanément pour un grand nombre de locus, ce sont les marqueurs de type empreinte génétique, ou *fingerprinting*.

#### LES MARQUEURS MULTIALLÉLIQUES

Les marqueurs multialléliques comprennent les isoenzymes, les RFLP et les microsatellites. Les allèles mis en évidence sont le plus souvent codominants : les deux allèles homologues sont observables chez les individus hétérozygotes.

#### Les isoenzymes et les RFLP

Pour les isoenzymes, un locus est défini par une fonction catalytique envers un substrat spécifique. Pour les RFLP, un locus correspond à la région du génome qui s'hybride avec une sonde, dont la taille va de quelques centaines à quelques milliers de paires de bases. Dans les deux cas, les allèles se distinguent par leur mobilité électrophorétique. Dans le cas le plus simple, un allèle est matérialisé par une seule bande de poids moléculaire spécifique. Néanmoins, plusieurs facteurs peuvent compliquer ce schéma.

- ❑ Pour les RFLP, un allèle peut correspondre à une combinaison de plusieurs bandes s'il existe un ou des sites de coupure de l'enzyme de restriction dans la séquence d'ADN homologue à celle de la sonde.
- ❑ Pour les isoenzymes, les enzymes polymériques peuvent engendrer de nouvelles bandes chez les hétérozygotes.
- ❑ Un substrat (isoenzymes) ou une sonde (RFLP) peuvent révéler simultanément les allèles de plusieurs locus paralogues, c'est-à-dire de locus présentant la même séquence d'ADN ou des séquences très proches.
- ❑ Pour les RFLP, si plusieurs enzymes de restriction sont utilisées en combinaison avec une même sonde, chaque allèle correspond en principe à une combinaison spécifique de bandes révélées avec les différentes enzymes. En pratique, les reconstructions d'allèles à partir d'analyses de polymorphisme multienzymatique sont rarement effectuées car trop complexes. On se contente dans ce cas de noter la présence ou l'absence de chaque bande ou de chaque profil.

Avec les isoenzymes et les RFLP, il est donc souvent nécessaire d'étudier l'hérédité des bandes dans des descendance contrôlée pour interpréter les profils observés en terme de locus et d'allèles. L'interprétation génétique est généralement plus simple chez les plantes autogames, pour lesquelles chaque individu est le plus souvent homozygote pour tous ses locus.

Pour les isoenzymes, les séquences ciblées sont par définition des gènes codant pour des enzymes. Le polymorphisme est essentiellement dû à des mutations ponctuelles qui induisent le remplacement d'un acide aminé par un autre, modifiant ainsi la charge électrique globale de la protéine ou sa masse moléculaire et, par conséquent, son comportement électrophorétique. Cependant, de nombreuses mutations sont silencieuses du fait de la dégénérescence du code génétique et une mutation induisant un changement d'acide aminé ne provoque pas systématiquement une modification de la mobilité électrophorétique de la protéine. Cela a été mis en évidence, par exemple, pour les  $\alpha$ -amylases de la drosophile (INOMATA *et al.*, 1995).

Pour les RFLP, la nature des séquences ciblées dépend du type de sonde choisi. Il est possible d'utiliser des sondes révélant le polymorphisme de séquences répétées ou simple copie, de séquences codantes ou non codantes, de séquences nucléaires ou cytoplasmiques. Nous nous intéresserons essentiellement aux séquences nucléaires simple copie. Le polymorphisme observé est dû à des mutations affectant la séquence homologue de la sonde et surtout celle de ses régions flanquantes sur environ 10 kilobases. Il peut s'agir de mutations ponctuelles dans les sites de coupure de l'enzyme ou bien d'insertions ou de délétions entre ces sites. L'importance relative des deux phénomènes peut varier d'une espèce à l'autre. Les mutations ponctuelles qui n'affectent pas le site de coupure de l'enzyme ou les petites insertions-délétions peuvent passer inaperçues. Que ce soit pour les isoenzymes ou pour

les RFLP, un allèle correspond potentiellement à un ensemble de plusieurs séquences, dans la mesure où le pouvoir résolutif de ces deux techniques ne garantit pas la mise en évidence de toutes les mutations qui affectent la séquence du locus ciblé.

### Les microsatellites

Pour les microsatellites, un locus est défini par un site microsatellite accompagné de ses séquences flanquantes. Des amorces définies dans ces séquences, de part et d'autre du site microsatellite, permettent d'amplifier le locus à l'exclusion de tout autre dans le génome. Cette spécificité est garantie par la longueur des amorces (20 à 25 nucléotides). Elle rend infiniment faible la probabilité d'amplifier une autre séquence dans le génome par le seul fait du hasard. La source de polymorphisme recherchée est la variation du nombre de répétitions du motif de base du site microsatellite. Dans la plupart des cas, chaque allèle sera matérialisé par une bande de poids moléculaire spécifique, qui peut être traduite en nombre de répétitions du motif de base du microsatellite si le gel est suffisamment résolutif. Dans certains cas cependant, une partie des allèles provient de variations de séquence entre une des amorces et le site microsatellite (ORTI *et al.*, 1997). Les taux de mutation associés aux sites microsatellites sont 100 à 1 000 fois plus élevés que pour les isoenzymes (JARNE et LAGODA, 1996). Quelle que soit la nature du polymorphisme, les profils observés sont directement interprétables en termes de locus et d'allèles.

### LES EMPREINTES GÉNÉTIQUES

Deux techniques d'empreinte génétique sont présentées : les RAPD et les AFLP. Le terme « empreinte génétique » est couramment adopté car ces techniques permettent de révéler simultanément le polymorphisme d'un grand nombre de locus si bien que chaque individu a de grandes chances d'avoir un profil multilocus qui lui soit propre, comme le sont les empreintes digitales chez l'homme.

Avec les techniques RAPD et AFLP, la similarité de séquence entre bandes révélées par un même couple d'amorces n'autorise pas à établir de relations d'homologie. Elle est de 20 paires de bases au plus, correspondant à la somme de la séquence des deux sites d'amorçage, pour les RAPD. Elle est encore plus faible pour les AFLP, dans la mesure où le seul point commun entre toutes les bandes révélées par un couple d'amorces est la séquence des deux sites de restriction flanquants et celle des bases sélectives, ce qui représente au plus 16 paires de bases. Typiquement, une amplification RAPD permet de révéler de 5 à 20 bandes et une amplification AFLP, de quelques dizaines à une centaine de bandes. L'interprétation génétique des profils observés revient à considérer chaque bande comme l'allèle d'un locus particulier. L'identité du locus et de l'allèle repose uniquement sur la mobilité électrophorétique de la bande. Pour chaque locus, le seul autre allèle possible est l'absence de la bande. La bande est donc un allèle dominant puisque les individus homozygotes qui pré-

sentent la bande ne sont pas discernables des hétérozygotes. Pour chaque locus, il est probable que l'absence de la bande corresponde à un ensemble hétérogène de séquences.

Bien que les bandes amplifiées par un même couple d'amorces RAPD ou AFLP aient peu de chance de correspondre à des séquences homologues, il est possible, dans un petit nombre de cas, que deux bandes différentes matérialisent deux allèles d'un même locus. Il n'est généralement pas possible de s'en apercevoir dans des études de diversité, mais ce phénomène peut être observé quand on utilise ces techniques pour faire de la cartographie génétique.

Les AFLP et les RAPD permettent, en principe, de cibler tous les compartiments du génome et tous les types de séquences. Le polymorphisme révélé par les RAPD correspond vraisemblablement à la modification d'un site d'amorçage à la suite de la substitution d'une base par une autre ou à un phénomène d'insertion-délétion. Il peut également s'agir d'une insertion entre les deux sites d'amorçage qui éloignerait suffisamment un site de l'autre pour rendre l'amplification impossible (WILLIAMS *et al.*, 1990). Le polymorphisme observé avec les AFLP correspond probablement à la modification d'un site de coupure d'une des deux enzymes de restriction à la suite d'une mutation ponctuelle (Vos *et al.*, 1995).

Plusieurs autres techniques d'empreinte génétique donnent des profils s'interprétant d'un point de vue génétique de la même façon que les RAPD et les AFLP. On peut citer les DAF (*DNA amplification fingerprinting* ; CAETANO-ANOLLES et GRESSHOFF, 1991), qui sont une variante des RAPD, les ISSR (*inter-simple sequence repeat analysis* ; ZIETKIEWICZ *et al.*, 1994), qui consistent à amplifier les séquences situées entre les sites microsatellites, ou encore l'utilisation de sondes microsatellites en RFLP (RUS-KORTEKAAS *et al.*, 1994).

## La comparaison des cinq techniques

Cette partie a pour but de comparer le polymorphisme et les similarités génétiques mesurés avec les différentes techniques de marquage sur la base de données expérimentales.

### LE POLYMORPHISME DES TECHNIQUES MULTIALLÉLIQUES

Il est possible de comparer le polymorphisme obtenu avec les différentes techniques multialléliques en utilisant des paramètres usuels de génétique des populations tels que le pourcentage de locus polymorphes, le nombre moyen d'allèles par locus et l'indice de diversité de Nei, appliqués à un même échantillon d'individus. Des exemples de comparaisons entre isoenzymes et RFLP et entre RFLP et microsatellites sont donnés dans le tableau 1. Ces résultats doivent être interprétés prudemment dans la mesure où les échantillons d'individus sont généralement faibles et où les nombres de locus observés sont souvent



Tableau 1. Comparaison du taux de polymorphisme, de la diversité de Nei et du nombre moyen d'allèles par locus entre isoenzymes (iso) et RFLP et entre microsatellites (SSR) et RFLP, dans différentes études. Pour les RFLP, une seule enzyme de restriction a été utilisée sauf mention contraire.

Espèce	Technique	Nombre d'individus étudiés	Nombre de locus étudiés	Taux de polymorphisme (%)	Diversité de Nei	Nombre moyen d'allèles par locus	Référence
<i>Brassica campestris</i>	Iso	285	5	60	0,41	2,7	MITCHELL et QUIROS (1992)
	RFLP	277	4	100***	0,60	8,5	
<i>Hordeum vulgare</i>	Iso	268	7	100	0,44	4,9	ZHANG <i>et al.</i> (1993)
	RFLP	240	13	100***	0,47	4,2	
<i>Populus tremuloides</i>	Iso	118	14	77	0,25	2,8	LIU et FURNIER (1993)
	RFLP	91	41	71	0,25	2,7	
<i>Populus grandidentata</i>	Iso	96	14	29	0,08	1,4	LIU et FURNIER (1993)
	RFLP	75	37	65	0,13	1,8	
<i>Zea mays</i>	Iso	21	22	68	—	2,1	MESSMER <i>et al.</i> (1991)
	RFLP	21	144 (79)*	94	—	3,34**	
<i>Zea mays</i>	Iso	31	27	—	—	2,2	GERDES et TRACY (1994)
	RFLP	43	71	—	—	4,1	
<i>Zea mays</i>	Iso	445	20	75	0,23	2,4	DUBREUIL et CHARCOSSET (1998)
	RFLP	285	35	100	0,61	6,3	
<i>Hordeum vulgare</i>	RFLP	18	114 (42)*	—	0,32	2,62**	RUSSELL <i>et al.</i> (1997)
	SSR	18	13	100	0,57	5,38	
<i>Glycine max</i>	RFLP	19	114	100***	0,38	2,15	MORGANTE <i>et al.</i> (1994)
	SSR	19	4	100	0,52	4,25	
<i>Glycine sp.</i>	RFLP	12	110	25	0,41	—	POWELL <i>et al.</i> (1996)
	SSR	12	36	100	0,60	4,28	
<i>Oryza sativa</i>	RFLP	14	18	33	0,10	2,3	WU et TANKSLEY (1993)
	SSR	14	8	100	0,70	5,2	
<i>Oryza sativa</i>	RFLP	20	12	83	0,32	2,5	OLUFOWOTE <i>et al.</i> (1997)
	SSR	20	10	100	0,62	7,4	
<i>Zea mays</i>	RFLP	12	96	—	0,58	—	TARAMINO et TINGEY (1996)
	SSR	12	34	100	0,76	6,56	

\* x (y) : x nombre de couples enzyme-sonde, y nombre de sondes.

\*\* Nombre moyen de bandes par couple enzyme-sonde.

\*\*\* Sondes RFLP prétriées pour leur polymorphisme.

déséquilibrés entre les techniques. Dans la plupart des études mentionnées, le polymorphisme RFLP est observé avec une seule enzyme. Même avec cette restriction, les RFLP apparaissent au moins aussi polymorphes que les isoenzymes et parfois nettement plus. Ce résultat est en accord avec ce que l'on sait de la nature du polymorphisme détecté. Les isoenzymes permettent de détecter le polymorphisme de séquences géniques uniquement au travers du filtre sévère que constituent la traduction et la dégénérescence du code génétique. Les RFLP permettent de détecter le polymorphisme de régions codantes et non codantes. De plus, quand des sondes correspondant à des gènes sont utilisées, c'est essentiellement le polymorphisme des régions flanquantes des gènes ciblés qui est révélé.

Dans le tableau 1, les microsatellites apparaissent systématiquement beaucoup plus polymorphes que les RFLP, ce qui est en accord avec leur taux de mutation particulièrement élevé. Ces observations sont confortées par de nombreuses autres études. SAGHAI-MAROOF *et al.* (1994), par exemple, ont trouvé jusqu'à 28 allèles différents pour un locus microsatellite chez l'orge.

#### LE POLYMORPHISME RÉVÉLÉ PAR L'ENSEMBLE DES TECHNIQUES

Le nombre d'allèles par locus et l'indice de diversité de Nei perdent une partie de leur signification génétique quand on les applique aux techniques d'empreinte. Le nombre d'allèles par locus est par nature limité à deux (présence et absence de la bande). En conséquence, l'indice de diversité de Nei est plafonné à 0,5. Par ailleurs, le calcul de l'indice de Nei repose sur la connaissance des fréquences alléliques. Du fait que les marqueurs sont dominants, ces fréquences sont directement accessibles dans seulement deux cas particuliers : si on fait l'hypothèse qu'il y a appariement au hasard des individus (l'équilibre de Hardy-Weinberg est respecté) ou bien s'il y a autogamie et que l'on fait l'hypothèse que tous les individus sont homozygotes pour tous les locus. Si ces hypothèses ne sont pas acceptables, le calcul des fréquences alléliques n'est pas possible à moins de disposer d'une estimation indépendante de l'indice de fixation  $F_{IS}$  (LYNCH et MILLIGAN, 1994 ; KREMER, 1998). Si on fait abusivement l'une de ces deux hypothèses, l'estimation de l'indice de Nei peut être biaisée. Cet indice est néanmoins parfois employé à titre indicatif sur les marqueurs de type empreinte génétique. Dans l'étude de LIU et FURNIER (1993) sur deux espèces de peuplier, les indices de Nei calculés pour les marqueurs RAPD sont d'environ 0,3 pour chacune des deux espèces. Ces valeurs sont supérieures à celles qui sont obtenues pour les isoenzymes et les RFLP sur les mêmes individus (tableau 1). Néanmoins, l'hypothèse faite sur l'équilibre de Hardy-Weinberg empêche de conclure de façon claire quant à une meilleure aptitude des RAPD à révéler du polymorphisme.

Dans le cas particulier des espèces autogames, les fréquences alléliques peuvent être directement estimées à partir de la fréquence des génotypes et l'indice de Nei peut être calculé. Sur le soja, par exemple, POWELL *et al.* (1996) ont trouvé

des indices de Nei de 0,31 et 0,32 pour les RAPD et les AFLP, respectivement. Ces valeurs sont plus faibles que celles qui sont obtenues pour les RFLP et les microsattellites sur le même échantillon d'individus (tableau 1).

Les techniques d'empreinte génétique ont aussi la particularité de révéler simultanément le polymorphisme de plusieurs locus par unité expérimentale. Une unité expérimentale correspond à une piste sur un gel. Pour tenir compte de cette caractéristique qui a une grande importance pratique pour l'expérimentateur, il peut être utile de comparer le polymorphisme non plus au niveau du locus mais au niveau de l'unité expérimentale. Plusieurs mesures de la diversité peuvent alors être envisagées, par exemple, le nombre moyen de bandes polymorphes par unité expérimentale, le PIC (*polymorphism information content* ; WEIR, 1990) et, pour les espèces autogames, l'index marqueur (POWELL *et al.*, 1996). Le PIC s'écrit comme l'indice de Nei, en remplaçant les fréquences alléliques par les fréquences des profils multilocus. L'index marqueur est défini comme le produit de l'indice de diversité de Nei par le nombre de bandes polymorphes par unité expérimentale. Il s'interprète comme le nombre moyen de bandes par unité expérimentale différenciant deux individus pris au hasard. Sur la base de ces critères, des comparaisons de techniques, effectuées chez l'orge (RUSSELL *et al.*, 1997), le soja (POWELL *et al.*, 1996) et le pois (LU *et al.*, 1996), montrent que les AFLP se distinguent par un nombre moyen de bandes polymorphes par unité expérimentale, un PIC et un index marqueur nettement plus élevés que pour les autres techniques. Les AFLP présentent un intérêt, non pas parce qu'ils révèlent plus de diversité à chaque locus, mais parce qu'ils permettent d'observer simultanément un grand nombre de locus polymorphes.

#### LES SIMILARITÉS GÉNÉTIQUES

Quand on observe un ensemble d'individus pour plusieurs locus avec une technique donnée, il est possible de calculer une similarité génétique pour tous les couples d'individus (PERRIER *et al.*, 1999). L'ensemble des similarités peut être synthétisé par une analyse multivariée ou une représentation arborée. Il est légitime de se demander si les différentes techniques de marquage donnent une même estimation de la proximité génétique entre les individus.

Les résultats expérimentaux permettant d'aborder ces questions sont encore peu nombreux. Quelques-uns ont été regroupés dans le tableau 2 sous la forme de corrélations entre indices de similarité calculés pour différents couples de techniques. Des exemples de relations entre distributions sont donnés dans la figure 1. L'absence d'homogénéité entre les tests, les niveaux de signification retenus, les gammes d'apparementement entre individus et la puissance des dispositifs (nombre d'individus étudiés) compliquent la synthèse des observations. Néanmoins, en règle générale, les corrélations entre les similarités calculées à partir de plusieurs techniques de marquage sont significatives mais d'intensité variable.

Tableau 2. Corrélation entre similarités calculées à partir de différentes techniques de marquage moléculaire.

Référence	Espèce	N*	Test**	Comparaison***	Correlation (r)	Seuil (P)
LU <i>et al.</i> (1996)	pois	10	Tm	RFLP - RAPD	0,5	< 0,05
				RFLP - AFLP	0,7	< 0,05
				RAPD - AFLP	0,6	< 0,05
POWELL <i>et al.</i> (1996)	soja	10	Tm	RFLP - RAPD	0,24	ns
				RFLP - AFLP	0,42	< 0,01
				RFLP - SSR	0,18	ns
				RAPD - AFLP	0,45	< 0,01
				RAPD - SSR	0,15	ns
				AFLP - SSR	0,14	ns
BEER <i>et al.</i> (1993)	<i>Avena sterilis</i>	177	Tm	RFLP - Iso	0,26	< 0,005
HEUN <i>et al.</i> (1994)	<i>Avena sterilis</i>	24	Tm	RFLP - Iso	0,36	< 0,01
PEAKALL <i>et al.</i> (1995)	<i>Buchloe dactyloides</i>	48	Tm	RAPD - Iso	0,63	< 0,001
MESSMER <i>et al.</i> (1991)	maïs	21	Cp	RFLP - Iso	0,23	< 0,01
GERDES et TRACY (1994)	maïs	31	Cp	RFLP - Iso	0,26	< 0,01
PRABHU <i>et al.</i> (1997)	soja	10	Cp	RFLP - DAF	0,70	< 0,01
RUSSELL <i>et al.</i> (1997)	orge	18	Cr	RFLP - RAPD	0,20	< 0,01
				RFLP - AFLP	0,71	< 0,01
				RFLP - SSR	0,51	< 0,01
				RAPD - AFLP	0,11	ns
				RAPD - SSR	0,24	< 0,01
				AFLP - SSR	0,52	< 0,01
DOS SANTOS <i>et al.</i> (1994)	choux	45	Cr	RFLP - RAPD	0,75	< 0,01
ENGQUIST et BEEKER (1994)	colza	17	Cr	RFLP - RAPD	0,76	< 0,01
				RFLP - Iso	0,53	< 0,01
				RAPD - Iso	0,67	< 0,01
THORMANN <i>et al.</i> (1994)	<i>Brassica spp.</i>	18	Tm	RFLP - RAPD	0,93	—

\* Nombre d'individus utilisés pour calculer la corrélation.

\*\* Tm : test de Mantel ; Cp : coefficient de corrélation de Pearson ; Cr : coefficient de corrélation de rang de Spearman.

\*\*\* Iso : isoenzymes ; SSR : microsatellites ; DAF : *DNA amplification fingerprinting*.

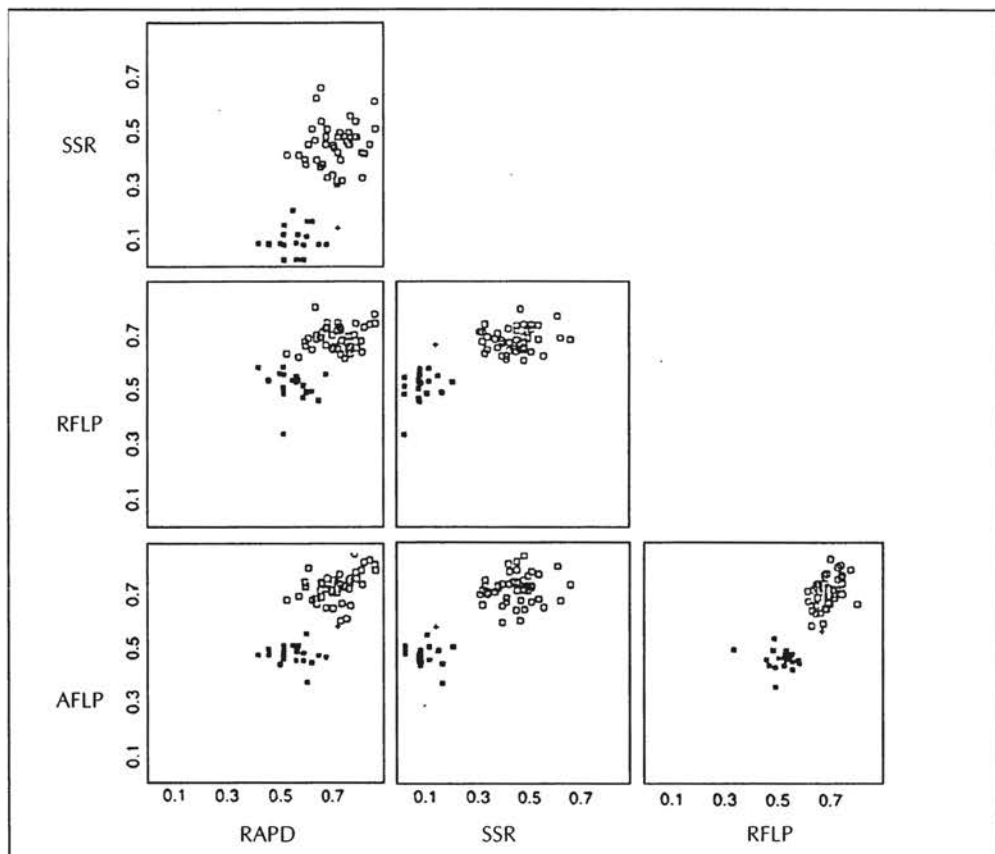


Figure 1. Relation à deux dimensions entre les similarités génétiques calculées à partir des RFLP, des AFLP, des microsatellites (SSR) et des RAPD chez le soja. Les carrés blancs représentent les similarités entre accessions de l'espèce *Glycine max*, les croix représentent les similarités entre les accessions de l'espèce *G. soja* et les carrés noirs les similarités entre accessions des deux espèces, d'après POWELL et al. (1996).

Les valeurs faibles de corrélation peuvent s'expliquer en grande partie par un biais dû à l'échantillonnage des locus. Cela est particulièrement vrai pour les isoenzymes dans la mesure où le nombre de locus analysés est le plus souvent inférieur à 20. MESSMER et al. (1991) ont calculé les écarts types des indices de similarité par la méthode du *jackknife* (MILLER, 1974). Ils ont montré que les valeurs des écarts types obtenues avec 22 locus isoenzymatiques étaient deux à trois fois supérieures aux valeurs obtenues avec 144 couples enzyme-sonde RFLP. Cela reste vrai pour les techniques moléculaires, qui permettent pourtant de révéler plus de locus. DOS SANTOS et al. (1994) ont montré par *bootstrap* (EFRON et TIBSHIRANI, 1986) que la distribution des corrélations entre indices calculés pour 100 échantillons aléatoires de 56 bandes RAPD pris deux à deux était largement superposée à la distribution des corrélations entre indices pour 100 échantillons aléatoires de 56 bandes RAPD et de 56 bandes RFLP pris deux à deux.

La qualité technique des manipulations a une répercussion directe sur la fiabilité du codage des données et peut donc être aussi une source de divergence dans l'estimation des similarités. Le codage peut devenir très laborieux si cette qualité n'est pas irréprochable, en particulier pour les AFLP qui révèlent simultanément de nombreuses bandes. Plusieurs études ont signalé des problèmes de reproductibilité pour les RAPD, qui imposent de répéter plusieurs fois l'analyse de chaque échantillon sous peine de commettre d'importantes erreurs de codage de données (SALIMATH *et al.*, 1995 ; YANG *et al.*, 1996). Les profils sont, en outre, difficiles à reproduire d'un laboratoire à l'autre, ce qui constitue un handicap pour les travaux menés en réseau (JONES *et al.*, 1997). Enfin, il est vraisemblable que la compétition pour les amorces puisse entraîner dans certains cas la non-amplification d'une bande présente à l'état hétérozygote, du fait de sa plus grande dilution (HALLDEN *et al.*, 1996).

Les microsatellites permettent de révéler plus d'allèles à un même locus que les autres techniques, ce qui se traduit par un indice moyen de similarité entre individus généralement beaucoup plus faible. Cela pourrait expliquer la faiblesse de la corrélation entre cette technique et les autres, surtout quand les individus sont peu apparentés (la corrélation n'est pas significative dans l'étude de POWELL *et al.* (1996) ; tableau 2).

Les RAPD semblent parfois donner une estimation de la similarité qui diffère de celle des autres techniques, en particulier quand les individus sont éloignés (appartenance à deux espèces différentes). Cette divergence a été constatée par THORMANN *et al.* (1994) dans le genre *Brassica* et par POWELL *et al.* (1996) dans le genre *Glycine*. THORMANN *et al.* (1994) ont montré par hybridation que, dans trois cas sur quinze, des bandes de même mobilité ne sont pas homologues chez les différentes espèces étudiées. Les RAPD pourraient être plus sensibles à ce type de confusion que les autres marqueurs.

L'influence des mécanismes évolutifs qui sous-tendent la variation observée ne peut pas être complètement écartée pour expliquer les différences de similarité entre marqueurs. N'GORAN *et al.* (1994), par exemple, ont montré que, chez le cacaoyer, la moitié des bandes amplifiées par RAPD correspondait à des séquences répétées et que les bandes amplifiant des séquences répétées et des séquences uniques ne donnaient pas la même structuration du matériel génétique.

#### LA RELATION AVEC LA GÉNÉALOGIE ET LES CARACTÈRES MORPHOLOGIQUES

On peut se demander si les similarités calculées à partir de marqueurs hiérarchisent de la même façon les relations entre individus que la généalogie ou les caractères morphologiques.

La relation entre la similarité évaluée avec les marqueurs et le coefficient d'apparentement de Malécot a été étudiée chez plusieurs plantes pour différents marqueurs (tableau 3). Dans l'ensemble, les corrélations entre les deux types d'information sont faibles mais significatives. La faiblesse des corréla-

Tableau 3. Corrélation entre le coefficient d'apparementement  $f$  de Malécot et la similarité calculée à partir de marqueurs.

Référence	Espèce	Technique	Test*	N**	Corrélation	Seuil ( $P$ )
COX <i>et al.</i> (1985)	soja	Isoenzymes	Cr	27	0,15	< 0,01
				32	0,40	< 0,01
				39	0,45	< 0,01
O'DONOUGHUE <i>et al.</i> (1994)	avoine	RFLP	Tm	55	0,32	< 0,001
GRANER <i>et al.</i> (1994)	orge	RFLP	Cr	21	0,21	< 0,05
				17	0,42	< 0,01
GERDES et TRACY (1994)	maïs	RFLP	Cp	42	0,54	< 0,01
		Isoenzymes	Cp	31	0,32	< 0,01
TINKER <i>et al.</i> (1993)	orge	RAPD	Cr	27	0,51	< 0,01
PLASCHKE <i>et al.</i> (1995)	blé	Microsatellites	Cr	40	0,55	< 0,001

\* Tm : test de Mantel ; Cp : coefficient de corrélation de Pearson ; Cr : coefficient de corrélation de rang.

\*\* Nombre d'individus utilisés pour calculer la corrélation.

tions pourrait s'expliquer par le fait que le calcul des coefficients d'apparementement repose sur des hypothèses irréalistes, en particulier l'absence d'apparementement entre les ancêtres lointains dont on ne connaît pas la généalogie. La dérive génétique et la sélection ne sont pas non plus prises en compte. Cela fait dire à GRANER *et al.* (1994) que l'apparementement calculé avec les marqueurs moléculaires, qui ne néglige pas ces paramètres, pourrait être un outil plus pertinent que le  $f$  de Malécot.

Des corrélations faibles mais significatives ont été observées entre similarités calculées à partir de marqueurs moléculaires et distances calculées à partir de variables morphologiques. Par exemple chez *Avena sterilis*, une corrélation de 0,24 ( $P < 0,005$ ) a été observée entre RFLP et morphologie et une corrélation de 0,13 ( $P < 0,005$ ), à été trouvée entre isoenzymes et morphologie (BEER *et al.*, 1993). Chez le maïs, des corrélations de 0,167 et 0,124 ( $P < 0,01$ ) ont été obtenues, dans ces deux cas respectivement (GERDES et TRACY, 1994). Plus précisément, DILLMANN *et al.* (1997) soulignent que la relation entre les deux types d'information n'est pas linéaire chez le maïs. Une faible distance calculée avec les marqueurs RFLP correspond toujours à une faible distance morphologique. En revanche, une forte distance calculée avec les marqueurs peut correspondre à une forte ou à une faible distance morphologique (figure 2). Cette relation triangulaire peut s'expliquer par le fait que, pour un caractère morphologique donné, différentes combinaisons de gènes peuvent correspondre à un même phénotype. Par ailleurs, le déséquilibre de liaison entre les marqueurs et les gènes impliqués dans le caractère est variable suivant l'échantillonnage des locus ; s'il est faible les deux types d'information seront indépendants.



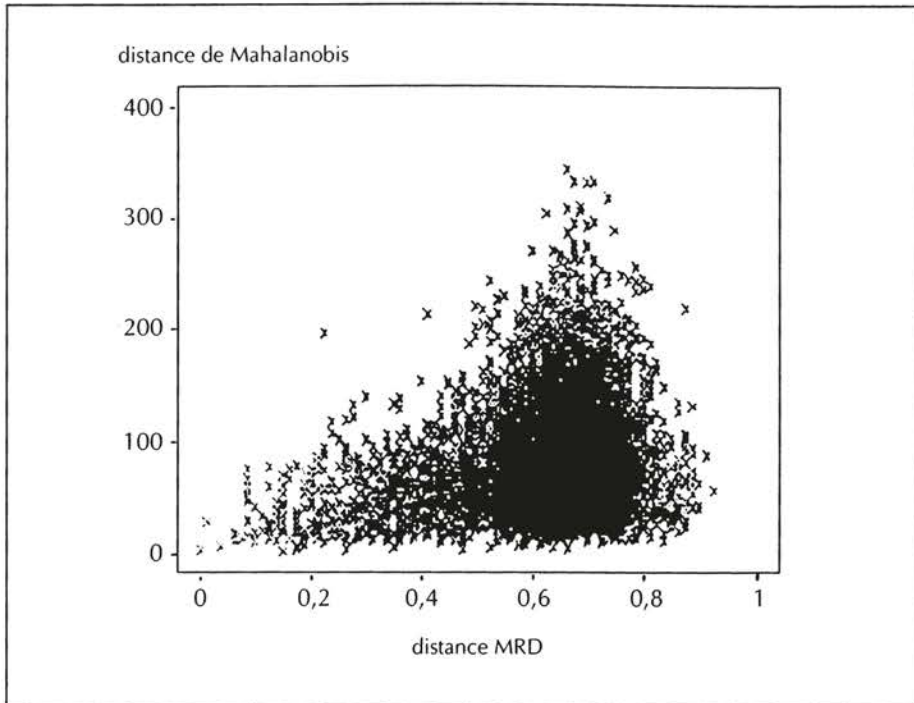


Figure 2. Relation entre la similarité (distance de Rogers, MRD) calculée à partir des marqueurs RFLP et la distance calculée à partir de données morphologiques (distance de Mahalanobis) chez le maïs, d'après DILLMANN et al. (1997).

## La mise en œuvre pratique des techniques de marquage

Deux aspects pratiques majeurs distinguent les techniques de marquage génétique : leur facilité de mise en œuvre et leur coût.

### La facilité de mise en œuvre

Les isoenzymes ne nécessitent pas de gros équipements de laboratoire et utilisent de petites quantités de matériel végétal. Leur mise en œuvre sur une espèce nouvelle peut être plus ou moins laborieuse. Elle dépend de la difficulté à trouver un équilibre entre le tampon d'extraction des protéines et le tampon d'électrophorèse. Cet équilibre est plus difficile à atteindre chez les plantes dont les organes sont riches en polyphénols. Le principal handicap des isoenzymes vient du faible nombre de locus qu'elles permettent de révéler. Ce nombre est toujours inférieur à 50 et le plus souvent inférieur à 20.

Pour les marqueurs moléculaires, on peut discuter la facilité d'utilisation des différentes techniques selon trois critères : la lourdeur de mise en œuvre lors du démarrage d'un programme, la quantité d'ADN nécessaire pour analyser chaque échantillon et les équipements de laboratoire indispensables pour une utilisation en routine.

Les techniques d'empreinte génétique, RAPD et AFLP, peuvent être utilisées directement sur n'importe quelle plante sans étape préliminaire. En revanche, pour les RFLP et les microsatellites, la difficulté de mise en œuvre dépend beaucoup du travail déjà réalisé par d'autres équipes sur la plante étudiée. Si rien n'a déjà été fait, ces techniques peuvent être très laborieuses à mettre en route.

Pour les RFLP, il faut disposer d'une source de sondes, ce qui nécessite la construction et l'entretien d'une banque d'ADN génomique ou d'ADNc. Dans les groupes taxonomiques très étudiés comme les poacées, il est cependant possible d'utiliser des sondes dites hétérologues, provenant d'autres espèces. Ainsi, un ensemble de 180 sondes de riz, d'orge et d'avoine, ayant un spectre d'hybridation large sur l'ensemble des poacées, a été défini ; il est distribué par l'université Cornell aux Etats-Unis (VAN DEYNE *et al.*, 1998). Ces sondes peuvent, en principe, être utilisées sur toutes les espèces de cette famille botanique.

Les microsatellites requièrent une première étape de criblage de banque et de séquençage pour définir des amorces dans les régions flanquantes des sites microsatellites, ce qui représente un lourd et coûteux travail de biologie moléculaire. Chez les plantes les plus étudiées, comme le blé, le maïs et le riz, les séquences des amorces de nombreux sites microsatellites sont publiées et cette étape n'est donc plus nécessaire, mais ce n'est pas le cas pour la plupart des plantes tropicales. Par ailleurs, le transfert de l'information de séquence d'une espèce à l'autre n'est souvent pas possible même entre espèces taxonomiquement proches.

La quantité d'ADN nécessaire peut être un facteur limitant lorsque l'ADN est difficile à extraire ou quand on travaille sur de grands échantillons. Cette quantité dépend beaucoup du fait que la technique utilisée repose ou non sur l'utilisation de la PCR. La PCR nécessite de faibles quantités d'ADN. Pour les RAPD et les microsatellites, quelques nanogrammes d'ADN sont nécessaires pour faire une amplification ; pour les AFLP, quelques centaines de nanogrammes suffisent. A titre de comparaison, il faut de 2 à 10 microgrammes d'ADN par piste pour les RFLP. Par ailleurs, les microsatellites peuvent être mis en œuvre avec de l'ADN de qualité médiocre obtenu avec une technique d'extraction simple du fait de la grande spécificité d'hybridation des amorces.

Les microsatellites et les RAPD peuvent être utilisés en routine dans un laboratoire équipé de façon rudimentaire. En revanche, les RFLP et les AFLP reposent sur le marquage radioactif ou biochimique de sondes ou d'amorces et requièrent donc des infrastructures de laboratoire plus importantes. Les AFLP peuvent éventuellement être révélés avec un colorant de l'ADN comme le nitrure

d'argent (CHO *et al.*, 1996). Cette technique est moins contraignante en terme d'infrastructure mais la lecture des profils obtenus est souvent plus délicate.

## Le coût

Le coût semble *a priori* un élément déterminant dans le choix d'une technique de marquage moléculaire. Le coût d'une opération de génotypage en routine dépend de la disponibilité d'un certain nombre d'outils moléculaires, comme les sondes pour les RFLP ou les amorces pour les microsatellites, dont la mise au point peut entraîner des frais importants et variables (fabrication et entretien d'une banque de sondes, définition d'amorces microsatellites). Si on fait abstraction de ces frais, pour une opération donnée, le calcul du coût des différentes techniques de marquage peut être divisé en deux composantes : le prix de revient du « point » élémentaire (donnée d'un marqueur pour un individu) et son temps d'obtention. Le prix de revient du point dépend du poids du laboratoire dans la négociation des fournitures et de sa localisation géographique par rapport au matériel végétal et aux fournisseurs. C'est la taille du laboratoire évaluée en termes de consommation de produits ou de volume de résultats obtenus pour un type de marqueur qui influence cette composante. Le temps d'obtention est lié aux coûts salariaux et aux coûts d'immobilisation du matériel. C'est la taille du laboratoire, mais évaluée en termes d'équipement et de masse critique d'opérateurs (technicité), qui influence cette seconde composante. Le coût est donc une information qu'il est très difficile de transposer d'un laboratoire à un autre.

RAGOT et HOISINGTON (1993) ont comparé dans le laboratoire du CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo, Mexique), le coût de deux types de marqueurs, RAPD et RFLP, en distinguant pour ces derniers le marquage radioactif des sondes (rRFLP) et le marquage biochimique (cRFLP). La conclusion de leur travail est que le coût relatif des différentes techniques dépend de la taille des échantillons traités. Les RAPD sont moins chers pour les petits échantillons et, même si leur coût diminue, ils deviennent plus chers en valeur relative pour de grands échantillons. Cela est principalement dû à l'économie d'échelle que permet la réutilisation des membranes pour les marqueurs RFLP. RAGOT et HOISINGTON (1993) estiment qu'au-delà de trois réutilisations le coût commence à baisser de façon importante. Pour les cRFLP, il est difficile de dépasser en routine une dizaine d'utilisations. Pour les rRFLP, le nombre d'utilisations peut osciller en routine entre 15 et 20 d'après notre expérience. Le coût de l'extraction de l'ADN peut, selon RAGOT et HOISINGTON (1993), représenter la moitié du prix de revient des RAPD. Pour abaisser ce coût, MOHAN *et al.* (1997) suggèrent d'utiliser la technique du *squash* de tissus végétaux (LANGRIDGE *et al.*, 1991). Les concentrations en polysaccharides et autres « polluants » de l'ADN présents dans les tissus de bon nombre de plantes tropicales rendent souvent cette technique impraticable. Il n'y a donc

pas de recette universelle, mais le ton très affirmatif de MOHAN *et al.* (1997) montre qu'il est très difficile d'apprécier l'ensemble des situations possibles.

RISTERUCCI (1997) a comparé le prix de revient en routine de différents types de marqueur dans le cadre particulier du programme Biotrop du Cirad. A titre d'exemple, pour le génotypage de 100 individus avec 100 marqueurs, le coût du point était de 1,10 franc pour les AFLP, de 1,71 franc pour les RAPD, de 3,14 francs pour les RFLP, de 3,21 francs pour les microsatellites et de 4 francs pour les isoenzymes. Il ressort de cette comparaison que le prix de revient des techniques multialléliques (RFLP, microsatellites, isoenzymes) est supérieur à celui des techniques d'empreinte génétique. Le coût élevé des microsatellites provient du choix qui a été fait d'utiliser la radioactivité pour les révéler. Par ailleurs, quand on observe l'évolution du coût du point en fonction de la taille de l'échantillon, il apparaît qu'au-delà de 25 individus étudiés les économies d'échelle sont beaucoup plus importantes avec les marqueurs RFLP qu'avec les marqueurs fondés sur la PCR comme l'ont constaté RAGOT et HOISINGTON (1993). Les RFLP coûtent à peu près le même prix que les techniques d'empreinte génétique pour un échantillon de l'ordre de 200 individus.

Bien que certaines tendances se dégagent, il semble illusoire de vouloir donner un coût standard et universel pour chaque technique de marquage. Plutôt que de choisir une technique en fonction d'un coût calculé par d'autres dans un contexte forcément très spécifique, il semble plus réaliste de choisir la technique en fonction de l'information dont on a besoin et de rechercher *a posteriori* le moyen de l'appliquer au moindre coût. A ce titre, la sous-traitance auprès de prestataires de service pour certaines étapes comme le séquençage occasionnel, les accords avec d'autres laboratoires de recherche compétents dans un domaine précis et pour de périodes limitées, le recours à l'utilisation de kits pour de courtes périodes ou pour des projets ponctuels peuvent aujourd'hui apparaître comme des éléments modérateurs de prix, qui peuvent aller jusqu'à rendre le choix du type de marqueur totalement indépendant de son prix de revient.

## Conclusion

Les marqueurs biochimiques et moléculaires permettent d'obtenir une image de la diversité non sélectionnée existant au sein d'une collection. Chaque type de marqueur présente des avantages et des inconvénients et le choix d'une technique doit être raisonné au cas par cas en tenant compte des objectifs, des moyens disponibles et de l'état d'avancement des connaissances sur l'espèce étudiée. Dans l'état actuel des connaissances, les différentes techniques semblent donner des images assez semblables de la structuration génétique. Les différences constatées peuvent être attribuées à l'échantillonnage des locus, à des problèmes de répétabilité et d'homoplasie pour certaines techniques

comme les RAPD et à la nature du polymorphisme révélé et des séquences ciblées par chaque technique. Il est probable que nous manquons de recul par rapport à ce dernier point.

La relation entre le polymorphisme neutre révélé par les marqueurs et le polymorphisme des caractères morphoagronomiques utiles n'est pas évidente. Les marqueurs révèlent, de façon plus ou moins précise, les similarités de séquences entre individus pour un échantillon de locus. Les caractères morphoagronomiques mesurent les ressemblances entre individus sur la base de variables dont le niveau d'expression dépend de nombreux gènes potentiellement épistatiques. Les marqueurs peuvent cependant constituer un outil intéressant pour mieux échantillonner les gènes d'intérêt agronomique. Si l'espèce est structurée en groupes génétiques, cela signifie que des entités (populations, ensembles de génotypes) ont eu des histoires évolutives indépendantes, caractérisées par des échanges de gènes limités. Il est alors possible que le jeu d'allèles déterminant l'expression de caractères d'intérêt agronomique soit différent d'un groupe à l'autre en réponse à la dérive génétique ou à des pressions de sélection différentes. Dans ce cas, la prise en compte de la structuration génétique révélée par les marqueurs devrait permettre un meilleur échantillonnage des sources de variations pour les caractères manipulés par le sélectionneur.

## Annexe

### Description des techniques moléculaires

Nous proposons ici une description succincte des techniques AFLP, RAPD, RFLP et microsatellites. Pour des descriptions plus détaillées, des publications comme celles de KARP *et al.* (1997) ou de DE VIENNE et SANTONI (1998) pourront être consultées ainsi que le site Internet de l'IPGRI (<http://www.cgiar.org/ipgri/training>).

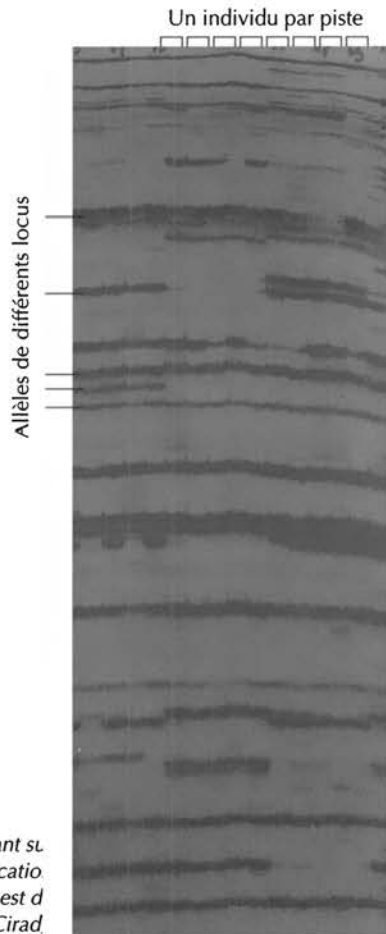
Les textes qui accompagnent la description des techniques comportent un certain nombre de mots signalés par un astérisque. Ces termes sont expliqués en utilisant un vocabulaire courant (quoique...) dans le glossaire (p. 35). Le site Internet de l'Institut Pasteur (<http://www.pasteur.fr/other/biology/francais/bio-docs-fr.html>) fournit des liens vers de nombreux dictionnaires.

Nous ne revenons pas ici sur les techniques isoenzymatiques, car bien qu'encore largement compétitives en terme d'analyse de diversité elles ont été déjà abondamment décrites.

### AFLP (*amplified fragment length polymorphism*)

Après extraction\*, l'ADN est hydrolysé\* sous l'action de deux enzymes de restriction\*. Des adaptateurs spécifiques des deux types de site de restriction\* sont ligués\* aux extrémités des fragments obtenus. L'ADN ainsi préparé est utilisé lors d'une amplification PCR\*, qui utilise deux amorces\* correspondant aux adaptateurs\* auxquels sont ajoutées de une à trois bases arbitraires en 3'. L'une des deux amorces est marquée\*. Les fragments produits par cette amplification sont séparés par électrophorèse\* sur un gel d'acrylamide dénaturant\*.

Le gel est exposé\* quelques jours au contact d'un film autoradiographique\*. Pour chaque individu, plusieurs dizaines de fragments, ou bandes, sont matérialisées. Chaque bande s'interprète comme un allèle d'un locus particulier.

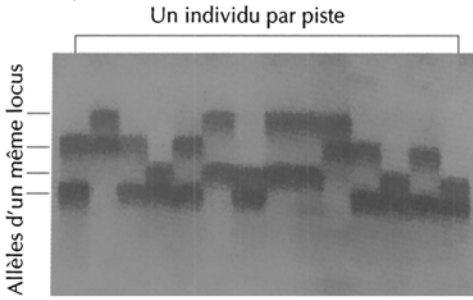


*Autoradiographie d'un gel d'acrylamide dénaturant sur lequel ont été séparés des fragments issus d'une amplification radioactive AFLP. La longueur totale du gel est de 40 centimètres. (photo R. Purba-Cirad)*

## Microsatellites

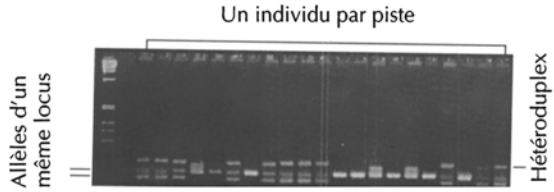
Après extraction\*, l'ADN est utilisé tel quel lors d'une amplification PCR\*, qui utilise deux amorces\* définies après séquençage\* dans les régions flanquantes\* d'un site microsatellite (répétition de séquences simples, SSR). Les fragments produits par cette amplification sont séparés par électrophorèse\* sur un gel d'agarose\* résolutif (concentré) ou, mieux encore, sur un gel d'acrylamide dénaturant\*.

C'est la variation du nombre de séquences simples répétées qui est mise en évidence.

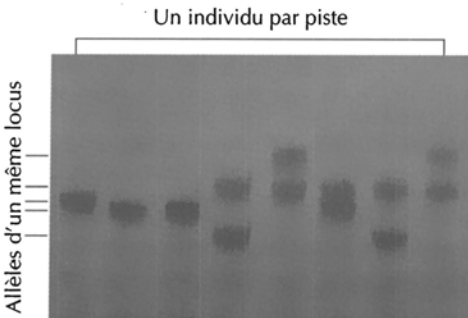


Autoradiographie d'un gel long (40 centimètres) d'acrylamide dénaturant après amplification par PCR radioactive d'un locus microsatellite d'hévéa. (photo M. Rodier-Goud et M. Seguin-Cirad).

La nature monocus des microsatellites permet une certaine souplesse quant au mode de mise en évidence des produits issus de l'amplification PCR\*. Les bandes observées matérialisent les allèles du locus ciblé. Chaque allèle correspond à un nombre particulier de répétitions du motif de base du microsatellite.



Photographie sous ultraviolets d'un gel d'agarose. Les produits d'amplification PCR d'un locus microsatellite d'hévéa sont colorés au bromure d'éthidium. L'intérêt de ce type de révélation réside dans sa facilité de mise en œuvre. Les inconvénients viennent du pouvoir résolutif plus faible que celui des gels d'acrylamide dénaturants, mais aussi de l'apparition de bandes supplémentaires appelées hétéroduplex, qui sont le fruit d'une réassociation artificielle en fin de PCR des allèles d'un individu hétérozygote. (photo M. Rodier-Goud-Cirad).



Photographie d'un gel court (12 centimètres) d'acrylamide dénaturant. Les produits d'amplification PCR d'un locus microsatellite de bananier sont colorés au nitrate d'argent. (photo J.L. Noyer-Cirad).



### RAPD (*random amplified polymorphic DNA*)

Après extraction\*, l'ADN est utilisé tel quel lors d'une amplification PCR\*, qui met en jeu une seule amorce\* de 10 bases définie arbitrairement. Pour que l'amplification ait lieu, cette amorce doit trouver deux séquences homologues\* inversées, situées à une distance telle que l'amplification puisse avoir lieu. Les fragments produits par cette amplification sont séparés par électrophorèse\* sur un gel d'agarose\* résolutif (concentré). Le résultat se présente sous la forme d'une photographie de gel coloré au bromure d'éthidium\*.

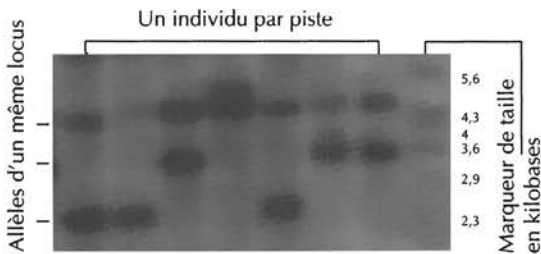
Chaque bande s'interprète comme un allèle d'un locus particulier.



Photographie sous ultraviolets d'un gel d'agarose à 1,4 %. Les produits d'amplification RAPD d'ADN de huit cacaoyers sont colorés au bromure d'éthidium. (photo J. N'goran-Cirad).

### RFLP (*restriction fragment length polymorphism*)

Après extraction\*, l'ADN est hydrolysé\* sous l'action d'une enzyme de restriction\*. Les fragments de taille variable sont ensuite séparés par électrophorèse\* sur un gel d'agarose\* puis transférés sur une membrane (Southern\*). Une sonde\* préalablement marquée\* est appliquée sur cette membrane. Elle s'hybride\* sur les fragments qui contiennent tout ou partie d'une séquence\* qui lui est homologue\*. Après plusieurs lavages\* de stringence\* croissante, la membrane est exposée\* quelques heures à quelques jours. Le résultat se présente sous la forme d'une autoradiographie\*. Les bandes observées matérialisent les différents allèles du locus ciblé. Chaque allèle correspond à une configuration particulière des sites de restriction autour de la région homologue de la sonde.



Autoradiographie des ADN de sept bananiers diploïdes hybridés avec une sonde marquée au  $^{32}P$ . (photo J.L. Noyer-Cirad).

## Glossaire

**3'-5'** : le double brin d'ADN est constitué de deux brins complémentaires antiparallèles, orientés 3'OH-5'P. La synthèse de l'ADN se faisant uniquement de 5' vers 3', les bases de l'amorce situées du côté 3' où s'effectue la synthèse auront plus d'importance dans la précision de l'homologie recherchée que les bases situées à l'autre extrémité de l'amorce lors des amplifications PCR.

**Adaptateur** : fragment d'ADN double brin, synthétisé, dont la séquence connue sert de site d'amorçage pour des amplifications PCR. Les adaptateurs sont au préalable ligués aux fragments d'ADN que l'on veut amplifier.

**Amorce** : fragment d'ADN simple brin, synthétisé, qui après s'être hybridé sur une séquence homologue d'un brin d'ADN dénaturé, sert de point d'ancrage pour démarrer la synthèse du brin complémentaire. Lors de l'hybridation sur de l'ADN génomique, c'est la longueur des amorces qui définit la spécificité du locus cible. Dans le cas des RAPD, une amorce d'environ dix bases, va trouver à s'hybrider plusieurs centaines, voire plusieurs milliers de fois dans chaque sens et permettre d'amplifier finalement quelques dizaines de fragments. Dans le cas des microsatellites, deux amorces de 20 bases vont statistiquement permettre d'amplifier un locus unique (hors duplications).

**Amplification PCR** : réaction de synthèse de fragments d'ADN à la suite d'une série de cycles de dénaturation, d'hybridation et de synthèse, qui permet d'amplifier en quantité une séquence d'ADN cible. En théorie, une amplification de  $n$  cycles donne  $2n$  fois le nombre de copies du fragment synthétisé au cours du premier cycle. Dans la pratique, c'est toujours moins. C'est la découverte d'ADN polymérase thermostables, résistant à plusieurs cycles de dénaturation de l'ADN à 95 °C, qui a entraîné le développement spectaculaire de cette technique.

**Autoradiographie** : film sensible impressionnable par un rayonnement radioactif ou « lumineux ». Ce rayonnement localisé sur un gel ou sur une membrane peut provenir d'une amorce marquée (marquage) à son extrémité ou d'un ADN marqué.

**Bromure d'éthidium** : agent intercalant de l'ADN qui se fixe à l'intérieur de la double hélice. Une fois fixé, il devient fluorescent sous l'action d'un rayonnement ultraviolet. Cette particularité est utilisée pour visualiser l'ADN sur un gel (produit dangereux).

**Electrophorèse** : technique qui permet de faire migrer sous l'action d'un champ électrique toute molécule polarisée. A densité de charge constante, les molécules les plus petites migrent plus loin de l'origine que les molécules les plus grosses. L'ADN étant chargé négativement en solution on peut donc séparer les fragments de taille différente.

**Endonucléase** : enzyme capable de catalyser en un point précis l'hydrolyse de l'ADN ce qui provoque une rupture des deux brins à l'intérieur de la chaîne (par opposition aux exonucléases qui provoquent des ruptures à l'extrémité de la chaîne).

**Enzyme de restriction** : endonucléase qui coupe l'ADN à un site spécifique. Ce site est défini par une séquence de bases. Les enzymes les plus fréquemment utilisées reconnaissent des séquences de 4 ou 6 bases.

**Exposition** : quand des fragments d'ADN ont été marqués (marquage) ils émettent un rayonnement lumineux ou radioactif plus ou moins intense. Le rayonnement est détecté au contact d'un film radiographique sensible (autoradiographie). Selon l'intensité du

signal, la détection demande un temps de contact plus ou moins long avec le support (gel, membrane...) qui contient l'ADN.

*Extraction* : il existe différentes techniques d'extraction de l'ADN contenu dans les végétaux. La forte proportion de « contaminants », qui accompagnent l'ADN lors de cette phase, oblige les manipulateurs à compléter l'extraction proprement dite par de longues étapes de purification. Cette difficulté d'extraction, particulière au monde végétal, a longtemps retardé le développement des techniques de biologie moléculaire par rapport au règne animal. Aujourd'hui encore certaines techniques demeurent difficiles d'accès à cause de la qualité parfois médiocre des ADN obtenus.

*Gel d'acrylamide dénaturant* : support de migration en électrophorèse, c'est un gel d'acrylamide, dans la composition duquel entre un chaotropique (en général de l'urée), qui maintient dissociés des brins complémentaires préalablement dénaturés. Cette technique de migration permet de séparer des fragments d'ADN qui, sans cela, ne pourraient pas pénétrer le gel au-delà d'une certaine taille.

*Gel d'agarose* : support de migration en électrophorèse. La résolution (pouvoir séparatif) augmente avec la concentration. Toutefois, on ne peut pas dépasser une concentration d'environ 4 %. La résolution ne dépasse pas une dizaine de paires de bases en routine même avec les agaroses les plus performants.

*Homologue* : voir séquence homologue.

*Hybrider-hybridation* : association de deux brins d'ADN complémentaires. La stringence peut être contrôlée par la température pour une charge saline donnée. Pour une amorce PCR de 10 à 30 bases, la température d'hybridation ( $T_m$ ) peut être calculée avec précision, bien que les formules de calcul aient été obtenues de façon empirique. Un abaissement de quelques degrés de la température calculée peut amener l'amorce à s'hybrider sur des sites qui ne lui sont pas totalement homologues en particulier dans son côté 5'. Pour les sondes utilisées en RFLP et qui sont de plus grande taille, le contrôle d'homologie n'est pas absolu (voir aussi séquence homologue).

*Hydrolyse* : voir endonucléase.

*Lavage* : en RFLP, technique qui permet d'éliminer d'une membrane les hybridations non spécifiques de sondes, c'est-à-dire celles qui ne présentent pas un degré d'homologie suffisamment élevé. Comme pour l'hybridation, c'est la charge saline et la température qui fixent la stringence.

*Ligier-ligation* : liaison de deux fragments d'ADN double brin par leurs extrémités. On utilise pour cela une ligase, enzyme capable d'effectuer cette liaison quelle que soit la nature de la coupure (bouts francs ou bouts collants) qui a produit les brins utilisés.

*Marquer-marquage* : technique qui permet de suivre et de localiser une molécule d'ADN par ajout ou remplacement d'un ou de plusieurs nucléotides par des nucléotides modifiés par un radio-isotope artificiel de type  $^{32}\text{P}$ ,  $^{33}\text{P}$ ,  $^{35}\text{S}$  ou une molécule chimiluminescente. Un fragment de grande taille (supérieur à 100-200 nucléotides) est marqué en remplaçant un nucléotide normal par un nucléotide modifié (*nick translation* : translation de coupure, ou *random priming labelling* : marquage par amorçage au hasard). Un fragment de petite taille ou oligonucléotide est marqué par l'ajout d'une base modifiée en position terminale (*end labelling* : marquage terminal).

*Régions flanquantes* : séquences d'ADN situées de part et d'autre d'une séquence cible.

*Séquençage* : permet de connaître la séquence d'un fragment d'ADN.

*Séquence* : nature et succession ordonnée des bases qui composent un brin d'ADN.

*Séquence homologue* : se dit de séquences identiques dans la nature et la succession des bases qui les composent. La notion d'homologie peut être relative et l'on peut fixer arbitrairement des seuils. Des séquences peuvent donc être considérées comme homologues, même si elles ne sont pas totalement identiques, on parle alors de pourcentage d'homologie (voir aussi hybridation).

*Site de restriction* : se dit d'une séquence reconnue par une enzyme de restriction.

*Sonde* : fragment d'ADN marqué dont la séquence et la fonction ne sont pas nécessairement connues (mais peuvent l'être) et qui est utilisé pour repérer des séquences homologues parmi d'autres séquences.

*Southern (blot)* : de T. Southern, inventeur de la technique. Technique de transfert d'ADN sur une membrane (*blot*) suivie d'une hybridation, qui permet ensuite de rechercher à l'aide d'une sonde marquée des fragments d'ADN cibles. Par analogie entre le nom de l'inventeur et les points cardinaux, on a mis au point le *Western blot* pour les protéines et le *Northern blot* pour les ARN. L'Est reste à conquérir.

*Stringence* : qualité d'un milieu d'hybridation ou de réassociation d'acides nucléiques. La stringence dépend principalement des conditions de température, de pH et de concentration saline. Une stringence forte impose que toutes les paires de bases de deux brins d'acides nucléiques à réassocier soient complémentaires. Quand la stringence baisse la réassociation se fait en tolérant un nombre croissant de paires de bases complémentaires (pourcentage d'homologie plus faible).

## Références bibliographiques

- BEER S.C., GOFFREDA J., PHILLIPS T.D., MURPHY J.P., SORRELLS M.E., 1993. Assessment of genetic variation in *Avena sterilis* using morphological traits, isozymes and RFLPs. *Crop Science*, 33 : 1386-1393.
- BENNETT M.D., LEITCH I., 1995. Nuclear DNA amounts in angiosperms. *Annals of Botany*, 76 : 113-176.
- BENNETT M.D., SMITH J.B., 1991. Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London, Series B*, 334 : 309-345.
- CAETANO-ANOLLES G.B., GRESSHOFF P.M., 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotides primers. *Bio/Technology*, 9 : 553-557.
- CHO Y.G., BLAIR M.W., PANAUD O., MCCOUCH S., 1996. Cloning and mapping of variety-specific rice genomic DNA sequences: amplified fragment length polymorphisms (AFLP) from silver-stained polyacrylamide gels. *Genome*, 39 : 373-378.
- COX T.S., KIANG Y.T., GORMAN M.B., RODGERS D.M., 1985. Relationships between coefficient of parentage and genetic similarity indices in the soybean. *Crop Science*, 25 : 529-532.
- VAN DEYNE A.E., SORELLO M.E., PARK W.D., AYRES N.M., FU H., CARTINHO S.W., PAUL E., MCCOUCH S.R., 1998. Anchor probes for comparative mapping of grass genera. *Theoretical and Applied Genetics*, 97 : 356-369.
- DILLMANN C., BAR-HEN A., GUERIN D., CHARCOSSET A., MURIGNEUX A., 1997. Comparison of RFLP and morphological distances between maize, *Zea mays* L., inbred lines: consequences for germplasm protection purposes. *Theoretical and Applied Genetics*, 95 : 92-102.
- DOLEZEL J., 1991. Flow cytometric analysis of nuclear DNA content in higher plants. *Phytochemical Analysis*, 2 : 143-154.
- DOS SANTOS J.B., NIENHUIS J., SKROCH P., TIVANG J., SLOCUM M.K., 1994. Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica oleracea* L. genotypes. *Theoretical and Applied Genetics*, 87 : 909-915.
- DUBREUIL P., CHARCOSSET A., 1998. Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. *Theoretical and Applied Genetics*, 96 : 577-587.
- EFRON B., TIBSHIRANI R., 1986. Bootstrap methods for standart errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1 : 54-77.
- ENGQVIST G.M., BECKER H.C., 1994. Genetic diversity for allozymes, RFLPs and RAPDs in resynthesized rape. In : *Biometrics in plant breeding: applications of molecular markers ; IXth meeting of the EUCARPIA section biometrics in plant breeding*, J. Jansen éd., Wageningen, Pays-Bas, EUCARPIA.
- GERDES J.T., TRACY W.F., 1994. Diversity of historically important sweet corn inbreds as estimated by RFLPs, morphology, isozymes and pedigree. *Crop Science*, 34 : 26-33.
- GRANER A., LUDWIG W.F., MELCHINGER A.E., 1994. Relationships among European barley germplasm. 2. Comparison of RFLP and pedigree data. *Crop Science*, 34 : 1199-1205.
- HALLDEN C., HANSEN M., NILSSON N.O., HJERDIN A., SALL T., 1996. Competition as a source of errors in RAPD analysis. *Theoretical and Applied Genetics*, 93 : 1185-1192.

HENRY A.M., DAMERVAL C., 1997. High rates of polymorphism and recombination at the *Opaque-2* locus in cultivated maize. *Molecular and General Genetics*, 256 : 147-157.

HEUN M., MURPHY J.P., PHILLIPS T.D., 1994. A comparison of RAPD and isozyme analyses for determining the genetic relationships among *Avena sterilis* L. accessions. *Theoretical and Applied Genetics*, 87 : 689-696.

INOMATA N., SHIBATA H., OKUYAMA E., YAMAZAKI T., 1995. Evolutionary relationships and sequence variation of  $\alpha$ -amylase variants encoded by duplicated genes in the *Amy* locus of *Drosophila melanogaster*. *Genetics*, 141 : 237-244.

JARNE P., LAGODA J.L., 1996. Microsatellites, from molecules to populations and back. *Tree*, 11 : 424-429.

JONES C.J., EDWARDS K.J., CASTAGLIONE S., WINFIELD M.O., SALA F., VAN DE WIEL C., BREDEMEIJER G., VOSMAN B., MATTHES M., DALY A., BRETTSCHEIDER R., BETTINI P., BUIATTI M., MAESTRI E., MALCEVSKI A., MARMIROLI N., AERT R., VOLCKAERT G., RUEDA J., LINACERO R., VAZQUEZ A., KARP A., 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, 3 : 381-390.

KARP A., ISAAC P.G., INGRAM D.S., 1998. *Molecular tools for screening biodiversity*. Londres, Royaume-Uni, Chapman and Hall, 498 p.

KARP A., KRESOVICH S., BHAT K.V., AYAD W.G., HODGKIN T., 1997. *Molecular tools in plant genetic resources conservation: a guide to the technologies*. Rome, Italie, IPGRI, Technical Bulletin n° 2, 47 p.

KARP A., SEBERG O., BUIATTI M., 1996. Molecular techniques in the assessment of botanical diversity. *Annals of Botany*, 78 : 143-149.

KREMER A., 1998. Les marqueurs moléculaires en génétique des populations. *In* : Les marqueurs moléculaires en génétique et biotechnologies végétales, D. de Vienne éd., Paris, France, Inra, 200 p.

LANGRIDGE U., SCHWALL M., LANGRIDGE P., 1991. Squashes of plant tissue as substrate for PCR. *Nucleic Acids Research*, 19 : 6954.

LIU Z., FURNIER G.R., 1993. Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. *Theoretical and Applied Genetics*, 87 : 97-105.

LU J., KNOX M.R., AMBROSE M.J., BROWN J.K.M., ELLIS T.H.N., 1996. Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. *Theoretical and Applied Genetics*, 93 : 1103-1111.

LYNCH M., MILLIGAN B.G., 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology*, 3 : 91-99.

MESSMER M.M., MELCHINGER A.E., LEE M., WOODMAN W.L., LEE E.A., LAMKEY K.R., 1991. Genetic diversity among progenitors and elite lines from the Iowa stiff stalk synthetic (BSSS) maize population: comparison of allozyme and RFLP data. *Theoretical and Applied Genetics*, 83 : 97-107.

MILLER R.G., 1974. The jackknife: a review. *Biometrika*, 61 : 1-15.

- MITCHELL M.J., QUIROS C., 1992. Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. *Theoretical and Applied Genetics*, 83 : 783-790.
- MOHAN M., NAIR S., BHAGWAT A., KRISHNA T.G., YANO M., BHATIA C.R., SASAKI T., 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding*, 3 : 87-103.
- MORGANTE M., RAFALSKI A., BIDDLE P., TINGEY S., OLIVIERI A.M., 1994. Genetic mapping and variability of seven soybean simple sequence repeat loci. *Genome*, 37 : 763-769.
- N'GORAN J.A.K., LAURENT V., RISTERUCCI A.M., LANAUD C., 1994. Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity*, 73 : 589-597.
- O'DONOUGHUE L.S., SOUZA E., TANKSLEY S.D., SORRELLS M.E., 1994. Relationships among North American oat cultivars based on restriction fragment length polymorphisms. *Crop Science*, 34 : 1251-1258.
- OLUFOWOTE J.O., XU Y., CHEN X., PARK W.D., BEACHELL H.M., DILDAY R.H., GOTO M., MCCOUCH S.R., 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. *Genome*, 40 : 370-378.
- ORTI G., PEARSE D.E., AVISE J.C., 1997. Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Sciences of the United States of America*, 94 : 10745-10749.
- PEAKALL R., SMOUSE P.E., HUFF D.R., 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. *Molecular Ecology*, 4 : 135-147.
- PERRIER X., FLORI A., BONNOT F., 1999. Les méthodes d'analyse des données. In : Diversité génétique des plantes tropicales cultivées, P. HAMON *et al.* éd., Montpellier, France, Cirad, collection Repères, p. 43-76.
- PLASCHKE J., GANAL M.W., RÖDER M.S., 1995. Detection of genetic diversity in closely related bread wheat using microsatellites markers. *Theoretical and Applied Genetics*, 91 : 1001-1007.
- POWELL W., MORGANTE M., ANDRE C., HANAFEY M., VOGEL J., TINGEY S., RAFALSKI A., 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, 2 : 225-238.
- PRABHU R.R., WEBB D., JESSEN H., LUK S., SMITH S., GRESSHOFF P.M., 1997. Genetic relatedness among soybean genotypes using DNA amplification fingerprinting (DAF), RFLP, and pedigree. *Crop Science*, 37 : 1590-1595.
- RAGOT M., HOISINGTON D.A., 1993. Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. *Theoretical and Applied Genetics*, 86 : 975-984.
- RISTERUCCI A.M., 1997. Coût des techniques de marquage moléculaire. In : Utilisation des marqueurs moléculaires en génétique et en amélioration des plantes ; séminaire de formation. Montpellier, France, Cirad-Biotrop.
- RUS-KORTEKAAS W., SMULDERS M.J.M., ARENS P., VOSMAN B., 1994. Direct comparison of levels of genetic variation in tomato detected by a gaca-containing microsatellite probe and by random amplified polymorphic DNA. *Genome*, 37 : 375-381.

- RUSSELL J.R., FULLER J.D., MACAULAY M., HATZ B.G., JAHOR A., POWELL W., WAUGH R., 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theoretical and Applied Genetics*, 95 : 714-722.
- SAGHAI-MAROOF M.A., BIYASHEV R.M., YANG G.P., ZHANG Q., ALLARD R.W., 1994. Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 91 : 5466-5470.
- SALIMATH S.S., DE OLIVEIRA O.A.C., GODWIN I.D., BENNETZEN J.L., 1995. Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. *Genome*, 38 : 757-763.
- SANTONI S., 1996. Les marqueurs moléculaires utilisables en amélioration des plantes. *Le Sélectionneur français*, 46 : 3-18.
- TARAMINO G., TINGEY S., 1996. Simple sequence repeats for germplasm analysis and mapping in maize. *Genome*, 39 : 277-287.
- THORMANN C.E., FERREIRA M.E., CAMARGO L.E.A., TIVANG J.G., OSBORN T.C., 1994. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. *Theoretical and Applied Genetics*, 88 : 973-980.
- TINKER N.A., FORTIN M.G., MATHER D.E., 1993. Random amplified polymorphic DNA and pedigree relationships in spring barley. *Theoretical and Applied Genetics*, 85 : 976-984.
- DE VIENNE D., SANTONI S., 1998. Les principales sources de marqueurs moléculaires. *In* : Les marqueurs moléculaires en génétique et biotechnologies végétales, D. de Vienne éd., Paris, France, Inra, 200 p.
- VOS P., HOGERS R., BLEEKER M., REIJANS M., VAN DE LEE T., HORNES M., FRIJTERS A., POT J., PELEMAN J., KUIPER M., ZABEAU M., 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23 : 4407-4414.
- WEIR B.S., 1990. *Genetics data analysis, methods for discrete population genetic data*. Sunderland, Etats-Unis, Sinaur Associates.
- WEISING K., NYBOM H., WOLFF K., MEYER W., 1995. *DNA fingerprinting in plants and fungi*. Londres, Royaume-Uni, CRC Press.
- WILLIAMS J.G.K., KUBELIK A.R., LIVAK K.J., RAFALSKI J.A., TINGEY S.V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18 : 6531-6535.
- WU K.S., TANKSLEY S.D., 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular and General Genetics*, 241 : 225-235.
- YANG W., DE OLIVEIRA A.C., GODWIN I., SCHERTZ K., BENNETZEN J.L., 1996. Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. *Crop Science*, 36 : 1669-1676.
- ZHANG Q., SAGHAI-MAROOF M.A., KLEINHOF A., 1993. Comparative diversity analysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* subsp. *spontaneum*. *Genetics*, 134 : 909-916.
- ZIETKIEWICZ E., RAFALSKI A., LABUDA D., 1994. Genome fingerprinting by simple-sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics*, 20 : 176-183.





# Les méthodes d'analyse des données

---

Xavier Perrier, Albert Flori, François Bonnot

L'analyse des données de diversité revient, le plus souvent, à l'analyse de tableaux croisant des unités taxonomiques — espèces, populations, cultivars... — et des variables de caractérisation de la diversité. Ces variables sont de nature diverse : caractères agronomiques quantitatifs, descripteurs morphologiques, le plus souvent qualitatifs, marqueurs biochimiques (isoenzymes) ou moléculaires (RFLP, RAPD...), souvent codés en présence ou absence. Les unités taxonomiques sont en général observées individuellement, mais elles peuvent aussi correspondre à des populations, les observations sont alors des moyennes de caractères, des fréquences alléliques... Dans tous les cas, l'analyse de ces données a pour objectif de repérer une éventuelle structuration des unités taxonomiques par l'analyse des ressemblances ou des dissemblances entre ces unités.

Cet objectif relève d'une action de classification, terme vague qui reçoit dans le langage courant deux acceptions différentes. La première concerne l'action de regrouper les objets en ensembles homogènes. Cette ambition a été celle de tous les grands systématiciens, d'Aristote à Linné. La seconde acception relève de l'action d'ordonner logiquement. On s'intéresse alors essentiellement aux relations d'ordre entre les unités taxonomiques, c'est-à-dire aux relations de filiation entre unités au sens de l'évolution, qui relèvent de la phylogénétique.

Ces deux démarches peuvent aboutir à des représentations graphiques sous forme d'arbre. Leur contenu informatif est cependant fondamentalement différent. Dans le premier cas, les nœuds de l'arbre correspondent à des concepts de regroupement. Dans le second, les nœuds sont des unités parentales non observées mais supposées existantes, par exemple des formes ancestrales disparues. Les liens incluent alors la dimension temporelle de l'évolution.

Le problème de la diversité traité ici relève en partie des deux démarches puisque l'on souhaite définir des groupes d'unités comparables, mais aussi décrire les relations de parenté entre ces groupes.

Deux approches peuvent être suivies pour traiter de l'organisation évolutive. La première, dite phénétiqque, décrit une organisation à partir de mesures objectives de dissemblance entre les unités et reporte à l'interprétation l'introduction d'hypothèses génétiques. Ces dissimilarités sont estimées globalement sur l'ensemble des caractères, c'est la multiplicité des caractères observés qui est censée donner une pertinence à ces mesures. Il est, le plus souvent, extrêmement difficile de détecter une organisation des unités par l'examen direct des dissimilarités. L'objet de l'analyse est alors de chercher un espace de représentation approchée mais facilement lisible de ces dissimilarités : plan factoriel ou structure arborée. Le gain de lisibilité se paie au prix d'une certaine perte d'information, perte que l'on cherchera bien sûr à minimiser.

La seconde approche, dite cladistique, s'intéresse directement aux caractères observés et tente de distinguer les caractères identiques hérités d'un ancêtre commun, qui sont seuls informatifs, des phénomènes de convergence accidentelle. Elle repose sur un modèle de transformation génétique et a été établie initialement sur des caractères morphologiques que l'on savait monogéniques et dont on savait ordonner les différentes modalités en termes d'évolution. On l'applique maintenant le plus souvent à des données de séquençage du génome munies d'un modèle génétique explicite, du type de celui de JUKES et CANTOR (1969) ou dérivé.

Dans le cadre de cet ouvrage, les méthodes cladistiques ou apparentées, comme les méthodes de compatibilité ou de vraisemblance, ne seront pas évoquées, on se référera, par exemple, à DARLU et TASSY (1993). En effet, la nature des marqueurs utilisés ne permet le plus souvent que des approches phénétiqques relevant de l'analyse des dissimilarités.

L'objet de ce chapitre est de présenter divers aspects méthodologiques relevant de ces approches. Les données de base seront constituées d'un tableau d'unités taxonomiques décrites par des marqueurs de la diversité. Il pourra aussi s'agir d'un empilement de tableaux lorsqu'un même ensemble d'unités est caractérisé par plusieurs séries de marqueurs de nature différente que l'on ne peut regrouper. Ces tableaux pourront être considérés séparément, mais on cherchera finalement à les analyser simultanément.

Ce chapitre est composé de trois grandes parties.

La première concerne l'étape initiale de toute analyse qui est la définition d'une mesure de la ressemblance ou de la dissemblance entre individus. Le choix de cette mesure est de toute première importance et répond au choix délibéré d'un point de vue sur les données. Lorsque les individus sont des populations et les observations des fréquences alléliques, diverses mesures de dissimilarité, dites génétiques, sont utilisables — distances de Nei, de Gregorius... Leurs propriétés ont été souvent étudiées, on se référera, par exemple, à LEFORT-BUSSON et DE VIENNE (1985). Quand les données sont des individus, décrits par une série de caractères, la nature des variables — quantitatives, qualitatives, binaires — conditionne les types de dissimilarité utilisables. On montrera, sur l'exemple des marqueurs moléculaires de type RFLP ou RAPD, que la connaissance de caractéristiques biologiques des marqueurs permet de justifier le choix d'un indice de dissimilarité.

La deuxième partie traite des méthodes factorielles. Une matrice de dissimilarité entre individus deux à deux étant établie, son analyse a pour objet d'en rechercher une représentation simplifiée mais fidèle. Si la dissimilarité est de nature euclidienne, les méthodes d'analyse factorielle peuvent être utilisées. Les bases des principales méthodes seront rappelées, on s'intéressera en particulier aux techniques d'analyse simultanée de plusieurs tableaux. Si la dissimilarité n'est pas euclidienne, les méthodes de positionnement multidimensionnel (*multidimensional scaling*) permettent de rechercher une décomposition sur un nombre fixé d'axes. Ces méthodes itératives demandent des temps de calcul important lorsque le nombre d'unités à traiter est élevé. Ces méthodes ne seront pas abordées dans le cadre de cet ouvrage, on pourra se référer, par exemple, à ESCOUFIER (1975).

La troisième partie s'intéresse aux représentations arborées. Un processus évolutif, par accumulation de mutations transmissibles, produirait une organisation représentable sous la forme d'un arbre. Cette représentation arborée, par son fondement biologique, est particulièrement pertinente pour l'analyse de la diversité, même si les modalités de l'évolution sont en général plus complexes. Diverses méthodes de création d'arbres seront présentées, ainsi que les techniques qui permettent de construire des arbres synthétiques à partir de plusieurs jeux de marqueurs.

Les méthodes factorielles et les méthodes de représentation arborée constituent deux approches très différentes de la représentation de la diversité. Elles doivent être considérées comme complémentaires plutôt que comme concurrentes. Les méthodes factorielles recherchent avant tout une représentation globale de la diversité, débarrassée autant que possible des effets individuels. A l'inverse, les méthodes arborées tendent, pour les plus utilisées, à représenter fidèlement les relations individuelles. Il s'agit donc bien de deux points de vue différents sur les données.

De nombreux logiciels statistiques proposent les diverses méthodes d'analyse factorielle. Les outils liés à l'analyse des dissimilarités et à leur représentation arborée sont moins fréquents, le logiciel Ntsys est sans doute le plus complet et le plus accessible (ROHLF, 1987). Diverses fonctions non disponibles ou des méthodes en partie remaniées ont été réunies dans un logiciel spécialisé, Darwin (*dissimilarity analysis and representation*), disponible auprès des auteurs (PERRIER, 1998).

## Le choix d'un indice de dissimilarité

Il s'agit de définir une mesure de la ressemblance (similarité) entre deux individus ou plutôt, de façon équivalente, de la dissemblance (dissimilarité) puisque l'on peut passer simplement de l'une à l'autre par une transformation linéaire.

Il existe, en fait, un grand nombre de mesures de dissemblance. En fonction de la nature des données, divers choix sont possibles. Certaines de ces dissimilarités ont des propriétés mathématiques intéressantes pour l'analyse ou possèdent des capacités de stabilité face à des erreurs de données, qui constituent des arguments pertinents pour les retenir. Le choix final reste relatif au réel contenu informatif des marqueurs utilisés et au point de vue que l'on souhaite adopter sur cette information.

Enfin, certaines dissimilarités, les ultramétriques et les distances additives, ont des propriétés remarquables de représentabilité sous forme d'arbre. Il n'y a, bien sûr, aucune raison pour que la mesure établie sur les données possède de telles propriétés. L'objet des méthodes de construction d'arbre est de transformer, en la déformant le moins possible, la mesure établie sur les données en une dissimilarité possédant des propriétés de représentabilité.

### Définitions et propriétés

L'objet mathématique le plus général traduisant la différence est une dissimilarité, c'est une fonction  $d$  de l'ensemble des paires  $(i, j)$  d'individus dans l'ensemble des réels positifs ou nuls, symétrique ( $d(i, j) = d(j, i)$ ) et telle que  $d(i, i) = 0$  pour tout  $i$ . Cette définition est peu contraignante et couvre un grand nombre de mesures possibles. En revanche, elle n'ouvre que peu de propriétés mathématiques et il est nécessaire d'ajouter d'autres contraintes pour acquérir certaines propriétés intéressantes.

Une distance, on dit parfois une métrique, est une dissimilarité particulière obtenue en ajoutant la condition  $d(i, j) = 0 \Rightarrow i = j$  et surtout l'inégalité triangulaire entre trois individus,  $d(i, j) \leq d(i, k) + d(j, k)$ . Cette condition, naturelle, tra-

duit simplement la possibilité de représenter tout triplet de points par un triangle dans le plan. Cette propriété, très intéressante, permet, en particulier, d'éviter le problème de longueurs d'arêtes négatives dans les méthodes de construction d'arbre.

#### DISTANCE DE MANHATTAN ET DISTANCE EUCLIDIENNE

Parmi les distances, une famille importante est constituée des distances dites de Minkowski d'ordre  $p$  ( $p \geq 1$ ). Une distance appartient à cette famille s'il existe un entier  $K$  et une série de  $K$  valeurs  $x_{ik}$  affectables à chaque individu  $i$ , la distance s'écrivant alors  $d(i, j) = (\sum_k |x_{ik} - x_{jk}|^p)^{1/p}$ .

Les seuls cas utilisés en pratique correspondent aux valeurs 1 et 2 de  $p$ . Quand  $p = 1$ , l'indice est connu sous le nom de distance de Manhattan ou *city-block*,  $d(i, j) = \sum_k |x_{ik} - x_{jk}|$ . Pour  $p = 2$ , on retrouve l'habituelle distance euclidienne,  $d(i, j) = [\sum_k (x_{ik} - x_{jk})^2]^{1/2}$ . Bien sûr, des distances de Manhattan ou euclidiennes sont obtenues, par construction même, lorsque, à partir d'un tableau individus  $\times$  variables, on calcule une distance en sommant, sur les variables, les valeurs absolues ou les carrés des écarts. Cependant, certains indices, par exemple ceux qui sont calculés sur des données de présence ou d'absence, peuvent être de l'un de ces deux types sans que cela découle de manière évidente du mode de construction.

Distance de Manhattan et distance euclidienne appartiennent à la même famille et sont liées par certaines relations. On sait montrer, par exemple, que toute distance euclidienne est une distance de Manhattan. On montre aussi que la racine carrée d'une distance de Manhattan est une distance euclidienne.

#### TRANSFORMATION PUISSANCE

Les transformations puissance d'une dissimilarité permettent, pour certaines puissances, de lui donner les propriétés d'une distance ou même d'une distance euclidienne. En préalable, il faut rappeler que deux dissimilarités  $d$  et  $d'$  sont équivalentes en ordre si  $d(i, j) \leq d(k, l) \Leftrightarrow d'(i, j) \leq d'(k, l)$ . Ces deux dissimilarités ordonnent les paires d'individus de la même façon et sont donc d'interprétation comparable. En particulier, si  $d'$  peut s'écrire comme une fonction monotone croissante de  $d$  alors  $d'$  et  $d$  sont équivalentes en ordre. C'est le cas des fonctions puissance  $\alpha$  ( $\alpha \geq 0$ ),  $d^\alpha$  et  $d$  sont équivalentes en ordre.

On sait montrer qu'il est toujours possible de trouver une valeur  $\alpha$  comprise entre 0 et 1 telle que, si  $d$  est une dissimilarité, alors pour tout  $\lambda$  ( $0 \leq \lambda \leq \alpha$ ),  $d^\lambda$  est une distance qui, par la propriété précédente, est équivalente en ordre avec  $d$ . Donc, si le contenu informatif des marqueurs amène à choisir une mesure qui n'est qu'une dissimilarité, on sait par une transformation puissance lui conférer les propriétés d'une distance dont on a signalé l'intérêt pour les méthodes de classification.

De même, on démontre que si  $d$  est une distance, on peut toujours trouver une valeur  $\alpha$  comprise entre 0 et 1 telle que, pour tout  $\lambda$  ( $0 \leq \lambda \leq \alpha$ ),  $d^\lambda$  est une distance euclidienne équivalente en ordre avec  $d$ . Une transformation puissance permet donc de passer d'une distance à une distance euclidienne, comme on passe d'une dissimilarité à une distance. Une application intéressante concerne les méthodes d'analyse factorielle. Bien qu'elles ne traitent que des distances euclidiennes, on pourra cependant utiliser ces méthodes à partir d'une dissimilarité quelconque après transformation puissance préalable. Cette transformation paraît préférable à la technique habituelle d'ajout d'une constante, éventuellement très grande, à toutes les valeurs des distances, technique moins respectueuse des données initiales.

Dans certains cas, on connaît les valeurs  $\alpha$  de ces deux transformations, par exemple, il a déjà été signalé que la racine carrée ( $\alpha = 1/2$ ) d'une distance de Manhattan est une distance euclidienne. Souvent on ne sait pas établir cette valeur qui doit être estimée numériquement sur les données.

#### ULTRAMÉTRIQUE ET DISTANCE ADDITIVE D'ARBRE

En ajoutant des contraintes supplémentaires à la définition d'une dissimilarité, on peut lui conférer la propriété d'être une distance représentable sous forme d'un arbre. Les méthodes de classification ont toutes pour objectif d'approcher au mieux, au sens d'un certain critère, la dissimilarité observée par une de ces distances représentables.

La plus connue de ces dissimilarités est l'ultramétrique, c'est une distance vérifiant l'inégalité  $d(i,j) \leq \max[d(i,k), d(j,k)]$ , pour trois individus quelconques  $i, j$  et  $k$ . Cette propriété exprime que, parmi les trois distances, les deux plus grandes sont égales : si  $d(i,j)$  est la plus petite, alors  $d(i,j) \leq d(i,k) = d(j,k)$ . Tout triplet de points forme donc un triangle isocèle « pointu ». Cette propriété permet une représentation sous la forme habituelle d'un dendrogramme (figure 1a), qui est un arbre possédant un point particulier, la racine, située à égale distance de toutes les feuilles de l'arbre. On montre qu'une ultramétrique est une distance euclidienne.

Dans les années 70, la distance dite additive d'arbre a été proposée. Elle vérifie l'inégalité  $d(i,j) + d(k,l) \leq \max[d(i,k) + d(j,l), d(i,l) + d(j,k)]$  pour quatre individus quelconques  $i, j, k$  et  $l$ . Cette propriété, dite des quatre points, est une extension à quatre points de la condition ultramétrique. Elle exprime que, parmi les trois sommes des distances deux à deux, les deux plus grandes sont égales. Cette propriété permet une représentation sous forme d'un arbre. On vérifie sur un exemple à quatre points (figure 2) que la condition est effectivement respectée puisque  $d(i,j) + d(k,l) = a_i + a_j + a_k + a_l$  et  $d(i,k) + d(j,l) = d(i,l) + d(j,k) = a_i + a_j + a_k + a_l + 2a_c$ . Cet arbre peut être représenté sous forme hiérarchique comme un dendrogramme (figure 1b). Cependant, en l'absence de racine objective, on lui préfère souvent une représentation radiale (figure 1c).

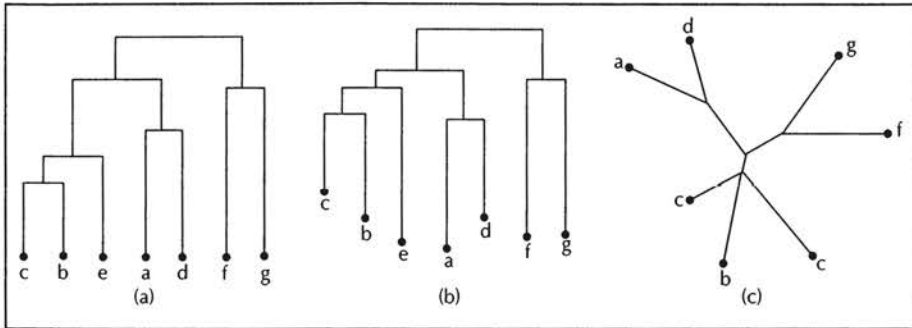


Figure 1. Ultramétrie (a) et distance additive en représentation hiérarchique (b) ou radiale (c).

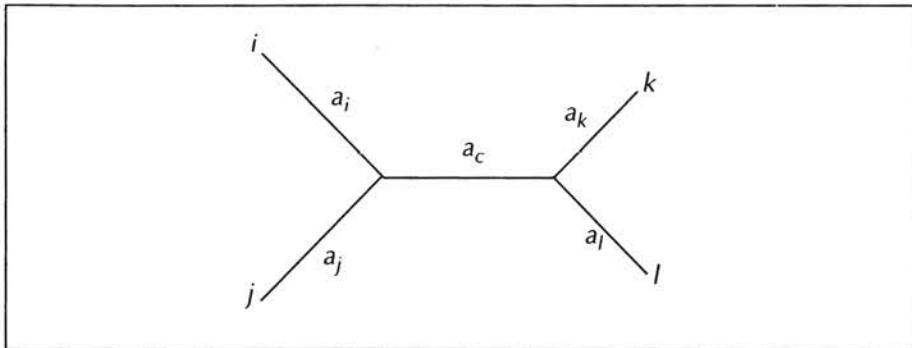


Figure 2. Arbre à quatre feuilles et longueurs des arêtes.

On montre que la condition ultramétrique n'est qu'un cas particulier de la condition des quatre points. L'ultramétrie apparaît alors comme une distance additive d'arbre particulière soumise à une contrainte plus forte. Cette contrainte supplémentaire correspond à l'obligation, sur trois individus, de l'égalité des deux plus grandes distances, contrainte absente d'une distance additive d'arbre, qui autorise des longueurs de branches inégales. Cette moindre contrainte permet donc une représentation plus fidèle des dissimilarités initiales.

#### RELATION D'INCLUSION ENTRE FAMILLES DE DISSIMILARITÉS

Ces différentes familles de dissimilarités ont été établies en ajoutant des contraintes supplémentaires à partir de la définition initiale d'une dissimilarité. Cela induit des relations d'inclusion entre ces différentes familles (figure 3). Ainsi, les ultramétriques sont des distances euclidiennes et des distances additives particulières. Elles sont elles-mêmes des distances de Manhattan, qui, à leur tour, sont des distances, cas particulier de dissimilarités. Quand on parcourt la figure 3 de haut en bas, on rencontre des indices supportant des conditions de plus en plus fortes et donc de moins en moins aptes à décrire



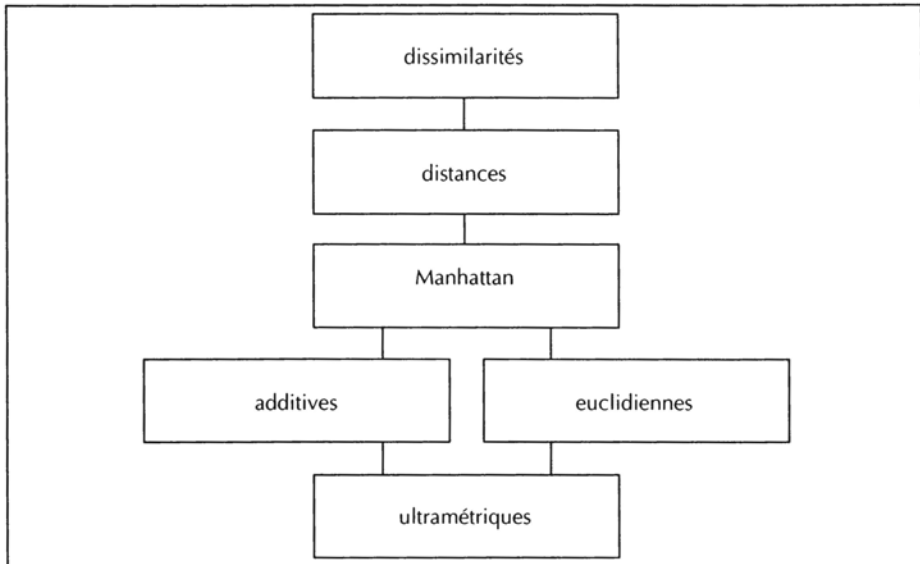


Figure 3. Relations entre familles de dissimilarités, d'après Critchley et Fichet (1994).

fidèlement les relations entre individus. Inversement, du bas vers le haut, les indices perdent leurs propriétés de représentabilité : les ultramétriques et les distances additives sont représentables sous forme d'arbre dans le plan, les distances euclidiennes et les distances de Manhattan sont représentables dans des espaces de dimension élevée. A l'opposé, distances et dissimilarités n'ont aucune propriété intéressante en la matière. Il apparaît alors que la distance de Manhattan est, parmi les distances représentables, celle qui est la moins contrainte. En l'absence d'indication contraire, le choix devrait logiquement se porter sur un tel indice.

Dans la pratique, on mène souvent en parallèle une analyse factorielle sur distance euclidienne et une représentation arborée. On note sur le graphique le niveau équivalent de la distance euclidienne et de la distance additive. La cohérence voudrait donc que la représentation arborée soit fondée sur une distance additive plutôt que sur une ultramétrique comme c'est souvent le cas.

### Mesures de dissimilarités sur variables quantitatives

La dissimilarité la plus souvent utilisée sur des données quantitatives est la distance euclidienne  $d(i, j) = [\sum_k (x_{ik} - x_{jk})^2]^{1/2}$  ( $x_{ik}$  : valeur de la  $k^e$  variable pour l'individu  $i$ ). Elle est très classique en statistique et peut être traitée par les méthodes factorielles. Si les variables sont de nature différente, il est souvent utile de les réduire et, pour rendre la valeur de la distance indépendante du nombre  $P$  de variables, on la pondère souvent par  $1/P$ . D'où l'expression générale  $d^p(i, j) = (\sum_k [(x_{ik} - x_{jk}) / \sigma_k]^2) / P$ .

L'intérêt mathématique de la distance de Manhattan,  $d(i,j) = \sum_k |x_{ik} - x_{jk}|$ , a été souligné. On observe que la distance euclidienne, en élevant au carré les différences, donne un poids fort aux écarts importants. Cet *a priori* n'a pas toujours de réelle justification, la distance de Manhattan, qui donne le même poids à toutes les différences, paraît alors préférable. On pondère en général cette distance par le nombre  $P$  de variables et, si l'ordre de grandeur des variables est très différent, il est nécessaire de les ramener à une même échelle, par exemple en les rapportant à leur amplitude.

## Les mesures de dissimilarités sur variables présence ou absence

Les descripteurs de la diversité génétique sont souvent des variables binaires, qui traduisent l'absence ou la présence d'un caractère. Pour les descripteurs biochimiques ou moléculaires, ces caractères codent en général la présence ou l'absence d'une bande sur un gel d'électrophorèse. Nous noterons conventionnellement 0 l'absence et 1 la présence.

De nombreuses mesures de ressemblance entre individus, ou coefficients d'association (SNEATH et SOKAL, 1973), ont été définies sur les données binaires. Elles ont le plus souvent été proposées par des chercheurs d'une discipline particulière — botanique, zoologie, paléontologie... — et avaient comme principale justification de bien traduire l'idée que se faisaient ces chercheurs de la ressemblance entre leurs objets d'étude. Nous nous limiterons ici à la présentation des indices dont on sait expliciter la logique de construction et qui présentent un comportement raisonnable (BEAULIEU, 1989) ; par exemple, si on ajoute un nouveau marqueur, la dissimilarité entre deux individus doit augmenter (ou diminuer) si les deux individus prennent, pour ce marqueur, des valeurs différentes (identiques).

On note, pour deux individus  $i$  et  $j$ ,  $a$  le nombre de marqueurs qui sont présents simultanément chez  $i$  et  $j$ . De même,  $d$  est le nombre d'absences en commun,  $b$  le nombre de présences chez  $i$  et d'absences chez  $j$  et  $c$  la situation inverse. Le tableau 1 donne la liste de treize de ces indices. Il est habituel de les exprimer sous la forme de la similarité  $S$ , la dissimilarité s'obtenant, pour ces indices, par  $D = 1 - S$ .

Le principe général de construction de tous ces indices est le même, la similitude est estimée par le nombre d'accords. Cependant, cette valeur n'a pas de sens absolu et doit être rapportée à une base de comparaison. Les indices divergent dans le mode d'estimation du nombre d'accords et dans le choix de la base de comparaison.

L'estimation des accords dépend du sens que l'on donne à la modalité absence. Si l'on considère que seule la modalité 1 est informative, la modalité 0 traduisant surtout une absence d'information, alors le nombre d'accords est  $a$ , le nombre de présences en commun (indices S1 à S9). Si 0 et 1 sont informatifs et

Tableau 1. Principaux indices de similarité sur variables présence ou absence : auteur, explication et propriétés ( $P = a + b + c + d$ ).

	Auteurs	Expression	Propriétés*			
			(1)	(2)	(3)	(4)
S1	Russel et Rao	$a / P$	o	o	o	
S2	Simpson	$a / \min[(a + b), (a + c)]$	n		n	
S3	Braun-Blanquet	$a / \max[(a + b), (a + c)]$				
S4	Dice	$a / [a + (b + c) / 2]$	n		n	S7, S8
S5	Ochiai	$a / [(a + b)(a + c)]^{1/2}$	n		o	
S6	Kulczynsky 1	$(a / 2)([1 / (a + b)] + [1 / (a + c)])$	n		n	
S7	Jaccard	$a / (a + b + c)$	o	o	o	S4, S8
S8	Sokal et Sneath un2	$a / [a + 2(b + c)]$	o	o	o	S4, S7
S9	Kulczynski 2	$a / (b + c)$				
S10	Sokal et Michener	$(a + d) / P$	o	o	o	S11, S
S11	Rogers et Tanimoto	$(a + d) / [a + d + 2(b + c)]$	o	o	o	S11, S
S12	Sokal et Sneath un1	$(a + d) / [a + d + (b + c) / 2]$	n		n	S10, S
S13	Sokal et Sneath un3	$(a + d) / (b + c)$				

\* (1) La dissimilarité associée est (o) ou n'est pas (n) une distance, (2) elle est (o) ou n'est pas (n) une distance de Manhattan, (3) sa racine carrée est euclidienne (o) ou ne l'est pas (n), (4) elle est équivalente en ordre avec les indices indiqués (l'absence de mention indique que l'on ne connaît pas aujourd'hui la réponse).

peuvent être considérés comme deux modalités d'une variable qualitative, alors le nombre d'accords est  $a + d$ , le nombre de présences et d'absences en commun (indices S10 à S13). Le choix entre ces deux attitudes dépend entièrement de la nature des caractères analysés et doit être préalable à toute réflexion sur le choix d'un indice de dissimilarité. Cependant, il n'est pas toujours facile de trancher entre ces deux points de vue. Concernant les marqueurs génétiques de la diversité, on verra que la connaissance biologique des marqueurs considérés permet d'orienter le choix vers un de ces modèles.

#### ACCORDS ESTIMÉS PAR LES PRÉSENCES EN COMMUN : INDICES S1 À S9

Le numérateur de ces indices est toujours  $a$ . Le dénominateur veut être une estimation du nombre d'accords que présenteraient deux individus identiques. L'indice S1 considère que ce nombre est le nombre de variables  $P$ . Ce choix n'est pas judicieux puisque deux individus ne peuvent alors être identiques que s'ils ont la valeur 1 pour toutes les variables. On préférera estimer le dénominateur à partir des nombres de présences chez  $i$ ,  $a + b$ , et chez  $j$ ,  $a + c$ . Bien sûr, ces deux valeurs ne sont pas égales, et il faut en définir une expression consensuelle. On peut en prendre le minimum (S2), le maximum (S3), ou une série de valeurs entre ces deux extrêmes : la moyenne arithmétique (S4), la moyenne géométrique (S5), la moyenne harmonique (S6). La connaissance des caractères peut parfois guider le choix d'un de ces indices. Cependant, en l'absence de justification claire, on retiendra de préférence le comportement neutre de l'indice de Dice (S4) fondé sur la moyenne arithmétique.

Une autre approche est de considérer que la base de comparaison est le nombre de présences rencontrées chez  $i$  ou chez  $j$ ,  $a + b + c$ , d'où l'indice de Jaccard (S7). Cela revient, comme pour S1, à comparer  $a$  au nombre de variables, mais en traitant les doubles absences comme des données manquantes. Ce point de vue est, dans bien des cas, très raisonnable et explique le succès de cet indice qui est certainement le plus utilisé.

L'indice S8, assez souvent rencontré, et l'indice S9 paraissent difficiles à justifier. On remarque que les indices S4, S7 et S8 peuvent être écrits sous la forme  $S = a / [a + \beta(b + c)]$  avec, respectivement,  $\beta = 1/2$ ,  $\beta = 1$  et  $\beta = 2$ . On pourrait construire de nouveaux indices en choisissant d'autres valeurs de  $\beta$ .

#### ACCORDS ESTIMÉS PAR LES PRÉSENCES ET LES ABSENCES EN COMMUN : INDICES S10 À S13

Ces indices sont obtenus par extension des indices S1, S8, S4, S9 en remplaçant simplement  $a$  par  $a + d$ . À l'évidence, ces indices doivent être symétriques en  $a$  et  $d$ , la notation 0 et 1 étant purement arbitraire. L'extension des autres indices conduirait à des indices non symétriques en  $a$  et  $d$ , ou parfois non symétriques en  $b$  et  $c$ , et donc inacceptables. L'indice S10, extension de S1, mais aussi de S7, l'indice de Jaccard, et S12, extension de S4 fondé sur la moyenne arithmétique, paraissent les mieux appropriés dans beaucoup de cas. Les indices S12, S10 et S11 peuvent être écrits sous la forme  $S = (a + d) / [(a + d) + \beta(b + c)]$  avec, respectivement,  $\beta = 1/2$ ,  $\beta = 1$  et  $\beta = 2$ .

#### PROPRIÉTÉS DE CES INDICES

En complément de la logique de construction, la possession de propriétés particulières peut guider le choix d'un indice. Le tableau 1 rappelle certaines propriétés des dissimilarités associées aux indices proposés. Une première caractéristique importante est le fait d'être une vraie distance. On montre aussi que certains de ces indices sont des distances de Manhattan. Si aucun de ces indices n'est une distance euclidienne, on sait que les distances de Manhattan ont des racines carrées euclidiennes. L'indice S5, sans être une distance de Manhattan, a lui aussi une racine carrée euclidienne. Ces indices, après transformation, peuvent ainsi être traités par des méthodes factorielles.

Enfin, plusieurs de ces indices, qui ont des modes de construction comparables, sont équivalents en ordre. Entre plusieurs indices donnant des ordres comparables, on préférera les indices possédant des propriétés intéressantes, S10 ou S11, plutôt que S12 par exemple.

Les algorithmes de construction d'arbres montrent une forte sensibilité à de petites variations des dissimilarités. Le comportement de ces différents indices face à des perturbations des données constitue donc un critère de choix important. Ce comportement peut être abordé en étudiant les propriétés d'une matrice soumise à une perturbation aléatoire transformant, avec une proba-

bilité  $t$ , les notations 1 en 0 et inversement. Il est alors possible d'estimer l'espérance  $D'$  de l'indice considéré pour une perturbation de taux  $t$  des données.

La figure 4 présente, pour les indices S10, S11 et S12, symétriques en 0 et 1, la valeur de  $D'$  en fonction de  $D$ . Le taux d'erreur utilisé est de 0,10, taux que l'on peut rencontrer dans certaines applications, sa valeur ne change d'ailleurs pas le sens des conclusions que l'on peut énoncer. Les erreurs d'observations peuvent entraîner pour l'indice S11 de fortes surestimations des petites valeurs de dissimilarité, alors que les grandes valeurs sont peu modifiées. L'indice S12 a un comportement inverse en laissant les petites valeurs peu modifiées et en sous-estimant les grandes valeurs. L'indice S10 a un comportement plus neutre la distorsion est nulle pour  $D = 0,5$  et augmente symétriquement de part et d'autre. Il constitue donc un compromis intéressant. Cependant, les algorithmes de construction d'arbre les plus utilisés procèdent par agrégation ascendante, les plus petites dissimilarités déterminent les premiers regroupements qui conditionneront l'ensemble de l'arbre. Il serait alors préférable de minimiser les perturbations sur les faibles valeurs comme le fait l'indice S12.

L'étude des indices estimant les accords par les seules présences est plus compliquée. Les distorsions observées sont plus fortes que pour les indices symétriques. Elles peuvent être très élevées (jusqu'à 50 % pour les faibles valeurs de  $d$ , même avec des taux d'erreur de 0,05) lorsque les fréquences de 0 sont élevées (figure 5). Cette situation est fréquente avec les marqueurs moléculaires très polymorphes, pour lesquels le nombre de 1 pour chaque individu est très faible. Cela doit conduire à une certaine prudence face à ce type de matrice, il serait alors nécessaire de vérifier la stabilité des structures en comparant les

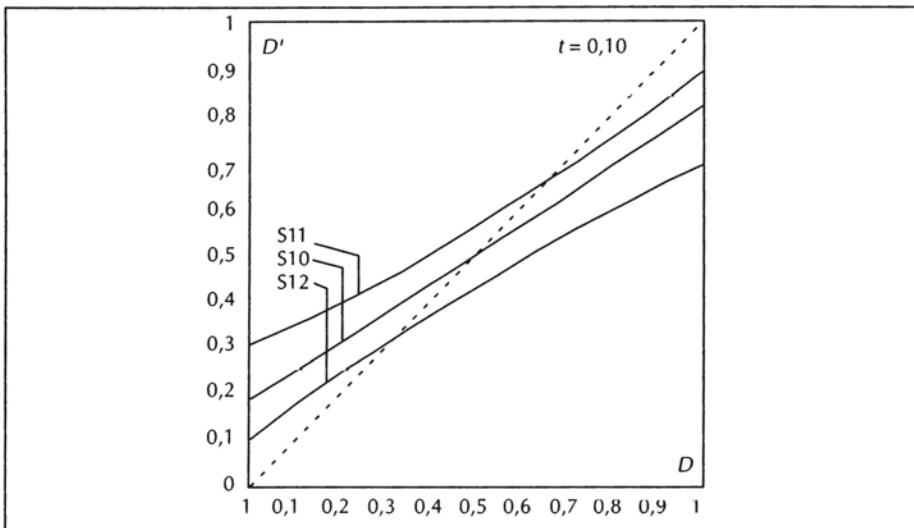


Figure 4. Estimation par simulation des dissimilarités  $D'$  associées à S10, S11 et S12, pour un taux d'erreur de 0,10 en fonction de la dissimilarité initiale  $D$ .

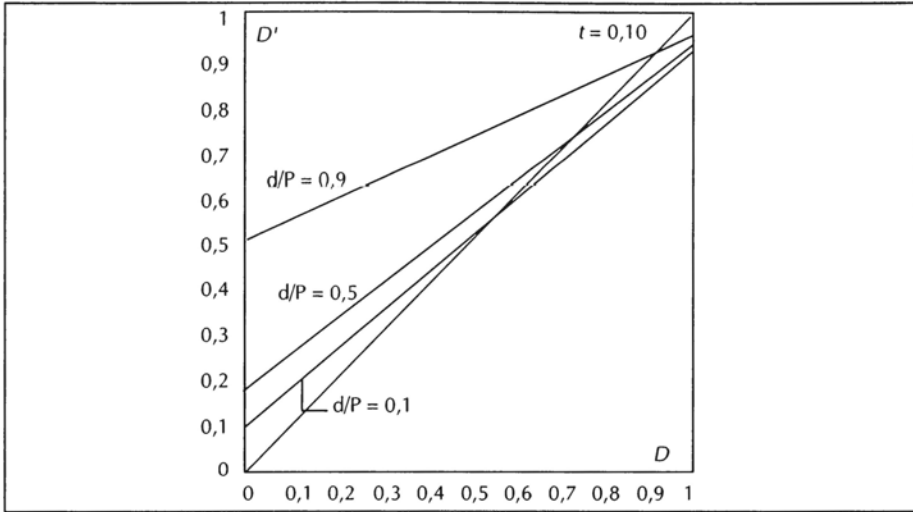


Figure 5. Estimation par simulation des dissimilarités  $D'$  associées à l'indice S7 (Jaccard), pour un taux d'erreur de 0,10 en fonction de la dissimilarité initiale  $D$ .

résultats obtenus sur la matrice complète et débarrassée des marqueurs particulièrement rares. Ils conduisent, en effet, pour la plupart des paires d'individus à augmenter le nombre d'absences en commun. Les indices S4, S5 et S6, correspondant aux trois types de moyennes du dénominateur, ont un comportement très comparable. Ils ont tendance à donner de plus faibles distorsions des petites valeurs de dissimilarité que l'indice de Jaccard, S7, lui-même plus favorable que S8. On les préférera donc pour cette raison.

## Mesures des dissimilarités sur variables qualitatives

Les caractères sont ici des variables qualitatives présentant un nombre fini de modalités : couleur des pétales, forme du stigmate... Intuitivement, deux individus sont proches s'ils possèdent la même modalité pour un grand nombre de variables.

Les indices utilisables correspondent à la généralisation des indices présentés pour les variables binaires symétriques en 0 et 1, qui sont en fait des variables qualitatives à deux modalités.

On note  $m$  le nombre de variables en accord,  $u$  le nombre de variables en désaccord,  $P$  étant toujours le nombre de variables. Les indices de similarité correspondants s'écrivent :

- $m / (m + u) = m / P$  (Sokal et Michener) ;
- $m / (m + 2u) = m / (P + u)$  (Rogers et Tanimoto) ;
- $2m / (2m + u) = 2m / (P + m)$  (Sokal et Sneath).

La dissimilarité s'obtient directement par  $d = 1 - s$ .

On utilise souvent la distance de  $\chi^2$  sur un tableau disjonctif complet. Pour une variable, on crée pour chacune de ses modalités une variable binaire prenant la valeur 1 si l'individu présente cette modalité, 0 sinon, et on calcule le  $\chi^2$  classique sur cette nouvelle matrice de données. Cette distance ne donne pas un même poids à toutes les modalités, en particulier les modalités rares prennent un poids très fort. Même si cet effet peut être parfois recherché, il paraît souvent indésirable.

## Le choix d'un indice de similarité sur marqueurs biochimiques et moléculaires

La ressemblance génotypique entre deux individus peut être mesurée par une similarité génétique définie à partir des formes alléliques des gènes observés. Certains types de marqueurs (isoenzymes, microsatellites...) donnent, en général, une information génétique directe et permettent un codage en allèles par identification de la composition allélique de chaque locus. La similarité génétique, telle qu'on la définit, est alors directement estimable. D'autres marqueurs ne permettent pas d'accéder à la totalité de l'information génétique et seul un phénotype est observé. La similarité génétique ne peut être calculée directement, mais les caractéristiques de ces marqueurs permettent de définir l'indice de similarité le plus pertinent. Il est intéressant d'évaluer de plus l'ordre de grandeur de l'erreur liée à cette perte d'information.

### LES MARQUEURS MULTIALLÉLIQUES

On fixe *a priori* une définition de la similarité génétique  $T_{ij}$  entre deux individus  $i$  et  $j$  comme la moyenne sur les  $L$  locus du rapport du nombre  $n_{ls}$  d'allèles du locus  $l$  présents simultanément chez les deux individus et du nombre  $n_{lc}$  d'allèles comparés :  $T_{ij} = (\sum_l n_{ls} / n_{lc}) / L$ . Si  $\pi$  est la ploïdie,  $n_{lc}$  vaut  $\pi$  pour tous les locus. En posant  $n_s = \sum_l n_{ls}$  et  $n_c = \pi L$ , alors  $T_{ij} = n_s / n_c = n_s / (\pi L)$ .

Pour une espèce diploïde, un locus peut être codé par autant de variables qu'il a d'allèles, chaque variable prenant les valeurs 2, 1 ou 0. Suivant ce codage, pour deux individus  $i$  et  $j$ , la combinaison des génotypes de chaque locus appartient à un des sept groupes, notés  $A$  à  $G$  (tableau 2). Le nombre de locus de chaque groupe est noté  $l_A$  à  $l_G$ . On peut alors écrire  $T_{ij}$  à partir de  $n_s = 2l_A + l_B + 2l_E + l_F$  et  $n_c = 2(l_A + l_B + l_C + l_D + l_E + l_F + l_G)$ .

On définit les paramètres  $n_{r,s}$  ( $r \geq s$ ) comme le nombre d'allèles présents  $r$  fois chez un individu et  $s$  fois chez l'autre. Chacun des groupes  $A$  à  $G$  contribue à ces paramètres. Par exemple, pour un locus à  $a_l$  allèles, une paire du groupe  $D$  participe une fois à  $n_{2,0}$ , deux fois à  $n_{1,0}$  et  $a_l - 3$  fois à  $n_{0,0}$ .

Tableau 2. Les sept combinaisons possibles des génotypes de deux individus diploïdes.

	A	B	C	D	E	F	G
Individu <i>i</i>	20..	200..	200..	2000..	110..	110..	1100..
Individu <i>j</i>	20..	110..	020..	0110..	110..	101..	0011..

On peut exprimer les effectifs  $I_A$  à  $I_G$  à partir de ces paramètres  $n_{r,s}$  d'où l'écriture de  $T_{ij}$  comme rapport de  $n_s = 2n_{2,2} + n_{2,1} + n_{1,1}$  et  $n_c = (2n_{2,2} + n_{2,1} + n_{1,1}) + (n_{2,1} + 2n_{2,0} + n_{1,0}) / 2$ .

On reconnaît alors dans l'expression de  $T_{ij}$  une généralisation de S4, l'indice de Dice. Le nombre d'accords est effectivement de 1 pour les génotypes contribuant à  $n_{2,1}$  et  $n_{1,1}$  et de 2 pour ceux contribuant à  $n_{2,2}$ . De même, les désaccords sont au nombre de 1 pour  $n_{2,1}$  et  $n_{1,0}$  et de 2 pour  $n_{2,0}$ . On note que pour une espèce haploïde, certains micro-organismes par exemple, on retrouverait directement l'indice de Dice.

L'indice de Dice appartient à la famille des indices qui ne prennent pas en compte l'information apportée par les doubles absences. Ce point de vue est logique puisque le nombre de doubles absences n'apporte aucune information sur la proximité des individus et ne dépend que du nombre d'allèles des locus.

#### LES MARQUEURS CODOMINANTS CODÉS EN BANDE

Pour des marqueurs de type RFLP ou les isoenzymes, il est parfois impossible de reconnaître les allèles appartenant à un même locus et seul un codage en bandes reste disponible. Pour une espèce diploïde, la présence d'une bande, codée 1, peut alors correspondre à un locus homozygote, qui devrait être codé 2, ou à un des allèles d'un locus hétérozygote, normalement codé 1. Les différents groupes de génotypes définis précédemment ne sont plus identifiables et le phénotype observé diffère du génotype réel (tableau 3).

Comme dans le cas précédent, l'information des doubles absences n'est pas pertinente et par analogie avec l'indice de Dice, on peut retenir, comme mesure de ressemblance, un indice de la famille  $S_{ij} = n_{1,1} / (n_{1,1} + \beta n_{1,0})$ . On

Tableau 3. Les sept combinaisons possibles des génotypes et les phénotypes observés par marqueurs codominants.

		A	B	C	D	E	F	G
Génotype	<i>i</i>	20..	200..	200..	2000.	110..	110..	1100..
	<i>j</i>	20..	110..	020..	0110.	110..	101..	0011..
Phénotype	<i>i</i>	10..	100..	100..	1000.	110..	110..	1100..
	<i>j</i>	10..	110..	010..	0110.	110..	101..	0011..



peut alors rechercher la valeur de  $\beta$  telle que, sous certaines hypothèses,  $S_{ij}$  approche au mieux  $T_{ij}$ .

Pour une espèce diploïde autogame qui serait homozygote pour tous les locus,  $S_{ij} = T_{ij}$  pour  $\beta = 1/2$ ,  $S_{ij}$  est alors directement l'indice de Dice.

On connaît, en général, pour une espèce donnée, une valeur de  $t$ , le taux d'hétérozygotie moyen. En admettant que cette valeur puisse s'appliquer à chaque locus, on peut définir les valeurs de  $\beta$ , fonction de  $t$ , qui annulent  $S_{ij} - T_{ij}$ . L'expression de cette différence est complexe et les relations entre  $\beta$  et  $t$  ont été recherchées numériquement pour différentes distributions du nombre d'allèles par locus, le maximum étant fixé à 7 (figure 6). Les distributions I, II et IV correspondent à des distributions extrêmes, la distribution III reproduit une distribution observée sur l'hévéa (M. Seguin, comm. pers.).

On remarque le cas des locus à faible nombre d'allèles (I ou II) où la similarité génétique n'est approchée, pour des hétérozygoties faibles, que pour des valeurs de  $\beta$  proches de 1.

Si l'on prend comme base de référence la distribution III, on pourrait proposer de fixer la valeur de  $\beta$  à 0,5 pour  $t \leq 0,1$ , à 0,6 pour  $0,1 < t \leq 0,3$ , à 0,7 pour  $t > 0,3$ .

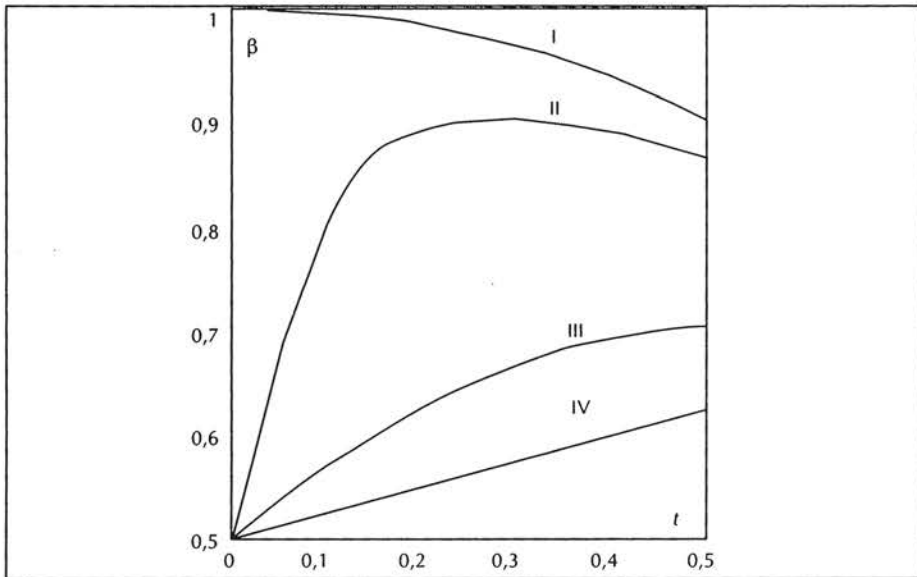


Figure 6. Valeurs de l'indice  $\beta$  de la dissimilarité  $D = a / [a + \beta(c + d)]$  telles que  $D_{ij}$  est égal à la dissimilarité génétique  $T_{ij}$  calculée à partir de l'information génétique complète, en fonction du taux d'hétérozygotie  $t$  et pour diverses distributions du nombre d'allèles :

I : tous les locus ont 2 allèles ;

II : locus à 2 ou 3 allèles dans les proportions 0,9 et 0,1 ;

III : locus de 2 à 7 allèles dans les proportions 0,27, 0,32, 0,19, 0,11, 0,07 et 0,03 ;

IV : tous les locus ont 7 allèles.

L'ordre de grandeur de la différence a été évalué par simulation, pour différents taux d'hétérozygotie et une série de valeurs de  $\beta$  entre 1/2 et 1, à partir d'une population de 40 individus décrits par 200 locus. L'écart  $S_{ij} - T_{ij}$  est composé d'un biais systématique, sans conséquence sur les méthodes d'analyse, et d'une erreur vraie qui seule nous intéresse ici. On exprime cette erreur par le demi-écart entre la plus faible et la plus forte des valeurs observées dans les simulations, elle est exprimée en pourcentage de la similarité génétique moyenne. Pour une distribution de type III, cette erreur maximale est de l'ordre de 3 % pour  $t$  voisin de 0,2, mais peut atteindre 5 % pour des espèces très hétérozygotes.

On montre, de plus, que cette erreur dépend surtout de  $t$  et peu de  $\beta$ . En pratique, on pourra le plus souvent retenir  $\beta = 0,5$ , correspondant à l'indice de Dice.

#### LES MARQUEURS DOMINANTS CODÉS EN BANDES

Les marqueurs dominants, RAPD ou AFLP, peuvent être considérés comme des locus à deux allèles particuliers puisqu'un seul allèle est repérable, l'autre étant un allèle nul non matérialisé sur le gel. Ces marqueurs sont dits dominants car il est impossible de connaître le nombre de copies d'un allèle pour un locus particulier et donc de distinguer les homozygotes des hétérozygotes.

Parmi les types  $A$  à  $G$  définis précédemment pour une espèce diploïde, seuls  $A$ ,  $B$ ,  $C$  et  $E$  existent puisque les locus n'ont que deux allèles. Les deux allèles ne jouant pas un rôle identique, les génotypes  $A$  et  $B$  doivent être subdivisés suivant les phénotypes, en distinguant l'allèle repérable de l'allèle nul. L'impossibilité du codage en allèles conduirait, comme pour les marqueurs RFLP, à un phénotype noté théorique dans le tableau 4. Comme l'allèle nul, noté par un point, n'est pas repérable, un seul des états alléliques est donc lisible, d'où le phénotype réellement lu.

Le nombre  $L$  de locus est le nombre  $P$  de bandes différentes observées dans la population. Il faut remarquer immédiatement que  $n_{0,0}$  le nombre de doubles absences, a ici un sens très clair. Il correspond au nombre de locus homozy-

Tableau 4. Génotypes, phénotypes théoriques et phénotypes observés par marqueurs dominants.

		$A1$	$A2$	$B1$	$B2$	$C$	$D$
Génotype	$i$	20	02	20	02	20	11
	$j$	20	02	11	11	02	11
Phénotype théorique	$i$	10	0.	10	0.	10	1.
	$j$	10	0.	1.	1.	0.	1.
Phénotype lu	$i$	1	0	1	0	1	1
	$j$	1	0	1	1	0	1

gotes pour l'allèle nul chez les deux individus et apporte autant d'information que le nombre de doubles présences. Il sera logique de retenir un indice symétrique en 0 et 1. Le dénominateur de la similarité génétique  $T_{ij}$  est égal, dans ce cas particulier, à  $P$ . Cette valeur conduit à retenir l'indice de Sokal et Michener, il s'exprime ici par  $S_{ij} = (n_{1,1} + n_{0,0}) / P$ .

La différence  $S_{ij} - T_{ij}$  qui mesure l'erreur d'estimation de la similarité génétique, s'exprime simplement par  $(I_{B1} - I_{B2}) / P$ . Cette différence est nulle si les fréquences des génotypes (2,0) et (0,2) sont égales, c'est-à-dire si les fréquences des deux allèles sont identiques pour tous les locus, elle restera faible si elles tendent à l'être, en moyenne, sur les locus. D'autre part, si l'espèce est très homozygote, les proportions de génotypes  $B$  sont faibles et  $I_{B1} - I_{B2}$  restera faible.

Pour des espèces plus hétérozygotes, cette différence peut être forte puisque les génotypes  $B$  peuvent représenter jusqu'à 50 % des effectifs. Comme précédemment, des simulations numériques ont permis d'en estimer l'ordre de grandeur pour différents niveaux d'hétérozygotie et pour différentes proportions des deux allèles d'un locus. L'erreur vraie est, comme précédemment, la demi-amplitude des extrêmes observés rapportée à la similarité génétique moyenne. Elle croît avec l'hétérozygotie et atteint 13 % pour  $t = 0,5$  et aucun déficit allélique, elle dépasse 18 % si les allèles non identifiables sont majoritaires et tombe à 10 % dans le cas contraire. Dans tous les cas, l'imprécision introduite par la perte d'information génétique est nettement plus importante que pour les marqueurs codominants.

On montre, sur ces exemples, que le contenu informatif des marqueurs permet de justifier le choix d'un indice de similarité, justification qui est le plus souvent absente des publications dans ce domaine.

## Les méthodes d'analyse factorielle

Les méthodes d'analyse factorielle ont pour objectif de repérer les structures les plus fortes dans les populations que l'on étudie et d'éliminer les particularités ponctuelles qui gênent la perception générale des phénomènes. Elles peuvent ainsi être d'une grande utilité lorsque l'on aborde l'étude de la diversité d'une espèce ou d'une population.

L'objet de ces méthodes est de produire une représentation géométrique des mesures de dissemblance entre unités. Cette représentation permet de manipuler de manière concrète la notion de diversité. On aboutit ainsi à une décomposition hiérarchisée de la diversité, qui permet de distinguer les tendances fondamentales des variations particulières.

L'utilisation de ces techniques nécessite de connaître les principes sur lesquels elles reposent et les conventions qu'elles utilisent. Après une présentation générale de l'analyse factorielle mettant l'accent sur ses aspects conventionnels, on verra comment elle s'intègre spécialement dans trois applications bien connues : l'analyse factorielle sur tableau de distances (AFTD), l'analyse en composantes principales (ACP) et l'analyse des correspondances multiples (ACM).

On abordera ensuite le problème du traitement des tableaux multiples, qui se rencontre en particulier en étudiant une population décrite par plusieurs types de marqueurs. La méthode de l'analyse factorielle multiple (AFM) permet, lorsque l'on dispose sur les individus d'informations provenant de différents points de vue, de mettre en évidence les caractères principaux, qui se manifestent de manière consensuelle, et ceux qui au contraire sont spécifiques d'un type de mesures.

## L'analyse factorielle : présentation générale

### NUAGE DE POINTS ET DISPERSION

Si la distance choisie pour mesurer les dissemblances entre les unités étudiées est euclidienne, l'ensemble des unités peut être représenté sous la forme d'un nuage de points dans un espace suffisamment vaste. Dans cet espace on peut distinguer un point  $B$ , appelé point moyen, ou barycentre, qui est défini comme étant le plus près possible de tous les points du nuage. Il correspond à un « type moyen » neutre dont chaque unité réelle diffère par ses caractéristiques propres.

Dans ces conditions, la diversité totale de la population est représentée par la dispersion du nuage autour de  $B$ . L'analyse factorielle convient de quantifier cette dispersion par l'inertie du nuage par rapport à  $B$ , c'est-à-dire par la somme des carrés des distances de chaque point à  $B$ .

### AXES PRINCIPAUX D'INERTIE ET COORDONNÉES DES INDIVIDUS

En général, l'espace dans lequel se trouve le nuage est un espace de dimension élevée. Pratiquement, cela signifie qu'en partant de  $B$ , on peut se diriger dans une multitude de directions. Autrement dit, il existe de très nombreuses façons pour un individu de se distinguer du type moyen et toutes les unités observées en diffèrent chacune à sa manière. Le but de l'analyse factorielle est de ramener approximativement toute particularité individuelle à la composition d'un petit nombre de directions très usitées et indépendantes les unes des autres.

Pour juger de la « fréquence d'usage » d'une direction  $U$ , on construit un critère en sommant pour tous les individus une valeur d'autant plus importante que l'angle que fait  $U$  avec la direction spécifique de l'individu est aigu et que

l'individu est à une distance élevée du point moyen B. Ce critère s'appelle l'inertie du nuage le long de  $U$ . La direction qui maximise ce critère s'appelle l'axe principal d'inertie du nuage. L'analyse factorielle produit successivement les autres axes en choisissant pour chacun la direction qui maximise le critère de l'inertie parmi les directions orthogonales aux axes déjà identifiés. L'orthogonalité des axes assure l'indépendance des variations résumées par ces axes.

Il résulte de ce processus un système d'axes indépendants expliquant successivement le maximum de variation. On sait calculer les coordonnées de chaque unité sur ces axes et ainsi les placer dans ce système orthogonal. L'ensemble des coordonnées de tous les individus sur un axe s'appelle un facteur.

Les coordonnées factorielles permettent de reconstituer la dispersion totale de la population. Cependant, et c'est l'intérêt essentiel des analyses factorielles, la sélection des premiers facteurs seulement permettra en général de reconstruire une grande part de la dispersion totale. Plusieurs critères plus ou moins empiriques ont été proposés pour déterminer le nombre d'axes à conserver (SAPORTA, 1990). Il convient cependant de rester circonspect en réalisant cette sélection et le choix des axes doit être raisonné dans chaque cas.

Sans autre information, les coordonnées des individus sur les axes n'ont guère de signification. Pour qu'une analyse factorielle soit interprétable, il faut disposer d'une source d'information « externe » que l'on pourra relier aux axes factoriels. Cette information externe peut ne pas être explicite et n'apparaître que lorsque l'on constate que tel axe rassemble ou, au contraire, oppose certains individus reconnaissables. Le plus souvent pourtant, un axe sera interprété en constatant qu'il est lié à des variables mesurées sur les individus, qu'elles aient ou non participé au calcul de la matrice de distance.

Ce n'est que lorsqu'un axe est interprété et que la part de dispersion qui lui est attachée est compréhensible que l'on considérera qu'il s'agit d'un des facteurs qui rentrent en ligne de compte dans l'élaboration de la diversité.

#### REPRÉSENTATIONS GRAPHIQUES

Pour avoir une représentation visuelle de la diversité, mais aussi pour faciliter l'interprétation des axes, les méthodes d'analyse factorielle proposent des représentations graphiques des individus et le cas échéant des variables. Ces graphiques constituent des apports déterminants de ces méthodes. Ils permettent de voir, au sens propre du terme, des oppositions, des regroupements et des tendances difficiles à percevoir avec des statistiques chiffrées. Chaque type de graphique a cependant ses règles de lecture propres qu'il convient de bien maîtriser pour éviter toute interprétation hasardeuse.

### Les méthodes classiques d'analyse factorielle

En pratique, l'analyse factorielle telle qu'elle vient d'être présentée ne constitue qu'une étape dans un processus qui part d'un tableau de données, opère

d'éventuels codages, calcule s'il le faut les distances, trouve les axes factoriels et produit des aides à l'interprétation adaptées à la nature des données. C'est l'ensemble de ce processus qui forme une « méthode » d'analyse factorielle. Selon le type de données dont on dispose, les méthodes les plus couramment employées sont : l'AFTD pour un tableau de distances, l'ACP pour un tableau de variables quantitatives et l'ACM pour un tableau de variables qualitatives.

#### L'ANALYSE FACTORIELLE SUR TABLEAU DE DISTANCES

L'analyse factorielle sur tableau de distances est utilisée pour traiter une matrice de distances calculées avec un indice original, choisi de manière adaptée à ses propres données. Il est cependant nécessaire que la distance obtenue soit une distance euclidienne ou rendue euclidienne par transformation.

Dans ce cas, deux individus  $i$  et  $j$  peuvent être représentés par deux points  $e_i$  et  $e_j$  d'un espace de dimension  $K$  inconnue tels que le carré de la distance  $d_{ij}$  peut s'écrire :  $d_{ij}^2 = (e_i - e_j)'(e_i - e_j)$ .

D'un point de vue calculatoire, on montre que les axes successifs de l'analyse factorielle et l'inertie du nuage sur chaque axe correspondent aux vecteurs propres et aux valeurs propres de la matrice  $W$  des produits scalaires entre éléments. Si l'origine des coordonnées est placée au point moyen du nuage, les produits scalaires  $w_{ij} = \langle e_i ; e_j \rangle$  sont entièrement déterminés par les  $d_{ij}$  par la formule de Torgerson :  $w_{ij} = -(d_{ij}^2 - d_i^2 - d_j^2 + d^2) / 2$ .

La diagonalisation de cette matrice  $W$  permet d'extraire les vecteurs propres et les valeurs propres associées.

Le nombre de vecteurs propres associés à une valeur propre non nulle est la dimension  $K$  de l'espace contenant le nuage de points.

L'AFTD constitue la mise en œuvre la plus simple de l'analyse factorielle. Dans ces conditions, la seule aide à l'interprétation disponible est le graphique des individus sur un choix quelconque de deux axes factoriels.

#### L'ANALYSE EN COMPOSANTES PRINCIPALES

L'analyse en composantes principales traite de variables quantitatives et impose une distance euclidienne entre individus. Les distances euclidiennes qui peuvent être calculées ne diffèrent que par les poids que l'on accorde aux variables. Le système de pondération le plus usuel consiste à ramener toutes les variables à une même échelle en les pondérant par l'inverse de leur écart type. On dit alors que l'ACP est normée, ou centrée-réduite.

De même que pour l'AFTD, les coordonnées des individus sur les composantes principales peuvent se calculer à partir de la matrice des produits scalaires. Celle-ci s'obtient immédiatement à partir du tableau des données centrées.

L'interprétation des axes fournis par une ACP est fortement enrichie puisque les corrélations entre les facteurs et les variables du tableau peuvent être calculées. L'examen de ces corrélations pour chaque axe permet de mettre en évidence la ou les variables les plus liées à cet axe. On met ainsi en évidence des faisceaux de variables qui varient conjointement et dont la variation conjointe provoque une forte discrimination des individus.

#### L'ANALYSE DES CORRESPONDANCES MULTIPLES

L'analyse des correspondances multiples (ACM) permet d'étudier des individus décrits par des variables qualitatives à modalités. Cette méthode calcule les distances entre individus par la distance du  $\chi^2$  sur le tableau disjonctif complet (voir ci-dessus). C'est-à-dire que deux individus sont d'autant plus distants l'un de l'autre qu'ils présentent de nombreux désaccords et que leurs désaccords leur font adopter à l'un ou à l'autre des modalités rares.

Cette méthode est aussi connue sous les noms d'analyse factorielle des correspondances multiples (AFCM), d'analyse factorielle des correspondances sur tableau disjonctif complet (AFC sur TDC), voire simplement d'analyse factorielle des correspondances (AFC).

L'interprétation des axes fournis par l'analyse factorielle de cette matrice de distances se déduit des relations qu'ils possèdent avec trois types d'objets : les individus, les variables et les modalités.

□ Les individus fournissent les mêmes interprétations que dans les autres méthodes d'analyse factorielle.

□ Comme en ACP, on peut évaluer le lien entre les variables du tableau et chaque axe. Pour une variable qualitative, ce lien s'apprécie par le rapport  $\eta^2$ , qui est égal à la part de l'inertie totale de l'axe reconstituée par la somme des carrés des écarts intermodalités. Un facteur peut alors être perçu comme une synthèse numérique d'un certain nombre de variables qui varient conjointement.

□ Une variable qualitative est un objet plus complexe qu'une variable quantitative et la diversité qu'elle induit se manifeste sur d'autant plus d'axes que la variable a de modalités. Une interprétation plus fine des axes et des liaisons entre variables doit donc être recherchée au niveau des modalités. Cependant, le choix de la distance du  $\chi^2$  fait que la forte contribution d'une modalité à un axe peut n'être due qu'à la rareté de cette modalité et donc à l'excentricité du petit nombre d'individus qui la possède. Lorsqu'il n'en est pas ainsi, un axe révèle un ensemble de modalités en association mutuelle, c'est-à-dire présentes ou absentes simultanément chez un grand nombre d'individus.

En ce qui concerne les représentations graphiques, la distance du  $\chi^2$  fait qu'en disposant sur un même graphique les individus et les modalités, un individu est proche des modalités qu'il possède et une modalité est proche des individus qui la possèdent.

## Le traitement des tableaux multiples et l'analyse factorielle multiple

### L'ANALYSE CONJOINTE DE PLUSIEURS TABLEAUX

Face à des données de nature différente, on peut souhaiter combiner l'information apportée par les différents tableaux de façon à mettre en évidence des structures plus fines dans la population.

Il peut arriver en effet qu'une organisation des individus qui apparaît lors de l'analyse d'un certain tableau et qui est clairement due à des variations de certains caractères importants de ce tableau ne se manifeste pas lors de l'analyse d'un autre type de données. Une approche pourrait être de procéder à l'analyse conjointe d'un supertableau juxtaposant les différentes données. On ferait ainsi apparaître toutes les oppositions possibles, et on repérerait celles qui sont communes à tous les types de données et celles qui, au contraire, sont spécifiques à un tableau donné.

Deux obstacles majeurs s'opposent cependant à l'analyse immédiate d'un tel supertableau :

- le type des variables — qualitatif ou quantitatif — n'est pas nécessairement le même pour tous les tableaux, il conviendrait *a priori* d'analyser les tableaux qualitatifs par l'ACM et les tableaux quantitatifs par l'ACP ;
- tous les tableaux ne possèdent pas nécessairement la même inertie et n'expriment pas forcément la même organisation de la diversité. Les tableaux les plus fortement structurés risquent donc d'influencer les résultats de façon prépondérante. L'apport des tableaux les moins structurés passerait alors inaperçu.

Un résultat intéressant permet de contourner le premier obstacle : on démontre en effet que l'on peut obtenir les mêmes facteurs que ceux qui sont produits par l'ACM en réalisant une ACP, pondérée de manière adéquate, du tableau formé par les indicatrices des variables qualitatives, c'est-à-dire du tableau codé sous forme disjonctive complète. Dans le cas des données moléculaires, cela signifie que chaque marqueur doit être représenté par deux colonnes : l'une indiquant sa présence et l'autre son absence.

Le second obstacle se contourne en adoptant une pondération des tableaux qui réduit l'influence des tableaux ayant une structure forte. Pour cela il faut quantifier *a priori* la part de l'inertie totale d'un tableau qui correspond à sa structure interprétable. Cette quantification est assez arbitraire et c'est l'un des points sur lesquels se distinguent les différentes méthodes d'analyse des tableaux multiples. La méthode Statis (LAVIT, 1988) choisit comme mesure la somme des carrés des valeurs propres de l'analyse séparée du tableau. L'AFM (ESCOFIER et PAGES, 1993) préfère la plus grande de ces valeurs propres. Sans entrer dans le détail, on peut dire que la première peut réduire l'importance d'un tableau dont, par hasard, les dernières valeurs propres seraient élevées.



La seconde au contraire peut donner plus de poids qu'il n'en faudrait à un tableau dont les premières valeurs propres ne sont pas très différentes les unes des autres.

#### L'ANALYSE FACTORIELLE MULTIPLE

L'analyse factorielle multiple est donc une ACP particulière, dans laquelle toutes les variables d'un tableau sont pondérées par l'inverse de la variance de l'axe principal d'inertie de l'analyse séparée du tableau.

En tant qu'ACP, elle permet de repérer les individus qui se ressemblent pour l'ensemble des variables et donc de faire une typologie des individus. Elle permet aussi d'attribuer le passage d'une classe de cette typologie à une autre, à la variation conjointe de quelques variables. Ce sont ces directions communes de variation que l'AFM produit comme axes d'inertie, sous le nom d'axes de l'analyse globale.

Cependant, l'organisation des données sous forme de juxtaposition de tableaux enrichit l'interprétation de l'AFM. D'une part, l'AFM fournit pour chaque axe de l'analyse globale des indices permettant de déterminer s'il n'est dû qu'à des variables d'un seul tableau ou s'il est commun à plusieurs types de données. D'autre part, l'AFM calcule les corrélations entre les axes de l'analyse globale et les axes que produiraient les analyses séparées de chaque tableau. Lorsque le critère précédent indique qu'un certain axe de l'analyse globale est commun à plusieurs tableaux, cela permet, pour aller plus loin dans l'analyse, de déterminer à quel caractère précisément au sein de chaque type de mesures, l'axe commun est lié.

Enfin, ces corrélations permettent de connaître, pour un certain axe d'une analyse séparée dont on saisit parfaitement le sens, le numéro de l'axe de l'analyse globale auquel il contribue essentiellement. Si cet axe global est commun à plusieurs tableaux, cela signifie que l'axe partiel se manifeste de manière consensuelle. Au contraire, si l'axe global est spécifique, cela signifie que l'axe partiel correspond à un caractère qui est propre au type d'observations du tableau séparé.

L'AFM possède par ailleurs un certain nombre de propriétés intéressantes pour ce qui est de la représentation des individus. En particulier, on montre qu'elle offre un compromis intéressant entre d'une part, la qualité de représentation des nuages correspondants à chaque tableau séparé et d'autre part, la proximité pour un même individu de ses représentations partielles.

## La représentation arborée des dissimilarités

Le principe de toute méthode de représentation arborée est d'approcher au mieux la dissimilarité  $\delta$ , choisie pour sa pertinence à décrire les relations entre

individus, par une distance  $d$  représentable sous forme d'un arbre, c'est-à-dire une ultramétrie ou une distance additive d'arbre (BARTHELEMY et GUENOCHÉ, 1988).

La recherche de la solution exacte nécessite l'énumération de toutes les configurations d'arbre possibles sur les  $n$  individus. Pour chaque structure d'arbre possible, on estime au sens des moindres carrés les longueurs des arêtes de l'arbre. On retient finalement l'arbre qui permet de minimiser la somme des carrés des écarts entre dissimilarité initiale et distance reconstituée dans l'arbre. On montre, par récurrence, que le nombre d'arbres binaires différents construits sur  $n$  individus est  $\prod_{i=3,n} (2i - 5)$ . Pour  $n = 10$ , on dénombre plus de  $2 \cdot 10^6$  arbres différents et plus de  $2 \cdot 10^{20}$  pour  $n = 20$ . L'énumération des tous les arbres devient donc irréalisable, même avec les moyens informatiques les plus puissants, dès que  $n$  dépasse quelques dizaines.

Dans ces situations, la seule possibilité est de construire, à partir d'heuristiques raisonnables, des solutions qui seront les meilleurs possibles mais que l'on ne pourra jamais assurer comme optimales. Diverses heuristiques ont été proposées, elles permettent de construire des algorithmes de plus ou moins grande complexité, la complexité d'un algorithme mesurant la croissance du temps de calcul lorsque  $n$  augmente. Dans le cadre de cet ouvrage, ne seront évoquées que les méthodes de regroupement qui sont de complexité suffisamment faible pour traiter quelques centaines d'individus.

Les individus étudiés sont souvent décrits par plusieurs types de variables qui ne peuvent être réunis dans le calcul d'une dissimilarité unique. Chaque jeu de variables est alors traité séparément et l'on s'interroge sur la présence de structures identiques dans les différents arbres. Deux méthodes de construction d'arbres synthétiques, les arbres consensus et les sous-arbres communs, seront présentées.

## Les méthodes de regroupement

Ce sont des méthodes itératives qui procèdent par agglomérations ascendantes successives, en construisant l'arbre pas-à-pas. Initialement, la matrice traitée a autant d'éléments que d'individus de la population étudiée et on fixe à l'arbre une structure en étoile. A chaque itération, deux éléments, individus ou groupes déjà formés, définis comme voisins, sont réunis pour former un groupe. Ils sont choisis pour que l'arbre qui découle de leur regroupement optimise un critère fixé. Ce groupe devient un nouvel élément fictif qui remplace les deux éléments réunis, la matrice est mise à jour et diminue alors d'une unité. Le processus est réitéré jusqu'à ce que tous les individus soient réunis dans un même groupe.

Les diverses méthodes de ce type se caractérisent par des choix différents aux trois points clés de chaque itération : la sélection des éléments à réunir, qui

dépend de la définition du voisinage retenue, la mise à jour de la matrice de dissimilarité par calcul d'une dissimilarité entre le groupe formé et les autres éléments, la construction des longueurs des deux arêtes issues des deux éléments réunis.

#### LA DÉFINITION DU VOISINAGE

La définition la plus naturelle est de choisir comme voisins les deux individus, ou groupes, présentant la plus petite dissimilarité. Les éléments  $i$  et  $j$  sont définis comme voisins si  $\delta(i,j)$  est la plus petite des valeurs de dissimilarité.

Ce critère permet de retrouver l'arbre solution dans un cas théorique où la distance initiale serait déjà une ultramétrie. En revanche, il ne permet pas forcément de retrouver la bonne structure si la dissimilarité initiale est une distance additive d'arbre. Sur l'exemple de la figure 2 sur quatre points seulement, on peut imaginer que la distance  $d(i,k) = a_i + a_c + a_k$  soit la plus petite des distances bien que  $i$  et  $k$  soient dans deux paires opposées par l'arête centrale. Ce critère regroupera obligatoirement  $i$  et  $k$  et ne permettra pas de retrouver l'arbre vrai. Pour cela d'autres critères, qui prennent en compte l'ensemble des distances pour juger d'un voisinage, doivent être utilisés.

SAITOU et NEI (1987), dans un contexte de diversité génétique, proposent un critère de voisinage fondé sur le principe de parcimonie, qui est à la base de l'approche phylogénétique. L'objectif est de créer un arbre qui soit de longueur totale minimale. On considère qu'une arête représente un nombre d'événements mutationnels et en vertu d'un principe de base qui veut que l'évolution procède toujours par modification génétique la plus simple possible, le nombre d'événements, et donc la longueur totale, doit être minimal. Ces considérations conduisent à une définition de voisinage relatif, défini en minimisant un critère  $Q(i,j)$ , fonction de  $\delta(i,j)$  et de la moyenne des dissimilarités de  $i$  et  $j$  aux  $n - 2$  autres éléments  $k$  :

$$Q(i,j) = \delta(i,j) - (\sum_k [\delta(i,k) + \delta(j,k)]) / (n - 2).$$

On montre que ce critère possède des propriétés d'optimalité en termes de moindres carrés et que son domaine d'application dépasse la cadre de la reconstruction phylogénétique. Il peut s'interpréter, très généralement, comme une pondération de la dissimilarité entre deux individus par leurs dissimilarités aux autres individus. Deux individus proches, qui diffèrent fortement des autres individus, sont plus ressemblants que deux individus proches, mais proches également des autres individus. Cette attitude est, dans bien des cas, justifiée et explique le succès de cette méthode.

SATTATH et TVERSKY (1977) adoptent une démarche très différente. Ils partent de la caractérisation d'une distance additive d'arbre par la condition des quatre points. Pour quatre individus quelconques  $i, j, k$  et  $l$ , si  $d(i,j) + d(k,l)$  est la plus petite des trois sommes des distances deux à deux alors les deux plus grandes sont égales :  $d(i,k) + d(j,l) = d(i,l) + d(j,k)$ . La dissimilarité initiale  $\delta$  n'est pas

une distance additive mais il est toujours possible de former les sommes des dissimilarités deux à deux. Parmi ces trois sommes, l'une est la plus petite,  $\delta(i,j) + \delta(k,l)$  par exemple, on considère alors que les paires de points  $(i,j)$  et  $(k,l)$  sont de bons candidats pour être des voisins. On leur attribue un score de 1 alors que les autres paires de points,  $(i,k)$ ,  $(j,l)$ ,  $(i,l)$  et  $(j,k)$ , se voient attribuer un score nul. On balaie tous les quadruplets de points que l'on peut former sur les  $n$  points et on retient finalement comme voisins les individus de la paire qui a le score total le plus grand. Cette définition de voisinage topologique est de nature ordinale puisque seule importe l'ordre des trois sommes et non pas directement les valeurs des dissimilarités qui les composent. Elle paraît particulièrement adaptée lorsque les valeurs des dissimilarités sont plus ou moins entachées d'erreur.

#### LA MISE À JOUR DES DISSIMILARITÉS

Deux groupes, composés éventuellement d'un seul individu, étant réunis pour former un nouveau groupe, il est nécessaire de définir une dissimilarité entre le nouvel élément fictif créé et les autres éléments présents à cette itération. Soient  $i$  et  $j$  les groupes réunis pour former un nouvel élément  $s$ ,  $c_i$  et  $c_j$  les effectifs de ces groupes, et  $k$  un autre élément, la définition la plus naturelle de  $\delta(s,k)$  est la moyenne arithmétique des dissimilarités entre  $k$  et les individus composant  $i$  et  $j$  :  $\delta(s,k) = [c_i \delta(i,k) + c_j \delta(j,k)] / (c_i + c_j)$ . Elle correspond à un critère non pondéré dans la mesure où tous les individus jouent le même rôle.

On utilise souvent aussi une moyenne arithmétique que l'on dit pondérée, au sens où un poids différent doit être attribué aux individus des groupes pour que la moyenne ne dépende pas des effectifs des groupes et s'écrive :  $\delta(s,k) = [\delta(i,k) + \delta(j,k)] / 2$ . Le choix entre ces deux critères dépend de la nature de la population étudiée. Si l'ensemble relève d'un réel processus d'échantillonnage dans une structure de diversité donnée, alors le nombre d'individus dans chaque élément a un sens et doit être pris en compte dans le calcul de la dissimilarité. Si l'ensemble n'est qu'une réunion circonstancielle d'unités sans caractère de représentativité, l'effectif de chaque groupe n'est que le fait du hasard et n'a pas à intervenir dans le calcul de la diversité.

Dans le cas d'ajustement à une distance additive, on utilise plutôt la formule  $\delta(s,k) = [\delta(i,k) + \delta(j,k) - \delta(i,j)] / 2$ , ou son équivalent pondéré. Cette formule permet, lors de la reconstruction des arêtes, de retrouver les bonnes longueurs d'arêtes lorsque l'arbre initial est déjà une distance additive. Cette modification est sans influence sur les choix des pas ultérieurs.

Il a parfois été proposé d'utiliser, pour l'ajustement à une ultramétrie, les critères dits du lien simple et du lien complet qui correspondent respectivement à  $\delta(s,k) = \min[\delta(i,k), \delta(j,k)]$  et  $\delta(s,k) = \max[\delta(i,k), \delta(j,k)]$ . Sauf cas très particulier, il est difficile de justifier ces critères dans des problèmes de diversité. Si le lien simple a été beaucoup étudié, c'est essentiellement parce qu'il entraîne des propriétés mathématiques remarquables de l'ultramétrie produite.

## LA RECONSTRUCTION DES ARÊTES

Les éléments  $i$  et  $j$  sont réunis pour former le nœud  $s$  de l'arbre, il reste à fixer les longueurs des arêtes  $l(i,s)$  et  $l(j,s)$ . On peut simplement répartir également la dissimilarité  $\delta(i,j)$  entre les deux arêtes :  $l(i,s) = l(j,s) = \delta(i,j) / 2$ . Ce mode de calcul correspond à un ajustement de la dissimilarité initiale par une ultramétrie.

Si l'on n'impose pas la condition ultramétrique, on accepte que les deux arêtes puissent avoir des longueurs différentes pour, qu'en moyenne, les dissimilarités des  $n - 2$  éléments  $k$  à  $i$  et à  $j$  soient représentées au mieux. La différence des longueurs d'arêtes, notée  $e$ , est la somme sur tous les éléments  $k$  de  $[\delta(i,k) - \delta(j,k)] / (n - 2)$ . D'où les formules de calcul des longueurs des arêtes :  $l(i,s) = [\delta(i,j) + e] / 2$  et  $l(j,s) = [\delta(i,j) - e] / 2$ . Ce mode de calcul correspond à un ajustement de la dissimilarité initiale par une distance additive d'arbre.

Ces formules de reconstruction des arêtes ne garantissent pas que ces estimations soient les meilleures et, par ailleurs, ces longueurs d'arêtes n'interviennent pas dans les étapes suivantes. Il est donc souvent préférable de réestimer *a posteriori* les longueurs d'arêtes lorsque la topologie de l'arbre a été établie. En effet, on sait estimer globalement, au sens des moindres carrés, la longueur de chaque arête d'un arbre de topologie donnée. Pour cela, on exprime la distance dans l'arbre entre deux individus  $i$  et  $j$  comme la somme des longueurs des arêtes appartenant au chemin de  $i$  à  $j$  et on souhaite que cette distance soit la plus proche, au sens des moindres carrés, de la dissimilarité initiale. On résout alors un système de  $n(n - 1) / 2$  équations à autant d'inconnues que d'arêtes. Cette réestimation fait perdre la propriété ultramétrique.

## QUELQUES ALGORITHMES CONNUS

Les algorithmes les plus connus correspondent à certaines combinaisons compatibles parmi les différentes combinaisons possibles des modes de définition du voisinage, des diverses formules de mise à jour des dissimilarités et des modes d'estimation des longueurs des arêtes.

L'ensemble des méthodes souvent qualifiées de classification ascendante hiérarchique correspond à la définition du voisinage selon la dissimilarité minimale, à un ajustement à une ultramétrie et propose les diverses formules de mise à jour. Parmi celles-ci, les formules de type moyenne ou moyenne pondérée sont les plus utilisées et les méthodes correspondantes sont fréquemment qualifiées par les abréviations anglo-saxonnes UPGMA et WPGMA pour *unweighted (ou weighted) pair group method using average*.

La méthode NJtree (pour *neighbor-joining*) proposée par SAITOU et NEI (1987) est souvent utilisée en diversité génétique. Elle utilise le critère de voisinage relatif, le critère de mise à jour des dissimilarités est celui de la moyenne non pondérée et les dissimilarités sont ajustées à une distance additive d'arbre. La complexité calculatoire est de l'ordre de  $n^3$  et permet de traiter des tableaux de

quelques centaines d'individus. De nombreuses études par simulation ont montré l'efficacité de cette méthode pour retrouver l'arbre vrai. GASCUEL (1997) propose une version UNJ (*unweighted neighbor-joining*) qui utilise un critère de moyenne pondérée. Le choix entre ces deux méthodes dépend, comme cela a déjà été signalé, de la nature de l'échantillon étudié.

SATTATH et IVERSKY (1977) proposent l'algorithme Addtree (ou méthode des scores) qui utilise la moyenne comme critère d'agrégation et fait un ajustement à une distance additive d'arbre. La notion de voisinage est celle du voisinage topologique dont on a déjà souligné l'intérêt. La complexité en  $n^5$  est nettement plus élevée que celle de NJtree. Cette méthode, dont on a également démontré l'efficacité, pourra être préférée pour son point de vue topologique.

## Arbres consensus et sous-arbres communs

Le même ensemble d'individus est fréquemment décrit par plusieurs types de marqueurs de la diversité. A chaque type de marqueur correspond une matrice de dissimilarité et un ajustement à un arbre. Chaque type de marqueur apporte une information différente, les arbres produits sont donc différents. Cependant, on peut espérer en dégager une structure forte, un arbre consensus, qui serait commune aux différents arbres.

Le problème de la comparaison d'arbres se rencontre également dans les approches de rééchantillonnage (*bootstrap* ou *jackknife*). L'objet de ces approches est de tester la stabilité des représentations arborées obtenues. Le principe est de comparer les arbres établis à partir de nombreux échantillons tirés aléatoirement dans l'ensemble des marqueurs observés (FELSENSTEIN, 1985). Si les marqueurs sont nombreux et de contenu informatif équivalent, des marqueurs moléculaires par exemple, chacun de ces échantillons est censé donner la même image de la diversité. Si tous les arbres obtenus sont très proches, on conclura à la stabilité de la structure mise en évidence. Si, au contraire, différents échantillons aléatoires des marqueurs donnent des arbres très différents, alors la structure obtenue devra être regardée avec la plus grande prudence. Une mesure de dispersion autour d'un arbre consensus synthétique quantifie les écarts entre ces arbres.

Considérant que la principale information sur la diversité relève beaucoup plus de la structure de l'arbre que de la longueur des arêtes, les méthodes présentées ici construisent des arbres synthétiques en terme de structure, sans que la longueur des arêtes soit prise en compte.

### LES ARBRES CONSENSUS

Diverses méthodes de consensus ont été proposées sur des arbres racinés. Une généralisation directe aux arbres non racinés ne pose pas de problème majeur.

La méthode du consensus strict ne retient que les arêtes qui sont présentes dans tous les arbres comparés. Cette méthode pouvant être considérée comme trop sévère, il a été proposé la règle du consensus majoritaire, qui retient les arêtes présentes dans plus de 50 % des arbres. Si l'on compare deux arbres seulement, ces deux consensus sont équivalents.

Une mesure de la différence entre deux arbres peut être définie comme la somme des différences de chacun à leur arbre consensus. L'arbre consensus est toujours plus simple en terme d'organisation que les arbres initiaux. On quantifie cette organisation par un indice de complexité  $v$ , la dissimilarité entre deux arbres  $H$  et  $H'$ , de consensus  $H_c$ , est alors :

$$d(H, H') = v(H) + v(H') - 2v(H_c).$$

Diverses définitions de complexité peuvent être considérées. La plus habituelle est de mesurer la complexité d'un arbre par son nombre d'arêtes. Pour les arbres binaires qui n'ont que des nœuds de degré 3, la complexité de  $H$  et  $H'$  vaut  $2n - 3$ , la complexité est la somme du nombre d'arêtes externes,  $n$ , et du nombre d'arêtes internes,  $I_c$ , d'où  $d(H, H') = 2n - 6 - 2I_c$ .

La distance, dite distance d'arêtes, varie donc suivant le nombre d'arêtes internes conservées dans l'arbre consensus. Si  $I_c$  vaut 0, le consensus est une étoile, les arbres comparés ne présentent aucune structure commune, à l'opposé si  $I_c$  vaut  $n - 3$ , les arbres initiaux sont identiques. On normalise souvent cette distance par  $2n - 6$  pour lui garder une variation entre 0 et 1.

Une interprétation de la valeur d'une distance ne peut se faire que par référence à la valeur qui serait obtenue sur des arbres tirés aléatoirement, donc sans autre structure commune que celle liée au hasard. On ne connaît la distribution de cet indice, pour des arbres aléatoires, qu'asymptotiquement, c'est-à-dire lorsque le nombre d'individus tend vers l'infini. On peut cependant estimer ces distributions, pour des effectifs finis, par simulation, et donc construire un test statistique approché. Le tableau 5 donne les valeurs seuils pour différentes valeurs de  $n$  et différents seuils de probabilité. Par exemple, pour un arbre sur 40 individus, il y a 5 % de chances seulement que le hasard donne des arbres de distance inférieure à 0,973. Cette distance apparaît très peu discriminante et on refusera le plus souvent l'hypothèse aléatoire même si les arbres n'ont en commun qu'un très petit nombre d'arêtes internes.

La distance d'arêtes porte le même jugement sur toutes les arêtes quelle que soit leur position dans l'arbre. La complexité d'un arbre ne dépend que de son nombre d'arêtes indépendamment de son organisation interne. Une idée intuitive serait de donner un poids différent aux arêtes suivant leur aptitude à « structurer » l'arbre. Par exemple, une arête externe qui concerne un seul élément est peu structurante, à l'inverse une arête qui partage l'ensemble en deux sous-groupes d'effectif  $n / 2$  est la plus structurante. On associe donc à une arête un poids qui est le produit des effectifs de chacun des sous-groupes définis par cette arête. La complexité est donc la somme de ces poids sur toutes les arêtes. On exprime alors, comme précédemment, une distance, dite

de bipartitions, en fonction de la complexité des arbres et de leur consensus strict. Le maximum atteint ne s'exprime pas simplement mais peut être calculé à partir de  $n$ , la distance est alors normalisée par ce maximum. Comme pour la distance en arêtes, la distribution sous un modèle aléatoire n'est pas connue mais on peut l'approcher par simulation. Le tableau 6 donne les valeurs seuils pour différentes probabilités d'un test statistique approché. Il s'avère nettement plus discriminant.

Tableau 5. Distance d'arêtes entre arbres. Valeurs seuils pour différents niveaux de probabilité. Test empirique établi à partir de la simulation de 20 000 paires d'arbres aléatoires à  $n$  feuilles.

n	Probabilité (%)			
	1	5	10	20
20	0,882	0,941	0,941	1
40	0,973	0,973	0,973	1
60	0,982	0,982	0,982	1
80	0,987	0,987	0,987	1
100	0,990	0,990	0,990	1

Tableau 6. Distance de bipartition entre arbres. Valeurs seuils pour différents niveaux de probabilité. Test empirique établi à partir de la simulation de 20 000 paires d'arbres aléatoires à  $n$  feuilles.

n	Probabilité (%)			
	1	5	10	20
20	0,727	0,753	0,766	0,782
40	0,580	0,600	0,611	0,626
60	0,492	0,511	0,522	0,537
80	0,435	0,453	0,464	0,477
100	0,394	0,412	0,421	0,434

#### LES SOUS-ARBRES COMMUNS

Il n'est pas rare, dans certaines applications, d'obtenir des arbres consensus d'aspect très étoilé, donc sans structure marquée, alors que l'examen direct des arbres initiaux révèle des analogies de structure évidentes.

L'arbre consensus fait l'hypothèse que tous les individus sont correctement représentés, le déplacement d'un individu d'un groupe à un autre est une information forte qui dénie une réelle séparation de ces groupes. Pour de nombreuses applications cette hypothèse n'est pas entièrement acceptable. Par exemple, on a montré précédemment que les dissimilarités estimées à partir de marqueurs moléculaires pouvaient être entachées d'une certaine erreur. Si la population étudiée regroupe des unités proches génétiquement, à l'échelle infraspécifique par exemple, l'imprécision de mesure des dissimilarités peut suffire à expliquer le caractère erratique de certaines unités.



Si l'on considère que quelques individus seulement peuvent masquer une structure commune, il devient intéressant de chercher à identifier ces individus « fluctuants » plutôt que de leur laisser jouer un rôle équivalent à celui des autres dans la détermination de la structure commune. On est alors amené à reformuler le problème de la recherche des structures communes à plusieurs arbres. Le problème devient celui de la détermination du plus petit sous-ensemble qu'il suffit d'élaguer dans chacun des arbres pour obtenir des arbres identiques ou, inversement, du plus grand sous-ensemble d'individus ayant la même structure dans tous les arbres. Ces individus forment le sous-arbre commun. La figure 7 permet de comparer, sur un exemple simple, ce sous-arbre commun et l'arbre consensus.

Cette approche est beaucoup moins répandue que l'approche consensus et a fait l'objet de peu de travaux. La simplicité d'énoncé du problème dissimule un problème algorithmique relativement complexe. Récemment, KUBICKA *et al.* (1995) ont publié les bases d'un algorithme qui donne une solution exacte, tout en restant de complexité suffisamment faible pour être utilisable en pratique. L'algorithme repose sur l'énumération de toutes les solutions possibles mais en limitant sur un critère d'arrêt la profondeur d'exploration des branches d'où le gain important en complexité.

L'ordre  $o$  du sous-arbre commun, c'est-à-dire le nombre d'individus conservés, peut être considéré comme une mesure de la ressemblance entre arbres. L'ordre maximum est  $n$ , il est obtenu pour deux arbres identiques. A l'inverse la valeur minimale de l'ordre est 3 puisqu'il n'y a qu'une topologie possible de trois points distincts, elle est donc commune aux deux arbres. D'où une définition de dissimilarité entre arbres  $D = n - o$ , que l'on peut normaliser par  $n - 3$  pour garder des variations entre 0 et 1. L'utilisation pratique de ce critère nécessite de connaître la distribution de  $o$  sous l'hypothèse d'indépendance des arbres. Cette distribution peut être approchée par simulation et a été

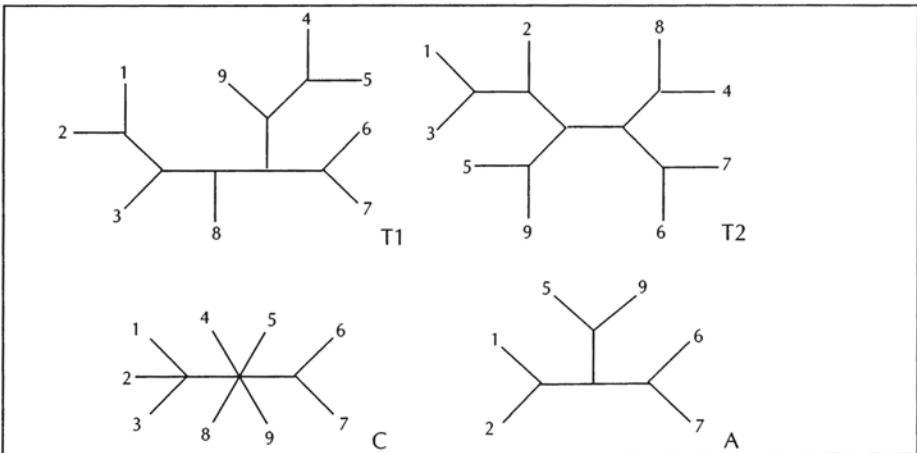


Figure 7. Arbre concensus (C) et sous-arbre commun (A) des arbres T1 et T2.

calculée pour des arbres binaires de 20 à 100 feuilles (tableau 7). Elle peut être utilisée pour fixer une règle d'interprétation. En fixant un seuil à 5 % par exemple, on considère que le hasard donnera moins de 5 fois sur 100 des arbres d'ordre supérieur ou égal à 10, 13, 16, 18 et 19 lorsque  $n$  varie de 20 à 100 par pas de 20.

Il faut retenir qu'arbre consensus et sous-arbre commun n'ont pas le même point de vue sur les données. Ils se justifient l'un et l'autre en fonction de l'interprétation que l'on fait des déplacements d'individus dans les arbres initiaux. S'il s'agit du déplacement accidentel de quelques individus au sein d'une structure commune forte alors le sous-arbre commun paraît le plus adapté. Si au contraire les déplacements traduisent réellement des différences de structure fondamentales alors l'arbre consensus est logiquement plus adapté. En pratique, il est rarement possible d'opter franchement pour une de ces deux hypothèses et les deux approches pourront être mises en œuvre de manière complémentaire.

Tableau 7. Distribution du nombre de feuilles des sous-arbres communs de 1 000 paires d'arbres binaires aléatoires à 20, 40, 60, 80 et 100 feuilles.

	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Moyenne
20	1	40	50	9											7,67
40				5	34	43	17	0	1						10,76
60						2	30	48	17	3					12,89
80								3	25	39	24	7	2		15,13
100									2	19	36	28	11	4	16,39

## Conclusion

Pour conclure ce chapitre, il faut souligner le dynamisme de ce secteur de recherche. On croyait, avec l'ouvrage de SNEATH et SOKAL (1973), que tout était dit, ou presque, en taxonomie numérique. Les récents développements théoriques, l'augmentation considérable des moyens informatiques, l'apparition de nouveaux types de marqueur ont complètement remis en question cette discipline, qui est actuellement un secteur de recherche très actif à l'interface entre les mathématiques et la biologie.

On aura compris à la lecture de ce chapitre la nécessaire intervention conjointe de ces deux domaines. La question naturelle qui surgit d'une telle présentation — certaines méthodes sont-elles meilleures que d'autres ? — n'a pas, dans ce domaine, de sens réel. Il n'existe pas de solution mathématique universelle, valable en toutes circonstances. Parmi la diversité des choix possibles, seule la connaissance de l'espèce végétale étudiée — niveau de diversité, mode de reproduction, hétérozygotie... —, de la nature de l'échantillon analysé, des caractéristiques des marqueurs utilisés permet de retenir les approches les plus pertinentes dans chaque contexte particulier.

## Références bibliographiques

- BARTHELEMY J.P., GUENOCHÉ A., 1988. Les arbres et les représentations des proximités. Paris, France, Masson, 239 p.
- BEAULIEU F.B., 1989. A classification of presence/absence based dissimilarity coefficients. *Journal of Classification*, 6 : 233-246.
- DARLU P., TASSY P., 1993. Reconstruction phylogénétique. Paris, France, Masson, 245 p.
- CRITCHLEY F., FICHET B., 1994. The partial order by inclusion of the principal classes of dissimilarity on a finite set and some of their basic properties. *In* : Classification and dissimilarity analysis, B. van Cussem éd., New York, Etats-Unis, Springer, Lecture Notes in Statistics, p. 5-65.
- ESCOFIER B., PAGES J., 1993. Analyses factorielles simples et multiples : objectifs, méthodes et interprétation. Paris, France, Dunod.
- ESCOUFIER Y., 1975. Le positionnement multidimensionnel. *Revue de statistique appliquée*, 23(4) : 5-14.
- FELSENSTEIN J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39(4) : 783-791.
- GASCUEL O., 1997. Concerning the NJ algorithm and its unweighted version, UNJ. *In* : Mathematical hierarchies and biology, DIMACS workshop. American Mathematical Society, Series in Discrete Mathematics and Theoretical Computer Science 37, p. 149-170.
- JUKES T.H., CANTOR C.R., 1969. Evolution of protein molecules. *In* : Mammalian protein metabolism, H.N. Munro éd., New York, Etats-Unis, Academic Press, p. 21-132.
- KUBICKA E., KUBICKI G., MCMORRIS F.R., 1995. An algorithm to find agreement subtrees. *Journal of Classification*, 12 : 91-99.
- LAVIT C., 1988. Analyse conjointe de tableaux quantitatifs. Paris, France, Masson.
- LEFORT-BUSSON M., DE VIENNE D., 1985. Les distances génétiques, estimations et applications. Paris, France, Inra, 181 p.
- PERRIER X., 1998. Analyse de la diversité génétique : mesures de dissimilarité et représentation arborée. Thèse de doctorat, université Montpellier II, Montpellier, France, 192 p.
- ROHLF F.J., 1987. Ntsys-pc numerical taxonomy and multivariate analysis system. New York, Etats-Unis, Applied Biostatistics Inc., Setauket.
- SAITOU N., NEI M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4) : 406-425.
- SAPORTA G., 1990. Probabilités, analyse de données et statistiques. Paris, France, Technip, 493 p.
- SATTATH S., TVERSKY A., 1977. Additive similarity trees. *Psychometrika*, 42(3) : 319-345.
- SNEATH P.H.A., SOKAL R.R., 1973. Numerical taxonomy. San Francisco, Etats-Unis, Freeman, 573 p.

# Une méthode de constitution de *core collections*

---

Michel Noirot, François Anthony,  
Stéphane Dussert, Serge Hamon

Il y a environ 10 000 ans, nos ancêtres découvraient l'agriculture et créaient de ce fait de nouvelles pressions de sélection sur les plantes sauvages : le processus de domestication était né. Au cours des millénaires qui ont suivi, ce processus a abouti à ce que l'on a coutume d'appeler les « variétés primitives ». A partir du XVI<sup>e</sup> siècle, avec le développement des migrations intercontinentales, la culture des plantes a très largement dépassé leur zone d'origine et de diversification. Enfin, le XX<sup>e</sup> siècle a été marqué par l'évolution rapide des techniques de sélection et des modes de culture et par la production de nouveaux idéotypes. C'est ainsi qu'une diversité considérable d'espèces et de formes a été engendrée.

VAVILOV (1935) a été parmi les premiers à démontrer l'importance des banques de gènes dans l'amélioration des plantes. Ce n'est cependant que quelques décennies plus tard qu'HARLAN (1970), FRANKEL et BENNETT (1970) puis PERNES (1984) ont encouragé la collecte et l'évaluation de la diversité des populations naturelles.

Les principales plantes cultivées, les espèces sauvages affines et les plantes cultivées mineures ont été collectées à l'initiative de l'IBPGR (International Board for Plant Genetic Resources). Ces prospections en routine ont entraîné rapidement des problèmes de gestion et de conservation des collections ainsi réunies.

A la fin des années 80, les collections sont devenues énormes et difficiles à régénérer et à maintenir. Cet accroissement considérable de la taille des collections, mais aussi l'insuffisance de la documentation disponible sur les échantillons ont souvent été cités comme une limite à l'utilisation effective des ressources génétiques (HOLDEN, 1984). Les utilisateurs potentiels ont besoin soit de populations représentatives de la diversité, soit d'accessions qui présentent des caractères agronomiques particuliers (résistance aux maladies, à la sécheresse...). Dans les deux cas les gestionnaires éprouvent des difficultés à répondre à leurs attentes.

FRANKEL et BROWN (1984) ont été les premiers à souligner la nécessité de constituer des collections de petite taille et de diversité maximale : les *core collections*. Pour la majorité des utilisateurs, constituer une *core collection* permet d'éviter la redondance de génotypes (doublons), souvent liée au mode de reproduction ou à la surreprésentation des variétés cultivées. Cette redondance est rare chez les espèces allogames, très fréquente chez les autogames et devient commune chez les espèces apomictiques ou chez les plantes à multiplication végétative.

En terme d'utilisation, les trois objectifs majeurs de la *core collection* sont de disposer d'une représentation aussi large que possible de la diversité génétique, de pouvoir réaliser des études approfondies sur un ensemble réduit de génotypes, d'espérer extrapoler les résultats ainsi obtenus pour faciliter la recherche de génotypes appropriés dans les collections de base.

Actuellement, pour constituer une *core collection*, la plupart des chercheurs s'accordent sur la nécessité d'une stratification préalable à l'échantillonnage. En d'autres termes, il s'agit de prendre en compte l'organisation de la variabilité en groupes et en sous-groupes (FRANKEL et BROWN, 1984 ; VAN HINTUM, 1995 ; YONEZAWA *et al.*, 1995).

En revanche, plusieurs démarches sont proposées pour l'échantillonnage au sein des groupes et des sous-groupes. L'une de ces démarches consiste à opérer un échantillonnage aléatoire dans chaque groupe préalablement défini (ou dans l'ensemble de la collection de base si l'organisation en groupes n'est pas connue ou n'existe pas). Ce type d'échantillonnage a l'avantage d'aboutir à une *core collection* statistiquement représentative de la collection de base.

Nous proposons dans ce chapitre une nouvelle méthode d'échantillonnage, la PCS (*principal component scoring*), dont le but est de maximiser la diversité échantillonnée. Cette diversité est mesurée en utilisant des variables quantitatives ou qualitatives, dont le choix est discuté. Nous examinons également les effets de la PCS sur la stratification de l'échantillonnage et sur la taille de l'échantillon. Nous présentons, enfin, quelques exemples de *core collections* réellement valorisées.

## Principes et méthodes

La diversité intrapopulation est déterminée par les différences entre individus pour un ou plusieurs caractères. Ces différences peuvent être estimées par une distance qui, dans notre cas, doit être une distance métrique. Le choix de cette distance dépend des caractères observés, quantitatifs ou qualitatifs.

### Les variables quantitatives

#### LE CHOIX DE LA DISTANCE, LA COLINÉARITÉ ET LA PONDÉRATION

Les caractères quantitatifs sont généralement hétérogènes. Ils correspondent à des longueurs (hauteur de la plante, diamètre de la tige, etc.), à des surfaces (surface foliaire, surface du stigmate, etc.), à des poids (biomasse aérienne, biomasse reproductive, etc.) ou à des temps (date de floraison, durée de fructification, etc.). Ils manifestent, de plus, différentes formes de variabilité. Afin de donner le même poids à chaque caractère  $j$ , la distance euclidienne est pondérée par l'inverse de l'écart type  $\sigma_j$ . La distance  $d_{ik}$  entre deux individus  $i$  et  $k$  pour les  $J$  caractères quantitatifs est définie par la formule suivante :

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

où  $x_{ij}$  est la valeur du caractère  $j$  observée sur l'individu  $i$  et  $x_{kj}$  la valeur du caractère  $j$  pour l'individu  $k$ .

La distance entre les individus est directement liée aux différences. Si ces différences proviennent de caractères fortement corrélés, positivement ou négativement, la distance entre certains individus est alors largement surestimée. Ainsi, si nous mesurons le diamètre du tronc d'un arbre à différentes hauteurs du sol (1 mètre, 1,10 mètre, 1,20 mètre, etc.), la distance euclidienne entre deux arbres sera fortement influencée par les différences de diamètre. Cet exemple est évident, mais un tel effet, appelé effet de colinéarité, se retrouve pour tous les caractères corrélés.

Pour éliminer les effets de colinéarité, l'analyse en composantes principales a été appliquée à des variables standardisées pour donner  $J$  nouvelles variables centrées, statistiquement indépendantes : les facteurs. La distance entre deux individus  $i$  et  $k$  pour les  $J$  facteurs est calculée en utilisant une formule similaire :

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

où la racine carrée de la valeur propre  $\lambda_j$  permet la pondération, et où  $z_{ij}$  et  $z_{kj}$  sont respectivement les coordonnées des individus  $i$  et  $k$  sur le facteur  $j$ .

Une telle procédure donne le même poids à tous les facteurs dans l'estimation de la distance, incluant aussi les composantes résiduelles, résultat du hasard ou des erreurs de notations. Les facteurs dont la valeur propre est inférieure à 1 (critère de Kaiser) sont éliminés afin d'éviter qu'ils n'interviennent dans le calcul de la distance.

#### LE CHOIX DES INDIVIDUS MAXIMISANT LA DIVERSITÉ

La somme des carrés généralisée (SCG) d'un lot de  $N$  individus dans l'espace factoriel des  $K$  variables standardisées (moyenne = 0 ; variance = 1) et indépendantes (coefficient de corrélation = 0) est égale au produit  $NK$  (LEBART *et al.*, 1977). La contribution  $P_i$  de l'individu  $i$  à la SCG est égale à la somme des carrés de ses  $K$  nouvelles coordonnées :

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

La contribution relative  $CR_i$  de l'individu  $i$  à la SCG de l'ensemble est donnée par :

$$CR_i = P_i / (NK).$$

Conserver la plus grande variabilité équivaut à maximiser le score du sous-ensemble d'individus échantillonnés en utilisant un estimateur SCG. La première étape consiste à prélever l'individu le plus éloigné du barycentre de l'ensemble, c'est-à-dire l'individu qui a la plus forte contribution relative. La sélection itérative des individus maximisant la diversité de la *core collection* augmente la taille de cette *core collection*. A chaque itération, la SCG cumulée de la *core collection*, exprimée en pourcentage de la SCG totale, est connue. La procédure peut être arrêtée soit selon la taille de la *core collection*, soit selon le pourcentage de diversité retenu. Les deux critères peuvent être pris simultanément en compte. Dans ce cas, le premier critère atteint provoque l'arrêt de l'échantillonnage.

### Les variables qualitatives

La méthode décrite ci-dessus est destinée aux données quantitatives. Les modifications nécessaires au traitement des données qualitatives portent sur les premières étapes de la PCS. Comme pour les données quantitatives, il existe des relations entre les variables. Par exemple, deux marqueurs moléculaires peuvent être liés génétiquement. Afin d'éliminer les effets de ce type de relation sur la distance et afin de donner le même poids à des variables indépendantes, une méthode d'analyse multivariée a été utilisée pour transformer les données initiales en coordonnées factorielles : il s'agit de l'analyse factorielle des correspondances (BENZECRI, 1972).

La distance du  $\chi^2$  est retenue à la place de la distance euclidienne et l'analyse utilise un tableau disjonctif complet. Dans ce tableau, la présence et l'absence d'un allèle sont considérées comme deux variables différentes prenant la valeur 1 et 0. Avec  $p$  marqueurs moléculaires observés sur  $N$  individus, nous obtenons un tableau  $2pN$ . Par conséquent, tous les individus montrent la même fréquence marginale égale à  $p$ . De plus, le terme  $\rho\lambda_i$  ( $\lambda_i$  est la valeur propre du facteur  $i$ ) est égal à la somme des rapports de corrélation de ce facteur avec les  $p$  variables. Ce terme est équivalent à la valeur propre observée dans l'analyse en composantes principales. La somme des  $\rho\lambda_i$  est égale au nombre de marqueurs (pour l'analyse en composantes principales sur données quantitatives, la somme des valeurs propres était égale au nombre de variables). Comme pour les données quantitatives, les coordonnées factorielles sont pondérées. Dans notre cas, les poids sont les racines carrées des  $\rho\lambda_i$  correspondants. Le critère de Kaiser pour le choix du nombre de facteurs est appliqué au terme  $\rho\lambda_i$ . Les étapes suivantes de la PCS sont les mêmes que pour les caractères quantitatifs.

## Discussion

### La stratification : conditions et conséquences

La méthode la plus simple pour créer une *core collection* est l'échantillonnage aléatoire sur l'ensemble de la collection de base. Lorsque la structure génétique de la collection de base est inconnue, un tel échantillonnage représente la meilleure solution (BROWN, 1989a). Néanmoins, son efficacité est faible pour les allèles communs localement mais rares dans l'ensemble de la collection. C'est pourquoi BROWN (1989b) a suggéré une stratification de l'échantillonnage.

PEETERS et MARTINELLI (1989), HOLBROOK et ANDERSON (1995), VAN HINTUM (1995) partagent cet avis, et suggèrent de prendre en compte, comme base de stratification, le pays d'origine du matériel végétal. PEETERS *et al.* (1993) conseillent d'utiliser des données écogéographiques précises. Celles-ci ont été prises en considération pour établir la *core collection* de soja (PERRY *et al.*, 1991). Les caractères agronomiques ou biologiques majeurs — mode de reproduction, durée du cycle... — sont également utilisés pour la stratification (SPAGNOLETTI-ZEULLI et QUALSET, 1987 ; HAMON et VAN SLOTEN, 1989 ; DIWAN *et al.*, 1994 ; HAMON *et al.*, 1995). Ces caractères sont des variables de nature à la fois quantitative (hauteur, diamètre...) et qualitative (couleur, aspect...), qu'il est difficile de traiter simultanément pour obtenir des matrices de distances, sauf si les données quantitatives sont recodées de manière à obtenir des



classes d'effectifs égaux ou d'amplitudes égales. Cependant, COLE-ROGERS *et al.* (1997) proposent une méthode originale, la *normed binary scale*, qui permet de calculer des matrices de distances intégrant ces deux types de variables.

Utilisés par LUX et HAMMER (1994), vivement conseillés par GEPTS (1995), les marqueurs moléculaires commencent seulement à être pris en compte pour la stratification. En effet, le pourcentage d'accessions évaluées à l'aide de ce type de marqueur est encore faible, les sélectionneurs y portent souvent peu d'intérêt et la structuration des ensembles déterminée par les marqueurs moléculaires ne coïncide pas toujours avec celle qui découle de la diversité morphoagronomique.

Cependant, l'échantillonnage aléatoire, même à l'intérieur des groupes, ne permet pas d'atteindre le premier objectif de la *core collection*, qui est d'échantillonner le maximum de diversité. A titre d'exemple, la production d'acide cyanhydrique chez le trèfle blanc (*Trifolium repens*) est contrôlée par deux locus indépendants. Ce caractère confère une résistance à plusieurs espèces d'insectes et de mollusques, et son expression est régulée par les paramètres climatiques (température, durée du jour, humidité). La collection de base du NPGS (National Plant Germplasm System) des Etats-Unis comporte 602 accessions de trèfle blanc. Une *core collection* de 91 accessions a par ailleurs été établie sur la base d'une stratification géographique et selon un choix aléatoire à l'intérieur des groupes. PEDERSON *et al.* (1996) ont déterminé la proportion de plantes cyanogéniques dans la collection de base et l'ont comparé à celle de la *core collection*. Aucune différence significative de fréquence n'a été trouvée entre les deux collections, ce qui prouve que la *core collection* n'avait pas « maximisé » la variabilité : elle était simplement une « image » en réduction de la collection de base.

En revanche, la PCS modifie le mode d'échantillonnage, qui n'est plus aléatoire, maximise la diversité et évite, dans la plupart des cas, les doublons. Cette méthode remplit ainsi les objectifs définis pour une *core collection*.

En prélevant préférentiellement les individus les plus éloignés, la méthode PCS requiert trois conditions pour être fonctionnelle et efficace. La première, et la plus importante, est que tous les individus de la *core collection* peuvent se croiser entre eux et donner des types intermédiaires. Cela implique une connaissance préalable de la structure génétique du complexe d'espèces (PERNES, 1984) et une bonne estimation du niveau des barrières reproductives entre les compartiments. La deuxième condition concerne l'efficacité de la PCS, qui suppose une additivité généralisée et une forte héritabilité (au sens large) pour les caractères quantitatifs. Dans les cas où cette hypothèse n'est pas valide (dominance, superdominance, haute plasticité, etc.), la sélection sur la diversité phénotypique ne mène pas nécessairement à la sélection sur la diversité génétique. L'échantillonnage peut être alors considéré comme aléatoire par rapport à la variabilité cachée. La troisième condition est l'absence d'une structuration polymodale de la collection de base. En effet,

l'existence de plusieurs groupes peut conduire à un prélèvement des individus alternativement dans les groupes les plus éloignés. Dans ce cas, maximiser la diversité augmente la redondance. La stratification de l'échantillonnage est ici une étape préliminaire déterminante, comme dans le cas de l'échantillonnage aléatoire.

Ainsi, pour la PCS, la stratification de l'échantillonnage doit dépendre de la structure génétique des populations et des limites à la recombinaison. Lorsque de telles données sont perdues ou absentes, la systématique devrait être prise en compte. Les informations bioclimatiques et biogéographiques devraient améliorer ensuite la structure avec l'établissement de sous-groupes correspondant à la différenciation génétique, en sous-espèces et en écotypes. La PCS est alors appliquée à l'intérieur de chaque groupe et sous-groupe.

## La taille de la *core collection* et de ses strates

La réflexion sur la taille de la *core collection* est toujours d'actualité. BROWN (1989a), en utilisant la théorie des allèles neutres (KIMURA et CROW, 1964) et celle de l'échantillonnage, a montré qu'un effectif de 10 % de la collection de base renferme au moins 80 % des allèles, avec un risque statistique d'erreur de 5 %. D'après cet auteur, les résultats sont fiables par rapport au type de distribution des fréquences des allèles à chaque locus. Cette valeur de 10 % n'est pas modifiée par notre mode d'échantillonnage.

La taille de chaque sous-groupe à l'intérieur de la *core collection* a été étudiée par BROWN (1989b). Trois méthodes ont été comparées pour déterminer cette taille : choisir le même nombre d'accessions par groupe, définir un nombre d'accessions proportionnel à l'effectif du groupe ou opter pour un nombre proportionnel au logarithme de l'effectif du groupe. L'auteur montre que la troisième solution constitue un bon compromis. Néanmoins, le choix de la taille du sous-échantillon d'après la taille du groupe suppose une relation entre la diversité du groupe et sa taille, ce qui est loin d'être toujours le cas dans les collections de base. Le rapport entre la diversité et la taille du groupe dépend, entre autres, du mode de reproduction et de l'importance économique de la plante (plante cultivée ou espèce affine). C'est ainsi, que, dans les collections de base du complexe d'espèces du genre *Coffea*, l'espèce cultivée et autogame *C. arabica* est surreprésentée par rapport à l'espèce sauvage et allogame d'Afrique de l'Est *C. sessiliflora* (NOIROT et al., 1993).

La méthode d'échantillonnage de la PCS permet de gérer la diversité de la *core collection*. L'arrêt des prélèvements d'individus pour alimenter la *core collection* peut se décider d'après le pourcentage de diversité déjà retenu. La prise en compte de la diversité déjà prélevée est particulièrement adaptée aux espèces à forte redondance naturelle (les complexes agamiques, les plantes autogames, les plantes à multiplication végétative).

## Les premières utilisations de *core collections*

Quelle que soit la stratégie utilisée, les *core collections* sont conçues pour aider les gestionnaires à conserver et à valoriser les ressources génétiques. Les deux exemples rapportés ici montrent que les utilisateurs peuvent y trouver leur intérêt.

L'USDA (United States Department of Agriculture), aux Etats-Unis, possédait en 1990 une collection de base de luzernes pérennes de 2 400 accessions. Afin d'en extraire une *core collection* de 200 accessions, BASIGALUP *et al.* (1995) ont retenu, parmi les huit méthodes testées, la sélection dirigée de génotypes après une stratification géographique. JUNG *et al.* (1997) ont par la suite utilisé cette *core collection* pour des recherches sur la composition protéique, la biodégradation des feuilles, la digestibilité et la composition en lignines. Ainsi, cette *core collection* s'est révélée utilisable pour d'autres caractères que ceux qui avaient déterminé sa constitution.

HOLBROOK *et al.* (1993) ont établi une *core collection* de 831 accessions d'arachide à partir de la collection de base, qui en compte 7 432. Cette collection comporte 70 % d'échantillons évalués sur des caractères morphoagronomiques et 30 % d'échantillons non évalués. Pour les accessions évaluées, les analyses multivariées ayant mis en évidence une structuration en groupes, 10 % des accessions de chaque groupe sont prélevées au hasard. Pour les accessions qui ne sont pas évaluées, 10 % sont prélevées aléatoirement après une stratification par pays. HOLBROOK et ANDERSON (1995) ont testé la pertinence de cette *core collection* par rapport à la collection de base pour la résistance à la cercosporiose due à *Cercosporidium personatum*. Il s'agissait de déterminer le nombre d'accessions résistantes que permettrait de détecter la *core collection* par rapport à la collection de base. La démarche comprenait deux étapes. Dans un premier temps, toute la *core collection* a été testée pour ce caractère, puis les groupes qui faisaient apparaître des résistances dans la *core collection* ont été examinés en détail dans la collection de base. Le taux d'efficacité, en terme de proportion d'accessions résistantes identifiées, passe alors de 1/64, pour la collection de base, à 1/8, pour la *core collection*. Ce résultat démontre l'utilité d'une *core collection* non seulement pour valoriser le matériel génétique, mais aussi pour améliorer l'efficacité de la recherche de caractéristiques particulières.

La méthode PCS de constitution de *core collections* permet de franchir un degré supplémentaire d'efficacité. Bien entendu, comme pour l'arachide, la mise en œuvre de cette démarche repose sur la disponibilité des données nécessaires aux sélections. HAMON *et al.* (1998) ont montré sur quatre plantes — le riz, les caféiers, le sorgho et l'hévéa — que la variabilité des caractères quantitatifs dans les *core collections* n'est que peu ou pas modifiée lorsque la sélection est qualitative. En revanche, les moyennes et les variances des caractères morphoagronomiques le sont fortement par une sélection quantitative. La sélection qualitative apparaît comme la plus efficace pour conserver les allèles

rare et augmenter la diversité globale avec des effets limités au niveau quantitatif. C'est ainsi que la sélection quantitative entraîne la perte de 6 % des allèles rares — de fréquence initiale dans la collection de base inférieure à 5 % — chez l'hévéa, de 11 %, pour le sorgho, de 12 %, pour le riz et de 33 %, pour le caféier. En utilisant la sélection qualitative les pertes sont réduites à 2 % pour le riz, le sorgho et l'hévéa et à 6 % pour le caféier. En d'autres termes, cette approche démontre qu'il est possible de maximiser la richesse allélique — autrement dit la variabilité neutre — dans la *core collection* tout en préservant la représentativité de la variabilité morphoagronomique.

Des *core collections* sont actuellement mises en place pour de nombreuses plantes. Un consensus semble se dégager sur leur taille (environ 10 % de la collection de base) et sur la nécessité d'une stratification. L'échantillonnage aléatoire conduit à conserver la variabilité contenue dans la collection de base tout en maintenant ses défauts (surreprésentation, redondance, biais d'échantillonnage...). Le principal avantage de la PCS est de permettre d'augmenter la diversité allélique neutre de la *core collection* sans modifier la représentativité agronomique ni changer l'intensité relative de l'échantillonnage (10 % de la collection de base). Les progrès actuels de la biologie moléculaire devraient faciliter l'utilisation des marqueurs moléculaires dans l'estimation et la structuration de la diversité génétique. La constitution de *core collections* devra alors prendre en compte les relations entre les différents niveaux de variabilité.

## Références bibliographiques

BASIGALUP D.H., BARNES D.K., STUCKER R.E., 1995. Development of a core collection for perennial *Medicago* plant introductions. *Crop Science*, 35 : 1163-1168.

BENZECRI J.P., 1972. Pratique de l'analyse des données : analyse des correspondances. Paris, France, Dunod, 424 p.

BROWN A.D.H., 1989a. Size and structure of collection: the case for core collection. *In* : The use of plant genetic resources, T. Hodgkin *et al.* éd., Chichester, Royaume-Uni, Wiley, p. 136-156.

BROWN A.D.H., 1989b. Core collections: a practical approach to genetic resources management. *Genome*, 31 : 818-824.

COLE-ROGERS P., SMITH D.W., BOSLAND P.W., 1997. A novel statistical approach to analyze genetic resources evaluations using *Capsicum* as an example. *Crop Science*, 37 : 1000-1002.

DIWAN N., BAUCHAN G.R., MCINTOSH M.S., 1994. A core collection for the United States annual *Medicago* germplasm collection. *Crop Science*, 34 : 279-285.

FRANKEL O.H., BENNETT E., 1970. Genetic resources. *In* : Genetic resources in plants, their exploration and conservation, O.H. Frankel et E. Bennett éd., Oxford, Royaume-Uni, Blackwell, 547 p.

FRANKEL O.H., BROWN A.H.D., 1984. Current plant genetic resources: a critical appraisal. *In* : Genetics, new frontiers (volume IV). New Delhi, Inde, Oxford and IBH.

GEPTS P., 1995. Genetic markers and core collections. *In* : Core collections of plant genetic resources, T. Hodgkin *et al.* éd., Chichester, Royaume-Uni, Wiley, p. 127-146.

HAMON S., DUSSERT J., DEU M., HAMON P., SEGUIN M., GLASZMANN J.C., GRIVET L., CHANTEREAU J., CHEVALLIER M.H., FLORI A., LASHERMES P., LEGNATE H., NOIROT M., 1998. Effects of quantitative and qualitative principal component score strategies on the structure of coffee, rice, rubber tree and sorghum core collections. *Genetics, Selection, Evolution*, 30 (suppl. 1) : 237-258.

HAMON S., NOIROT M., ANTHONY F., 1995. Developing a coffee core collection using the principal components score strategy with quantitative data. *In* : Core collections of plant genetic resources, T. Hodgkin *et al.* éd., Chichester, Royaume-Uni, Wiley, p. 117-126.

HAMON S., VAN SLOTEN D.H., 1989. Characterization and evaluation of okra. *In* : The use of plant genetic resources, A.D.H. Brown et O. Frankel éd., Cambridge, Royaume-Uni, Cambridge University Press, p. 173-196.

HARLAN J.R., 1970. The evolution of cultivated plants. *In* : Genetic resources in plants, their exploration and conservation, O.H. Frankel et E. Bennett éd., Oxford, Royaume-Uni, Blackwell, 547 p.

VAN HINTUM T.J.L., 1995. Hierarchical approaches to the analysis of genetic diversity of crop plants. *In* : Core collections of plant genetic resources, T. Hodgkin *et al.* éd., Chichester, Royaume-Uni, Wiley, p. 23-34.

HOLBROOK C.C., ANDERSON W.F., 1995. Evaluation of a core collection to identify resistance to late leafspot peanut. *Crop Science*, 35 : 1700-1702.

HOLBROOK C.C., ANDERSON W.F., PITTMAN R.N., 1993. Selection of a core collection from the US germplasm collection of peanut. *Crop Science*, 33 : 859-861.

HOLDEN J.H.W., 1984. The second ten years. *In* : Crop genetic resources: conservation and evaluation, J.H.W. Holden et J.T. Williams éd., Londres, Royaume-Uni, George Allen and Unwin, 296 p.

JUNG H.G., SHEAFFER C.C., BARNES D.K., HALGERSON J.L., 1997. Forage quality variation in the US alfalfa core collection. *Crop Science*, 37 : 1361-1366.

KIMURA M., CROW J.F., 1964. The number of alleles that can be maintained in a finite population. *Genetics*, 49 : 725-738.

LEBART L., MORINEAU A., TABARD N., 1977. Techniques de la description statistique : méthodes et logiciels pour l'analyse des grands tableaux. Paris, France, Dunod.

LUX H., HAMMER K., 1994. Molecular markers and genetic diversity: some experience from the genebank. *In* : EUCARPIA meeting on evaluation and exploitation of genetic resources pre-breeding. Clermont-Ferrand, France, EUCARPIA, p. 49-53.

NOIROT M., HAMON S., ANTHONY F., 1993. L'obtention d'une *core collection* de caféiers : définition des groupes d'échantillonnage et méthodologie. In : XVI<sup>e</sup> Colloque scientifique international sur le café. Paris, France, ASIC.

PEDERSON G.A., FAIRBROTHER T.E., GREENE S.L., 1996. Cyanogenesis and climatic relationships in the US white clover germplasm and core subset. *Crop Science*, 36 : 427-433.

PEETERS J.P., MARTINELLI J.A., 1989. Hierarchical cluster analysis as a tool to manage variation in germplasm collections. *Theoretical and Applied Genetics*, 78 : 42-48.

PEETERS J.P., WILKES H.G., GALWEY N.W., 1993. The use of ecogeographical data in the exploitation of variation from gene bank. *Theoretical and Applied Genetics*, 80 : 110-112.

PERNES J., 1984. Gestion des ressources génétiques des plantes. Paris, France, ACCT, 212 p.

PERRY M.C., MCINTOSH M.S., STONER A.K., 1991. Geographical patterns of variation in the USDA soybean germplasm collection. 2. Allozyme frequencies. *Crop Science*, 31 : 1356-1360.

SPAGNOLETTI-ZEULI P.L., QUALSET C.O., 1987. Geographical diversity for quantitative spike characters in a world collection of durum wheat. *Crop Science*, 27 : 235-241.

VAVILOV N.I., 1935. The origin, variation, immunity and breeding of cultivated plants. *Chronica Botanica*, 13 (6 volumes).

YONEZAWA K., NOMURA T., MORISHIMA H., 1995. Sampling strategies for use in stratified germplasm collections. In : *Core collections of plant genetic resources*, T. Hodgkin *et al.* éd., Chichester, Royaume-Uni, Wiley, p. 35-53.



# Les agrumes

---

Patrick Ollitrault, Camille Jacquemond,  
Cécile Dubois, François Luro

Les agrumes constituent la première production fruitière mondiale avec 89,6 millions de tonnes, dont près de 60 millions de tonnes d'oranges en 1996 (FAO, 1997). Le volume de fruits transformés est en augmentation : les jus d'orange concentrés et congelés absorbent une grande partie de la production des Etats-Unis et du Brésil.

Domestiqués dans le Sud-Est asiatique il y a plusieurs millénaires, les agrumes ont été diffusés dans le monde entier (figure 1). Le cédratier a été la première espèce cultivée dans le bassin méditerranéen, quelques siècles avant notre ère, tandis que les autres espèces n'y ont été introduites qu'au cours du deuxième millénaire. Les agrumes ont gagné l'Amérique à la suite de la découverte du Nouveau Monde au xv<sup>e</sup> siècle. L'aire agrumicole est aujourd'hui très étendue, elle se situe approximativement entre 40° de latitude nord et de latitude sud.

La culture des agrumes est confrontée à des contraintes biotiques et abiotiques croissantes dans les principales régions de production. La *tristeza* — maladie de dégénérescence provoquée par le *citrus tristeza virus* —, *Phytophthora* sp. et les nématodes se rencontrent aujourd'hui dans la quasi-totalité de l'aire de culture. D'autres contraintes ont, en revanche, un caractère régional : le froid et le *blight* — maladie de dégénérescence d'origine encore indéterminée —, aux Etats-Unis, la *citrus variegated chlorosis* due à *Xilela fastidiosa*, au Brésil,



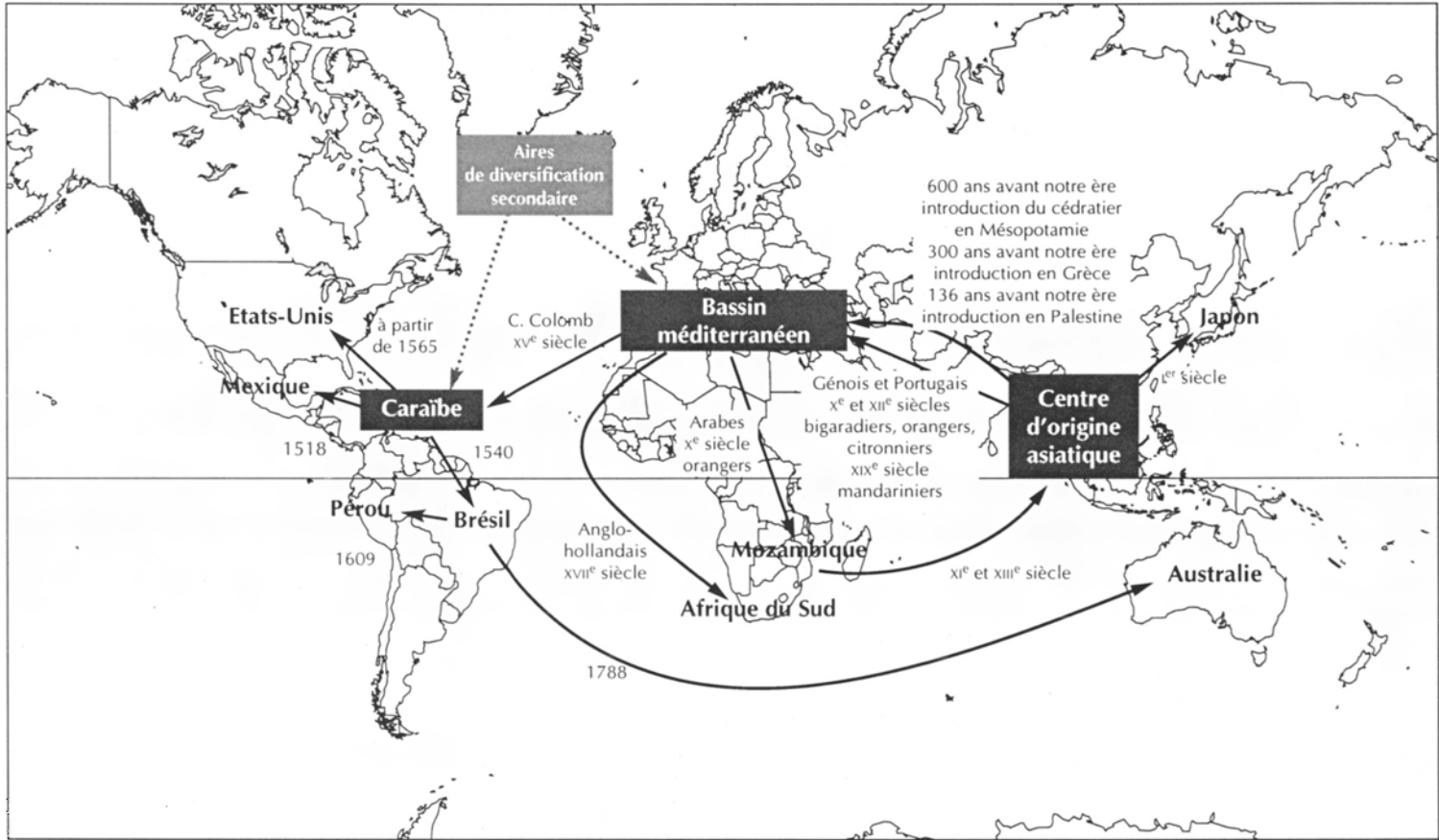


Figure 1. Région d'origine, dispersion et zones de diversification des agrumes cultivés.

la cercosporiose provoquée par *Phaeoramularia angolensis*, en Afrique, le *greening*, ou *citrus huanglongbing*, en Asie. Parmi les contraintes abiotiques, la salinité et les sols calcaires sont les problèmes majeurs du bassin méditerranéen. La pratique quasi généralisée de la culture de plants greffés permet, dans une certaine mesure, de répondre aux contraintes liées aux sols (calcaire, salinité, parasitisme tellurique) et à la *tristeza* grâce aux porte-greffe. La sélection des scions repose quant à elle sur des aspects qualitatifs et, dans certains pays, sur des caractères de tolérance à la *citrus variegated chlorosis* ou à la cercosporiose (OLLITRAULT et LURO, 1997).

## La botanique et les ressources génétiques

### La botanique et la taxonomie

L'apomixie partielle par embryogenèse nucellaire, associée à une large compatibilité sexuelle, a conduit à la production de populations clonales d'hybrides interspécifiques, qui ont souvent été assimilées à de nouvelles espèces par les taxonomistes. Les classifications botaniques sont ainsi généralement complexes. TANAKA (1961) identifie 156 espèces tandis que SWINGLE et REECE (1967) n'en distinguent que 16. La correspondance entre ces deux classifications et les noms communs est donnée dans le tableau 1 pour les taxons étudiés dans ce chapitre. Chez tous les agrumes et genres apparentés, le nombre de chromosomes de base ( $n$ ) est égal à 9 (KRUG, 1943). La quasi-totalité des agrumes est diploïde et seuls quelques polyploïdes naturels ont été identifiés, comme *Fortunella hindsii* ou la lime Tahiti.

### Les ressources génétiques

De nombreuses collections d'agrumes existent à travers le monde. Elles ont deux vocations, souvent divergentes quant au choix du matériel végétal à conserver : préserver à long terme la diversité des agrumes et des genres apparentés ; constituer des parcs à bois pour fournir des greffons de variétés commerciales. La collection de l'Okitsu Branch (Fruit Tree Research Station), au Japon, est la plus importante pour le matériel cultivé des zones d'origine, tandis que le conservatoire de l'université de Malaisie est remarquable par sa collection d'*Aurantioideae* du Sud-Est asiatique. Les collections de l'USDA (United States Department of Agriculture) et de l'université de Californie, aux Etats-Unis, celles de l'IVIA (Instituto Valenciano de Investigaciones Agrarias), en Espagne, et de l'université d'Adana, en Turquie, renferment certaines rutacées apparentées aux agrumes mais sont surtout régulièrement alimentées par

Tableau 1. Accessions de *Citrus* analysées et caractéristiques génétiques.

Code	Cultivar	Nom d'espèce		Marqueur morpho.	iso.	Taille du génome (pg/2C)	Génotype enzymatique										
		SWINGLE et REECE (1967)	TANAKA (1961)				ADH-1 AAT	IDH LAP	MDH-1 MDH-2	PGI PGM-1	PGM-2 PER	SKDH					
• Mandariniers (M)																	
Mks	King of Siam	<i>reticulata hybrid</i>	<i>nobilis</i>	1	1	0,760	22	11	33	44	33	22	33	34	12	12	33
Msw*	Satsuma Wase	<i>reticulata</i>	<i>unshiu</i>	1	1	0,737	22	11	33	34	33	22	23	33	22	22	33
Mso	Satsuma Owari	<i>reticulata</i>	<i>unshiu</i>	1	1	—											
Mda	Dancy	<i>reticulata</i>	<i>tangerina</i>	1	1	0,736	22	11	33	44	33	22	33	33	22	12	33
Mte	Temple	<i>reticulata hybrid</i>	<i>temple</i>	1	1	0,748	22	11	33	24	33	22	23	23	22	22	33
Mcl	Cléopâtre	<i>reticulata</i>	<i>reshni</i>	1	1	0,733	22	11	33	44	33	22	33	22	22	22	33
Mpo	Ponkan	<i>reticulata</i>	<i>reticulata</i>	1	1	0,744	22	11	33	44	33	22	33	33	22	11	22
Mco	Commun	<i>reticulata</i>	<i>deliciosa</i>	1	1	0,730	22	11	33	45	33	22	34	23	22	12	23
M63	Clémentinier SRA63	<i>reticulata</i>	<i>clementina</i>	1	1	0,750	22	11	23	24	33	22	24	33	22	22	33
Mmu	Murcott	<i>reticulata hybrid</i>		1	1	0,746	22	11	33	44	33	22	33	33	22	22	33
• Pamplémoussiens (P)																	
Pme	Menara	<i>grandis</i>	sp.	1	1	0,751	22	12	33	22	33	22	22	13	11	12	23
Prk	Reinking	<i>grandis</i>	<i>maxima</i>	1	1	0,774	22	22	33	45	33	22	22	44	11	11	12
Pkp	Kao Pan	<i>grandis</i>	<i>maxima</i>	1	1	0,767	22	12	23	35	33	22	22	13	11	11	12
Psn	Sunshine	<i>grandis</i>	<i>maxima</i>	1	1	0,794	22	22	23	25	33	12	22	33	11	11	11
Ppi	Pink	<i>grandis</i>	<i>maxima</i>	1	1	0,779	22	11	33	55	33	22	22	13	11	11	11
Psp	Sans pépins	<i>grandis</i>	<i>maxima</i>	1	1	0,787	22	12	33	25	33	22	23	11	11	11	12
Pin	Inde	<i>grandis</i>	<i>maxima</i>	1	1	0,787	22	22	22	55	33	22	22	33	11	11	12
Pah	Tahiti	<i>grandis</i>	<i>maxima</i>	0	1	—	22	12	33	25	33	22	22	11	11	11	11
Pph	Philippines	<i>grandis</i>	<i>maxima</i>	0	1	—	22	22	33	55	33	22	22	11	11	11	11
Psu	Surinam	<i>grandis</i>	<i>maxima</i>	0	1	—	22	12	23	35	33	22	22	13	11	11	11
Pei	Eingedi	<i>grandis</i>	<i>maxima</i>	1	0	0,763	—	—	—	—	—	—	—	—	—	—	—
Pch	Chandler	<i>grandis</i>	<i>maxima</i>	1	0	0,764	—	—	—	—	—	—	—	—	—	—	—
• Limettiers (L)																	
Lbs	Brazil Sweet	<i>aurantifolia</i>	<i>limettioides</i>	1	1	0,756	22	12	33	36	13	22	23	22	12	11	12
Lga	Gallet	<i>aurantifolia</i>	<i>aurantifolia</i>	1	1	0,787	12	12	13	36	13	12	22	22	22	11	12
Lta	Tahiti	<i>aurantifolia</i>	<i>latifolia</i>	1	1	1,170											
Lme	Mexicain	<i>aurantifolia</i>	<i>aurantifolia</i>	1	1	0,779											

Tableau 1. Suite.

Code	Cultivar	Nom d'espèce		Taille du génome (pg/2C)		Génotype enzymatique											
		SWINGLE et REECE (1967)	TANAKA (1961)			Marqueur morpho. iso.		ADH-1 AAT		IDH	MDH-1 LAP		MDH-2	PGI PGM-1		PGM-2 PER	SKDH
Lel	Elkseur	<i>aurantifolia</i>	<i>latifolia</i>	1	1	1,170											
Lbe	Bears	<i>aurantifolia</i>	<i>latifolia</i>	1	1	1,170	22	22	13	36	13	12	22	23	12	11	22
Lca	Calédonie	<i>aurantifolia</i>	<i>aurantifolia</i>	1	1	0,784											
Lki	Kirk	<i>aurantifolia</i>	<i>aurantifolia</i>	1	1	0,779											
Lra	Rangpur	<i>aurantifolia</i>	<i>limonia</i>	1	1	0,772	22	12	13	36	13	22	23	22	22	11	13
Lka	Kanghzi	<i>aurantifolia</i>	<i>aurantifolia</i>	0	1	—	22	22	12	36	13	22	22	22	12	11	22
Lsr	IAC SRA618	<i>aurantifolia</i>	<i>aurantifolia</i>	1	0	1,170	—	—	—	—	—	—	—	—	—	—	—
• Citronniers (C)																	
Cme	Meyer	<i>limon</i>	<i>meyeri</i>	1	1	0,772	22	12	23	46	13	22	23	23	12	12	12
Cfi	Fino	<i>limon</i>	<i>limon</i>	1	1	0,784											
Cve	Verna	<i>limon</i>	<i>limon</i>	1	1	—											
Cad	Adamapoulos	<i>limon</i>	<i>limon</i>	1	1	0,769											
Cdx	Doux	<i>limon</i>	<i>limon</i>	1	1	0,778											
Cli	Lisbon	<i>limon</i>	<i>limon</i>	1	1	0,786	22	12	13	46	13	22	24	23	12	12	12
Cvi	Villafranca	<i>limon</i>	<i>limon</i>	1	1	0,776											
Cmo	Monachello	<i>limon</i>	<i>limon</i>	1	1	0,787											
Ceu	Euréka	<i>limon</i>	<i>limon</i>	1	1	0,777											
Clu	Lunari	<i>limon</i>	<i>limon</i>	0	1	—											
Cst	Santa Teresa	<i>limon</i>	<i>limon</i>	1	0	0,786	—	—	—	—	—	—	—	—	—	—	—
• Orangers (O)																	
Owa	Washington Navel	<i>sinensis</i>	<i>sinensis</i>	1	1	0,757											
Odf	Double Fine	<i>sinensis</i>	<i>sinensis</i>	1	1	0,778											
Ota	Tarocco	<i>sinensis</i>	<i>sinensis</i>	1	1	0,772											
Onh	New Hall	<i>sinensis</i>	<i>sinensis</i>	1	1	0,778											
Ona	Navelina	<i>sinensis</i>	<i>sinensis</i>	1	1	0,755	22	11	23	24	33	22	23	33	12	22	12
Oha	Hamlin	<i>sinensis</i>	<i>sinensis</i>	1	1	0,749											
Osh	Shamouti	<i>sinensis</i>	<i>sinensis</i>	1	1	0,756											
Opb	Parson Brown	<i>sinensis</i>	<i>sinensis</i>	1	1	0,756											
Oca	Cadenera	<i>sinensis</i>	<i>sinensis</i>	1	1	0,751											

Tableau 1. Suite.

Code	Cultivar	Nom d'espèce		Marqueur morpho.	Taille du génome (pg/2C)	Génotype enzymatique											
		SWINGLE et REECE (1967)	TANAKA (1961)			iso.	ADH-1 AAT	IDH LAP	MDH-1 MDH-2	PGI PGM-1	PGM-2 PER	SKDH					
Ovl	Valencia Late	<i>sinensis</i>	<i>sinensis</i>	1	1	0,757											
• Bigaradiers et bouquetiers (B)																	
Bfe	Ferrando	<i>aurantium</i>	<i>aurantium</i>	1	1	0,755											
Bfl	Floride	<i>aurantium</i>	<i>aurantium</i>	1	1	0,755											
Bse	Sans épines	<i>aurantium</i>	<i>aurantium</i>	1	1	0,779											
Bma	Maroc	<i>aurantium</i>	<i>aurantium</i>	1	1	0,750											
Bqn	Nice (bouquetier)	<i>aurantium</i>	<i>aurantium</i>	1	1	0,756											
Bqf	Fleurs (bouquetier)	<i>aurantium</i>	<i>aurantium</i>	1	1	0,750	22	11	33	44	33	22	24	13	12	12	22
Bbs	Brazil Sour	<i>aurantium</i>	<i>aurantium</i>	1	1	—											
Bdd	Dai Dai	<i>aurantium</i>	<i>aurantium</i>	1	1	0,756											
Btu	Tuléar	<i>aurantium</i>	<i>aurantium</i>	1	1	—											
Bav	Avanito	<i>aurantium</i>	<i>aurantium</i>	0	1	—											
Bgr	Granito	<i>aurantium</i>	<i>aurantium</i>	1	0	0,753	—	—	—	—	—	—	—	—	—	—	
• Cédraiers (K)																	
Kdc	Corse	<i>medica</i>	<i>medica</i>	1	1	0,814											
Ket	Etrog	<i>medica</i>	<i>limonimedica</i>	1	1	0,821											
Kde	Digité	<i>medica</i>	<i>medica</i>	1	1	0,815	22	22	22	66	11	22	22	22	11	22	
Kpc	Poncire	<i>medica</i>	<i>medica</i>	1	1	0,807											
Kdi	Diamante	<i>medica</i>	<i>medica</i>	1	0	—	—	—	—	—	—	—	—	—	—	—	
• Pomelos (G)																	
Gsh	Shambar	<i>paradisi</i>	<i>paradisi</i>	1	1	0,749											
Gce	Cecily	<i>paradisi</i>	<i>paradisi</i>	1	1	0,778											
Gal	Alanoek	<i>paradisi</i>	<i>paradisi</i>	1	1	0,759											
Gre	Reed	<i>paradisi</i>	<i>paradisi</i>	1	1	0,772											
Gsr	Star Ruby	<i>paradisi</i>	<i>paradisi</i>	1	1	0,772	22	12	33	25	33	22	22	13	11	12	23
Grb	Red Blush	<i>paradisi</i>	<i>paradisi</i>	1	1	0,788											
Glr	Little River	<i>paradisi</i>	<i>paradisi</i>	1	1	0,784											
Gth	Thomson	<i>paradisi</i>	<i>paradisi</i>	1	1	0,784											

Tableau 1. Suite.

Code	Cultivar	Nom d'espèce		Marqueur		Taille du génome (pg/2C)	Génotype enzymatique										
		SWINGLE et REECE (1967)	TANAKA (1961)	morpho.	iso.		ADH-1 AAT	IDH LAP	MDH-1 MDH-2	PGI PGM-1	PGM-2 PER	SKDH					
Gma	Marsh	<i>paradisi</i>	<i>paradisi</i>	1	1	0,783											
Gru	Ruby	<i>paradisi</i>	<i>paradisi</i>	1	0	0,781	—	—	—	—	—	—	—	—	—	—	
• Autres Citrus																	
ROL	<i>rough lemon</i>	<i>limon</i>	<i>jambhiri</i>	1	1	0,777	12	12	23	46	13	22	23	22	11	22	
PEC	<i>pectinifera</i>	<i>reticulata hybrid</i>	<i>depressa</i>	1	1	0,751	22	11	33	24	23	22	33	33	22	12	33
JUN	—	<i>ichang austera</i>	<i>junos</i>	1	1	0,810	22	12	33	24	23	22	33	13	12	11	22
GUL	—	<i>maxima</i>	<i>pseudogulgul</i>	1	1	0,745	22	12	33	24	33	22	23	11	11	11	22
ICH	<i>ichangensis lemon</i>	<i>ichangensis</i>	<i>ichangensis</i>	1	1	0,774	22	11	22	44	23	22	23	34	12	11	22
BGM	bergamotier	<i>aurantifolia</i>	<i>bergamia</i>	1	1	0,771	22	12	13	44	13	22	24	13	12	11	12
PDC	poire du commandeur	<i>limon</i>	<i>lumia</i>	1	1	—	22	12	33	24	33	22	22	34	12	12	22
COM	<i>combava</i>	<i>hystrix</i>	<i>hystrix</i>	1	1	0,803	22	12	33	14	12	22	22	33	11	11	12
INT	—	<i>paradisi</i>	<i>intermedia</i>	1	1	0,764	22	12	33	23	33	22	23	23	11	12	22
MAC	—	<i>aurantifolia</i>	<i>macrophylla</i>	1	1	0,798	22	22	23	44	12	22	12	23	12	11	22
PEN	—	<i>aurantifolia</i>	<i>pennivesiculata</i>	1	1	0,813	22	22	13	22	11	12	22	23	12	11	12
EXE	—	<i>aurantifolia</i>	<i>excelsa</i>	1	1	0,793	22	22	23	44	33	12	12	23	22	11	22
SIA	siamelo	<i>hybrid</i>	<i>hybrid</i>	1	1	0,745	22	11	33	24	33	22	23	33	12	11	12
KPA	<i>khasi papeda</i>	<i>latipes</i>	<i>latipes</i>	1	1	0,780	22	12	34	44	23	23	12	34	12	11	12
HAL	—	<i>halimii</i>	<i>halimii</i>	1	1	0,778	22	22	22	44	22	23	12	34	22	11	22
VOL	—	<i>limon</i>	<i>limonia</i>	1	1	0,764	12	12	13	46	13	22	23	22	22	11	12
NAS	<i>nasnaran</i>	<i>reticulata hybrid</i>	<i>amblycarpa</i>	0	1	—	12	12	33	44	33	22	13	33	11	11	12

\* Les codes notés en caractères gras représentent le type enzymatique commun dans les analyses. Tous les individus d'une même zone grisée présentent le même profil enzymatique.

les variétés nouvellement sélectionnées dans le monde. La station de l'Inra et du Cirad de San Giuliano, en France, possède un statut unique du fait du contexte phytosanitaire favorable de la Corse. Elle abrite une importante collection de matériel végétal sain, qui inclut de nombreuses accessions d'Asie du Sud-Est, et peut être évaluée en plein champ. Le logiciel de gestion de bases de données Egid, élaboré par le Cirad et l'Inra (COTTIN *et al.*, 1995) à partir des descripteurs de l'IPGRI (International Plant Genetic Resources Institute), a par ailleurs été adopté par la FAO (Food and Agriculture Organization of the United Nations) pour mettre en place un réseau global de gestion des ressources génétiques des agrumes.

## L'organisation de la diversité

### La variabilité agromorphologique

La variabilité agromorphologique des agrumes est considérable. Elle concerne aussi bien les caractères pomologiques et organoleptiques que les résistances aux facteurs biotiques et abiotiques. Le genre *Citrus* renferme ainsi de nombreuses sources de tolérance aux contraintes biotiques et abiotiques, qui ouvrent des perspectives intéressantes pour l'utilisation des ressources génétiques en amélioration variétale.

Pour les facteurs abiotiques on peut citer : la tolérance au froid des mandariniers Satsuma ; la tolérance à la salinité du limettier Rangpur et du mandarinier Cléopâtre ; la tolérance aux sols calcaires de *C. jambhiri*, de *C. macrophylla*, de *C. volkameriana*, de *C. amblycarpa* et du bigaradier ; la tolérance à la sécheresse du limettier Rangpur. Des tolérances aux principaux ravageurs et aux maladies sont également identifiées : la tolérance à *Phytophthora* sp. des pamplemoussiers, des bigaradiers, de *C. volkameriana* et de *C. amblycarpa* ; la tolérance à la cercosporiose africaine des agrumes des pamplemoussiers, des citronniers et des mandariniers Satsuma et Beauty ; la tolérance à la *tristeza* du mandarinier Cléopâtre, de *C. amblycarpa*, du limettier Rangpur, de *C. jambhiri* et de *C. volkameriana* ; la tolérance au *blight* de l'oranger ; la tolérance au *greening* des pamplemoussiers et de certains mandariniers originaires de la zone tropicale ; la tolérance au chancre citrique dû à *Xanthomonas campestris* de *C. junos* et de certains mandariniers (Satsuma, Dancy...) ; la résistance aux acarions phytophages du pomelo Marsh et des mandariniers Satsuma et Dancy. Au vu des exemples précédents, il ne semble pas exister de liens entre la répartition des sources de résistance aux facteurs biotiques et la structuration spécifique du genre *Citrus*.

A l'inverse, la variabilité morphophysiological est fortement marquée entre les espèces, même si certains caractères sélectionnés par l'homme présentent une

forte diversité intraspécifique (précocité, calibre, coloration des fruits). A titre d'exemple, au sein du genre *Citrus*, le diamètre des fruits varie de quelques centimètres pour certains mandariniers et limettiers à plus de 30 centimètres pour certains pamplemoussiers. L'albédo est quasi inexistant chez les mandariniers mais constitue l'essentiel du fruit chez les cédratiers. La pulpe du fruit est verte, orange, jaune ou rouge, son acidité est nulle pour certaines oranges douces et très forte pour les limes et les citrons. Si les feuilles de toutes les espèces du genre *Citrus* sont monofoliées, leur taille et leur forme ainsi que la morphogénèse des arbres varient considérablement suivant les espèces.

Une étude plus fine de la structuration de la diversité morphologique dans le genre *Citrus* a été réalisée à partir de 20 descripteurs de l'appareil végétatif observés sur 74 cultivars. Elle sert de support à l'analyse des relations entre la diversité morphologique et la diversité moléculaire présentée dans ce chapitre.

## La variabilité biochimique et moléculaire

Les huiles essentielles et les polyphénols ont été les premiers marqueurs utilisés pour caractériser les variétés (TATUM *et al.*, 1974) et pour étudier la phylogénie des agrumes (SCORA, 1988). Les isoenzymes sont employées en routine pour identifier les plants de semis zygotiques et nucellaires (SOOST *et al.*, 1980 ; KHAN et ROOSE, 1988 ; OLLITRAULT *et al.*, 1992). Elles ont également permis de préciser les relations phylogéniques entre les espèces (TORRES *et al.*, 1982 ; HIRAI *et al.*, 1986 ; OLLITRAULT et FAURE 1992 ; HERRERO *et al.*, 1996 ; 1997). Les techniques d'analyse directe du polymorphisme de l'ADN — RFLP, RAPD, VNTR — ont principalement été appliquées dans des programmes de cartographie du génome (DURHAM *et al.*, 1992 ; JARREL *et al.*, 1992 ; LURO *et al.*, 1994b) ou de caractérisation variétale et de taxonomie (LURO *et al.*, 1994a ; LURO *et al.*, 1995 ; FANG et ROOSE, 1996). Toutefois, le déterminisme allélique de ces marqueurs est parfois difficile à clarifier, ce qui limite leur utilisation pour des études de génétique des populations s'intéressant à l'hétérozygotie et aux index de fixation ou de déséquilibre gamétique.

Des études cytogénétiques et des analyses par cytométrie en flux ont par ailleurs démontré l'existence de fortes variations entre les espèces quant à la taille des chromosomes (NAIR et RANDHAWA, 1969 ; OLLITRAULT *et al.*, 1994). Elles ont également mis en évidence de nombreux cas d'hétérozygotie structurale (RAGHUVANSHI, 1969 ; GMITTER *et al.*, 1992 ; GUERRA, 1993 ; MIRANDA *et al.*, 1997). Ces éléments sur la structure des génomes des différents taxons sont déterminants pour analyser l'organisation de la diversité allélique en termes évolutifs.

Afin d'étudier les paramètres de structure des populations, l'analyse de la diversité allélique présentée dans ce chapitre repose sur le polymorphisme de 9 systèmes isoenzymatiques. La diversité structurale nucléaire est pour sa part examinée en évaluant la taille des génomes par cytométrie en flux. L'échan-



tillonnage variétal des formes cultivées est le même que pour l'étude de la diversité morphologique. Dix-sept *Citrus* non cultivés complètent l'analyse.

#### LA DIVERSITÉ ISOENZYMATIQUE

Trente-cinq allèles ont été identifiés pour 11 locus polymorphes. Seuls 5 de ces allèles ne sont pas observés chez les cultivars. L'allèle nul du locus *LAP* (*LAP-6*), identifié à l'état homozygote chez les cédratiers, a pu être mis en évidence à l'état hétérozygote chez un certain nombre d'agrumes acides (citronniers, limes...) par l'examen d'hybrides contrôlés. De nombreux cultivars d'une même espèce peuvent présenter des profils identiques. C'est en particulier le cas pour les orangers, les bigaradiers, les pomelos et les citronniers. Les 74 cultivars sont ainsi regroupés en 30 génotypes isoenzymatiques (tableau 1).

La diversité intraspécifique analysée pour les espèces comestibles apparaît très contrastée (tableau 2). Les cédratiers présentent une diversité allélique nulle due à une forte homozygotie et à l'absence de polymorphisme entre les cultivars. Les pomelos, les orangers et les bigaradiers ont des structures intraspécifiques similaires. La diversité allélique et l'hétérozygotie y sont modérées ; le polymorphisme intercultivar est inexistant. Les citronniers sont très hétérozygotes et leur polymorphisme intervariétal est très faible puisqu'un seul cultivar, le citron Meyer, se distingue des neuf autres. Les limettiers sont également très hétérozygotes mais manifestent un plus fort polymorphisme intervariétal que les citronniers. Les pamplemoussiers et les mandariniers offrent une très grande richesse allélique, principalement due à un fort polymorphisme intervariétal. Les deux espèces qui possèdent une forte diversité intercultivar — les mandariniers et les pamplemoussiers — ne présentent pas d'écart significatif à la pan-

Tableau 2. Structuration de la diversité allélique intraspécifique observée pour 11 locus codant pour des isoenzymes.

	Effectif	Nombre moyen d'allèles par locus	Diversité totale	Diversité inter-cultivar	Hétérozygotie observée	Ecart à la panmixie
Cédratier	4	1,00	0,00	0,00	0,00	—
Pomelo	10	1,45	0,23	0,00	0,45	*** (5 locus)
Bigaradier	10	1,36	0,18	0,00	0,36	*** (4 locus)
Oranger	10	1,45	0,23	0,00	0,45	*** (5 locus)
Citronnier	10	2,00	0,42	0,02	0,82	*** (9 locus)
Limettier	10	2,09	0,34	0,08	0,54	** (2 locus)
Pamplemoussier	10	2,09	0,25	0,13	0,24	ns
Mandarinier	10	2,00	0,19	0,10	0,17	ns

ns : non significatif au seuil de 5 % ; \*\* : significatif au seuil de 1 % ; \*\*\* : significatif au seuil de 1 %.

mixie, ce qui témoigne sans doute d'importants brassages génétiques au sein de ces taxons. Toutes les autres espèces, à l'exception des cédratiers qui sont totalement fixés, présentent un excès d'hétérozygotes.

La diversité totale au sens de Nei (1973) de l'échantillon des agrumes cultivés est de 0,45. Elle se décompose de manière équilibrée en termes de diversité intraspécifique (0,23) et de diversité interspécifique (0,22), soit une valeur élevée du coefficient  $G_{ST}$  (0,49). Cette valeur traduit une différenciation allélique marquée entre les taxons cultivés. Celle-ci est en effet significative pour 10 des 11 locus analysés. Cette différenciation entre taxons, observée pour la quasi-totalité des locus, se retrouve dans la structuration multilocus évaluée à partir des 30 génotypes de *Citrus* cultivés. Les déséquilibres de liaison concernent ainsi 23 couples de locus sur 55 et intéressent 9 locus sur 11.

Cette forte structuration observée au sein des cultivars est confirmée lorsque l'on s'intéresse aux 47 génotypes enzymatiques identifiés, qui associent aux 30 génotypes des cultivars ceux des 17 autres *Citrus*. Neuf locus sur 11 présentent en effet un écart à la panmixie significatif avec un défaut d'hétérozygotes. Ce type d'écart est classiquement lié à la structuration en sous-populations (effet Walhund) et aux systèmes de reproduction limitant les flux de gènes.

Le fort niveau de structuration génétique constaté grâce aux paramètres de génétique des populations se retrouve dans l'analyse factorielle sur tableau de distances réalisée sur les génotypes des cultivars, où 50,4 % de la variance totale est représentée sur le plan 1-2 (figure 2). La diversité des *Citrus* cultivés se structure autour de trois pôles : le premier regroupe les mandariniers, le deuxième associe les pamplemoussiers et les pomelos, le troisième est formé par les cédratiers, qui présentent un apparentement marqué avec les limettiers. Les orangers et les bigaradiers sont voisins des mandariniers avec une probable introgression des pamplemoussiers. Les citronniers, très hétérozygotes, pourraient être issus d'une hybridation entre le groupe constitué par les mandariniers, les orangers et les bigaradiers et le groupe des limettiers. L'analyse factorielle permet en effet, pour cette population très structurée, d'identifier les formes hybrides et leurs parents potentiels.

Cette structuration des formes cultivées en trois pôles n'est pas remise en cause par l'introduction des formes non cultivées comme le montre l'arbre de diversité construit (figure 3). Certains *Citrus* non cultivés sont associés aux groupes formés par les cultivars : *C. pectinifera* aux mandariniers ; le siamelo aux orangers ; le *pseudogulgul* et *C. intermedia* au groupe des pamplemoussiers et des pomelos ; *C. pennivesiculata*, *C. volkameriana* et le *rough lemon* au groupe des limettiers. Les autres se distinguent de ces groupes soit parce qu'ils portent des allèles qui n'ont pas été observés chez les cultivars — c'est le cas de *C. macrophylla*, de *C. excelsa*, de *C. junos*, de l'*ichangensis lemon*, du *khasi papeda*, du *combava* et du *nasnaran* —, soit parce qu'ils présentent des structures alléliques recombinées originales, comme le bergamotier et la poire du commandeur.

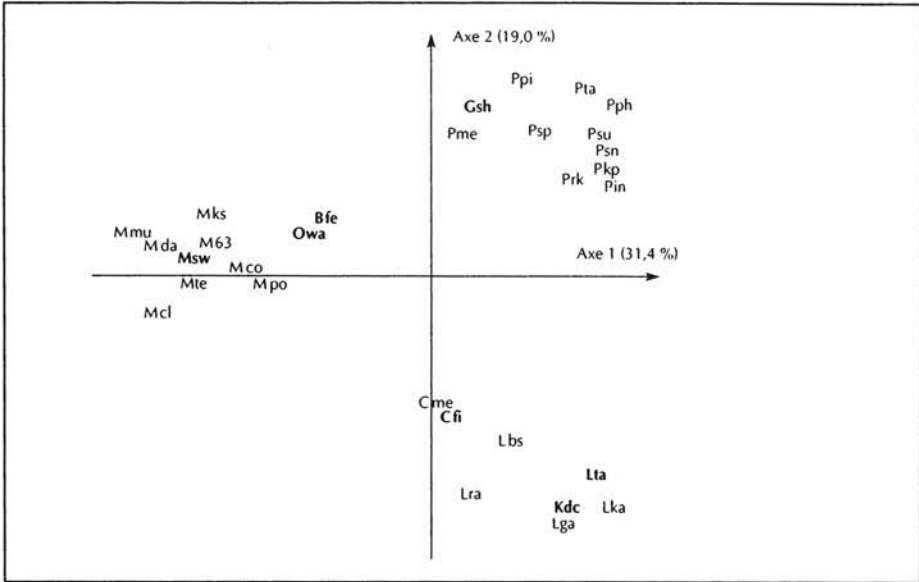


Figure 2. Diversité isoenzymatique des agrumes cultivés sur la base de 11 locus : représentation du premier plan factoriel de l'AFTD réalisée sur une matrice de dissimilarité de Dice entre les 30 génotypes différents identifiés parmi 74 cultivars. Les codes utilisés sont ceux du tableau 1.

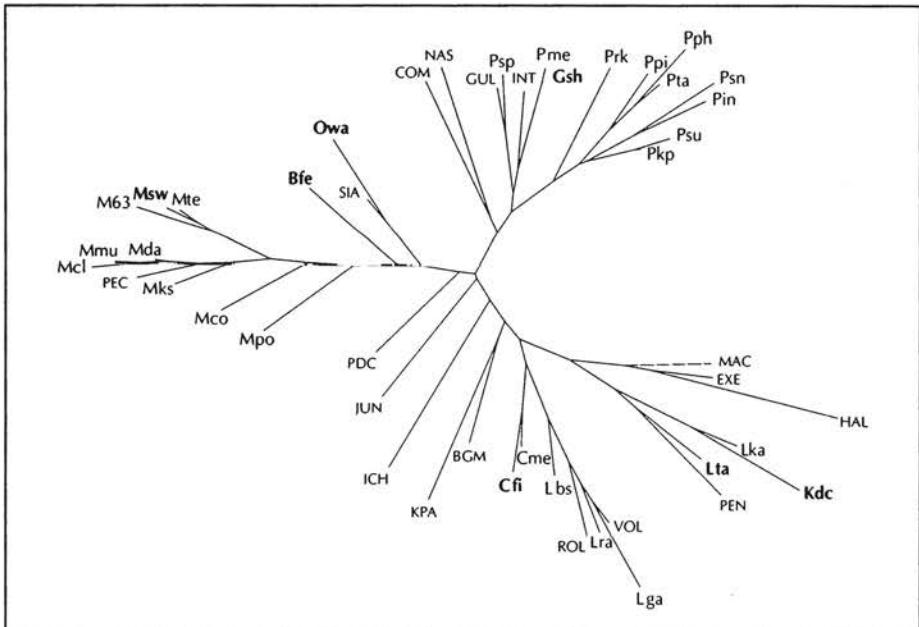


Figure 3. Diversité isoenzymatique du genre Citrus sur la base de 11 locus : représentation arborée, selon la méthode NJ, réalisée sur une matrice de dissimilarité de Dice entre 47 génotypes (30 génotypes cultivés et 17 autres Citrus). Les codes utilisés sont ceux du tableau 1.

## LA TAILLE DES GÉNOMES

Les tailles des génomes nucléaires des individus sont données dans le tableau 1. Les génotypes diploïdes ont des génomes de taille relativement petite, comprise entre 0,73 et 0,82 picogramme d'ADN par génome diploïde (figure 4). Les valeurs de 1,17 picogramme correspondent à des génotypes triploïdes ; elles ont été observées pour quatre cultivars de limettier, Tahiti, Bears, Elkseur, IAC SRA618. Parmi les espèces comestibles, les différences interspécifiques sont statistiquement significatives et représentent un écart de 10 % entre les mandariniers et le cédratiers (figure 5). Les autres espèces se répartissent en deux groupes de tailles intermédiaires. L'un associe les orangers et les bigaradiers, l'autre, correspondant à des tailles supérieures, regroupe les citronniers, les limettiers, les pamplemoussiers et les pomelos. Les types non comestibles présentent eux aussi des tailles de génome comprises entre celle des mandariniers et celle des cédratiers. Ainsi, deux des trois taxons qui structurent la

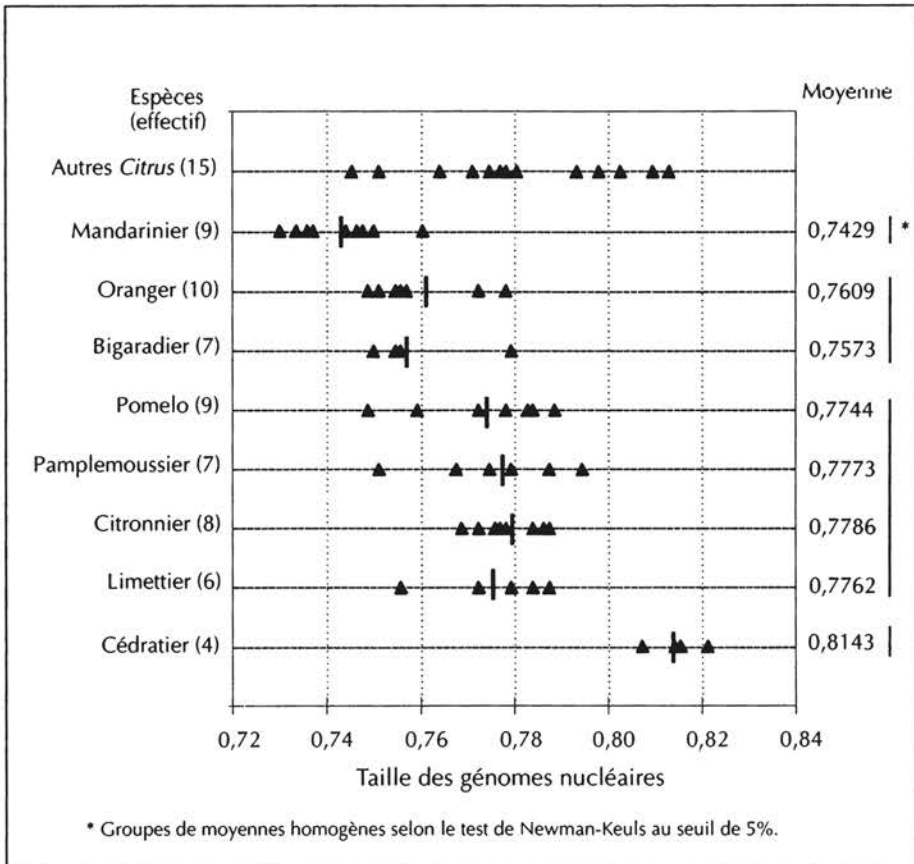


Figure 4. Taille des génomes nucléaires de 75 individus, dont 60 cultivars comestibles regroupés par espèces (moyennes de 3 mesures données en picogrammes d'ADN par génome diploïde).

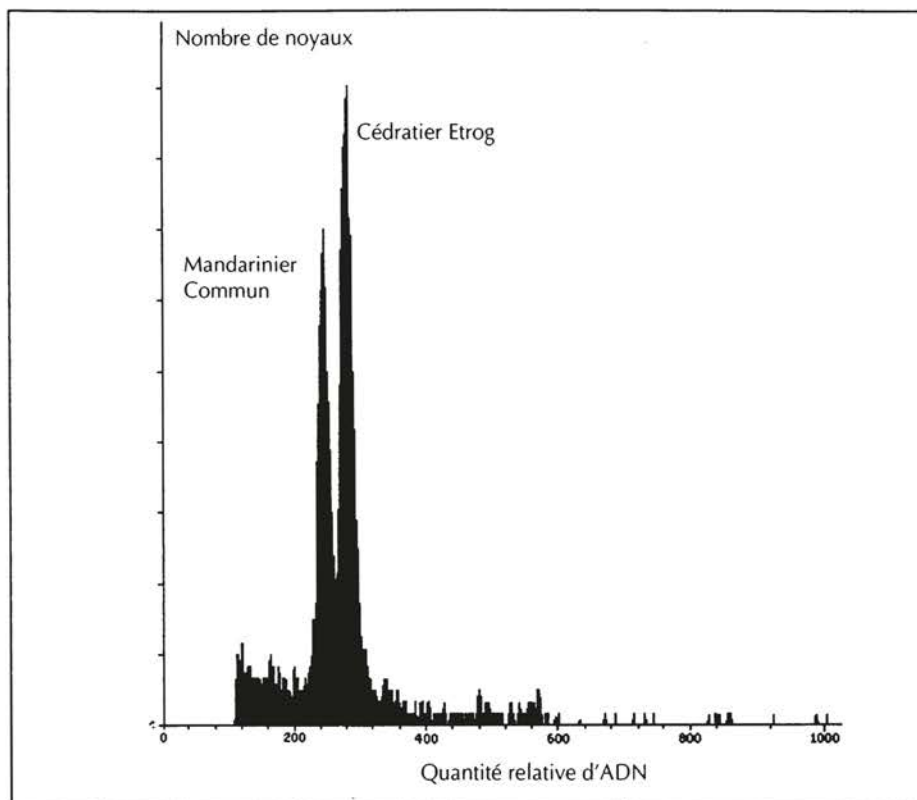


Figure 5. Tailles relatives des génomes nucléaires du cédratier Etrog et du mandarinier Commun : cytométrie en flux d'un mélange de noyaux colorés à l'iodure de propidium.

diversité, le mandarinier et le cédratier, possèdent les tailles de génome extrêmes observées dans le genre *Citrus*. Les autres taxons ont des tailles de génome qui concordent avec les affinités génétiques déterminées par les analyses isoenzymatiques.

## Les relations entre les différents niveaux de variabilité

L'analyse de la diversité morphologique réalisée à partir de 20 descripteurs végétatifs permet de retrouver la structuration globale autour des trois pôles précédemment identifiés d'après les données isoenzymatiques (figure 6). Les positions relatives des espèces cultivées autour de ces trois pôles sont dans l'ensemble conservées. En revanche, les espèces monomorphes du point de vue enzymatique présentent une dispersion morphologique équivalente à celle des espèces polymorphes sur le plan moléculaire (figure 7). Deux niveaux

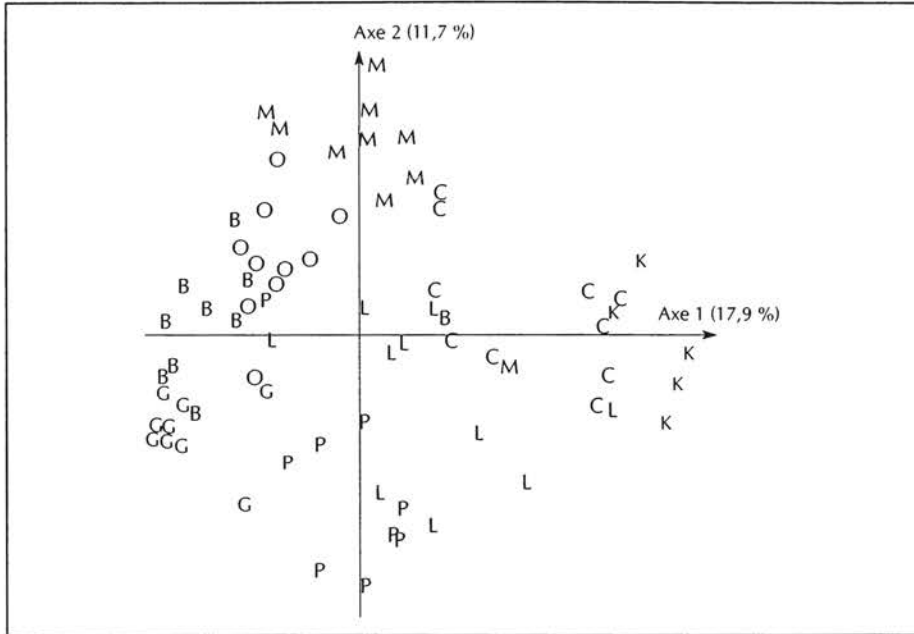


Figure 6. Diversité morphologique : représentation du premier plan factoriel de l'AFTD réalisée sur une matrice de distance de Sokal et Michener entre 74 cultivars sur la base de 20 descripteurs végétatifs. Les codes utilisés sont ceux du tableau 1.

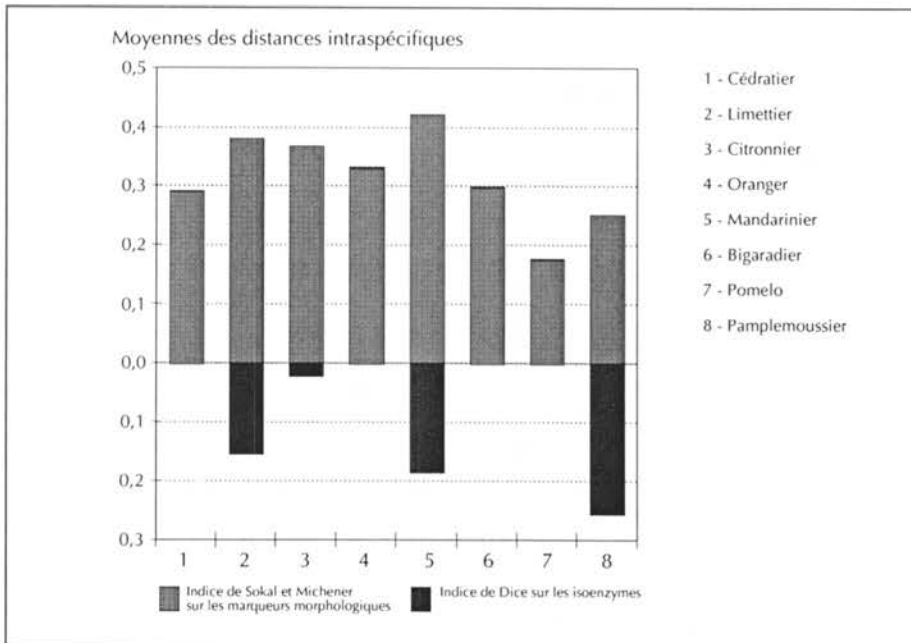


Figure 7. Illustration simultanée de la dispersion intraspécifique calculée avec les marqueurs isoenzymatiques et morphologiques.

coexistent donc dans la structuration de la diversité morphologique : un niveau majeur, qui répond à des contraintes affectant l'évolution du génome dans son ensemble, et un niveau secondaire, dissocié de l'évolution moléculaire visualisée par les isoenzymes.

## La structuration interspécifique

Hormis le système d'auto-incompatibilité gamétophytique, il n'existe pas d'incompatibilité sexuelle au sein du genre *Citrus* : des hybrides sont obtenus aisément pour l'ensemble des combinaisons interspécifiques. Il était ainsi envisageable de remettre en cause la notion même de différenciation spécifique. Il apparaît toutefois que ce genre est très fortement structuré dans la mesure où des déséquilibres gamétiques généralisés ont été identifiés pour les isoenzymes et où les grands axes de structuration moléculaire et morphologique apparaissent similaires. Cela traduit une organisation en sous-populations, entre lesquelles les flux de gènes sont limités, comme le confirment les écarts à la panmixie observés pour la quasi-totalité des locus.

La structuration des agrumes autour de trois pôles — pamplemoussiers, mandariniers, cédratiers — obtenue dans nos analyses est compatible avec les données de GREEN *et al.* (1986) pour l'ADN chloroplastique et de HANDA *et al.* (1986) pour les protéines totales. Elle confirme les résultats de taxonomie numérique de BARRET et RHODES (1976), qui avaient suggéré que ces trois taxons étaient à l'origine de l'ensemble des *Citrus* cultivés. La différenciation entre ces taxons sexuellement compatibles s'explique par un processus de fondation dans trois zones géographiques et par une évolution allopatrique. Les pamplemoussiers seraient en effet originaires de l'archipel malais et de l'Indonésie, les cédratiers auraient évolué dans le nord-est de l'Inde et dans les régions voisines de Birmanie et de Chine, tandis que les mandariniers se seraient diversifiés dans une région qui couvre le Vietnam, la Chine du Sud et le Japon (WEBBER, 1967 ; SCORA, 1975).

Les autres espèces cultivées — orangers, bigaradiers, citronniers, pomelos, limettiers — seraient apparues ensuite par recombinaison entre ces taxons de base, mis en contact au gré des échanges commerciaux et des migrations humaines. Les données enzymatiques — hétérozygotie généralement élevée et absence de polymorphisme intervariétal — prouvent qu'il s'agit de cas typiques de fausses espèces, pour lesquelles la diversification variétale s'est produite à partir d'un hybride ancestral par accumulation de mutations sans qu'intervienne la recombinaison sexuée. Il est à noter que tous les cultivars de ces espèces sont polyembryonnés ce qui a permis de fixer l'hétérozygotie et de conserver le type morphologique et pomologique alors que les méthodes de propagation végétative, comme le marcottage, le bouturage ou le greffage, n'existaient pas.

La forte structuration, encore observée aujourd'hui tant à l'échelle moléculaire que morphologique, indique que les brassages génétiques entre les trois groupes originels sont limités. L'apomixie partielle, liée à la polyembryonie, a très certainement été un élément essentiel dans la limitation des flux de gènes. D'autres facteurs, comme la différenciation structurelle des génomes, ont dû également favoriser le maintien de déséquilibres gamétiques en limitant la recombinaison sur de larges portions du génome. Cette différenciation dans la taille de génome est en accord avec les observations cytogénétiques de NAIR et RANDHAWA (1969) et de RAGHUVANSHI (1969). Elle témoigne du stade avancé qu'ont atteint les trois taxons de base sur la voie d'une réelle spéciation.

## La diversification intraspécifique

Le polymorphisme morphologique intervariétal, relativement important au sein des orangers, des bigaradiers, des pomelos, des citronniers et des limettiers, s'explique en grande partie par la sélection humaine. Celle-ci est particulièrement marquée pour les critères pomologiques et phénologiques. Elle peut entraîner une évolution morphophysologique rapide, indépendante de l'évolution moléculaire analysée grâce aux isoenzymes. L'exemple le plus flagrant est celui du clémentinier. Apparu il y a environ un siècle dans un semis de mandarinier Commun réalisé par le père Clément, il a connu depuis une diversification considérable. Cette diversification, résultat d'une simple sélection en verger des mutations de bourgeons, concerne tant la précocité — la période de production s'étend aujourd'hui d'octobre à mars — que les caractères pomologiques, comme le calibre, la coloration et la présence de pépins (BONO *et al.*, 1982).

En revanche, la recombinaison sexuée a dû jouer un rôle déterminant dans la diversification des pamplemoussiers, pour lesquels les cultivars sont tous monoembryonnés, et dans celle des mandariniers, dont certains cultivars sont monoembryonnés. Ils présentent en effet un polymorphisme isoenzymatique intervariétal élevé sans écart significatif à la panmixie.

## La gestion des ressources génétiques

La situation des agrumes illustre bien l'intérêt et les limites des marqueurs moléculaires pour la construction de *core collections*. On retrouve en effet dans l'évolution du genre *Citrus* des facteurs qui concourent, à l'échelle globale, à une bonne corrélation entre la structuration de la diversité phénotypique et la structuration de la diversité moléculaire (effet de fondation, évolution allopatrique et limitation des flux de gènes autorisant le maintien de déséquilibres gamétiques globaux). Pour les espèces secondaires, il existe également, sur le plan intraspécifique, des mécanismes évolutifs, tels que la reproduction somatique et les fortes pressions de sélection sur les mutations affectant les caractères morphophysologiques, qui conduisent à dissocier les deux



niveaux d'évolution. Dans le cas des agrumes, le principal intérêt des études de marquage réside dans l'identification des séquences et des facteurs évolutifs à l'origine des taxons et de leur diversification. La réflexion sur la constitution d'une *core collection* doit donc reposer davantage sur ces enseignements généraux que sur la constitution allélique d'individus.

Parmi les trois espèces de base, les pamplemoussiers et les mandariniers présentent un polymorphisme moléculaire important. L'amélioration variétale intraspécifique peut être réalisée traditionnellement par hybridation sexuée. La gestion des ressources génétiques intraspécifiques peut donc se raisonner classiquement sous forme de *core collections*. Les résultats obtenus sur une collection d'une centaine de mandariniers témoignent de l'existence d'une forte structuration à l'échelle intraspécifique, qui pourrait permettre d'établir, pour partie, une stratégie d'échantillonnage sur la base des données moléculaires.

Les ensembles de caractères définissant les autres espèces cultivées — orangers, bigaradiers, pomelos, citronniers — reposent sur des génotypes qui ont une hétérozygotie relativement élevée mais qui sont stabilisés par la multiplication végétative. La conservation des ressources génétiques de chacune de ces espèces doit s'appuyer sur la constitution de collections de génotypes. Cette diversité intraspécifique peut difficilement être recombinaisonnée par voie sexuée pour améliorer l'« espèce » puisque les caractères définissant l'« espèce » sont alors recombinaisonnés. Ces collections, qui visent à conserver la plus large diversité adaptative et morphophysiological au sein de chaque « espèce », doivent en revanche permettre de proposer aux agrumiculteurs les cultivars les mieux adaptés aux différentes régions de production. Le marquage moléculaire n'apportant aucune information à ce niveau compte tenu des mécanismes d'évolution intraspécifique préalablement décrits, la stratification doit se fonder principalement sur les critères géographiques et sur les données agromorphologiques.

Si l'on s'intéresse aux agrumes en général, la gestion des ressources génétiques peut se raisonner également en termes de conservatoire de gènes. Le groupe de trois taxons identifiés comme étant à l'origine des diverses formes cultivées constitue alors un réservoir essentiel puisqu'une grande part de la diversité allélique y réside au niveau intercultivar. Les mandariniers et les pamplemoussiers paraissent à ce titre devoir posséder une importance plus grande dans les conservatoires. Par ailleurs, certains agrumes non cultivés apportent, comme le montre notre étude, un enrichissement de la diversité allélique. Ces taxons ne sont donc pas des combinaisons génotypiques particulières issues d'hybridation entre les trois taxons de base des formes cultivées. Leur conservation apparaît essentielle, en particulier pour les tolérances aux facteurs biotiques ou abiotiques qu'ils peuvent apporter dans l'amélioration des porte-greffe. Enfin, le développement des biotechnologies, en particulier de l'hybridation somatique, élargit considérablement le pool génique exploitable pour les porte-greffe (GROSSER *et al.*, 1996). Il convient donc aujourd'hui de raisonner la conservation des ressources génétiques des agrumes à l'échelon de la tribu des *Citreae*.

# Annexe

## Matériel végétal

Soixante-quatorze cultivars représentant les 8 espèces cultivées pour leurs fruits (SWINGLE et REECE, 1967) et 17 types non consommés, mais dont certains sont utilisés comme porte-greffe, ont servi de base à l'étude enzymatique (tableau 1). Dans la mesure du possible 10 cultivars ont été retenus pour chaque espèce cultivée, à l'exception du cédratier pour lequel nous ne disposons que de 4 géotypes en collection. Les arbres, indemnes de toute maladie virale ou viroïdale, sont cultivés sur la station de recherche agronomique de l'Inra et du Cirad de San Giuliano, en Corse. Quatre-vingt-dix de ces géotypes y ont fait l'objet d'une description morphologique.

## Analyses enzymatiques

Neuf systèmes enzymatiques ont été analysés par électrophorèse sur gel d'amidon ou de polyacrylamide (OLLITRAULT *et al.*, 1992) : alcool déshydrogénase (ADH), malate déshydrogénase (MDH), isocitrate déshydrogénase (IDH), shikimate déshydrogénase (SKDH), phosphoglucomutase (PGM), phosphogluco-isomérase (PGI), peroxydases (PER), leucine aminopeptidase (LAP) et aspartate aminotransférase (AAT). Pour le locus *PGM-2*, seules deux positions alléliques ont été retenues. Pour les autres systèmes, l'interprétation et la nomenclature allélique reprennent celles de OLLITRAULT *et al.* (1992) et sont en accord avec l'interprétation donnée par TORRES *et al.* (1978 ; 1982) pour les MDH, IDH, PGI et LAP.

## Analyses par cytométrie en flux

La taille du génome nucléaire de chacun des géotypes diploïdes a été estimée par la moyenne de trois mesures relatives à celle d'un cultivar triploïde (la lime Tahiti), utilisé comme témoin interne. Des morceaux de feuille de l'échantillon et du témoin sont préparés en mélange et colorés à l'iodure de propidium selon le protocole décrit par OLLITRAULT *et al.* (1994). Deux mille noyaux sont ensuite analysés sur un cytomètre Fasca. La taille du génome nucléaire de chaque géotype est estimée en picogrammes par génome diploïde à partir de la moyenne des valeurs relatives multipliée par 1,17 picogramme, qui correspond à la taille du génome de la lime Tahiti estimée par OLLITRAULT *et al.* (1994).

## Etudes morphologiques

Vingt descripteurs qualitatifs de l'appareil végétatif (tableau 3) ont été étudiés. L'ensemble des données sur la morphologie des agrumes est géré par le système de gestion du matériel génétique des agrumes en réseau, Egid (COTTIN *et al.*, 1995).

## Analyses statistiques

L'étude des paramètres de structuration génétique est réalisée à l'aide du logiciel Genepop pour l'analyse des écarts à la panmixie, de la différenciation entre taxons cultivés (étude de la répartition des allèles dans les espèces par le test exact de Fisher) et des déséquilibres gamétiques. Les paramètres descriptifs de la diversité — diversité totale, diversité entre taxons, diversité entre individus,  $G_{ST}$  — sont ceux qui ont été proposés par NEI (1973). Les représentations arborées et les analyses factorielles sur tableaux de distances sont réalisées sur la base de matrice de distance de Dice pour les données enzymatiques et de matrice de distance de Sokal et Michener pour les données morphologiques. Les arbres sont construits par la méthode du *neighbor-joining* à l'aide du logiciel Darwin (PERRIER *et al.*, 1999).

Tableau 3. Les vingt descripteurs qualitatifs morphologiques.

A. Port de l'arbre	M. Longueur des épines
1 érigé	1 nulle
2 sphéroïde	2 très courte (0 à 5 mm)
3 élipsoïde aplati	3 courte (5 à 15 mm)
B. Position des branches	4 moyenne (15 à 40 mm)
1 érigée	5 longue (> 40 mm)
2 étalée	N. Forme de la section des jeunes rameaux
3 retombante	1 angulaire
4 pleureuse	2 ronde
C. Densité de frondaison	O. Bordure de la feuille
1 éparse	1 crénelée
2 dense	2 dentelée
D. Surface du tronc	3 entière
1 lisse	4 ondulée
2 rugueuse	P. Forme de la feuille
E. Couleur de la face supérieure des feuilles	1 elliptique
1 vert clair	2 ovale
2 verte	3 ovale inversée
3 vert foncé	4 lancéolée
F. Couleur de la face inférieure par rapport à la face supérieure	5 orticulaire
1 identique	Q. Longueur du pétiole
2 plus claire	1 nulle
G. Nervures sur la face supérieure	2 courte (0 à 10 mm)
1 proéminantes	3 moyenne (10 à 15 mm)
2 non proéminantes	4 longue (15 à 35 mm)
H. Angle de la base foliaire	5 très longue (> 35 mm)
1 aigu	R. Forme des ailettes
2 obtu	1 absente
I. Angle de l'apex foliaire	2 cordiforme
1 aigu	3 deltoïde
2 obtu	4 ovale
J. Articulation de la feuille	S. Taille des ailettes
1 présente	1 insignifiante
2 absente	2 petite
K. Attache du pétiole au rameau	3 moyenne
1 droite	4 grosse
2 coudée	5 énorme (égal au limbe)
L. Densité des épines	T. Couleur des jeunes pousses
1 nulle	1 anthocyanée
2 faible	2 verte
3 modérée	
4 forte	

## Références bibliographiques

- BARRET H.C., RHODES A.M., 1976. A numerical taxonomic study of affinity relationships in cultivated *Citrus* and its close relatives. *Systematic Botany*, 1 : 105-136.
- BONO R., FERNANDEZ DE CORDOVA L., SOLER J., 1982. Arrufatina, Esbal and Guillermina, three Clementine mandarin mutations recently appearing in Spain. *Proceedings of the International Society of Citriculture*, 1 : 94-96.
- COTTIN R., ALLENT V., JACQUEMOND C., 1995. Gestion informatique des ressources génétiques : Egid. *In* : Symposium méditerranéen sur les mandarines. San Giuliano, France, Inra, p. 2.
- DURHAM R.E., LIU P.C., GMITTER R.G., MOORE G.A., 1992. Linkage map of restriction fragment length polymorphisms and isozymes in *Citrus*. *Theoretical and Applied Genetics*, 84 : 39-48.
- FANG D., ROOSE M.L., 1996. Fingerprinting citrus cultivars with inter-SSR markers. *Proceedings of the International Society of Citriculture* : 185-188.
- FAO, 1997. *Annuaire production : 1996*. Rome, Italie, FAO.
- GMITTER F.G., DENG X.X., HEARN C.J., 1992. Cytogenetic mechanism underlying reduced fertility and seedlessness in *Citrus*. *In* : VIIIth international citrus congress, p. 113-116.
- GREEN R.M., VARDI A., GALUN E., 1986. The plastome of *Citrus*: physical map, variation among *Citrus* cultivars and species and comparison with related genera. *Theoretical and Applied Genetics*, 72 : 170-177.
- GROSSER J.W., MOURAO-FO A.A., GMITTER F.G.JR., LOUZADA E.S., JIANG J., BAERGEN K., QUIROS A., CABASSON C., SCHELL J.L., CHANDLER J.L., 1996. Allotetraploid hybrids between *Citrus* and seven related genera produced by somatic hybridization. *Theoretical and Applied Genetics*, 92 : 577-582.
- GUERRA M.S., 1993. Cytogenetics of Rutaceae. 5. High chromosomal variability in *Citrus* species revealed by CMA/DAPI staining. *Heredity*, 71 : 234-241.
- HANDA T., ISHIZAWA Y., OOGAKI C., 1986. Phylogenetic study of fraction I protein of *Citrus* and its close related genera. *Journal of Genetics*, 61 : 15-24.
- HERRERO R., ASINS M.J., CARBONELL E.A., NAVARRO L., 1996. Genetic diversity in the orange subfamily Aurantiodeae. 1. Intraspecies and intragenus genetic variability. *Theoretical and Applied Genetics*, 92 : 599-609.
- HERRERO R., ASINS M.J., PINA J.A., CARBONELL E.A., NAVARRO L., 1997. Genetic diversity in the orange subfamily Aurantiodeae. 2. Genetic relationships among genera and species. *Theoretical and Applied Genetics*, 93 : 1327-1334.
- HIRAI M., KOZAKI I., KAJIURA I., 1986. Isozyme analysis and phylogenetic relationship of *Citrus*. *Japanese Journal of Breeding*, 36 : 377-389.
- JARREL D.C., ROOSE M.L., TRAUUGH S.N., KUPPER R.S., 1992. A genetic map of *Citrus* based on the segregation of isozymes and RFLPs in an intergeneric cross. *Theoretical and Applied Genetics*, 84 : 49-56.

- KHAN I.A., ROOSE M.L., 1988. Frequency and characteristics of nucellar and zygotic seedlings in three cultivars of trifoliolate orange. *Journal of the American Society for Horticultural Science*, 113 : 105-110.
- KRUG C.A., 1943. Chromosome numbers in the subfamily Arantioideae, with special reference in the genus *Citrus*. *Citrus Botanical Gazette*, 104 : 602-611.
- LURO F., LAIGRET F., BOVE J.M., OLLITRAULT P., 1994a. Application of RAPD to *Citrus* genetics and taxonomy. *In* : VIIIth international citrus congress, p. 225-228.
- LURO F., LAIGRET F., OLLITRAULT P., BOVE J.M., 1995. DNA amplified fingerprinting (DAF); an useful tool for determination of genetic origin and diversity analysis in *Citrus*. *HortScience*, 30 : 1063-1067.
- LURO F., LORIEUX M., LAIGRET F., BOVE J.M., OLLITRAULT P., 1994b. Genetic mapping of an intergeneric *Citrus* hybrid using molecular markers. *Fruits*, 49 : 404-408.
- MIRANDA M., IKEDA F., ENDO T., MORIGUCHI T., OMURA M., 1997. Chromosome markers and alterations in mitotic cells from interspecific *Citrus* somatic hybrids analysed by fluorochrome staining. *Plant Cell Reports*, 16 : 807-812.
- NAIR P.K.R., RANDHAWA G.S., 1969. Chromosome morphology of the pachytene stage with respect to different *Citrus* types. *In* : Ist international citrus symposium, p. 215-223.
- NEI M., 1973. Analysis of gene diversity in subdivided population. *Proceedings of the National Academy of Science of the United States of America*, 70 : 3321-3323.
- OLLITRAULT P., DAMBIER D., LURO F., DUPERRAY C., 1994. Nuclear genome size variations in *Citrus*. *Fruits*, 49 : 390-393.
- OLLITRAULT P., FAURE X., 1992. Système de reproduction et organisation de la diversité génétique dans le genre *Citrus*. *In* : Complexes d'espèces, flux de gènes et ressources génétiques des plantes. Paris, France, BRG, p. 133-151.
- OLLITRAULT P., FAURE X., NORMAND F., 1992. Citrus rootstocks characterization with bark and leaf isozymes: application for distinguishing nucellar from zygotic trees. *In* : VIIIth international citrus congress, p. 338-341.
- OLLITRAULT P., LURO F., 1997. Les agrumes. *In* : L'amélioration des plantes tropicales, A. Charrier *et al.* éd., Montpellier, France, Cirad-Orstom, p. 13-36.
- PERRIER X., FLORI A., BONNOT F., 1999. Les méthodes d'analyse des données. *In* : Diversité génétique des plantes tropicales cultivées, P. Hamon *et al.* éd., Montpellier, France, Cirad, collection Repères, p. 43-76.
- RAGHUVANSHI S.S., 1969. Cytological evidence bearing on evolution in *Citrus*. *In* : Ist international citrus symposium, p. 207-214.
- SCORA R.W., 1975. On the history and origin of citrus. *In* : Symposium on the biochemical systematics, genetics and origin of cultivated plants. *Bulletin of the Torrey Botanical Club*, 102 : 369-375.
- SCORA R.W., 1988. Biochemistry, taxonomy and evolution of modern cultivated citrus. *In* : VIth international citrus congress, p. 277-289.
- SOOST R.K., WILLIAMS T.E., TORRES A.M., 1980. Identification of nucellar and zygotic seedlings with leaf isozymes. *HortScience*, 15 : 728-729.

SWINGLE W.T., REECE P.C., 1967. The botany of *Citrus* and its wild relatives. *In* : The citrus industry. 1. History, world distribution, botany and varieties, W. Reuther *et al.* éd., Berkeley, Etats-Unis, University of California Press, p. 190-430.

TANAKA T., 1961. Citrologia: semi centennial commemoration papers on *Citrus* studies. Osaka, Japon, Citrologia Supporting Foundation, 114 p.

TATUM J.H., BERRY R.E., HEARN C.I., 1974. Characterization of citrus cultivars and separation by thin layer chromatography. Proceedings of the Florida State Horticultural Society, 87 : 75-81.

TORRES A.M., SOOST R.K., DIEDENHOFEN U., 1978. Leaf isozymes as genetic markers in *Citrus*. American Journal of Botany, 65 : 869-881.

TORRES A.M., SOOST R.K., MAU-LASTOVICKA T., 1982. Citrus isozymes: genetic and distinguishing nucellar from zygotic seedlings. Journal of Heredity, 73 : 335-339.

WEBBER H.J., 1967. History and development of the citrus industry. *In* : The citrus industry. 1. History, world distribution, botany and varieties, W. Reuther *et al.* éd., Berkeley, Etats-Unis, University of California Press, p. 1-39.



# Les bananiers

---

Christophe Jenny, Françoise Carreel,  
Kodjo Tomekpe, Xavier Perrier, Cécile Dubois,  
Jean-Pierre Horry, Hugues Tézenas du Montcel

La banane, dont la production mondiale était estimée à 84 millions de tonnes en 1996, est le premier fruit produit au monde et échangé sur le marché international, devant la pomme, l'orange, le raisin et le melon. Les bananiers sont cultivés dans plus de cent vingt pays tropicaux et subtropicaux, sur les cinq continents. Les productions bananières ont un rôle important du point de vue alimentaire, social et économique, mais aussi environnemental.

Il existe deux grandes filières de production : celle des bananiers destinés à l'exportation et celle des bananiers réservés à l'approvisionnement des marchés locaux.

Les bananiers cultivés pour l'exportation, connus sous les noms de Grande Naine, Poyo et Williams, appartiennent au sous-groupe des bananiers triploïdes de type Cavendish. Ils ne diffèrent entre eux que par des mutations somatiques telles que la hauteur de la plante et la conformation des régimes et des fruits. Leur production repose sur un système de monoculture intensive de type agro-industriel, qui ne fait pas intervenir de rotations, mais recourt à de nombreux intrants.

La culture des bananiers pour la consommation locale est fondée sur un grand nombre de cultivars, adaptés aux différents contextes de production ainsi qu'aux utilisations et aux goûts variés des consommateurs. Les systèmes de production de ces bananiers n'utilisent généralement que peu ou pas d'intrants.



Des bananiers diploïdes, proches des formes sauvages ancestrales, sont encore cultivés en Asie du Sud-Est. Sur les autres continents, les cultivars triploïdes appartenant à différents sous-groupes — Plantain, Figue Pomme, Lujugira, Gros Michel, Pisang Awak — sont les plus répandus.

Les bananes font l'objet de multiples utilisations. Elles sont consommées sous forme de fruits frais, mais aussi cuites, comme les bananes Plantain. Elles sont soumises à de nombreuses transformations : chips, frites, beignets, purées, confitures, ketchup, alcools (vin et bière de banane, dont la production est particulièrement importante en Afrique de l'Est). La consommation quotidienne par personne de bananes varie de 30 grammes à plus de 500 grammes dans certains pays d'Afrique de l'Est. Hormis le fruit, d'autres parties de la plante sont utilisées : le pseudo-tronc, pour ses fibres et comme flotteurs (*M. textilis* appelé abacca) aux Philippines ; les feuilles pour fabriquer des abris, des couvertures ou des emballages pour la cuisson (les Plantain en Afrique). En Thaïlande, les bourgeons floraux de variétés particulières (Pisang Awak) sont utilisés dans de multiples préparations culinaires. Enfin, on attribue à certaines variétés des propriétés médicinales.

Cultivés dans le monde entier, les bananiers sont menacés par de nombreuses maladies et différents ravageurs. Les contraintes les plus fortes sont exercées par des nématodes — *Radopholus similis* et plusieurs représentants du genre *Pratylenchus* — et par le charançon noir du bananier, *Cosmopolites sordidus*. Différentes maladies fongiques sont aussi des contraintes majeures de la production industrielle et, à un moindre degré, des productions locales. Ainsi, la cercosporiose jaune, due à *Mycosphaerella musicola*, et la maladie des raies noires, causée par *M. fijiensis*, provoquent dans les grandes plantations industrielles des pertes de rendement et entraînent des coûts de traitement élevés. Dans certaines zones de production, la maladie de Panamá, due à un champignon du sol, *Fusarium oxysporum* f. sp. *cubense*, interdit la culture des variétés de type Gros Michel. Enfin, les maladies virales connaissent une expansion récente, peut-être simplement liée à une meilleure détection. Les plus préoccupantes sont dues au CMV (*cucumber mosaic virus*), au BSV (*banana streak virus*), au BBTV (*banana bunchy top virus*) et au BBMV (*banana bract mosaic virus*).

## La botanique et la taxonomie

### La botanique

Les bananiers sont des monocotylédones originaires du Sud-Est asiatique, qui appartiennent au genre *Musa*, famille des musacées, ordre des zingibérales. Le bananier est une herbe géante, dont le pseudo-tronc, formé par l'emboîtement

des gaines foliaires, mesure de 1 à 8 mètres. Les feuilles sont émises par le méristème terminal de la tige vraie, improprement appelée bulbe, qui est souterraine et de taille réduite. Le bourgeon situé à l'aisselle de chaque feuille est susceptible de se développer en rejet. A la fin de la phase végétative, le changement de fonction du méristème central induit la croissance et l'allongement de la tige vraie au cœur du pseudo-tronc, puis l'émergence de l'inflorescence. L'inflorescence, qui peut être verticale, pendante ou subhorizontale, forme une grappe. Elle est constituée de spathes imbriquées, disposées en hélice, à l'aisselle desquelles émergent les rangées simples ou doubles de fleurs.

Les fleurs des premières rangées sont femelles (ovaire développé, présence de staminodes) ou, dans quelques rares cas, hermaphrodites (ovaire et étamines développés). La cavité des ovaires infères peut se remplir de graines et de pulpe pour former le fruit. Ces rangées de fleurs, ou « mains », forment alors le régime. Après les mains femelles, apparaissent les mains de fleurs neutres (ni mâle ni femelle), puis les mains de fleurs mâles (ovaire réduit, étamines bien développées). Chez certains cultivars, la croissance du méristème apical s'interrompt après l'apparition des mains femelles. Mais, en général, la croissance de l'inflorescence se poursuit indéfiniment pour former ce que l'on appelle improprement le bourgeon mâle.

Les bananiers sauvages sont séminifères, leurs fruits remplis de graines contiennent peu de pulpe. En revanche, les bananiers cultivés sont de type parthénocarpique vrai. La cavité des ovaires infères se remplit de pulpe pour former le fruit, sans pollinisation ni formation de zygote. La stérilité femelle de la majorité de ces bananiers est très forte, voire totale, il n'y a donc généralement pas de graines dans leurs fruits. Pour l'ensemble des bananiers, la pérennité de la plante est assurée grâce à une multiplication végétative naturelle par rejetonnage, chaque variété est donc un clone.

## Les bases de la classification

Le genre *Musa* est composé de quatre sections : *Australimusa* ( $2n = 20$ ), *Callimusa* ( $2n = 20$ ), *Rhodochlamys* ( $2n = 22$ ) et *Eumusa* ( $2n = 22$ ). Dans cette dernière section se trouvent presque tous les bananiers cultivés à l'exception des Fe'i, bananiers cultivés diploïdes de la zone du Pacifique appartenant aux *Australimusa*. Les sections *Callimusa* et *Rhodochlamys* contiennent essentiellement des espèces d'intérêt floral.

La morphotaxonomie a permis de caractériser les variétés de bananiers et de dresser les bases de la classification botanique adoptée aujourd'hui. Kurz a proposé dès 1865 l'hypothèse d'une origine bispécifique — *Musa acuminata* et *M. balbisiana* — des cultivars (figure 1). SIMMONDS et SHEPHERD (1955), en se fondant sur une méthode de scores, ont précisé la contribution relative des deux espèces de départ dans la constitution des cultivars. Parmi les nombreux

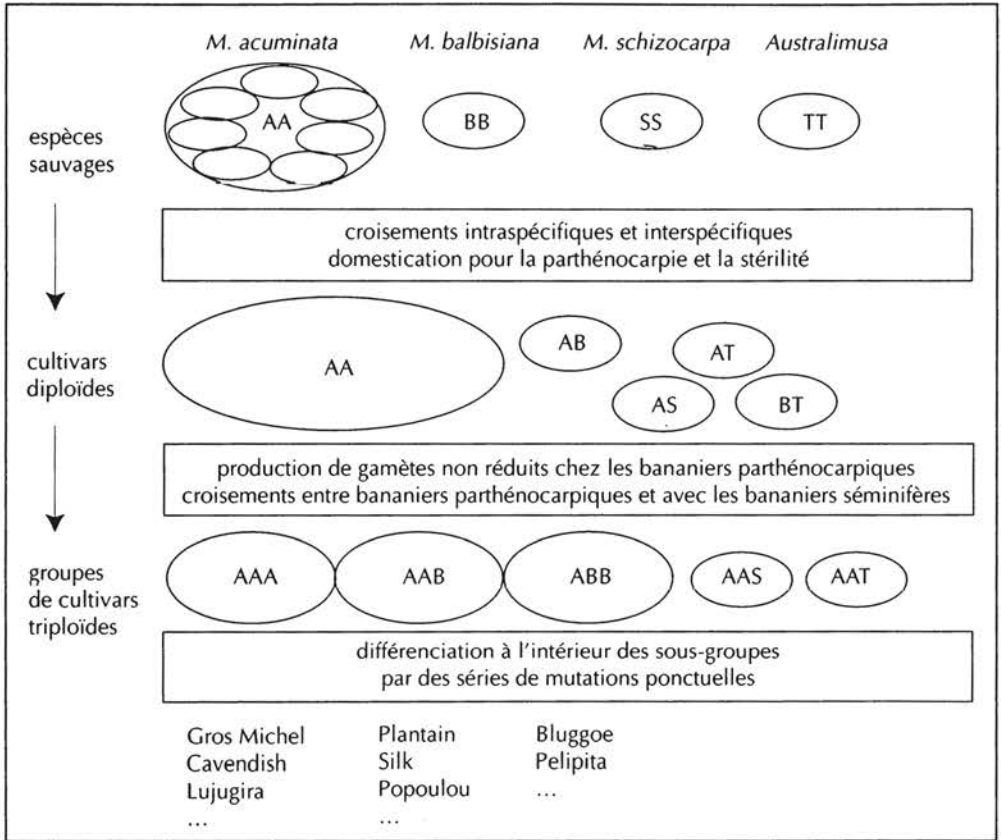


Figure 1. Schéma de domestication des bananiers.

caractères morphologiques qui permettent de caractériser un bananier, ces auteurs en ont retenu quinze, choisis en raison de leur stabilité et de leur capacité à discriminer les différents groupes de bananiers cultivés. Chaque caractère a été quantifié sur une gamme de 1 à 5, où 1 correspond à une expression phénotypique des bananiers sauvages de l'espèce *M. acuminata*, notée A, et 5 à celle des bananiers sauvages de l'espèce *M. balbisiana*, notée B. Pour chaque cultivar, le niveau de ploïdie et le score obtenu par l'addition des notes pour chacun des quinze caractères déterminent sa constitution génomique et, par conséquent, son appartenance à un groupe donné (tableau 1). La majorité des cultivars se classent dans les groupes AA, AAA, AAB et ABB.

Au sein de chaque groupe génomique, sont rassemblés en sous-groupes les cultivars qui présentent une forte proportion de caractères morphotaxonomiques communs et qui dérivent souvent les uns des autres par mutation (tableau 2).

Les informations fournies par les marqueurs moléculaires ont démontré l'implication marginale de *M. schizocarpa* (génome S, section *Eumusa*) et de *M. textilis* (génome T, *Australimusa*,  $2n = 2x = 20$ ) dans la constitution génomique de quelques cultivars, qui sont classés AS, AT ou AAT.

Tableau 1. Déduction de la constitution génomique d'une variété par la connaissance de son niveau de ploïdie et de son score. Les intervalles des scores sont indiqués entre parenthèses, d'après SIMMONDS et SHEPHERD (1955).

Score théorique	Niveau de ploïdie		
	2x	3x	4x
15	AA (16-23)	AAA (15-21)	AAAA (15-20)
30			AAAB (27-35)
35		AAB (26-46)	
45	AB (46-49)		AABB (45-48)
55		ABB (59-63)	
60			ABBB (63-67)
75	BB	BBB	BBBB

Tableau 2. Les différents types de bananiers triploïdes cultivés.

Groupe génomique	Sous-groupe	Type*	Répartition géographique actuelle
AAA	Gros Michel**	dessert	monde, zones
	Cavendish**	dessert	de production industrielle
	Figue Rose	dessert	Pacifique, Antilles, Philippines
	Lujugira-Mutika	à cuire, à bière	Afrique de l'Est
	Ibota	dessert	Thaïlande, Afrique centrale
AAB	Figue Pomme	dessert	Extrême-Orient,
	Pome-Prata	dessert	Amérique latine, Caraïbe
	Mysore	dessert	Inde, Australie, Hawaïi,
	Pisang Kelat	mixte	Brésil, Afrique
	Pisang Rajah	mixte	Inde, Brésil
	Plantain**	à cuire	Inde, Malaisie
	Popoulou-Maïa Maoli	à cuire	Malaisie, Indonésie
Laknao	à cuire	Philippines, Amérique latine	
Iholena		à cuire	Afrique du Centre
		à cuire	et de l'Ouest, Caraïbe
ABB	Bluggoe**	à cuire	Pacifique
	Monthan	à cuire	Philippines
	Pelipita	à cuire	Pacifique, Amérique latine,
	Pisang Awak	dessert	Caraïbe, Afrique de l'Est
	Peyan	à cuire	Inde
	Saba	à cuire	Inde
		à cuire	Philippines

\* La qualité dessert ou à cuire d'un fruit est très subjective, la plupart des types dessert peuvent être consommés cuits, mais l'inverse n'est que très rarement vrai. \*\* Sous-groupes les plus répandus.

## La reproduction sexuée et l'interfertilité

Les bananiers sauvages se reproduisent aussi bien par voie sexuée que par multiplication végétative. Quelques espèces, comme *M. schizocarpa* et *M. acuminata* sous-espèce *banksii*, sont préférentiellement autogames, les premières mains de fleurs de l'inflorescence étant hermaphrodites. Chez les autres espèces, les fleurs basales sont femelles et les bananiers allogames. Toutefois, ces espèces ont souvent une croissance en bouquet, qui permet une autopolli-nisation par rejets interposés (géitonogamie).

En Asie, où les formes séminifères sont toujours présentes, quelques hybrides entre les sections ont été identifiés, entre *M. acuminata* (*Eumusa*) et *M. laterita* (*Rhodochlamys*) et entre *M. balbisiana* (*Eumusa*) et *M. textilis* (*Australimusa*). Plusieurs hybrides interspécifiques, AB et AS, ont également été trouvés.

L'espèce *M. acuminata* présente une grande diversité de formes. Elle a été structurée en sous-espèces interfertiles sur la base des caractéristiques morphologiques et de la répartition géographique. Par ailleurs, SHEPHERD (1987) a établi, à partir d'observations cytologiques, une nomenclature de cette espèce en sept groupes à l'intérieur desquels les clones sont homozygotes de structure. Les hybrides entre ces groupes sont hétérozygotes pour une à quatre translocations ou inversions, ce qui limite les introgressions. Ces groupes recourent en grande partie la structuration en sous-espèces.

Les bananiers cultivés sont fortement stériles. Plusieurs phénomènes provoquant des erreurs méiotiques entraînent une stérilité gamétique. Ils sont d'ordre génomique (BAKRY *et al.*, 1990) — homologie partielle des deux génomes *acuminata* et *balbisiana* —, de nature chromosomique structurale — ces clones sont hétérozygotes structuraux —, ou chromosomique numérique — la triploidie de la majorité des cultivars conduisant à la formation de gamètes déséquilibrés. D'autres phénomènes qui peuvent être qualifiés de géniques entraînent des erreurs morphologiques et physiologiques : asynchronismes, décalage de réceptivité des fleurs.

## Les ressources génétiques

Le complexe d'espèces des bananiers présente des originalités qui entraînent une vision particulière de la gestion de ses ressources génétiques. La multiplication clonale, associée à de fréquentes stérilités ou incompatibilités, et la coexistence de plusieurs niveaux de ploïdie rendent indispensable une connaissance approfondie des potentialités de chaque clone et des relations de phylogénie entre les différents types reconnus, pour exploiter les ressources génétiques en amélioration variétale. De nombreux outils ont été développés à cet effet.

Les principaux programmes d'amélioration des bananiers à travers le monde possèdent des collections *ex situ*. Parmi les plus importantes et les mieux documentées, on peut citer celles du CRBP (Centre de recherches régionales sur bananiers et plantains), au Cameroun, de l'IITA (International Institute of Tropical Agriculture), au Nigeria, de la FHIA (Fundación Hondureña de Investigación Agrícola), au Honduras, de l'EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária), au Brésil, du MARDI (Malaysian Agricultural Research and Development Institute), en Malaisie, du CRIH (Central Research Institute for Horticulture), en Indonésie, et celle du Cirad, à la Guadeloupe. Elles renferment chacune entre 300 et 500 clones. Outre la conservation du patrimoine génétique, ces collections servent de base et de référence pour l'amélioration variétale. Les clones conservés *in vivo* dans chacune de ces collections dépendent des objectifs du programme de création variétale qui leur est associé. Les collections du Cameroun, du Nigeria et du Brésil sont largement tournées vers les variétés locales d'autoconsommation : Plantain en Afrique, Pome et Figue Pomme au Brésil. La FHIA et le Cirad ont des collections plus larges, plutôt vouées à l'amélioration des variétés d'exportation. La collection du Cirad est particulièrement riche en variétés diploïdes apparentées aux types dessert, matériel sur lequel repose le schéma de création d'hybrides triploïdes développé à la Guadeloupe.

À l'échelle internationale, l'INIBAP (International Network for the Improvement of Banana and Plantain) a un rôle de coordination entre les différents programmes de recherche, mais aussi une fonction d'intermédiaire pour l'échange de matériel certifié. L'INIBAP maintient au nom de la communauté internationale une collection *in vitro* de 1 100 accessions au centre international de transit installé sur le campus de l'université catholique de Leuven, en Belgique. Il a, d'autre part, mis en place deux centres d'indexation du matériel végétal pour sécuriser les échanges internationaux, l'un au QDPI (Queensland Department of Primary Industry), en Australie, et l'autre au Cirad, à Montpellier, en France.

Au cours des vingt dernières années, de nombreux marqueurs de la diversité ont été mis successivement au point et appliqués au bananier. Les descripteurs morphotaxonomiques ont été les premiers marqueurs développés et affinés chez les bananiers. Les variations des organes végétatifs portent principalement sur la couleur du pseudo-tronc, la présence de macules à la base des pétioles et leur couleur, la forme de la section du canal pétiolaire, la taille et le port de la plante. On reconnaît également des variations particulières dues au nanisme vrai (engorgement des plantes, aspect trapu des feuilles, inhibition des rejets) et à des chimères de couleurs. Les variations les plus importantes sont cependant celles de l'inflorescence et, par conséquent, du régime. La taille des fruits, leur forme, leur couleur ainsi que celle de la pulpe sont autant de critères qui permettent de différencier les fruits entre eux. Ainsi, les Plantain possèdent une pulpe jaune orangé, très ferme, que l'on ne retrouve pas chez les autres bananiers à cuire — Laknao, Popoulou, Bluggoe, Monthan (TEZENAS

DU MONTCEL, 1979). Les bananes d'Afrique de l'Est sont très spécifiques, elles sont destinées, selon les clones, à la cuisson ou à la fabrication de bière. Les parfums des bananes desserts sont variés ainsi que leurs goûts : très sucrés chez certains cultivars diploïdes comme Pisang Mas, doux acidulés dans le sous-groupe Figue Pomme, largement cultivé en Inde et au Brésil, standard et consensuel pour la banane de type Cavendish, dite d'exportation. La variabilité morphologique du bourgeon mâle s'exprime, pour sa part, par des différences dans la forme et la couleur des bractées et des pièces florales.

Un ensemble de 119 descripteurs agromorphotaxonomiques a été défini comme norme de description des bananiers (IPGRI-INIBAP et CIRAD, 1996). Ces descripteurs servent de base au système d'échange d'informations entre collections, MGIS (Musa Germplasm Information System), animé par l'INIBAP. Des outils informatiques d'aide à l'identification fondés sur ces descripteurs ont été développés (PERRIER et TEZENAS DU MONTCEL, 1990).

Ces 119 descripteurs ont été observés pour 273 clones de la collection du Cirad et pour 223 clones de la collection du CRBP, 99 clones complètement décrits étant communs aux deux collections. L'analyse statistique a été réalisée à partir d'un indice de dissimilarité pondéré par les probabilités d'erreur d'observation (voir annexe).

Depuis dix ans, différents types de marqueurs moléculaires ont été utilisés sur le bananier (LAGODA *et al.*, 1998) : polyphénols (HORRY, 1989), isoenzymes (LEBOT *et al.*, 1993), RFLP (JARRET et LITZ, 1986 ; CARREEL *et al.*, 1994) et, récemment, microsattellites (GRAPIN *et al.*, 1998). Au sein du Cirad, ces données moléculaires sont regroupées dans la base de données Tropgene (RABOIN et LAGODA, 1998). D'autres outils telles la cytométrie en flux et l'hybridation *in situ* ont également contribué à la caractérisation des ressources génétiques en permettant une détermination plus fine de la nature génomique des accessions étudiées : ploïdie, aneuploïdie et nombre de chromosomes affiliés à chacune des espèces d'origine (DOLEZEL *et al.*, 1994 ; JENNY *et al.*, 1997 ; HORRY *et al.*, 1998 ; D'HONT *et al.*, à paraître). Les résultats les plus complets concernent l'étude des génomes cytoplasmiques et nucléaire à l'aide des marqueurs RFLP.

Les deux génomes cytoplasmiques se caractérisent par une hérédité monoparentale. Au sein de l'espèce *M. acuminata*, FAURE *et al.* (1994) ont montré que le génome mitochondrial était transmis par voie paternelle et le génome chloroplastique, par voie maternelle. A ce jour, tous les hybrides monospécifiques et interspécifiques étudiés confirment cette transmission originale du génome cytoplasmique chez le bananier.

Les génomes cytoplasmiques et nucléaire de clones issus de la collection du Cirad à la Guadeloupe et de la collection *in vitro* de l'INIBAP ont été analysés par RFLP pour 3 couples enzyme-sonde chloroplastiques, 14 couples enzyme-sonde mitochondriaux et 51 couples enzyme-sonde nucléaires, repartis sur tout le génome. Au total, 115 accessions sauvages et 243 cultivars ont été étudiés. Cent cinquante-huit accessions, dont 89 diploïdes comprenant 34 sau-

vages et 69 triploïdes, ont fait l'objet d'une comparaison avec les descriptions morphologiques de la collection de Guadeloupe. La liste des clones étudiés en commun du point de vue tant morphologique que moléculaire est présentée en annexe (tableau 5).

Ces apports successifs ont permis de structurer les ressources génétiques des bananiers et d'en affiner la caractérisation. Ce chapitre présente une synthèse de ces travaux.

## La diversité des diploïdes

Les bananiers diploïdes, sauvages et cultivés, sont à l'heure actuelle beaucoup moins répandus que les triploïdes cultivés. On les trouve cependant encore à l'état endémique dans tout le Sud-Est asiatique. La présence de graines rend souvent leurs fruits impropres à la consommation et à la commercialisation. Ces bananiers sont généralement plus chétifs et donnent des rendements en fruits plus faibles que les triploïdes. Seuls les cultivars de type Figue Sucrée, qui ont de petits fruits à saveur très sucrée, sont cultivés en dehors de leur zone d'origine. Ces clones diploïdes sont toutefois indispensables pour les programmes d'amélioration génétique, notamment du fait de la forte stérilité des triploïdes.

### LES BANANIERES SAUVAGES

Les bananiers sauvages séminifères du genre *Musa* se rencontrent dans les vallées et les clairières humides mais bien drainées des forêts de faible et moyenne altitude, dans la zone tropicale, en Asie du Sud et du Sud-Est et dans le Pacifique, de la péninsule indienne aux îles Samoa. Plus de 25 espèces ont été décrites et rattachées au genre *Musa*. Seules celles qui ont contribué au génome des bananiers parthénocarpiques ont retenu notre attention.

Les espèces de la section *Australimusa* et *M. schizocarpa*, de la section *Eumusa*, sont présentes à l'est du berceau des *Musa* : est de l'Indonésie, Papouasie-Nouvelle-Guinée et Pacifique. Les *Australimusa* se reconnaissent à leur inflorescence érigée. Les différentes espèces de cette section ont été décrites comme apparentées et morphologiquement très proches par CHEESMAN (1947) et ARGENT (1976). Les 16 accessions étudiées à l'échelle moléculaire se sont révélées peu variables.

*M. schizocarpa* se caractérise par un tronc de couleur vert d'eau, identifiable chez les hybrides interspécifiques, ainsi que par la couleur verte des bractées du bourgeon mâle. Les deux accessions prospectées présentes en collection sont très similaires du point de vue tant morphologique que moléculaire.

*M. balbisiana*, de la section *Eumusa*, se rencontre de l'Inde aux Philippines en passant par la péninsule indochinoise. La représentation arborée des données morphotaxonomiques (figure 2) selon la méthode NJtree appliquée sur la



dissimilarité  $\Delta_{pp}$  montre bien la différenciation de *M. balbisiana* par rapport à *M. acuminata*, ainsi que leur faible variabilité (voir annexe). L'analyse isoenzymatique révèle également un polymorphisme réduit. HORRY (1989) classe en quatre types distincts les huit accessions présentes en collection, ce que confirme l'analyse par les RFLP. Il est étonnant que *M. balbisiana*, répandu sur une vaste aire géographique, soit si peu variable, mais cela peut être dû à un manque de prospections.

L'aire d'extension de *M. acuminata*, de la section *Eumusa*, couvre la majorité de l'aire de répartition du genre *Musa*, d'est en ouest, de la Birmanie à la Nouvelle-Guinée. Sa classification botanique fait référence à 7 sous-espèces. La représentation arborée des données morphotaxonomiques laisse apparaître quatre pôles (figure 2), qui correspondent à *M. a. malaccensis*, à *M. a. banksii*, à *M. a. microcarpa* et aux accessions des deux sous-espèces *M. a. siamea* et *M. a. burmanicca*. L'analyse par les RFLP nucléaires confirme ces résultats et permet d'identifier la nature intersubspécifique de certains représentants de

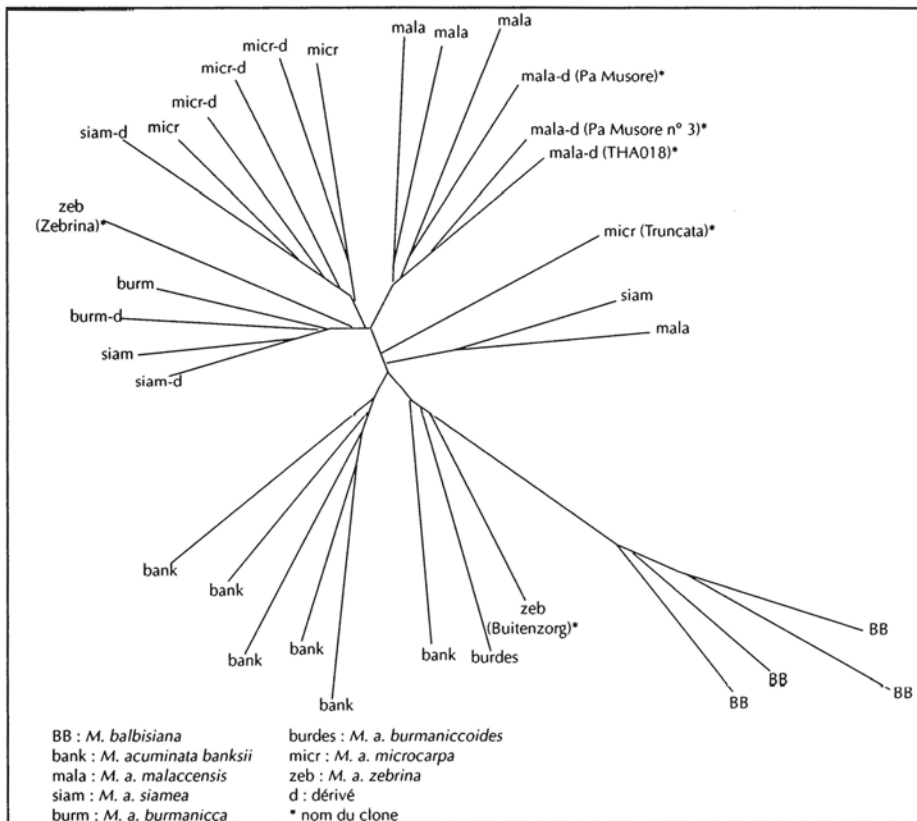


Figure 2. Diversité morphologique des bananiers séminifères (collection de Guadeloupe) : représentation arborée selon la méthode NJtree, réalisée sur une matrice de dissimilarités entre 32 accessions sur la base de 99 descripteurs.

*M. acuminata*. L'analyse en parallèle des deux types de marqueurs structure l'espèce en cinq pôles.

Le pôle *M. a. banksii* constitue un ensemble très homogène, du point de vue moléculaire plus encore que sur le plan morphologique. Ses représentants sont fortement autogames et ont divergé des autres sous-espèces de façon isolée, en Papouasie-Nouvelle-Guinée. Il est alors logique de constater qu'ils sont fortement homozygotes et morphologiquement à la fois originaux et proches.

Le pôle *M. a. burmanicca* et *M. a. siamea* comprend également la sous-espèce *burmanicoides*. Ces espèces sont originaires d'une vaste zone continentale, allant du nord de l'Inde à la Thaïlande. Quelques caractères morphotaxonomiques très particuliers empêchent de regrouper les sous-espèces *burmanicca* et *burmanicoides*, ce qui a conduit DE LANGHE et DEVREUX (1960) à les dissocier. L'analyse à l'aide des marqueurs moléculaires va dans le sens de SHEPHERD (1990), qui les regroupe.

Le cas des accessions de *M. a. zebrina* est un peu particulier : le Buitenzorg, jusqu'alors considéré comme l'holotype, est très atypique. L'analyse par RFLP prouve sa nature hybride. Il n'est pas étonnant dès lors de voir ces accessions diverger au sein de l'arbre. Le pôle *zebrina* est constitué en collection d'une seule accession, Zebrina.

Pour le pôle correspondant à *M. a. malaccensis*, il est nécessaire d'introduire la notion d'accessions dérivées d'une sous-espèce. Cette appellation découle de l'observation phénotypique d'accessions proches des holotypes, mais non identiques à ceux-ci. Ces accessions (THA018, Pa Musore...) sont génétiquement identifiées comme très proches de la sous-espèce d'origine par les marqueurs RFLP nucléaires et cytoplasmiques. Elles sont de plus fortement homozygotes et sont donc dès lors classées au sein même de *M. a. malaccensis*.

La même notion d'accessions dérivées existe pour *M. a. microcarpa*. Le regroupement est aisé du point de vue morphologique. Toutefois, ces accessions sont polymorphes du point de vue moléculaire et l'analyse par RFLP démontre leur nature d'hybride intersubspécifique. Il est quasiment reconnu que l'accession *Truncata*, classée précédemment au sein de *M. a. microcarpa*, constitue une sous-espèce à part entière sur le plan tant morphologique que moléculaire.

*M. acuminata* est donc l'espèce la plus variable et la mieux structurée. La topographie de son aire d'origine a conduit à un isolement géographique et donc reproductif, source d'une différenciation. Les modifications chromosomiques de type translocation observées par SHEPHERD (1987) entre les sous-espèces de *M. acuminata* sont à associer à cet isolement reproductif.

#### LES DIPLOÏDES CULTIVÉS

Pour les diploïdes cultivés, et malgré le grand nombre de clones connus, aucune structuration en sous-groupes n'a été officiellement établie. Les bananiers cultivés diploïdes présents en collection sont pour la plupart uniquement

apparentés à l'espèce *M. acuminata*. Sur 135 clones diploïdes de la collection du Cirad, seule une dizaine d'accessions ont été identifiées comme interspécifiques et classées dans les groupes AB, AS ou AT, grâce aux allèles spécifiques des espèces.

Les analyses factorielles, sur données morphotaxonomiques et moléculaires nucléaires, révèlent un nuage quasiment continu, sans dégager de structuration. En revanche, l'analyse des génomes cytoplasmiques et des taux d'hétérozygotie apporte des éléments de structuration.

Les cultivars AA se répartissent dans quatre des neuf profils chloroplastiques identifiés (notés de I à IX) lors de l'analyse de l'ensemble des *Musa*. De même, ils présentent 47 des 111 profils mitochondriaux identifiés. Ces derniers ont pu être regroupés en neuf ensembles de clones ayant des profils mitochondriaux comparables (notés de  $\alpha$  à  $\iota$ ). Les cultivars AA se répartissent dans cinq de ces ensembles mitochondriaux. L'information sur le génome chloroplastique, témoin d'un apparentement maternel, et celle sur le génome mitochondrial, à hérédité paternelle, sont complémentaires et peuvent être croisées (tableau 3). Les cultivars AA se répartissent ainsi en neuf types cytoplasmiques, ou cytotypes, mais la plupart d'entre eux correspondent à trois cytotypes : II- $\alpha$ , V- $\alpha$  ou V- $\Phi$ .

L'étude du génome nucléaire a également permis d'estimer le taux d'hétérozygotie à l'aide de 43 couples enzyme-sonde assimilables à des locus, et donc d'apprécier la complexité du génome et l'origine intersubspécifique ou non des clones. Deux groupes de clones peuvent être distingués au sein des trois principaux cytotypes. Par exemple, au sein des II- $\alpha$ , qui regroupent principalement des clones issus des prospections en Papouasie-Nouvelle-Guinée, plusieurs clones ont un faible taux d'hétérozygotie (de 10 à 20 %) et sont très apparentés aux *acuminata banksii*, sur le plan tant cytoplasmique que nucléaire. Les autres cultivars ont des taux d'hétérozygotie bien supérieurs (de 40 à 70 %), caractéristiques d'une origine intersubspécifique.

Les regroupements en cytotypes et les taux d'hétérozygotie permettent de définir douze ensembles de clones au sein des cultivars AA. Cette structuration se retrouve globalement lors de l'analyse du génome nucléaire ou de l'analyse morphologique. Toutefois, ces analyses montrent que les limites en sont parfois arbitraires car la variation est continue. Cette structuration permet cependant de définir les populations de base des programmes de sélection.

## Les bananiers triploïdes

Des analyses portant sur les données morphologiques et moléculaires ont été conduites sur les bananiers triploïdes (figure 3). Les deux types de marqueurs différencient nettement les cultivars triploïdes purement *acuminata* des cultivars interspécifiques. SIMMONDS et SHEPHERD (1955) ont fondé la classification des bananiers en groupes génomiques sur l'observation de quinze descripteurs morphotaxonomiques. L'utilisation de descripteurs plus complets et leur ana-

Tableau 3. Structuration des bananiers selon leur cytotype.

Profil chloroplastique	Profil mitochondrial									
	$\alpha$	$\beta$	$\chi$	$\delta$	$\epsilon$	$\phi$	$\gamma$	$\eta$	$\iota$	
I	<i>M. a. zebrina</i>									
II	<i>M. a. errans</i> AAcv (36)* AAA (Orotova, Red, Cavenish, Gros Michel) ABA (Mysore, Nadan, Pome)	AAcv (1)*	<i>M. a. burmanicca</i> <i>M. a. burmaniccoides</i> <i>M. a. siamea</i>	<i>M. a. malaccensis</i> AAcv (2)* AB ABA (Silk)	AAcv (1)*	AAcv (3)* AAA (3)**	<i>M. a. microcarpa</i>	ABB (Saba, Bluggoe, Ney, Mannan)		
III	AAcv (4)* AAA (1)**	<i>M. a. malaccensis</i>								
IV	AAcv (1)*	<i>M. a. siamea</i>								
V	AAcv (49)* AAA (4)** ABA (Laknao, Maia Maoli) ABB (1)**				AAA (Lujugira-Mutika)	<i>M. a. banksii</i> AAcv (18)* ABA (Plantain, Popoulou, Laknao)	ASs (2)*	ABB (Pelipita, Saba)		
VI	AScv (1)*					ASs (1)* AScv (2)*	<i>M. schizocarpa</i>			
VII									<i>M. balbisiana</i> type 2	
VIII	BAA (P. Rajah Bulu)				BAA (P. Kelat)				<i>M. balbisiana</i> type 1, 3, 4 BAB (Peyan, Saba, P. Awak)	
IX									BAB (P. Awak)	

\* Nombre de clones ayant ce cytotype. \*\* Nombre de clones de sous-groupe indéterminé. En italique : bananiers séminifères. s : sauvage, cv : cultivé.

lyse par des méthodes statistiques multivariées permettent de retrouver aisément cette structuration en groupes génomiques, aussi bien sur les observations provenant des collections du Cameroun que sur celles de la Guadeloupe.

En revanche, les données moléculaires ne mettent pas en évidence une distinction aussi nette entre les AAB et les ABB que les marqueurs morphologiques. *M. balbisiana* se révèle peu polymorphe avec les marqueurs moléculaires et les deux génomes B des ABB sont rarement différenciés. Plusieurs

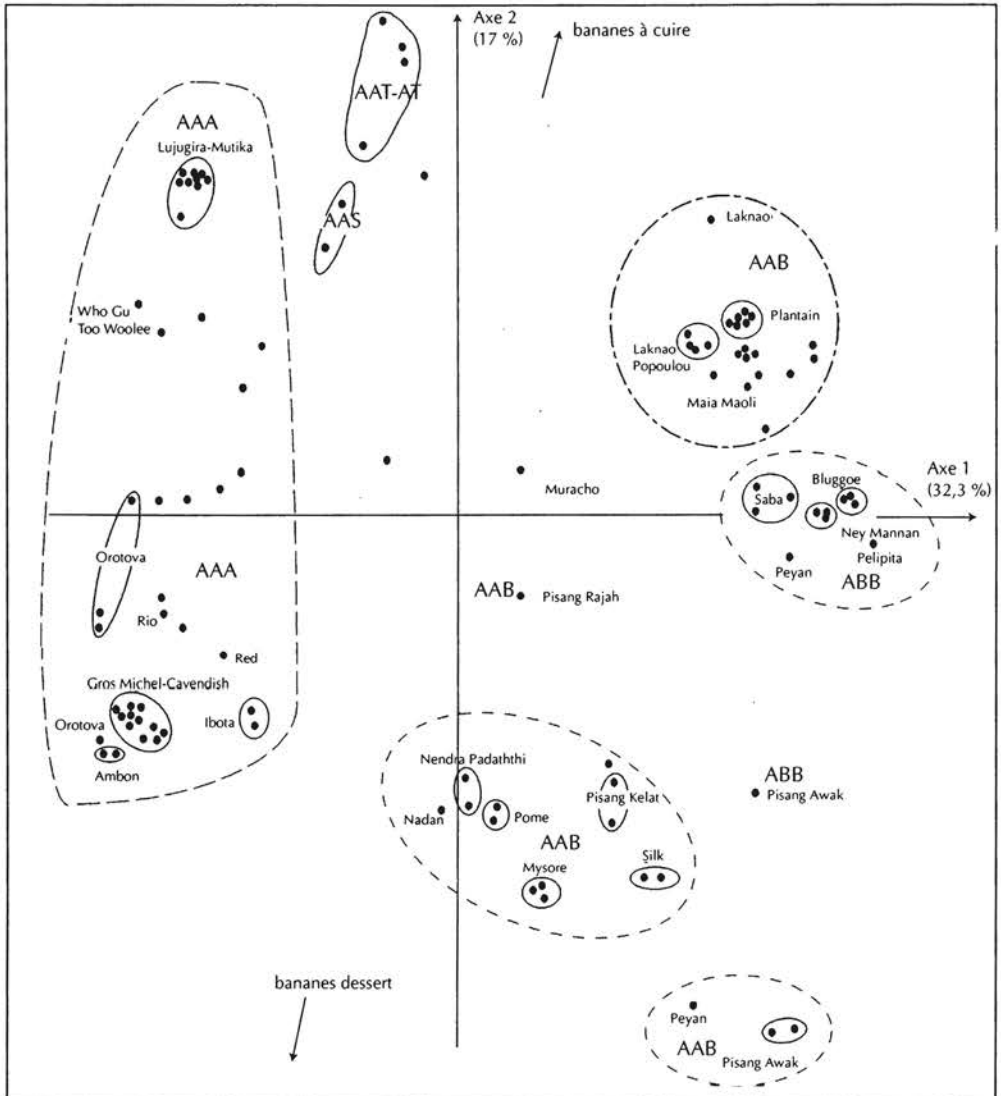


Figure 3. Diversité moléculaire nucléaire des bananiers triploïdes selon leur groupe et sous-groupe génomique : premier plan de l'analyse factorielle sur une matrice de distances de Jaccard entre 109 cultivars sur la base de 267 variables.

AAB ont également leurs deux génomes A quasi identiques. Dans ces deux cas, les clones AAB et ABB ont des profils moléculaires de type AB, et seule une lecture de l'intensité relative des bandes RFLP — qui n'est possible que pour quelques couples enzyme-sonde — permettrait de différencier clairement les AAB des ABB.

Si certains sous-ensembles se dégagent de l'analyse morphologique, on ne peut pas toujours identifier les sous-groupes aussi aisément que pour les Plantain ou les Cavendish. D'une part, les variations environnementales influent fortement sur la stabilité des descriptions morphotaxonomiques. D'autre part, les mutations — aussi petites soient-elles — qui sont à l'origine de la différenciation des cultivars à l'intérieur des sous-groupes provoquent des modifications phénotypiques suffisamment fortes (le nanisme, par exemple) pour gommer tout espoir de structuration avancée. A l'inverse, certains phénomènes de convergence rendent illusoire ce type d'analyse à l'aide des seuls caractères morphotaxonomiques. L'un des exemples les plus frappants est probablement celui du cultivar Mnalouki, AAB des Comores, dont l'aspect rappelle à s'y méprendre un Plantain de type French : les marqueurs moléculaires et le simple goût du fruit prouvent cependant qu'il ne s'agit en rien d'un Plantain.

Cette structuration botanique en sous-groupes se retrouve plus au niveau moléculaire, les cultivars ayant des profils nucléaires et cytoplasmiques identiques ou très similaires dans la plupart des cas. L'existence de quelques sous-groupes, comme les AAA Orotova, est toutefois remise en question. Certains clones qui ne se rattachaient à aucun sous-groupe du point de vue morphologique s'apparentent désormais à l'un d'eux, c'est le cas du AAB Muracho associé aux Pisang Rajah. D'autres clones non regroupés précédemment ont été identifiés comme très proches entre eux, il en est ainsi des AAA, Who Gu et Too Woolee.

Il est également possible de distinguer les clones de type dessert des clones de type à cuire au sein des AAB, sur une base morphologique, et au sein des trois groupes AAA, AAB et ABB, grâce aux marqueurs moléculaires. Ainsi, parmi les AAB, les bananiers des sous-groupes Mysore, Pome-Prata et Pisang Kelat se différencient des bananiers typiquement à cuire : Plantain, Popoulou, Maia Maoli et Laknao. Il est à noter que le clone Pisang Raja Bulu du sous-groupe Pisang Rajah, de type mixte, a un profil intermédiaire entre les types desserts et les types à cuire. Cette structuration est imputable au génome A de chacun de ces sous-groupes.

La structuration des bananiers triploïdes est beaucoup plus aisée à établir que celle des clones diploïdes en raison du mode d'évolution des clones triploïdes. A ce stade, tous ces cultivars sont fortement stériles et la multiplication devient exclusivement végétative. La différenciation des clones entre eux ne se fait plus qu'à l'occasion de petites mutations, conduisant rapidement à l'identification de sous-groupes vrais. Le degré de variabilité à l'intérieur de chaque sous-groupe dépend dès lors de l'intensité avec laquelle chaque type de clone sera

utilisé et donc multiplié. La variabilité morphologique la plus forte est relevée dans deux sous-groupes particulièrement exploités en Afrique : les Plantain, dans toute la zone d'Afrique centrale et d'Afrique de l'Ouest, et les Lujugira, en Afrique de l'Est. En revanche, à l'intérieur de ces sous-groupes, les profils RFLP nucléaires et cytoplasmiques sont identiques ou très similaires.

## Les relations entre les différents niveaux de variabilité

Les croisements réalisés en sélection font intervenir des clones de ploïdie différente. Les hybrides tétraploïdes sont issus du croisement entre le cultivar triploïde à améliorer et un géniteur diploïde porteur de résistances. Le schéma élaboré par le Cirad vise à produire des cultivars triploïdes proches des variétés cultivées à partir de géniteurs diploïdes (BAKRY *et al.*, 1997). Dans ces deux situations, une bonne connaissance de l'origine des bananiers cultivés est indispensable. Chez le bananier, la compréhension des processus de domestication — parthénocarpié, stérilité et processus d'évolution vers la triploïdie — est également nécessaire pour mieux gérer et exploiter les ressources génétiques. A travers l'étude de l'organisation génétique des bananiers, on cherche à identifier les populations de diploïdes actuels les plus apparentées aux différents types de cultivars triploïdes.

Plusieurs ressemblances morphologiques entre clones diploïdes et triploïdes sont connues. Ainsi, le régime de plusieurs cultivars AA — Pongani, par exemple — qui proviennent de prospections en Papouasie-Nouvelle-Guinée rappelle celui des Plantain. Le goût et le type de consommation d'autres cultivars AA, comme IDN110, les apparentent aux triploïdes de type Figue Pomme. Ces apparentements ont été confirmés par l'analyse moléculaire des génomes cytoplasmiques et nucléaire, et d'autres rapprochements ont été mis en évidence.

L'analyse des génomes cytoplasmiques, présentée pour les clones diploïdes cultivés, a été étendue à l'ensemble des clones sauvages et cultivés, diploïdes ou triploïdes. Dans le tableau 3, deux clones figurant sur une même ligne sont donc apparentés du côté maternel et deux clones placés dans une même colonne sont apparentés du côté paternel, indépendamment de leur niveau de ploïdie. Beaucoup de triploïdes sont du même groupe cytoplasmique que des diploïdes sauvages ou cultivés. Ainsi, les cultivars des sous-groupes Cavendish, Gros Michel, Ambon, Rio et Ibota ont les mêmes profils cytoplasmiques que plusieurs cultivars AA. Les clones AAB de type dessert du sous-groupe Figue Pomme ont le même profil cytoplasmique que les AB de type Figue Pomme. Les clones AAB de type à cuire des sous-groupes Plantain, Popoulou et Laknao ont leurs deux génomes cytoplasmiques identiques à celui de *M. a. banksii*. Cet apparentement entre les deux génomes A des Plantain et *M. a. banksii* avait été mis en évidence par des observations morphologiques (TEZENAS DU MONTCEL, 1990) ; il a été confirmé tant par l'analyse du génome à l'aide des isoenzymes (HORRY, 1989)

que par le marquage RFLP. Pour trois sous-groupes de triploïdes, aucun clone diploïde n'appartient au même groupe cytoplasmique, ni n'a le même profil mitochondrial. Parmi eux, les clones AAA d'Afrique de l'Est, qui appartiennent au sous-groupe Lujugira-Mutika, s'apparentent du côté maternel à *M. a. banksii* et du côté paternel à *M. a. zebrina*, ce qui démontre une origine intersubspécifique, confirmée par l'analyse de leur génome nucléaire. Cette origine bisubspécifique a également été trouvée pour les cultivars AA d'origine géographique voisine, de Madagascar ou des Comores.

L'ensemble des clones a également été analysé à l'aide des mêmes marqueurs RFLP nucléaires. La projection des données relatives aux triploïdes cultivés sur la structuration obtenue pour *M. acuminata* et les diploïdes cultivés met en évidence les apparentements entre les génomes A des clones sauvages, des diploïdes cultivés et des triploïdes (figure 4). Il est intéressant de constater que les cultivars diploïdes et triploïdes des ensembles proches du pôle *banksii*, *errans* et *microcarpa* Bornéo ont des fruits farineux et donc consommés cuits,

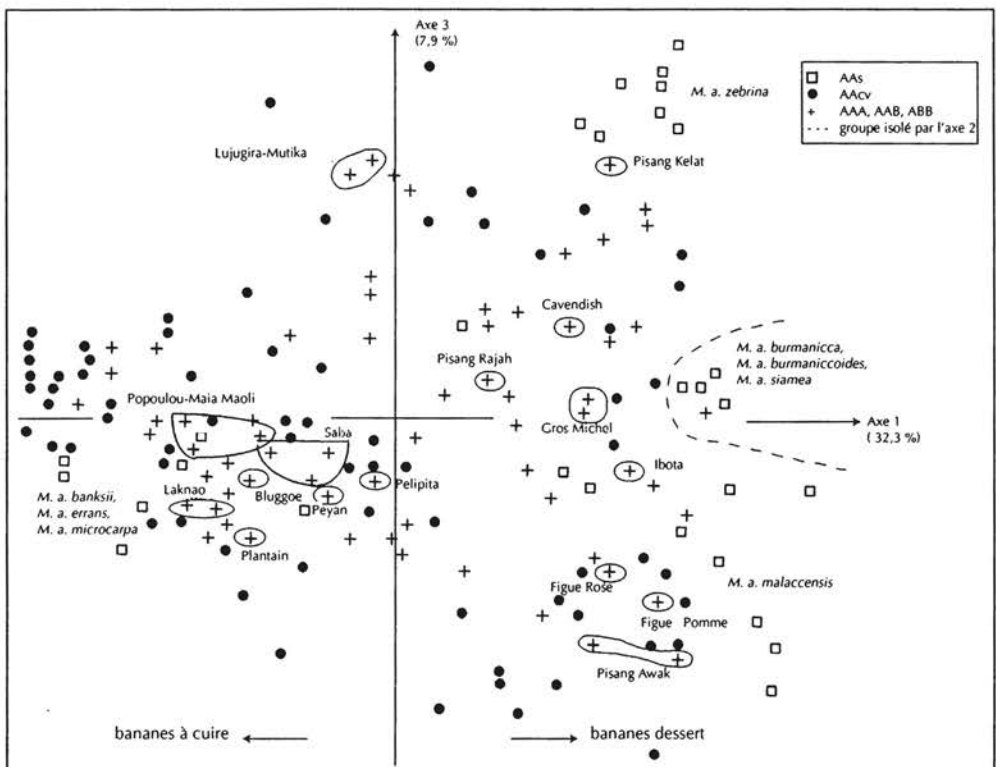


Figure 4. Apparentement moléculaire nucléaire entre les génomes A des bananiers triploïdes avec les diploïdes sauvages et cultivés : projection des données relatives aux triploïdes sur le plan 1-3 de l'analyse factorielle sur une matrice de distances de Jaccard des diploïdes sauvages et cultivés sur la base des données issues des 19 couples enzyme-sonde étudiés qui révèlent des différences significatives entre les cytotypes V- $\alpha$  et V- $\Phi$ .



tandis que les bananiers des ensembles proches des pôles *malaccensis* et *zebrina* ont des fruits à saveur plus sucrée.

L'étude du génome nucléaire a également permis d'estimer les taux d'hétérozygotie à partir d'un profil à deux niveaux de bandes, pour l'ensemble des clones, et d'un profil à trois niveaux de bandes, pour les triploïdes, à l'aide de 43 couples enzyme-sonde assimilables à des locus. Ces taux permettent d'apprécier la complexité du génome des triploïdes à améliorer. Les différents génomes A d'un même génotype sont plus apparentés — leur taux d'hétérozygotie est plus faible — pour les types à cuire que pour les bananiers de type dessert, qui auraient donc une origine plus complexe.

L'ensemble de ces données cytoplasmiques et nucléaires sur diploïdes et triploïdes permet de cibler les géniteurs diploïdes les plus proches des triploïdes à améliorer. Ainsi, sur la base des marqueurs moléculaires, des schémas de croisement à partir de diploïdes, doublés ou non à la colchicine, ont été établis afin d'obtenir des hybrides proches des triploïdes à améliorer, tels que ceux qui sont proposés par RABOIN *et al.* (à paraître) pour les Gros Michel et Cavendish.

## Conclusion

A partir de l'ensemble des données morphologiques et moléculaires, des hypothèses quant à l'origine de la parthénocarpie et des principaux sous-groupes de triploïdes ont été avancées.

## La dispersion et l'évolution des bananiers

SIMMONDS (1962) place le centre de domestication primaire des bananiers cultivés dans la région malaise. Or, plusieurs résultats nous font supposer que ce sont les génomes des accessions *M. a. banksii* de Papouasie-Nouvelle-Guinée et de l'accession Agutay des Philippines, représentant *M. a. errans*, qui sont à l'origine de la parthénocarpie et donc des bananiers cultivés. En effet, les génomes cytoplasmiques ou nucléaire de ces deux sous-espèces ont été identifiés chez quasiment tous les bananiers parthénocarpiques, diploïdes ou triploïdes. De plus, les distances génétiques sont minimales entre *M. a. banksii* et les seuls cultivars AA homozygotes. SIMMONDS (1962) signale d'ailleurs chez un représentant de la sous-espèce *banksii* un développement important de la pulpe, qui serait un indice de parthénocarpie. La région des Philippines et de la Nouvelle-Guinée, d'où sont originaires les deux sous-espèces *banksii* et *errans*, apparaît donc comme le centre de domestication primaire.

Les fruits de certains bananiers sauvages étaient consommés immatures avant que les graines ne deviennent dures. Les bananiers *M. a. banksii* qui présentaient un début de parthénocarpie ont dû être sélectionnés par l'homme. Transportés en Asie du Sud-Est, ils se sont hybridés avec d'autres sous-espèces de *M. acuminata* et ont ainsi acquis les hétérozygoties génétique et structurale qui caractérisent l'ensemble des cultivars. Ils ont intégré des gènes de stérilité et des gènes régulateurs de la parthénocarpie pour donner les premiers diploïdes comestibles. Pour ces diploïdes, la stérilité des cultivars a dû limiter l'obtention d'une unité panmictique par région et la multiplication végétative a conservé des formes intermédiaires. Ainsi, malgré l'isolement géographique, aucune structuration forte n'apparaît. La fertilité partielle de ces cultivars a toutefois permis, peu à peu et sous l'action de l'homme, l'établissement d'un cline. La multiplication végétative a permis de conserver les génotypes issus de chacune de ces étapes.

Les Plantain AAB, présents dans toute la zone forestière de l'Afrique (CHAMPION, 1967), seraient d'introduction très ancienne sur ce continent. Cette introduction aurait une origine austronésienne et daterait d'au moins 3000 ans avant notre ère (DE LANGHE, 1995). Champion note qu'elle a eu lieu à une époque où les autres bananiers n'existaient pas, sinon ils auraient également été introduits en Afrique. Ces bananiers, tout comme les Popoulou aussi appelés Plantain du Pacifique, ont leurs deux génomes A très proches de celui de *M. a. banksii* et ne sont apparentés à aucune autre sous-espèce. Les bananiers parthénocarpiques de la sous-espèce *banksii* ont pu se croiser avec *M. balbisiana* dans les aires de recouvrement des deux espèces. Les différences génomiques entre *M. a. banksii* et *M. balbisiana* sont suffisantes pour que leurs hybrides aient donné des gamètes diploïdes issus d'une simple restitution. Les AAB issus de ces quelques croisements se seraient répandus vers l'ouest, en Inde puis en Afrique, où ils se sont différenciés et propagés. L'Afrique centrale est donc le centre de diversification secondaire des Plantain. Les croisements entre AA d'origine *banksii* et *M. balbisiana* ont également donné naissance aux Laknao, aux Philippines, et aux Popoulou-Maia Maoli, dans le Pacifique.

Les AAA Lujugira-Mutika n'ont été observés qu'en Afrique de l'Est. Ils sont apparentés aux deux sous-espèces *banksii* et *zebrina*. Ces clones ou les diploïdes dont ils sont issus auraient été introduits directement d'Indonésie, zone d'origine de *M. a. zebrina*, peut-être via les Comores et la Tanzanie, comme Simmonds en a fait l'hypothèse. Transportés en Indonésie par les Austronésiens, les bananiers partiellement parthénocarpiques *banksii* provenant de Papouasie-Nouvelle-Guinée ont pu se croiser avec *M. a. zebrina*. Les hybrides étant des hétérozygotes structuraux, leur stérilité partielle a pu être suffisante pour donner les gamètes diploïdes nécessaires à la formation des triploïdes.

La formation des cultivars triploïdes de type dessert repose sur l'existence de cultivars hybrides entre les génomes *banksii*, *errans*, *malaccensis* et *zebrina*. Des cultivars diploïdes puis triploïdes qui associent ces génomes ont pu apparaître progressivement en Malaisie. Cette région serait donc, selon notre

hypothèse, un centre de domestication secondaire et non, comme Simmonds l'a proposé, le centre primaire.

## La conservation des ressources génétiques

Les clones cultivés prospectés — un millier, ce qui est peu comparé à d'autres plantes — sont conservés *in vitro* et gérés par l'INIBAP. Chaque centre de recherche maintient une collection *in vivo* de clones représentatifs de la diversité et de nombreux clones correspondant au type de fruit qu'il cherche à améliorer. Dans le milieu naturel, la variabilité est surtout d'origine somaclonale en raison de la forte stérilité des cultivars. La variabilité issue de la recombinaison résulte essentiellement des croisements forcés réalisés lors de la création variétale. Une bonne caractérisation des clones créés et sa comparaison avec celle des clones d'origine naturelle devraient permettre de vérifier qu'il n'y a pas dérive génétique.

La situation est tout autre pour les bananiers séminifères. De nombreuses espèces ou sous-espèces du genre *Musa*, très peu prospectées, ne sont représentées en collection que par des holotypes, qui sont maintenus par multiplication végétative. La prospection de ces variétés est rendue difficile par la situation géographique et politique de leurs pays d'origine, mais devient urgente du fait de la déforestation accélérée des zones qui les abritent. Nombre de populations sauvages risquent, à plus ou moins long terme, de disparaître. La conservation et la connaissance de ces variétés sont cependant indispensables pour les programmes d'amélioration futurs, qui rechercheront de nouvelles sources de résistance aux menaces parasitaires.

## Annexe

### Choix d'une distance pour l'analyse des descripteurs morphotaxonomiques

L'une des caractéristiques des descripteurs morphotaxonomiques est leur subjectivité. Leur notation est en effet liée, d'une part, à l'expression du caractère chez la plante en fonction du milieu, d'autre part, à la notation de cette expression par l'observateur. Malgré tous les efforts qui peuvent être apportés à la réduction de ces variations induites, leur élimination est impossible. Ces variations constituent cependant un moyen d'accès aux réponses de la plante à cet environnement. Une dissimilarité particulière, tenant compte de ces erreurs d'appréciation, a été choisie.

### Définition de la dissimilarité $\Delta_{pp}$

Les descripteurs retenus sont tous de type qualitatif. Afin d'affiner les réponses, une matrice de probabilité d'erreur a été attribuée à chaque descripteur. Dans le logiciel Musaid (PERRIER ET TEZENAS DU MONTCEL, 1990), l'utilisation de cette matrice permet de ne pas éliminer systématiquement une réponse négative, mais plutôt de la conserver en lui attribuant un coefficient de probabilité. Ainsi dans le tableau 4, qui concerne la notation de la couleur de la face ventrale de la nervure, la notation vert clair pourrait aisément avoir été notée ou confondue avec jaune ou verte, alors que la notation noire est sans équivoque.

Tableau 4. Matrice de probabilité pour le descripteur « couleur de la face ventrale de la nervure ».

Réponse	Probabilité d'identité					
1 - jaune	43	30	10	05	01	01
2 - vert clair	30	33	30	05	01	01
3 - verte	20	30	53	05	01	01
4 - rose	05	05	05	64	10	01
5 - rouge violacé	01	01	01	01	86	01
6 - noire	01	01	01	01	01	95

L'indice de dissimilarité choisi dérive de l'indice de Sokal et Michener (ou *simple matching coefficient*, noté  $\Delta_{SM}$ ) qui, pour deux individus, est égal au rapport du nombre de réponses égales sur le nombre de questions comparées. Ici notre indice, noté  $\Delta_{pp}$ , est pondéré par les probabilités d'erreur à l'observation. La contribution du descripteur  $v$  à la dissimilarité entre deux individus  $i$  et  $j$  s'exprime ainsi :

$$\Delta_v = \frac{1}{2} \left[ \frac{P_v(v_i, v_i) - (P_v(v_i, v_j) + P_v(v_j, v_i))}{P_v(v_i, v_i) + (P_v(v_i, v_j) + P_v(v_j, v_i))} + \frac{P_v(v_j, v_j) - (P_v(v_j, v_i) + P_v(v_i, v_j))}{P_v(v_j, v_j) + (P_v(v_j, v_i) + P_v(v_i, v_j))} \right]$$

où  $P_v$  représente la matrice de probabilité associée au descripteur  $v$ , et où  $v_i$  et  $v_j$  correspondent aux réponses des deux individus pour ce descripteur. La dissimilarité globale  $\Delta_{pp}$  entre  $i$  et  $j$  est égale à la moyenne des  $\Delta_v$  sur tous les descripteurs.

$\Delta_v$  varie entre 0 et 1, il est nul lorsque les deux individus ont la même modalité de réponse et d'autant plus grand que les probabilités attachées aux réponses  $v_i$  et  $v_j$  sont différentes, sans toutefois atteindre 1.

Ainsi, pour le descripteur concernant la couleur de la face ventrale de la nervure, on obtient si  $v_i = 2$  (vert clair) et  $v_j = 3$  (vert) :

$$\Delta_{pp} = \frac{1}{2} \left( \frac{33 - 30}{33 + 30} + \frac{53 - 30}{53 + 30} \right) = 0,16$$

si  $v_i = 2$  (vert clair) et  $v_j = 6$  (noir) :

$$\Delta_{pp} = \frac{1}{2} \left( \frac{33 - 1}{33 + 1} + \frac{95 - 1}{95 + 1} \right) = 0,96$$

Dans ces deux situations, l'indice  $\Delta_{SM}$  est égal à 1.

L'indice  $\Delta_{pp}$  tend à minimiser les différences observées entre individus. Il présente l'avantage d'éviter une augmentation excessive de l'indice calculé par accumulation de différences minimales entre critères morphotaxonomiques, qui sont souvent délicats à évaluer. Ce gain de précision pour les faibles distances est un avantage recherché puisque ces dernières conditionnent entièrement le déroulement des algorithmes ascendants de classification.

La représentation graphique de  $\Delta_{pp}$  en fonction de  $\Delta_{SM}$  indique que la corrélation linéaire entre les indices est bonne ( $r = 0,896$ ). Les deux indices ont une tendance nette à mesurer dans le même sens les divergences entre individus. Les points extrêmes sont les mêmes, mais les valeurs sont en moyenne divisées par un facteur de l'ordre de 2 avec l'indice  $\Delta_{pp}$ .

### Comparaison de l'indice $\Delta_{pp}$ dans les différentes collections

Nous avons utilisé pour cette analyse les descripteurs officiels des bananiers (IPGRI-INIBAP et CIRAD, 1996). Les observations ont été réalisées sur deux collections, celle du Cameroun (CRBP) et celle de la Guadeloupe (Cirad). Afin de conserver un nombre maximal d'individus, les descripteurs comportant le plus de réponses manquantes ont été exclus pour n'en retenir que 92 sur 119. Seuls les bananiers pour lesquels les descriptions étaient complètes dans les deux collections ont été utilisés. Ainsi, 99 accessions ont été retenues : 58 diploïdes et 41 triploïdes.

Les moyennes de l'indices  $\Delta_{pp}$  ont été calculées par collection et par groupe génomique. L'observation des mêmes cultivars sur deux sites permet de distinguer et de quantifier la variabilité observée entre différents cultivars au sein d'une collection (intercultivar) de la variabilité entre collections pour un même cultivar (intersite). La comparaison des résultats permet de dégager quelques points marquants.

La divergence intersite (ou intracultivar) est inférieure à la divergence intrasite (ou intercultivar) pour les collections en général et, plus précisément, pour les groupes AAA et AAB. A ces échelles, l'environnement n'a pas d'influence sur la classification des clones en groupe génomique. En revanche, la divergence devient supérieure pour les sous-groupes. Pour un sous-groupe dans lequel l'homogénéité entre cultivars est grande, les variations liées à l'environnement ont plus d'influence que les différences entre clones. C'est le cas des Cavendish, qui diffèrent entre eux essentiellement par des variations de taille souvent très fines. L'influence de l'environnement sur ces variations est forte, et la divergence sera plus grande entre les descriptions du clone Grande Naine réalisées au Cameroun et à la Guadeloupe qu'entre la description d'une Grande Naine et celle d'un Williams dans une même collection. Les clones répondant dans le même sens à l'influence du milieu, la variation à l'intérieur d'un même site est éventuellement déplacée, mais elle n'est pas amplifiée.

L'indice  $\Delta_{pp}$ , qui tient compte des erreurs possibles à l'aide de tableaux de probabilités, permet une détermination plus fiable que l'indice  $\Delta_{SM}$  du groupe génomique des accessions, même en utilisant les données d'une autre collection. En revanche, pour l'identification d'un cultivar, il n'est pas toujours possible d'utiliser une base de référence (Cameroun, par exemple) pour déterminer une accession décrite dans une autre base (Guadeloupe), surtout lorsqu'il s'agit d'accessions issues de mutations somaclonales. Cela conforte les recommandations quant à l'utilisation du logiciel Musaid pour identifier une accession.

Tableau 5. Liste des clones étudiés en commun du point de vue tant morphologique que moléculaire.

Section • Groupe Espèce	Sous-groupe Sous-espèce	Clone Variété	GLP	
			CMR*	RFLP**
<i>Rhodochlamys</i>				
• <i>Musa velutina</i>		Velutina	x	x
• <i>Musa laterita</i>		Jamaïque		x
<i>Eumusa</i>				
• <i>Musa acuminata</i>	<i>malaccensis</i>	Pahang, Sélangor	x	x
		Cici (Brésil), Malaccensis		x
	<i>burmanicca</i>	Long Tavoy		x
	<i>burmanicoides</i>	Calcutta 4	x	x
	<i>banksii</i>	Banksii, Madang, Paliama	x	x
		Hawain 2, Higa, Waigu		x
	<i>siamea</i>	Khæ (Phrae), Pa Rayong	x	x
	<i>microcarpa</i>	Bornéo, Microcarpa, Truncata	x	x
	<i>zebrina</i>	Zebrina	x	x
		Buitenzorg		x
	<i>malaccensis</i> dérivé	Pa (Musore) n° 3, THA018	x	x
		Pa (Musore)		x
	<i>burmanicca</i> dérivé	Pisang Prentel		x
	<i>siamea</i> dérivé	Pa (Songhkla)	x	x
		Pa (Abyssinea)		x
	<i>microcarpa</i> dérivé	Pisang Cici Alas	x	x
		AAs-IDO113, EN13		x
• <i>Musa balbisiana</i>		Balbisiana (CMR), Balbisiana (HND), Klue Tani, Pisang Batu	x	x
• AA cultivés		Akondro Mainty, Ato, Bie Yeng, Chicame, Galeo, Gorop, Guyod, IDN110, Kenar, Khai Nai On, Kirun, M48, M53, Manameg Red, Mapua, Niyarma Yik, Pa (Patthalong), Pisang Bangkahulu, Pisang Berlin, Pisang Sasi, Pongani, Sa, Samba, Sepi, SF215, SF265, Sowmuk, Thong Det, Tomolo, To'o, Tuu Gia, Vudu Papua, Wikago	x	x

Tableau 5. *Suite.*

Section	Sous-groupe	Clone	GLP	
			CMR*	RFLP**
• Groupe	Sous-espèce	Variété		
Espèce				
<i>Eumusa (suite)</i>				
• AA cultivés		Bebeck, Beram, Dibit, Fako Fako, Gu Nin Chiao, Hom, Inori, Khi Maeo, Kumburgh (Kunburg), Manang, N° 110-THA052, Padri, Pallen Berry, Pisang Gigi Buaya, Pisang Lilin, Pisang Sapon, Pitu, Saing Todloh, Ta, Wudi Yali Yalua		x
• AB cultivés		Figue Pomme (Ekona), Safet Velchi Kunnan	x	x
• AAA	Gros Michel	Cocos, Gros Michel	x	x
		Bout Rond, Dougoufoui, Highgate		x
	Cavendish	Americani, Mutant Rouge, Padji, Seredou	x	
		Grande Naine, Petite Naine, Poyo	x	x
		IRFA901, Williams		x
	Figue Rose	Figue Rose Naine		x
	Orotova	Pisang Sri	x	x
		Hom (Sakhon Nakhon), Orotova, Pisang Kayu, Pisang Umbuk		x
	Ibota	Khom Bao, Yangambi km5	x	x
	Lujugira	Foulah	x	x
		Bui Se Ed, Nakitengwa, Nshika		x
	Ambon	Hom (Thong Mokho), Pisang Ambon, Pisang Bakar		x
	AAA indéterminé	Ouro Mel, Too Woolee, Who Gu	x	x
		Lagun Vunalir, Palang		x
• AAB	Plantain	Amou, Kelong Mekintu, Kwa, Madre del Platano	x	
		Big Ebanga		x
	Pisang Kelat	Pisang Kelat, Pisang Pulut		x
	Pisang Rajah	Pisang Raja Bulu	x	x
	Popoulou	Poingo, Popoulou (CIV)		x
	Nendra Padaththi	Pisang Rajah	x	x
		Rajapuri India		x
	Mysore	Pisang Ceylan	x	x
	Figue Pomme	Kingala 1, Supari	x	
		Figue Pomme Naine		x
	Nadan	Lady Finger		x
	Pome	Guindy, Rois	x	
		Foconah, Prata Ana		x
	Laknao	Laknao	x	x
		Kune, Mugus		x

Tableau 5. Suite.

Section • Groupe Espèce	Sous-groupe Sous-espèce	Clone Variété	GLP		
			CMR*	RFLP**	
<i>Eumusa (suite)</i>					
• AAB (suite)	AAB indéterminé	Mnalouki	x		
		Pisang Nangka	x	x	
• ABB	Bluggoe	Kupulik, Muracho, Teeb Kum, Tomnam		x	
		Pisang Awak	Dole, Poteau Nain	x	
			Bom, Praha	x	
	Monthan	Namwa Knom, Pisang Kepok		x	
		Monthan	x	x	
		Ney Mannan	Ice Cream	x	x
	Radjah			x	
	Peyan	Brazza IV, Pisang Kepok Bung		x	
		Saba	IDN107		x
	ABB indéterminé		Saba	x	x
		Auko-PNG034, Dwarf Kalapua	x	x	
		Auko-PNG125, Klue Tiparot***		x	
• AAAA		Champa Nasik	x		
• AAAB		Langka 08	x	x	
		Ouro da Mata, Platina	x		
• AABB		Pisang Slendang		x	
• ABBT		Yawa 2		x	
Total			99	158	
dont	diploïdes cultivés		37	55	
	diploïdes sauvages		21	34	
	triploïdes et tétraploïdes		41	69	

\* Population commune aux collections de la Guadeloupe et du Cameroun, descripteurs morphotaxonomiques.

\*\* Accessions de Guadeloupe décrites à la fois par marqueurs RFLP et descripteurs morphotaxonomiques.

\*\*\* Nouvelle classification selon JENNY *et al.* (1997).



## Références bibliographiques

- ARGENT G.C.G., 1976. The wild bananas of Papua New Guinea. Royal Botanic Garden Edinburgh, 35(1) : 77-114.
- BAKRY F., CARREEL F., CARUANA M.L., COTE F.X., JENNY C., TEZENAS DU MONTCEL H., 1997. Les bananiers. In : L'amélioration des plantes tropicales, A. Charrier et al. éd., Montpellier, France, Cirad-Orstom, collection Repères, p. 109-139.
- BAKRY F., HORRY J.P., TEISSON C., TEZENAS DU MONTCEL H., 1990. L'amélioration génétique des bananiers à l'Irfa-Cirad. Fruits, n° spéc. : 25-40.
- CARREEL F., NOYER J.L., GONZALEZ DE LEON D., LAGODA P.J.L., PERRIER X., BAKRY F., TEZENAS DU MONTCEL H., LANAUD C., HORRY J.P., 1994. Evaluation de la diversité génétique chez les bananiers diploïdes (*Musa* sp.). Génétique, sélection, évolution, 26(suppl. 1) : 125s-136s.
- CHAMPION J., 1967. Les bananiers et leur culture. 1. Botanique et génétique. Paris, France, IFAC, 212 p.
- CHEESMAN E.E., 1947. Classification of the bananas. 2. The genus *Musa* L. Kew Bulletin, 2 : 106-117.
- D'HONT A., PAGET-GOY A., ESCOUTE J., CARREEL F. The interspecific genome structure of cultivated banana, *Musa* spp. revealed by genomic DNA *in situ* hybridization (à paraître).
- DOLEZEL J., DOLEZELOVA M., NOVAK F.J., 1994. Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). Biologia Plantarum, 36(3) : 351-357.
- FAURE S., NOYER J.L., CARREEL F., HORRY J.P., BAKRY F., LANAUD C., 1994. Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). Current Genetics, 25 : 265-269.
- GRAPIN A., NOYER J.L., CARREEL F., DAMBIER D., BAURENS F.C., LAGODA P.J.L., 1998. Diploid *Musa acuminata* genetic diversity with STMS. Electrophoresis, 19 : 1374-1380.
- HORRY J.P., 1989. Chimiotaxonomie et organisation génétique dans le genre *Musa*. Thèse de doctorat, université Paris XI, Orsay, France, 105 p.
- HORRY J.P., DOLEZEL J., DOLEZELOVA M., LYSAK M.A., 1998. Les bananiers naturels tétra-ploïdes A × B existent-ils ? InfoMusa, 7(1) : 5-6.
- IPGRI-INIBAP, CIRAD, 1996. Descripteurs pour le bananier (*Musa* spp.). Rome, Italie, IPGRI, 55 p.
- JARRET R.L., LITZ R.E., 1986. Enzyme polymorphism in *Musa acuminata* Colla. Journal of Heredity, 77 : 183-188.
- JENNY C., CARREEL F., BAKRY F., 1997. Revision on banana taxonomy: Klue Tiparot (*Musa* sp.) reclassified as a triploid. Fruits, 52(2) : 83-91.
- LAGODA P.J.L., AMSON C., AUBERT G., AUBOIRON E., BAKRY F., BAURENS F.C., CARREEL F., COSTET L., DAMBIER D., FAURE S., FOURE E., GANRY J., GONZALES DE LEON D., GRAPIN A., HORRY J.P., JENNY C., NOYER J.L., OLLITRAULT P., RABOIN L.M., RESCHŒUR F., TEZENAS DU MONTCEL H., TOMEKPE K., LANAUD C., 1998. Acronyms to study the Musaceae, reviewed paper. Acta Horticulturae, n° 461 : 113-121.

- DE LANGHE E., 1995. Is plantain the oldest fruit crop in the world ? *In* : Celebration of the King Baudouin award to IITA. Ibadan, Nigeria, IITA, p. 44-47.
- DE LANGHE E., DEVREUX M., 1960. Une sous-espèce nouvelle de *Musa acuminata* Colla. *Bulletin du Jardin Botanique de Bruxelles*, 30 : 375-388.
- LEBOT V., ARADHYA M.K., MANSHARDT R.M., MEILLEUR B.A., 1993. Genetic relationships among cultivated bananas and plantains from Asia and the Pacific. *Euphytica*, 67 : 163-175.
- PERRIER X., TEZENAS DU MONTCEL H., 1990. Musaid, a computerized determination system. *In* : Identification of genetic diversity in the genus *Musa*, R.L. Jarret éd., Montpellier, France, INIBAP, p. 76-91.
- RABOIN L.M., CARREEL F., NOYER J.L., BAURENS F.C., HORRY J.P., BAKRY F., TEZENAS DU MONTCEL H., GANRY J., LANAUD C., LAGODA P.J.L. The putative common diploid ancestor of Cavendish and Gros Michel banana cultivars (à paraître).
- RABOIN L.M., LAGODA P.J.L., 1998. Eléments moléculaires d'aide à la décision pour la sélection de la banane et du plantain. Montpellier, France, Cirad-Agetrop.
- SHEPHERD K., 1987. Translocations in *Musa acuminata*. Document reprographié, 3 p.
- SHEPHERD K., 1990. Observations on *Musa* taxonomy. *In* : Identification of genetic diversity in the genus *Musa*, R.L. Jarret éd., Montpellier, France, INIBAP, p. 158-165.
- SIMMONDS N.W., 1962. The evolution of the bananas. Londres, Royaume-Uni, Longmans, 170 p.
- SIMMONDS N.W., SHEPHERD K., 1955. The taxonomy and origins of the cultivated bananas. *Journal of the Linnean Society of London, Botany*, 55 : 302-312.
- TEZENAS DU MONTCEL H., 1979. Les plantains du Cameroun : propositions pour leur classification et dénominations vernaculaires. *Fruits*, 34(2) : 83-97.
- TEZENAS DU MONTCEL H., 1990. *Musa acuminata* subspecies *banksii*: status and diversity. *In* : Identification of genetic diversity in the genus *Musa*, R.L. Jarret éd., Montpellier, France, INIBAP, p. 211-218.



# Le cacaoyer

---

Claire Lanaud, Juan-Carlos Motamayor,  
Olivier Sounigo

La domestication du cacaoyer est ancienne. Sa culture était pratiquée il y a 2 000 à 3 000 ans par les civilisations précolombiennes, en particulier par les Mayas et les Aztèques (PARADIS, 1979). Les fèves de cacao étaient utilisées dans leur breuvage, appelé « chocolatl », où étaient mélangés et broyés cacao, maïs, piment et autres plantes aromatiques. Les fèves de cacao étaient si appréciées qu'elles constituaient aussi une monnaie d'échange. Avant même l'arrivée des Espagnols, le cacao a voyagé en empruntant les routes de commerce des Mayas et des Aztèques, mais aussi grâce aux Pipil-Nicaraos (YOUNG, 1994 ; COE et COE, 1996). En 1585, les Espagnols ont expédié du cacao en Espagne, où le secret du chocolat — une version sucrée du breuvage — a été gardé pendant une quarantaine d'années. La mode de cette boisson chocolatée s'est ensuite répandue en Europe. Pour répondre à une demande croissante, les plantations de cacaoyers se sont étendues en Amérique centrale et de nouvelles plantations ont été mises en place dans plusieurs îles de la Caraïbe, comme à la Trinité, en 1525, et à la Jamaïque. Des cacaoyers provenant d'Amérique centrale, en particulier du Costa Rica, ont été introduits au Venezuela par les Espagnols, mais il est possible que des cacaoyers aient déjà été cultivés dans la région sud-ouest du pays avant la colonisation espagnole (PITTIER, 1933 ; BERGMAN, 1969). Vers 1750, les Français ont planté des cacaoyers à la Martinique et en Haïti et les Portugais à Belém et à Bahia.

Les cacaoyers ont gagné l'Asie et le Pacifique dès 1560 (WOOD, 1991 ; YOUNG, 1994). A cette époque, des Criollo du Venezuela ont été introduits dans l'île de Célèbes puis à Java, par les Hollandais. En 1614, des Criollo provenant du Mexique ont été implantés aux Philippines par les Espagnols. En 1798, la culture du cacaoyer a gagné Madras, en Inde, à partir de l'île d'Amboine, et Ceylan, à partir de la Trinité, grâce aux Anglais. De Ceylan, le cacaoyer a été introduit à Singapour et aux Fidji, en 1880, aux Samoa, en 1883, dans le Queensland, en 1886, à Bombay et à Zanzibar, en 1887. Le cacaoyer était cultivé en Malaisie dès 1778 et à Hawaii en 1831.

Son introduction en Afrique est plus récente. Les navigateurs espagnols ou portugais l'ont importé dans l'île de São Tomé, en 1822, puis dans l'île de Fernando Poo, en 1855 (BURLE, 1952). D'autres introductions ont été réalisées ensuite par des missionnaires suisses à partir du Surinam. Les premières graines de cacaoyer ont été semées sur le continent africain en 1857. Les premiers cacaoyers répandus en Afrique de l'Ouest à partir du Ghana étaient originaires de basse Amazonie (Forastero Amelonado) puis, en 1920, des formes hybrides Trinitario et des Criollo ont été importés et se sont hybridés avec les Amelonado locaux (TOXOPEUS, 1985). Chacun de ces transferts était effectué avec un nombre très restreint de génotypes, en conséquence la base génétique des cacaoyers initialement cultivés en Afrique de l'Ouest est très étroite et leur origine, imprécise.

La technique de fabrication du chocolat s'est perfectionnée avec l'invention en 1828 des presses pour extraire le beurre de cacao, ce qui a permis d'améliorer la consistance du chocolat (ENRIQUEZ, 1985). La mise en œuvre de la fermentation et de la torréfaction des fèves a permis de développer l'arôme du chocolat à partir de n'importe quel cultivar de cacaoyer autre que les variétés de Criollo traditionnellement cultivées, qui, elles, nécessitent peu de fermentation.

Au début du siècle, près de 80 % de la production mondiale de cacao, soit 115 000 tonnes, était réalisée par l'Amérique centrale et l'Amérique du Sud (BRAUDEAU, 1969). En 1997, la production atteignait près de 2,7 millions de tonnes. La Côte d'Ivoire est actuellement le premier producteur mondial avec 1,12 million de tonnes produites par an, viennent ensuite le Ghana et l'Indonésie, avec 330 000 tonnes chacun.

## La taxonomie et les ressources génétiques

### La taxonomie

Les cacaoyers appartiennent à la famille des sterculiacées et au genre *Theobroma*. *Theobroma cacao* L. ( $2n = 2x = 20$ ) est un arbre originaire des régions tropicales humides du nord de l'Amérique du Sud et de l'Amérique centrale.

Bien que le premier foyer de domestication et de culture soit localisé en Amérique centrale, CHEESMAN (1944) considère que le centre d'origine du cacaoyer se situe dans le cours supérieur de l'Amazone, près des Andes équatoriennes. C'est en effet dans cette région que POUND (1938) a observé la plus forte variabilité morphologique. Les cacaoyers ont un petit génome dont la taille a été estimée à 0,4 picogramme par génome haploïde (FIGUEIRA *et al.*, 1992 ; LANAUD *et al.*, 1992).

La taxonomie du genre *Theobroma* a été étudiée par les botanistes dès la fin du XIX<sup>e</sup> siècle (BERNOULLI, 1869 ; SCHUMANN, 1886 ; PITTIER, 1930 ; CHEVALIER, 1946). L'étude la plus récente et la plus complète est celle de CUATRECASAS (1964), qui divise le genre *Theobroma* en 6 sections et 22 espèces, qui se répartissent sur le continent américain entre 18° de latitude nord et 15° de latitude sud. L'une des sections est constituée par la seule espèce *T. cacao*. La classification proposée par cet auteur est fondée sur le mode de germination, l'architecture des arbres et les caractères des fruits et des fleurs. Parmi toutes ces espèces, seule *T. cacao* a une importance économique. *T. grandiflora*, le cupuassu, est cependant aussi exploitée au Brésil, la pulpe de ses fruits entrant dans la fabrication de boissons et de sorbets.

L'espèce *T. cacao* se compose d'un grand nombre de populations interfertiles, très variables sur le plan morphologique. Les plantes sont autogames ou allogames en fonction de leur origine génétique. Un système d'auto-incompatibilité gamétophytique renforce l'allogamie dans certaines populations (KNIGHT et ROGERS, 1955 ; BOUHARMONT, 1960 ; COPE, 1962 ; GLENDINNING, 1966).

Les classifications proposées pour l'espèce *T. cacao* n'ont jamais été pleinement satisfaisantes. Elles sont trop souvent schématiques étant donné l'intense brassage génétique entre populations d'origine génétique différente. MORRIS (1882) a été le premier à proposer une classification des cacaoyers en deux groupes : Criollo et Forastero<sup>1</sup>. PITTIER (1930) a fait de chaque groupe une espèce : *T. leiocarpum*, pour les Forastero, et *T. cacao*, pour les Criollo. CUATRECASAS (1964) les a considérés comme deux sous-espèces : *T. cacao* subsp. *cacao*, pour les Criollo, et *T. cacao* subsp. *sphaerocarpum*, pour les Forastero. Actuellement, une classification en deux groupes morphogéographiques, les Criollo et les Forastero, est admise. Un troisième groupe, les Trinitario, rassemble les formes hybrides entre les deux premiers.

1. Les termes Criollo et Forastero viennent du vocabulaire vénézuélien, qui distinguait les cacaoyers locaux traditionnellement cultivés, les Criollo, des cacaoyers étrangers, les Forastero, aussi appelés Trinitario, introduits plus tardivement à partir de la Trinité. Ainsi, à l'origine, le terme Forastero était relatif à toutes les variétés différentes des Criollo. Par extension, le terme Criollo, associé à un cacao de haute qualité issu de fèves épaisses, claires et aromatiques, a été utilisé pour désigner des cacaos de même qualité provenant d'autres pays comme le Nicaragua et le Mexique.

## LES FORASTERO

Le groupe des Forastero rassemble un grand nombre de populations sauvages et de variétés cultivées originaires d'Amérique du Sud. Les cacaoyers de ce groupe se rencontrent de l'Equateur jusqu'aux Guyanes. Ce sont des arbres vigoureux qui présentent de nombreuses résistances aux maladies. Leurs fèves, généralement violettes et aplaties, mais parfois blanches et bombées dans certaines populations, donnent un cacao de qualité moyenne à bonne. Les variétés de type Nacional, originaires d'Equateur, dont la classification dans ce groupe est actuellement remise en question (ENRIQUEZ, 1992), produisent un cacao fin, aromatique, très recherché par les chocolatiers.

### Les Forastero du Brésil et du Venezuela

Les Forastero de basse Amazonie et de la vallée de l'Orénoque sont largement cultivés dans tout le bassin amazonien. La forme la plus typique en est l'Amelonado. C'est ce type de cacaoyer, autocompatible, qui a été tout d'abord introduit en Afrique. De nombreuses prospections réalisées dans le bassin amazonien au Brésil ont montré une forte diversité morphologique (BARRIGA et al., 1984).

### Les Forastero du Pérou, d'Equateur et de Colombie

C'est au cours de plusieurs expéditions à la recherche de génotypes résistants au balai de sorcière, maladie due au champignon *Crinipellis perniciosa*, que POUND (1938 ; 1943) a décrit pour la première fois les populations du Pérou et de l'Equateur. Celles-ci présentent une forte diversité morphologique, avec des cabosses de forme et de couleur variables, vertes ou vert pâle ou légèrement pigmentées dans les zones proches de la Colombie. Des prospections plus récentes en haute Amazonie et, en particulier, à l'est de l'Equateur (ALLEN et LASS, 1983) ont mis en évidence des populations indigènes et sauvages de cacaoyers. Contrairement aux observations de Pound au Pérou, toutes les populations de l'est de l'Equateur apparaissent morphologiquement assez semblables. Elles se caractérisent par des arbres de grande taille, aux jeunes feuillés vert pâle et aux cabosses vertes et rugueuses contenant une forte proportion de fèves blanches. Ces cacaoyers ont des caractéristiques communes avec les variétés de Criollo d'Amérique centrale ou d'Amérique du Sud, toutefois, les cabosses des Criollo peuvent être plus allongées et complètement rouges, contrairement aux cabosses des cacaoyers équatoriens, qui ne sont que parfois légèrement pigmentées.

On trouve aussi en Equateur des formes hybrides, en particulier, les Refractorio. Ce sont des descendants d'une introduction de Trinitario effectuée en Equateur vers 1890 à partir de quelques cabosses venant de la Trinité. Ce matériel particulièrement vigoureux et précoce, même sur des sols pauvres, a été utilisé par la plupart des planteurs. Un large brassage génétique s'est alors établi entre les différentes origines. A l'arrivée de la maladie du balai de sorcière, qui s'est répandue dans presque tous les districts en provoquant de

lourdes pertes de production dès les années 20-30, quelques arbres, appelés *Refractario*, sont apparus plus tolérants. Des replantations ont alors été réalisées à partir d'un faible pourcentage de semences issues de ces arbres rescapés.

### Les Forastero des Guyanes

Les premières observations de cacaoyers sauvages au Surinam ont révélé des populations très homogènes pour la forme de leurs cabosses, de type *Amelonado*, et distinctes des types cultivés dans la région (MYERS, 1930). Des cacaoyers spontanés avaient été observés dans le sud-est de la Guyane française dès 1729 (LECOMTE et CHALLOT, 1897). Des prospections organisées en Guyane française ont mis en évidence certaines populations, en particulier celles du haut Camopi, qui présentent une variabilité phénotypique importante (SALLEE, 1987). Ces populations se distinguent des Forastero bas-amazoniens par leurs cabosses, dont la forme va de celle de l'*Amelonado* à celle de l'*Angoleta*, avec une verrucosité plus ou moins marquée et une taille moyenne à élevée (LACHENAUD et SALLEE, 1993). L'origine des types cultivés en Guyane est incertaine. Ils proviendraient, selon GUIBAN (1825), des forêts de cacaoyers naturels situées au-dessus du Camopi.

### LES CRIOLLO

Les Criollo ont été les premiers cacaoyers domestiqués par les civilisations mayas et aztèques. Ce sont les seules variétés qui ont été cultivées dans toute l'Amérique centrale et dans le nord de l'Amérique du Sud jusqu'au XVIII<sup>e</sup> siècle.

Les Criollo, autocompatibles, se rencontrent du Mexique jusqu'en Colombie et au Venezuela sous forme cultivée ou subsponnée (SORIA, 1973 ; REYES, 1992). Ils ont une croissance lente, sont plus sensibles aux maladies et aux insectes que les Forastero et manifestent une forte diversité morphologique (SORIA, 1970a ; 1970b). Leurs fruits sont allongés, avec une pointe acuminée, une surface lisse ou rugueuse, rouge ou verte avant maturité, et un cortex peu lignifié. Leurs fèves sont de taille variable mais, le plus souvent, grandes et épaisses et de couleur blanche ou rosée. Les fèves de Criollo sont généralement bombées et non aplaties, contrairement à celles des Forastero de basse Amazonie et du Pérou. Ces caractéristiques des fèves, associées à une moindre lignification du cortex, sont habituellement utilisées pour classer les clones de cacaoyer dans le groupe des Criollo. Les formes cultivées de Criollo varient depuis un type à cabosse lisse, comme le *Porcelana* du Venezuela, jusqu'au type *Cundeamor*, à cabosses très verruqueuses, trouvé au Mexique, en Colombie et au Venezuela. Une forme particulière, le *Pentagona*, se rencontre dans les anciennes plantations du Mexique, du Guatemala, du Nicaragua et du Venezuela.

Outre quelques variétés traditionnelles de Criollo présentes dans les collections, des types Criollo plus vigoureux et productifs ont été collectés dans les



plantations des dernières décennies. Ce sont ces clones de Criollo, représentant en général ce groupe génétique dans les collections, qui seront appelés « Criollo actuels » dans la suite de cette étude.

## LES TRINITARIO

Les Trinitario regroupent toutes les formes hybrides entre Criollo et Forastero de basse Amazonie ou de la vallée de l'Orénoque, qui sont à l'origine de ces variétés. Les Trinitario auraient été cultivés tout d'abord à la Trinité, puis au Venezuela et en Amérique centrale, où ils se seraient mélangés aux Criollo dans les plantations traditionnelles.

## Les ressources génétiques

Une quarantaine de prospections ont été réalisées depuis 1930. Elles ont principalement porté sur les Forastero du Pérou, du Brésil, d'Equateur, de Colombie, du Venezuela et de Guyane et, plus récemment, sur les Criollo d'Amérique centrale, du Venezuela et de Colombie.

Les collections les plus importantes issues de ces prospections sont conservées à la CRU, Cocoa Research Unit, à la Trinité (2 500 génotypes), à la CEPLAC, Comissão Executiva do Plano de Lavoura Cacaueira, au Brésil (2 000 génotypes originaux) et au CATIE, Centro Agronómico Tropical de Investigación y Enseñanza, au Costa Rica (700 génotypes). Chacune des collections a sa spécificité : celle de la Trinité est riche en Forastero de haute Amazonie, celle du Brésil regroupe le matériel provenant des prospections de la région amazonienne du Brésil et des variétés cultivées à Bahia, celle du CATIE rassemble surtout des Trinitario et des Criollo. La collection du Cirad en Guyane française contient les descendants d'environ 200 pieds mères spontanés collectés dans ce pays.

Les échanges internationaux de matériel végétal sont soumis à une quarantaine de deux ans afin de dépister les maladies virales et fongiques. Trois stations de quarantaine sont opérationnelles : celle de l'université de Reading, au Royaume-Uni, celle du Cirad, en France, et celle de la CRU, à la Barbade.

Une base de données internationale a été constituée au début des années 90. Elle rassemble des informations sur plus de 14 000 génotypes conservés dans 43 collections et fait l'objet d'une actualisation régulière (END *et al.*, 1992). Une autre base de données, Tropgène-db, actuellement mise en place au Cirad, regroupe les données agronomiques et moléculaires de près de 400 génotypes.

L'amélioration génétique du cacaoyer, ciblée sur des objectifs différents selon les pays — qualité, production, résistance —, repose toujours sur la création d'hybrides entre géniteurs appartenant à des populations différentes. Comme il est difficile d'établir clairement une classification uniquement sur la base de

caractères morphologiques, nous avons développé l'utilisation de marqueurs biochimiques et moléculaires pour préciser l'organisation génétique de l'espèce et pour étudier les processus de domestication.

## L'organisation de la diversité génétique

### La diversité morphologique

Plusieurs auteurs ont tenté de définir les descripteurs morphologiques les plus discriminants qui prennent en compte la variance intra et interclonale (ENRIQUEZ et SORIA, 1968 ; ENGELS *et al.*, 1980 ; SORIA et ENRIQUEZ, 1981 ; ENGELS, 1983a ; 1983b ; BEKELE, 1992 ; BEKELE *et al.*, 1994 ; RABOIN et PAULIN, 1993). L'IBPGR (1981) a publié une liste de 65 descripteurs morphologiques pour caractériser les ressources génétiques du cacaoyer. Cependant, afin de réduire le temps nécessaire à la caractérisation dans les études portant sur un grand nombre de génotypes, cette liste a souvent été écourtée.

Ainsi, ENGELS (1986) a étudié la diversité de 294 clones de la collection du CATIE à l'aide de 39 descripteurs qualitatifs et quantitatifs portant sur la morphologie des fleurs, des feuilles et des fruits. Cette analyse met en évidence une structuration entre les types Criollo et Forastero ainsi qu'une diversité importante dans chacun de ces groupes. Dans cette analyse, les Trinitario se regroupent avec les Criollo.

N'GORAN (1994) a analysé la diversité de 52 clones appartenant aux groupes Forastero, Criollo et Trinitario pour 9 caractères des fèves et des cabosses. Cette analyse a confirmé la structuration observée par Engels, avec une différenciation entre les Forastero, d'une part, les Criollo et Trinitario, d'autre part.

Une étude plus récente (BEKELE et BEKELE, 1998), réalisée sur 100 clones issus de 24 populations, met en évidence une structuration des populations en fonction de leur origine géographique.

L'ensemble de ces études montre que les marqueurs morphologiques permettent de structurer globalement la diversité des différentes populations de cacaoyers en collection. Ils sont accessibles à tous, mais leur coût en temps et leur variabilité en fonction de l'environnement rendent leur utilisation souvent délicate.

### La diversité enzymatique

La diversité enzymatique de clones ou de populations de cacaoyers a été évaluée par plusieurs auteurs (LANAUD et BERTHAUD, 1984 ; LANAUD, 1986a ;

1986b ; 1987 ; AMEFIA, 1986 ; ATKINSON *et al.*, 1986 ; YIDANA *et al.*, 1987 ; RONNING et SCHNELL, 1994 ; WARREN, 1994 ; SOUNIGO *et al.*, 1996).

LANAUD (1987) a analysé la diversité de 296 géotypes à l'aide de 6 systèmes enzymatiques polymorphes représentant 9 locus. Sur l'ensemble des échantillons analysés, un total de 30 allèles a pu être identifié. A l'exception de *PAC1*, tous les allèles se retrouvent dans les populations originaires de haute Amazonie. *PAC1* est en revanche fréquent parmi les Criollo et les Trinitario. Un autre allèle, *MDHA1*, est lui aussi fréquent parmi les Criollo et les Trinitario et rare parmi les Forastero.

Les populations qui présentent le plus grand nombre d'allèles par locus sont celles de haute Amazonie, avec 1,8 à 2,2 allèles par locus. Les populations du Venezuela et de Guyane sont les moins variables avec de 1 à 1,5 allèles par locus. Le pourcentage de locus polymorphes est le plus élevé dans les populations de haute Amazonie, où il est supérieur à 50 %. L'hétérozygotie moyenne est la plus élevée parmi les Trinitario américains (0,36), les Criollo actuels (0,29) et les Refractorio EQX d'Equateur (0,35). L'hétérozygotie est la plus faible pour les Amelonado africains (0,04), les Forastero de Guyane (0,06) et les deux populations du Venezuela analysées (0,1 et 0).

Cette variabilité des taux d'hétérozygotie au sein des populations de Forastero pourrait résulter du système de reproduction des arbres et de leur système d'auto-incompatibilité. En effet, les cacaoyers de haute Amazonie sont fortement auto-incompatibles et préférentiellement allogames, alors que les Forastero de basse Amazonie sont autocompatibles et donc préférentiellement autogames.

Une analyse factorielle des correspondances (AFC) a été réalisée à l'aide de 31 variables observées sur 286 individus. Aucune structuration nette entre Forastero et Criollo ou Trinitario n'apparaît sur le premier plan de l'AFC alors qu'une légère structuration est visible sur le plan 3-4. Les Forastero issus de haute Amazonie présentent la plus grande diversité génétique. Ceux d'Equateur sont aussi variables que ceux du Pérou, contrairement aux observations morphologiques de ALLEN et LASS (1983). Comparativement, la diversité des autres populations de Forastero analysées est plus réduite. Les Forastero cultivés en Guyane se différencient nettement des Forastero sauvages prospectés dans le sud de la Guyane sur le premier plan de l'AFC (figure 1). L'utilisation directe des cacaoyers spontanés locaux semble donc exclue pour expliquer l'origine des cacaoyers cultivés, dont la provenance demeure inconnue. Une origine vénézuélienne est cependant probable (LACHENAUD et SALLEE, 1993). La diversité de la population sauvage de Guyane est relativement faible compte tenu du grand nombre d'individus analysés (92). Les deux populations de Guyane, cultivée et sauvage, se différencient bien des Forastero bas-amazoniens (Amelonado) d'Afrique de l'Ouest et des populations de la vallée de l'Orénoque au Venezuela. Les cacaoyers guyanais cultivés sont cependant plus proches des Forastero du Venezuela. Sur le premier axe de l'AFC, les

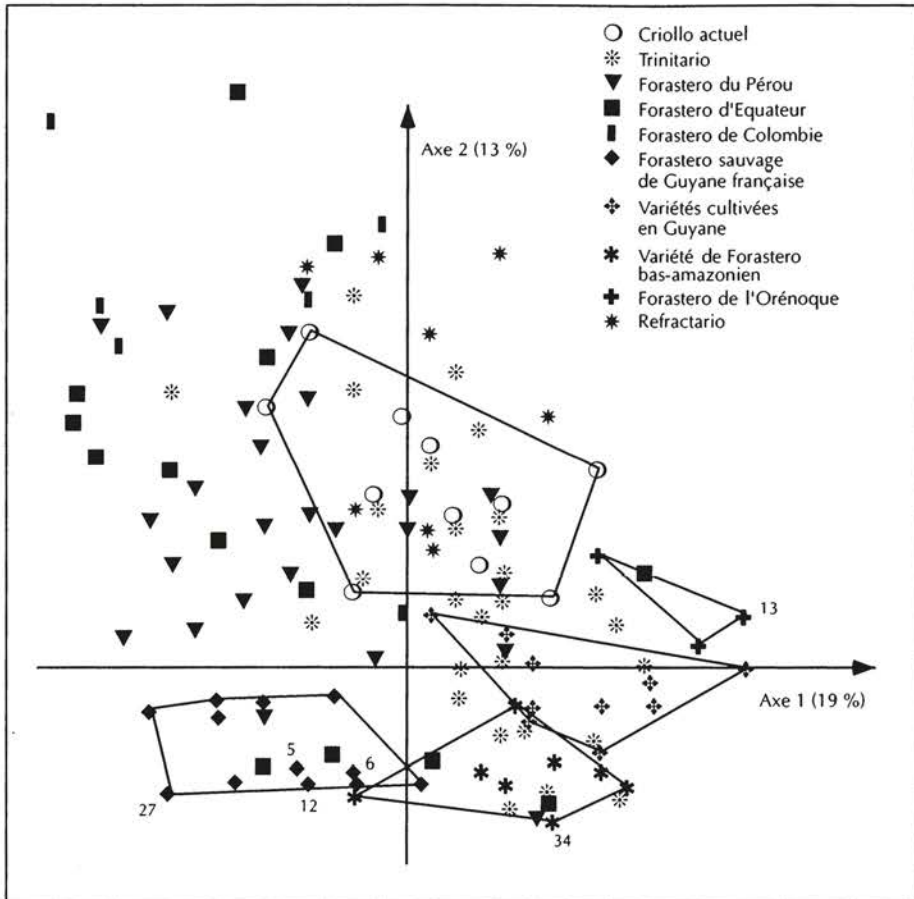


Figure 1. Premier plan de l'AFC de 296 individus analysés par isoenzymes, d'après LANAUD (1987). Les chiffres notés sur la figure indiquent le nombre d'individus superposés sur un point donné.

cacaoyers guyanais sauvages se rapprochent plus de certains Forastero du Pérou et d'Equateur que des Forastero du Venezuela et des formes cultivées de Guyane.

Il existe une grande proximité génétique entre les variétés cultivées actuelles de Criollo et les Trinitario cultivés en Amérique et en Afrique, ce qui s'explique facilement par les brassages génétiques importants qui existent dans les plantations. La variabilité des Trinitario cultivés en Afrique est très forte et s'étend des types Forastero Amelonado jusqu'aux Criollo. Cette situation est liée à l'histoire de l'introduction du cacaoyer en Afrique avec une première vague de Forastero Amelonado puis une deuxième vague de Trinitario. Les brassages génétiques qui se sont produits par la suite ont conduit à certains types hybrides plus proches des Forastero Amelonado.

A l'aide de 4 systèmes enzymatiques, WARREN (1994) a analysé la diversité de sept populations de Forastero de Colombie, d'Equateur et du Pérou et de deux populations de Trinitario, à raison de 10 individus par population. Il observe un grand nombre d'allèles pour le système PGI (10 allèles contre 5 dans l'étude de LANAUD, 1987). Les indices de diversité de Shannon indiquent une plus forte diversité dans les populations de Forastero d'Equateur que dans celles du Pérou. Les formes hybrides Trinitario apparaissent aussi variables. L'auteur en conclut que, s'il existe réellement un centre de diversité des cacaoyers sauvages, celui-ci doit se situer non pas au Pérou, comme le suggère CHEESMAN (1944), mais plus au nord, en Equateur et en Colombie.

RONNING et SCHNELL (1994) ont étudié la diversité enzymatique de 86 clones de cacaoyer issus des groupes Forastero, Criollo actuels, Trinitario et d'un groupe hybride non défini, à l'aide de 6 systèmes enzymatiques correspondant à 8 locus. Les fréquences alléliques et les distances génétiques obtenues confirment la différenciation entre Forastero et Criollo. Les valeurs des paramètres de diversité génétique sont globalement similaires à celles d'autres espèces ligneuses pérennes :  $H_T = 0,295$ ,  $H_S = 0,266$ ,  $G_{ST} = 0,096$  (HAMRICK et GODT, 1990).

SOUNIGO *et al.* (1996) ont étudié la diversité enzymatique de 487 clones de cacaoyer appartenant à 28 populations ou groupes d'accessions présents dans la collection de la CRU à la Trinité, à l'aide de 5 systèmes enzymatiques polymorphes. L'analyse des paramètres de la diversité génétique montre des richesses alléliques et des hétérozygoties très variables. Certaines populations telles que les Forastero de Guyane et les Trinitario de la République dominicaine ou de la Martinique sont peu diversifiées, avec en moyenne de 1,3 à 1,8 allèles par locus et de 6 à 16 % d'hétérozygotie moyenne par locus et par individu. L'inverse est observé pour certaines populations hybrides de Trinitario et de Refractario, qui présentent 2,2 à 2,5 allèles par locus et 42 à 50 % d'hétérozygotie moyenne par locus et par individu. D'autres populations de Trinitario montrent une forte hétérozygotie mais un faible indice de diversité de Shannon ou de richesse allélique. En revanche, certaines populations de Forastero et de Refractario d'Equateur ont un faible niveau d'hétérozygotie mais une grande richesse allélique.

La faible diversité de certaines populations, telles que celles de la Martinique et de la République dominicaine, reflète l'homogénéité du matériel issu probablement d'un petit nombre d'introductions réalisées dans ces îles de la Caraïbe. A l'inverse, les valeurs de diversité élevées des Refractario témoignent bien de leurs origines multiples.

Estimées à partir de l'indice de diversité de Nei, la diversité moyenne intrapopulation,  $H_{pop}$ , est de 0,77 et la diversité interpopulation, de 0,23. Ainsi, les trois quarts de la diversité totale se situent au niveau intrapopulation. Ces résultats sont en accord avec ceux de RONNING et SCHNELL (1994) et de RUSSEL *et al.* (1993) ; ils sont caractéristiques des espèces pérennes allogames pour

lesquelles d'importants brassages génétiques ont lieu (HAMRICK *et al.*, 1992). La classification ascendante hiérarchique (CAH) fondée sur les distances génétiques de Nei est représentée sur la figure 2. Deux groupes majeurs sont observés à  $d = 0,2$ . L'un rassemble les populations de Refractario et de Trinitario ainsi que la population de Forastero IMC du Pérou. Ce groupe est lui-même structuré en deux sous-groupes : le premier comprend des Trinitario différenciés selon leur origine géographique, le second rassemble tous les Refractario ainsi que les populations IMC (Forastero), ICS et CC (Trinitario). L'autre groupe majeur réunit toutes les autres populations de Forastero, y compris celle des clones sauvages de Guyane, qui apparaît proche des populations MO et PA du Pérou. Dans cette analyse, la population des Scavina est proche de la population des Forastero d'Equateur LCTEEN.

Cette analyse, malgré le faible nombre de locus analysés, permet donc de révéler une structure globale, correspondant aux différents groupes génétiques, et une structuration géographique des populations à l'intérieur de certains groupes.

## La diversité moléculaire

### LE NIVEAU D'HÉTÉROZYGOTIE DES CLONES

Les pourcentages de locus hétérozygotes de 300 clones ont été évalués par RFLP après hybridation de 33 sondes génomiques et d'ADNc. Ces sondes ont été cartographiées sur la carte de référence (LANAUD *et al.*, 1995), ce qui a permis de connaître le déterminisme génétique des marqueurs utilisés.

Les clones dont le taux d'hétérozygotie est le plus élevé sont les Trinitario (86 % pour les clones UIT). Une forte hétérozygotie est aussi observée pour certaines variétés actuelles de Criollo (73 % pour CHO42), ainsi que chez certains Forastero de haute Amazonie tel IMC105 (42 %). Les niveaux les plus bas d'hétérozygotie se rencontrent dans des variétés anciennes de Criollo (Porcelana, 3 %) ou de Forastero bas-amazoniens (Pará, 4 %). Certains Forastero de Guyane (GU154), du Venezuela (VENC20), de Colombie (EBC5), du Pérou (P2) et de basse Amazonie (Catongo) sont totalement homozygotes.

Ces variations observées dans le niveau d'hétérozygotie des clones peuvent refléter soit le brassage génétique qui s'est produit entre populations d'origine différente, soit le mode de reproduction préférentiellement autogame ou allo-game des cacaoyers.

### LA DIVERSITÉ MOLÉCULAIRE RÉVÉLÉE PAR LES RFLP

La diversité moléculaire de populations de cacaoyers, révélée par RFLP au niveau nucléaire ou cytoplasmique, a été évaluée par plusieurs auteurs (LAURENT *et al.*, 1993a ; 1993b ; 1994 ; FIGUEIRA *et al.*, 1994 ; LERCETEAU *et al.*, 1997 ; MOTAMAYOR *et al.*, 1997).

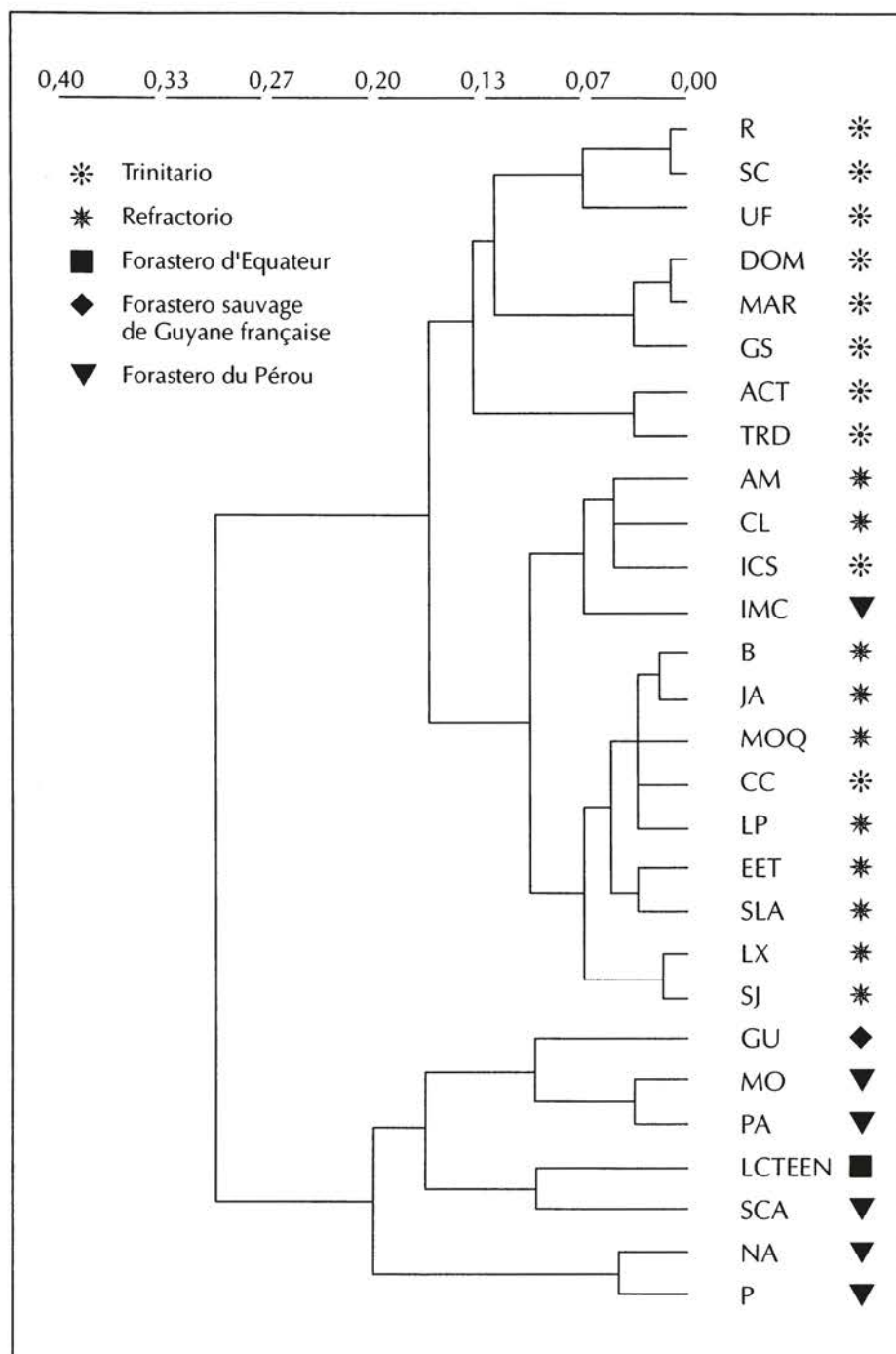


Figure 2. Arbre construit par la méthode du neighbor joining à partir des distances de Nei non biaisées des différentes populations analysées à l'aide d'isoenzymes, d'après SOUNIGO et al. (1996).



## La diversité nucléaire

LAURENT *et al.* (1994) ont analysé 201 génotypes appartenant à différents groupes morphogéographiques. La diversité de l'ADN nucléaire total a été analysée à l'aide de 31 sondes d'ADNc qui ont permis de révéler 87 bandes polymorphes. Malgré une variation continue entre les groupes, due en particulier à un grand nombre d'hybrides Trinitario, une structuration assez claire apparaît sur l'axe 1 d'une AFC (figure 3) entre Forastero, d'une part, Criollo actuels et Trinitario, d'autre part. Les populations de Forastero haut-amazoniens et les variétés de Criollo sélectionnées présentent une grande variabilité. Les variétés de Criollo cultivées se superposent au pool des hybrides Trinitario.

Dans cette analyse, une différenciation marquée apparaît entre les Forastero sauvages de Guyane, d'une part, et les Forastero bas-amazoniens ou du Venezuela, d'autre part. Les Refractario se répartissent sur l'aire de distribution des Forastero. Les deux clones de Nacional sont proches de certains Refractario.

Une deuxième étude par RFLP a été centrée sur la diversité des formes de Criollo cultivées en Amérique centrale et au Venezuela (MOTAMAYOR *et al.*, 1997). Elle portait sur 208 génotypes, qui provenaient de collectes réalisées dans les plus anciennes plantations du Venezuela, sans tenir compte des critères agronomiques de production, dans la forêt Lacandona du Mexique, près de sites archéologiques mayas où l'on trouve des cacaoyers subspontanés qui sont probablement les descendants des cacaoyers anciennement cultivés par les Mayas, et dans le Yucatán. Ces collectes comprenaient des représentants des variétés pures de Criollo anciennement cultivées, qui présentent des formes de cabosse variées : ovale et lisse, comme celle des Porcelana, ou très verruqueuse, telle celle des Pentagona. L'analyse de ce matériel a été complétée par celle de Trinitario, de Forastero et de variétés actuelles de Criollo présentes dans les collections du Venezuela, du Mexique et du Costa Rica.

Une cinquantaine de clones communs et une trentaine de sondes communes à l'étude précédente ont ensuite permis de comparer les résultats des deux types d'étude. Les fréquences alléliques de 26 locus cartographiés et l'hétérozygotie moyenne par locus pour chaque groupe montre un degré élevé de polymorphisme parmi les Forastero de haute Amazonie et les Trinitario. En effet, le plus grand nombre d'allèles par locus (2,2) est observé dans les groupes des Trinitario et des Forastero haut-amazoniens, de même, la plus forte valeur de diversité génétique de Nei (0,41) se retrouve dans le groupe des Trinitario. Des valeurs semblables de diversité génétique moyenne (0,37) sont relevées parmi les Forastero haut-amazoniens étudiés et le groupe de clones de Criollo actuellement cultivés. Cette diversité des formes actuelles de Criollo a été aussi signalée par d'autres auteurs (ENGELS, 1986 ; FIGUEIRA *et al.*, 1994 ; LAURENT *et al.*, 1994 ; LERCETEAU *et al.*, 1997).

Au contraire, les Criollo correspondant aux anciens types cultivés sont très peu polymorphes et montrent les plus faibles valeurs de diversité génétique (0,02). De même, ils présentent un faible niveau d'hétérozygotie moyenne par locus



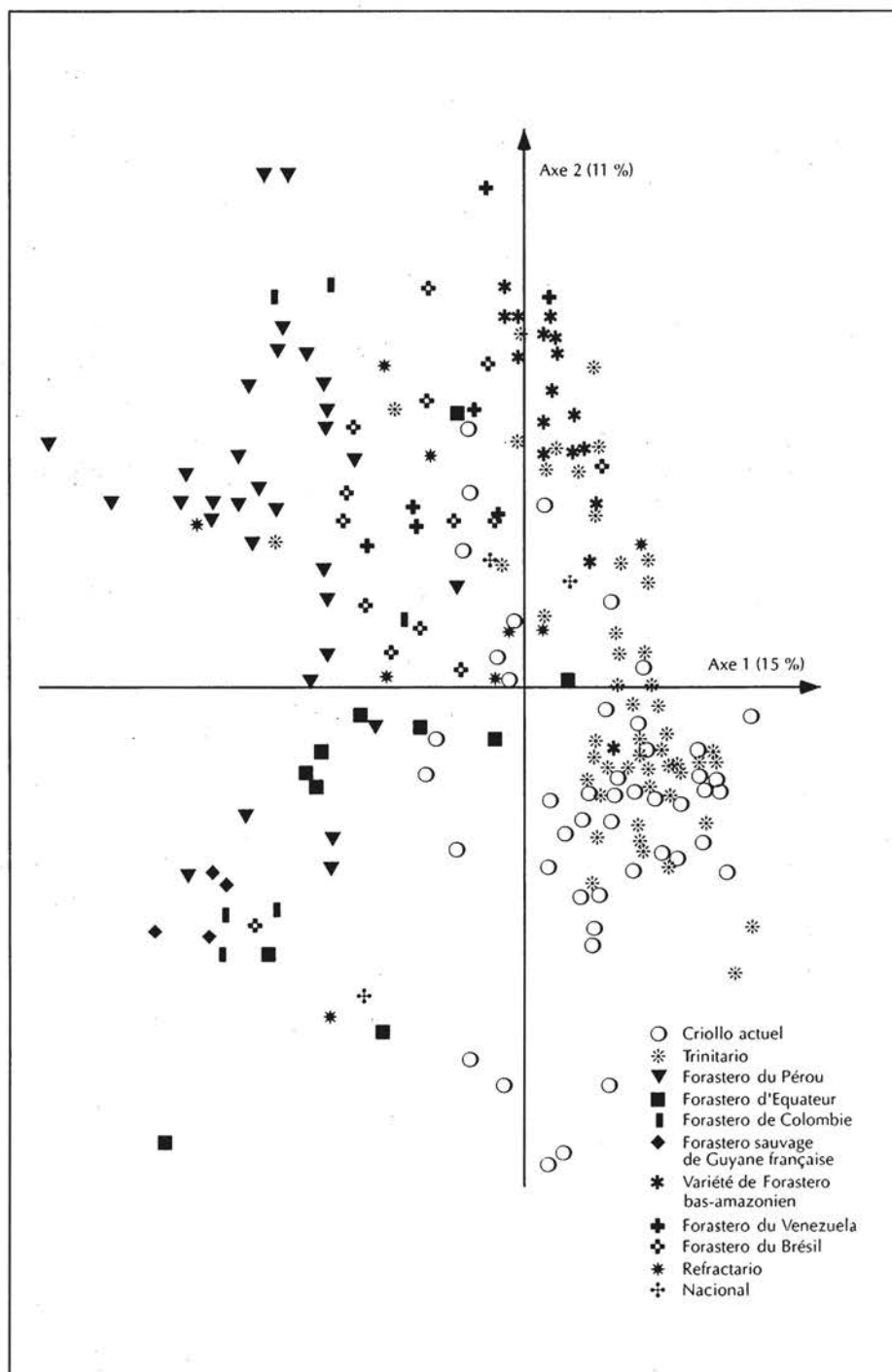


Figure 3. Premier plan de l'AFC de 201 individus étudiés pour leur diversité nucléaire à l'aide de 31 sondes d'ADNc, d'après LAURENT et al. (1994).

(0,02), contrairement aux variétés actuelles de Criollo (0,43) qui présentent avec les Trinitario (0,40) les valeurs les plus élevées. De plus, au sein de ces variétés anciennes, presque aucune différence génotypique n'a été trouvée entre des types morphologiques pourtant très contrastés tels que les Porcelana, les Pentagona et les Guasare du Venezuela ou les Criollo de la forêt Lacandona au Mexique.

Une AFC réalisée à partir de 64 bandes RFLP polymorphes révèle une différenciation très nette entre les Forastero et les variétés anciennes de Criollo (figure 4). Les Trinitario se superposent aux variétés actuelles de Criollo comme dans les analyses précédentes.

Les clones de Criollo actuels présents dans les collections apparaissent généralement plus hétérozygotes que les variétés anciennes, ce qui peut s'expliquer par une sélection de clones réalisée non seulement sur des caractères de qualité des fèves, mais aussi sur certains caractères agronomiques de vigueur, de production ou de résistance aux maladies. Ces types plus vigoureux pourraient correspondre à des formes plus ou moins hybrides résultant d'introgessions de gènes de Forastero dans les types de Criollo purs anciennement cultivés. La diversité observée parmi les variétés actuelles de Criollo pourrait s'expliquer à la fois par des introgessions de gènes de Forastero — plus ou moins importantes et portant sur différentes portions du génome — et par la diversité des Forastero à l'origine de ces introgessions.

Dans l'étude de LAURENT *et al.* (1994), la diversité révélée sur le premier plan de l'AFC montrait une différenciation entre Forastero, d'une part, Criollo et Trinitario, d'autre part, le long de l'axe 1. Cependant, la variabilité apparaît identique pour ces deux groupes le long de l'axe 2, ce qui laisse penser que la variabilité des Forastero pour les caractères de l'axe 2 pourrait être à l'origine de la diversité révélée au sein des Criollo actuels et des Trinitario. Or, cette deuxième hypothèse semble pouvoir être écartée. En effet, dans l'étude de MOTAMAYOR *et al.* (1997), pour laquelle des sondes supplémentaires ont été utilisées, la diversité révélée par le premier plan de l'AFC se structure de façon différente. La diversité des Forastero ne peut expliquer celle des Criollo actuels et des Trinitario, et la majorité des clones de ces deux derniers groupes s'étale des Criollo anciens jusqu'aux variétés bas-amazoniennes. De plus, la nature très homozygote des anciennes variétés de Criollo, mise en évidence par les RFLP et par les microsatellites répartis sur tout le génome (LANAUD *et al.*, 1995 ; RISTERUCCI *et al.*, 1996), permet de disposer d'un génotype de référence pour l'ensemble des chromosomes caractérisant les Criollo purs. L'observation d'une vingtaine de Trinitario et de variétés actuelles de Criollo à l'aide de 19 marqueurs microsatellites très polymorphes semble indiquer que le même génotype ou qu'une même population homogène de Forastero bas-amazoniens est à l'origine de la majorité des Trinitario (18 sur les 19 analysés). En effet, pour tous les locus, c'est toujours le même allèle Forastero, différent de l'allèle présent chez les Criollo anciens, qui est trouvé, alors que ces microsatellites révèlent une forte diversité allélique dans le groupe des Forastero. Il semble

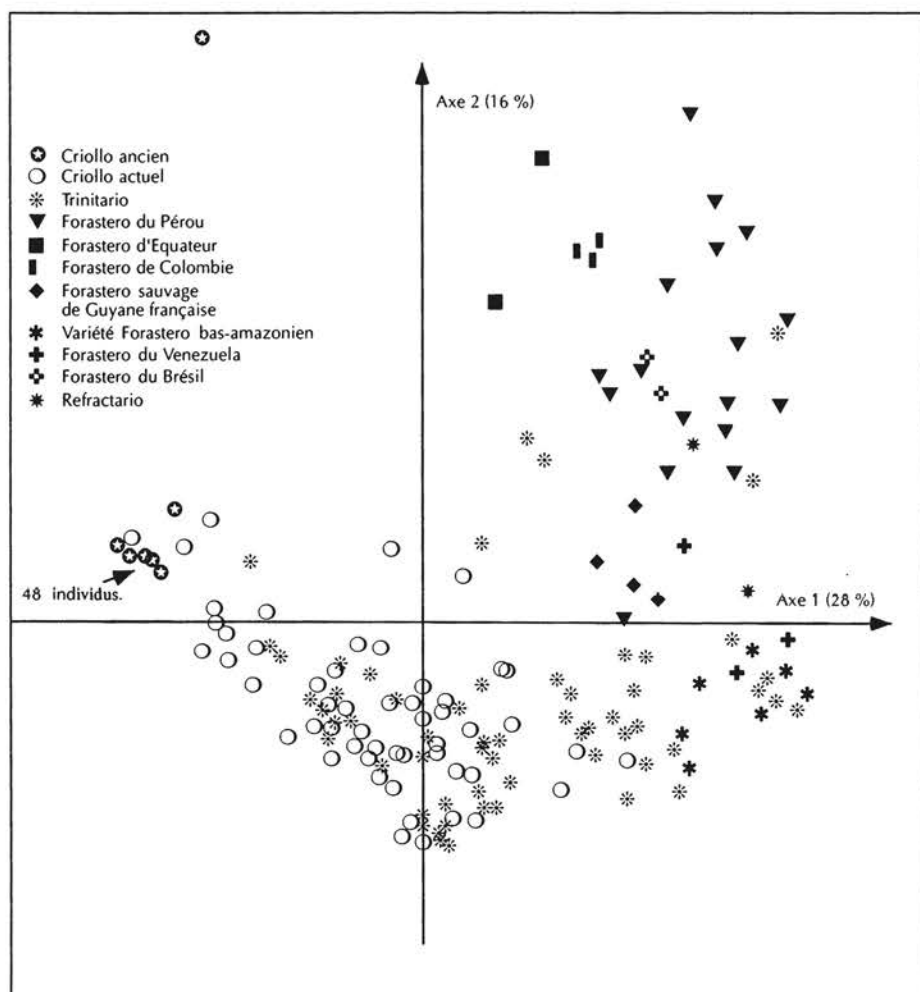


Figure 4. Premier plan de l'AFC de 208 individus comprenant des Criollo anciens et des variétés nouvelles de Criollo, étudiés pour leur diversité nucléaire à l'aide de 26 sondes RFLP nucléaires, d'après MOTAMAYOR et al. (1997).

donc d'après ces premiers résultats que ce serait plutôt les combinaisons différentes de parties du génome Forastero introgressées dans les Criollo purs qui seraient à l'origine de la variabilité observée au sein des Trinitario et des variétés modernes de Criollo. Les résultats de LAURENT et al. (1994), qui montrent une variabilité identique entre les groupes révélée par l'axe 2 de l'AFC, sont sans doute liés au faible polymorphisme révélé par les marqueurs RFLP qui, pour la plupart, ne présentent que deux allèles possibles, communs à tous les groupes. L'utilisation d'un plus grand nombre de marqueurs RFLP et de marqueurs plus polymorphes, comme les microsatellites qui ont révélé de nombreux allèles, nous a permis d'identifier les origines génétiques probables du

pool des Criollo actuels et des Trinitario. Ces résultats mettent donc en évidence une structure génétique très semblable entre ce qu'on appelle les Criollo, qui regroupent en majorité des Criollo actuels hybrides, et les Trinitario.

LERCETEAU *et al.* (1997) ont analysé la diversité génétique de 59 clones Nacional d'Equateur, mais aussi de 29 Forastero, de 29 Trinitario et de 9 variétés actuelles de Criollo. Quarante-trois sondes génomiques, codées en terme de locus, ont été utilisées. La diversité génétique intragroupe est identique pour les Forastero (0,33), les Trinitario (0,31) et les Criollo actuels (0,31). Elle est la plus faible chez les Nacional (0,19). L'hétérozygotie est la plus élevée chez les Trinitario et les Criollo, et la plus faible pour les Nacional. Parmi ceux-ci, certains génotypes prélevés dans de très vieilles plantations d'Equateur apparaissent presque totalement homozygotes.

Des CAH ont été réalisées à l'aide des distances de Rogers modifiées, calculées soit entre les groupes morphogéographiques, soit entre les origines géographiques seules des clones. La discrimination entre groupes est meilleure quand les clones sont classés en fonction de leur origine géographique ( $G_{ST} = 0,23$  contre 0,16). En considérant les groupes morphogéographiques, les Criollo et les Trinitario sont les groupes les plus proches alors que les Nacional sont les plus distants de tous les autres groupes. Ces résultats confirment la spécificité génétique des Nacional et leur différenciation marquée par rapport aux Criollo.

Lorsque seule l'origine géographique des clones est prise en compte, les clones de Forastero guyanais apparaissent comme les plus éloignés de tous les autres, puis les clones d'Equateur se différencient de tous les autres groupes. La séparation des cacaoyers guyanais peut s'expliquer non seulement par la différenciation de cette population, mais aussi par le fait qu'aucun type hybride n'est inclus dans ce groupe, ce qui n'est pas le cas pour les autres groupes géographiques.

Le pourcentage de locus hétérozygotes des clones a été estimé sur la base de 31 sondes RFLP. Parmi les clones Nacional, on peut identifier deux groupes d'individus, BCH et SA, dont le taux d'homozygotie très élevé varie de 90 à 100 % et qui contiennent les mêmes allèles. Ces arbres proviennent de deux plantations qui datent d'une centaine d'années, situées à 450 kilomètres l'une de l'autre dans le nord et dans le sud de l'Equateur. Les Nacional constituent donc une population très hybride et, de façon identique à la genèse des Criollo actuels et des Trinitario, les clones BCH et SA, très homozygotes, pourraient représenter les ancêtres de cette population.

### La diversité de l'ADN cytoplasmique

YEOH *et al.* (1990), dans une étude sur la structure du génome chloroplastique du cacaoyer, ont signalé la petite taille de ce génome, qui est de l'ordre de  $100 \cdot 10^3$  bases.

La diversité cytoplasmique de 177 génotypes a été analysée au niveau mitochondrial et chloroplastique par LAURENT *et al.* (1993b) à l'aide de sondes hétérologues mitochondriales (ATP-synthétase de tournesol, cytochrome-oxydase de blé) et chloroplastiques (Rubisco d'épinard).

Deux profils chloroplastiques, A et B, ont été observés. Soixante-dix pour-cent des clones ont le profil A et appartiennent à tous les groupes morphogéographiques. Les clones ayant le profil B sont en majorité des Criollo actuels, des Trinitario et quelques Forastero de basse Amazonie, de Colombie et du Pérou (Scavina).

Les sondes mitochondriales ont permis de révéler 44 profils mitochondriaux, dont 35 ne regroupent pas plus de 5 clones. Ces types très variables et mineurs sont essentiellement constitués de Criollo et Trinitario. Parmi les 9 types restants, le type 1 regroupe les deux tiers des clones de Forastero, incluant des génotypes de Guyane, du Venezuela, du Brésil, du Pérou, de Colombie et d'Equateur. L'autre type majoritaire, le type 2, regroupe une majorité de Criollo et Trinitario (26 clones) et quelques génotypes bas-amazoniens. Une AFC permet de visualiser la diversité de l'ADN mitochondrial des clones étudiés (figure 5). Il est intéressant de constater que, contrairement à l'ADN nucléaire, la diversité de l'ADN mitochondrial est beaucoup plus importante chez les variétés actuelles de Criollo que chez les Forastero, qui apparaissent assez peu variables. Récemment, ce polymorphisme mitochondrial a été mis en évidence également parmi le pool de Criollo anciens du Mexique et du Venezuela, qui présentent par ailleurs une homozygotie quasi totale au niveau nucléaire.

#### LA DIVERSITÉ MOLÉCULAIRE RÉVÉLÉE PAR LES RAPD

La facilité d'accès aux marqueurs RAPD a permis à de nombreux auteurs de les utiliser pour identifier ou pour étudier la diversité des cacaoyers (WILDE *et al.*, 1992 ; RUSSEL *et al.*, 1993 ; FIGUEIRA *et al.*, 1994 ; N'GORAN *et al.*, 1994 ; RONNING et SCHNELL, 1994 ; DE LA CRUZ *et al.*, 1995 ; RONNING *et al.*, 1995 ; SOUNIGO *et al.*, 1996 ; LERCETEAU *et al.*, 1997 ; WHITKUS *et al.*, 1998).

FIGUEIRA *et al.* (1994) ont analysé la diversité de génotypes appartenant aux trois groupes, Criollo, Trinitario et Forastero, à l'aide de 128 marqueurs RAPD correspondant à des fragments amplifiés provenant de 23 amorces. Ils ont observé une variation continue entre ces groupes, sans pouvoir les différencier. En revanche, ils ont constaté une structuration assez nette entre clones sauvages et clones cultivés. Cette structuration est confirmée par l'analyse de l'ADN ribosomique. Ces résultats amènent ces auteurs à proposer une nouvelle classification des cacaoyers qui s'appuie non pas sur les trois groupes traditionnels, Criollo, Trinitario et Forastero, mais sur les groupes « cacaoyers sauvages » et « cacaoyers cultivés ».

Dans le même temps, RUSSEL *et al.* (1993) ont étudié la diversité de 25 clones issus de trois populations de Forastero du Pérou et d'Equateur à l'aide de

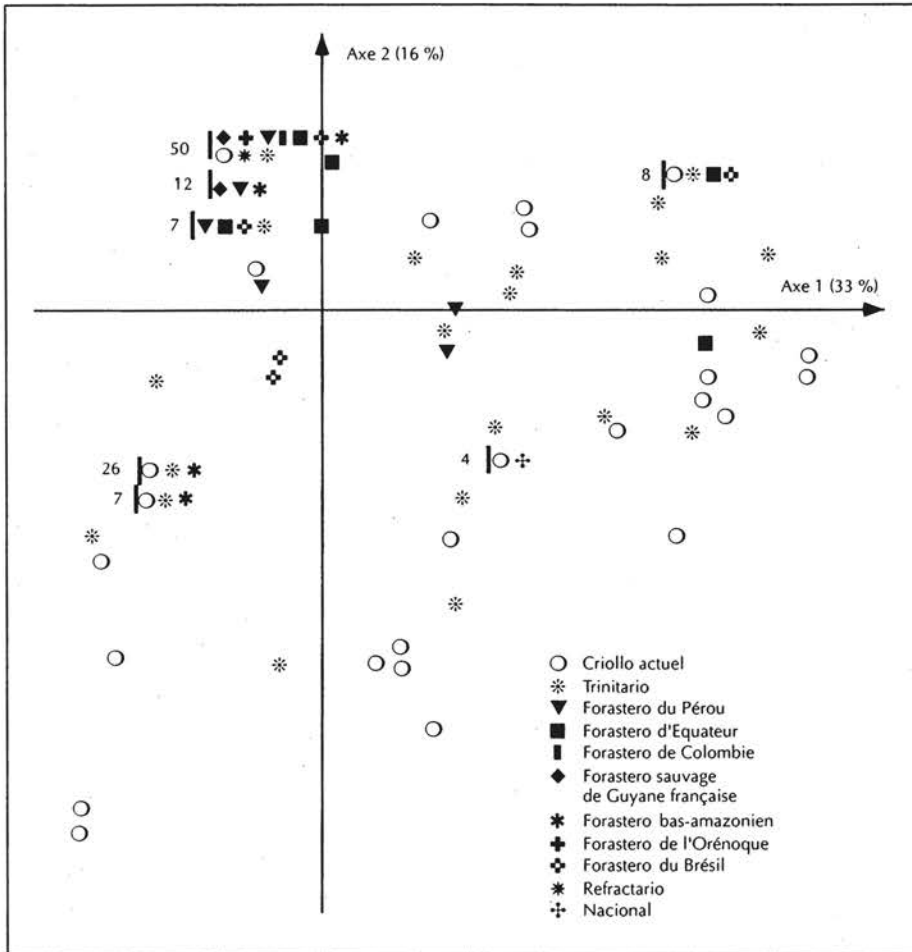


Figure 5. Diversité de l'ADN mitochondrial de 177 individus appartenant aux différents groupes morphogéographiques et représentée sur le premier plan d'une AFC, d'après LAURENT et al. (1993b).

9 amorces RAPD ayant généré 75 bandes. Malgré l'échantillonnage réduit, les indices de diversité de Shannon révèlent une diversité intrapopulation supérieure à la diversité interpopulation. Les analyses multivariées, AFC et CAH, révèlent une différenciation génétique entre les trois populations de Forastero haut-amazoniens en relation avec leur origine géographique.

N'GORAN *et al.* (1994) ont analysé la diversité génétique de 106 géotypes qui appartiennent aux différents groupes morphogéographiques incluant des variétés actuelles de Criollo, à l'aide de 19 amorces. Après un tri sévère des bandes, 49 bandes polymorphes et reproductibles ont été retenues pour les analyses. Sur 36 bandes hybridées à de l'ADN total restreint, 12 correspondent à des séquences hautement répétées et dispersées, 12, à des séquences

uniques et 12, à des séquences faiblement répétées. La CAH, établie à partir des coordonnées factorielles d'une AFC (figure 6), montre une structuration claire sur le premier plan entre Forastero et Criollo actuels, ainsi qu'une différenciation nette entre Forastero haut-amazoniens et Forastero bas-amazoniens (N'GORAN *et al.*, 1994). Si l'on ne considère que les bandes correspondant aux séquences uniques, aucune structuration claire n'apparaît. En revanche, une structuration apparaît entre Criollo actuels et Forastero en ne considérant que les bandes correspondant aux séquences hautement répétées alors qu'elles ne différencient pas du tout les Forastero bas-amazoniens des Forastero haut-amazoniens.

LERCETEAU *et al.* (1997) ont étudié la diversité de 155 clones appartenant aux groupes Nacional, Forastero, Trinitario et Criollo actuels, à l'aide de 40 bandes RAPD issues de 18 amorces. Une analyse en composantes principales (ACP) a permis de structurer la diversité. L'axe 2 met en évidence essentiellement la spécificité des Forastero sauvages de Guyane, tandis que le plan 1-3 révèle la structuration des autres populations. Malgré les variations continues d'un groupe à l'autre, dues aux brassages génétiques intenses, on observe une différenciation claire entre Nacional et Criollo. La mise en évidence de cette structuration ainsi que l'information apportée par les RFLP sur ces mêmes clones amènent les auteurs à penser que les clones de Nacional très homozygotes tels que les BCH et SA, issus de très vieilles plantations, pourraient représenter une partie du pool originel des Nacional. Ces clones se seraient ensuite largement hybridés avec d'autres clones introduits. Ce niveau d'homozygotie élevé pourrait être un caractère commun aux variétés pures de Nacional.

SOUNIGO *et al.* (1996) ont analysé par RAPD la diversité de 149 clones appartenant aux groupes Forastero, Trinitario et Refractario. Le calcul de l'indice de diversité de Shannon montre que les populations les plus variables sont LCTEEN (Forastero d'Equateur, 0,85), IMC (Forastero du Pérou, 0,72) et CL (Refractario, 0,71) et les moins variables sont les populations B, JA (Refractario, 0,29 et 0,47), NA (Forastero du Pérou, 0,42) et GU (Forastero de Guyane, 0,50). Un arbre additif construit à partir des distances de Rogers montre une structuration des populations en trois groupes. Le premier rassemble les trois populations du Pérou, NA, IMC, SCA. Le deuxième associe les populations de Forastero, PA et MO du Pérou, LCTEEN d'Equateur et GU de Guyane. Le troisième regroupe les populations de Refractario et de Trinitario.

Dans cette analyse à l'aide des RAPD, les populations de Forastero se regroupent de manière différente de celle qui est observée avec les marqueurs enzymatiques. En particulier, la population IMC est proche des autres populations de Forastero, alors qu'elle était associée à des Trinitario dans l'analyse isoenzymatique. La population SCA apparaît plus proche des IMC et des NA du Pérou, alors que, avec les isoenzymes, elle était plus proche de la population d'Equateur LCTEEN. La population de cacaoyers guyanais sauvages apparaît la plus distante de toutes les autres populations (figure 7), alors qu'avec les isoenzymes elle était assez proche des populations MO et NA.

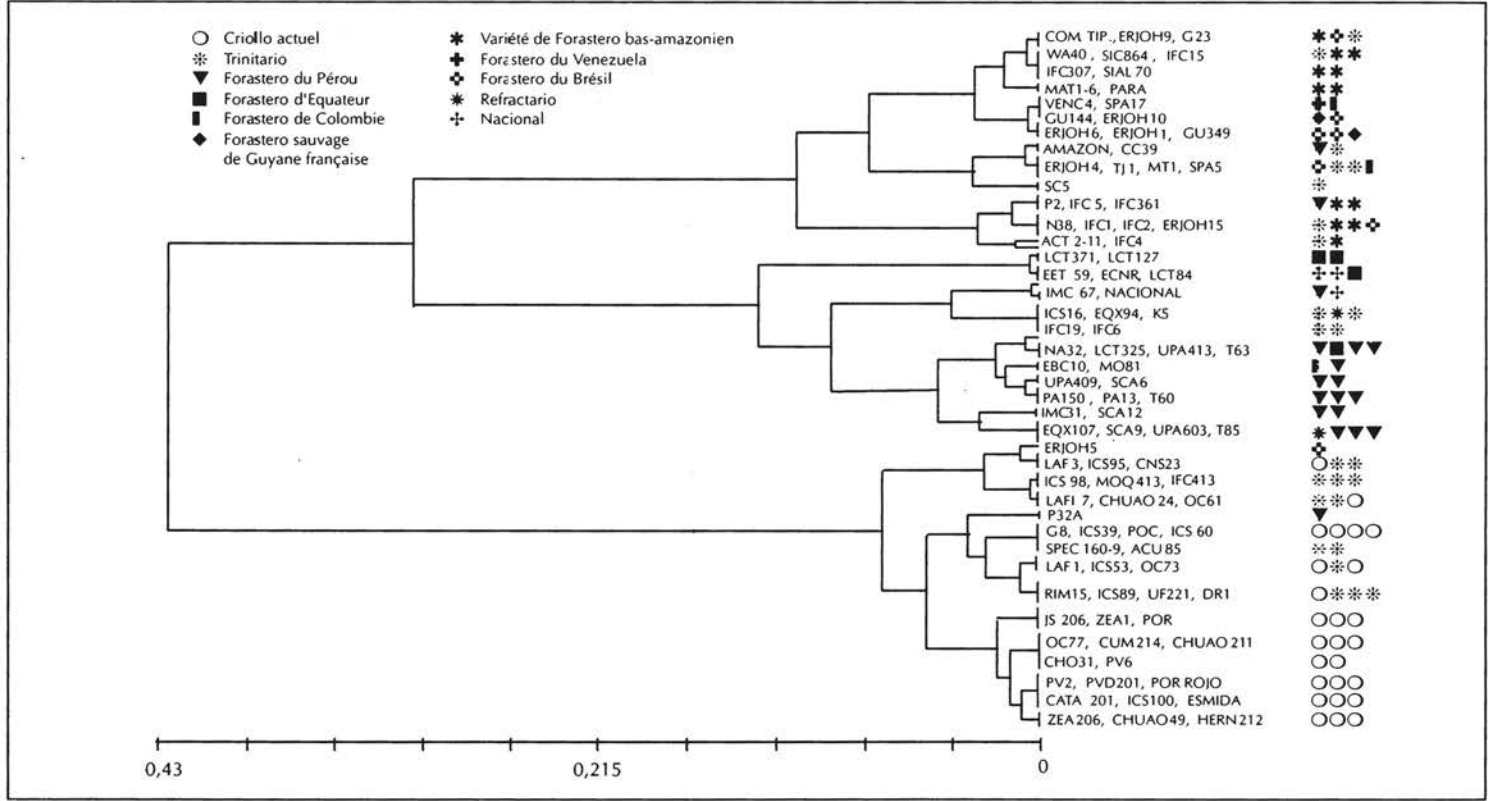


Figure 6. Analyse ascendante hiérarchique construite par la méthode du neighbor joining à partir des coordonnées des individus sur les sept premiers axes de l'AFC, d'après N'GORAN et al. (1994). La diversité des individus a été évaluée à l'aide de 49 bandes RAPD polymorphes.



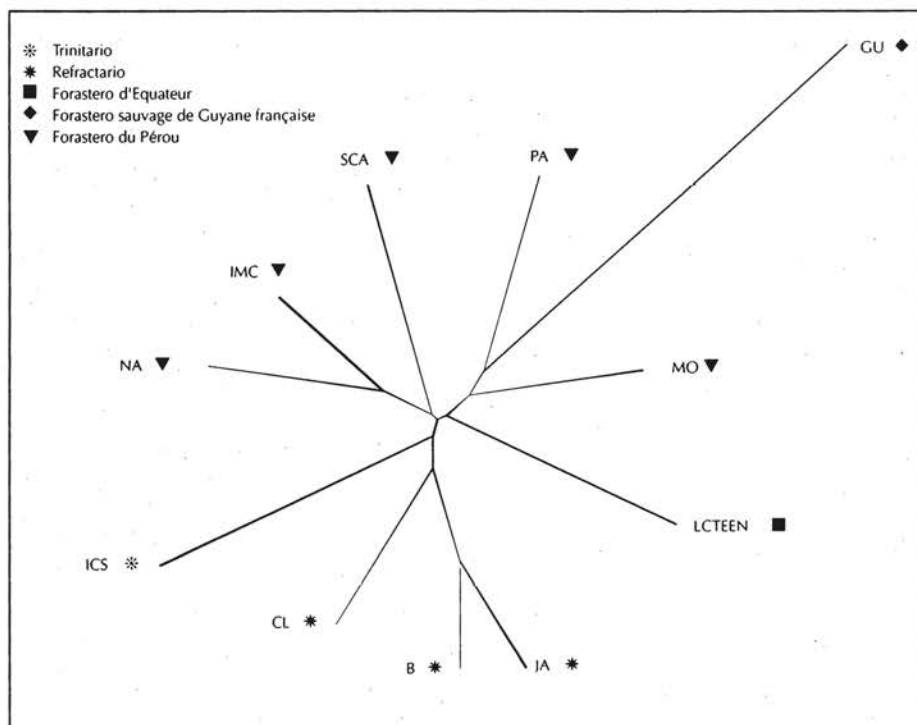


Figure 7. Arbre construit par la méthode du neighbor joining à partir des distances de Rogers obtenues sur les données RAPD de SOUNIGO et al. (1996).

DE LA CRUZ *et al.* (1995) ont analysé à l'aide des RAPD 42 génotypes correspondant à des plantes sauvages, provenant soit du Mexique (îlots forestiers du Yucatán et du Chiapas), soit de haute Amazonie, à des cultivars de Criollo, de Forastero et de Trinitario, et à un représentant d'un genre voisin *Herrania*. Les calculs d'indices de dissimilarité ont permis de construire un arbre phylogénétique selon la méthode des plus proches voisins. Ils montrent une plus grande similarité entre les Criollo cultivés et les cacaoyers sauvages d'Amérique du Sud qu'entre les Criollo cultivés et les plantes sauvages du Mexique. Les auteurs suggèrent que les arbres sauvages trouvés dans les anciens bois sacrés des Mayas n'existent pas dans les collections actuelles et qu'ils pourraient être les représentants les plus proches des anciens cultivars mayas.

Dans une étude plus récente, WHITKUS *et al.* (1998) ont analysé à l'aide de RAPD un échantillon de 86 individus, qui inclut, en particulier, 26 génotypes sauvages collectés dans la forêt Lacandona dans l'Etat du Chiapas et 5 individus provenant d'îlots forestiers du Yucatán, au Mexique, ainsi que des clones sauvages et cultivés d'Amérique centrale. Contrairement aux populations et aux cultivars d'Amérique centrale où la diversité intrapopulation est toujours plus élevée que la diversité interpopulation, la diversité trouvée entre les deux

populations mexicaines est plus élevée que celle de chaque région. Cette situation reflète le faible niveau de polymorphisme trouvé dans ces deux populations. Une structuration de la diversité semblable à celle de DE LA CRUZ *et al.* (1995) est observée. Les deux populations du Mexique apparaissent bien différenciées et originales. Ces résultats confortent l'hypothèse de CUATRECASAS (1964) quant à la répartition naturelle des cacaoyers en Amérique centrale. Les auteurs soulignent toutefois le manque d'affinité entre les Criollo cultivés et les plantes sauvages observées au Mexique. Selon eux, les cacaoyers collectés dans la forêt Lacandona pourraient représenter les cacaoyers sauvages. Les plants échantillonnés dans les îlots forestiers anciennement exploités par les Mayas pourraient en revanche y avoir été introduits à partir des populations sauvages. Dans ce cas, ils représenteraient un sous-échantillonnage des populations présentes dans la forêt Lacandona, ce qui expliquerait la différence entre les deux populations mexicaines. Ils correspondraient aussi au matériel le plus proche des cacaoyers cultivés par les Mayas.

Nos propres analyses, réalisées récemment à la fois sur la population de la forêt Lacandona et sur celle du Yucatán, indiquent, contrairement à WHITKUS *et al.* (1998), une parfaite similitude génétique entre ces deux populations. Cette contradiction trouve peut-être son origine dans le fait que les analyses de WHITKUS *et al.* (1998) ont probablement inclus dans leurs analyses de jeunes plants sauvages, morphologiquement très semblables à ceux de *T. cacao*, mais qui n'appartiennent pas à cette espèce. Ces plants, observés dans le Yucatán, pourraient expliquer les distances génétiques obtenues par WHITKUS *et al.* (1998) entre les cacaoyers issus des deux localités.

## Conclusion

Les brassages génétiques intenses survenus au cours des trois derniers siècles entre les populations de cacaoyers sauvages et cultivés rendent souvent difficile leur classification.

Un certain nombre de descripteurs morphologiques permettent de différencier des populations d'origine géographique différente et de structurer leur diversité. Cependant, ils donnent parfois une image biaisée de la diversité des ressources génétiques. Ainsi au cours de nos analyses, il est apparu que les variétés anciennes de Criollo, très contrastées quant à leur morphologie, qui va du type Porcelana aux cabosses lisses au type Pentagona aux cabosses particulièrement verruqueuses, étaient très homogènes du point de vue génétique et très homozygotes. La sélection humaine a pu dans ce cas participer à la fixation et à la conservation de types morphologiques très différents résultant de mutations ponctuelles.

Les marqueurs génétiques, qui reflètent la structure de l'ensemble du génome, sont des outils de choix pour révéler la diversité et les apparentements entre cacaoyers sauvages ou cultivés. Les différents types de marqueurs, nucléaires (RAPD, RFLP provenant de sondes génomiques ou d'ADNc, microsatellites) ou cytoplasmiques, apportent chacun des informations différentes et complémentaires en révélant la variation de séquences dont les vitesses d'évolution sont différentes. Les marqueurs codominants tels que les RFLP et les isoenzymes ont fourni une information plus précise sur la structure génétique des populations et des individus.

L'identification des allèles présents dans les géotypes ainsi que les gradients observés dans les niveaux d'hétérozygotie des clones au sein de populations cultivées hybrides telles que les Nacional d'Equateur ou les Criollo et les Trinitario ont permis de comprendre la genèse de ces populations et d'identifier leurs ancêtres les plus probables (LERCETEAU *et al.*, 1997 ; MOTAMAYOR *et al.*, 1997). Dans les deux cas, ces ancêtres sont probablement constitués par une base génétique très étroite.

Une forte diversité génétique a été mise en évidence entre les populations sauvages de Forastero et au sein de certaines de ces populations originaires de haute Amazonie, d'Equateur en particulier. Une variation continue peut être observée entre elles. Cependant, dans plusieurs analyses, les Forastero de Guyane apparaissent différenciés des autres types de Forastero, y compris des Forastero de basse Amazonie ou de l'Orénoque, dont ils sont les plus proches géographiquement. Une structuration entre Forastero et Criollo a été observée avec l'ensemble des marqueurs, avec une différenciation encore plus marquée entre Forastero et Criollo anciens révélée par les RFLP nucléaires. L'analyse des Nacional a aussi montré leur originalité à la fois par rapport aux Forastero et par rapport aux Criollo, dont ils se distinguent nettement.

Ainsi, la diversité de l'espèce pourrait se structurer, de façon schématique, autour de quatre pôles d'importance inégale : les Criollo anciens, les variétés pures de Nacional, les Forastero sauvages de Guyane et un pool plus large constitué des Forastero amazoniens (de haute et de basse Amazonie) et de l'Orénoque, même si ce pool comprend des populations qui se différencient les unes des autres selon leur origine géographique. Cette structuration diffère de la classification habituelle des cacaoyers en trois groupes : Forastero, Criollo et Trinitario. En effet, d'une part la structure génétique des Trinitario est semblable à celle du groupe des Criollo qui, dans cette classification, englobe une majorité de Criollo actuels. D'autre part, au sein du groupe appelé Forastero sont inclus des cacaoyers bien différenciés tels que les populations sauvages de Guyane ou ceux de la variété Nacional d'origine.

La structuration que nous proposons peut être étayée par l'histoire paléoclimatique du continent sud-américain à l'ère quaternaire, époque à laquelle des bouleversements écologiques considérables se sont produits. La succession de périodes de glaciation entrecoupées de phases de réchauffement du climat

s'est traduite, dans les régions tropicales, par une alternance de périodes sèches et de périodes humides. Durant les périodes de sécheresse, la forêt tropicale s'est rétrécie jusqu'à ne plus former que des îlots forestiers, favorisant ainsi la différenciation de populations et la disparition de certaines autres. Au cours des phases humides, les forêts se sont étendues de nouveau à partir de ces zones refuges. De telles zones refuges ont été identifiées en Guyane, au Brésil, en Colombie, au Venezuela ainsi qu'en haute Amazonie, où une vaste zone refuge a été mise en évidence. Ce phénomène, avéré pour de nombreuses espèces animales et végétales (SIMPSON et HAFFER, 1978), pourrait expliquer la diversité et la différenciation de certaines populations de cacaoyers. L'étude moléculaire de la séquence de certains gènes devrait permettre d'affiner l'analyse des apparentements et de la différenciation entre populations et de dater leur temps de divergence ; il sera ainsi possible de situer la différenciation des populations par rapport aux grands événements écologiques et géographiques.

On peut souligner l'importance de l'échantillonnage dans les conclusions tirées des différentes analyses. En effet, les collectes de ressources génétiques ont souvent été réalisées avec des critères d'intérêt agronomique, qui peuvent donner une image biaisée de la diversité réellement présente dans une région donnée : c'est le cas des Forastero collectés au Pérou avec un objectif bien précis, celui de la résistance au balai de sorcière. Pour les Criollo, les premières études n'ont porté que sur les clones présents dans les collections et collectés aussi pour leurs caractères agronomiques. La forte diversité révélée au sein de ce groupe reflétait en fait la diversité des introgressions réalisées à partir d'un type unique, ancestral, identifié par la suite grâce aux marqueurs et aux prospections effectuées directement dans les plus vieilles plantations ou près des vestiges des civilisations mayas, les premières à avoir domestiqué les cacaoyers. Un plus large échantillonnage et une meilleure représentation de certaines populations permettraient de conforter la structuration entre les différentes origines de cacaoyer qui est apparue dans ces analyses.

Ces analyses ont confirmé que seule une faible part de la diversité génétique est exploitée en sélection, comme l'ont mentionné plusieurs auteurs (BARTLEY, 1979 ; LOCKWOOD et END, 1992). Ainsi, les formes hybrides entre Criollo et Forastero, qui correspondent le plus souvent aux formes traditionnelles de Trinitario, ont impliqué une base génétique très réduite dans chacun des deux groupes d'origine. D'autres types d'hybrides entre Criollo et Forastero exploitant toute la richesse génétique des Forastero pourraient donc être utilisés en sélection. Par ailleurs, peu de méioses séparent les Criollo anciens de la formation des types hybrides, et les déséquilibres de liaison entre caractères et marqueurs devraient être maintenus. Cette situation devrait faciliter l'exploitation des ressources génétiques de type Criollo et Trinitario en utilisant les marqueurs qui encadrent les gènes d'intérêt des Criollo comme marqueurs de sélection précoce. De même pour les Forastero, une très faible part de la diversité génétique a été réellement mise à contribution dans les programmes de

sélection, et un nombre réduit de Forastero du Pérou, ayant souvent des liens forts de parenté, ont été largement diffusés et intégrés dans les programmes de sélection de tous les pays producteurs.

Même si la diversité présente dans les collections est encore loin d'avoir été exploitée, l'importance de la diversité révélée sur un échantillon parfois restreint de certaines populations plaide en faveur de la poursuite des prospections dans les régions riches en diversité pour préserver les ressources génétiques des cacaoyers.

## Annexe

### Matériel végétal

Les analyses enzymatiques de LANAUD (1987) ont porté sur 12 Criollo actuels, 17 Trinitario sélectionnés en Amérique, 22 Trinitario sélectionnés en Afrique, 64 Forastero bas-amazoniens sélectionnés en Afrique, 19 Forastero prospectés par LANAUD (1986) le long de l'Orénoque au Venezuela (VENC), 92 Forastero sauvages prospectés par SALLEE (1987) en Guyane française, 19 cacaoyers cultivés prospectés par CLEMENT (1986) en Guyane française, 74 Forastero prospectés par POUND (1938 ; 1943) au Pérou (39 individus GO, 8 IMC, 11 P, 7 NA, 8 PA, 1 SCA), 10 Refractario collectés en Equateur par CHALMERS (1968) (6 individus EQX) et par POUND (1938) (4 individus MOQ), 7 Forastero prospectés lors de l'expédition colombienne d'OCAMPO (1985) en Colombie (EBC), 18 Forastero prospectés par J.B. Allen entre 1980 et 1985 en Equateur (LCTEEN).

Les analyses enzymatiques de SOUNIGO *et al.* (1996) ont porté sur 482 clones de cacaoyer appartenant à 28 populations ou groupes d'accessions présents dans la collection de la Trinité : 6 populations de Forastero prospectées au Pérou (IMC, MO, PA, NA, P, SCA : 112 individus), 1 population de Forastero provenant d'Equateur (LCTEEN : 16 individus), 1 population de Forastero prospectée en Guyane française (16 individus), 9 populations correspondant à des Refractario (AM, B, CL, JA, LP, LX, MOQ, SJ, SLA : 245 individus), 10 populations de Trinitario (ACT, CC, DOM, ICS, GS, MAR, R, SC, TRD, UF : 114 individus).

Les analyses par RFLP (LAURENT *et al.*, 1993b ; 1994) ont porté sur 201 génotypes appartenant aux différents groupes morphogéographiques : 45 Criollo actuels (ICS, CATA, BOC, CHUAO, OC, POR, JS, MT, PV, ZEA, LAF), 20 Forastero du Pérou (P, PA, NA, MO, IMC, SCA), 12 Forastero sélectionnés en Afrique à partir de clones du Pérou (T, UPA), 11 Forastero d'Equateur (LCTEEN), 6 Forastero de Colombie (EBC, SPA), 7 Forastero de la vallée de l'Orénoque au Venezuela (VENC), 4 Forastero sauvages de Guyane française (GU), 22 Forastero du Brésil (ERJOH, Comun, Pará, SIAL, SIC, MAT, Catongo), 9 Forastero bas-amazoniens et sélectionnés en Afrique (Amelonado) (IFC, SF), 37 Trinitario sélectionnés en Amérique du Centre et du Sud et dans la Caraïbe (ACT, ICS, GS, CHUAO, CNS, WA, RIM, MT, TJ, SC, SGU, CC, UF), 4 Trinitario sélectionnés en Asie (WA, LAFI, DR, G), 12 Trinitario sélectionnés en Afrique (SNK, IFC, N, W, ACU, K), 10 Refractario (EQX, MOQ), 2 Nacional.

### Analyses enzymatiques

Les méthodes utilisées ainsi que le déterminisme génétique et les protocoles d'extraction et de révélation sont indiqués dans LANAUD (1986a). Dans les analyses de LANAUD (1987), 6 systèmes enzymatiques (PGI, PGM, ADH, MDH, PAC, ICD) polymorphes, correspondant à 9 locus, ont été utilisés. Les analyses de SOUNIGO *et al.* (1996) ont porté sur 5 systèmes enzymatiques (PGI, ADH, MDH, PAC, ICD) polymorphes, correspondant à 6 locus.

### Analyses par RFLP

Une banque génomique et une banque d'ADNc ont été construites. Entre 31 et 50 de ces sondes, pour la plupart cartographiées (LANAUD *et al.*, 1995), ont été utilisées pour étudier la diversité nucléaire. Le niveau d'hétérozygotie des clones a été déterminé à

l'aide de marqueurs cartographiés dont le déterminisme génétique est connu. Par ailleurs, des sondes hétérologues ont été mises en œuvre pour étudier la diversité cytoplasmique (mitochondriale et chloroplastique). Les protocoles sont décrits dans LAURENT *et al.* (1993a ; 1993b ; 1994).

### Analyses par RAPD

Après une sélection sévère des amorces issues des kits Operon qui génèrent des bandes reproductibles, 19 amorces produisant 49 bandes polymorphes ont été utilisées pour analyser la diversité (N'GORAN *et al.*, 1994). La nature — séquence unique ou répétée — des bandes RAPD observées a été testée par hybridation Southern sur de l'ADN total de cacaoyer restreint par des enzymes de restriction.

### Analyses des données

Plusieurs paramètres génétiques de la diversité ont été calculés sur les données enzymatiques et moléculaires : pourcentage de locus polymorphes, pourcentage d'hétérozygotie, nombre moyen d'allèles par locus, indice de diversité de Shannon fondé sur les fréquences génotypiques, indice de diversité de Nei (NEI, 1978) fondé sur les fréquences alléliques.

A partir des indices de Nei et de Shannon, la diversité totale a été décomposée en diversité intrapopulation et diversité interpopulation selon les formules :

$$H_{\text{intrapop}} = H_{\text{pop}} / H_{\text{total}}$$

$$H_{\text{interpop}} = (H_{\text{total}} - H_{\text{pop}}) / H_{\text{total}}$$

Des analyses multivariées — analyses factorielles des correspondances, AFC (BENZECRI, 1973), ou analyses en composantes principales, ACP, pour les caractères quantitatifs — ont été généralement effectuées pour toutes les données.

Pour les données enzymatiques, les distances génétiques de Nei sur les fréquences alléliques ont été utilisées pour mesurer les distances entre populations. Une classification ascendante hiérarchique, CAH, a été construite selon la méthode UPGMA.

Pour les données RAPD (N'GORAN *et al.*, 1994), une CAH a été construite à partir des coordonnées des clones sur les sept premiers axes de l'AFC.

Pour les données RAPD de SOUNIGO *et al.* (1996), les distances de Rogers et Tanimoto ont été calculées, et un arbre a été construit selon la méthode du *neighbor joining*.

## Références bibliographiques

- ALLEN J.B., LASS R.A., 1983. London cocoa trade Amazon project: final report, phase 1. *Cocoa Grower's Bulletin*, 34 : 1-71.
- AMEFIA Y.K., 1986. Sur le polymorphisme enzymatique des cacaoyers introduits au Togo. Thèse de doctorat, INA, Paris-Grignon, France, 146 p.
- ATKINSON M.D., WITHERS L.A., SIMPSON M.J.A., 1986. Characterisation of cacao germplasm using isoenzyme markers. 1. A preliminary survey of diversity using starch gel electrophoresis and standardisation of the procedure. *Euphytica*, 35 : 741-750.
- BARRIGA J.P., MACHADO P.R.F., DE ALMEIDA C.M.V.C., DE ALMEIDA C.F.G., 1984. A preservação e utilização dos recursos genéticos de cacau na Amazônia brasileira. *In* : IX<sup>e</sup> conférence internationale sur la recherche cacaoyère. Lagos, Nigeria, Cocoa Producers' Alliance, p. 73-79.
- BARTLEY B.G., 1979. Global concepts for genetic resources and breeding in cacao. *In* : VII<sup>e</sup> conférence internationale sur la recherche cacaoyère. Lagos, Nigeria, Cocoa Producers' Alliance, p. 519-525.
- BEKELE F.L., 1992. Use of botanical descriptors for cocoa characterization: CRU experiences. *In* : International workshop on conservation, characterisation and utilisation of cocoa genetic resources in the 21st century. Port of Spain, Trinité-et-Tobago, CRU, p. 77-91.
- BEKELE F., BEKELE I., 1998. A sampling of the phenetic diversity of cacao in the international cocoa gene bank of Trinidad. *Crop Science*, 36 : 57-64.
- BEKELE F.L., KENNEDY A.J., MCDAVID C., LAUCKNER B., BEKELE I., 1994. Numerical taxonomic studies on cacao (*Theobroma cacao* L.) in Trinidad. *Euphytica*, 75 : 231-240.
- BENZECRI J.P., 1973. L'analyse des données. 2. L'analyse des correspondances. Paris, France, Dunod, 616 p.
- BERGMAN J.F., 1969. The distribution of cacao cultivation in pre-Columbian America. *Annals of the Association of American Geographers*, 59 : 85-96.
- BERNOULLI G., 1869. Übersicht der bis jetzt bekannten Arten von *Theobroma*. *Neue Denkschriften der allgemeinen schweizerischen Gesellschaft für die gesammten Naturwissenschaften*, 24 : 1-15.
- BOUHARMONT J., 1960. Recherches cytologiques sur la fructification et l'incompatibilité chez *Theobroma cacao* L. INEAC, Série scientifique n° 89, 117 p.
- BRAUDEAU J., 1969. Le cacaoyer. Paris, France, Maisonneuve et Larose, 304 p.
- BURLE L., 1952. La production du cacao en Afrique occidentale française. Centre de recherches agronomiques de Bingerville, Bulletin n° 5, p. 3-21.
- CHALMERS W.S., 1968. Cacao collecting trip to the Oriente of Ecuador (November 1968). Port of Spain, Trinité-et-Tobago, CRU, 4 p.
- CHEESMAN E.E., 1944. Notes on the nomenclature, classification and possible relationships of cacao populations. *Tropical Agriculture*, 21 : 144-159.



- CHEVALIER A., 1946. Révision du genre *Theobroma* d'après l'herbier du Muséum national d'histoire naturelle de Paris. *Revue de botanique appliquée*, 26 : 265-285.
- CLEMENT D., 1986. Cacaoyers de Guyane : prospections. *Café, cacao, thé*, 30 : 11-35.
- COE S.D., COE M.D., 1996. *The true history of chocolate*. Londres, Royaume-Uni, Thames and Hudson, 280 p.
- COPE F.W., 1962. The mechanism of pollen incompatibility in *Theobroma cacao*. *Heredity*, 17 : 157-182.
- DE LA CRUZ M., WHITKUS R., GOMEZ-POMPA A., MOTA-BRAVO L., 1995. Origins of cacao cultivation. *Nature*, 375 : 542-543.
- CUATRECASAS J., 1964. Cacao and its allies: a taxonomic revision of the genus *Theobroma*. *Contributions from the United States Herbarium*, 35 : 379-614.
- END M.J., WADSWORTH R.M., HADLEY P., 1992. The international cocoa germplasm database. *In* : International workshop on conservation, characterisation and utilisation of cocoa genetic resources in the 21st century. Port of Spain, Trinité-et-Tobago, CRU, p. 48-52.
- ENGELS J.M.M., 1983a. A systematic description of cacao clones. 1. The discriminative value of quantitative characteristics. *Euphytica*, 32 : 377-385.
- ENGELS J.M.M., 1983b. A systematic description of cacao clones. 2. The discriminative value of qualitative characteristics and the practical compatibility of the discriminative value of quantitative and qualitative descriptors. *Euphytica*, 32 : 719-733.
- ENGELS J.M.M., 1986. The systematic description of cacao clones and its significance for taxonomy and plant breeding. Thèse PhD, Agricultural University, Wageningen, Pays-Bas, 125 p.
- ENGELS J.M.M., BARTLEY B.G.D., ENRIQUEZ G.A., 1980. Cacao descriptors their states and modus operandi. *Turrialba*, 30 : 209-218.
- ENRIQUEZ G.A., 1985. *Cursos sobre el cultivo del cacao*. Turrialba, Costa Rica, CATIE, Serie Materiales de Enseñanza, n° 22, 239 p.
- ENRIQUEZ G.A., 1992. Characteristics of cacao Nacional of Ecuador. *In* : International workshop on conservation, characterisation and utilisation of cocoa genetic resources in the 21st century. Port of Spain, Trinité-et-Tobago, CRU, p. 269-278.
- ENRIQUEZ G.A., SORIA J., 1968. The variability of certain bean characteristics of cacao (*Theobroma cacao* L.). *Euphytica*, 17 : 114-120.
- FIGUEIRA A., JANICK J., GOLDSBROUGH P., 1992. Genome size and DNA polymorphism in *Theobroma cacao*. *Journal of the American Society for Horticultural Science*, 117 : 673-677.
- FIGUEIRA A., JANICK J., LEVY M., GOLDSBROUGH P., 1994. Re-examining the classification of *Theobroma cacao* L. using molecular markers. *Journal of the American Society for Horticultural Science*, 119 : 1073-1082.
- GLENDINNING D.R., 1966. The incompatibility alleles of cocoa: report for the period 1st October 1963-31st March 1965. Ghana, Cocoa Research Institute.

- GUISAN, 1825. De la culture du cacaoyer. *In* : Traité sur les terres noyées de Guyane (2<sup>e</sup> éd.). Imprimerie du Roy, p. 170-189.
- HAMRICK J.L., GODT M.J.W., 1990. Allozyme variation in plant species. *In* : Plant population genetics, breeding, and genetic resources, A.H.D. Brown *et al.* éd., Sunderland, Etats-Unis, Sinauer, p. 43-63.
- HAMRICK J.L., GODT M.J.W., SHERMAN-BROYLES S.L., 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests*, 6 : 95-124.
- IBPGR, 1981. Report of the IBPGR working group on genetic resources of cocoa. Rome, Italie, IBPGR, ACP-IBPGR/80/56, 28 p.
- KNIGHT R., ROGERS H., 1955. Incompatibility in *Theobroma cacao*. *Heredity*, 9 : 67-69.
- LACHENAUD P., SALLEE S., 1993. Les cacaoyers spontanés de Guyane : localisation, écologie et morphologie. *Café, cacao, thé*, 37 : 101-114.
- LANAUD C., 1986a. Utilisation des marqueurs enzymatiques pour l'étude génétique du cacaoyer *Theobroma cacao* L. 1. Contrôle génétique et linkage de neuf marqueurs enzymatiques. *Café, cacao, thé*, 30 : 259-270.
- LANAUD C., 1986b. Utilisation des marqueurs enzymatiques pour l'étude génétique du cacaoyer *Theobroma cacao* L. 2. Etude du polymorphisme de six systèmes enzymatiques. *Café, cacao, thé*, 30 : 271-280.
- LANAUD C., 1987. Nouvelles données sur la biologie du cacaoyer (*Theobroma cacao* L.) : diversité des populations, systèmes d'incompatibilité, haploïdes spontanés ; leurs conséquences pour l'amélioration génétique de cette espèce. Thèse de doctorat, université Paris XI, Orsay, France, 106 p.
- LANAUD C., BERTAUD J., 1984. Mise en évidence de nouveaux marqueurs génétiques chez *Theobroma cacao* L. par les techniques d'électrophorèse. *In* : IX<sup>e</sup> conférence internationale sur la recherche cacaoyère. Lagos, Nigeria, Cocoa Producers' Alliance, p. 249-253.
- LANAUD C., HAMON P., DUPERRAY C., 1992. Estimation of nuclear DNA content of *Theobroma cacao* L. by flow cytometry. *Café, cacao, thé*, 36 : 3-8.
- LANAUD C., RISTERUCCI A.M., N'GORAN A.K.J., CLEMENT D., FLAMENT M.H., LAURENT V., FALQUE M., 1995. A genetic linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics*, 91 : 987-993.
- LAURENT V., RISTERUCCI A.M., LANAUD C., 1993a. Variability for nuclear ribosomal genes within *Theobroma cacao*. *Heredity*, 71 : 96-103.
- LAURENT V., RISTERUCCI A.M., LANAUD C., 1993b. Chloroplast and mitochondrial DNA diversity in *Theobroma cacao*. *Theoretical and Applied Genetics*, 87 : 81-88.
- LAURENT V., RISTERUCCI A.M., LANAUD C., 1994. Genetic diversity in cacao revealed by cDNA probes. *Theoretical and Applied Genetics*, 88 : 193-198.
- LECOMTE H., CHALLOT C., 1897. Le cacaoyer et sa culture. Paris, France, Carré et Naud, 121 p.
- LERCETEAU E., ROBERT T., PETIARD V., CROUZILLAT D., 1997. Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD and RFLP markers. *Theoretical and Applied Genetics*, 95 : 10-19.

- LOCKWOOD G., END M., 1992. History, technique and future needs for cacao collection. *In* : International workshop on conservation, characterisation and utilisation of cocoa genetic resources in the 21st century. Port of Spain, Trinité-et-Tobago, CRU, p. 1-14.
- MORRIS D., 1882. Cacao: how to grow and how to cure it. Jamaica, p. 1-45.
- MOTAMAYOR J.C., RISTERUCCI A.M., LAURENT V., MORENO A., LANAUD C., 1997. The genetic diversity of Criollo cacao and the consequences for quality breeding. *In* : 1<sup>o</sup> congreso venezolano del cacao y su industria. Aragua, Venezuela.
- MYERS J.G., 1930. Notes on wild cacao in Surinam and in British Guiana. *Kew Bulletin*, 1 : 1-10.
- NEI M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89 : 583-590.
- N'GORAN J.A.K., 1994. Contribution à l'étude génétique du cacaoyer par les marqueurs moléculaires : diversité génétique et recherche de QTLs. Thèse de doctorat, université Montpellier II, Montpellier, France, 131 p.
- N'GORAN J.A.K., LAURENT V., RISTERUCCI A.M., LANAUD C., 1994. Comparative genetic diversity of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity*, 73 : 589-597.
- OCAMPO R.F., 1985. Informe sobre recolección de germoplasma de *Theobroma cacao* L. El Cacaotero Colombiano, 31 : 24-29.
- PARADIS L., 1979. Le cacao précolombien : monnaie d'échange et breuvage des dieux. *Journal d'agriculture traditionnelle et de botanique appliquée*, 26 : 3-4.
- PITTIER H., 1930. A propos des cacaoyers spontanés. *Revue de botanique appliquée*, 10 : 777.
- PITTIER H., 1933. Degeneration of cacao through natural hybridization. *Journal of Heredity*, 36 : 385-90.
- POUND F.J., 1938. Cocoa and witches' broom disease (*Marasmius perniciosus*) of South America with notes on other species of *Theobroma*. Port of Spain, Trinité-et-Tobago, Yuille's Printery.
- POUND F.J., 1943. Cocoa and witches' broom disease: report on a recent visit to the Amazon territory of Peru. *In* : The archives of cocoa research, volume 1, H. Toxopeus éd., Londres, Royaume-Uni, ACRI-IOCC.
- RABOUIN L.M., PAULIN D., 1993. Etude de l'héritabilité de descripteurs floraux. *In* : XI<sup>e</sup> conférence internationale sur la recherche cacaoyère. Lagos, Nigeria, Cocoa Producers' Alliance, p. 449-450.
- REYES H., 1992. Criollo cacao germplasm in Venezuela. *In* : International workshop on conservation, characterisation and utilisation of cocoa genetic resources in the 21st century. Port of Spain, Trinité-et-Tobago, CRU.
- RISTERUCCI A.M., LANAUD C., N'GORAN J.A.K., PIERETTI I., 1996. A saturated linkage map of *Theobroma cacao*. *In* : XII<sup>e</sup> conférence internationale sur la recherche cacaoyère. Lagos, Nigeria, Cocoa Producers' Alliance.
- RONNING C.M., SCHNELL R.J., 1994. Allozyme diversity in a germplasm collection of *Theobroma cacao* L. *Journal of Heredity*, 85 : 291-295.

- RONNING C.M., SCHNELL R.J., KUHN D.N., 1995. Inheritance of random amplified polymorphic DNA (RAPD) markers in *Theobroma cacao* L. *Journal of the American Society for Horticultural Science*, 120 : 681-686.
- RUSSEL J.R., HOSEIN F., WAUGH R., POWELL W., 1993. Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. *Molecular Ecology*, 2 : 89-97.
- SALLEE B., 1987. Compte rendu de prospection dans les populations naturelles de *Theobroma cacao* L. du Haut Camopi (Guyane française). Kourou, Guyane, Cirad, 45 p.
- SCHUMANN K., 1886. Fam. Sterculiaceae g. *Theobroma*. In : *Flora brasiliensis*, volume XII, part III, p. 70-77.
- SIMPSON B.B., HAFFER J., 1978. Speciation patterns in the Amazonian forest biota. *Annual Review of Ecology and Systematics*, 9 : 497-518.
- SORIA J., 1970a. Principal varieties of cocoa cultivated in tropical America. *Cocoa Growers' Bulletin*, n° 19 : 12-21.
- SORIA J., 1970b. The present status and perspectives for cocoa cultivars in Latin America. *Proceedings of the Tropical Region American Society for Horticultural Science*, 14.
- SORIA J., 1973. Survey of crop genetics resources in their centers of diversity: first report. Rome, Italie, FAO, p. 119-125.
- SORIA J., ENRIQUEZ G.A., 1981. International cacao cultivar catalogue. Turrialba, Costa Rica, CATIE, Technical Series, Technical Bulletin n° 6, 156 p.
- SOUNIGO O., CHRISTOPHER Y., UMAHARAN R., 1996. Genetic diversity assessment of *Theobroma cacao* L. using isoenzyme and RAPD analyses. In : *Annual report 1996*, CRU. Port of Spain, Trinité-et-Tobago, CRU, p. 35-51.
- TOXOPEUS H., 1985. Botany, types and populations in cocoa. In : *Cocoa* (4th ed.), G.A.R. Wood et R.A. Lass éd., Londres, Royaume-Uni, Longman, p. 11-37.
- WARREN J.M., 1994. Isozyme variation in a number of populations of *Theobroma cacao* L. obtained through various sampling regimes. *Euphytica*, 72 : 121-126.
- WHITKUS R., DE LA CRUZ M., MOTA-BRAVO L., 1998. Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico. *Theoretical and Applied Genetics*, 96 : 621-627.
- WILDE J., WAUGH R., POWELL W., 1992. Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. *Theoretical and Applied Genetics*, 83 : 871-877.
- WOOD G.A.R., 1991. A history of early cocoa introductions. *Cocoa Growers' Bulletin*, 44 : 7-12.
- YEOH H.H., CHUNG D.K., FRITZ P.J., 1990. *Theobroma cacao* chloroplast DNA: isolation, molecular cloning and characterization. *Café, cacao, thé*, 34 : 173-178.
- YIDANA J.A., KENNEDY A.J., WITHERS L.A., 1987. Variation in peroxidase isozymes of cocoa (*Theobroma cacao* L.). In : *X<sup>e</sup> conférence internationale sur la recherche cacaoyère*. Lagos, Nigeria, Cocoa Producers' Alliance, p. 719-723.
- YOUNG A.M., 1994. *The chocolate tree: a natural history of cacao*. Washington, Etats-Unis, Smithsonian Institution Press, 200 p.



# Le caféier, *Coffea canephora*

---

Stéphane Dussert, Philippe Lashermes,  
François Anthony, Christophe Montagnon,  
Pierre Trouslot, Marie-Christine Combes,  
Julien Berthaud, Michel Noiroto, Serge Hamon

Le café est le premier produit agricole d'exportation (CHARRIER et ESKES, 1997). Il est produit par deux espèces : *Coffea arabica* L. et *C. canephora* Pierre. *C. arabica* est connu pour ses qualités gustatives. Sa culture s'étend sur les hauts plateaux tropicaux humides, en Amérique latine et en Afrique de l'Est, essentiellement. *C. canephora* est renommé pour sa robustesse agronomique, d'où son nom commun de Robusta. Il est cultivé principalement dans les zones tropicales humides de basse altitude et représente 30 % de la production mondiale de café. Il provient surtout du Brésil, d'Indonésie et de Côte d'Ivoire. Sa production en Asie du Sud-Est — Philippines et Vietnam — et en Inde est en expansion.

Au cours des XVIII<sup>e</sup> et XIX<sup>e</sup> siècles, seul le café Arabica a été produit et ce, principalement en Amérique tropicale, dans la Caraïbe et en Asie (CHARRIER et ESKES, 1997). Cependant, cette espèce est apparue très sensible aux aléas parasitaires, notamment à la rouille orangée. C'est pourquoi, en Afrique, au cours du XIX<sup>e</sup> siècle, les formes spontanées d'autres espèces de caféiers, notamment *C. canephora*, ont été mises en culture localement. Pour *C. canephora*, c'est surtout au Congo belge (actuellement, la République démocratique du Congo) et en Ouganda que des caféiers issus de populations sylvestres locales, de type Robusta, ont été cultivés. Ils ont été transférés à Java, principal centre de sélection de *C. canephora* de 1900 à 1930 (MONTAGNON *et al.*, 1998). A la même époque, en Afrique, l'utilisation des formes spontanées locales a permis

d'élargir l'éventail du matériel mis en culture : Kouilou en Côte d'Ivoire, Niaouli au Togo et au Bénin, Nana en République centrafricaine. Le matériel sélectionné à Java a été réintroduit au Congo belge vers 1916 à l'INEAC (Institut national pour l'étude agronomique du Congo belge), qui est devenu le principal centre de sélection de *C. canephora* de 1930 à 1960 (MONTAGNON *et al.*, 1998). Cependant, si les performances globales des arbres cultivés ont sensiblement augmenté à travers ces quelques cycles de sélection à Java et au Congo belge, les cultivars n'en sont pas moins restés très proches génétiquement des individus des populations naturelles d'origine. Par ailleurs, dans les pays africains de l'aire d'origine de l'espèce où *C. canephora* est cultivé, des formes spontanées locales ont pu se croiser avec les formes introduites et des plantes cultivées ont pu retourner vers des formes sauvages.

## La botanique et les ressources génétiques

### La botanique et le mode de reproduction

*C. canephora* appartient à la famille des rubiacées, genre *Coffea* L., sous-genre *Coffea* Bridson. Le genre *Coffea* a une aire de répartition limitée au continent africain, à Madagascar et aux îles Mascareignes. Il contiendrait près de 80 espèces, dont 25 seraient endémiques d'Afrique (BRIDSON et VERDCOURT, 1988). Les espèces du genre *Coffea* les plus proches génétiquement de *C. canephora* sont *C. congensis* et *C. brevipes* (LASHERMES *et al.*, 1997). De plus, *C. canephora* ou une forme ancestrale de *C. canephora* est l'une des deux espèces diploïdes parentales de *C. arabica* (LASHERMES *et al.*, 1997). Toutes les espèces du genre *Coffea* sont diploïdes à l'exception de *C. arabica*, qui est allotétraploïde. De même, elles possèdent toutes un mode de reproduction strictement allogame à l'exception de *C. arabica*, seule espèce autogame. Les études menées sur *C. canephora* ont mis en évidence un système d'auto-incompatibilité gamétophytique (BERTHAUD, 1980).

Pour les espèces diploïdes du genre *Coffea*, la quantité d'ADN par génome diploïde, mesurée par cytométrie en flux, varie de 0,95 à 1,78 picogramme (CROS *et al.*, 1995). Elle est de 1,54 picogramme pour *C. canephora* et de 2,61 picogrammes pour *C. arabica*.

*C. canephora* possède l'une des aires de répartition les plus larges du sous-genre *Coffea* : elle s'étend, d'ouest en est, de la Guinée au Soudan et, du nord au sud, du Cameroun à l'Angola (BERTHAUD, 1986).

La croissance des caféiers de cette espèce est dimorphe. Les tiges principales (axes orthotropes) ont une croissance verticale et les branches (axes plagiotropes), une croissance horizontale. Le bouturage horticole est relativement

aisé. L'espèce peut fleurir une ou plusieurs fois par an, après une pluie d'au moins 10 millimètres qui suit une période de stress hydrique. La maturation des baies dure de huit à douze mois selon la variété et l'environnement.

Les graines de *C. canephora* n'ont pas un comportement orthodoxe (ROBERTS, 1973) lorsqu'elles sont déshydratées ou stockées à basse température (COUTURON, 1980). Leur longévité n'est que de un à deux ans à l'état hydraté et à température ambiante.

## Les ressources génétiques

Etant donné le comportement des graines de *C. canephora*, la conservation à long terme des ressources génétiques de cette espèce s'effectue en champs.

Des collections plus ou moins représentatives des introductions les plus répandues, du matériel prélevé en plantations et des formes locales sont conservées en Côte d'Ivoire, au Cameroun, en Ouganda, en Inde, en Indonésie et au Brésil. Mais l'unique collection de référence pour les formes sauvages de *C. canephora* est celle de Divo, en Côte d'Ivoire. Elle contient plus de 700 génotypes sauvages collectés par l'Orstom (Institut français de recherche scientifique pour le développement en coopération) en collaboration avec le Cirad, l'IBPGR (International Board for Plant Genetic Resources), la FAO (Food and Agriculture Organization of the United Nations) et le MNHN (Muséum national d'histoire naturelle) entre 1975 et 1987, dans cinq pays africains : la Côte d'Ivoire et la Guinée, en Afrique de l'Ouest ; le Cameroun, le Congo et la République centrafricaine, en Afrique centrale. La gestion de cette collection repose sur la duplication clonale en champs de chaque génotype. Le remplacement d'un arbre mort est réalisé par bouturage horticole à partir de l'autre représentant du même génotype. Parallèlement, le Cirad a constitué une importante collection de matériel cultivé, également conservée sur la station expérimentale de Divo. Cette collection de travail contient plus de 600 accessions d'origines diverses : variétés et populations locales, prélèvements dans des plantations villageoises, matériel sélectionné.

## La structure de la diversité génétique

### La variabilité isoenzymatique

La première analyse de la diversité génétique de l'espèce *C. canephora* à partir du polymorphisme enzymatique a été réalisée par BERTHAUD (1986). Quinze échantillons ont été classés à l'aide des distances génétiques calculées à partir des fréquences alléliques de chaque échantillon. Douze des quinze



échantillons correspondaient à des populations sylvestres : neuf populations prospectées en Côte d'Ivoire et trois en République centrafricaine. Pour les trois autres échantillons, il a été nécessaire de regrouper des individus de populations ou d'origines différentes. C'est ainsi qu'un des trois échantillons était constitué par l'ensemble du matériel prospecté au Cameroun, qu'un autre regroupait tous les génotypes cultivés de la collection de travail du Cirad impliqués dans les essais agronomiques et que le dernier rassemblait des caféiers cultivés de type Ebobo, qui seraient originaires de Côte d'Ivoire (il n'existe plus de représentants du type Ebobo aujourd'hui en collection). Cette étude a mis en évidence, pour la première fois, une structuration génétique dans l'espèce *C. canephora*. Deux groupes ont été identifiés : le groupe Guinéen, composé de populations sauvages de Côte d'Ivoire, et le groupe Congolais, qui comprend le matériel sauvage de République centrafricaine et du Cameroun et le matériel cultivé. Par la suite, en augmentant le nombre de génotypes analysés et en classant la collection de matériel cultivé en 11 échantillons, MONTAGNON *et al.* (1992) ont identifié deux sous-groupes au sein du groupe Congolais : SG1 et SG2.

Dans notre étude, nous avons tenu compte des individus — 60 sauvages et 50 cultivés — et non pas des « populations » (tableaux 1 et 2). Au total, 29 allèles ont été détectés pour les 8 locus polymorphes, avec de 2 à 6 allèles par locus et un nombre moyen d'allèles par locus de 3,6. Il n'y a pas de différence significative entre les individus sauvages et les individus cultivés pour le nombre moyen d'allèles par locus.

La classification des 60 génotypes sauvages indique une structuration en deux groupes (figure 1a). Le groupe 1 ne contient que des individus collectés en Afrique de l'Ouest (Côte d'Ivoire et Guinée). Le groupe 2 renferme tous les individus originaires d'Afrique centrale (Cameroun, Congo et République centrafricaine) et deux génotypes provenant de Côte d'Ivoire. Ces deux groupes correspondent, par leur composition, aux groupes Guinéen et Congolais de Berthaud (1986).

Tableau 1. Origine du matériel sauvage étudié : pays, année de collecte, nombre de populations sylvestres échantillonnées et nombre de génotypes analysés.

Pays	Année de prospection	Nombre de populations	Nombre de génotypes	Référence
Cameroun	1983	10	15	ANTHONY <i>al.</i> (1985)
Congo	1985	7	13	DE NAMUR <i>et al.</i> (1988)
Côte d'Ivoire	1975-1986	21	36	BERTHAUD (1983) ; LE PIERRES <i>et al.</i> (1989)
Guinée	1987	1	2	LE PIERRES <i>et al.</i> (1989)
République centrafricaine	1975	6	11	BERTHAUD et GUILLAUMET (1978)
Total		45	77	

Tableau 2. Origine du matériel cultivé étudié : type d'introduction, dénomination en collection, institut donateur ou référence de la collecte, pays (d'origine pour les donations, de culture pour les prélèvements en plantations) et nombre de génotypes analysés (N).

Type d'introduction	Dénomination	Donateur ou collecteur	Pays	N
Donation	Aboisso	Aboisso <sup>1</sup> , Côte d'Ivoire	Gabon	6
	Niaouli	Bingerville <sup>2</sup> , Côte d'Ivoire	Togo	3
	Kouilou de Madagascar	Bingerville <sup>3</sup> , Côte d'Ivoire	Gabon	4
	C10 Man	—	Rép. du Congo	2
	INEAC	INEAC <sup>4</sup> , Rép. du Congo	Rép. du Congo	12
Prélèvement dans des plantations	Côte d'Ivoire	BERTHAUD (1983)	Côte d'Ivoire	7
	Guinée	LE PIERRES <i>et al.</i> (1989)	Guinée	9
	Togo	LE PIERRES <i>et al.</i> (1989)	Togo	2
	Hybrides		Côte d'Ivoire	6
	Inconnu	Robusta A1	Inconnu	4
<b>Total</b>				<b>55</b>

1 Introduction à Aboisso (Côte d'Ivoire), par Beynis en 1910, de matériel cultivé au Gabon (CORDIER, 1961).

2 Introduction au Jardin d'essais de Bingerville (Côte d'Ivoire), en 1914, de matériel cultivé au Togo (CORDIER, 1961).

3 Introduction à Bingerville (Côte d'Ivoire), en 1951, de matériel sélectionné à Madagascar et originaire du Gabon (CORDIER, 1961).

4 Introduction en Côte d'Ivoire, en 1935, de matériel sélectionné à l'INEAC au Congo belge (CORDIER, 1961).

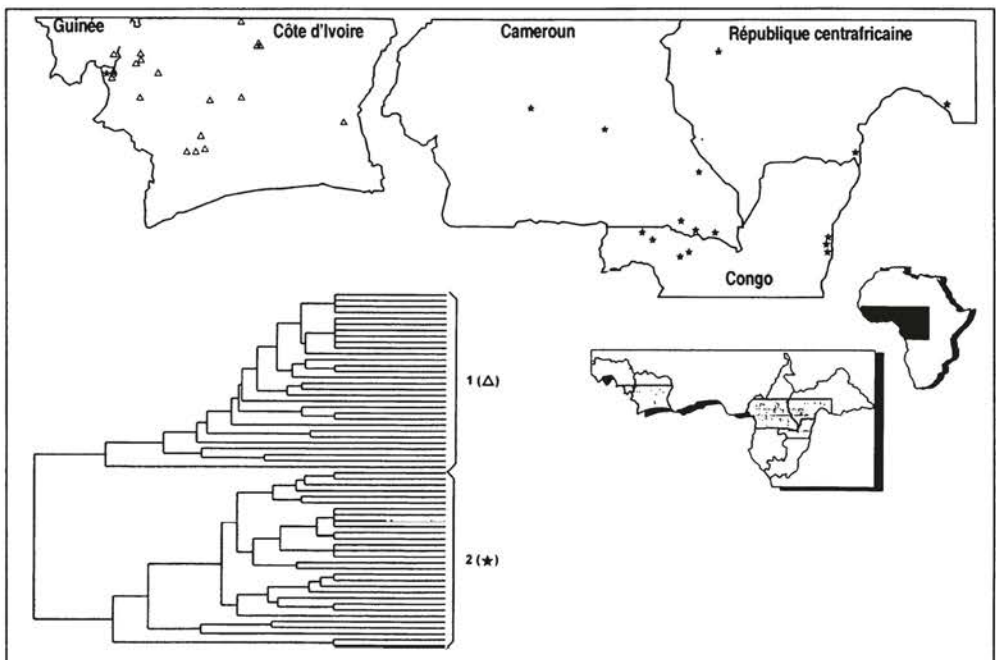


Figure 1a. Classification de 60 génotypes sauvages suivant l'indice de similarité de Dice et la méthode d'agrégation UPGMA à partir des données observées pour 29 marqueurs isoenzymatiques.

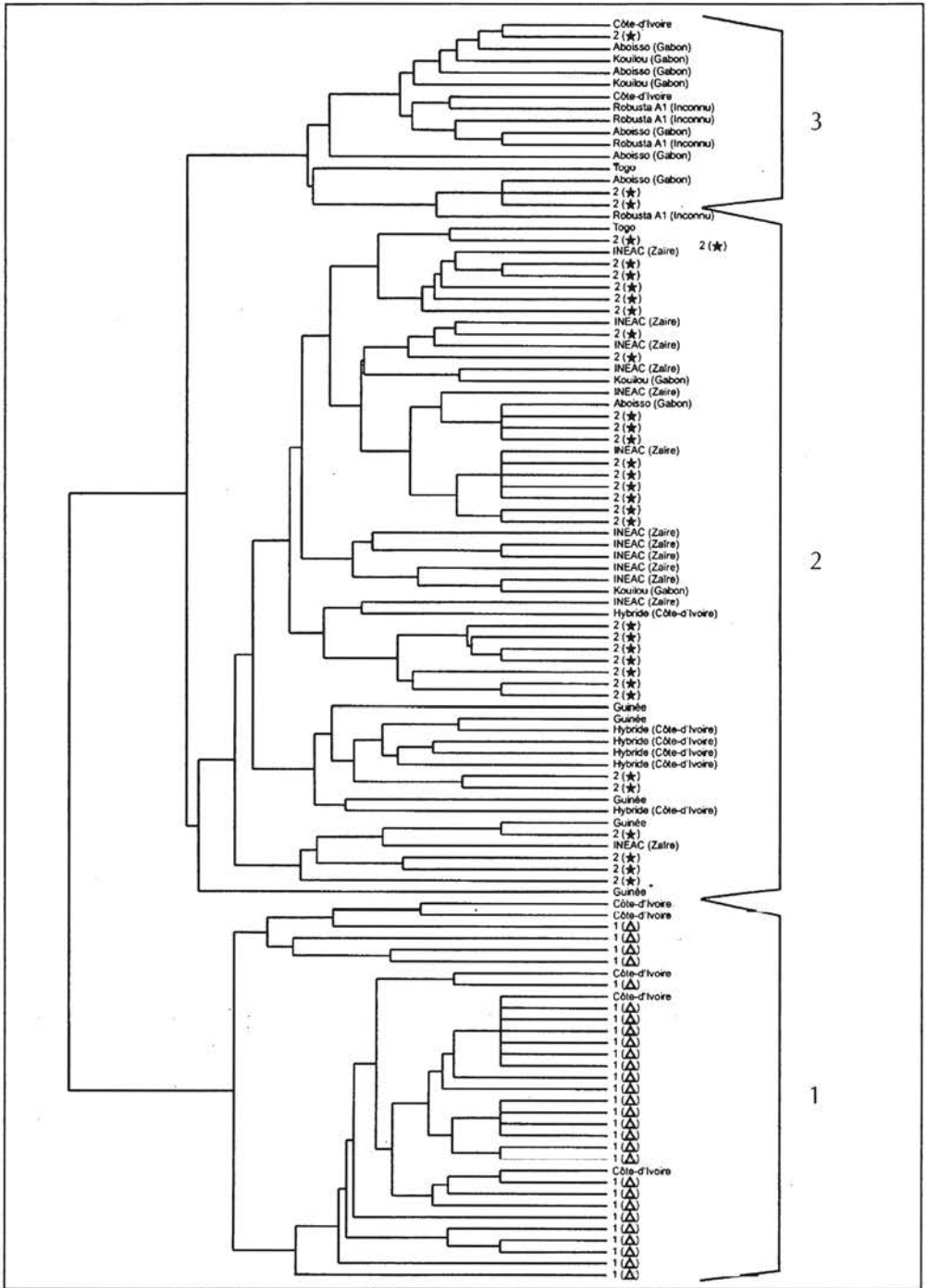


Figure 1b. Classification de 110 génotypes sauvages et cultivés suivant l'indice de similarité de Dice et la méthode d'agrégation UPGMA à partir des données observées pour 29 marqueurs isoenzymatiques.

Le dendrogramme obtenu avec l'ensemble des génotypes étudiés, sauvages et cultivés, présente une structure en trois groupes (figure 1b). Aucune séparation entre formes sauvages et formes cultivées n'apparaît. Le groupe le plus distant des deux autres (groupe 1) contient tous les individus du groupe sauvage 1 (Afrique de l'Ouest) et des génotypes prélevés dans des plantations en Côte d'Ivoire. Le deuxième groupe (groupe 2) comprend les génotypes du groupe sauvage 2 (à l'exception de trois génotypes), tout le matériel cultivé originaire de la République du Congo, les individus du groupe « hybrides » et ceux qui ont été prélevés dans des plantations en Guinée. Le matériel cultivé originaire du Gabon (sauf trois génotypes), les individus du groupe Robusta A1, deux génotypes prélevés dans des plantations en Côte d'Ivoire, un génotype prélevé au Togo et trois individus du groupe sauvage 2 forment le troisième groupe (groupe 3). Au regard de l'origine du matériel qui les compose, notre groupe 2 correspond au sous-groupe SG2 et notre groupe 3, au sous-groupe SG1 de MONTAGNON *et al.* (1992).

## La variabilité moléculaire

L'utilisation des marqueurs moléculaires pour l'étude de la diversité génétique du genre *Coffea* est récente. Les premiers travaux ont porté sur l'analyse de la diversité de *C. arabica* à l'aide de marqueurs RAPD (LASHERMES *et al.*, 1996). Les résultats que nous présentons constituent les premières données sur la variabilité moléculaire de *C. canephora*.

Sur 26 sondes homologues testées, 10 se sont révélées monolocus et polymorphes. Elles ont permis de révéler de 2 à 14 allèles par locus soit 66 allèles au total et un nombre moyen d'allèles par locus polymorphe de 6,6. Le nombre total d'allèles observés pour les génotypes sauvages (62) est significativement supérieur ( $\chi^2 = 4,55$  ;  $P = 0,0329$ ) à celui des génotypes cultivés (54).

Le dendrogramme obtenu à partir des données moléculaires met en évidence une structuration du matériel sauvage en cinq groupes (figure 2a). Les génotypes d'une population du nord-ouest du Congo et d'une population du sud-est du Cameroun constituent le groupe A. Le groupe B comprend tous les génotypes collectés le long de la frontière sud de la République centrafricaine. Les individus du groupe C sont répartis dans les trois pays d'Afrique centrale : nord-ouest du Congo, sud-est du Cameroun et sud-ouest de la République centrafricaine. Le groupe D est constitué par tous les génotypes collectés en Guinée et en Côte d'Ivoire, à l'exception de quatre individus de l'ouest de la Côte d'Ivoire. Le groupe E contient les génotypes collectés dans le nord-est du Congo, ceux qui appartiennent à des populations du nord-ouest du Congo et du sud du Cameroun et des individus provenant de trois populations de l'ouest de la Côte d'Ivoire.

Une analyse globale du dendrogramme indique que le matériel sauvage originaire d'Afrique de l'Ouest est rassemblé dans un seul groupe, alors que le matériel collecté en Afrique centrale se structure en quatre groupes. Le groupe E est le plus distant des quatre autres groupes. Le groupe d'Afrique centrale (groupe C) le plus proche de celui d'Afrique de l'Ouest (groupe D) possède la répartition géographique la plus large.

Lorsque le matériel cultivé est pris en compte dans l'analyse, la structure de l'espèce en cinq groupes est conservée (figure 2b). De même, la position des groupes les uns par rapport aux autres demeure inchangée. Pour chacun des groupes, la composition en matériel sauvage reste identique à celle qui a été définie précédemment. Cependant, aucun matériel cultivé n'est présent dans les groupes sauvages B et C. Les individus provenant de la République du Congo, à l'exception d'un individu, s'insèrent dans le groupe E, auquel se rattachent également des individus collectés dans des plantations de Guinée et un hybride prélevé dans des plantations de Côte d'Ivoire. Aux individus sauvages du groupe A s'ajoutent la plupart des génotypes cultivés originaires du Gabon, ceux du Togo (prélèvements dans des plantations et groupe Niaouli) et les individus du groupe Robusta A1. Enfin, la quasi-totalité des individus hybrides, les génotypes prélevés dans des plantations de Côte d'Ivoire et la plupart des génotypes provenant des plantations de Guinée sont inclus dans le groupe D.

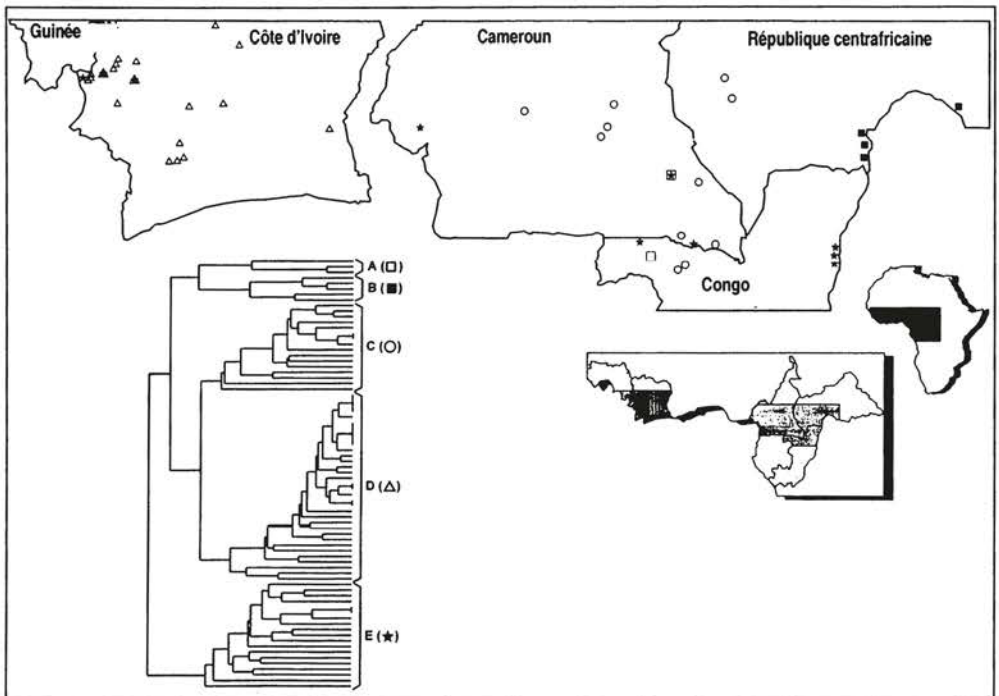


Figure 2a. Classification de 77 génotypes sauvages suivant l'indice de similarité de Dice et la méthode d'agrégation UPGMA à partir des données observées pour 66 marqueurs RFLP.

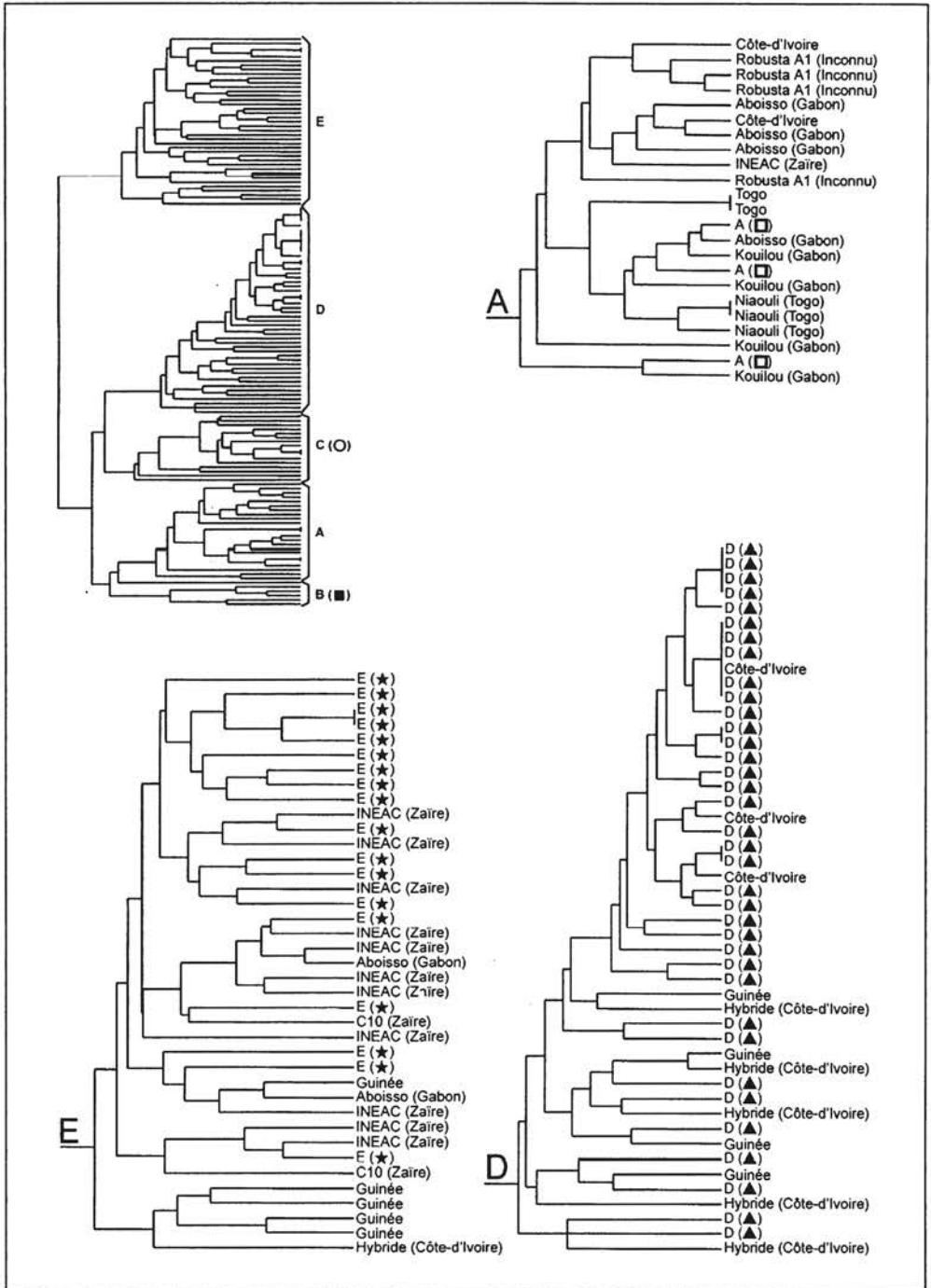


Figure 2b. Classification de 132 génotypes sauvages et cultivés suivant l'indice de similarité de Dice et la méthode d'agrégation UPGMA à partir des données observées pour 66 marqueurs RFLP.



## La variabilité agromorphologique

A notre connaissance, seules deux études ont porté sur l'analyse de la diversité agromorphologique de *C. canephora* (MONTAGNON *et al.*, 1992 ; LEROY *et al.*, 1993). Dans les deux cas, les analyses en composantes principales réalisées ne permettent pas de révéler une forte structuration de l'espèce. En revanche, des comparaisons *a posteriori* entre les groupes établis sur la base des isoenzymes ont été effectuées dans plusieurs études et pour divers jeux de variables agromorphologiques (BERTHAUD, 1986 ; MONTAGNON *et al.*, 1992 ; LEROY *et al.*, 1993 ; MONTAGNON et LEROY, 1993 ; MONTAGNON *et al.*, 1993 ; MOSCHETTO *et al.*, 1996). Des différences significatives entre les moyennes des groupes isoenzymatiques ont ainsi pu être mises en évidence pour certains caractères : morphologie foliaire, longueur des entre-nœuds, ramification, sensibilité à la sécheresse, phénologie de la fructification, sensibilité à la rouille orangée due à *Hemileia vastatrix*. En revanche, il n'est pas possible de déterminer, en plantation, l'appartenance d'un caféier à un groupe génétique d'après sa morphologie.

Dans notre étude, l'analyse des données agromorphologiques conduit à une classification du matériel sauvage en deux groupes majeurs auxquels s'ajoute un troisième groupe comportant deux individus relativement proches l'un de l'autre mais très distants des autres génotypes sauvages (figure 3a). Cette

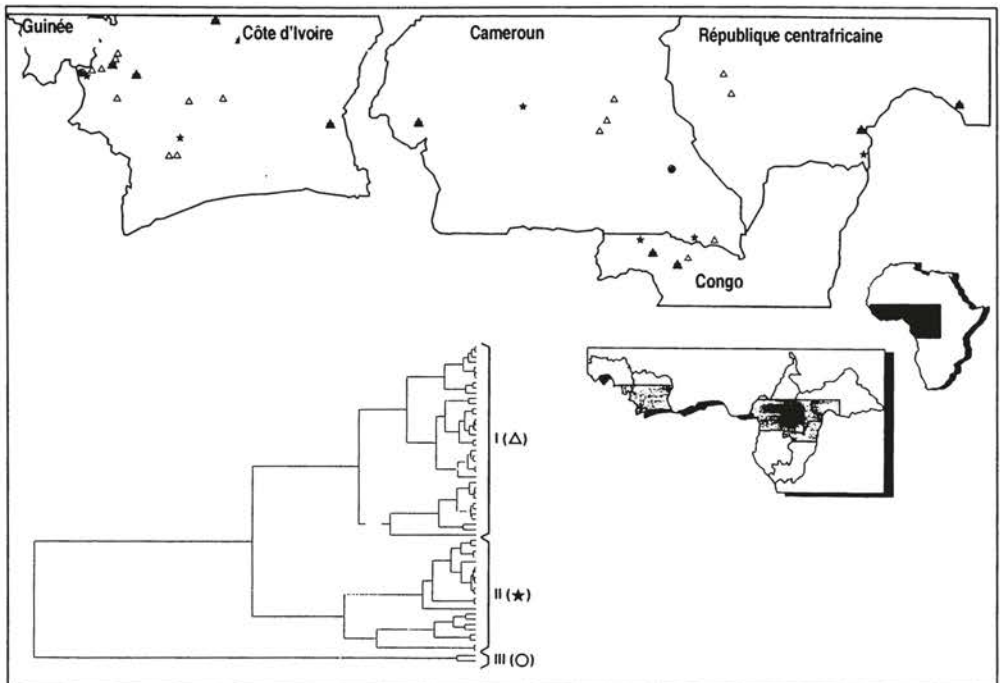


Figure 3a. Classification de 61 génotypes sauvages suivant la distance euclidienne et la méthode d'agrégation UPGMA à partir des données observées pour 11 marqueurs agromorphologiques.

structuration n'est pas d'ordre géographique : tous les pays prospectés comportent des représentants de chacun des groupes I et II. De plus, pour environ un tiers des populations, les individus provenant d'une même population se répartissent entre les deux groupes majeurs I et II. Des analyses de variance réalisées pour chacun des marqueurs agromorphologiques étudiés révèlent une différence significative entre les moyennes des groupes I et II pour la longueur du cycle de fructification, la granulométrie et la morphologie foliaire.

Lorsque le matériel cultivé est pris en compte, la structure de la diversité est globalement semblable à celle qui est obtenue lorsque seuls les génotypes sauvages sont analysés (figure 3b). La répartition des individus sauvages dans les groupes demeure inchangée à l'exception de deux génotypes du groupe sauvage II qui s'insèrent dans le groupe III. Aucun individu cultivé ne se retrouve dans le groupe III. Tous les génotypes originaires du Gabon, à l'exception d'un génotype du groupe Aboisso, ainsi que les individus des groupes Robusta A1 et hybrides s'intègrent dans le groupe I. Les individus de la République du Congo, excepté un génotype, et les génotypes prélevés dans des plantations en Côte d'Ivoire se retrouvent dans le groupe II.

## Les relations entre les différents niveaux de diversité

La distribution des individus des groupes biochimiques au sein des groupes moléculaires et des groupes agromorphologiques (tableau 3) permet de comparer les structurations observées. Les groupes biochimiques présentent des diversités différentes. Le groupe biochimique 1, relativement homogène du point de vue moléculaire, correspond uniquement au groupe moléculaire D. Le groupe biochimique 2 présente une large diversité et comporte des

Tableau 3. Distributions des génotypes sauvages, d'une part, des génotypes sauvages et cultivés, d'autre part, classés suivant leur appartenance aux groupes biochimiques au sein des groupes moléculaires et des groupes agromorphologiques. Pour chacune des comparaisons, seuls les génotypes communs aux analyses correspondant aux deux marqueurs comparés ont été pris en compte.

	Groupes biochimiques	Groupes moléculaires					Groupes agromorphologiques		
		A	B	C	D	E	I	II	III
Génotypes sauvages	1	0	0	0	<b>30</b>	0	14	7	0
	2	<b>2</b>	<b>4</b>	<b>12</b>	1	<b>12</b>	7	7	1
Génotypes sauvages et cultivés	1	0	0	0	<b>32</b>	1	17	9	0
	2	6	<b>4</b>	<b>12</b>	8	<b>29</b>	16	20	2
	3	<b>13</b>	0	0	0	1	8	1	0



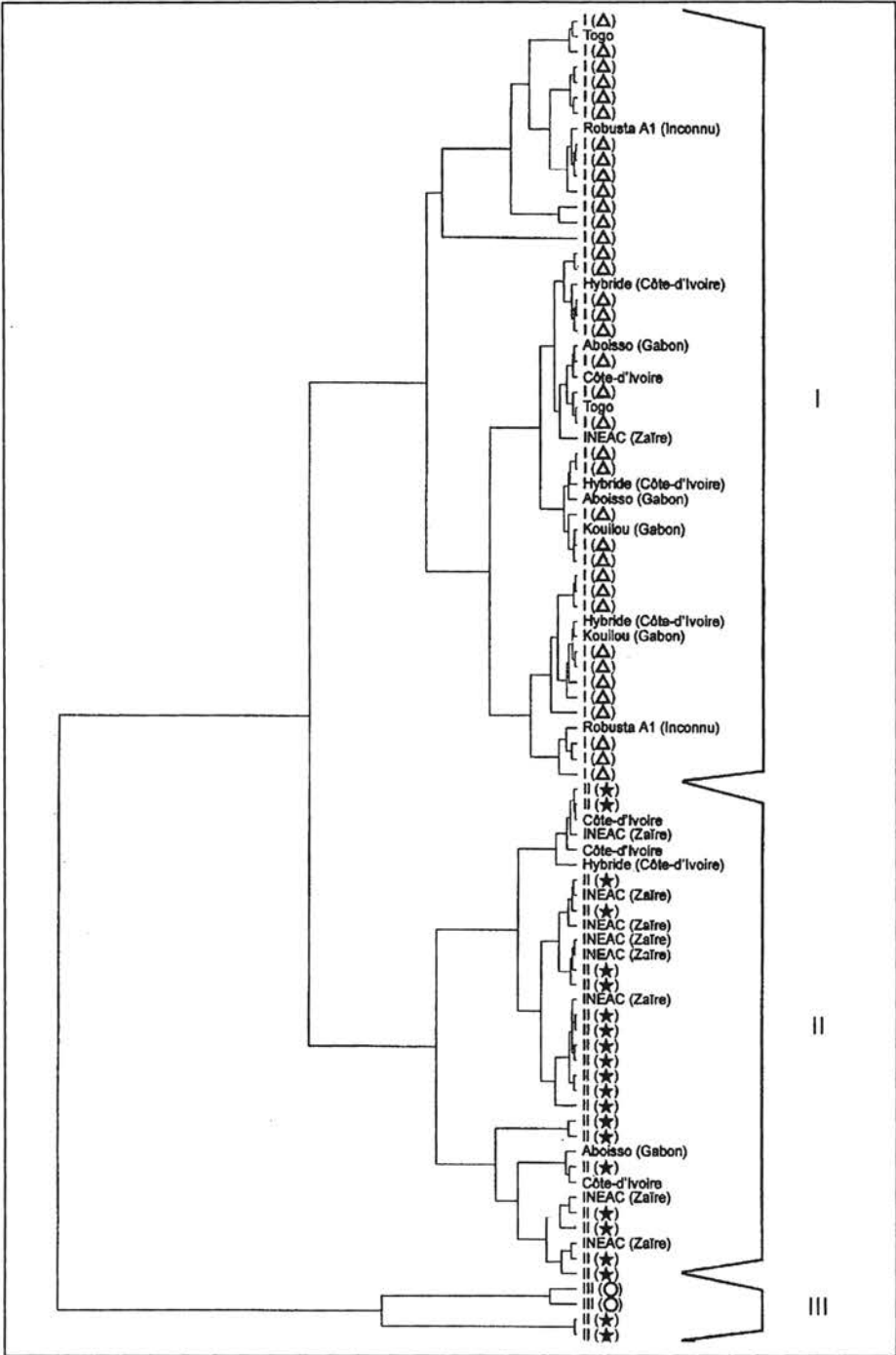


Figure 3b. Classification de 87 génotypes sauvages et cultivés suivant la distance euclidienne et la méthode d'agrégation UPGMA à partir des données observées pour 11 marqueurs agromorphologiques.

représentants des cinq groupes moléculaires. Lorsque seuls les génotypes sauvages sont pris en compte, une concordance quasi parfaite (un individu fait exception) entre le groupe biochimique 1 et le groupe moléculaire D est observée ; les individus du groupe biochimique 2 se répartissant dans les groupes moléculaires A, B, C et E. Lorsque les génotypes sauvages et cultivés sont considérés simultanément, les concordances mises en évidence avec les génotypes sauvages seuls ne sont pas modifiées dans leurs grandes lignes. La variabilité biochimique des génotypes cultivés est plus large que celle des génotypes sauvages. Le groupe biochimique 3, propre aux génotypes cultivés, correspond essentiellement au groupe moléculaire A.

Les marqueurs agromorphologiques mettent en évidence une structuration relativement forte de la diversité de *C. canephora* en deux groupes majeurs. Cependant, aucune concordance entre les deux groupes biochimiques et les trois groupes agromorphologiques ne peut être établie. De même, la structure agromorphologique ne coïncide pas avec celle qui est issue des marqueurs de type RFLP. Le fait que chaque groupe morphologique inclut des représentants des différents groupes biochimiques et moléculaires rend compte de l'impossibilité de distinguer les groupes Guinéen et Congolais d'après leurs caractéristiques morphologiques. Des pressions de sélection similaires se sont probablement exercées en Afrique de l'Ouest et en Afrique centrale, qui ont abouti à des formes morphologiquement indifférenciées.

La forte concordance des structures observées avec les deux types de marqueur neutre utilisés, isoenzymes et RFLP, conforte la classification au sein de *C. canephora* (tableau 4). De plus, les résultats de notre étude prouvent l'intérêt

Tableau 4. Concordance des structures de la diversité sauvage et cultivée observées à l'aide de marqueurs isoenzymatiques et RFLP au cours de trois études successives et composition en matériel sauvage, prélevé dans des plantations et sélectionné, des groupes observés.

BERTHAUD (1986)		MONTAGNON et al. (1992)		Notre étude		
Marqueurs				Matériel		
Isoenzymes	Isoenzymes	Isoenzymes	RFLP	Sauvage	Plantation	Sélectionné
Guinéens	Guinéens	1	D	Côte d'Ivoire Guinée	Côte d'Ivoire Guinée	Hybride (Côte d'Ivoire)
			B	Centrafrique E		
			C	Centrafrique O		
	SG2	2	C	Cameroun Congo NO		
Congolais			E	Congo Cameroun S	Guinée	INEAC (Rép. du Congo) C10 (Rép. du Congo)
	SG1	3	A	Congo NO Cameroun SE	Togo Côte d'Ivoire	Kouilou (Gabon) Aboisso (Gabon) Niaouli (Togo) Robusta A1 (inconnu)

des marqueurs moléculaires pour la structuration de *C. canephora*. D'une part, les marqueurs RFLP permettent, pour un nombre de locus équivalent, de révéler un nombre moyen d'allèles par locus polymorphe plus important que les marqueurs biochimiques (6,6 contre 3,6). D'autre part, les marqueurs RFLP permettent d'affiner la structuration en mettant en évidence une structuration intra-groupe : le groupe Congolais défini par BERTHAUD (1986) correspond à quatre groupes moléculaires.

## La valorisation et la gestion des ressources génétiques

### La structuration de la diversité et la valorisation des ressources génétiques

A l'aide des locus isoenzymatiques discriminant les groupes Guinéen et Congolais, BERTHAUD (1986) a identifié des hybrides intergroupes dans le matériel cultivé. Parmi les 12 hybrides intergroupes repérés, 4 correspondent aux 6 meilleurs clones vulgarisés. Cette observation a amené J. Berthaud à proposer un schéma de sélection récurrente réciproque de *C. canephora* fondé sur l'utilisation de ces deux groupes. L'efficacité de ce schéma a été démontrée par la suite (LEROY *et al.*, 1993), ce qui confirme qu'il est important de connaître la structure de la diversité pour exploiter au mieux les ressources génétiques.

Les résultats de l'étude sur les marqueurs moléculaires devraient permettre d'améliorer ce schéma de sélection. En effet, d'une part, il faudrait s'assurer qu'il n'existe pas de combinaisons à l'intérieur du groupe Congolais présentant une hétérosis plus forte que l'hétérosis moyenne entre les groupes Guinéen et Congolais. D'autre part, si les hybrides entre Guinéen et Congolais restent les plus intéressants, des résultats préliminaires semblent indiquer que la valeur de l'hétérosis entre les groupes Guinéen et Congolais dépend du groupe moléculaire auquel appartient l'individu du groupe Congolais. Ainsi, en se fondant uniquement sur les distances génétiques établies à l'aide des marqueurs RFLP, l'hétérosis entre les individus classés dans les groupes D et C pourrait être moins forte que celle entre les individus classés dans les groupes E et C.

### La structuration de la diversité et la gestion des ressources génétiques

Les résultats de cette étude permettent d'avancer un certain nombre de recommandations pour assurer une meilleure gestion *ex situ* des ressources génétiques de *C. canephora*.

La plupart des collections de *C. canephora* rassemblent surtout du matériel cultivé et sont fortement redondantes. Or notre étude montre qu'une grande partie de la diversité incluse dans le matériel sauvage n'est pas représentée dans le matériel cultivé. De plus, le matériel cultivé ne recèle pas une diversité originale par rapport au matériel sauvage. Par conséquent, le matériel sauvage collecté au cours des nombreuses missions de prospection et conservé en Côte d'Ivoire constitue réellement la plus grande source de variabilité disponible pour cette espèce. Des efforts devraient donc être consentis pour préserver cette collection que ce soit en réalisant la duplication du matériel végétal ou en envisageant d'autres modes de conservation.

En Côte d'Ivoire jusqu'à présent, les deux collections de caféiers cultivés et sauvages sont gérées de manière indépendante, ce qui multiplie le travail de gestion. Nos résultats indiquent qu'il est possible d'envisager la collection dans sa globalité et de la hiérarchiser sur la base des groupes moléculaires. Par ailleurs, l'utilisation des algorithmes de sélection qui maximisent la diversité intragroupe, tels que celui proposé par NOIROT *et al.* (1996), devrait permettre, en définissant une *core collection*, d'optimiser la gestion de la collection globale et d'établir des priorités tant pour la conservation que pour l'évaluation, l'utilisation et la diffusion des ressources en collection.

Actuellement, le matériel végétal est uniquement maintenu sous forme de collection de génotypes au champ. La création de *core collections* stratifiées et de petite taille pourrait permettre de conserver des gènes plutôt que des génotypes en constituant rationnellement des mélanges de graines au sein de chacun des groupes génétiques mis en évidence dans notre étude. La conservation de plantules issues de mélanges de graines en vitrothèque a montré ses limites : des groupes génétiques sont rapidement perdus, notamment pour *C. canephora* (DUSSERT *et al.*, 1997). En revanche, la cryoconservation de graines, déjà utilisée pour *C. arabica* (DUSSERT *et al.*, 1998), constituerait une solution intéressante par rapport à la conservation au champ.

## Conclusion

Notre étude a permis de montrer que les marqueurs moléculaires de type RFLP permettent de structurer la diversité des caféiers *C. canephora*. Cette structuration est largement en accord avec celle qui a été obtenue précédemment, mais aussi avec la structuration issue de notre étude par les marqueurs biochimiques. Cependant, les marqueurs moléculaires révèlent une différenciation plus importante que les autres marqueurs.

Bien que notre travail ait porté sur un échantillon réduit des collections conservées en Côte d'Ivoire, il est intéressant de constater que le matériel cultivé ne se différencie généralement pas du matériel sauvage et qu'une partie

de la diversité de ce matériel sauvage n'a pas encore été exploitée pour la culture de *C. canephora*. De plus, il existe une différenciation à l'intérieur du groupe Congolais, le matériel originaire d'Afrique centrale se structurant en plusieurs groupes moléculaires. D'autre part, le matériel originaire d'Afrique de l'Ouest, du groupe Guinéen, classé dans un seul groupe moléculaire, n'est pas plus distant des groupes d'Afrique centrale que ces derniers ne le sont entre eux.

Nos résultats permettent d'envisager de nouvelles stratégies aussi bien pour la sélection variétale que pour une gestion rationnelle des ressources génétiques des caféiers *C. canephora*.

# Annexe

## Matériel végétal

Un échantillonnage de 132 génotypes a été réalisé au sein des collections de matériel sauvage et cultivé conservées en Côte d'Ivoire (tableaux 1 et 2). Les 77 génotypes sauvages ont été échantillonnés afin d'avoir une représentation de chacune des 45 populations sylvestres prospectées (tableau 1). Pour le matériel cultivé un échantillonnage aléatoire proportionnel a été réalisé pour chacune des dix origines principales identifiées en collection (tableau 2). La nature des regroupements est très hétérogène. Pour le matériel ayant fait l'objet d'une donation, la dénomination des groupes correspond au nom, à la localisation de la station expérimentale donatrice ou au type variétal du matériel (Robusta ou Kouilou). Pour ces groupes, le pays d'origine peut correspondre à la zone de culture ou au centre de sélection. Pour le matériel collecté dans les plantations, le pays mentionné est celui dans lequel les prélèvements ont été effectués. Le groupe dénommé « hybrides » rassemble des génotypes pour lesquels on a montré *a posteriori* qu'il s'agissait d'hybrides entre des formes originaires d'Afrique de l'Ouest et des formes provenant d'Afrique centrale (BERTHAUD, 1983). Enfin, l'historique de l'introduction du groupe Robusta A1 n'a pu être retracée. L'origine de ce groupe reste donc inconnue.

## Analyse par RFLP

L'extraction de l'ADN génomique total a été réalisée selon la méthode décrite par AGWANDA *et al.* (1997). La technique de marquage moléculaire utilisée est celle qui a été décrite par LASHERMES *et al.* (1995). Deux enzymes de restriction ont été utilisées : *EcoRI* et *HindIII*. Les 26 sondes testées proviennent d'une banque génomique de *C. arabica*. Parmi celles-ci, dix sondes ont été retenues pour leur caractère polymorphe et monocus. Chaque sonde a été utilisée après digestion par l'une ou l'autre des enzymes de restriction. La présence et l'absence des 66 bandes correspondant à 66 allèles ont été codées 1 et 0, respectivement.

## Analyse enzymatique

Au sein de l'échantillon total de 132 individus (tableaux 1 et 2), l'analyse du polymorphisme isoenzymatique a été réalisée sur 60 individus sauvages et 50 individus cultivés. Parmi les 60 individus sauvages, 48 sont communs aux analyses réalisées sur les marqueurs agromorphologiques. Pour le matériel cultivé, le nombre d'individus communs aux analyses isoenzymatiques et agromorphologiques est de 26. Les techniques d'extraction, d'électrophorèse et de révélation des isoenzymes sont celles de BERTHAUD (1986). Les analyses ont porté sur cinq systèmes enzymatiques révélant huit locus : estérases  $\alpha$  et  $\beta$  (3 locus), 6-phosphogluconate déshydrogénase (2 locus), isocitrate déshydrogénase (1 locus), phosphoglucomutase (1 locus) et phosphoglucoisomérase (1 locus). Les 29 allèles identifiés ont été codés en présence ou en absence (1 ou 0).

## Etude agromorphologique

Au sein des 132 génotypes étudiés (tableaux 1 et 2), 61 génotypes sauvages et 26 génotypes cultivés ont été évalués à partir de 11 marqueurs agromorphologiques. Quatre

classes de marqueurs peuvent être distinguées : les marqueurs morphologiques (longueur, largeur, surface et forme des feuilles, longueur de l'acumen, longueur du pétiole) ; les marqueurs technologiques (poids de 100 grains, taux de grains caracolis, rendement en café marchand, taux de loges vides, pour le matériel sauvage uniquement) ; les marqueurs phénologiques (pour les génotypes sauvages uniquement, longueur du cycle de fructification et étalement de la maturation des cerises) ; les marqueurs agronomiques (production annuelle de cerises). Pour une description détaillée des marqueurs, on peut se reporter à ANTHONY (1992).

### Analyses statistiques de classification

Pour les marqueurs RFLP et isoenzymatiques, les matrices de distance entre individus ont été calculées en utilisant l'indice de similarité de DICE (1945). La distance euclidienne a été utilisée pour les marqueurs agromorphologiques. Pour les trois types de marqueurs, la méthode d'agrégation utilisée pour construire les dendrogrammes est la méthode UPGMA.

## Références bibliographiques

- AGWANDA O.A., LASHERMES P., TROUSLOT P., COMBES M.C., CHARRIER A., 1997. Identification of RAPD markers for resistance to coffee berry disease, *Colletotrichum kahawae*, in arabica coffee. *Euphytica*, 97 : 241-248.
- ANTHONY A., 1992. Les ressources génétiques des caféiers : collecte, gestion d'un conservatoire et évaluation de la diversité génétique. Montpellier, France, Orstom, collection Travaux et documents, 320 p.
- ANTHONY F., COUTURON E., DE NAMUR C., 1985. Les caféiers sauvages du Cameroun : résultats d'une mission de prospection effectuée par l'Orstom en 1983. In : XI<sup>e</sup> colloque scientifique international sur le café. Paris, France, ASIC, p. 495-501.
- BERTHAUD J., 1980. L'incompatibilité chez *Coffea canephora* : méthode de test et déterminisme génétique. *Café, cacao, thé*, 24 : 267-274.
- BERTHAUD J., 1983. Liste du matériel provenant des prospections de Côte d'Ivoire. Paris, France, Orstom (document interne).
- BERTHAUD J., 1986. Les ressources génétiques pour l'amélioration des caféiers africains diploïdes. Montpellier, France, Orstom, collection Travaux et documents, 379 p.
- BERTHAUD J., GUILLAUMET J.L., 1978. Les caféiers sauvages en Centrafrique : résultats d'une mission de prospection (janvier-février 1975). *Café, cacao, thé*, 3 : 171-186.
- BRIDSON D.M., VERDCOURT B., 1988. Rubiaceae (Part 2). In : Flora of tropical East Africa, R.M. Polhill éd., Rotterdam, Pays-Bas, Balkema, 727 p.
- CHARRIER A., ESKES B., 1997. Les caféiers. In : L'amélioration des plantes tropicales, A. Charrier et al. éd., Montpellier, France, Cirad-Orstom, collection Repères, p. 171-196.
- CORDIER L., 1961. Les objectifs de la sélection caféière en Côte d'Ivoire. *Café, cacao, thé*, 5 : 147-159.
- COUTURON E., 1980. Le maintien de la viabilité des graines de caféiers par le contrôle de leur teneur en eau et de la température de stockage. *Café, cacao, thé*, 24 : 27-32.
- CROS J., COMBES M.C., CHABRILLANGE N., DUPERRAY C., MONNOT DES ANGES A., HAMON S., 1995. Nuclear DNA content in the subgenus *Coffea*: inter- and intra-specific variation in African species. *Canadian Journal of Botany*, 73 : 14-20.
- DICE L.R., 1945. Measures of the amount of ecologic association between species. *Ecology*, 26 : 297-302.
- DUSSERT S., CHABRILLANGE N., ANTHONY F., ENGELMANN F., RECALT C., HAMON S., 1997. Variability in storage response within a coffee (*Coffea* spp.) core collection under slow growth conditions. *Plant Cell Report*, 16 : 344-348.
- DUSSERT S., CHABRILLANGE N., ENGELMANN F., ANTHONY F., LOUARN J., HAMON S., 1998. Cryopreservation of seeds of four coffee species (*Coffea arabica*, *C. costatifructa*, *C. racemosa* and *C. sessiliflora*): importance of water content and cooling rate. *Seed Science Research*, 8 : 9-15.



LASHERMES P., COMBES M.C., CROS J., 1995. Use of non-radioactive digoxigenin-labelled DNA probes for RFLP analysis in coffee. *In* : Techniques et utilisations des marqueurs moléculaires, A. Berville et M. Tressac éd., Paris, France, Inra, p. 21-25.

LASHERMES P., COMBES M.C., TROUSLOT P., CHARRIER A., 1997. Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theoretical and Applied Genetics*, 94 : 947-955.

LASHERMES P., TROUSLOT P., ANTHONY F., COMBES M.C., CHARRIER A., 1996. Genetic diversity for RAPD markers between cultivated and wild accessions of *C. arabica*. *Euphytica*, 87 : 59-64.

LE PIERRES D., CHARMETANT P., YAPO A., LEROY T., COUTURON E., BONTEMS S., TEHE H., 1989. Les caféiers sauvages de Côte d'Ivoire et de Guinée : bilan des missions de prospection effectuées de 1984 à 1987. *In* : XIII<sup>e</sup> colloque scientifique international sur le café. Paris, France, ASIC, p. 420-428.

LEROY T., MONTAGNON C., CHARRIER A., ESKES A., 1993. Reciprocal recurrent selection applied to *Coffea canephora* Pierre. 1. Characterization and evaluation of breeding populations and value of intergroup hybrids. *Euphytica*, 67 : 113-125.

MONTAGNON C., LEROY T., 1993. Réaction à la sécheresse de jeunes caféiers *Coffea canephora* de Côte d'Ivoire appartenant à différents groupes génétiques. *Café, cacao, thé*, 37 : 179-190.

MONTAGNON C., LEROY T., CILAS C., ESKES A.B., 1993. Differences among clones of *Coffea canephora* in resistance to the scolytid coffee twig-borer. *International Journal of Pest Management*, 39 : 204-209.

MONTAGNON C., LEROY T., ESKES A.B., 1998. Amélioration variétale de *Coffea canephora*. 2. Les programmes de sélection et leurs résultats. *Plantations, recherche, développement*, 5 : 89-98.

MONTAGNON C., LEROY T., YAPO A., 1992. Etude complémentaire de la diversité génotypique et phénotypique des caféiers de l'espèce *C. canephora* en collection en Côte d'Ivoire. *In* : XIV<sup>e</sup> colloque scientifique international sur le café. Paris, France, ASIC, p. 444-450.

MOSCHETTO D., MONTAGNON C., GUYOT B., PERRIOT J.J., LEROY T., ESKES A.B., 1996. Studies on the effect of genotype on cup quality of *Coffea canephora*. *Tropical Science*, 36 : 18-31.

DE NAMUR C., COUTURON E., SITA P., ANTHONY F., 1988. Résultats d'une mission de prospection des caféiers sauvages du Congo. *In* : XII<sup>e</sup> colloque scientifique international sur le café. Paris, France, ASIC, p. 397-404.

NOIROT M., ANTHONY F., HAMON S., 1996. The principal component scoring: a new method of constituting a core collection using quantitative data. *Genetic Resources and Crop Evolution*, 43 : 1-6.

ROBERTS E.H., 1973. Predicting the storage life of seeds. *Seed Science and Technology*, 1 : 499-514.

# La canne à sucre

---

Jean Christophe Glaszmann, Nazeema Jannoo,  
Laurent Grivet, Angélique D'Hont

La canne à sucre est une production majeure des zones tropicales et subtropicales. C'est autour de sa culture que l'industrie sucrière est née et s'est développée. Actuellement, la production mondiale de saccharose dépasse 100 millions de tonnes par an, 70 % environ provenant de la canne et 30 %, de la betterave. Les plus gros producteurs sont l'Inde, le Brésil, Cuba et la Chine.

La canne cultivée a une morphologie classique de graminée pérenne. Elle est multipliée par bouturage de tige à partir du développement des bourgeons axillaires. L'accumulation du saccharose dans la tige survient à la fin de la période de végétation, après la floraison quand elle a lieu. Elle est déclenchée par une action combinée du froid, même relatif, et d'une baisse de l'alimentation hydrique (FAUCONNIER, 1991). Dans les climats équatoriaux, où il n'y a pas de saison sèche marquée, la richesse en sucre des tiges est souvent faible.

Le cycle entre deux récoltes varie entre 10 et 24 mois en fonction du climat et des choix économiques. Le cycle entre deux plantations est très variable et dépend surtout de critères socio-économiques. Il est, par exemple, confondu avec le cycle de récolte à Hawaii et peut s'étendre sur plus de dix repousses dans certaines zones défavorisées de la Réunion. La pousse issue directement des boutures s'appelle la « vierge ». Le rendement est généralement maximal à la première repousse et tend ensuite à décroître au fur et à mesure des récoltes.

## La taxonomie et les ressources génétiques

### La biologie, la taxonomie et la répartition géographique

La canne est une monocotylédone de la famille des poacées. L'inflorescence est une panicule lâche et ramifiée, où les fleurs (épillet) sont disposées par paires, l'une étant sessile, l'autre pédonculée. Chaque épillet est bisexué et possède trois étamines et un ovule. Toutes les tiges ne fleurissent pas forcément ; l'intensité de la floraison dépend de facteurs génétiques et climatiques. La pollinisation est anémophile comme pour la plupart des graminées. La floraison est généralement considérée comme défavorable à la production et donc contre-sélectionnée. Le fruit est un caryopse. Il n'est utilisé que pour les besoins de la sélection et ne sert jamais de semence dans les champs de production.

Le genre *Saccharum* appartient à la tribu des andropogonées, au même titre que deux céréales majeures, le maïs et le sorgho. Il trouve son origine en Asie (figure 1). Les taxonomistes distinguent cinq espèces de base.

*S. officinarum*, la première espèce cultivée, est probablement originaire de Papouasie-Nouvelle-Guinée. Les clones de cette espèce présentent des tiges à fort diamètre, très riches en sucre.

*S. barberi* est originaire d'Inde et *S. sinense*, de Chine. Les clones de ces deux espèces sont généralement plus rustiques que ceux de *S. officinarum*. Ils ont des tiges plus fines, plus fibreuses et moins riches en sucre. On pense qu'ils proviennent d'hybridations spontanées entre *S. officinarum* et l'espèce sauvage *S. spontaneum*, qui se seraient produites en Inde pour *S. barberi* et en Chine pour *S. sinense*.

*S. spontaneum* est une espèce sauvage à distribution géographique très vaste, qui englobe presque toute l'Asie, de l'Afghanistan aux îles du Pacifique. Les différents écotypes peuvent être annuels ou pérennes. Ils présentent une très grande variabilité morphologique.

*S. robustum*, autre espèce sauvage, est probablement l'ancêtre de *S. officinarum*. Elle se rencontre essentiellement en Papouasie-Nouvelle-Guinée, où elle forme des peuplements denses le long des rivières.

Toutes les espèces du genre *Saccharum* sont polyploïdes. Les clones de *S. officinarum* ont 80 chromosomes. Ce nombre a été établi sur la base de centaines d'observations, et il est vraisemblable que les quelques clones qui ne présentent pas 80 chromosomes sont en fait des hybrides avec d'autres espèces (BREMER, 1924). Les clones de *S. barberi* et de *S. sinense* ont des nombres chromosomiques variant entre 82 et 124. La plupart sont probablement aneu-ploïdes. Pour *S. spontaneum*, le nombre chromosomique varie entre 40 et 128 suivant les clones, et pas moins de 21 cytotypes différents ont été observés en

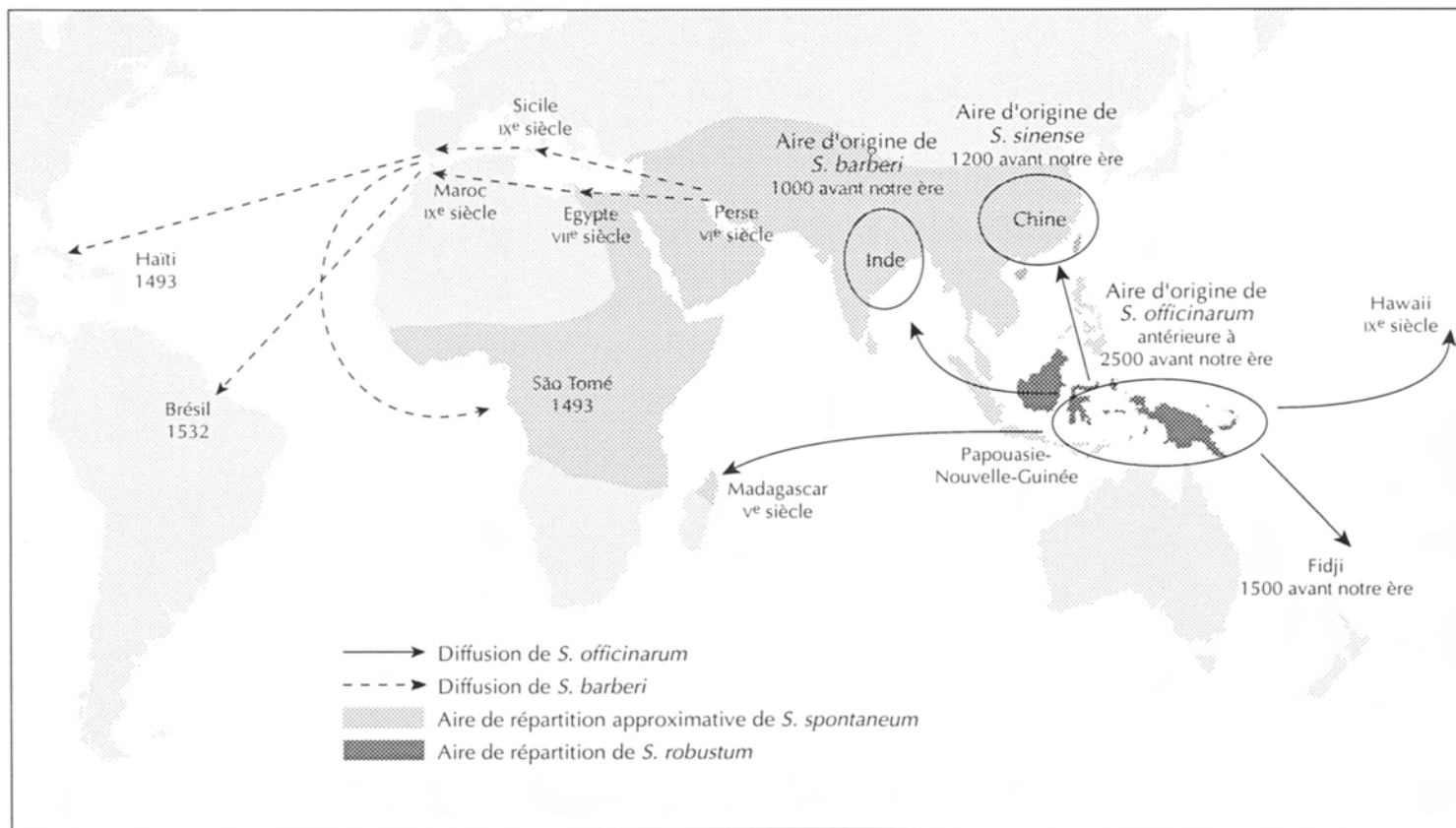


Figure 1. Aire d'origine des trois espèces de canne à sucre domestiquées, *S. officinarum*, *S. barberi* et *S. sinense*, et dispersion jusqu'au début du xv<sup>e</sup> siècle, d'après BLUME (1985), DANIELS et ROACH (1987) et MEYER (1989).

Inde. De nombreux clones seraient aneuploïdes, mais les clones qui ont un nombre chromosomique multiple de huit sont les plus fréquents. Pour *S. robustum*, il existe deux cytotypes majoritaires :  $2n = 60$  et  $2n = 80$  (SREENIVASAN *et al.*, 1987). Les travaux récents d'hybridation *in situ* sur chromosomes (*fluorescent in situ hybridization* ou FISH) apportent la preuve que le nombre de base  $x$  est égal à 8 pour *S. spontaneum* et à 10 pour *S. officinarum* et *S. robustum* (D'HONT *et al.*, 1998).

Dans toutes les espèces du genre *Saccharum*, les chromosomes s'apparient principalement sous la forme de bivalents (PRICE, 1963). Des irrégularités lors de la méiose, comme la formation d'univalents ou de multivalents, sont néanmoins fréquemment observées (BURNER, 1991 ; BURNER et LEGENDRE, 1993).

## L'évolution des formes cultivées

### LA DOMESTICATION ET LA DIFFUSION DES CLONES

La plupart des chercheurs s'accordent sur le fait que la domestication de *S. officinarum* est survenue en Papouasie-Nouvelle-Guinée et dans les îles proches (figure 1). D'une part, il existe dans cette région une exceptionnelle diversité morphologique des clones de *S. officinarum*. D'autre part, l'espèce sauvage locale *S. robustum*, qui serait à l'origine de *S. officinarum*, y est présente (DANIELS et ROACH, 1987). La domestication de la canne serait antérieure à 2500 avant notre ère. La canne était cultivée pour être mâchée et n'a vraisemblablement pas fait l'objet d'une véritable industrie d'extraction du sucre dans son aire d'origine.

Les migrations austronésiennes auraient dispersé la culture de *S. officinarum* vers l'est, dans les îles du Pacifique sud, et vers le nord-ouest, en Inde et en Chine, aux alentours de 1500 à 1000 avant notre ère. Les espèces *S. barberi* et *S. sinense* seraient apparues à cette époque, en Inde pour la première et en Chine pour la seconde. Ces deux pays sont vraisemblablement les centres d'origine de l'industrie d'extraction du sucre.

La dispersion de la culture hors de l'aire d'origine a connu différentes étapes importantes. *S. officinarum* a probablement été disséminé avec les migrations humaines en tant que « canne de bouche », de la Mélanésie vers une grande partie des îles tropicales du Pacifique au cours du premier millénaire de notre ère. Les Européens ne découvriront cette espèce qu'au XVIII<sup>e</sup> siècle lors des premiers voyages exploratoires dans le Pacifique. *S. barberi* a été diffusé vers l'ouest à partir de l'Inde, en tant que matière première de l'industrie sucrière. En 500, la Perse était un foyer réputé pour la production de sucre. Les Arabes ont étendu sa culture à l'Afrique du Nord et aux îles méditerranéennes. Les Portugais puis les Espagnols l'ont ensuite introduite au XV<sup>e</sup> siècle dans les îles de l'Atlantique (Madère, Canaries, îles du Cap-Vert, São Tomé). Enfin, cette espèce a franchi l'Atlantique lors du second voyage de Christophe Colomb et a

été acclimatée pour la première fois en Amérique sur l'île d'Haïti. Au cours des deux siècles suivants, l'extension de la culture de la canne en Amérique, surtout au Brésil puis dans la Caraïbe, a été étroitement liée à la colonisation européenne et a eu pour corollaire l'économie de plantation et la traite des esclaves.

La propagation vers l'ouest de la culture de la canne à partir de l'Inde pendant le premier millénaire, son introduction en Amérique puis le développement des plantations jusqu'au milieu du XVIII<sup>e</sup> siècle n'ont vraisemblablement été réalisés qu'à partir d'un seul clone — ou d'un très petit nombre de clones — baptisé Créole aux Antilles. Il pourrait s'agir d'un clone de *S. barberi* ou d'un hybride entre *S. barberi* et *S. officinarum*.

Au milieu du XVIII<sup>e</sup> siècle, les explorateurs européens ont rapporté des clones de *S. officinarum* du Pacifique sud. Leur culture s'est développée rapidement en Amérique du Sud et du Centre. Ces clones, en raison de leur richesse en sucre, ont été appelés cannes « nobles ». Leur utilisation s'est rapidement développée dans les plantations. Le clone Bourbon, aussi dénommé Vellai, Otaheite et Lahaina, a été exploité presque partout en monoculture jusqu'au milieu du XIX<sup>e</sup> siècle puis, sous la pression parasitaire, il a été remplacé par d'autres clones comme Lousier, la série des Cheribons ou encore Tanna (STEVENSON, 1965).

Si les prospections en Asie du Sud-Est et dans le Pacifique sud ont joué un rôle important dans le renouvellement clonal, il est intéressant de constater que les mutants naturels des variétés cultivées ont également obtenu un certain succès. Ainsi, Lousier serait un mutant de Bourbon et la série des Cheribons correspondrait à un ensemble de mutants de coloration issus d'un même clone.

#### L'UTILISATION DE LA REPRODUCTION SEXUÉE

L'inflorescence de la canne a été reconnue en tant que telle et décrite dès le XVIII<sup>e</sup> siècle mais ce n'est qu'au milieu du XIX<sup>e</sup> siècle, dans l'île de la Barbade, que des graines ont été observées pour la première fois (STEVENSON, 1965). Les premiers programmes de création variétale ont débuté simultanément à la Barbade et à Java vers 1890 et, au début du XX<sup>e</sup> siècle, il existait déjà six stations de sélection dans le monde. Les sélectionneurs se sont dans un premier temps concentrés sur les croisements entre clones nobles de *S. officinarum* et ont remporté quelques succès. A Java, les clones POJ100 et EK28 sont issus de ces programmes de croisements intraspécifiques. Ils ont permis des avancées notables dans la production de sucre de l'île.

Les premiers travaux d'hybridation interspécifique ont débuté à Java dès l'installation de la station de sélection Proefstation Oost Java, au début du XX<sup>e</sup> siècle. Ils reposaient sur la « nobilisation », terme créé par les Hollandais pour désigner le processus qui consiste à croiser un clone noble de *S. officinarum*, riche en sucre, avec un clone d'une espèce apparentée, vigoureux ou

résistant à une maladie, puis à rétrocroiser, éventuellement plusieurs fois, l'hybride sur l'espèce noble de façon à récupérer un phénotype cultivable tout en conservant le caractère intéressant apporté par le clone apparenté.

A l'époque, les plantations de Java étaient ravagées par la mosaïque, une maladie causée par un potyvirus, et par la maladie de sereh, probablement d'origine virale mais qui n'existerait plus aujourd'hui (RANDS et ABBOTT, 1964). Comme aucune source de résistance n'avait été trouvée chez *S. officinarum*, les sélectionneurs ont utilisé Chunnee, un clone résistant de *S. barberi* importé d'Inde. Les descendants n'étaient plus sensibles à la maladie de sereh, mais ils avaient de mauvais rendements en sucre et restaient sensibles à la mosaïque. Certains descendants, comme POJ213, ont été cependant cultivés à grande échelle dans d'autres régions du monde et utilisés avec succès comme géniteurs dans plusieurs stations de sélection, notamment en Inde.

La preuve de l'intérêt des croisements interspécifiques a été apportée dans les années 20. A cette époque, les sélectionneurs ont découvert à Java le clone Kassoer, qui était probablement un hybride spontané entre Black Cheribon, un clone cultivé de *S. officinarum*, et Glagah, un clone local de *S. spontaneum*. Kassoer a été nobilisé une première fois par POJ100 puis une seconde fois par EK28. Parmi les descendants, les chercheurs hollandais ont sélectionné POJ2878, un clone exceptionnel, riche en sucre et résistant à la fois à la mosaïque et à la maladie de sereh. Huit ans seulement après le croisement, POJ2878 occupait 90 % de la sole cannière de Java et s'est ensuite répandu dans le monde entier. Ce clone a connu également un succès phénoménal comme géniteur dans la plupart des stations de sélection.

En Inde, à la station de Coimbatore, les croisements interspécifiques ont commencé également dès le début du xx<sup>e</sup> siècle. Un premier hybride commercial, Co205, a été obtenu après une seule génération de nobilisation (hybride F<sub>1</sub>) entre Bourbon et un clone local de *S. spontaneum*. Ce succès est un exemple unique d'acquisition d'un phénotype commercial intéressant après un croisement simple sans rétrocroisement sur un clone noble. Les sélectionneurs ont développé ensuite des hybrides trispécifiques en croisant leurs hybrides F<sub>1</sub> *S. officinarum*-*S. spontaneum* avec les hybrides *S. officinarum*-*S. barberi* du type POJ213 produits à Java. Les meilleures variétés produites à Coimbatore l'ont été par cette voie.

Les premiers hybrides interspécifiques créés à Proefstation Oost Java et à Coimbatore (POJ2878, Co290, etc.) sont dans la généalogie de presque toutes les variétés cultivées actuelles.

Malgré ces succès, l'étroitesse de la base génétique des variétés commerciales reste une préoccupation majeure pour nombre de sélectionneurs. ARCENEUX (1967) a étudié la généalogie de 114 variétés créées dans les principales stations de sélection au cours de la période 1940-1964. Il a montré que les clones à la base de ces variétés étaient en nombre limité : 19 clones de *S. officinarum*, dont 4 ont joué un rôle particulièrement important (Black Cheribon,

Bandjarmasin Hitam, Loethers et Crystalina) ; quelques clones de *S. spontaneum*, notamment un clone à  $2n = 112$  originaire de Java (Glagah) intervenu par l'intermédiaire d'un seul gamète de l'hybride interspécifique Kassoer, et un ou plusieurs clones à  $2n = 64$  originaires d'Inde, baptisés Coimbatore local ; un clone de *S. barberi* (Chunnee) ; un clone de *S. robustum*, seulement présent dans la généalogie de quelques variétés produites à Hawaii. Ces chiffres contrastent avec les centaines de clones des différentes espèces qui ont été prospectés et qui sont conservés dans différentes collections (BERDING et ROACH, 1987).

Face à cette situation, les travaux de nobilisation ont repris dans plusieurs stations, en Australie, à la Barbade et en Louisiane, dans les années 60 (ROACH, 1978 ; ROACH, 1986 ; BERDING et ROACH, 1987). Plus récemment, des clones appartenant aux genres *Erianthus* et *Miscanthus* ont été utilisés comme source de matériel sauvage. Ces tentatives d'élargissement intergénérique de la base génétique n'ont pas encore donné de résultats significatifs.

#### LA STRUCTURE DU GÉNOME

Les variétés issues de nobilisation ont permis un bond prodigieux des rendements en sucre. *S. spontaneum* a certainement apporté des facteurs de résistance à plusieurs maladies. Le succès rencontré dans le monde entier par les premiers clones hybrides permet de penser qu'ils ont également acquis une meilleure adaptation générale aux conditions de culture, avec notamment une vigueur et un tallage plus importants et une meilleure résistance à la sécheresse et au froid (PANJE, 1972 ; ROACH, 1986). Sur le plan du génome, la contribution de *S. spontaneum* a été déterminée par des mécanismes de transmission particuliers. Les premières générations de croisements interspécifiques et de rétrocroisements ont vu la transmission de  $2n$  chromosomes par le clone de *S. officinarum* utilisé comme parent femelle, alors que le parent mâle transmettait le nombre gamétique normal  $n$ . Il en résulte que les cultivars modernes ont un nombre de chromosomes compris entre 100 et 130 suivant les clones ; environ 10 % de ces chromosomes proviennent de l'espèce sauvage.

L'hybridation *in situ* (*genomic in situ hybridization* ou GISH) permet maintenant de différencier les chromosomes selon leur origine parentale (D'HONT *et al.*, 1996). A titre d'exemple, les études réalisées sur la variété R570 ( $2n \approx 112$ ) montrent qu'à peu près 10 % des chromosomes viennent de l'espèce *spontaneum* et 10 % sont issus de recombinaisons entre des chromosomes des deux espèces parentales. Pour la variété NCo376, on dénombre environ 112 chromosomes, dont à peu près 25 viennent de *S. spontaneum* et 11 sont issus de recombinaisons interspécifiques.

La cartographie moléculaire réalisée à partir des RFLP (GRIVET *et al.*, 1996) indique que les deux espèces ancestrales, qui n'ont pas le même nombre de base, ne se différencient que par quelques réarrangements chromosomiques simples. L'appariement des chromosomes à la méiose semble de type polyso-



mique, comportement typique des autopolyploïdes. Cependant, on a constaté quelques appariements préférentiels entre certains chromosomes issus de l'espèce *S. spontaneum*. Cela peut expliquer le nombre relativement limité de recombinaisons entre chromosomes des deux espèces.

## Les ressources génétiques

La canne est propagée au champ par bouturage de tiges. La production de graines est souvent possible, mais très déstructurante sur le plan génotypique chez cette plante hautement polyploïde et hétérozygote. Les échanges de matériel se font donc essentiellement sous forme de boutures.

La communauté scientifique et interprofessionnelle est très organisée. L'ISSCT (International Society of Sugar Cane Technologists) regroupe la plupart des institutions de recherche et des chercheurs travaillant sur la canne à sucre. Elle est relayée par des sociétés nationales dans chaque pays. Outre des échanges d'informations, l'ISSCT assure la coordination et parfois le financement d'actions d'intérêt général. Dans le domaine des ressources génétiques, l'ISSCT a participé à différentes activités comme la collecte de matériel et la publication de normes à respecter pour les échanges de matériel. Il existe donc une autorité et un cadre pour la conservation et la circulation de ressources génétiques, ainsi qu'une tradition d'échanges entre stations d'amélioration variétale.

Les opérations de collecte ont reposé sur une coopération internationale entre les instituts de recherche sur la canne, l'ISSCT et l'IPGRI (International Plant Genetic Resources Institute), avec l'autorisation des pays prospectés.

Les échantillons prospectés ont été déposés dans deux collections mondiales : celle de l'Inde (Cannanore et Coimbatore) et celle des Etats-Unis (Miami et Canal Point). Ces collections ont reçu aujourd'hui près de 2 500 accessions chacune.

L'intérêt et le soutien de la communauté pour la collecte des ressources génétiques de la canne à sucre n'ont cependant pas été suivis des efforts nécessaires pour leur conservation, leur évaluation et leur utilisation. Les principales banques de gènes ont atteint une taille difficilement compatible avec le maintien et l'évaluation systématique de toutes les accessions. Ainsi, un grand nombre de clones de la collection des Etats-Unis ont été perdus : près de 100 % pour les premiers clones mis en collection à près de 50 % pour les clones collectés en 1976 et en 1977, principalement à cause de maladies ou de catastrophes naturelles (COMSTOCK *et al.*, 1996).

La collection de l'Inde bénéficie de bien meilleures conditions (ROACH, 1992 ; ALEXANDER et VISWANATHAN, 1996). Elle comporte 3 345 accessions, dont plus de la moitié est directement issue des collections. Elle est maintenue dans trois régions aux environnements complémentaires : à Coimbatore, par le Sugarcane Breeding Institute, pour les espèces résistantes à la mosaïque, notamment

les *Erianthus* et la plupart des clones de *S. spontaneum* ; à Cannanore, également par cet institut, pour la plupart des autres matériels (sauf *Miscanthus*) car cette région est exempte de mosaïque ; à Willington, par l'Indian Agricultural Research Institute, pour les clones qui ne supportent pas la conservation à basse altitude, en particulier les représentants du genre *Miscanthus*. Très peu de clones conservés dans ces régions sont perdus. Par ailleurs, pour compenser le risque d'accident dans la culture des clones au champ, une collection *in vitro* a commencé à être mise en place. Il est important de signaler que la collection indienne ne jouit d'aucun soutien financier international.

Il existe de petites collections dans les pays producteurs, mais ce sont plus des collections de travail pour les sélectionneurs que des collections de conservation des ressources génétiques.

Les collections mondiales alimentent régulièrement les programmes de sélection en partie fondés sur l'élargissement de la base génétique. Les échanges et le transport du matériel végétal sous forme de boutures présentent des risques graves de transfert de pathogènes qu'il convient de contrôler. Des règles très strictes sont adoptées et des sites localisés hors des régions de culture ont été identifiés pour implanter des services de quarantaine.

Récemment, la communauté internationale a manifesté son intérêt pour la conservation et l'échange de ressources génétiques au cours de deux ateliers : un atelier de l'ISSCT, qui s'est tenu au Cirad à Montpellier en mars 1994 sur le thème des ressources génétiques de la canne à sucre, et un atelier international sur la conservation et les échanges de matériel génétique, organisé par la communauté cannière australienne, à Brisbane, en juin 1995.

La première réunion a été l'occasion de formuler des priorités en matière de gestion des ressources génétiques — notamment en faveur d'inventaires et d'échanges d'informations plus systématiques — et de standardisation des méthodes de description, en particulier pour les marqueurs moléculaires. L'intérêt d'une *core collection* a été affirmé ; celle-ci pourrait être construite à partir du regroupement et de l'analyse des données existantes, puis diffusée à travers le monde pour une caractérisation complémentaire. La seconde réunion a pris plus spécifiquement en compte les contraintes phytosanitaires qui limitent les échanges et représentent un défi permanent pour les chercheurs et les services de protection des végétaux (CROFT, 1996).

## Le cadre de l'application des marqueurs moléculaires

L'analyse de la diversité génétique de la canne à sucre au moyen des marqueurs moléculaires a été entreprise dès la fin des années 60 avec les isoenzymes (HEINZ, 1969) et les flavonoïdes (WILLIAMS *et al.*, 1974 ; DANIELS et DANIELS, 1975). Ces études, ainsi que celles qui ont suivi, ont apporté des informations importantes sur la structure du genre *Saccharum* et sur ses rela-

tions avec d'autres genres (DANIELS et ROACH, 1987 ; GLASZMANN *et al.*, 1989 ; EKSONTRAMAGE *et al.*, 1992). Les marqueurs liés au polymorphisme de l'ADN ont commencé à être utilisés à la fin des années 80 pour étudier la diversité au sein du genre *Saccharum* (GLASZMANN *et al.*, 1990 ; BURNQUIST *et al.*, 1992). Des études ont été conduites par la suite sur du matériel plus particulier, allant de quelques cultivars à des représentants de genres apparentés à la canne (AL-JANABI *et al.*, 1994 ; SOBRAL *et al.*, 1994 ; HARVEY et BOTHA, 1996 ; BESSE *et al.*, 1997 ; BURNER *et al.*, 1997). Les chercheurs du Cirad ont réalisé différents travaux, en collaboration avec plusieurs partenaires, liés aux objectifs de l'amélioration génétique.

L'application des marqueurs moléculaires a eu pour objectifs de mieux comprendre l'histoire évolutive ayant abouti aux formes cultivées et de rechercher dans quelle mesure la diversité moléculaire peut avoir une valeur prédictive pour des caractères utiles aux sélectionneurs.

La diversité d'intérêt agronomique au sein du genre *Saccharum* n'a pas fait l'objet d'études de grande envergure. Cette situation tient vraisemblablement à la très grande plasticité des caractères, qui rend leur évaluation difficile. Par ailleurs, les interprétations génétiques sont limitées, puisque sont juxtaposées des sources de variation aussi différentes que le nombre de chromosomes, qui va du simple au triple chez *S. spontaneum*, et des mutations ponctuelles telles que celles qui ont accompagné l'évolution clonale des variétés cultivées. De plus, de telles études chez les espèces de base ont un impact faible car la description morphologique du matériel utilisé pour une hybridation interspécifique n'a presque aucune valeur prédictive pour les descendances qui en sont issues (SIMMONDS, 1993).

L'analyse la plus détaillée de la variation morphoagronomique chez les cultivars a été réalisée très récemment à Cannanore, au sud de l'Inde (NAIR *et al.*, 1998). Elle porte sur près de 400 cultivars issus de 10 origines géographiques et concerne essentiellement des caractères quantitatifs qui participent au rendement en sucre de la culture. Il apparaît clairement que les cultivars issus des différentes origines ne présentent pas la même adaptation au site de test. Schématiquement, les deux premiers facteurs de l'analyse multivariée expriment un niveau très variable de performance dans ce milieu, l'un étant construit à partir des corrélations entre diverses mesures de la production de canne (hauteur de tige, poids de canne, rendement en canne), l'autre par des corrélations entre différentes mesures de la richesse en saccharose. Le troisième facteur est déterminé par l'opposition classique entre le tallage et le diamètre des tiges. Il est probable que cette part de la variabilité agromorphologique des cultivars modernes soit essentiellement conditionnée par leur origine hybride interspécifique et l'équilibre entre les diverses composantes génomiques issues de *S. officinarum* et de *S. spontaneum*.

L'application des marqueurs moléculaires a eu plusieurs objectifs spécifiques. Tout d'abord, on a cherché à tester les hypothèses phylogénétiques formulées

pour expliquer les relations entre les espèces du genre *Saccharum*. On a ensuite analysé la diversité nucléaire révélée au moyen des RFLP au sein du matériel cultivé aujourd'hui. Ce matériel étant issu d'hybridations interspécifiques qui impliquent les espèces *S. officinarum* et *S. spontaneum*, les résultats ont été examinés en référence à la diversité de ces deux espèces. Dans une troisième phase, on a cherché à déterminer si la diversité avait conservé les traces de ce qui fait la caractéristique principale de l'origine des cultivars modernes de canne à sucre, à savoir le recours à l'hybridation interspécifique à partir d'un nombre très limité d'accessions suivie de quelques générations seulement d'intercroisements à partir des premiers produits interspécifiques. La conséquence attendue est l'existence de déséquilibres de liaison associant certains marqueurs ; la conséquence espérée est la possibilité d'étendre ce raisonnement à des gènes d'intérêt agronomique et de cibler les études moléculaires futures de sorte qu'elles affinent la compréhension des bases génétiques de la diversité utile pour les sélectionneurs.

## L'organisation de la diversité moléculaire

### Les relations entre les espèces du genre *Saccharum* et les cultivars

Avant de considérer la diversité moléculaire parmi les cultivars, il est utile de les situer par rapport aux principales espèces.

#### LA DIVERSITÉ CYTOPLASMIQUE

La diversité cytoplasmique a été étudiée par D'HONT *et al.* (1993) à l'aide de sondes chloroplastiques et mitochondriales hétérologues pour révéler les RFLP parmi 58 clones représentant différents groupes du complexe *Saccharum*, ainsi que quelques cultivars. La sonde chloroplastique utilisée, bien que couvrant près de 20 % du génome du chloroplaste chez le blé, a permis de différencier seulement les genres, *Saccharum*, *Erianthus* et *Miscanthus*. Les huit sondes mitochondriales utilisées ont permis de différencier 10 types de profil. Parmi les 18 clones de *S. spontaneum*, on a révélé une grande variabilité, avec l'existence de six types de profil suivant une distribution 11, 2, 2, 1, 1, 1. Aucune relation claire entre cette diversité et l'origine géographique des clones n'est apparue. Les 15 clones de *S. robustum* ont montré deux profils, qui se distinguaient par une seule bande selon une distribution 13, 2. Les clones des trois espèces cultivées, *S. officinarum*, *S. barberi* et *S. sinense*, ont montré un seul profil, identique au profil dominant de *S. robustum*. La diversité du génome mitochondrial est en accord avec les hypothèses taxonomiques entre les

espèces sauvages. Les résultats sont compatibles avec l'hypothèse selon laquelle *S. officinarum* est issu de *S. robustum*. Ils concordent aussi avec une origine hybride pour *S. barberi* et *S. sinense* par introgression entre *S. officinarum* et *S. spontaneum*, *S. officinarum* étant le parent femelle. Les quelques cultivars étudiés ont montré le même profil que les clones de *S. officinarum*.

#### LA DIVERSITÉ NUCLÉAIRE

La diversité nucléaire a été étudiée par Lu *et al.* (1994a ; 1994b) à l'aide de sondes nucléaires simple copie sur la base d'une collection de 51 clones représentant différentes espèces de *Saccharum* et 39 cultivars. Les profils d'hybridation ont montré pour chaque clone un grand nombre de bandes avec des intensités variables, ce qui reflète la structure polyploïde complexe des espèces. La plupart des combinaisons enzyme-sonde ont révélé entre 10 et 60 bandes parmi les clones de la collection et entre 10 et 40 bandes parmi les cultivars. Au total, 1 106 bandes polymorphes ont été notées à partir de 36 combinaisons enzyme-sonde. La plupart des bandes étaient présentes chez seulement quelques clones puisque 61 % ont été trouvées chez moins de cinq génotypes et 25 % chez seulement un génotype. Cela a permis de révéler une grande variabilité à l'intérieur de la collection, en particulier chez les espèces sauvages. Par contraste, les bandes étaient généralement plus fréquentes entre les cultivars, ce qui montre une plus grande similarité entre les génotypes cultivés. La matrice formée par les 90 individus et les 1 106 bandes comportant certaines données manquantes, différentes analyses factorielles des correspondances (AFC) ont été réalisées avec certains individus ou certains couples enzyme-sonde considérés comme inactifs pour obtenir des matrices complètes. Ces analyses ont révélé des images globales similaires, qu'elles soient fondées sur l'utilisation de 5, 10, 20 ou 30 combinaisons enzyme-sonde. La figure 2 montre la distribution globale des génotypes obtenus avec 13 combinaisons enzyme-sonde pour lesquelles les données étaient complètes pour presque tout le matériel. Les trois espèces de base, *S. spontaneum*, *S. officinarum* et *S. robustum*, sont clairement différenciées. La distribution selon l'axe 1 sépare les clones de *S. spontaneum* de ceux de *S. robustum* et de *S. officinarum*. *S. robustum* et *S. officinarum* peuvent être distingués selon l'axe 2. La plus grande diversité est observée parmi les génotypes de *S. spontaneum* puis au sein de l'échantillon de *S. robustum*. Les représentants des deux espèces *S. barberi* et *S. sinense* se distribuent entre *S. officinarum* et *S. spontaneum* à proximité des génotypes *S. officinarum*. Ces résultats sont en accord avec l'hypothèse selon laquelle *S. officinarum* a été introduit en Inde et en Chine et pollinisé par des formes locales de *S. spontaneum* pour produire *S. barberi*, en Inde, et *S. sinense*, en Chine. Les cultivars se répartissent entre les clones de *S. officinarum* et de *S. spontaneum*, mais sont plus proches de la première espèce. Cela reflète leur origine interspécifique, mais aussi la nature du schéma de nobilisation dont ils sont issus et qui leur a été appliqué pour leur conférer les principales caractéristiques des cannes nobles.

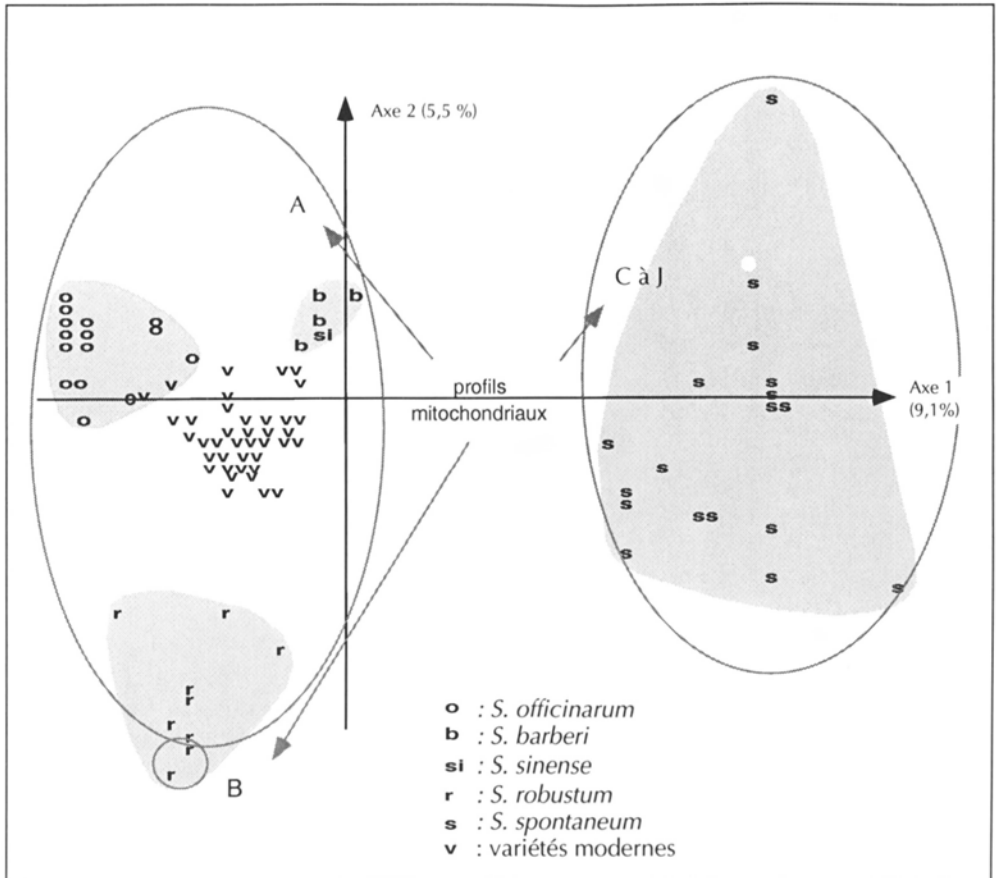


Figure 2. Distribution de 89 clones de canne à sucre cultivés ou sauvages dans le plan 1-2 d'une analyse factorielle des correspondances parmi 463 bandes RFLP polymorphes obtenues avec 13 sondes correspondant à des séquences nucléaires uniques. La distribution des divers cytotypes répertoriés grâce à huit sondes mitochondriales est indiquée.

## La diversité parmi les cultivars

La diversité au sein des cultivars est conditionnée par la diversité au sein des espèces ancestrales et par la transmission de cette diversité lors des croisements interspécifiques, ainsi que par des facteurs de structuration plus fine liés à l'intervention des sélectionneurs et à l'organisation du génome.

### LA DIVERSITÉ AU SEIN DES ESPÈCES ANCESTRALES

Les travaux de LU *et al.* (1994a) ont révélé une diversité considérable au sein de l'espèce *S. spontaneum*. Ainsi, la probabilité qu'une bande présente chez un clone soit aussi présente chez un autre clone qu'on lui compare (indice de Dice) est de 0,31. On peut cependant observer une structuration significative ;

L'analyse multivariée des données extraites pour *S. spontaneum* permet de différencier les génotypes d'Inde à faible nombre chromosomique des génotypes d'Asie du Sud-Est et d'Asie de l'Est, ce qui est en accord avec la classification cytogéographique de PANJE et BABU (1960), qui distingue les génotypes de la région centrale (Inde, Afghanistan) des génotypes de la région orientale (Chine, Asie du Sud-Est). La grande diversité observée parmi les génotypes du groupe oriental suggère la possibilité d'une subdivision plus fine.

La diversité au sein de l'espèce *S. officinarum* a été étudiée par LU *et al.* (1994a) et de façon plus précise par JANNOO *et al.* (1999b). Pour cette dernière étude, un échantillon de 53 clones de *S. officinarum* a été analysé par RFLP au moyen de 11 sondes nucléaires. Les clones représentaient quatre sous-ensembles particuliers : la Nouvelle-Guinée, qui est considérée comme le centre d'origine de l'espèce ; différentes îles indonésiennes (Molouques, Célèbes, Bornéo) ; plusieurs îles du Pacifique (Fidji, Nouvelle-Calédonie) ; des clones d'origine incertaine mais qui entrent largement dans la constitution des cultivars actuels. Les sondes nucléaires étaient réparties sur tous les groupes de liaison connus du génome de la canne à sucre et ont été utilisées en combinaison avec une ou deux enzymes de restriction. Un total de 305 bandes a été révélé. Sur les 53 clones analysés, deux couples de clones totalement identiques ont été relevés et 51 profils uniques ont été conservés pour les étapes suivantes de l'analyse. Les clones présentaient en moyenne entre 4,5 et 7,5 bandes par profil. Ce fort nombre de bandes caractérise le haut niveau de ploïdie de l'espèce et une forte hétérozygotie générale. L'analyse de la distribution de ce paramètre a fait apparaître un sous-ensemble de neuf clones, dont l'hétérozygotie est nettement plus élevée que celle des autres. La figure 3 montre la distribution des clones sur le plan 1-2 d'une AFC réalisée sur ces données. On distingue deux cas. Si l'on considère tous les clones (figure 3a), la structure est essentiellement déterminée par les génotypes qui possèdent le plus grand nombre de bandes : 7 génotypes regroupés dans la partie droite du plan et 2 génotypes en position extrême dans sa partie haute. Ces clones représentent des formes particulières de l'espèce, qui présentent une concentration d'allèles peu fréquents. Leur plus forte hétérozygotie (nombre plus élevé de bandes) peut être l'indicateur d'une origine hybride avec d'autres compartiments du genre ou du complexe *Saccharum*. Si l'on exclut les clones qui présentent le plus fort nombre de bandes afin de limiter l'analyse à un compartiment *S. officinarum* plus homogène, la distribution des clones devient plus continue (figure 3b). On retrouve quelques formes particulières, originaires de Nouvelle-Calédonie notamment, apparentées à un des génotypes en position extrême sur la figure 3a et retirés de cette dernière analyse. Les clones de Nouvelle-Guinée sont distribués dans toute la partie basse du plan, avec cependant une concentration forte au centre. Les clones des îles indonésiennes et ceux qui interviennent dans la généalogie des cultivars ont une distribution voisine de celle des clones de Nouvelle-Guinée. Les clones de Fidji sont intermédiaires entre les formes de Nouvelle-Calédonie et le reste des clones.

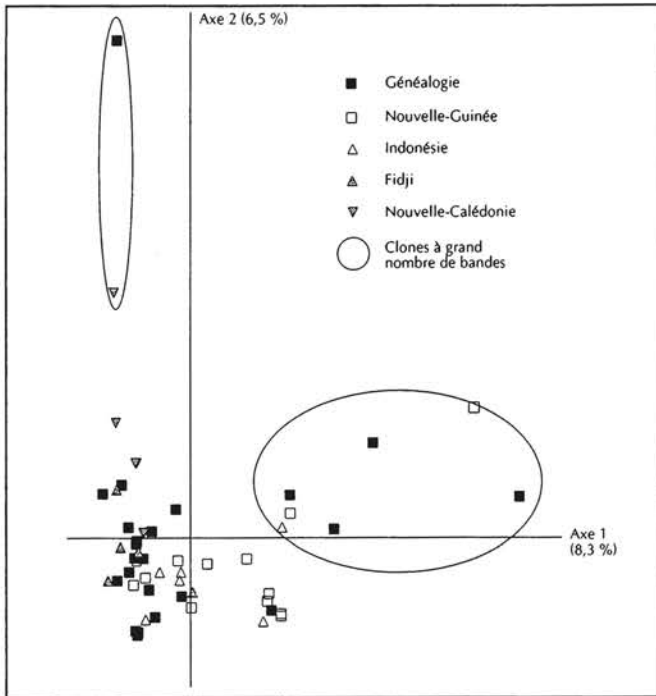


Figure 3a. Distribution de 51 génotypes de *S. officinarum* dans le plan 1-2 d'une analyse factorielle des correspondances parmi 305 bandes RFLP polymorphes obtenues avec 11 sondes correspondant à des séquences nucléaires uniques.

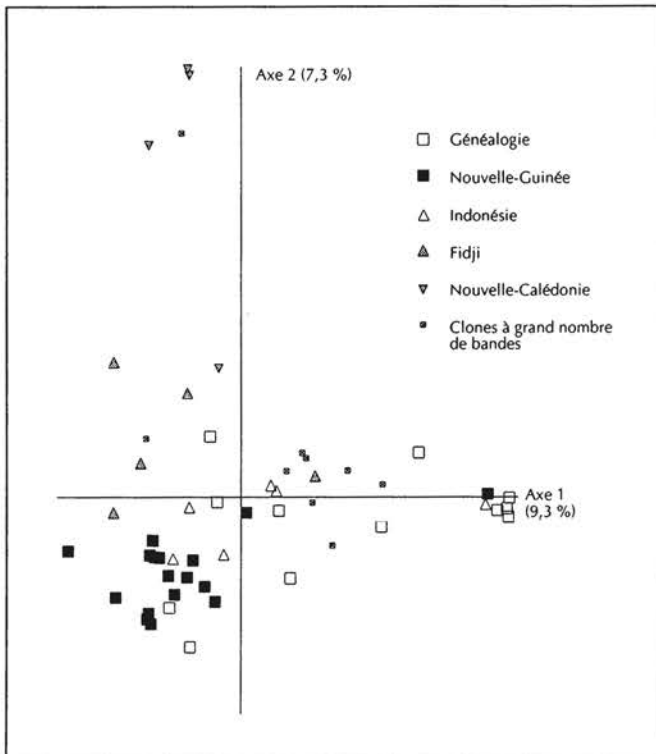


Figure 3b. Distribution de 42 génotypes de *S. officinarum* dans le plan 1-2 d'une analyse factorielle des correspondances parmi 252 bandes RFLP polymorphes obtenues avec 11 sondes correspondant à des séquences nucléaires uniques. Neuf clones ayant un grand nombre de bandes sont projetés en supplémentaires.



On constate donc chez *S. officinarum* une structuration générale probablement déterminée par des introgressions venant d'autres espèces ou d'autres genres. Parmi les formes qui semblent exemptes d'une telle influence, la variation est importante, puisqu'on observe une forte hétérozygotie, mais elle semble peu structurée. Les génotypes utilisés pour créer les cultivars modernes occupent une grande part de la distribution de l'espèce et assurent vraisemblablement une bonne représentation de la diversité de *S. officinarum* au sein du génome des cultivars.

#### LA DÉCOMPOSITION DE LA DIVERSITÉ CHEZ LES CULTIVARS

La diversité moléculaire au sein des cultivars a été étudiée par LU *et al.* (1994b), sur la base d'un échantillon de 39 variétés d'origine très variée, puis par JANNOO *et al.* (1999b) à partir de 109 cultivars principalement issus des programmes de sélection de la Barbade et de Maurice. Cette dernière étude a porté sur 11 sondes combinées avec une ou deux enzymes de restriction et a révélé 336 bandes polymorphes. La variabilité au sein des cultivars est caractérisée tout d'abord par un nombre de bandes plus important que chez *S. officinarum* : 7,4 contre 5,5 bandes par couple enzyme-sonde. Le traitement des données par AFC révèle plusieurs éléments importants.

La variabilité présente chez les clones de *S. officinarum* impliqués dans la généalogie des cultivars est représentée sans biais apparent (figure 4a). Une analyse plus détaillée des fréquences de bandes montre que la majorité des marqueurs se retrouve au sein des cultivars.

La partie structurante de la diversité chez les cultivars est essentiellement apportée par *S. spontaneum* (figure 4b). Les marqueurs qui contribuent le plus aux principaux axes de l'AFC sont généralement absents de l'ensemble des clones de *S. officinarum*.

L'origine des cultivars peut constituer une composante de variabilité significative, même si elle repose sur quelques marqueurs seulement. Le décalage entre les clones issus de la Barbade et ceux de Maurice est porté par le premier axe de l'AFC (figure 4b), bien qu'il soit lié à une différence de fréquence de bande pour 13 marqueurs seulement. Les résultats ne permettent pas de déterminer quels sont les facteurs responsables de cette différenciation, en particulier s'ils s'apparentent aux pratiques des sélectionneurs, comme le recours récurrent à certains géniteurs, ou s'ils dénotent un effet d'adaptation différentielle à des environnements contrastés.

#### LA STRUCTURATION FINE DU POLYMORPHISME

La recherche de déséquilibres de liaison a été conduite sur 59 cultivars, cultivés à l'île Maurice ou utilisés en croisement (JANNOO *et al.*, 1999a). En restreignant l'analyse au matériel issu d'un seul programme de sélection, on limite les facteurs de variation associés à l'origine géographique. On favorise ainsi la

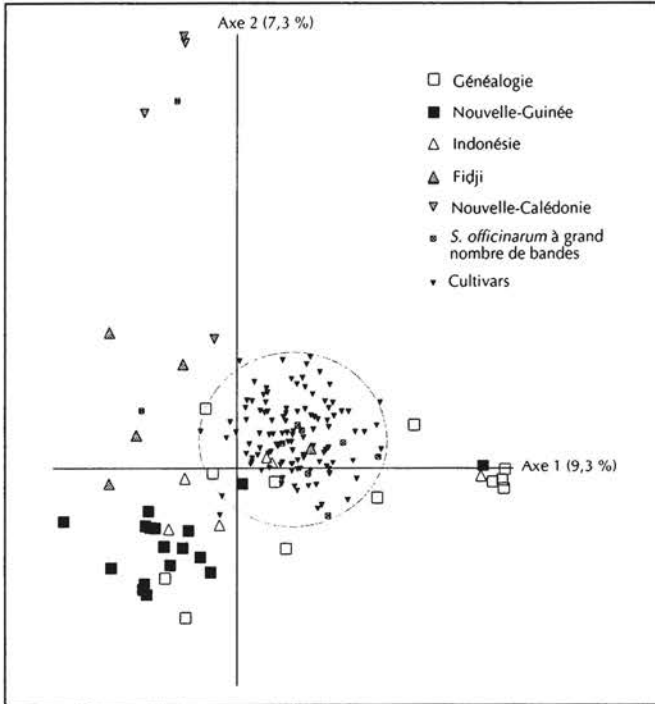


Figure 4a. Distribution de 42 génotypes de *S. officinarum* dans le plan 1-2 d'une analyse factorielle des correspondances parmi 252 bandes RFLP polymorphes obtenues avec 11 sondes correspondant à des séquences nucléaires uniques. Cent neuf cultivars sont projetés en supplémentaires.

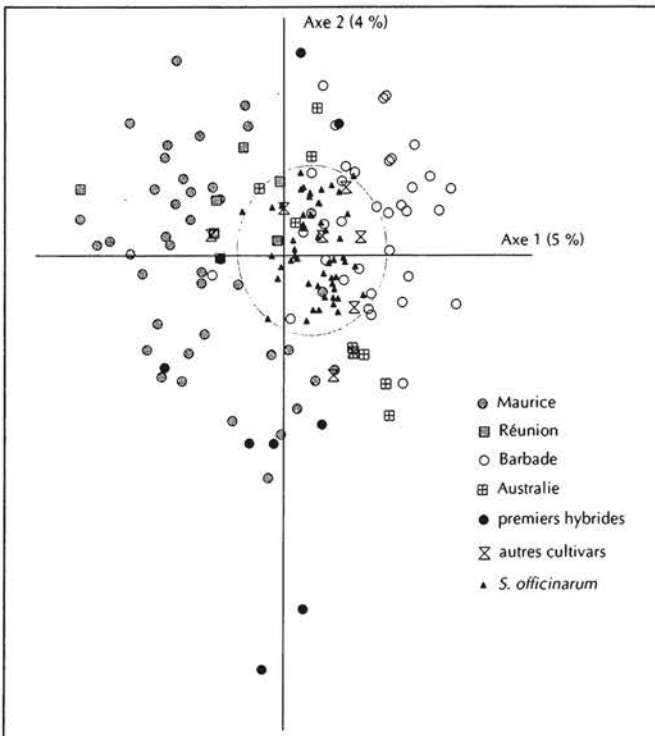


Figure 4b. Distribution de 109 cultivars de canne à sucre dans le plan 1-2 d'une analyse factorielle des correspondances parmi 336 bandes RFLP polymorphes obtenues avec 11 sondes correspondant à des séquences nucléaires uniques. Cinquante et un clones de *S. officinarum* sont projetés en supplémentaires.

révélation d'associations entre marqueurs imputables à la seule liaison physique sur les chromosomes.

Il ressort de cette analyse que les associations impliquent généralement des locus séparés par moins de 10 centimorgans (figure 5). Quarante-deux cas d'association entre au moins deux locus liés ont été répertoriés, représentant des formules alléliques multilocus présentes chez au moins un des premiers géniteurs et donc probablement transmises par lui à l'origine. Environ deux tiers des associations impliquent des marqueurs qui semblent issus de *S. spontaneum*.

Cette composante de la structuration de la diversité au sein des cultivars est interprétée comme le résultat d'un effet de fondation (associé au goulet d'étranglement) intervenu lors du recours à quelques hybrides interpécifiques seulement. Les associations ainsi créées ont pu être maintenues au cours des croisements successifs lorsque la liaison physique était suffisamment forte.

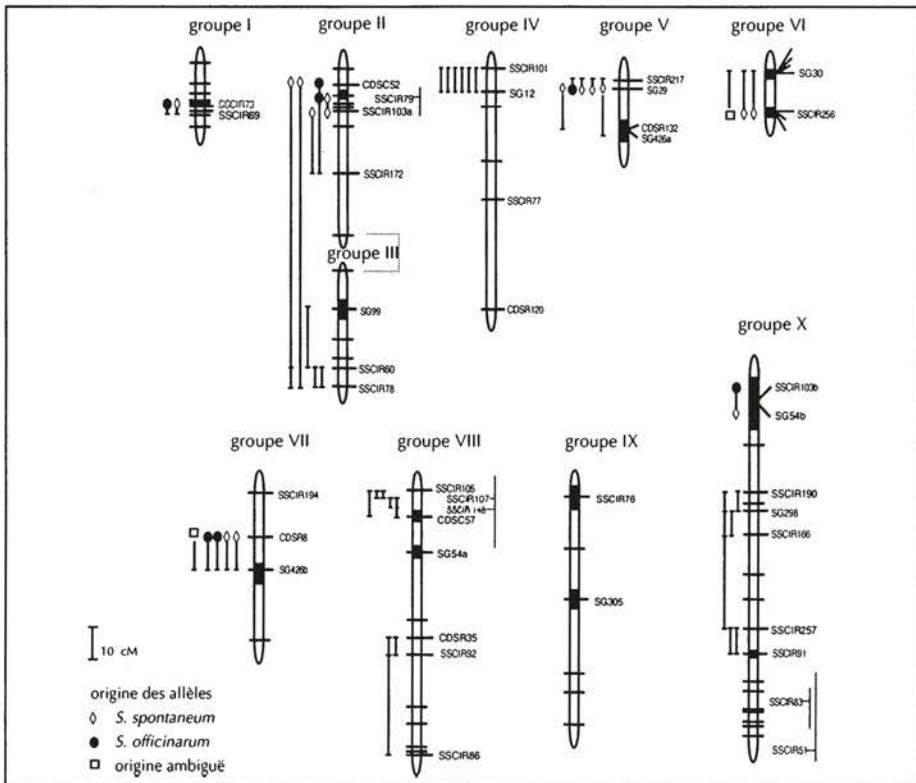


Figure 5. Distribution des déséquilibres de liaison détectés parmi 59 cultivars le long de la carte composite du cultivar R570. Les 38 sondes correspondent aux 41 locus indiqués sur la carte. Les locus dont la position est imprécise sont identifiés par une barre en T sur la droite lorsqu'il s'agit de locus isolés, par une barre noire lorsqu'il s'agit d'un groupe de locus. Les locus impliqués dans un déséquilibre de liaison sont signalés par une barre à gauche du groupe de liaison. Plusieurs barres entre les locus indiquent que plusieurs associations alléliques préférentielles sont observées.

## Conclusion

Les marqueurs moléculaires utilisés lors de ces études sont des RFLP. Ils permettent une caractérisation très détaillée du matériel. Quel que soit l'échantillon considéré, plus de la moitié des bandes révélées sont polymorphes. Du fait de la polyploïdie du matériel et d'une très forte hétérozygotie, chaque sonde révèle généralement plus de cinq bandes par individu. En prenant en compte quelques sondes, on porte rapidement le nombre de bandes comparées au-delà du seuil de discrimination entre les individus. A titre d'exemple, Lu *et al.* (1994b) ont proposé un ensemble de cinq sondes qui, combinées à une seule enzyme de restriction, permettent d'identifier toutes les variétés cultivées sans aucune ambiguïté. Pour les mêmes raisons, il apparaît que quelques sondes seulement suffisent à donner accès à la structure générale de la diversité interspécifique.

Les résultats confirment les hypothèses phylogénétiques antérieures et apportent des éléments complémentaires à l'échelle intraspécifique. Au sein de *S. spontaneum*, on observe un cline géographique qui oppose les formes méridionales dont le nombre chromosomique est élevé aux formes septentrionales, la variation au niveau cytoplasmique présente cependant un profil spécifique indépendant de l'origine géographique. La variation chez *S. officinarum* révèle un centre de diversité étendu sans structure forte et permet de repérer une diversité secondaire, peut-être associée à des introgressions avec d'autres compartiments du complexe.

Ces résultats indiquent que les marqueurs moléculaires, en particulier les RFLP, seront d'une grande utilité pour constituer une *core collection* de ressources génétiques de la canne à sucre. La nature du dispositif international de conservation et de gestion de ces ressources justifierait un investissement de laboratoire proche de la collection indienne, actuellement la plus riche et la mieux gérée.

En revanche, les marqueurs RFLP se révèlent inefficaces pour examiner un niveau taxonomique supérieur, en particulier celui qui permettrait d'élucider les événements, amphi ou autopolyploïdisations par exemple, qui ont abouti au complexe polyploïde *Saccharum*. D'autres modes de marquage plus globaux, comme la GISH, seront probablement plus utiles.

Parmi les cultivars modernes, les marqueurs RFLP permettent en premier lieu d'évaluer la base génétique mise en valeur par rapport aux ressources disponibles, d'analyser sa structure et de la rapporter à différentes composantes et à divers phénomènes. Malgré le faible nombre de géniteurs effectivement utilisés lors des croisements interspécifiques, on constate une très forte diversité des variétés cultivées aujourd'hui : la polyploïdie a assuré le maintien d'une large base génétique. Les déséquilibres de liaison entre marqueurs proches ouvrent des perspectives pour évaluer la contribution qualitative des premiers géniteurs en matière de gènes d'intérêt agronomique. Les progrès de la cartographie du

génomique et du marquage de ces gènes dans des descendance modèles permettront de suivre leur transmission dans le matériel sélectionné et de mettre en relation la diversité moléculaire pour des locus choisis et la diversité des caractères utiles en sélection.

# Annexe

## Analyse par RFLP

Les données produites se ramènent à l'identification de bandes RFLP polymorphes au sein de l'échantillon étudié et à la construction de matrices codées 0-1, correspondant à l'absence ou à la présence de ces bandes chez les accessions analysées. Comme il s'agit de clones, chaque accession a été représentée par un seul individu. L'indice de Dice a été calculé pour quantifier la similarité entre deux accessions, qui correspond au pourcentage de bandes communes par rapport au nombre de bandes présentes chez au moins une des deux accessions comparées. Les matrices ont été traitées par des AFC. Ces AFC ont été réalisées avec l'ensemble des données, puis après avoir retiré certains marqueurs ou certains individus des éléments actifs dans l'analyse. En effet, les marqueurs trop rares ou trop fréquents peuvent donner un poids excessif à certains individus et ainsi masquer une structure générale. Dans tous les cas d'analyse présentés ici, les marqueurs montrant une fréquence inférieure à 5 % ou supérieure à 95 % ont été mis en supplémentaires. La possibilité de placer certains individus en supplémentaires permet de retirer des clones très particuliers ou de situer un ensemble de clones donné dans un cadre de référence fondé sur la diversité des clones maintenus comme actifs. Nous avons utilisé cette possibilité pour examiner les contributions respectives des espèces ancestrales à la structuration de la diversité des cultivars.

## Recherche de déséquilibres de liaison

Trente-huit sondes se cartographiant en 41 locus répartis sur l'ensemble du génome ont été utilisées, ce qui a permis de révéler 1 057 bandes polymorphes. Un test exact de Fisher a été effectué sur l'ensemble des données pour comparer les fréquences d'association entre les locus selon qu'ils appartiennent ou non au même groupe de liaison. Le même test a ensuite été appliqué en limitant les comparaisons aux marqueurs du même groupe de liaison, pour tester si les liaisons fortes étaient plus fréquemment porteuses d'associations alléliques bilocus.

## Références bibliographiques

- ALEXANDER K.C., VISWANATHAN R., 1996. Conservation of sugarcane germplasm in India given the occurrence of new viral diseases. *In* : Sugarcane germplasm conservation and exchange, B.J. Croft *et al.* éd., Brisbane, Australie, ACIAR Proceedings n° 67, p. 19-21.
- AL-JANABI S.M., HONEYCUTT R.J., SOBRAL B.W.S., 1994. Chromosome assortment in *Saccharum*. *Theoretical and Applied Genetics*, 16 : 167-172.
- ARCENEUX G., 1967. Cultivated sugarcanes of the world and their botanical derivation. *In* : XIIIth congress of the International Society of Sugar Cane Technologists, p. 844-854.
- BERDING N., ROACH B.T., 1987. Germplasm collection, maintenance, and use. *In* : Sugarcane improvement through breeding, D.J. Heinz éd., Amsterdam, Pays-Bas, Elsevier, p. 143-210.
- BESSE P., MCINTYRE C.L., BERDING N., 1997. Characterisation of *Erianthus* sect. *Ripidium* and *Saccharum* germplasm (Andropogoneae: Saccharinae) using RFLP markers. *Euphytica*, 93 : 283-292.
- BLUME H., 1985. Geography of sugarcane. Berlin, Allemagne, Albert Bartens, 391 p.
- BREMER G., 1924. The cytology of sugarcane: a cytological investigation of some cultivated kinds and of their parents. *Genetica*, 5 : 97-148 ; 273-326.
- BURNER D.M., 1991. Cytogenetic analyses of sugarcane relatives (Andropogoneae: Saccharinae). *Euphytica*, 54 : 125-133.
- BURNER D.M., LEGENDRE B.L., 1993. Chromosome transmission and meiotic stability of sugarcane (*Saccharum* spp.) hybrid derivatives. *Crop Science*, 33 : 600-606.
- BURNER D.M., PAN Y.B., WEBSTER R.D., 1997. Genetic diversity of North American and Old World *Saccharum* assessed by RAPD analysis. *Genetic Resources and Crop Evolution*, 44 : 235-240.
- BURNQUIST W.L., SORRELS M.E., TANKSLEY S., 1992. Characterization of genetic variability in *Saccharum* germplasm by means of restriction fragment length polymorphism (RFLP) analysis. *In* : XXIst congress of the International Society of Sugar Cane Technologists, volume 2, p. 355-365.
- COMSTOCK J.C., SCHNELL R.J., MILLER J.D., 1996. Current status of world germplasm collection in Florida. *In* : Sugarcane germplasm conservation and exchange, B.J. Croft *et al.* éd., Brisbane, Australie, ACIAR Proceedings n° 67, p. 17-18.
- CROFT B.J., 1996. Review of restrictions to free access to germplasm exchange facing Australian and other international sugar industries. *In* : Sugarcane germplasm conservation and exchange, B.J. Croft *et al.* éd., Brisbane, Australie, ACIAR Proceedings n° 67, p. 6-9.
- DANIELS J., DANIELS C.A., 1975. Geographical, historical and cultural aspects of the origin of the Indian and Chinese sugarcanes *S. barberi* and *S. sinense*. *Sugarcane Breeding Newsletter*, 36 : 4-23.
- DANIELS J., ROACH B.T., 1987. Taxonomy and evolution. *In* : Sugarcane improvement through breeding, D.J. Heinz éd., Amsterdam, Pays-Bas, Elsevier, p. 7-84.

- D'HONT A., GRIVET L., FELDMANN P., RAO S., BERDING N., GLASZMANN J.C., 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular and General Genetics*, 250 : 405-413.
- D'HONT A., ISON D., ALIX K., ROUX C., GLASZMANN J.C., 1998. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of RNA genes. *Genome*, 41 : 221-225.
- D'HONT A., LU Y.H., FELDMANN P., GLASZMANN J.C., 1993. Cytoplasmic diversity in sugarcane revealed by heterologous probes. *Sugar Cane*, 1 : 12-15.
- EKSOMTRAMAGE T., PAULET F., NOYER J.L., FELDMANN P., GLASZMANN J.C., 1992. Utility of isozymes in sugarcane breeding. *Sugar Cane*, 3 : 14-21.
- FAUCONNIER R., 1991. La canne à sucre. Paris, France, Maisonneuve et Larose, Le Technicien d'agriculture tropicale, 168 p.
- GLASZMANN J.C., FAUTRET A., NOYER J.L., FELDMANN P., LANAUD C., 1989. Biochemical genetic markers in sugarcane. *Theoretical and Applied Genetics*, 78 : 537-543.
- GLASZMANN J.C., LU Y.H., LANAUD C., 1990. Variation of nuclear ribosomal DNA in sugarcane. *Journal of Genetics and Breeding*, 44 : 191-198.
- GRIVET L., D'HONT A., ROQUES D., FELDMANN P., LANAUD C., GLASZMANN J.C., 1996. RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a high polyploid and aneuploid interspecific hybrid. *Genetics*, 142 : 987-1000.
- HARVEY M., BOTHA F.C., 1996. Use of PCR-based methodologies for determination of DNA diversity between *Saccharum* varieties. *Euphytica*, 89 : 257-265.
- HEINZ D.J., 1969. Isozyme prints for variety identification. *Sugarcane Breeding Newsletter*, 24 : 8.
- JANNOO N., GRIVET L., DOOKUN A., D'HONT A., GLASZMANN J.C., 1999a. Linkage disequilibrium among sugarcane cultivars. *Theoretical and Applied Genetics* (sous presse).
- JANNOO N., GRIVET L., SEGUIN M., PAULET F., DOMAINGUE R., ROA P.S., DOOKUN A., D'HONT A., GLASZMANN J.C., 1999b. Molecular investigation of the genetic base of sugarcane cultivars. *Theoretical and Applied Genetics* (sous presse).
- LU Y.H., D'HONT A., PAULET F., GRIVET L., ARNAUD M., GLASZMANN J.C., 1994b. Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica*, 78 : 217-226.
- LU Y.H., D'HONT A., WALKER D.I.T., RAO P.S., 1994a. Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. *Euphytica*, 78 : 7-18.
- MEYER J., 1989. Histoire du sucre. Paris, France, Desjonquères, 335 p.
- NAIR N.V., BALAKRISHNAN R., SREENIVASAN T.V., 1998. Variability for quantitative traits in exotic hybrid germplasm of sugarcane. *Genetic Resources and Crop Evolution*, 45 : 459-464.
- PANJE R.R., 1972. The role of *Saccharum spontaneum* in sugarcane breeding. *In* : XIVth congress of the International Society of Sugar Cane Technologists, p. 217-223.



- PANJE R.R., BABU C.N., 1960. Studies in *Saccharum spontaneum*: distribution and geographical association of chromosome number. *Cytologia*, 25 : 152-172.
- PRICE S., 1963. Cytogenetics of modern sugarcane. *Economic Botany*, 17 : 97-105.
- RANDS R.D., ABBOT E.V., 1964. Sereh. *In* : Sugarcane diseases of the world, C.G. Hughes *et al.* éd., Amsterdam, Pays-Bas, Elsevier, p. 183-189.
- ROACH B.T., 1978. Utilisation of *Saccharum* in sugarcane breeding. *In* : XVIth congress of the International Society of Sugar Cane Technologists, p. 43-58.
- ROACH B.T., 1986. Evaluation and breeding use of sugarcane. *In* : XIXth congress of the International Society of Sugar Cane Technologists, p. 492-501.
- ROACH B.T., 1992. The case for a core collection of sugarcane germplasm. *In* : XXIst congress of the International Society of Sugar Cane Technologists.
- SIMMONDS N.W., 1993. Introgression and incorporation strategies for the use of crop genetic resources. *Biological Review*, 68 : 539-562.
- SOBRAL B.W.S., BRAGA D.P.V., LAHOOD E.S., KEIM P., 1994. Phylogenetic analysis of chloroplast restriction enzyme site mutations in the Saccharinae Griseb. subtribe of the Andropogoneae Dumort. tribe. *Theoretical and Applied Genetics*, 87 : 843-853.
- SREENIVASAN T.V., AHLOOWALIA B.S., HEINZ D.J., 1987. Cytogenetics. *In* : Sugarcane improvement through breeding, D.J. Heinz éd., Amsterdam, Pays-Bas, Elsevier, p. 211-253.
- STEVENSON G.C., 1965. Genetics and breeding of sugarcane. Londres, Royaume-Uni, Longman, 284 p.
- WILLIAMS C.A., HARBORNE J.B., SMITH P., 1974. The taxonomic significance of leaf flavonoids in *Saccharum* and related genera. *Phytochemistry*, 13 : 1141-1149.

# Le cocotier

---

Patricia Lebrun, Yavo-Pierre N'cho,  
Roland Bourdeix, Luc Baudouin

**S**i le cocotier est une plante emblématique des paysages côtiers tropicaux, il représente également une ressource vitale pour de nombreuses populations de ces régions. Sa culture couvre environ 11 millions d'hectares, dont 94 % se situent en Asie et dans le Pacifique (BOURDEIX *et al.*, 1997). Les principaux pays producteurs sont les Philippines et l'Indonésie. C'est une culture essentiellement paysanne, le plus souvent plantée sur de petites surfaces. Les grandes plantations représentent moins de 10 % du total.

Presque toutes les parties du cocotier sont exploitées et leurs utilisations sont innombrables (PERSLEY, 1992). Le bois, bien que difficile à mettre en œuvre, est d'excellente qualité. Les palmes sont employées pour couvrir les habitations. Les nervures des folioles servent à fabriquer des balais. La sève tirée des inflorescences donne du sucre et des boissons alcoolisées. Les racines fournissent des colorants et sont utilisées en médecine traditionnelle.

Mais c'est son fruit qui fait la réputation du cocotier : l'épiderme de cette drupe recouvre une bourre épaisse dont les fibres sont largement utilisées sous le nom de *coir*. À l'intérieur se trouve une graine volumineuse — la noix de coco —, qui comprend une coque brune lignifiée et un albumen dont la partie périphérique est solidifiée à maturité. La cavité restante renferme la partie liquide de l'albumen : l'eau de coco. Consommée immature, la noix de coco

donne une boisson rafraîchissante et sucrée. Mûre, elle est surtout appréciée pour son albumen solide, qui peut être consommé directement ou après diverses transformations. L'albumen séché fait l'objet, sous le nom de coprah, d'un commerce international important. C'est en effet la source d'une des principales huiles de type laurique, particulièrement appréciée en savonnerie. Cette huile est également utilisée en alimentation et dans les cosmétiques.

## La taxonomie et les ressources génétiques

### La botanique et la taxonomie

Le cocotier (*Cocos nucifera* L.) est une monocotylédone arborescente diploïde ( $2n = 2x = 32$ ) de la famille des arecacées. Il forme un genre monospécifique, sans espèce sauvage étroitement apparentée. C'est un palmier dont le stipe non ramifié porte un bouquet de palmes produites au rythme d'environ une par mois. A l'aisselle de chaque palme, une inflorescence ramifiée est émise. Chacun des épillets porte quelques fleurs femelles à la base et de nombreuses fleurs mâles au sommet.

La floraison est protandre et le déroulement de ses cycles explique le comportement reproductif des deux principaux types de cocotiers, dans lesquels se rangent les différents cultivars (ROGNON, 1976). Les cocotiers « Grands » sont principalement allogames : la floraison femelle commence après la fin de la floraison mâle. Un certain taux d'autogamie est cependant possible lorsque la floraison mâle chevauche la floraison femelle de l'inflorescence précédente. Les cocotiers Grands se caractérisent aussi par une croissance rapide, la présence d'un bulbe volumineux à la base du stipe et des cicatrices foliaires espacées. Les cocotiers « Nains » sont le plus souvent autogames : la floraison femelle se déroule entièrement (ou, comme pour le Nain Vert Brésil, en grande partie) avant que la floraison mâle ne s'achève. Ils représentent une faible partie de la population mondiale et sont généralement situés près des habitations. Outre leur autogamie, les Nains se distinguent par une croissance plus lente, des cicatrices foliaires rapprochées, une plus grande précocité, des noix plus petites et à l'épiderme souvent vivement coloré. Signalons une variété de cocotiers semblable aux Nains par la taille, mais allogame : le Nain Niu Leka.

### Les ressources génétiques

#### L'ORIGINE ET L'AIRE DE DISTRIBUTION

Bien qu'une origine américaine ait été évoquée, la plupart des auteurs s'accordent aujourd'hui pour rattacher le cocotier au centre d'origine indo-malais tel que Vavilov l'a défini (CHILD, 1964 ; ZOHARY, 1970). Il est à présent répandu

dans l'ensemble des zones côtières intertropicales. Les graines auraient été disséminées par flottaison vers le Pacifique sud, mais surtout par l'homme, qui est depuis longtemps son principal agent de dissémination. On signale çà et là quelques peuplements spontanés, mais l'immense majorité des cocotiers ont été plantés.

#### LA DIVERSITÉ DES POPULATIONS

On recense dans les collections conservées à travers le monde plus de 300 variétés locales (« cultivars » ou « écotypes ») de cocotier, identifiées d'après leur provenance géographique et la morphologie de la plante et du fruit. Outre la distinction entre Nains et Grands, HARRIES (1978) propose de regrouper les cultivars Grands en deux sous-types : le type Niu Kafa, au stipe élancé et aux fruits allongés, de section triangulaire et riches en bourre, représenterait le type sauvage, adapté à la dissémination par les courants marins ; le type Niu Vai, au stipe plus massif, au port plus rigide et aux fruits arrondis, riches en eau de coco, aurait été domestiqué anciennement. Un certain nombre de populations qui possèdent des caractères intermédiaires proviendraient d'une introgression des populations « sauvages » par des cultivars domestiqués. Cela dit, les deux types conviennent à la production de coprah et ont été largement plantés à partir de la fin du XIX<sup>e</sup> siècle.

Différentes méthodes ont été employées pour caractériser les cultivars de cocotier et étudier les relations entre populations. L'évaluation des caractères morphologiques et agronomiques, pour lesquels N'CHO *et al.* (1993) apportent des informations précieuses tant sur la diversité que sur l'utilisation potentielle des cultivars, a une efficacité limitée du fait de l'influence de l'environnement. Les marqueurs enzymatiques étudiés par BENOIT et GHESQUIERE (1984) révèlent un faible polymorphisme et apportent peu d'informations. Ils ont été repris récemment, en Indonésie (HARTANA *et al.*, 1993) et au Sri Lanka (FERNANDO et GAMINI, 1997), avec plus de succès, mais le nombre de systèmes utilisables reste faible. Les polyphénols ont fourni des résultats prometteurs (JAY *et al.*, 1989), mais qui se sont révélés peu reproductibles d'un environnement à l'autre. Les marqueurs moléculaires font l'objet de recherches actives selon différentes méthodes : ISTR (ROHDE *et al.*, 1995), RAPD (ASHBURNER *et al.*, 1997), AFLP (PERERA *et al.*, 1998), RFLP (LEBRUN *et al.*, 1998a ; 1998b), micro-satellites.

#### L'UTILISATION EN AMÉLIORATION

Les travaux d'hybridation entre cocotiers Nains ont débuté à Fidji (MARECHAL, 1928), et PATEL (1938) a réalisé les premiers hybrides Nain × Grand. Dans les années 40 à 60, des tests comparatifs comprenant des hybrides Nain × Grand et Grand × Grand ont permis de démontrer la supériorité des hybrides en terme de potentiel de production (DE NUCE DE LAMOTHE et BENARD, 1985). THAMMES (1955) a proposé d'interplanter en isolement deux variétés, l'une

utilisée comme mâle et l'autre comme femelle, ce qui a rendu possible la production en masse des variétés hybrides. Ultérieurement, la pollinisation assistée, qui consiste à effectuer un apport massif de pollen exogène sur les inflorescences émasculées, a remplacé progressivement cette méthode (DE NUCE DE LAMOTHE et ROGNON, 1972).

Une stratégie d'amélioration du cocotier a été proposée par BOURDEIX *et al.* (1990 ; 1991a ; 1991b). Elle s'inspire du schéma de sélection récurrente réciproque mis au point par COMSTOCK *et al.* (1949) pour le maïs. Son originalité réside dans la conduite simultanée de deux axes de sélection, l'un orienté vers les hybrides Grand  $\times$  Grand et l'autre, vers les hybrides Nain  $\times$  Grand (figure 1).

La caractérisation des nombreux cultivars disponibles est indispensable à leur utilisation en amélioration et à la recherche d'hétérosis, mais aussi à leur conservation comme source de variabilité. Ce chapitre revient sur les résultats de l'étude de LEBRUN *et al.* (1998b) et tente de dégager ce que les outils de marquage moléculaire peuvent apporter à la connaissance et à l'amélioration du cocotier par rapport aux méthodologies précédemment utilisées.

## L'organisation de la diversité

### La diversité nucléaire et mitochondriale révélée par les RFLP

Pour les RFLP nucléaires, les 25 couples enzyme-sonde utilisés permettent de révéler 60 bandes polymorphes parmi les 289 arbres analysés. La sonde Cox 1 met en évidence deux profils mitochondriaux.

Les analyses factorielles des correspondances (AFC) et les informations géographiques permettent de constituer des groupes de cultivars relativement homogènes et ensuite d'identifier les marqueurs distinctifs de chaque groupe.

#### LES DEUX ZONES MAJEURES DE DIVERSITÉ DES COCOTIERS GRANDS

On distingue sur la droite de l'AFC (figure 2) un premier groupe qui correspond à l'ensemble des Grands de la zone écogéographique du Pacifique, comprenant l'Asie du Sud-Est et le Pacifique sud, auxquels se rattachent tous les cultivars Nains et les cocotiers du Panamá : ce premier groupe est appelé « groupe pacifique ». Un autre ensemble, situé à gauche, correspond aux cocotiers de l'Inde, du Sri Lanka et de l'Afrique de l'Ouest : c'est le « groupe indo-atlantique ». Ces deux groupes sont séparés par un ensemble de trois cultivars, qui forment un troisième groupe, riverain de l'océan Indien, le « groupe océan Indien ».

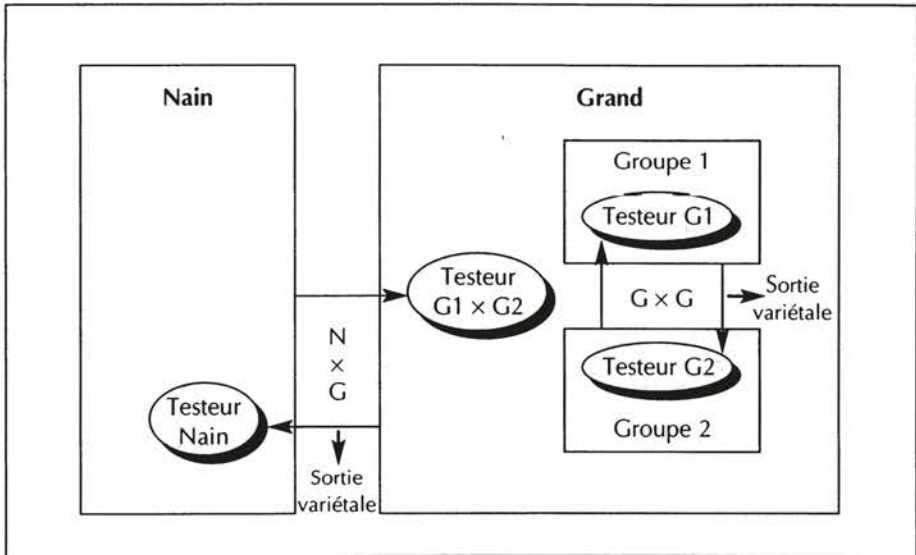


Figure 1. Articulation des deux axes de la sélection réciproque chez le cocotier, d'après BOURDEIX et al. (1991a).

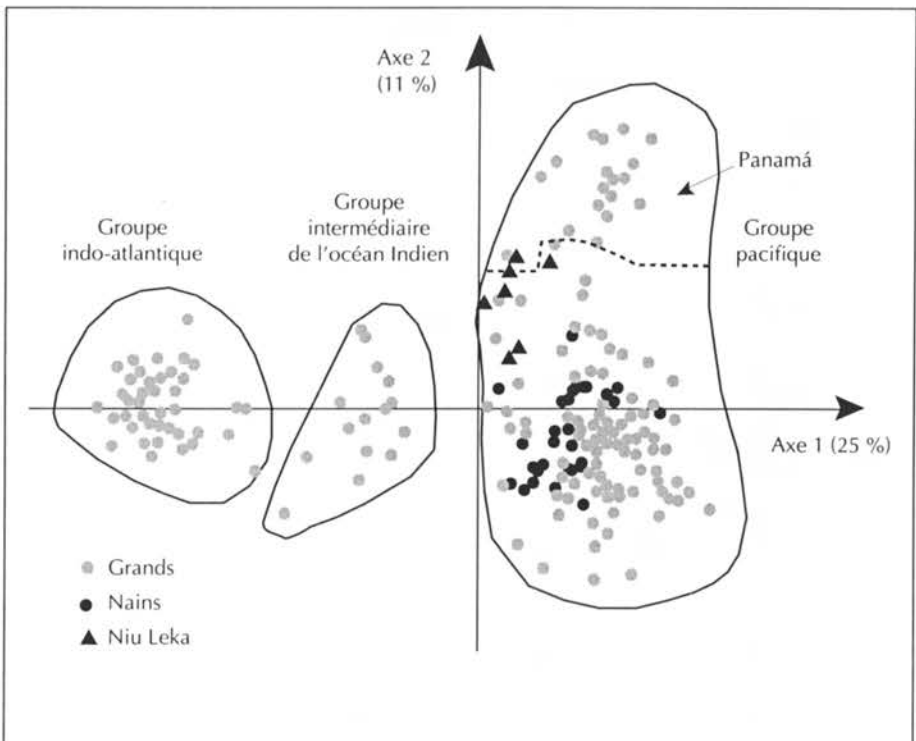


Figure 2. Synthèse des données RFLP : les grands groupes de cultivars.

En termes de marqueurs nucléaires, le groupe indo-atlantique se distingue par la présence prédominante de cinq marqueurs, qui sont quasiment absents ailleurs, et de trois autres, qu'il partage avec les seules populations d'Asie du Sud-Est. Vingt-et-un marqueurs sont absents ou présentent une faible fréquence dans ce groupe, alors qu'ils sont bien représentés chez la plupart des autres Grands. Enfin, l'allèle « rapide » du marqueur cytoplasmique Cox 1 semble caractéristique de ce groupe : sur les 45 individus qui le composent, 35 présentent en effet cet allèle, alors que tous les individus du groupe pacifique, sauf un individu Grand Tonga, possèdent l'allèle « lent ». Le Grand Kappadam constitue un cas particulier puisque les cinq individus qui le représentent portent l'allèle « lent ». En outre, bien que ses marqueurs nucléaires soient pour l'essentiel caractéristiques du groupe indo-atlantique, on observe une fréquence plus élevée que celle attendue pour quelques marqueurs nucléaires généralement trouvés dans le groupe pacifique. Nous verrons plus loin que ces deux anomalies apparentes pourraient constituer la clé de l'origine du Grand Kappadam, dont les caractères du fruit sont très particuliers pour cette région.

Ces deux groupes se distinguent également par leur degré de polymorphisme : le groupe pacifique présente plus de marqueurs communs aux différents cultivars — 42 contre 33 pour le groupe indo-atlantique — et plus de marqueurs de fréquence intermédiaire, comprise entre 10 % et 90 %, — 20 à 38 bandes selon le cultivar, contre 11 à 20 pour les indo-atlantiques.

Le groupe intermédiaire de l'océan Indien présente presque tous les marqueurs nucléaires des deux premiers groupes, c'est donc celui qui regroupe la plus large diversité moléculaire. L'allèle cytoplasmique est l'allèle « lent », sauf pour deux individus du Grand Mozambique.

#### LA DIVERSITÉ DANS LE GROUPE PACIFIQUE

Afin de mieux caractériser la diversité du groupe pacifique, une seconde AFC a été réalisée (figure 3). En premier lieu, on note une gradation géographique entre l'Asie du Sud-Est et le Pacifique sud selon le second axe : les cocotiers de Polynésie apparaissent en bas de la figure, ceux du Sud-Est asiatique, en haut, et ceux de Mélanésie, en position intermédiaire. Contrairement aux divergences relevées entre les groupes, les différences portent ici moins sur l'existence d'allèles spécifiques que sur des variations de fréquence. On remarque également un certain degré de chevauchement entre sous-régions voisines.

On constate en second lieu que les trois populations du Panamá s'individualisent très nettement sur le premier axe. Elles se rattachent aux autres cultivars du groupe pacifique quant à leur profil RFLP, dans la mesure où tous les marqueurs fréquents au Panamá sont présents dans ce groupe. Cependant, 19 bandes de ce groupe sont absentes ou rares au Panamá. Certains marqueurs très fréquents dans les populations du Panamá sont plus fréquents en Polynésie que dans le reste du groupe pacifique, pour d'autres, c'est le contraire. Les



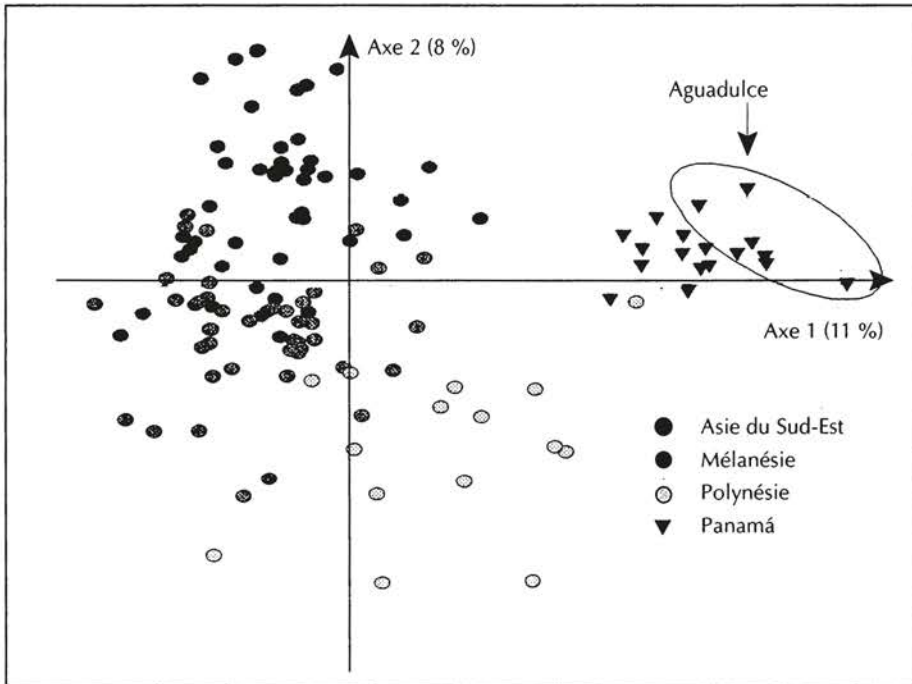


Figure 3. Synthèse des données RFLP : la diversité au sein du groupe pacifique.

marqueurs de fréquence intermédiaire sont peu nombreux (de 12 à 17). La population Aguadulce se distingue légèrement des deux autres populations panaméennes, en raison notamment de la présence de quelques bandes spécifiques des indo-atlantiques, avec une faible fréquence. En revanche, les deux autres populations, Monagre et Bowden (Jamaïque) semblent indiscernables.

#### LES COCOTIERS NAINS

Les cocotiers Nains sont essentiellement autogames. Cette caractéristique se traduit par la quasi-absence des bandes de fréquence intermédiaire et par un taux d'hétérozygotie très faible. Les deux exceptions sont le Niu Leka, seul cultivar Nain franchement allogame, et le Nain Vert Malais, connu pour être partiellement allogame. Toutes les bandes communes aux Nains se retrouvent dans le groupe pacifique, mais 13 bandes présentes dans ce groupe sont absentes de la plupart des Nains. Parmi celles-ci, plusieurs correspondent à la fixation d'allèles de fréquence intermédiaire chez les Grands, ce qui suggère fortement une origine commune.

Quatre cultivars Nains d'origines et de couleurs distinctes ont un profil quasiment identique : Nains Verts Sri Lanka et Kiribati, Nains Bruns Ternate et Madang. D'un autre côté, les Nains Malais forment un groupe homogène, qui se distingue par la fixation de cinq allèles différents de ceux qui se trouvent



chez les autres Nains. Les sept Nains Jaunes Ghana et les quinze Nains Jaunes Malais ont exactement le même profil, ce qui confirme l'identité de ces deux cultivars. Les trois cultivars Nains des Philippines, ainsi que le Nain Vert Brésil, présentent également des traits communs, tandis que le profil de deux Nains Rouges du Pacifique sud et celui du Nain Rouge Cameroun (un cultivar prélevé au Cameroun mais probablement exotique) apparaissent nettement divergents bien que certains traits les rapprochent de l'ensemble des Nains.

Enfin, le Nain Niu Leka présente un profil qui rappelle celui des Grands de sa région d'origine (Grand Tonga et Grand Rotuma) et porte l'allèle « rapide » de Cox 1, rare dans la région, mais présent chez un individu Grand Tonga. La présence de cet allèle typique du groupe indo-atlantique dans cette région ne s'explique ni au vu des données historiques, ni au vu des caractères morpho-agronomiques. Il est vraisemblable que les deux régions possèdent deux allèles différents mais indiscernables. Quoi qu'il en soit, le Niu Leka se distingue des autres Nains, qui semblent se rapprocher plus des cocotiers d'Asie du Sud-Est ou de Mélanésie.

## Les polyphénols

La figure 4 résume les principaux résultats de l'étude des polyphénols réalisée par JAY *et al.* (1989). Il s'agit d'une analyse discriminante, qui fait intervenir 32 cultivars (ou écotypes), chacun d'entre eux étant représenté en moyenne par cinq individus. Les cultivars étudiés sont inclus dans l'analyse RFLP à l'exception du Nain Vert Thaïlande et du Grand de Polynésie Tahiti. A l'inverse de l'analyse factorielle des correspondances réalisée sur les marqueurs moléculaires, l'origine des individus figure explicitement dans les données fournies à l'analyse statistique. Même en tenant compte de cette différence, les principales conclusions de cette étude convergent avec celles des données moléculaires. On peut par exemple noter que les cultivars Grands se répartissent autour de trois grandes zones : Afrique, Pacifique sud et Extrême-Orient, ces deux dernières se recouvrant partiellement. En fait, les cocotiers indiens sont répartis entre plusieurs groupes. Ceux des Laccadives, un archipel situé au sud-ouest de l'Inde, sont proches des cocotiers africains alors que ceux des Andaman se situent entre les populations d'Extrême-Orient et celles du Pacifique. Les cultivars d'Afrique de l'Est sont eux aussi décalés vers le centre de la figure par rapport aux autres cultivars africains. En outre, tous les Nains apparaissent proches des Grands d'Extrême-Orient et de Mélanésie. Cependant, il n'y a pas de lien apparent entre l'origine géographique d'un Nain et son profil polyphénolique. C'est en particulier le cas de ceux qui ont été collectés au Sri Lanka et en Afrique, et dont l'origine exotique est clairement confirmée.

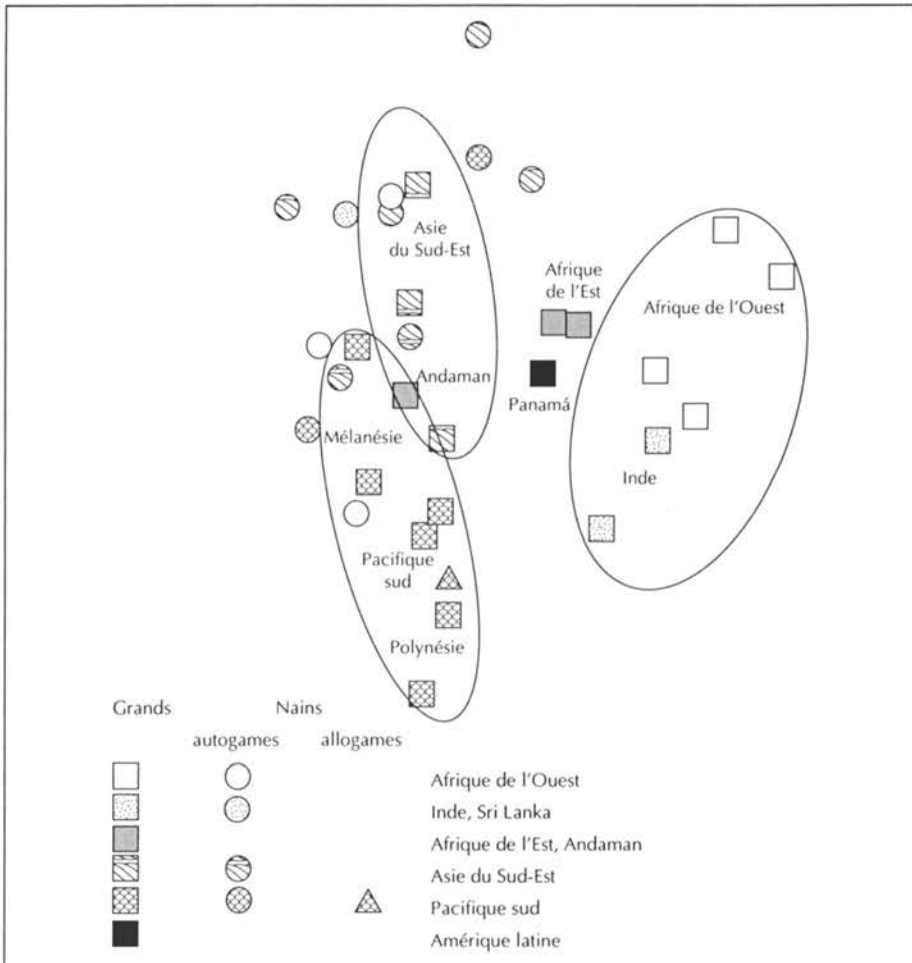


Figure 4. Données polyphénoliques : analyse discriminante de l'ensemble des individus selon le critère cultivar (plan 1-2), d'après JAY et al. (1989).

Ainsi, en dépit des différences dans la méthode d'analyse statistique et dans le mode de présentation, les résultats des deux approches sont globalement convergents. La différence la plus notable tient à la situation du Grand Panamá, qui apparaît ici intermédiaire et relativement proche de l'Afrique.

## Les données agromorphologiques

La figure 5 présente les deux premiers axes d'une analyse discriminante réalisée sur des données morphologiques et de production pour 17 cultivars Grands représentés chacun par 30 individus (N'CHO et al., 1993). Le premier

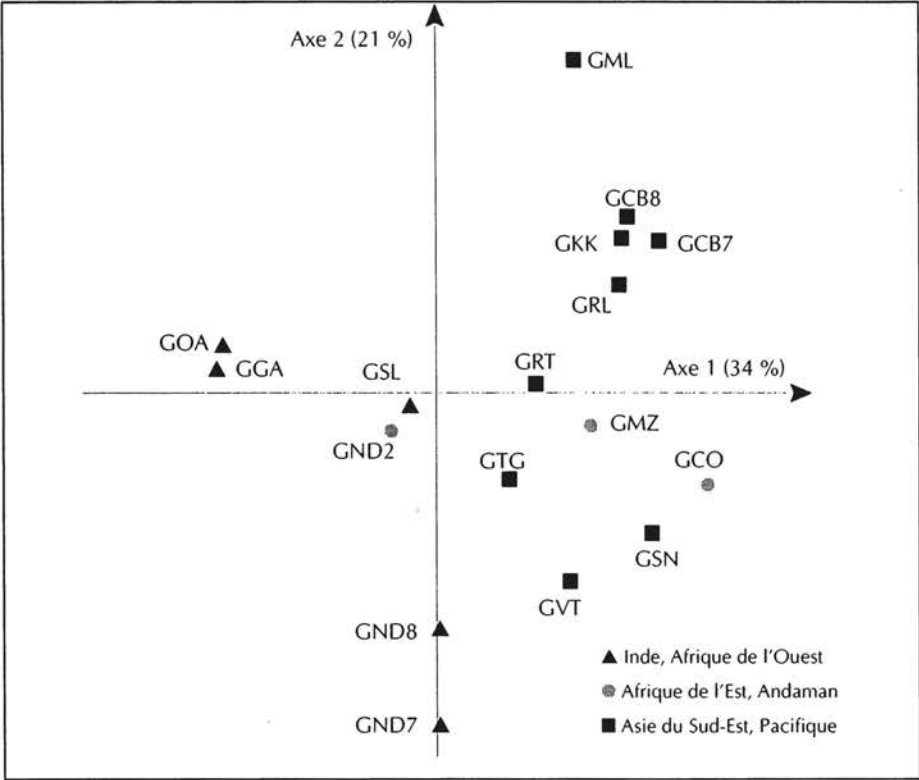


Figure 5. Synthèse des données morphologiques, d'après N'CHO et al. (1993).

axe représente 34 % de la variabilité totale et comporte des variables liées à la vigueur générale de la plante et, plus particulièrement, à celle du stipe. Le second (21 % de la variabilité) est positivement lié à la taille du fruit et de la cavité laissée par l'albumen (correspondant au volume d'eau de coco dans la noix presque mûre). Il est également associé négativement au nombre de noix produites.

Suivant cette représentation, les cultivars d'Inde et d'Afrique de l'Ouest sont caractérisés par un port relativement grêle et par de nombreux fruits de petite taille, renfermant un faible volume d'eau, ceux d'Afrique de l'Est et des îles mélanésiennes se distinguent par un port plus robuste, avec des fruits de petite dimension (sauf le Grand Rennell), alors que ceux d'Asie du Sud-Est et de Papouasie associent robustesse et taille des noix. Les cultivars de Polynésie occupent une position médiane. Un troisième axe non représenté (13 % de la variabilité), associé à la longueur des éléments de l'inflorescence, isole le seul Grand Rennell.

On retrouve comme pour les marqueurs polyphénoliques et RFLP une zonation est-ouest, quoique moins évidente. La position intermédiaire des cultivars d'Afrique de l'Est n'est pas observée.

## Discussion et conclusion

### La contribution des marqueurs moléculaires à l'histoire de la diversification du cocotier

Les analyses moléculaires présentées ici sont donc globalement en accord avec les résultats des études réalisées précédemment. Les principales différences tiennent à la meilleure reproductibilité des résultats et à l'indépendance du polymorphisme des RFLP à l'égard de l'environnement, ce qui se traduit par une plus grande précision. Les marqueurs RFLP contribuent ainsi à une meilleure connaissance des principaux événements qui ont marqué la diversification du cocotier (LEBRUN *et al.*, 1998b). Afin de faire ressortir les contributions spécifiques des marqueurs moléculaires, nous les replacerons tout au long d'une brève histoire du cocotier (figure 6).

La région d'origine du cocotier la plus probable se situe entre l'Asie du Sud-Est et la Papouasie (ZOHARY, 1970 ; CHILD, 1964). La noix de coco est adaptée à la dispersion par flottaison, et la répartition des îles lui permet d'atteindre facilement une côte favorable à sa germination. Il existe sans doute, dès le début, une variation notable de la composition du fruit. Peu à peu, le cocotier aurait migré spontanément par flottaison vers le Pacifique et, éventuellement, de proche en proche vers l'Inde. Dans ces migrations lointaines, c'est sans doute la forme la plus adaptée au transport marin à longue distance qui a été favorisée, le type Niu Kafa de HARRIES (1978).

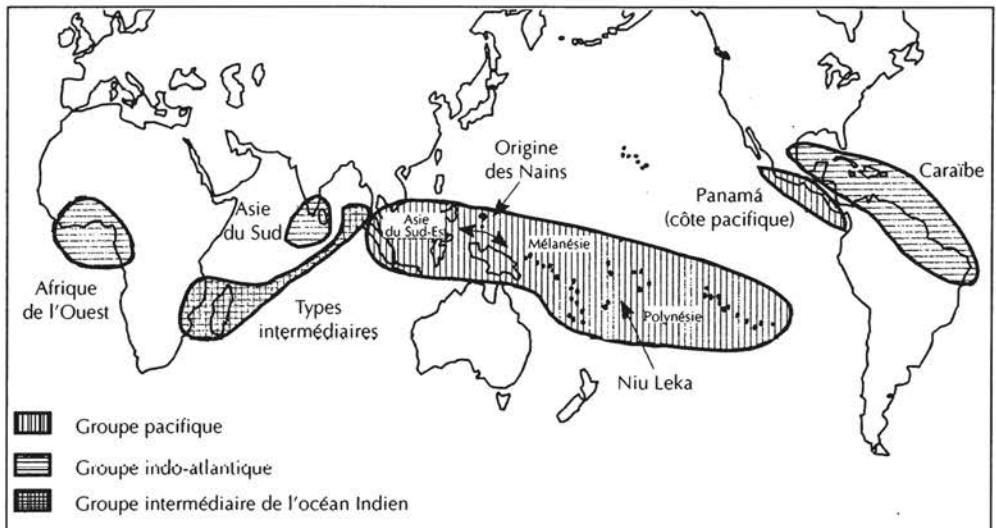


Figure 6. Localisation géographique des grands groupes de cocotier.

A une date très reculée et inconnue, les populations typiques du sous-continent indien se sont trouvées isolées du groupe pacifique. Bien qu'il ne soit pas considéré comme une plante indigène, le cocotier serait présent en Inde depuis au moins 3 000 ans (CHILD, 1964). Le fait qu'on retrouve un type Niu Kafa prédominant en Inde pourrait suggérer un transport par flottaison, sans qu'on puisse le prouver. Les apports ultérieurs de l'Asie du Sud-Est ont été limités et n'ont que peu modifié la structure génétique.

□ La répartition des cultivars en deux grands ensembles n'est pas une nouveauté. En revanche, de nombreux marqueurs RFLP présentent un fort déséquilibre de fréquence, qui peut aller jusqu'à la spécificité, entre les groupes pacifique et indo-atlantique. Cela indique une divergence ancienne probablement due à un isolement géographique prolongé.

À partir de 1300 avant notre ère, la colonisation humaine du Pacifique sud consolide l'implantation du cocotier. La bourre fournit des cordes nécessaires à la construction d'embarcations souples et résistantes. Les fruits immatures (eau) et matures (albumen) font partie des vivres. Le mouvement est progressif et ne s'est jamais interrompu. Les deux formes du fruit, Niu Vai et Niu Kafa, coexistent et sont probablement sélectionnées pour leurs qualités propres. Cette sélection se fait à partir d'un même stock initial polymorphe, dont le fond génétique est similaire à celui de l'Asie du Sud-Est et de la Papouasie.

□ Les populations du groupe pacifique possèdent les mêmes marqueurs, et les différences entre les sous-régions portent surtout sur des différences de fréquences.

À partir de 600 avant notre ère, des migrations malayo-polynésiennes se sont produites d'Asie du Sud-Est vers Madagascar. Les immigrants importent le cocotier avec leurs coutumes et leur langue, y compris le nom de la plante (*voanio*). De là, le cocotier atteint l'Afrique de l'Est et les Comores.

Depuis 800 ou un peu avant, les Arabes commercent entre l'Inde et l'Afrique de l'Est (IBN BATTUTA, 1351). Ils transportent des cocotiers typiques de ces régions. Des populations à structure génétique intermédiaire se constituent en Afrique de l'Est par hybridation entre les cocotiers importés et locaux.

□ Le groupe intermédiaire de l'océan Indien cumule dans son génome les marqueurs qui caractérisent les groupes indo-atlantique et pacifique. Sa répartition en deux zones géographiques distinctes et les informations historiques permettent de considérer ce groupe intermédiaire comme le résultat d'une fusion entre des populations issues de ces deux groupes.

À une période indéterminée, des flux en provenance de l'Inde se sont établis vers l'Asie du Sud-Est (peut-être par l'intermédiaire des voyageurs arabes). Les populations des Andaman, îles indiennes à l'ouest de Sumatra, ont de ce fait une structure génétique proche de celle de l'Afrique de l'Est. Plus loin vers l'est, de l'Indonésie aux Philippines, des échanges très limités ont pu se pro-

duire. Un échange en sens inverse serait à l'origine du Grand Kappadam, qui se distingue des populations de la même région par ses fruits volumineux et ronds.

□ Trois marqueurs caractéristiques des populations indo-atlantiques sont observés à faible fréquence en Asie du Sud-Est, mais absents plus à l'est. Ils témoignent d'un faible apport génétique d'origine indienne.

□ Le Grand Kappadam présente un marqueur cytoplasmique exotique et, pour les marqueurs nucléaires, quelques traces d'introgression en provenance de l'est, sur un fond génétique typiquement indo-atlantique. La réunion de ces particularités pourrait s'expliquer de la façon suivante : ce cultivar serait issu d'une introduction originaire d'Asie du Sud-Est, suivie par une sélection des arbres mères présentant des fruits ronds pendant plusieurs générations successives. La pollinisation libre, assurée majoritairement par des populations locales, lui aurait conféré à la longue un génotype nucléaire proche de celui du groupe indo-atlantique.

Vers 1498, à la suite de Vasco de Gama, les Portugais contournent l'Afrique et atteignent l'Inde par voie marine. Le cocotier est du voyage de retour et s'installe en Afrique de l'Ouest.

Entre 1525 et 1550, directement ou peut-être à partir d'Afrique de l'Ouest (Cap-Vert), il atteint la Caraïbe et la côte atlantique de l'Amérique. Des arguments historiques forts et la ressemblance morphologique entre les cocotiers du Brésil, de la zone caraïbe (non inclus dans cette étude) et de l'Afrique permettent de considérer qu'ils appartiennent au même groupe.

□ La similitude des profils RFLP des cultivars ouest-africains et indiens apparaît comme une confirmation des résultats obtenus par d'autres méthodes.

À l'époque précolombienne, le cocotier est déjà présent sur la côte pacifique du Panamá, comme le relate Oviedo en 1535 dans *La historia general y natural de las Indias, Islas y Tierra Firme del Mar Océano* (cité par ZIZUMBO et QUERO, 1997), mais absent du Mexique. Les populations actuelles pourraient provenir de ces cocotiers. La distance entre la côte latino-américaine et les îles du Pacifique, ainsi que le caractère typiquement Niu Vai du matériel local semble exclure un transport par flottaison. En revanche, le transport humain semble possible, soit par les Polynésiens, soit par des populations d'Asie du Sud-Est, dont des traces archéologiques auraient été trouvées jusqu'en Equateur (LANGDON, 1995).

Au XVI<sup>e</sup> siècle, des importations en provenance du Panamá, des Philippines et des îles Salomon sont signalées au Mexique.

□ Les particularités des cocotiers du Panamá sont mises en évidence par les RFLP. Leur origine se trouve de toute évidence dans le groupe pacifique, sans qu'on puisse pour l'instant trancher entre les routes polynésienne et du Sud-Est asiatique. Leur grande homogénéité et leur homozygotie témoignent d'un effet fondateur.

□ Au sein de l'origine Panamá, la population Aguadulce possède, à une faible fréquence, des allèles spécifiques du groupe indo-atlantique, ce qui est la signature très probable d'une introgression de gènes venant de populations de la côte atlantique.

A une date difficile à préciser, mais après la séparation des stocks pacifique et indo-atlantique, les Nains apparaissent entre le Sud-Est asiatique et la Papouasie, vraisemblablement aux Philippines, eu égard au nombre et à la diversité des Nains qu'on y trouve. La différenciation ultérieure s'opère sur une base régionale au cours de la diffusion de ce matériel, probablement par intercroisement avec des Grands locaux.

□ Des indices sérieux, liés au nombre d'allèles qu'ils ont fixé, sont apportés par les RFLP en faveur d'une origine unique des Nains, exception faite du Niu Leka. Les cultivars Grands dont la composition allélique est la plus compatible avec celle des Nains se situent dans une zone allant de l'Asie du Sud-Est à la Papouasie.

A une date indéterminée, le Niu Leka, cocotier Nain allogame, apparaît dans la région des Fidji, des Samoa et de Tonga.

□ La ressemblance entre le profil RFLP du Niu Leka et celui des Grands Tonga et Rotuma, originaires de la même région, donne des arguments en faveur d'une origine locale — et donc indépendante de celle des autres Nains — pour ce cultivar.

## Une approche de la relation entre distance génétique et hétérosis

On observe chez de nombreuses plantes que les croisements réalisés entre des populations génétiquement éloignées sont supérieurs à ceux qui proviennent d'une base génétique étroite. Cette « hétérosis interpopulation » est largement exploitée dans les schémas de sélection récurrente réciproque. Il convient de vérifier si un tel phénomène s'observe chez le cocotier. Le tableau 1 présente les résultats de tests d'hybrides Grand  $\times$  Grand en fonction de l'origine des parents. Les populations de l'océan Indien intermédiaires entre pacifiques et indo-atlantiques sont représentées par le Grand Mozambique.

Dans cinq essais comparant des hybrides Grand  $\times$  Grand, tous les hybrides sont supérieurs au témoin Grand Ouest Africain. Parmi eux, les croisements entre cultivars du même groupe ont les rendements les plus bas, tandis que les croisements entre les cultivars indo-atlantiques et ceux du groupe pacifique ont généralement la meilleure production. Le groupe de l'océan Indien est représenté par le Grand du Mozambique. Ses croisements avec des cultivars des deux groupes précédents ont des performances intermédiaires entre celles des croisements intragroupes et celles des croisements intergroupes.



Tableau 1. Production de coprah de quelques hybrides Grand × Grand classés selon les groupes définis par les RFLP : résultats de cinq essais.

Essai	PBGC1	PBGC 3	PBGC 7	PBGC 8	PBGC 9
Année de plantation	1965	1969-1970	1971	1972-1973	1971
Pacifique × indo-atlantique	(2)* 130 %	(2) 182 %	(2) 174 %	(3) 212 %	(3) 182 %
Océan Indien × pacifique	(2) 124 %	(2) 151 %	(2) 153 %		
Océan Indien × indo-atlantique	(1) 136 %				
Pacifique × pacifique		(2) 138 %	(1) 137 %	(3) 142 %	
Indo-atlantique × indo-atlantique					(3) 142 %
Témoin GOA (%)	(1) 100 %	(1) 100 %	(1) 100 %	(1) 100 %	(1) 100 %
Production (kg/arbre/an)	23,0	15,8	11,5	8,2	11,5
Témoin autofécondé	(1) 88 %				

\* Pour chaque essai, le nombre de croisements par type de recombinaison est donné entre parenthèses.

En revanche, bien que les données moléculaires suggèrent fortement que les Nains trouvent leur origine dans le groupe pacifique, ceux-ci peuvent donner des hybrides excellents avec des partenaires de l'un ou l'autre des deux principaux groupes définis plus haut. C'est par exemple le cas des hybrides Nain Jaune Malais × Grand Ouest Africain (PB121) et Nain Rouge Malais × Grand Rennell. Quels que soient les facteurs responsables de l'« hétérosis interpopulations » évoquée plus haut, il semble que le régime de reproduction autogame qui prévaut chez les Nains ait induit une divergence génétique — due à la fixation d'un allèle par locus, plutôt qu'à l'apparition d'allèles nouveaux — suffisamment importante par rapport à leur groupe d'origine pour que cette hétérosis ait lieu aussi bien avec celui-ci qu'avec les indo-africains.

## Conclusion

Les résultats obtenus à l'aide des marqueurs moléculaires concordent largement avec les résultats trouvés précédemment par d'autres méthodes. Ils apportent cependant une plus grande précision dans l'étude des relations génétiques entre populations. Les études portant sur les critères agronomiques et morphologiques sont indispensables pour la caractérisation de la variabilité d'une espèce comme le cocotier, dans la mesure où elles fournissent des éléments essentiels pour l'utilisation des cultivars étudiés en amélioration. L'image de l'organisation génétique des populations à partir de ces seuls critères est cependant moins précise que ce que permettent les marqueurs moléculaires : d'une part, les effets de l'environnement biaisent les comparaisons lorsque les études sont réalisées dans des conditions différentes de lieux ou de périodes ; d'autre part, la sélection naturelle ou humaine a pu dans certains cas aboutir à des phénotypes semblables à partir de populations d'origine distincte. La similitude des phénotypes peut masquer de réelles complémentarités génétiques. Enfin, l'expression de la valeur génétique peut être affectée par le degré plus ou moins élevé de consanguinité des populations.



Les marqueurs polyphénoliques se sont révélés très performants dans une première étude. Cependant, leur répétabilité est faible et les résultats obtenus n'ont pu être reproduits dans des environnements différents. Chez les plantes, les polyphénols sont impliqués dans les réactions à différents stress. On peut donc s'attendre à ce que les profils obtenus varient en fonction des stimulus extérieurs. Enfin, leur implication dans la défense contre les stress suggère qu'ils pourraient ne pas être neutres à l'égard de la sélection. A l'inverse, les marqueurs RFLP sont choisis indépendamment de l'activité de la séquence concernée. Bien que difficile à prouver en toute rigueur, l'hypothèse de neutralité est plus facilement soutenable.

Les marqueurs moléculaires offrent de multiples avantages pour l'amélioration du cocotier. Ce sont des outils précieux pour gérer les collections. Chez une plante pérenne comme le cocotier, il est important de rassembler sur une surface limitée la plus large variabilité génétique possible. Dans certaines situations, le polymorphisme des RFLP permet d'identifier très précisément les variétés. Il est même possible, dans certaines populations, de détecter la présence de gènes d'origine étrangère et d'en expliquer la cause probable. Compte tenu des effectifs restreints utilisés, cette technique s'est révélée très puissante.

En outre, la répartition du stock génétique de l'espèce en deux grands groupes géographiques ayant connu un isolement génétique prolongé apporte une base solide au choix d'un dispositif de sélection récurrente réciproque pour la production d'hybrides Grand  $\times$  Grand. Elle permet d'assigner *a priori* une place à la plupart des cultivars dans ce schéma, afin de maximiser l'hétérosis. Les quelques cultivars intermédiaires situés en Afrique de l'Est et aux confins de l'Asie du Sud et de l'Extrême-Orient peuvent être utilisés pour enrichir la variabilité de l'un des deux groupes, notamment les indo-atlantiques, dont la base génétique et la variabilité phénotypique est la plus restreinte. Au sein de chacun des groupes hétérotiques, il est possible de choisir les cultivars à recombiner en priorité afin de maximiser la variabilité sélectionnable.

Enfin, les marqueurs moléculaires fournissent dans certains cas des indices précis sur l'histoire du matériel végétal, ce qui peut être utile pour rechercher de nouvelles sources de caractères particuliers. L'étude des populations de la côte pacifique d'Amérique latine en relation avec la tolérance au *lethal yellowing* dans la zone caraïbe illustre ce domaine d'application. Les populations du Panamá semblent en effet présenter des facteurs de tolérance à cette maladie. Le profil RFLP de trois de ces populations suggère qu'elles remontent à une population initiale de petite taille. Ainsi, l'étude, à partir d'un échantillonnage plus large, de la diversité et de la dynamique des populations chez les Grands de la côte du Pacifique permettrait de caractériser plus précisément la nature des apports initiaux et d'identifier avec une plus grande certitude des populations susceptibles de transmettre une tolérance à la maladie. Cet élargissement de la recherche peut se révéler capital dans la mesure où le Grand Panamá est sensible au *Phytophthora* qui sévit dans la même région.

# Annexe

## Matériel végétal

Les échantillons foliaires prélevés sur 289 arbres représentent 26 cultivars Grands et 16 cultivars Nains provenant des collections de Côte d'Ivoire (station Marc Delorme), du Vanuatu (station de Saraoutu) et de Jamaïque. Ils recouvrent la plupart des zones de culture du cocotier, à l'exception, pour les écotypes Grands, de la zone caraïbe et de la façade est de l'Amérique. La liste des cultivars, avec leur origine géographique, est donnée dans le tableau 2.

## Analyses par RFLP

Les extraits d'ADN total sont réalisés à partir de folioles lyophilisées, prélevées sur la feuille n° 1 (plus jeune feuille chlorophyllienne). La méthode utilisée est celle au CTAB (cétyle-triméthyl ammonium bromide), adaptée sur le maïs par HOISINGTON (1992). Les restrictions sont effectuées par quatre enzymes : *EcoR1*, *EcoRV*, *Bgl2* et *Sst1*. L'ADN restreint migre dans un gel d'agarose à 0,8 % en tampon TAE (tris-acétate EDTA), puis est transféré sur une membrane de Nylon. Les sondes utilisées pour l'hybridation moléculaire suivant le protocole de HOISINGTON (1992) sont marquées au <sup>32</sup>P. Les résultats sont lus à partir d'autoradiogrammes.

## Origine des sondes

Cette étude a été réalisée grâce à 20 sondes d'ADNc et une sonde mitochondriale (Cox 1). Parmi les sondes nucléaires se trouve un ADNc de cocotier, les autres étant des sondes hétérologues de riz, de palmier à huile et de maïs, dont l'origine est citée dans LEBRUN *et al.* (1998b).

## Analyse des données

Chaque bande est codée comme un marqueur dominant : 10 pour sa présence et 01 pour son absence. La matrice binaire bandes × individus ainsi obtenue nous a permis de réaliser des analyses factorielles des correspondances (BENZECRI, 1973) à l'aide du logiciel Addad (ADDAD, 1983). Les représentations graphiques de ces analyses multivariées nous ont permis de décrire la structuration génétique du matériel.

Tableau 2. Effectif et répartition géographique des cultivars prélevés.

Origine	Grands	Code	Effectif*	Nains	Code	Effectif*
Afrique de l'Ouest	4 cultivars			2 cultivars		
• Côte d'Ivoire	Grand Ouest Africain	GOA	5			
	Grand Ouest Africain Mensah	GOA04	10			
• Bénin	Grand Ouest Africain Ouidah	GOA06	10			
• Cameroun	Grand Cameroun Kribi	GCA	5	Nain Rouge Cameroun	NRC	5
• Ghana				Nain Jaune Ghana	NJG	7
Afrique de l'Est	2 cultivars					
• Comores	Grand des Comores Moheli	GCO	5			
• Mozambique	Grand du Mozambique	GMZ	5			
Asie du Sud	4 cultivars			1 cultivar		
• Inde	Grand Laccadives Micro	GND07	5			
	Grand Kappadam	GND05	5			
	Grand Andaman Ordinaire	GND02	4			
• Sri Lanka	Grand Sri Lanka	GSL	5	Nain Vert Sri Lanka	NVS	5
Asie du Sud-Est	8 cultivars			7 cultivars		
• Thaïlande	Grand de Thaïlande	GTH	5			
• Philippines	Grand Baybay	GPH04	5	Nain Vert Catigan	NVP02	5
	Grand Tagnanan	GTN	5	Nain Vert Pilipog	NVP05	5
				Nain Vert Tacunan	NVP03	5
• Cambodge	Grand du Cambodge	GCB	10			
• Indonésie	Grand Tenga	GDO02	5	Nain Brun Ternate	NBO	5
	Grand Palu	GDO03	5			
	Grand Takome	GDO04	5			

Tableau 2. Suite.

Origine	Grands	Code	Effectif*	Nains	Code	Effectif*
• Malaisie	Grand de Malaisie	GML	11	Nain Jaune Malais Nain Vert Malais Nain Rouge Malais	NJM NVM NRM	15 5 5 + 5**
Pacifique sud	9 cultivars			5 cultivars		
• Papouasie- Nouvelle-Guinée	Grand Karkar Grand Markham Valley Grand Gazelle	GNG01 GNG03 GNG04	5 5 5 + 5**	Nain Brun Madang	NBN	5
• Iles Salomon	Grand Rennell Grand Salomon	GRL GSL	7 + 5** 6			
• Polynésie française	Grand de Polynésie Rangiroa	GPY01	5	Nain Rouge Polynesie	NRY	5
• Fidji	Grand Rotuma	GRT	5	Nain Niu Leka	NNL	7
• Tonga	Grand Tonga	GTG	5			
• Vanuatu	Grand du Vanuatu	GVT	5	Nain Rouge du Vanuatu	NRV	5**
• Kiribati				Nain Vert Kiribati	NVK	5**
Amérique latine	3 cultivars			1 cultivar		
• Panamá	Grand Panamá Grand Panamá (Aguadulce) Grand Panamá (Monagre)	GPA GPA01 GPA02	10*** 6 6			
• Brésil				Nain Vert Brésil	NVB	5

\* Tous les échantillons proviennent de Côte d'Ivoire, sauf mention contraire : \*\* prélevé au Vanuatu ; \*\*\* prélevé en Jamaïque.

## Références bibliographiques

- ADDAD, 1983. Manuel de référence. Paris, France, Association pour le développement et la diffusion de l'analyse des données.
- ASHBURNER G.R., THOMPSON W.K., HALLORAN G.M., 1997. RAPD analysis of South Pacific coconut palm populations. *Crop Science*, 37 : 992-997.
- BENOIT H., GHESQUIERE M., 1984. Electrophorèse, compte rendu cocotier. IV. Déterminisme génétique. Montpellier, France, Cirad-Irho, 11 p. (document interne).
- BENZECRI J.P., 1973. L'analyse des données. Tome II. L'analyse des correspondances. Paris, France, Dunod, 616 p.
- BOURDEIX R., BAUDOUIN L., BILLOTTE N., LABOUISSÉ J.P., NOIRET J.M., 1997. Le cocotier. *In* : L'amélioration des plantes tropicales, A. Charrier *et al.* éd., Montpellier, France, Cirad-Orstom, p. 217-239.
- BOURDEIX R., MEUNIER J., N'CHO Y.P., 1991a. Une stratégie de sélection du cocotier *Cocos nucifera* L. 2. Amélioration des hybrides Grand × Grand. *Oléagineux*, 46(7) : 267-282.
- BOURDEIX R., MEUNIER J., N'CHO Y.P., 1991b. Une stratégie de sélection du cocotier *Cocos nucifera* L. 3. Amélioration des hybrides Nain × Grand. *Oléagineux*, 46(10) : 361-374.
- BOURDEIX R., N'CHO Y.P., LE SAINT J.P., 1990. Une stratégie de sélection du cocotier. 1. Synthèse des acquis. *Oléagineux*, 45(8-9) : 359-371.
- CHILD R., 1964. Coconuts (2nd edition). Londres, Royaume-Uni, Longman, 335 p.
- COMSTOCK R.E., ROBINSON H.F., HARVEY P.H., 1949. A breeding procedure to make maximum use of both general and specific combining ability. *Agronomy Journal*, 41 : 360-367.
- FERNANDO W.M.U., GAMINI G., 1997. Profil des variations isoenzymatiques chez les populations de cocotier (*Cocos nucifera* L.) utilisées pour la sélection des variétés améliorées. *Plantations, recherche, développement*, 4(4) : 256-263.
- HARRIES H.C., 1978. The evolution, dissemination and classification of *Cocos nucifera* L. *The Botanical Review*, 44 : 265-320.
- HARTANA A., HENGKY, DWI ASMONO, 1993. Analisis Keragaman dan pewarisan pola pita isozim tanaman kelapa. *Journal Matematika dan Sains*, 1(suplement D) : 63-76.
- HOISINGTON D., 1992. Laboratory protocols. Mexico, Mexique, CIMMY. CIMMYT Applied Molecular Genetics Laboratory.
- IBN BATTUTA, 1351. Voyages et périples. *In* : Les voyageurs arabes. Paris, France, Gallimard, collection La Pléiade, 1408 p.
- JAY M., BOURDEIX R., POTIER F., SANLAVILLE C., 1989. Premiers résultats de l'étude des polyphénols foliaires du cocotier. *Oléagineux*, 44(3) : 151-161.
- LANGDON R., 1995. The banana as a key to early American and Polynesian history. *Journal of Pacific History*, 28(1) : 15-35.

- LEBRUN P., GRIVET L., BAUDOIN L., 1998b. Dissémination et domestication du cocotier à la lumière des marqueurs RFLP. *Plantations, recherche, développement*, 5(4) : 233-245.
- LEBRUN P., N'CHO Y.P., SEGUIN M., GRIVET L., BAUDOIN L., 1998a. Genetic diversity in coconut (*Cocos nucifera* L.) revealed by restriction fragment length polymorphism (RFLP) markers. *Euphytica*, 101 : 103-108.
- MARECHAL H., 1928. Observation and preliminary experiments on the coconut palm with a view to developing improved seed for Fiji. *Fiji Agricultural Journal*, 1 : 16-45.
- N'CHO Y.P., SANGARE A., BOURDEIX R., BONNOT F., BAUDOIN L., 1993. Evaluation de quelques écotypes de cocotier par une approche biométrique. 1. Etude des populations de Grands. *Oléagineux*, 48(3) : 121-132.
- DE NUCE DE LAMOTHE M., BENARD G., 1985. L'hybride de cocotier PB121 (ou Mawa) (NJM x GOA). *Oléagineux*, 40(5) : 261-266.
- DE NUCE DE LAMOTHE M., ROGNON F., 1972. La production de semences hybrides chez le cocotier par pollinisation assistée. *Oléagineux*, 27(10) : 539-544.
- PATEL J.S., 1938. The coconut: a monograph. Madras, Inde, Government Press, 350 p.
- PERERA L., RUSSEL J.R., PROVAN J., MCNICOL J.W., POWELL W., 1998. Evaluating genetic relationships between indigenous coconut (*Cocos nucifera* L.) accessions from Sri Lanka by means of AFLP profiling. *Theoretical and Applied Genetics*, 96 : 545-550.
- PERSLEY G.J., 1992. Replanting the tree of life. Wallingford, Royaume-Uni, CAB International, 156 p.
- ROGNON F., 1976. Biologie florale du cocotier. *Oléagineux*, 31(1) : 13-18.
- ROHDE W., KULLAYA A., RODRIGUEZ J., RITTER E., 1995. Genome analysis of *Cocos nucifera* L. by PCR amplification of spacer sequences separating a subset of *copia*-like *EcoRI* repetitive elements. *Journal of Genetics and Breeding*, 49 : 179-186.
- THAMMES P.L.M., 1955. Review of coconut selection in Indonesia. *Euphytica*, 4 : 17-24.
- ZIZUMBO V.D., QUERO H.J., 1997. Re-evaluation of early observations on coconut in the New World. *Economic Botany*, 52(1) : 68-77.
- ZOHARY D., 1970. Centers of diversity and centers of origin. *In* : Genetic resources in plants, their exploration and conservation, O.H. Frankel et E. Bennett éd., Oxford, Royaume-Uni, Blackwell, 547 p.



# L'hévéa

---

Marc Seguin, Albert Flori,  
Hyacinthe Legnaté, André Clément-Demange

La répartition taxonomique des plantes à latex dans le règne végétal est très dispersée. Parmi les 12 500 espèces à latex, qui appartiennent à 900 genres, 1 000 espèces, réparties dans 76 familles, produisent du caoutchouc, ou cis-polyisoprène (POLHAMUS, 1962). Au sein du genre *Hevea*, qui offre des conditions favorables à l'exploitation, l'espèce *Hevea brasiliensis* est la source quasi exclusive du caoutchouc naturel en raison de sa productivité et de la qualité de son caoutchouc.

Le caoutchouc naturel, sous forme de latex ou coagulé, est collecté et transporté vers une usine de séchage, qui produit des feuilles fumées ou des balles de caoutchouc granulé sec. Il est le plus souvent exporté vers les zones de transformation finale proches des marchés de consommation. La production mondiale s'élève à près de 6 millions de tonnes par an et provient de la zone intertropicale, où la culture couvre environ 8 millions d'hectares. Elle est assurée à plus de 90 % par l'Asie du Sud-Est et dominée par trois pays : la Thaïlande, l'Indonésie et la Malaisie. L'Afrique réalise 6 % de la production. En Amérique du Sud, le développement de la culture est limité par la maladie sud-américaine des feuilles, provoquée par le champignon *Microcyclus ulei*.

Le caoutchouc naturel représente le tiers de la production totale de caoutchouc. Il possède, par rapport au caoutchouc synthétique issu de la pétrochimie, des qualités spécifiques, telles que la résistance à l'échauffement.



L'industrie des pneus d'avion, par exemple, utilise uniquement du caoutchouc naturel, et l'industrie des pneumatiques en général consomme 70 % de la production, les 30 % restants étant employés dans une multitude de produits.

## La botanique et les ressources génétiques

### La botanique et la taxonomie

L'hévéa est un arbre à croissance rythmique et à ramification orthotrope, qui peut dépasser 30 mètres de hauteur et atteindre 3 mètres de circonférence dans son milieu naturel, le bassin amazonien. En plantation, les arbres greffés n'atteignent pas ces dimensions. C'est une essence tropicale de lumière, qui nécessite ensoleillement et humidité et tolère mal les altitudes supérieures à 600 mètres. Les feuilles comportent trois folioles de forme ovoïde, disposées à l'extrémité d'un long pétiole ; elles sont renouvelées chaque année par une défoliation-refoliation naturelle intervenant en saison sèche. C'est une plante monoïque préférentiellement allogame, dont les fleurs unisexuées, de couleur jaune, ont une dimension de quelques millimètres et sont groupées en racèmes. Le fruit de l'hévéa, dont la taille varie de 0,5 à 5 centimètres selon les espèces, est caractéristique du fruit des euphorbiacées, appelé tricoque, composé de trois loges contenant chacune une graine.

Au sein du genre *Hevea* (Willd.), bien défini, les espèces sont difficiles à différencier, ce qui a été à l'origine de beaucoup de confusions, de désaccords et de fluctuations entre les auteurs. Hormis pour l'espèce *H. brasiliensis*, les observations ont été réalisées dans des conditions souvent difficiles sur le terrain et sont généralement assez sommaires. SCHULTES (1990) présente la synthèse la plus récente et distingue dix espèces, dont trois sont subdivisées en quatre variétés. On ne connaît pas de barrière biologique à la recombinaison entre ces espèces. Les hybridations naturelles paraissent limitées en milieu forestier non perturbé, mais de nombreuses variations morphologiques intraspécifiques sont observées. Il conviendrait donc de parler d'un complexe d'espèces selon la définition de Pernès (PERNÈS, 1984) ; les neuf espèces sauvages apparentées faisant partie du pool génique primaire de l'espèce cultivée *H. brasiliensis*. Elles sont donc d'un grand intérêt pour les ressources génétiques de l'hévéa cultivé, d'autant qu'elles présentent des caractéristiques utiles telles que des résistances génétiques à certaines maladies (SCHULTES, 1977 ; SCHULTES, 1990).

### La structure du génome

A l'exception d'un clone triploïde d'*H. guianensis* et de l'existence possible d'une race d'*H. pauciflora* à 18 chromosomes (BALDWIN, 1947), toutes les espèces du genre ont un nombre de chromosomes de  $2n = 36$ .

De fortes similitudes entre différentes paires de chromosomes suggèrent une duplication ancienne du stock chromosomique. A la méiose, on observe de temps en temps la formation de multivalents, et notamment de quadrivalents, ce qui indique un certain degré d'affinité entre paires de chromosomes homéologues. L'hévéa serait donc un genre amphidiploïde, c'est-à-dire un tétraploïde qui se comporte à la méiose comme un diploïde, résultant du croisement de deux espèces sauvages ancestrales diploïdes non identifiées (BOUHARMONT, 1960 ; ONG, 1985). Le nombre chromosomique de base serait donc égal à 9, ce qui semble plus proche du nombre chromosomique habituel des euphorbiacées, qui varie de 6 à 11 (ONG, 1985).

Cependant, l'analyse des ségrégations de marqueurs isoenzymatiques ou RFLP révèle une majorité de locus non dupliqués (SEGUIN *et al.*, 1996). Ainsi, l'hévéa pourrait être un amphidiploïde, dont les deux génomes homéologues auraient fortement divergé. Il n'est pas possible d'affirmer que les rares duplications observées dans le génome de l'hévéa (SEGUIN *et al.*, 1998) témoignent de l'origine polyploïde de l'hévéa. Ces duplications pourraient aussi provenir de remaniements chromosomiques postérieurs à la polyploïdisation. La même situation s'observe chez le manioc, une autre euphorbiacée à  $2n = 36$  chromosomes (FREGENE *et al.*, 1997).

Une première approche de la caractérisation globale du génome nucléaire d'*H. brasiliensis* a été proposée (LOW et BONNER, 1985). L'étude de la cinétique de réassociation de courts fragments d'ADN marqué (300 nucléotides) indique la présence de 43 % d'ADN simple copie et de 32 % de séquences moyennement répétées (fréquence de répétition d'environ 1 000), le reste de l'ADN étant hautement répété ou palindromique. La cinétique de réassociation d'ADN simple copie permet d'estimer la taille globale du génome à  $6 \cdot 10^8$  paires de bases. Cependant, une étude ponctuelle que nous avons menée par cytométrie en flux a conduit à une estimation de la taille du génome haploïde de l'hévéa de  $2 \cdot 10^9$  paires de bases, identique chez les cinq espèces étudiées, *H. brasiliensis*, *H. benthamiana*, *H. guianensis*, *H. pauciflora* et *H. spruceana*. Cette valeur a également été obtenue par microdensitométrie chez *H. brasiliensis* et *H. camargoana* par BENNETT et LEITCH (1997).

## Les ressources génétiques

### LES PROSPECTIONS, LES INTRODUCTIONS ET LES COLLECTIONS

En 1747, Fresneau donne la première description de l'arbre « caoutchouc », qui sera plus tard nommé hévéa. D'abord produit de cueillette récolté par les *seringueiros* en Amérique du Sud, le caoutchouc naturel connaît son essor après la mise au point de la mastication et de l'imperméabilisation, puis la découverte de la vulcanisation en 1830 et le développement des plantations (SERIER, 1993).

En 1876, sur l'insistance de Marckham, qui avait déjà réussi le transfert du quinquina d'Amérique du Sud en Asie, Wickham, planteur britannique installé au Brésil, fait recueillir 70 000 graines près de Boim, à l'embouchure de l'Amazone et du Tapajós, en amont de Santarém (Etat du Pará, d'où l'appellation initiale d'arbre du Pará pour l'espèce *H. brasiliensis*). Il les convoie par bateau vers le jardin botanique de Kew, au Royaume-Uni, où seulement 4 % des graines germent après trois semaines de traversée. Plusieurs envois de plantules sont ensuite réalisés vers Ceylan et Singapour avec des succès divers. On considère que les 22 arbres plantés au jardin botanique de Singapour en 1877 sont ainsi à l'origine de la presque totalité des plantations d'hévéa dans le monde. Les variétés (clones de greffe) issues de sélection en Asie à partir de ces plants introduits à la fin du XIX<sup>e</sup> siècle sont appelés de ce fait clones Wickham. Même s'il est probable que d'autres introductions de génotypes ont eu lieu par la suite, l'histoire n'a retenu que celle, exemplaire, de Wickham et l'on admet le rôle prépondérant de ce matériel dans la composition génétique des cultivars actuels.

Les clones Wickham se sont avérés très sensibles à *Microcyclus* en Amérique du Sud. D'importants efforts de sélection ont alors été déployés par les compagnies Ford et Firestone, au Brésil et au Guatemala, puis par les centres de recherche brésiliens afin d'obtenir des clones productifs et résistants (clones F, FB, FX, FDR, MDF, MDX, IAN, CNSAM). L'introgression chez *H. brasiliensis* de gènes de résistance totale provenant d'autres espèces du genre n'a pas été couronnée de succès, ce type de résistance étant facilement contourné par le pathogène. La stratégie actuelle vise à associer par croisements plusieurs composantes de la résistance partielle pour atteindre un niveau suffisant de résistance génétique durable à la maladie (RIVANO, 1992).

L'é étroitesse supposée de la base génétique du matériel Wickham introduit puis sélectionné en Asie et en Afrique, qu'illustre la forte sensibilité générale de ce matériel à *Microcyclus*, a conduit à organiser des prospections au sein de l'aire de répartition de l'espèce *H. brasiliensis*.

En Malaisie, le RRIM (Rubber Research Institute of Malaysia) a importé du Brésil, en 1951-1952 puis en 1966, des seedlings issus d'arbres sauvages de sept espèces du genre *Hevea*, ainsi que des clones sélectionnés par Ford (F, FB, FX) et par l'Instituto Agrônômico do Norte (IAN) au Brésil (ONG, 1987).

En 1974, l'Irca (Institut de recherches sur le caoutchouc, France), en coopération avec l'EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária, Brésil), a réalisé une première prospection en Amazonie, dans les Etats de l'Acre (22 clones) et du Rondônia (20 clones), sous la forme de bois de greffe collectés en forêt sur des arbres jugés exceptionnels. A cette prospection viennent s'ajouter 18 clones de la prospection Firestone dans le Madre de Dios (MDF) péruvien.

La prospection internationale organisée en 1981 par l'IRRDB (International Rubber Research and Development Board), pour l'espèce *H. brasiliensis*, dans 16 districts et 60 localités des Etats de l'Acre (LINS *et al.*, 1981), du Rondônia (GONÇALVES, 1981) et du Mato Grosso (DE PAIVA, 1981) au Brésil, a permis de mettre en collection 9 800 génotypes sauvages au centre asiatique de Malaisie (ONG *et al.*, 1995) et 2 500 génotypes au centre africain de Côte d'Ivoire (CHAPUSET *et al.*, 1995). La localisation géographique des districts prospectés est donnée dans le tableau 1. Cette prospection a collecté essentiellement des graines, mais aussi 130 génotypes issus de bois de greffe récoltés sur des arbres jugés exceptionnels, appelés ortets. Des échanges entre les deux centres et une distribution de matériel génétique aux autres pays membres de l'IRRDB ont été réalisés. La plupart des génotypes ont été greffés et sont conservés sous forme de clones en collection vivante.

Après 1945, Schultes avait rassemblé, entre autres sur les sites de Calima et de Palmira, d'importantes collections de matériel prospecté en Colombie. En 1985, avec l'accord du gouvernement colombien, l'Irca a transféré, d'abord à la Guadeloupe puis en Côte d'Ivoire, 341 génotypes « SCH », issus de ces deux sites (NICOLAS, 1985).

Tableau 1. Code et origine des principales populations d'*H. brasiliensis* en collection.

Code	Prospection	Etat et pays d'origine	Lieu d'origine (districts)	Bassin fluvial (affluents de l'Amazone)
AC/T	IRRDB 1981	Acre, Brésil	Tarauaca	rio Juruá
AC/F	IRRDB 1981	Acre, Brésil	Feijo	rio Juruá
AC/S	IRRDB 1981	Acre, Brésil	Sena Madureira	rio Purus
AC/B	IRRDB 1981	Acre, Brésil	Brasília	rio Purus
AC/X	IRRDB 1981	Acre, Brésil	Xapuri	rio Purus
RO/A	IRRDB 1981	Rondônia, Brésil	Ariquemes	rio Madeira
RO/C	IRRDB 1981	Rondônia, Brésil	Calama	rio Madeira
RO/CM	IRRDB 1981	Rondônia, Brésil	Costa Marques	rio Madeira
RO/J	IRRDB 1981	Rondônia, Brésil	Jarù	rio Madeira
RO/JP	IRRDB 1981	Rondônia, Brésil	Jiparana	rio Madeira
RO/OP	IRRDB 1981	Rondônia, Brésil	Ouro Preto	rio Madeira
RO/PB	IRRDB 1981	Rondônia, Brésil	Pimenta Bueno	rio Madeira
MT/A	IRRDB 1981	Mato Grosso, Brésil	Aracatuba	rio Tapajós
MT/C	IRRDB 1981	Mato Grosso, Brésil	Juruena	rio Tapajós
MT/IT	IRRDB 1981	Mato Grosso, Brésil	Itauba	rio Tapajós
MT/VB	IRRDB 1981	Mato Grosso, Brésil	Vila Bella	rio Madeira
MDF	Firestone	Madre de Dios, Pérou	?	rio Madeira
SCH	Schultes coll. Calima et Palmira	Colombie	?	?
W	Wickham	Pará, Brésil	Santarém	rio Tapajós

L'Irca a également reçu du Brésil et introduit en Côte d'Ivoire en 1986, 23 clones CNSAM provenant de la collection de l'EMBRAPA à Manaus et comprenant des génotypes *H. pauciflora* et les seuls *H. brasiliensis* prospectés dans l'Etat brésilien de l'Amazonas (GONÇALVES *et al.*, 1983).

En conclusion, on peut considérer que l'espèce *H. brasiliensis* est bien représentée dans les collections vivantes, en Malaisie et en Côte d'Ivoire, par plusieurs milliers de génotypes. Cependant, l'aire de répartition de cette espèce a été échantillonnée de façon très hétérogène : les Etats de l'Amazonas et du Pará au Brésil, qui en constituent la plus grande part, ont été très peu prospectés. Il serait également nécessaire de réaliser une prospection des autres espèces du genre afin de créer un centre de conservation et d'étude de l'ensemble du genre *Hevea*. Les observations disponibles sur les collections Schultes et CNSAM montrent qu'il existe des génotypes d'espèces autres qu'*H. brasiliensis* dans ces collections.

#### LA CONSERVATION, LA MULTIPLICATION ET L'ÉCHANGE DES ACCESSIONS

A l'état naturel, l'hévéa se reproduit par graines. La mise au point du greffage par van Helten en 1918 a permis de développer la sélection clonale et de conserver les génotypes sous forme de collections vivantes, durables et relativement peu coûteuses à créer et à entretenir en jardins à bois de greffe. La multiplication s'effectue en pépinière par greffage sur des porte-greffe issus de graines.

Le délai maximal de transport est d'une semaine pour le bois de greffe et de trois semaines pour les graines (les graines sont dites récalcitrantes : la durée du pouvoir germinatif est faible). Au stade expérimental, la conservation du pollen ne dépasse pas un mois et les quantités récoltables sont très faibles ; on ne peut donc envisager de transfert génétique sous cette forme.

## L'organisation de la diversité génétique

Ce chapitre fait le point des connaissances actuelles sur l'organisation de la diversité de l'hévéa. Il se fonde sur les acquis d'études antérieures et présente des résultats complémentaires issus d'analyses isoenzymatiques sur un plus grand nombre de génotypes. Ces données supplémentaires nous ont permis de développer, sur un nombre de génotypes suffisant, une démarche originale de comparaison entre variabilité morphoagronomique et diversité génétique de marqueurs neutres, par une analyse factorielle multiple.

Les recherches sur la diversité génétique de l'hévéa avaient pour objectifs de mieux connaître l'importance et l'organisation de la diversité et la relation entre variabilité agronomique et diversité génétique. Ces études visaient

à répondre aux problèmes des sélectionneurs, qui souhaitent définir une stratégie optimale de conservation des ressources génétiques, échantillonner au mieux les populations pour leur évaluation agronomique et choisir les plans de croisement et les géniteurs.

## La variabilité agromorphologique

### L'ÉVALUATION AGRONOMIQUE

L'évaluation agronomique du matériel issu de prospection a été conduite afin de constituer, avant toute recombinaison génétique, deux collections de taille réduite. L'une, appelée collection de travail, vise à concentrer rapidement les allèles d'intérêt agronomique en vue d'apprécier les possibilités de sortie clonale utilisable, notamment après croisement avec le matériel Wickham. Cette approche peut se satisfaire de la seule évaluation agronomique, mais le marquage génétique doit permettre d'apprécier la réduction de variabilité opérée par cette sélection intense et très directionnelle. L'autre devrait être construite selon le concept de *core collection*, c'est-à-dire en cherchant à maintenir la plus grande diversité possible dans un échantillon de taille réduite. Cette démarche nécessite la mise en œuvre de toutes les méthodes d'évaluation : agronomique, morphologique, isoenzymatique et moléculaire, pour les génomes nucléaire et cytoplasmique. Des recherches méthodologiques sont en cours, notamment sur l'hévéa, afin d'optimiser la sélection des génotypes tout en intégrant ces différents types d'information (HAMON *et al.*, 1998).

Un premier essai portant sur les 2 500 génotypes issus de la prospection de l'IRRDB dans les Etats de l'Acre (AC), du Rondônia (RO) et du Mato Grosso (MT) au Brésil, à raison d'un arbre par génotype, a permis de réaliser une première évaluation précoce globale de la production et de la croissance des arbres. La production moyenne de ce matériel sauvage est très faible, elle ne représente que 12 % de celle du clone témoin Wickham, GT1. De même, son taux de croissance est en moyenne inférieur à celui de GT1. L'aspect architectural des arbres se caractérise par une prédominance des clones filant en hauteur, avec peu ou pas de branches ; ce caractère correspondrait à une adaptation naturelle à l'environnement forestier, qui permet aux jeunes arbres d'accéder rapidement à la lumière. On trouve cependant une faible proportion de clones qui ont une production voisine de celle de GT1, une bonne vigueur ou une abondante ramification (CHAPUSET *et al.*, 1995 ; CLEMENT-DEMANGE *et al.*, 1997).

Du point de vue agronomique, il apparaît un gradient de différenciation d'est en ouest de l'aire de répartition des génotypes prospectés, comme l'illustre le premier plan de l'analyse en composantes principales (ACP) réalisée sur les valeurs moyennes de production, de croissance et d'architecture, par localité (Vi Cao, comm. pers. ; figure 1). Les génotypes du Mato Grosso se distinguent

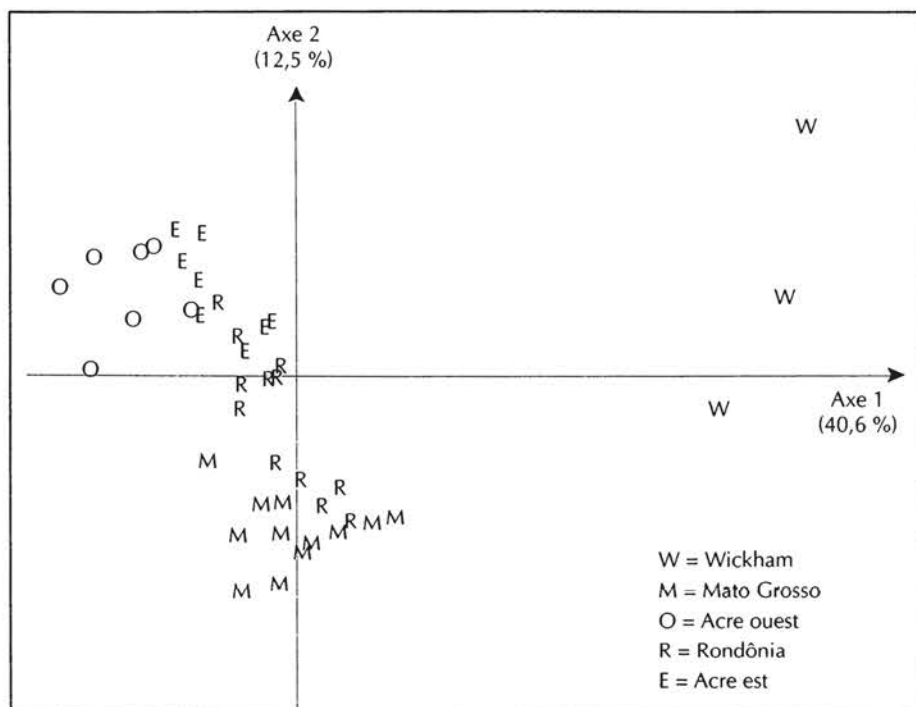


Figure 1. Variabilité agronomique du matériel amazonien et des clones Wickham. Plan 1-2 de l'ACP sur les valeurs moyennes de trois clones Wickham et de 41 localités de la collection de l'IRRDB, calculée sur 2 300 clones amazoniens.

notamment de ceux de l'Acre par une meilleure aptitude à la ramification et une production plus forte, mais ils semblent plus sensibles à *Colletotrichum*, champignon responsable d'une maladie foliaire.

#### LA MORPHOLOGIE FOLIAIRE

Une première étude de la structure génétique de la prospection de l'IRRDB dans les Etats de l'Acre, du Rondônia et du Mato Grosso a été réalisée par une analyse biométrique de la morphologie des unités de croissance et des feuilles. Les observations sont phénotypiques et 34 % de la variance est due aux effets de l'environnement et aux erreurs. La plus grande part de la variabilité génétique (42 %) se situe entre les génotypes d'une même localité de prospection, tandis que les différences entre Etats ou entre districts apparaissent faibles. Néanmoins, l'Acre se distingue des deux autres Etats (NICOLAS *et al.*, 1988).

#### La diversité biochimique et moléculaire

La diversité génétique d'un échantillon des collections d'*H. brasiliensis* a été caractérisée à l'aide des marqueurs biochimiques et moléculaires. Les effectifs

utilisés dans ces études sont donnés dans le tableau 2. Les autres espèces du genre *Hevea* étant quasiment absentes des collections, il n'a pas été possible d'aborder l'étude de leur diversité. Un ou deux génotypes des quatre espèces *H. benthamiana*, *H. guianensis*, *H. pauciflora* et *H. spruceana* ont pu être caractérisés par les isoenzymes ou les RFLP, ce qui a permis d'identifier des allèles absents de l'espèce *H. brasiliensis*.

Tableau 2. Nombre de clones, par population et par type de marqueur, utilisés dans les études de diversité génétique.

Populations	isoenzymes	RFLP	
		nucléaires	ADNmt
Wickham	203	73	28
Hybrides (Wickham × Amazoniens)	81	0	21
Pérou (MDF)	14	0	12
Colombie (SCH)	48	0	83
Brésil EMBRAPA-Irca 1974	55	0	30
Brésil IRRDB 1981	486	92	220
Total	887	165	395

## LES ISOENZYMES

L'analyse isoenzymatique a mis en évidence 14 locus polymorphes, à l'aide de 12 systèmes isoenzymatiques, chez *H. brasiliensis* (CHEVALLIER, 1988 ; CHEVALLIER *et al.*, 1988 ; SEGUIN *et al.*, 1995a).

Le premier niveau d'analyse a porté sur la variabilité intergénotype, c'est-à-dire entre les clones. L'analyse de diversité la plus précise, en terme d'échantillonnage des populations, a été réalisée sur 486 clones de la collection de l'IRRDB, qui proviennent de 16 districts des trois Etats du Brésil. Les génotypes ont été caractérisés pour 8 locus polymorphes, sans données manquantes, et 25 allèles ont été révélés.

Le premier axe de l'AFC réalisée sur les données d'allèles représente 20 % de la variabilité totale (figure 2). Il révèle une séparation en deux groupes de clones : le premier groupe rassemble presque exclusivement la quasi-totalité des génotypes de quatre districts — trois districts du Mato Grosso (MT/A, MT/I et MT/C) et un du Rondônia (RO/PB) — ; le second groupe est constitué des génotypes appartenant aux douze districts restants — cinq du Rondônia, les cinq de l'Acre et le district MT/VB du Mato Grosso.

Le second axe, qui représente 10 % de la variabilité, montre une structuration intra-Etat avec la différenciation d'un troisième groupe constitué des deux districts les plus à l'ouest de l'Acre, AC/T et AC/F (figure 2). Comme pour les données morphoagronomiques, la variabilité génétique est faible entre les



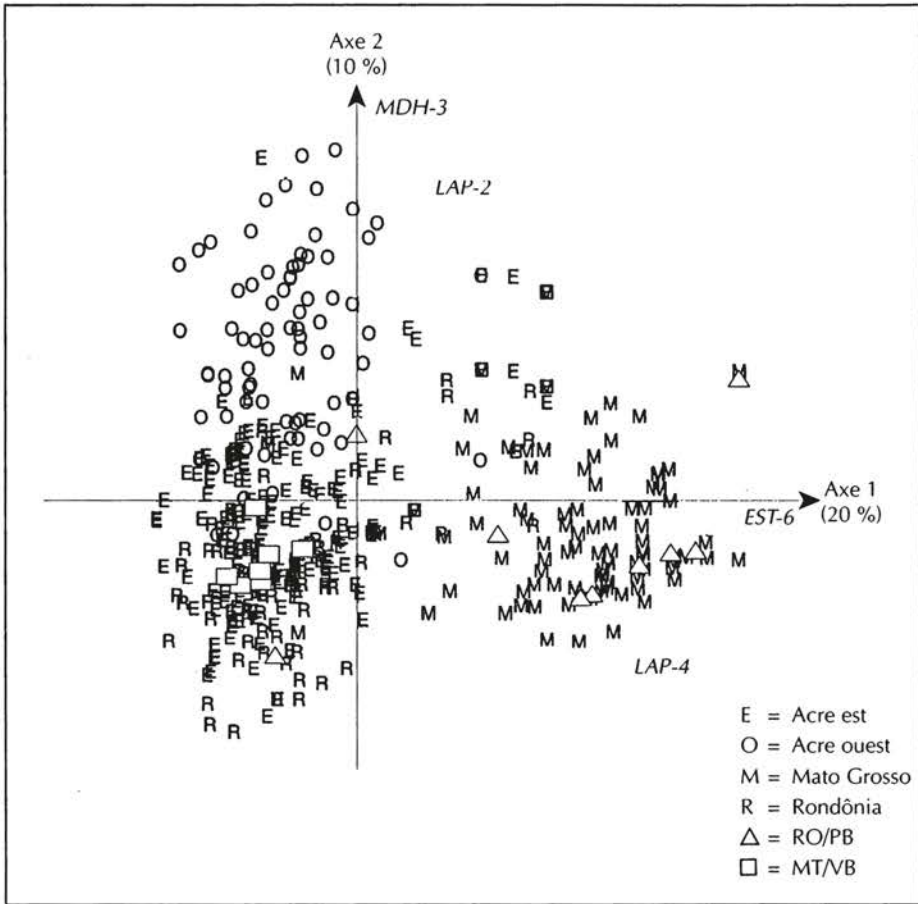


Figure 2. Diversité génétique du matériel amazonien révélée par les isoenzymes. Plan 1-2 de l'AFC sur les données de présence-absence de 25 allèles de 8 locus, chez 486 clones de la collection de l'IRRDB. MDH-3, LAP-2, LAP-4, EST-6 : allèles contribuant le plus fortement aux deux premiers axes.

groupes ou les origines et forte à l'intérieur des groupes. Malgré cette part relativement faible de diversité intergroupe — 30 % de l'inertie totale sur le premier plan de l'AFC —, la différenciation génétique en trois groupes apparaît très clairement.

Cette structuration en trois groupes avait également été mise en évidence dans une étude portant sur un échantillon plus réduit de génotypes, mais pour 11 locus isoenzymatiques (BESSE *et al.*, 1994 ; SEGUIN *et al.*, 1996). Cette étude ne comprenait pas le district RO/PB, mais incluait la population Wickham. La population Wickham se révèle génétiquement proche du groupe Mato Grosso. Elle présente un haut niveau d'hétérozygotie et de polymorphisme, très inférieur, cependant, à celui des populations amazoniennes, dont la richesse allélique est plus grande. Aucun allèle n'est spécifique de la collection Wickham, ni de l'un ou l'autre des trois Etats de la prospection amazonienne. La diversité

génétique des clones Wickham a par ailleurs permis de mettre au point une méthode isoenzymatique d'identification variétale (LECONTE *et al.*, 1994).

Le second niveau d'analyse concerne la diversité entre populations de génotypes, les populations étant définies en fonction de leur localisation. Le niveau de subdivision dépend de la finesse d'analyse souhaitée, mais surtout de l'effectif des génotypes étudiés par population.

La diversité interpopulation a été caractérisée en utilisant la distance de Nei (1978) sur les fréquences alléliques de 8 locus isoenzymatiques. La figure 3 montre l'arbre UPGMA construit pour seize districts du Brésil, une population du Pérou (MDF), deux collections de Colombie (populations Schultes) et la population Wickham, soit 590 génotypes au total. Pour les populations brésiliennes, cette analyse par population aboutit aux mêmes conclusions que l'analyse factorielle sur les génotypes pris individuellement. On retrouve le regroupement des districts en trois sous-ensembles : l'ouest de l'Acre ; l'est de l'Acre, le Rondônia et le district MT/VB ; le Mato Grosso et le district RO/PB.

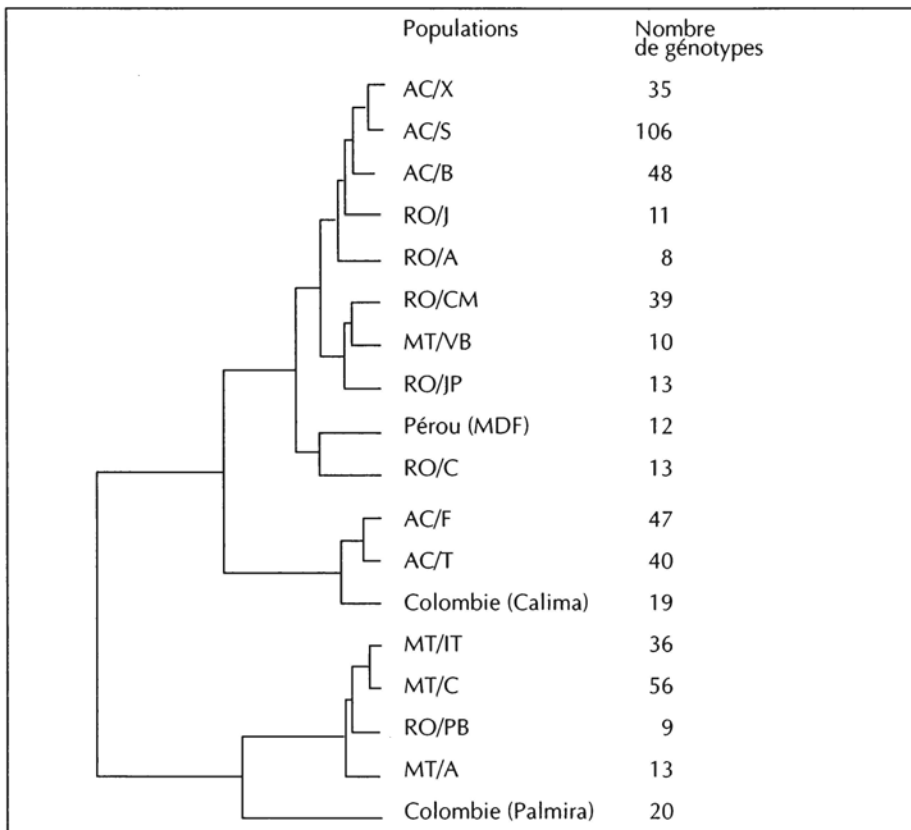


Figure 3. Distances génétiques entre populations de clones amazoniens. Arbre UPGMA construit à partir de la distance de Nei (1978) calculée sur les fréquences alléliques par population, pour 8 locus enzymatiques.

La population péruvienne est regroupée en premier avec un district du Rondônia (RO/C). Les deux populations colombiennes (SCH) paraissent génétiquement très éloignées l'une de l'autre. La première, Calima, se rapproche des deux districts de l'ouest de l'Acre ; la seconde, Palmira, est plus proche du groupe Mato Grosso que des autres populations, mais possède des allèles spécifiques, caractéristiques d'autres espèces qu'*H. brasiliensis*, qui la distinguent nettement.

L'étude isoenzymatique a été réalisée en identifiant les allèles à chaque locus grâce à la connaissance du déterminisme génétique des zymogrammes. Cela a permis d'aborder l'analyse de la diversité génétique de l'hévéa à l'aide de la génétique des populations. Le calcul des *F* statistiques de Wright montre ainsi que 75 % de la diversité du matériel brésilien se situe au niveau intrapopulation, et donc que 25 % de la variabilité correspond à la différenciation entre les districts. Lorsque l'on considère les trois groupes génétiques identifiés sur l'AFC comme des unités populationnelles, la part de diversité intragroupe atteint 80 %.

De plus, les populations constituées par les districts prospectés présentent toutes un écart à la panmixie avec un déficit d'hétérozygotes, les valeurs de l'indice de fixation  $F_{IS}$  étant comprises entre 0,10 et 0,60. Compte tenu de l'étendue géographique des districts, il est probable que ce déficit n'est pas dû à une différenciation intradistrict en sous-populations, mais plutôt à des croisements préférentiels entre arbres voisins liés au faible taux de dissémination du pollen et des graines. Les données passeport étant disponibles pour la collection amazonienne de l'IRRDB jusqu'à l'échelon de la localité de prospection dans chaque district, il serait possible d'aborder cette question, sous réserve d'analyser un plus grand nombre de génotypes.

#### LES RFLP NUCLÉAIRES

L'étude par RFLP a été réalisée sur un nombre plus restreint de génotypes : 73 Wickham et 92 amazoniens provenant des quinze districts de prospection, le district RO/PB n'étant pas représenté. En revanche, elle a porté sur un plus grand nombre de locus que pour les isoenzymes. Le déterminisme génétique des profils RFLP n'ayant pas été établi, le nombre exact de locus polymorphes n'est pas connu. Une même sonde peut en effet révéler deux locus ou davantage. Vingt-cinq sondes ayant été utilisées, le nombre de locus polymorphes ne doit cependant pas être très éloigné de 25. La cartographie du génome de l'hévéa a montré que l'on révèle généralement un seul locus avec ce type de sonde génomique (SEGUEIN *et al.*, 1996 ; 1998).

Les résultats obtenus grâce aux RFLP soulignent l'enrichissement génétique apporté par les populations amazoniennes et l'existence d'une structuration génétique forte en fonction de l'origine des prospections. De plus, par rapport aux isoenzymes, en permettant l'accès à une plus grande part du génome, les RFLP nucléaires apportent une information spécifique : ils révèlent une

différenciation très marquée des cinq districts de l'Acre, qui forment un groupe génétique distinct. En revanche, les RFLP ne mettent pas en évidence l'individualisation du groupe de l'ouest de l'Acre, révélé par les isoenzymes.

Dans l'ensemble, l'étude par les RFLP nucléaires confirme et renforce les observations réalisées avec les isoenzymes. On retrouve, en particulier, la même partition de la diversité en intra et intergroupe, avec également 30 % de la variabilité sur le premier plan de l'AFC correspondant à la différenciation en groupes (BESSE *et al.*, 1994 ; SEGUIN *et al.*, 1995b ; 1996). Avec les deux types de marqueurs, le premier axe des AFC correspond à la différenciation du groupe de trois districts du Mato Grosso, ce groupe apparaissant comme le plus proche de la population Wickham.

### LES RFLP CYTOPLASMIQUES

Les génomes cytoplasmiques ont le plus souvent une hérédité uniparentale (REBOUD et ZEYL, 1994 ; MOGENSEN, 1996), qui interdit la recombinaison génétique, et une vitesse d'évolution propre à chaque génome (WOLFE *et al.*, 1987 ; PALMER et HERBON, 1988). L'étude de la diversité mitochondriale et chloroplastique est donc complémentaire de celle du génome nucléaire.

L'étude du génome mitochondrial (ADNmt) a porté sur 395 génotypes. Un fort polymorphisme a été révélé par les RFLP (LUO et BOUTRY, 1995 ; LUO *et al.*, 1995). On retrouve, comme à l'échelle nucléaire, la divergence des populations de l'Acre ainsi qu'une séparation entre les génotypes de l'ouest de l'Acre (districts AC/T et AC/F) par rapport à ceux de l'est (AC/B, AC/S et AC/X) comme avec les isoenzymes. Les génotypes du Pérou (MDF) se regroupent préférentiellement avec les génotypes du district RO/C du Rondônia, comme avec les isoenzymes.

En revanche, les populations du Mato Grosso et du Rondônia sont hétérogènes et ne se regroupent pas en fonction de leur origine. De même, les populations colombiennes (SCH) apparaissent très polymorphes et ne s'individualisent pas. De plus, à la différence des isoenzymes et des RFLP nucléaires, la diversité de l'ADNmt est pratiquement nulle au sein des clones Wickham, qui présentent tous, sauf le clone GT1, un même type mitochondrial spécifique. Le clone cultivé GT1, mâle-stérile, présente un type mitochondrial unique très différent du type Wickham dominant.

La diversité du génome chloroplastique (ADNcp) a également été étudiée par les RFLP, avec des sondes d'ADNcp de féverole, sur 217 des 395 génotypes précédents. Comme pour beaucoup d'autres espèces, l'ADNcp présente, chez l'hévéa, un polymorphisme très faible. Deux types chloroplastiques seulement ont été trouvés, contre 126 types mitochondriaux détectés sur le même échantillon de clones. Le type minoritaire, présent chez 37 génotypes, est caractéristique des trois districts de l'est de l'Acre, auxquels s'ajoutent trois des six génotypes du district RO/C présents dans l'étude (LUO *et al.*, 1995).

## L'ORGANISATION DE LA DIVERSITÉ BIOCHIMIQUE ET MOLÉCULAIRE

Il apparaît donc clairement que l'apport des marqueurs moléculaires a été considérable pour la caractérisation des ressources génétiques de l'hévéa. Ils ont permis de révéler l'importance de la diversité d'*H. brasiliensis*. Les différents types de marqueur utilisés ont donné des résultats globalement concordants sur l'organisation de cette diversité, qui se structure en fonction de l'origine géographique des génotypes. De l'ensemble de ces études, il est possible de dégager les conclusions suivantes.

Sur la base des similarités génétiques entre les génotypes ou les populations, six groupes génétiques ont été identifiés au sein du matériel d'*H. brasiliensis* étudié (Seguin *et al.*, 1996) :

- le groupe 1 rassemble deux districts de l'ouest de l'Acre et la population Calima de Colombie ;
- le groupe 2 correspond aux trois districts de l'est de l'Acre ;
- le groupe 3 comprend six districts du Rondônia, le district MT/VB du Mato Grosso et la population MDF du Pérou ;
- le groupe 4 réunit trois districts du Mato Grosso et le district RO/PB du Rondônia ;
- le groupe 5 est constitué de la collection Palmira de Colombie ;
- le groupe 6 correspond à la population Wickham.

La localisation géographique des populations et des groupes génétiques est présentée sur la figure 4.

Ces groupes génétiques rassemblent des populations géographiquement proches. De plus, la relation entre la distance géographique et la distance génétique ne se retrouve pas seulement à l'intérieur d'un même groupe, mais également entre les groupes. Ainsi, la différenciation moléculaire du groupe 6 (Wickham) tient essentiellement à l'existence d'un type mitochondrial particulier, mais au niveau nucléaire, ce groupe apparaît proche du groupe 4, le plus proche géographiquement. La place des populations Schultes de Colombie est plus difficile à discuter faute de données passeport précises.

Cette organisation de la diversité d'*H. brasiliensis* apparaît très clairement liée au réseau hydrographique du bassin amazonien (tableau 1 ; figure 4), qui détermine les flux de gènes, et donc la similarité génétique, entre populations ou groupes de populations. Pour les populations brésiliennes, dont les données passeport sont disponibles et précises, on peut associer à chaque groupe un bassin fluvial (figure 4) : au groupe 1, le rio Juruá ; au groupe 2, le rio Purus ; au groupe 3, le rio Madeira et ses prolongements — le rio Guaporé au Brésil et le rio Madre de Dios au Pérou — ; au groupe 4, le rio Tapajós.

De même, certaines similarités entre les groupes peuvent s'expliquer par des flux géniques suivant les rivières ou les fleuves. Ainsi, la proximité entre la population RO/C et le groupe 2 est liée au rio Abuna, affluent du rio Madeira, et celle du groupe 4 et du groupe 6, au rio Tapajós.

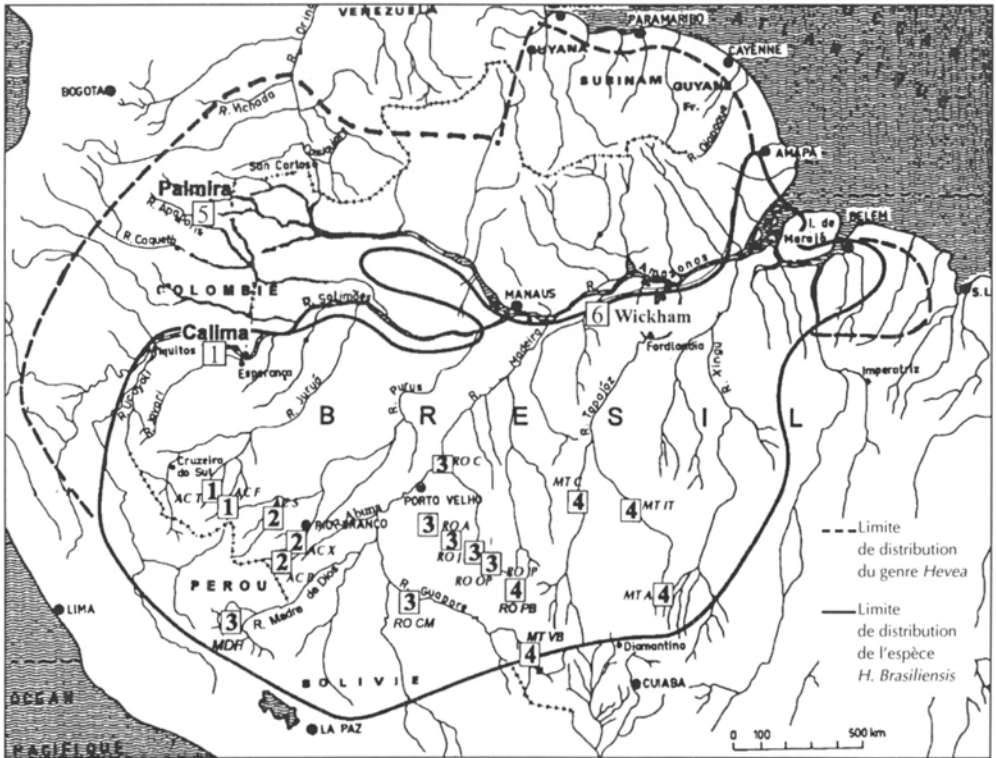


Figure 4. Correspondance entre origine géographique et similarité génétique chez *Hevea brasiliensis*. Les numéros 1 à 6 correspondent aux groupes génétiques identifiés par les marqueurs moléculaires. Le lieu exact de collecte n'est pas connu pour les clones prospectés par E. Schultes en Colombie (collections de Calima et de Palmira).

Cette situation justifierait que l'on désigne les groupes génétiques de l'hévéa par le nom du fleuve correspondant. Mais, pour simplifier la présentation, nous désignerons les groupes par leurs numéros de 1 à 6. Le district RO/PB fait un peu exception : localisé sur un affluent du rio Madeira, il se rattache au groupe 4, du rio Tapajós, et non au groupe 3 (figures 2 et 3). La zone prospectée dans ce district est en fait proche de la limite entre les deux bassins fluviaux (figure 4) et l'on peut admettre qu'une dissémination importante se fasse via le bassin du rio Tapajós. La diversité des clones cultivés Wickham est réduite comparée à celle des ressources génétiques amazoniennes, notamment du point de vue cytoplasmique. La diversité nucléaire du groupe Wickham apparaît néanmoins plus forte que ne le laissait prévoir l'histoire de la domestication de l'hévéa en Asie.

Ces conclusions découlent de l'ensemble des études de diversité par marqueurs, ce qui ne veut pas dire que tous les types de marqueur ont donné exactement les mêmes résultats, chaque étude apportant en réalité des informations spécifiques. Plusieurs facteurs permettent d'expliquer les différences observées avec les différents types de marqueur.

Il n'est pas surprenant que des marqueurs génétiques de nature différente apportent des informations différentes, qui ne sont pas pour autant contradictoires. Ainsi, les marqueurs cytoplasmiques, du fait de leur hérédité uniparentale (REBOUD et ZEYL, 1994 ; MOGENSEN, 1996), permettent de mieux révéler des flux migratoires dont la trace disparaît rapidement au niveau nucléaire à la suite de la recombinaison génétique entre populations indigènes et immigrantes (PALMER, 1987 ; CROZIER, 1990). Pour l'hévéa, par exemple, c'est le cas de la similarité entre le district RO/C et le groupe génétique 2 révélée spécifiquement par les marqueurs chloroplastiques. Il existe donc des flux géniques, probablement liés à la flottaison des graines suivant le cours du rio Apuna, depuis les districts AC/X et AC/S du groupe 2 jusqu'au nord du Rondônia, mais ces migrations seraient peu intenses et la divergence génétique entre les deux groupes s'est maintenue.

La vitesse d'évolution varie beaucoup selon le génome ou le type de séquence considérés (WOLFE *et al.*, 1987 ; PALMER et HERBON, 1988). Avec une évolution très lente, il est peu probable qu'apparaisse une mutation associée à une différenciation récente en sous-populations. Inversement, une évolution moléculaire très rapide conduit au phénomène d'homoplasie, ou de convergence évolutive : un même caractère, une même forme d'un marqueur, a une probabilité élevée d'apparaître indépendamment dans plusieurs populations isolées et n'est donc pas caractéristique d'une population. Cela dépend donc aussi du temps de divergence impliqué dans la différenciation des unités étudiées, populations ou espèces. Pour l'hévéa, ce phénomène pourrait expliquer l'organisation complexe des types mitochondriaux entre les groupes 3 et 4, pourtant bien différenciés sur le plan nucléaire. Cependant, cette situation peut aussi traduire l'existence de flux migratoires, faibles mais divers, entre ces deux groupes et il n'est pas possible de conclure. La vitesse d'évolution trop rapide expliquerait également pourquoi aucune structuration n'a été observée au sein de la collection de l'IRRDB dans une étude portant sur les variations de longueur du gène codant pour les ARN ribosomiques (BESSE, 1993 ; BESSE *et al.*, 1993).

Enfin, et c'est sans doute un facteur déterminant dans nombre d'études de diversité moléculaire, la qualité des résultats dépend du niveau d'échantillonnage de la diversité. Les différences de résultats obtenus avec les isoenzymes ou les RFLP chez l'hévéa ne proviennent pas d'une différence de nature. On peut considérer que les deux types de marqueurs sont neutres à l'égard de la sélection, qu'ils correspondent au même type de séquence nucléaire et qu'ils ont une vitesse d'évolution similaire en moyenne. En revanche, ces différences peuvent s'expliquer par le mode d'échantillonnage des populations — le nombre et la représentativité des génotypes étudiés —, ou du génome, c'est-à-dire le nombre de locus étudiés et leur répartition sur l'ensemble du génome. Le fait que les groupes 1 et 2 ne se distinguent pas dans l'étude par les RFLP, est probablement dû à un effectif trop faible des génotypes originaires de l'Acre, comme le suggère l'ACP sur les isoenzymes



réalisée sur le même échantillon (figure 5). Parallèlement, l'absence de divergence entre les groupes 2 et 3 avec les isoenzymes viendrait de ce que l'on n'a pas détecté de locus ou d'allèle marqueur de cette différenciation, les marqueurs isoenzymatiques obtenus chez l'hévéa, peu nombreux, ne couvrant pas suffisamment l'ensemble du génome.

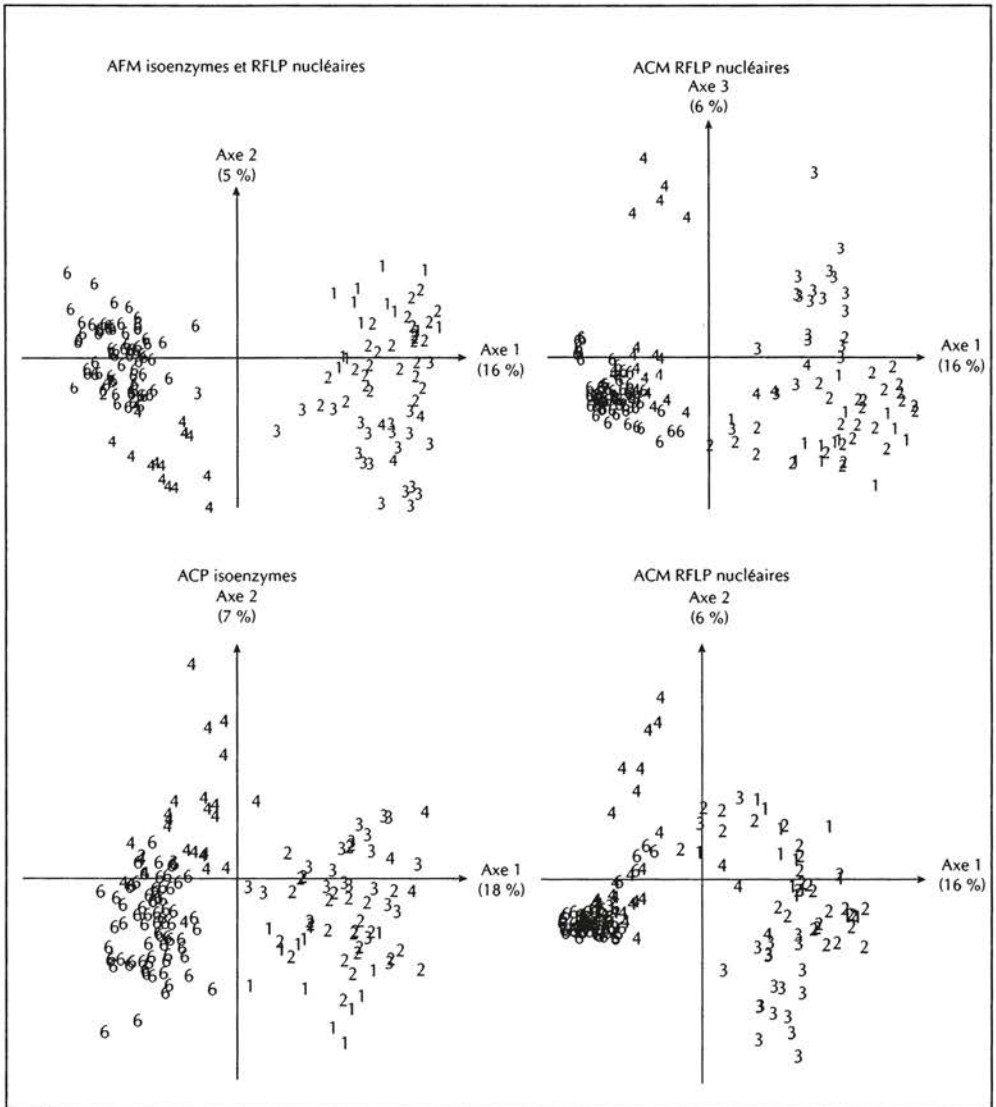


Figure 5. Comparaison par AFM de la diversité révélée par les isoenzymes et par les RFLP nucléaires. Analyse de 72 clones Wickham et de 89 clones amazoniens de la collection de l'IRRDB. Représentation graphique des premiers axes de l'AFM et des analyses partielles (ACP isoenzymes et ACM RFLP). Les numéros 1 à 6 correspondent aux groupes génétiques identifiés par les marqueurs moléculaires.



## La diversité moléculaire et la variabilité morphoagronomique : comparaison par analyse factorielle multiple

L'analyse factorielle multiple (AFM) a été élaborée pour analyser simultanément et comparer des variables, qualitatives ou quantitatives, organisées en groupes de même nature. Cette méthode est donc bien adaptée à l'analyse simultanée de données moléculaires, qualitatives, de différents types (isoenzymes, RFLP...) ou de données agronomiques, quantitatives. Le principe de l'AFM est présenté dans cet ouvrage (PERRIER *et al.*, 1999) et plus en détail dans ESCOPIER et PAGES (1988).

### LA COMPARAISON ENTRE ISOENZYMES ET RFLP NUCLÉAIRES

Nous avons mis en œuvre l'AFM, en premier lieu, pour obtenir une mesure de la relation entre les différents marqueurs moléculaires utilisés chez l'hévéa. Une première analyse par AFM a ainsi porté sur 161 génotypes (89 clones brésiliens de la collection de l'IRRDB et 72 clones Wickham) caractérisés par deux groupes de variables moléculaires : les isoenzymes (44 allèles pour 11 locus) et les RFLP nucléaires (113 bandes pour 25 sondes génomiques).

Les résultats de l'AFM sont donnés dans le tableau 3 et la figure 5. Les valeurs de corrélation correspondent aux corrélations entre les coordonnées des points (génotypes d'hévéa) sur les axes des analyses factorielles. Le nombre d'axes à prendre en compte dépend en partie de leur valeur propre. Pour l'AFM, une valeur propre du premier axe supérieure à 1 indique que plus d'un groupe de variables y contribuent ou, en d'autres termes, qu'il existe un axe d'inertie commun aux groupes de variables (ESCOPIER et PAGES, 1988). En revanche, pour les axes suivants de l'AFM, la valeur propre n'est pas un critère suffisant et le choix de retenir un axe ou non dépend, comme pour toute ACP, de l'interprétation biologique que l'on peut apporter (ESCOPIER et PAGES, 1988). La qualité de la représentation et la contribution d'un axe partiel à un axe de l'AFM sont également des paramètres à prendre en compte comme pour l'interprétation d'une ACP classique.

Les deux premiers axes de l'AFM témoignent de la forte similitude entre les structurations révélées par les deux types de marqueurs, alors que ce sont essentiellement les RFLP qui contribuent au troisième axe de l'AFM (tableau 3). Le premier axe de l'AFM correspond à l'opposition entre les groupes 4 et 6 (rio Tapajós), d'un côté, et les groupes 1, 2 et 3, de l'autre. Cette opposition apparaît très nettement sur le premier axe des deux analyses partielles (figure 5). Le deuxième axe de l'AFM oppose les groupes 3 et 4 aux groupes 1, 2 et 6. Cette opposition se manifeste au second rang pour les isoenzymes et au troisième pour les RFLP (tableau 3b ; figure 5). Elle est moins marquée avec les RFLP comme en témoignent les valeurs de corrélation, de qualité de la représentation et de contribution (tableau 3b). Bien qu'il ait une

Tableau 3. AFM sur les isoenzymes et les RFLP nucléaires pour 161 clones d'hévéa. a : valeurs propres et part d'inertie des premiers axes des analyses factorielles ; b : corrélation, qualité de la représentation et contribution des axes des analyses partielles (ACP et ACM) aux deux premiers axes de l'AFM.

a

Analyse factorielle	Facteurs (n° axe)	Valeur propre	Inertie (%)
AFM	1	1,9	16
	2	0,6	5
	3	0,5	4
ACP isoenzymes	1	8,1	18
	2	3,2	7
	3	2,5	6
ACM RFLP	1	20	16
	2	7,4	6
	3	7,2	6

b

Analyse factorielle	Facteurs (n° axe)	Axe 1 de l'AFM			Axe 2 de l'AFM			Axe 3 de l'AFM		
		corrélation	qualité	contribution	corrélation	qualité	contribution	corrélation	qualité	contribution
ACP isoenzymes	1	0,97	0,94	0,50	0,05	0	0	0,08	0	0,01
	2	0,04	0	0	0,87	0,75	0,53	0,08	0	0
	3	0,03	0	0	0,14	0,02	0	0,58	0,34	0,22
ACM RFLP	1	0,97	0,93	0,50	0,04	0	0	0,09	0	0,02
	2	0,01	0	0	0,17	0,03	0,02	0,80	0,64	0,49
	3	0	0	0	0,73	0,53	0,34	0,07	0	0

valeur propre inférieure à 1, ce deuxième axe de l'AFM, qui correspond à un axe d'inertie commun aux deux types de marqueurs, est à retenir. Le troisième axe de l'AFM correspond à l'opposition entre le groupe 3 et le groupe 4. Cette opposition est clairement mise en évidence par les RFLP (deuxième axe de l'analyse des correspondances multiples, ACM), mais n'apparaît pas nettement sur le troisième axe de l'ACP des isoenzymes, qui présente une corrélation et une contribution plus faibles (tableau 3b).

Ces éléments de structuration se retrouvent sur le premier plan des analyses partielles — ACP isoenzyme et ACM RFLP séparées — mais de façon moins marquée. De plus, l'AFM montre que la corrélation entre les deux types de marqueur est très forte pour la part la plus significative et la plus structurée de leur variabilité.

#### LA COMPARAISON ENTRE ISOENZYMES ET CARACTÈRES MORPHOAGRONOMIQUES

L'application de l'AFM a ensuite été étendue à la comparaison entre diversité moléculaire et variabilité phénotypique. L'analyse a porté sur 171 génotypes amazoniens de la collection de l'IRRDB pour 51 allèles isoenzymatiques (13 locus) et 26 variables morphoagronomiques (production, croissance, architecture), sans données manquantes.

Les résultats de l'AFM sont donnés dans le tableau 4 et la figure 6. Le premier axe de l'AFM a une valeur propre de 1,2 — supérieure à 1 mais peu élevée —,

Tableau 4. AFM sur les données isoenzymatiques et morphoagronomiques pour 171 clones d'hévéa. a : valeurs propres et part d'inertie des premiers axes des analyses factorielles ; b : corrélation, qualité de la représentation et contribution des axes des analyses partielles (ACP) au premier axe de l'AFM.

a				
Analyse factorielle	Facteurs (n° axe)	Valeur propre	Inertie (%)	
AFM	1	1,2	10,5	
	2	0,9	8,3	
ACP isoenzymes	1	5,6	11	
	2	3,9	7,6	
ACP morphoagronomique	1	12,7	49	
	2	2,2	8,5	

b				
Analyse factorielle	Facteurs (n° axe)	Facteurs corrélation	Axe 1 de l'AFM	
			qualité	contribution
ACP isoenzymes	1	0,67	0,45	0,56
	2	0	0,01	0,01
ACP morphoagronomique	1	0,81	0,67	0,57
	2	0,03	0,00	0

ce qui signifie que la corrélation entre les deux groupes de variables est faible. Elle n'est cependant pas nulle en raison du caractère particulier des génotypes du groupe 4, très nettement différenciés par les isoenzymes (axe 1 de la figure 6) et dont les valeurs de production ont tendance à être supérieures à la moyenne. Le second axe de l'AFM ne doit pas être pris en compte dans cette analyse car il ne permet pas de dégager une interprétation claire de la relation entre les deux groupes de variables.

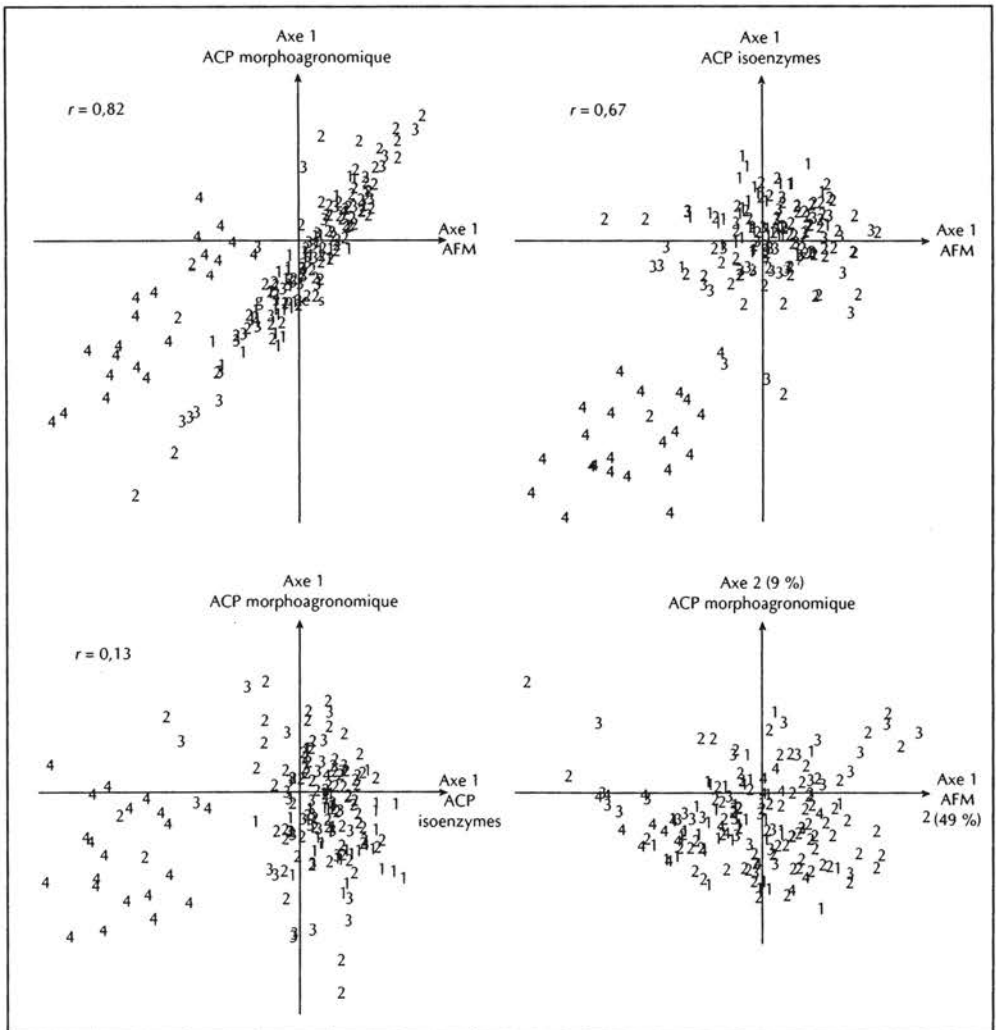


Figure 6. Comparaison par AFM de la diversité des isoenzymes et de la variabilité morphoagronomique chez 183 clones amazoniens de la collection de l'IRRDB. Représentation graphique des premiers axes de l'AFM et des analyses partielles (ACP). Les numéros 1 à 6 correspondent aux groupes génétiques identifiés par les marqueurs moléculaires.

Les résultats de l'AFM suggèrent donc qu'il existe une relation entre la diversité des marqueurs neutres et la variabilité phénotypique chez l'hévéa. Cette relation peut paraître en contradiction avec la faible valeur de corrélation entre les premiers axes des deux analyses partielles (figure 6). Compte tenu de l'effectif, cette corrélation de 0,13 est en fait significative à 10 %. Ainsi l'AFM fait apparaître une relation difficile à identifier sur les analyses séparées. Les données morphoagronomiques sont peu précises car il s'agit de mesures monoarbre, sans répétitions. C'est pourquoi, une différenciation n'apparaît avec ces données que lorsque l'on travaille sur de grands effectifs et en valeurs moyennes, à l'échelle de la population (figure 1). La relation entre diversité moléculaire et variabilité phénotypique apparaîtrait sans doute plus forte si l'on disposait d'une meilleure évaluation des caractères morphoagronomiques. Un examen attentif de l'ACP morphoagronomique (figure 6) montre que les clones du groupe 4 sont, parmi des génotypes d'autres groupes, plutôt situés d'un même côté du premier axe ; cette répartition ne peut être retenue comme significative d'après cette seule analyse partielle.

## Conclusion et perspectives

L'analyse fine de la diversité génétique de l'hévéa avec différents outils, tant moléculaires que phénotypiques, a conduit à une bonne connaissance de l'organisation génétique de cette espèce. L'apport des marqueurs moléculaires a été déterminant. Les résultats obtenus indépendamment par différentes techniques convergent vers une image claire de la structuration génétique, qui peut être considérée comme quasiment définitive, au moins pour les populations du Brésil les plus étudiées.

Une forte corrélation a été trouvée entre les différents outils moléculaires. Pour les populations amazoniennes, la relation établie entre diversité nucléaire et diversité mitochondriale est particulièrement intéressante. Elle prouve que la différenciation entre les populations, si elle n'est pas très forte, est bien réelle et que des flux géniques entre les populations existent mais restent peu intenses. Le bassin amazonien ne présente pas de barrière géographique à la migration, mais l'immensité de la zone considérée et la dispersion limitée des graines conduisent à un maintien de la structuration génétique héritée de la fragmentation de la forêt au pléistocène selon le modèle des refuges forestiers (HAFFER, 1982 ; PRANCE, 1987).

Les relations de similarité génétique entre populations d'hévéa, révélées par les marqueurs cytoplasmiques et nucléaires, montrent que les flux de gènes sont déterminés principalement par le réseau hydrographique. La dispersion des graines d'hévéa se ferait suivant les cours d'eau, soit sous l'action de l'homme (SEIBERT, 1948), soit par flottaison naturelle des graines comme cela se produit pour d'autres espèces amazoniennes d'arbre (GOULDING, 1993). Une même

structuration de la diversité, en fonction du réseau hydrographique, a d'ailleurs été trouvée chez une autre espèce pérenne amazonienne : le palmier à huile américain *Elaeis oleifera* (BARCELOS, 1998).

Les différents types de marqueurs moléculaires ont donné des résultats dans l'ensemble concordants, ce que confirme l'AFM, mais chacun apporte des informations spécifiques. L'hérédité maternelle des marqueurs cytoplasmiques permet de mieux apprécier les flux géniques entre populations — entre le groupe 2 et le district RO/C, par exemple — ou les effets de fondation, comme pour la populations Wickham, quasiment monomorphe sur le plan mitochondrial.

Certaines différenciations entre groupes n'ont été clairement révélées que par les isoenzymes ou les RFLP. Sur le plan nucléaire, cela est imputable à des différences dans le niveau d'échantillonnage des populations (nombre de génotypes analysés) ou du génome (nombre de marqueurs utilisés). Il serait donc utile de caractériser la diversité des ressources génétiques de l'hévéa en utilisant des marqueurs cartographiés et distribués sur l'ensemble du génome. Le développement des marqueurs microsatellites (LESPINASSE, 1993 ; SEGUIN *et al.*, 1997) permet d'envisager une telle étude sur un échantillon suffisant de génotypes. L'intégration de tous les marqueurs sur une carte synthétique est en cours (LESPINASSE *et al.*, 1999)

Il est intéressant de noter l'existence d'une relation entre diversité moléculaire et variabilité agronomique chez l'hévéa. La corrélation est cependant beaucoup plus faible qu'entre les différents marqueurs moléculaires eux-mêmes et l'AFM s'est avérée utile pour la faire apparaître sur un sous-échantillon de génotypes. Il serait nécessaire de poursuivre l'étude de la relation entre les deux niveaux de variabilité — génétique et phénotypique — sur des données phénotypiques plus précises, recueillies sur un large échantillon de génotypes. Il serait également intéressant d'appliquer l'AFM en utilisant l'analyse factorielle sur tableau de distances (AFTD), ce qui permettrait de mieux raisonner le choix des indices de distance, les distances du  $\chi^2$  (en AFC ou en ACM) et euclidienne (en ACP) n'étant pas forcément les plus adaptées.

Nos résultats indiquent que la différenciation des populations d'hévéa est assez forte pour toucher l'ensemble des génomes et concerne un grand nombre de gènes, qu'il s'agisse de marqueurs neutres ou de gènes déterminant des caractères phénotypiques. En conséquence, dans l'état actuel de nos connaissances, la prise en compte de la structuration d'*Hevea brasiliensis* en six groupes génétiques sera très utile pour définir une stratégie d'échantillonnage de la diversité génétique de cette espèce.

## Annexe

### Matériel végétal

Les génotypes retenus dans cette étude sont des clones d'hévéa de la collection internationale de l'IDEFOR (Institut des forêts), en Côte d'Ivoire. L'origine géographique des populations de clones est précisée dans le tableau 1. Tous ces clones ont été évalués au champ pour les caractères de production de latex, de croissance en épaisseur et d'architecture, mais dans des essais différents (CHAPUSET *et al.*, 1995). L'essai le plus important comprend 2 500 clones de la prospection de l'IRRDB de 1981, sans répétitions (1 arbre par clone). Les marqueurs moléculaires ont été appliqués à des échantillons de la collection, de taille et de recouvrement variables. Les effectifs utilisés dans ces différentes études sont résumés dans le tableau 2.

### Méthodes de révélation de la diversité

Les protocoles utilisés pour l'étude de la diversité génétique et agronomique sont décrits par CHAPUSET *et al.* (1995) pour l'évaluation morphoagronomique sur 26 variables, par LEBRUN et CHEVALLIER (1990) pour les isoenzymes, par BESSE (1993) et BESSE *et al.* (1994) pour les RFLP nucléaires avec 25 sondes génomiques et par LUO *et al.* (1995) pour les RFLP cytoplasmiques, ADNmt et ADNcp.

La technique des isoenzymes utilise 12 systèmes enzymatiques, qui permettent de révéler jusqu'à 14 locus polymorphes dans l'espèce *H. brasiliensis* (CHEVALLIER, 1988 ; CHEVALLIER *et al.*, 1988 ; SEGUIN *et al.*, 1995a ; 1995b). Plusieurs centaines de génotypes ont été caractérisés par les isoenzymes (tableau 2) à l'aide de ces 12 systèmes, mais avec un nombre variable de données manquantes selon les génotypes ou les séries de manipulations. C'est pourquoi l'analyse des données n'a porté que sur 8 à 13 locus isoenzymatiques selon le dispositif. Pour avoir un effectif de génotypes suffisant, il a souvent été nécessaire de réduire le nombre de locus pour limiter ou éliminer les données manquantes.

Pour les RFLP nucléaires, 25 sondes homologues correspondant à des séquences uniques ou peu répétées dans le génome de l'hévéa (banque génomique *Pst*I) ont été utilisées. Ces sondes ont été couplées à deux enzymes de restriction, *Eco*RI et *Sst*I (BESSE *et al.*, 1994 ; SEGUIN *et al.*, 1995a). Pour les RFLP cytoplasmiques, les sondes proviennent de banques de féverole (LUO *et al.*, 1995).

### Analyse de données

Le traitement des données a été réalisé principalement par l'analyse factorielle, dont l'intérêt est discuté dans cet ouvrage (PERRIER *et al.*, 1999). De plus, l'interprétation des résultats est plus facile avec une analyse factorielle qu'avec les méthodes arborées lorsque l'on travaille sur un grand nombre d'individus.

Les données agronomiques, quantitatives, ont été traitées par ACP. Les données de présence/absence de bandes RFLP ont été traitées par AFC sur tableau disjonctif complet, ce qui revient à une analyse des correspondances multiples (ACM). Les données d'allèles isoenzymatiques ont été codées, 0 pour l'absence, 1 pour la présence à l'état hétérozygote et 2 pour la présence à l'état homozygote, pour chaque génotype. Ces données ont été traitées par AFC ou par ACP dans l'AFM (PERRIER *et al.*, 1999). Pour les

AFC, n'ont été retenus comme variables actives que les bandes ou les allèles de fréquence comprise entre 2 et 98 %, dans l'échantillon total. En AFM, tous les allèles ont été retenus quelle que soit leur fréquence.

L'analyse de la diversité a également été menée à l'échelle de la population, par analyse en classification ascendante hiérarchique (CAH) en utilisant, pour les isoenzymes, la distance de Nei (1978) sur les fréquences alléliques et, pour les données agronomiques, quantitatives, la distance de Manhattan sur les moyennes par population. Les arbres ont été construits par la méthode UPGMA. L'étude de l'organisation du matériel génétique en sous-populations a été complétée par les calculs de  $F$  statistiques, qui permettent d'évaluer l'écart à la panmixie et de quantifier le rapport des diversités intra et interpopulations.



## Références bibliographiques

- BALDWIN J.J.T., 1947. Hevea: a first interpretation. *Journal of Heredity*, 38 : 54-64.
- BARCELOS E., 1998. Etude de la diversité génétique du genre *Elaeis* (*E. oleifera* Cortès et *E. guineensis* Jacq.) par marqueurs moléculaires (RFLP et AFLP). Thèse de doctorat, ENSAM, Montpellier, France, 138 p.
- BENNETT M.D., LEITCH I.J., 1997. Nuclear DNA amounts in Angiosperms: 583 new estimates. *Annals of Botany*, 80 : 169-196.
- BESSE P., 1993. Identification des clones cultivés et analyse de la diversité génétique chez *Hevea brasiliensis* par RFLP. Thèse de doctorat, université Paris XI, Orsay, France, 114 p.
- BESSE P., SEGUIN M., LEBRUN P., CHEVALLIER M.H., NICOLAS D., LANAUD C., 1994. Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis. *Theoretical and Applied Genetics*, 88 : 199-207.
- BESSE P., SEGUIN M., LEBRUN P., LANAUD C., 1993. Ribosomal DNA variations in wild and cultivated rubber tree (*Hevea brasiliensis*). *Genome*, 36 : 1049-1057.
- BOUHARMONT J., 1960. Recherches taxonomiques et caryologiques chez quelques espèces du genre *Hevea*. Congo, INEAC, 64 p.
- CHAPUSET T., LEGNATE H., DOUMBIA A., CLEMENT-DEMANGE A., NICOLAS D., KELI J., 1995. Agronomical characterisation of the 1981 germplasm in Côte d'Ivoire: growth, production, architecture and leaf disease sensibility. *In* : IRRDB symposium on physiological and molecular aspects of the breeding of *Hevea brasiliensis*. Brickendonbury, Royaume-Uni, IRRDB, p. 112-122.
- CHEVALLIER M.H., 1988. Genetic variability of *Hevea brasiliensis* germplasm using isoenzymes markers. *Journal of Natural Rubber Research*, 3 : 42-53.
- CHEVALLIER M.H., LEBRUN P., NORMAND F., 1988. Approach of the genetic variability of germplasm using enzymatic markers. *In* : Colloque exploitation, physiologie et amélioration de l'hévéa, J.L. Jacob et J.C. Prévôts éd., Montpellier, France, Cirad-Irca, p. 365-376.
- CLEMENT-DEMANGE A., SEGUIN M., LESPINASSE D., LEGNATE H., CHAPUSET T., NICOLAS D., 1997. Germplasm, genetic improvement and marker assisted selection of the rubber tree. *In* : Seminar-workshop on the biochemical and molecular tools for exploitation diagnostic and rubber tree improvement. Cirad-Orstom-University of Mahidol (sous presse).
- CROZIER R., 1990. From population genetics to phylogeny uses and limits of mitochondrial DNA. *Australian Systematic Botany*, 3 : 111-124.
- ESCOFIER B., PAGES J., 1988. Analyses factorielles simples et multiples. Paris, France, Dunod, 241 p.
- FREGENE M., ANGEL F., GOMEZ R., RODRIGUEZ F., CHAVARRIAGA P., ROCA W., TOHME J., BONIERBALE M., 1997. A molecular map of cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics*, 95 : 431-441.

- GONÇALVES P.S., 1981. Expedição internacional à Amazônia no Território Federal de Rondônia para coleta de material botânico de seringueira (*Hevea brasiliensis*). Manaus, Brésil, EMBRAPA-CNSPD, 60 p.
- GONÇALVES P.S., DE PAIVA J.R., DE SOUZA R.A., 1983. Retrospectiva e actualidade do melhoramento genético da seringueira (*Hevea* spp.) no Brasil e em países asiáticos. Manaus, Brésil, EMBRAPA-CNSPD, Documentos nº 2, 69 p.
- GOULDING M., 1993. Les forêts inondables d'Amazonie. Pour la science, 187 : 70-77.
- HAFFER J., 1982. General aspects of the refuge theory. In : Biological diversification in the tropics, G.T. Prance éd., New York, Etats-Unis, Columbia University Press, p. 6-26.
- HAMON S., DUSSERT J., DEU M., HAMON P., SEGUIN M., GLASZMANN J.C., GRIVET L., CHANTEREAU J., CHEVALLIER M.H., FLORI A., LASHERMES P., LEGNATE H., NOIROT M., 1998. Effects of quantitative and qualitative principal component score strategies on the structure of coffee, rice, rubber tree and sorghum core collections. Genetics, Selection, Evolution, 30 (suppl. 1) : 237-258.
- LEBRUN P., CHEVALLIER M.H., 1990. Starch and polyacrylamide gel electrophoresis of *Hevea brasiliensis*, a laboratory manual. Montpellier, France, Cirad-Irca, 55 p.
- LECONTE A., LEBRUN P., NICOLAS D., SEGUIN M., 1994. Electrophoresis: application to *Hevea* clone identification. Plantations, recherche, développement, 1 : 28-36.
- LESPINASSE D., 1993. Recherche de marqueurs en vue d'une cartographie génétique chez *Hevea brasiliensis*. Mémoire de DEA, ENGREF, Paris, France, 66 p.
- LESPINASSE D., RODIER-GOUD M., GRIVET L., LECONTE A., LEGNATE H., SEGUIN M., 1999. A saturated genetic linkage map of rubber-tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite, and isozyme markers. Theoretical and Applied Genetics (sous presse).
- LINS A.C.R., SILVA G.P., NICOLAS D., ONG S.H., MELO C.C., SANTOS M.R., 1981. Report of the Acre team in the 1981 joint IRRDB-Brazil *Hevea* germoplasm expedition. Manaus, Brésil, EMBRAPA-CNSPD, 24 p.
- LOW F.C., BONNER J., 1985. Characterisation of the nuclear genome of *Hevea brasiliensis*. In : International rubber conference 1985. Kuala Lumpur, Malaisie, IRRDB, p. 127-136.
- LUO H., BOUTRY M., 1995. Phylogenetic relationships within *Hevea brasiliensis* as deduced from a polymorphic mitochondrial DNA region. Theoretical and Applied Genetics, 91 : 876-884.
- LUO H., VAN COPPENOLLE B., SEGUIN M., BOUTRY M., 1995. Mitochondrial DNA polymorphism and phylogenetic relationships in *Hevea brasiliensis*. Molecular Breeding, 1 : 51-63.
- MOGENSEN H.L., 1996. The hows and whys of cytoplasmic inheritance in seed plants. American Journal of Botany, 83 : 383-404.
- NEI M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, 89 : 583-590.
- NICOLAS D., 1985. Acquisition of *Hevea* material derived from Colombian Schultes collections. In : International rubber conference 1985. Kuala Lumpur, Malaisie, IRRDB.

NICOLAS D., CHEVALLIER M.H., CLEMENT-DEMANGE A., 1988. Contribution to the study and evaluation of new germplasm for use in *Hevea* genetic improvement. In : Colloque exploitation, physiologie et amélioration de l'hévéa, J.L. Jacob et J.C. Prévôts éd., Montpellier, France, Cirad-Irca, p. 335-352.

ONG S.H., 1985. Chromosome morphology at the pachytene stage in *Hevea brasiliensis*: a preliminary report. In : International rubber conference 1985. Kuala Lumpur, Malaisie, IRRDB, p. 3-12.

ONG S.H., 1987. Utilization of *Hevea* genetic resources in the RRIM. Malaysian Applied Biology, 16 : 145-155.

ONG S.H., OTHMAN R., BENONG M., 1995. Status report on the 1981 *Hevea* germplasm collection. In : IRRDB symposium on physiological and molecular aspects of the breeding of *Hevea brasiliensis*. Brickendonbury, Royaume-Uni, IRRDB, p. 95-105.

DE PAIVA J.R., 1981. I coleta de material sexuado a assexuado nos seringais nativos do Estado do Mato Grosso. Manaus, Brésil, EMBRAPA-CNSPD, 26 p.

PALMER J.D., 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. American Naturalist, 130 : S6-S29.

PALMER J.D., HERBON L.A., 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. Journal of Molecular Evolution, 28 : 87-97.

PERNES J., 1984. Gestion des ressources génétiques des plantes. Paris, France, ACCT, 212 p.

PERRIER X., FLORI A., BONNOT F., 1999. Les méthodes d'analyse des données. In : Diversité génétique des plantes tropicales cultivées, P. HAMON *et al.* éd., Montpellier, France, Cirad, collection Repères, p. 43-76.

POLHAMUS L.G., 1962. Rubber: botany, production and utilization. Londres, Royaume-Uni, Leonard Hill, 449 p.

PRANCE G.T., 1987. Biogeography of neotropical plants. In : Biogeography and Quaternary history in tropical America, I.C. Whitmore et G.T. Prance éd., Oxford, Royaume-Uni, Clarendon Press, p. 46-65.

REBOUD X., ZEYL C., 1994. Organelle inheritance in plants. Heredity, 72 : 132-140.

RIVANO F., 1992. La maladie sud-américaine des feuilles de l'hévéa : étude en conditions naturelles et contrôlées des composantes de la résistance partielle à *Microcyclus ulei*. Thèse de doctorat, université Paris XI, Orsay, France, 260 p.

SCHULTES R.E., 1977. Wild *Hevea*: an untapped source of germplasm. Journal of the Rubber Research Institute of Sri Lanka, 54 : 227-257.

SCHULTES R.E., 1990. A brief taxonomic view of the genus *Hevea*. Kuala Lumpur, Malaisie, MRRDB, 57 p.

SEGUIN M., BESSE P., LEBRUN P., CHEVALLIER M.H., 1995a. *Hevea* germplasm characterization using isozymes and RFLP markers. In : Population genetics and genetic conservation of forest trees, P. Baradat *et al.* éd., Amsterdam, Pays-Bas, SPB Academic Publishing, p. 129-134.

SEGUIN M., BESSE P., LESPINASSE D., LEBRUN P., RODIER-GOUD M., NICOLAS D., 1995b. Characterization of genetic diversity and *Hevea* genome mapping by biochemical and molecular markers. In : IRRDB symposium on physiological and molecular aspects of the breeding of *Hevea brasiliensis*. Brickendonbury, Royaume-Uni, IRRDB, p. 19-30.

SEGUIN M., BESSE P., LESPINASSE D., LEBRUN P., RODIER-GOUD M., NICOLAS D., 1996. *Hevea* molecular genetics. Plantations, recherche, développement, 3 : 77-88.

SEGUIN M., LESPINASSE D., RODIER-GOUD M., LEGNATE H., TROISPOUX V., PINARD F., CLEMENT-DEMANGE A., 1998. Genome mapping in connection with resistance to the South American leaf blight in rubber tree (*Hevea brasiliensis*). In : IIIrd ASAP conference on agricultural biotechnology. Bangkok, Thaïlande, Biotech, volume 1.

SEGUIN M., RODIER-GOUD M., LESPINASSE D., 1997. Mapping SSR markers in rubber tree (*Hevea brasiliensis*) facilitated and enhanced by heteroduplex formation and template mixing. In : Plant and animal genome V, D. Bigwood *et al.* éd., Washington, Etats-Unis, USDA, poster n° 61, p. 66.

SEIBERT R.J., 1948. The use of *Hevea* for food in relation to its domestication. Annals of the Missouri Botanical Garden, 35 : 117-121.

SERIER J.B., 1993. Histoire du caoutchouc. Paris, France, Desjonquières, 273 p.

WOLFE K.H., LI W.H., SHARP P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences of the United States of America, 84 : 9054-9058.



# Le manioc

---

Gérard Second, Jean-Pierre Raffailac,  
Carlos Colombo

**L**e manioc est cultivé pour ses racines qui tubérisent, souvent appelées tubercules. Il entre dans le régime alimentaire d'une partie des pays les plus pauvres de la zone tropicale, particulièrement en Afrique, qui assure plus de la moitié de la production mondiale (figure 1). Les jeunes feuilles prélevées en cours de végétation sont également consommées : elles constituent un complément protéique, vitaminique et minéral appréciable pour les populations d'Afrique centrale et du nord-est brésilien dont l'alimentation est à base de racines de manioc. Les feuilles sont parfois employées comme fourrage.

Le manioc est cultivé le plus souvent dans des systèmes traditionnels, qui font rarement appel aux techniques culturales améliorées. C'est dans les pays asiatiques, qui utilisent le manioc comme culture de rente pour l'exportation et l'industrie, que les innovations techniques, dont l'emploi de variétés améliorées, sont adoptées le plus facilement par les agriculteurs. Le cas le plus typique est celui de la Thaïlande, qui en deux décennies s'est placée parmi les principaux producteurs mondiaux aux côtés du Brésil, du Nigeria, du Zaïre et de l'Indonésie.

Le matériel de plantation employé est une bouture de tige d'environ 20 centimètres, mais la domestication a sans aucun doute impliqué l'utilisation de plantes issues de la reproduction sexuée selon des techniques qui sont encore

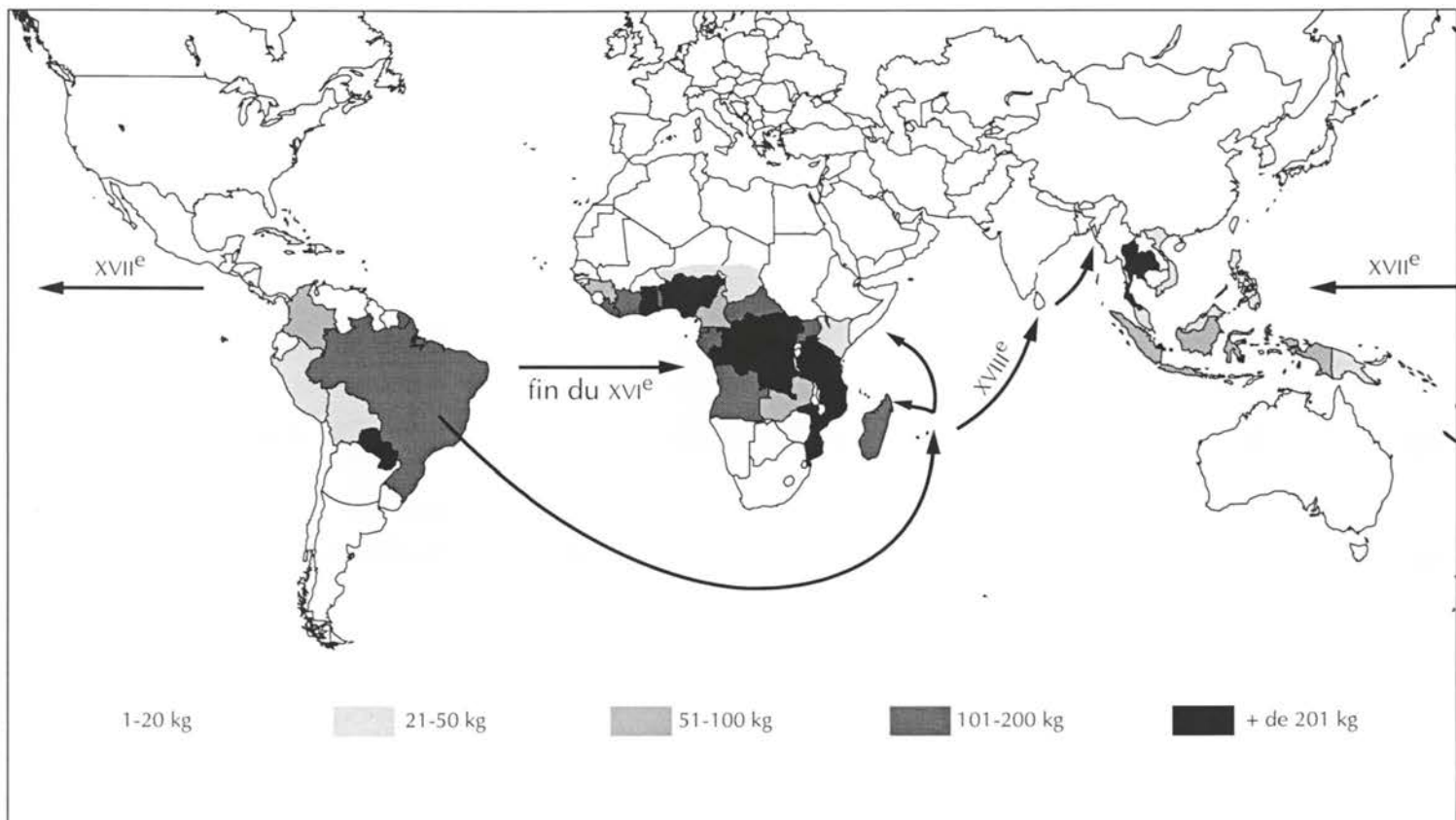


Figure 1. Production de manioc, d'après les données de la FAO (en kilos par an et par habitant, moyenne sur 3 ans : 1995, 1996 et 1997). Les flèches indiquent la diffusion du manioc à partir de son centre d'origine, le Brésil, et aussi probablement du Mexique (PURSEGLOVE, 1992).

observées en Afrique et en Amazonie (EMPERAIRE *et al.*, 1998). L'emploi des graines est réservé à l'amélioration variétale, mais des recherches sur l'amélioration des techniques pour leur utilisation directe en culture sont entreprises (IGLESIAS *et al.*, 1994).

La greffe en début de cycle d'un manioc sauvage à fort développement végétatif et résistant aux maladies foliaires, tel que *Manihot glaziovii*, sur un manioc cultivé porte-greffe permet de tripler le rendement individuel. C'est le système *mukibat*, facile à réaliser sur de petites superficies, décrit en Indonésie par DE BRUIJN et DHAMAPUTRA (1974).

Le manioc préfère des sols légers et bien drainés à dominante sableuse, mais il peut s'accommoder de sols plus lourds, argileux, s'ils sont ameublés. La fertilisation minérale reste encore peu pratiquée. Les réponses aux différents engrais sont cependant bien connues : l'azote favorise le développement des parties aériennes, parfois au détriment de la tubérisation, tandis que la fertilisation potassique augmente fortement les rendements (HOWELER, 1990). La présence d'endomycorhizes sur les racines améliore la nutrition phosphorée.

Les contraintes phytosanitaires majeures sont la mosaïque africaine — maladie à géminivirus transmis par la mouche blanche *Bemisia* —, la bactériose vasculaire, dont l'agent est *Xanthomonas campestris* pv. *manihoti*, et les attaques de cochenilles farineuses (*Phenacoccus manihoti*) et d'acariens (*Mononychellus tanajoa*). Elles sont importantes en Afrique.

La durée du cycle cultural varie selon les variétés et les facteurs du milieu — température, longueur de la saison sèche, ensoleillement... — de six mois, pour les variétés très précoces, à trois ans et plus, lorsque les conditions climatiques sont peu favorables. Dans certains systèmes de production, la récolte partielle d'un ou de plusieurs tubercules sur le plant en place est pratiquée pour une consommation domestique.

Le manioc est une source d'amidon de bonne qualité, qui se prête aux utilisations les plus variées : alimentation humaine de base, pâtisserie moderne, fabrication traditionnelle de bière, utilisations industrielles de l'amidon, alimentation animale. Il se place au cinquième rang mondial des productions végétales alimentaires avec 166 millions de tonnes (FAO, 1998). Cependant, son amélioration génétique n'en est qu'à ses débuts (MCKEY et BECKERMAN, 1996 ; RAFFAILLAC et SECOND, 1997).

La racine de manioc produit de l'acide cyanhydrique en quantités variables selon la variété et le milieu. Certaines variétés peuvent être dangereuses sur le plan nutritionnel si des techniques de transformation ne sont pas mises en œuvre pour en éliminer la plus grande partie. La cyanogénèse est connue comme un système de défense chimique chez de nombreuses plantes (JONES, 1998).



# La botanique et la taxonomie du genre *Manihot*

## La morphologie

Les plants de manioc provenant de boutures développent simultanément une ou plusieurs tiges principales à partir des bourgeons, alors que les plants issus de graines ont une tige unique. L'appareil végétatif du manioc se caractérise par deux types de ramification orthotrope (MEDARD *et al.*, 1992). Les ramifications proleptiques, ou rameaux secondaires, sont issues du développement de bourgeons latéraux après la levée de la dominance apicale. Les ramifications sylleptiques, ou rameaux de floraison, proviennent de la transformation du méristème végétatif en méristème floral : 2 à 4 branches se développent simultanément. L'aptitude à fleurir, qui détermine l'architecture de la plante à la fin du cycle, est l'un des critères d'identification des variétés. Dans la figure 5 (en annexe), qui présente les principaux caractères morphologiques d'identification des variétés, le port final de la tige est illustré par quelques exemples tirés des travaux de COURS (1951).

Les feuilles de manioc sont fortement hétéroblastiques : au cours de la vie de la plante, le nombre de lobes, impair, augmente au début du cycle, puis diminue pour aboutir à un lobe unique à la fin du cycle chez certaines variétés. La morphologie de la feuille est l'un des critères d'identification des variétés (tableau 1, en annexe). COURS (1951) souligne que seule la comparaison entre variétés plantées en un même lieu et dans les mêmes conditions peut avoir une valeur.

Le système racinaire se compose de racines nodales et de racines basales, plus nombreuses. Ces dernières donnent les plus gros tubercules. Le poids des tubercules d'un même plant est souvent hétérogène, mais leur teneur en matière sèche est identique. Toutes les racines sont susceptibles d'évoluer en tubercules. Le pourcentage des racines qui tubérisent varie considérablement selon les variétés et dépend largement des facteurs du milieu et des techniques culturales. Un plant issu de graine émet une racine séminale pivot, qui tubérise en premier, mais qui présente l'inconvénient de contenir une proportion de fibres souvent trop forte pour être facilement consommée. Il produit ensuite des racines secondaires, rattachées au pivot, qui évoluent en tubercules.

Sur le plan physiologique, le manioc est une plante en  $C_3$ , qui possède cependant quelques caractéristiques de plante en  $C_4$  (EL-SHARKAWY et COCK, 1990). La tubérisation des racines se déclenche au cours des premières semaines de croissance. Le nombre de tubercules est fixé entre le 2<sup>e</sup> et le 4<sup>e</sup> mois du cycle, sauf accident végétatif. Il ne dépend pas de la photopériode (KEATING *et al.*, 1985).

Le port des espèces sauvages varie d'une forme herbacée en rosette jusqu'à un port d'arbre de plus de 15 mètres de hauteur, en passant par des formes buissonnantes, de fausses lianes et d'arbustes.

## La reproduction

Toutes les espèces du genre *Manihot* sont pérennes, monoïques et allogames. Leur pollinisation est entomo-anémophile. La floraison peut être très précoce et fréquente au cours de la croissance ou, à l'inverse, tardive et unique, voire inexistante. Généralement les premiers axes floraux avortent. Les inflorescences sont des grappes de 20 à 60 fleurs unisexuées et monopérianthées. Les fleurs femelles, situées à la base de l'inflorescence, s'épanouissent avant les fleurs mâles, qui sont 10 à 20 fois plus nombreuses (COURS, 1951). Pour chaque inflorescence, de un à six fruits, ou capsules, deviennent matures en trois à cinq mois. Ils éclatent en période sèche et libèrent trois graines au tégument marbré, marron à gris, avec une caroncule.

## La taxonomie et la répartition du genre *Manihot*

Le genre *Manihot*, distribué naturellement dans toute la zone tropicale du Nouveau Monde, se rattache à la famille des euphorbiacées, qui est divisée en cinq sous-familles. Il est placé dans celle des Crotonoïdées avec l'hévéa (WEBSTER, 1975).

Les limites du genre sont bien définies, mais les limites entre les espèces sont très difficiles à établir. La dernière monographie (ROGERS et APPAN, 1973 ; ROGERS et FLEMING, 1973) réduit le nombre d'espèces de 171 à 98 sur la base de l'analyse multivariée de 44 caractères végétatifs observés sur herbier, dans laquelle les caractéristiques des feuilles sont importantes. La distribution des espèces est remarquablement indiquée. Dix-sept sections sont décrites, dont l'une ne comprend que le manioc cultivé. Celui-ci est proche de deux sections : la section d'Amérique centrale, *Parvibracteatae*, qui inclut l'espèce *M. aesculifolia*, considérée alors comme l'espèce la plus apparentée au manioc, et la section d'Amérique du Sud, *Heterophyllae*, qui comprend la plupart des taxons maintenant rassemblés dans l'espèce *M. esculenta* par ALLEM (1994), tels que le manioc, subsp. *esculenta*, et les formes sauvages, subsp. *flabellifolia* et subsp. *peruviana*, considérées par cet auteur comme les ancêtres directs du manioc.

Des 77 espèces reconnues, présentes au Brésil et décrites par ROGERS et APPAN (1973), seules 38 sont retenues comme non synonymes par A.C. ALLEM (comm. pers.) ; elles sont classées dans 16 groupes d'affinité. En comparant ces deux classifications, on constate de grandes différences en ce qui concerne les regroupements par affinité, ce qui suggère d'énormes difficultés pour définir

des discontinuités nettes au sein du genre, tout du moins dans sa zone de diversité principale (SECOND *et al.*, 1997).

Dans sa zone d'origine, le genre *Manihot* s'étend du sud de l'Arizona, aux Etats-Unis, au nord de l'Argentine. Les espèces ne sont jamais dominantes dans la végétation mais plutôt sporadiques. La plupart d'entre elles se trouvent dans des régions à longue saison sèche, mais certaines se rencontrent aussi dans les forêts humides ou, plus souvent, en bordure et dans les clairières, naturelles ou de déforestation. Certaines espèces ont aussi une distribution d'adventices, dans des zones perturbées, le long des routes, par exemple.

La distribution naturelle du genre ne s'étend pas au-delà de 2 000 mètres d'altitude, mais quelques espèces se rencontrent en zones subtropicales. Une seule espèce, *M. brachyloba*, a une distribution à cheval sur l'Amérique centrale et l'Amérique du Sud et se retrouve aussi en République dominicaine. Cette distribution exceptionnelle pourrait être liée à sa capacité à être transportée dans l'eau modérément salée : des graines ayant conservé leur pouvoir germinatif peuvent être récoltées sur la côte de la Guyane française, où l'on observe une influence forte de l'Amazone et des fleuves locaux (D. Loubry, comm. pers.).

L'introduction du manioc dans l'Ancien Monde date du xvi<sup>e</sup> siècle et celle de *M. glaziovii*, de la fin du xix<sup>e</sup> siècle. *M. glaziovii*, le *manicoba*, a fait l'objet d'une tentative de développement en Afrique dans les années 30 pour la production de latex et il y a subsisté pour d'autres raisons (LEFEVRE, 1989 ; SERIER, 1989).

Depuis les travaux de ROGERS et APPAN (1973), la distinction entre *M. utilisima*, variétés amères, et *M. dulcis*, ou *M. aipi*, variétés douces, est rejetée, et ce d'autant que plusieurs facteurs du milieu ont un impact significatif sur la présence des deux glucosides responsables de la production d'acide cyanhydrique, et donc sur l'amertume.

## La structure du génome

La vingtaine d'espèces de *Manihot* observées possèdent le même nombre de chromosomes que le manioc, soit 18 paires (LEFEVRE, 1989 ; BAI *et al.*, 1993), et il est probable qu'il en est ainsi pour toutes les espèces du genre. A partir de l'examen des chromosomes au pachytène, MAGOON *et al.* (1969) ont suggéré que le manioc cultivé était un allopolyploïde segmentaire. L'observation des proportions de locus dupliqués, tant pour les RFLP que pour les microsatellites, ne corrobore pas cette hypothèse (CHAVARRIAGA-AGUIRE *et al.*, 1998). Cette allopolyploidie présumée pourrait cependant être très ancienne et correspondre au genre entier et non à l'espèce cultivée en particulier. La taille du génome du manioc est petite, elle est estimée à 1,68 picogramme par génome diploïde (MARIE et BROWN, 1993).

## Les ressources génétiques

L'ensemble du genre *Manihot* peut être considéré comme recelant des ressources génétiques utilisables pour l'amélioration du manioc par voie sexuée. En effet, aucune barrière forte à l'hybridation n'est connue. De plus, des hybridations interspécifiques ont été fréquemment obtenues ou observées dans la nature. La plus grande partie des collections *ex situ* est néanmoins constituée de cultivars conservés généralement dans des collections maintenues au champ et reconduites chaque année végétativement. Une partie de ces collections de cultivars est en outre maintenue *in vitro*. L'utilisation des graines pour la conservation et les échanges concerne principalement des hybrides et des espèces sauvages récemment récoltées.

On estime à près de 25 000 le nombre d'accessions, toutes espèces confondues, conservées en collection dans le monde. Mais ce nombre est probablement surestimé en raison de la perte récente de plusieurs collections et de la duplication de certaines autres. Actuellement, les deux principales collections sont celles du Brésil et du CIAT (Centro Internacional de Agricultura Tropical), en Colombie. Un ensemble de réseaux, formels et informels, travaille à la conservation et à l'utilisation des ressources génétiques du manioc, par continent ou à l'échelle mondiale (SECOND *et al.*, 1998b). Une très faible proportion de ce matériel génétique a été utilisée en croisement pour l'amélioration variétale.

Dans une bibliographie de travaux antérieurs à 1973 (CAB, 1974), on relève l'existence de collections anciennes au Congo belge (INEAC, Institut national pour l'étude agronomique du Congo belge), en Afrique de l'Ouest et en Afrique centrale (IRAT, Institut de recherches agronomiques tropicales et des cultures vivrières), à Madagascar (IRAM, Institut de recherches agronomiques à Madagascar), au Kenya (EAAFRRO, East African Agricultural and Forestry Research Organization), en Tanzanie, en Indonésie, en Inde (CTCRI, Central Tuber Crops Research Institute), en Malaisie (MARDI, Malaysian Agricultural Research and Development Institute), au Costa Rica (IICA, Instituto Interamericano de Ciencias Agrícolas), au Venezuela, au Mexique, au Brésil, en Colombie, en Argentine... Les échanges se faisaient parfois sans contrôle sanitaire rigoureux. Chaque station de recherche employait des techniques de gestion de la collection différentes — mode de plantation, qualité des boutures, durée du cycle, fertilisation, densité de plantation... — et utilisait sa propre classification des variétés.

Depuis le début des années 80, l'IITA (International Institute of Tropical Agriculture), au Nigeria, tente de superviser les activités de plusieurs pays en Afrique et collabore, à ce titre, avec le CIAT. C'est ainsi qu'une grande partie de la collection de manioc du CIAT a été transférée au Nigeria. De son côté, la Thaïlande a fait de même pour disposer de la forte diversité des clones de la collection du CIAT.

En ce qui concerne l'intérêt des ressources génétiques sauvages pour l'amélioration du manioc, on peut citer : les sources de résistance aux maladies et aux parasites — l'espèce *M. glaziovii* en est un bon exemple car elle a été largement utilisée pour la recherche de résistance à la bactériose et à la mosaïque virale africaine (HAHN, 1984) — et aux stress tels que les basses températures ; une architecture des cultivars dans un milieu donné et une morphologie des feuilles adaptées aux contraintes culturales comme la concurrence avec les adventices et la culture associée ; les teneurs en amidon et ses propriétés physico-chimiques, qui sont d'une grande importance pour son utilisation mais n'ont été que peu étudiées chez les espèces sauvages.

Dans ce chapitre, nous tenterons de montrer comment l'utilisation des marqueurs moléculaires complète les études effectuées antérieurement sur une base morphologique et physiologique. Elle permet, entre autres, de préciser l'origine génétique du manioc : parent sauvage le plus direct, autres espèces proches parentes et autres espèces ayant participé à la constitution génétique du manioc par introgression. Elle offre les moyens de comprendre le processus de domestication et de gestion de la diversité génétique telle qu'on peut l'observer en milieu traditionnel. Enfin, elle permet de saisir l'ampleur et la structure de la diversité génétique des cultivars et de l'ensemble du genre *Manihot*, dont le centre principal de diversité, au Brésil, sera particulièrement étudié.

## L'organisation de la diversité

### La variabilité agromorphologique

Les agriculteurs traditionnels, et en particulier les Amérindiens, distinguent un très grand nombre de cultivars (EMPERAIRE *et al.*, 1998). Ils utilisent plusieurs critères de classification des clones : le goût, lié à la concentration en glucosides cyanogéniques dans le cortex des racines, pour laquelle il existe un continuum de variation allant de concentrations très basses à des concentrations extrêmement toxiques (MCKEY et BECKERMAN, 1996) ; la précocité — les variétés dites à cycle court sont récoltées entre 6 et 8 mois alors que d'autres nécessitent un cycle supérieur à 18 mois — ; l'aspect de la plante — le port de la partie aérienne, la coloration des feuilles, des tiges et des racines, leur forme...

On dispose pour le manioc d'un grand nombre de marqueurs morphologiques et chimiques, dont la description la plus remarquable, et qui a inspiré les travaux postérieurs, est celle de COURS (1951). Cet auteur propose un classement des variétés présentes à Madagascar en huit sections botaniques. Les caractéristiques

tères utilisés par G. Cours et dans des travaux plus récents sont indiqués dans le tableau 1 (en annexe). GULICK *et al.* (1983) les ont rassemblés dans un document de l'IBPGR (International Board for Plant Genetic Resources) et ont proposé une liste universelle. On dispose ainsi d'un catalogue exhaustif, qui rassemble les descripteurs classiques utilisables pour l'identification des variétés cultivées de manioc.

La classification des variétés de manioc établie par ROGERS et APPAN (1973) à partir de l'analyse multivariée de 44 caractères conduit à la mise en synonymie de toutes les dénominations spécifiques ou subs spécifiques du manioc, *M. utilisissima* devient *M. esculenta*. Elle aboutit à la constitution de 19 « groupes de similarité », destinés à faciliter le travail du sélectionneur, qui se sont néanmoins avérés d'utilisation difficile. En effet, en dehors des facteurs génétiques, le manioc cultivé est soumis à deux types de facteurs de variation importants : l'état de la bouture, sa coupe et son mode de plantation ; la fertilité du milieu et les techniques appliquées pour la gestion de la parcelle. Ces facteurs compliquent l'évaluation agromorphologique et la comparaison de cultivars qui ont poussé dans des environnements distincts ou au cours d'années différentes.

#### LA PLASTICITÉ AGROMORPHOLOGIQUE EN RELATION AVEC LE BOUTURAGE

La qualité de la bouture joue sur la structuration du peuplement végétal. C'est d'elle que dépendent en premier le taux de reprise, donc la densité de peuplement, et le nombre de tiges par plant. Par ailleurs, l'enracinement primaire est lui aussi conditionné par l'état de la bouture, au niveau tant nodal (nombre de nœuds en contact avec le sol humide) que basal (orientation et nombre de racines primaires liés à la nature de la coupe et à la position de la bouture dans le sol).

Par rapport à une bouture prélevée vers le haut de la tige, une bouture de base présente à la fois un meilleur taux de reprise, un début de croissance accéléré, un plus grand nombre de tiges par plant et un meilleur enracinement nodal et basal (RAFFAILLAC, 1992). De même, pour une bouture de base, l'âge de la tige intervient sur la qualité de la bouture. En Côte d'Ivoire, on est passé de 13 900 tiges par hectare avec des boutures de tiges de 6 mois à 27 400 avec des boutures de tiges de 12 mois. Une bouture de 6 mois totalise en moyenne 17,2 racines contre 32,7 pour une bouture de 12 mois (RAFFAILLAC, 1998).

Les tiges coupées à la récolte peuvent être stockées plusieurs semaines avant l'installation d'une nouvelle parcelle. Ce stockage modifie le comportement des jeunes plants : une bouture prélevée sur une tige stockée pendant quatre à huit semaines s'enracine plus rapidement. De plus, le nombre de tiges par plant augmente (RAFFAILLAC et NEDELEC, 1988). Ces quelques exemples montrent qu'il convient donc de procéder à un choix rigoureux des lots de boutures lorsque l'on souhaite étudier les comportements variétaux sur la base de

critères tels que la vitesse de reprise, la vigueur de la plantule au départ, la structuration du peuplement végétal et les potentialités racinaires.

#### LA PLASTICITÉ AGROMORPHOLOGIQUE EN RELATION AVEC LA GESTION DE LA PARCELLE ET LE MILIEU

Le nombre de tiges par plant dépend du mode de plantation de la bouture et des conditions climatiques qui prévalent dans les semaines qui suivent (pluviométrie, rayonnement global). Certaines variétés développent plus de tiges par plant si la plantation est horizontale au lieu d'être oblique. Lorsque le rayonnement global reçu en début de plantation est limité, le nombre de tiges par bouture diminue (RAFFAILLAC, 1998). Néanmoins, pour une même variété, une parcelle de manioc est toujours composée d'un mélange de plants monocaulés et de plants multicaules (à 2 tiges, à 3 tiges...), avec le plus souvent un type dominant. Une plantation oblique combinée à une coupe en biais de la bouture permet de localiser les axes racinaires primaires dans un secteur du sol alors que, à l'inverse, une plantation verticale avec une coupe droite favorise une dispersion radiale des racines autour du plant (RAFFAILLAC, 1992). Le regroupement des racines, et donc des futurs tubercules, dans un même secteur facilite l'arrachage du plant. Il est illusoire de procéder à des recherches sur l'amélioration variétale de ce caractère dans la mesure où les techniques de plantation permettent aisément de le contrôler.

Les variétés de manioc présentent des différences dans l'aptitude à développer des rameaux secondaires, dont la présence reste liée aux facteurs du milieu. Ces rameaux sont plus nombreux quand le sol est riche et le rayonnement élevé. A l'inverse, de faibles écartements entre plants ou la présence simultanée d'autres plantes (adventices ou plantes associées) limitent leur apparition.

La floraison peut être inexistante au cours d'un cycle cultural ou intervenir plusieurs fois en relation avec les facteurs du milieu : ainsi sur des sols carencés en potassium le nombre de floraisons est supérieur à celui qui est observé sur des sols riches (HOWELER, 1990 ; RAFFAILLAC, 1998). Cet effet de la fertilité du sol se retrouve pour la morphologie et la qualité des racines tubérisées. A poids égal, un tubercule obtenu sur un sol carencé en potassium est plus long et plus fin qu'un tubercule issu d'un sol fertilisé (RAFFAILLAC, 1997). La fertilisation potassique permet d'augmenter fortement le rendement, mais la teneur en matière sèche des racines tubérisées diminue. La pluviométrie reçue pendant les dernières semaines qui précèdent l'arrachage du manioc modifie également la qualité des tubercules : la teneur en amidon est d'autant plus faible que la quantité d'eau reçue est élevée (RAFFAILLAC, 1985).

La plasticité des espèces sauvages est également connue. Nous pouvons en rappeler deux exemples. Le premier a trait aux formes sauvages de *M. esculenta*, qui présentent naturellement deux types biologiques selon qu'elles croissent en forêt ou en savane (ou sur la défriche de forêts). Dans le premier cas,



elles poussent comme des fausses lianes, qui s'appuient sur les branches des arbres. Dans le second cas, elles prennent un port d'arbuste. Des graines prélevées sur la forme en liane et cultivées en pleine lumière ou des graines ayant germé sur défriche de forêt donnent des arbustes sans traces de port en liane. Le second exemple concerne l'espèce *M. quinquepartita*, qui forme des feuilles entières ou fortement indentées selon qu'elle pousse en pleine lumière ou à l'ombre, comme on peut le constater à Saül, en Guyane française. Les différences morphologiques sont telles que, en herbier, les deux formes peuvent être considérées comme des espèces différentes, même par un expert.

## La diversité moléculaire

### L'ANALYSE PAR RAPD DES MANIOCS CULTIVÉS

L'analyse par RAPD a porté sur 126 accessions (COLOMBO, 1997). Celles-ci se répartissent en 71 accessions issues des collections de l'IAC (Instituto Agronômico de Campinas) et de l'EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) et originaires des différentes zones géographiques du Brésil, en 33 accessions représentant une collection *core de core* de la diversité mondiale disponible au CIAT (HERSHEY *et al.*, 1994) et en 42 plantes représentant 18 ethnovariétés en provenance d'un seul « abattis » du moyen rio Negro dans le bassin amazonien. Cet échantillon peut être considéré comme représentatif : le Brésil recouvre probablement une bonne part de la zone d'origine du manioc et une majorité de cultivars y trouvent leur source, mais ses collections sont déficitaires en origines amazoniennes. L'échantillon reste cependant limité, avec seulement 7 accessions d'Amérique centrale et de la Caraïbe et 6 accessions non américaines.

A partir de 21 amorces, 193 bandes RAPD ont été observées, dont 88 sont polymorphes en présence-absence, ce qui indique une forte diversité de l'espèce. Le coefficient de similarité de Jaccard varie de 0,99 à 0,45 avec une moyenne de 0,67 (les bandes polymorphes sont seules prises en compte). La distribution des 126 génotypes sur les deux premiers axes factoriels d'une analyse des coordonnées principales (AFTD, analyse factorielle sur tableau de distances) est donnée sur la figure 2a. Le test de Lebart (LEBART *et al.*, 1977) sur une analyse factorielle des correspondances (AFC) a été utilisé pour caractériser les bandes qui contribuent de manière significative aux premiers axes : 46 bandes, soit 52 % seulement, contribuent aux cinq premiers axes de l'AFC.

La variation est continue et aucune structure forte n'apparaît. Cependant, les ethnovariétés issues de l'abattis amazonien sont séparées du reste de la collection par le premier axe. Les autres variétés qui s'en rapprochent le plus sont originaires de la région amazonienne du Brésil, de pays qui partagent le bassin amazonien (Venezuela, Colombie) ou de pays qui possèdent des régions à climat similaire, comme la Malaisie et la Thaïlande.



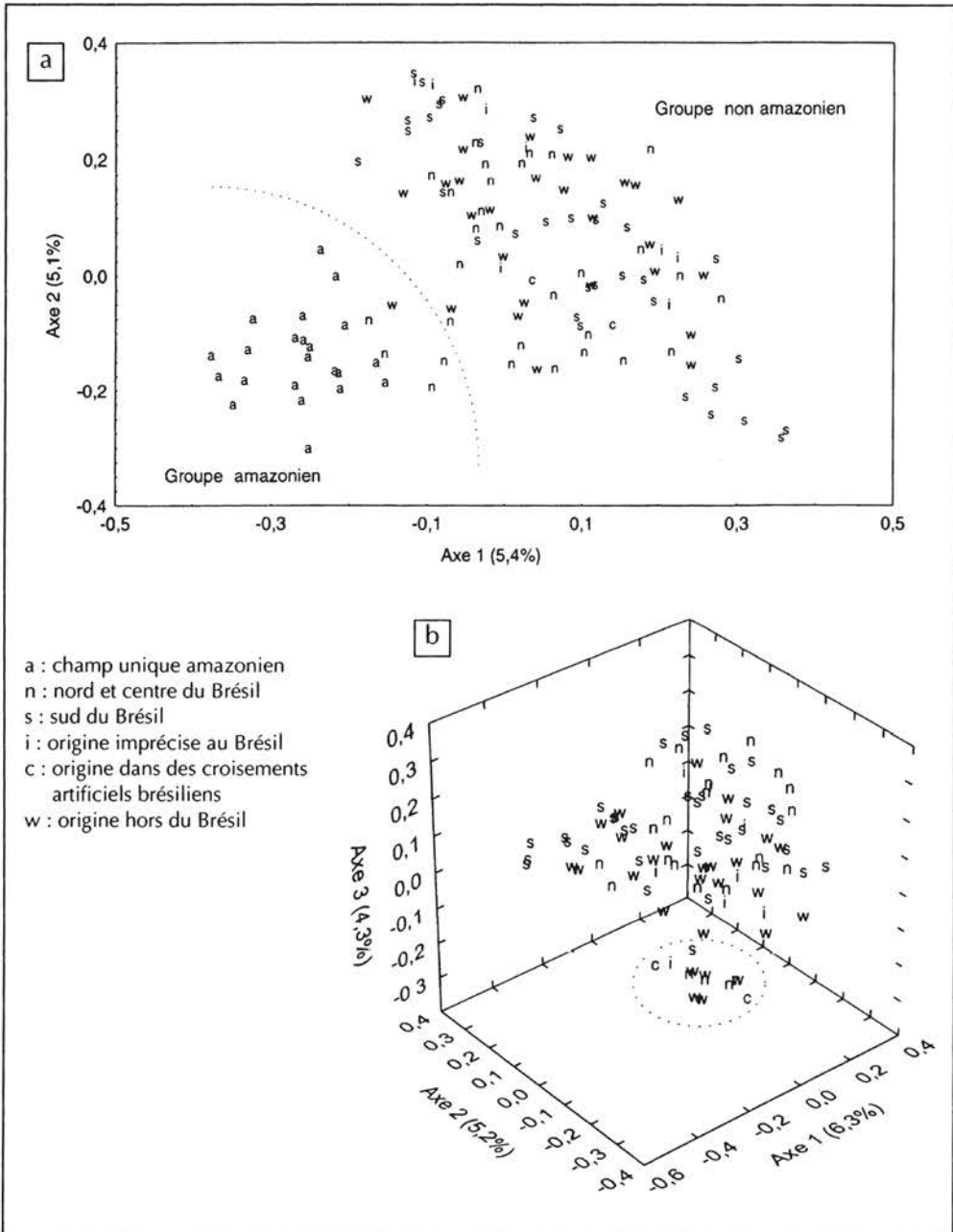


Figure 2. Analyse en coordonnées principales de la matrice de similarité (Jaccard) fondée sur la présence ou l'absence de 88 fragments RAPD polymorphes dans les comparaisons de génotypes deux à deux : plan ou volume définis par les 2 ou 3 premiers axes. a : distribution de 126 génotypes représentant une collection mondiale et un champ unique amazonien (les pointillés indiquent la séparation d'un groupe amazonien du reste de la collection). b : analyse du groupe non amazonien seul (le cercle pointillé indique un groupe remarquable).

L'AFTD réalisée sur un groupe « non amazonien » (figure 2b), constitué en éliminant de l'échantillon les variétés de l'abattis et celles qui s'en rapprochent le plus, met en évidence, à partir des trois premiers axes, un ensemble de 13 variétés, qui présentent quelques particularités remarquables. Leurs origines sont variées : 6 sont originaires du Brésil (dont 4 du nord-est brésilien), 2 sont les croisements artificiels brésiliens inclus dans l'échantillon et 5 proviennent de divers pays. Parmi les variétés originaires du nord-est brésilien, la variété SRT1316 présente une affinité remarquable avec des hybrides interspécifiques entre le manioc et *M. glaziovii*. Parmi les variétés originaires de divers pays, l'une, NGA2 a été introgressée artificiellement par *M. glaziovii* pour acquérir une résistance à la mosaïque virale africaine et à la bactériose : c'est l'un des parents du croisement utilisé pour construire la première carte génétique du manioc (FREGENE *et al.*, 1997). Une autre, Col 22, est une variété traditionnelle, qui entre dans la généalogie de plusieurs variétés récemment sélectionnées (voir tableau 2 dans SECOND *et al.*, 1998b). L'hypothèse selon laquelle une introgression de *M. glaziovii* serait à l'origine du regroupement de ces variétés mérite confirmation.

D'autres particularités peuvent être décelées par l'examen de certains axes, qui, malgré les faibles valeurs de variance totale extraites, se révèlent significatifs. Ainsi, les faibles valeurs de l'axe 4 (3,7 %) isolent les quatre variétés provenant de Thaïlande et de Malaisie, les variétés du Nouveau Monde qui leur sont les plus proches étant originaires de Colombie.

#### L'ANALYSE DES ETHNOVARIÉTÉS AMAZONIENNES ET D'UNE COLLECTION DE CULTIVARS PAR LES MARQUEURS AFLP ET RAPD

La collection des ethnovariétés issue de l'abattis permet d'aborder la question de la nature d'une ethnovariété reconnue par un nom. Il s'agit de savoir si la multiplicité des noms rapportée par les anthropologues recouvre une diversité génétique réelle et si un nom identifie un clone ou une famille de clones. Plus généralement, il s'agit de vérifier l'hypothèse d'une gestion dynamique traditionnelle de la diversité du manioc, qui pourrait nous renseigner sur le processus de sa domestication (EMPERAIRE *et al.*, 1998 ; SECOND *et al.*, 1998).

Les marqueurs AFLP ont été utilisés pour analyser les 42 plantes représentant les 18 variétés échantillonnées dans le même abattis, à raison de 1, 2 ou 10 plantes par variété, et 40 accessions d'origine géographique distincte, provenant de la collection mondiale et représentant la variabilité décrite précédemment. Les deux combinaisons d'amorces employées ont permis d'observer 60 bandes polymorphes sur un total de 132. L'analyse révèle une structuration de la diversité similaire à celle qui est observée sur les mêmes échantillons à l'aide des marqueurs RAPD. Une analyse globale a donc été réalisée en combinant les bandes polymorphes observées avec les deux techniques, soit 143 bandes.

Le dendrogramme UPGMA des 82 plantes étudiées, établi grâce au coefficient de similarité de Jaccard (figure 3), révèle la présence de deux groupes. L'un

correspond à la collection mondiale, l'autre à l'abattis. Une seule variété de la collection mondiale se retrouve dans le groupe des plantes de l'abattis ; elle est originaire de l'Etat brésilien du Pará dans le bassin amazonien. Les deux groupes, dont l'origine géographique est très différente, ont une diversité totale sensiblement équivalente. Seules 10 bandes (7 % du total), dont la fréquence est inférieure à 0,45 dans la collection mondiale, ne sont pas observées parmi les variétés de l'abattis. On trouve un coefficient de corrélation significatif de 0,7 entre les fréquences de bandes observées dans l'abattis et dans la collection mondiale. Une analyse de variance moléculaire (Amova) révèle que 80 % de la diversité génétique totale se trouve à l'intérieur des groupes et 20 %, entre les deux groupes.

Tous les noms correspondent à des génotypes différents et, en général, les plantes portant le même nom ne peuvent être différenciées : elles représentent probablement des clones. Il y a cependant quatre exceptions pour lesquelles un même nom — F, Mn, Bu et Sn sur la figure 3 — correspond à plusieurs génotypes. Les variétés Fino (F) et Manipeba (Mn) sont représentées chacune par deux plantes qui correspondent à deux génotypes, avec respectivement 16 et 26 bandes différentes sur les 143 étudiées. La variété Buia (Bu) est représentée par trois génotypes, qui diffèrent par 31, 38 et 40 bandes ; malgré la dispersion des génotypes Buia dans le dendrogramme de la figure 3, une AFTD, non présentée, confirme une tendance au regroupement des génotypes portant le même nom par rapport à la diversité totale du champ. Les trois variétés « sans nom » (Sn) correspondent chacune à un génotype, ce qui est en accord avec leur origine présumée : elles proviendraient de trois plantes issues de graines, en cours d'évaluation par le cultivateur (EMPERAIRE *et al.*, 1998) ; leur regroupement relatif concorde avec celui des génotypes observés respectivement dans les trois variétés F, Mn et Bu.

Ainsi, ces résultats préliminaires confirment l'existence d'une gestion dynamique de la diversité du manioc : regroupement dans un même champ d'une diversité considérable, recombinaison liée à la reproduction sexuée, sélection de génotypes performants rassemblés ensuite selon leurs affinités sous un même nom.

Dans cette optique, une variété traditionnelle peut donc être considérée comme une famille de clones partageant certaines caractéristiques d'intérêt direct pour les cultivateurs, telles que la productivité, la qualité et la résistance aux maladies, aux parasites et aux stress divers, mais aussi des caractères d'intérêt moins immédiat mais qui permettent de reconnaître les variétés et de leur attribuer un nom, comme la couleur, la forme, l'aspect général de la plante... Ces caractéristiques ne suffisent pas à elles seules à la caractérisation d'une variété, qui n'est possible que par une évaluation allant jusqu'à la consommation.

Pour l'échantillon représentatif de la collection mondiale, les analyses par AFLP ne font que confirmer — mais avec un nombre plus réduit d'accèsions —

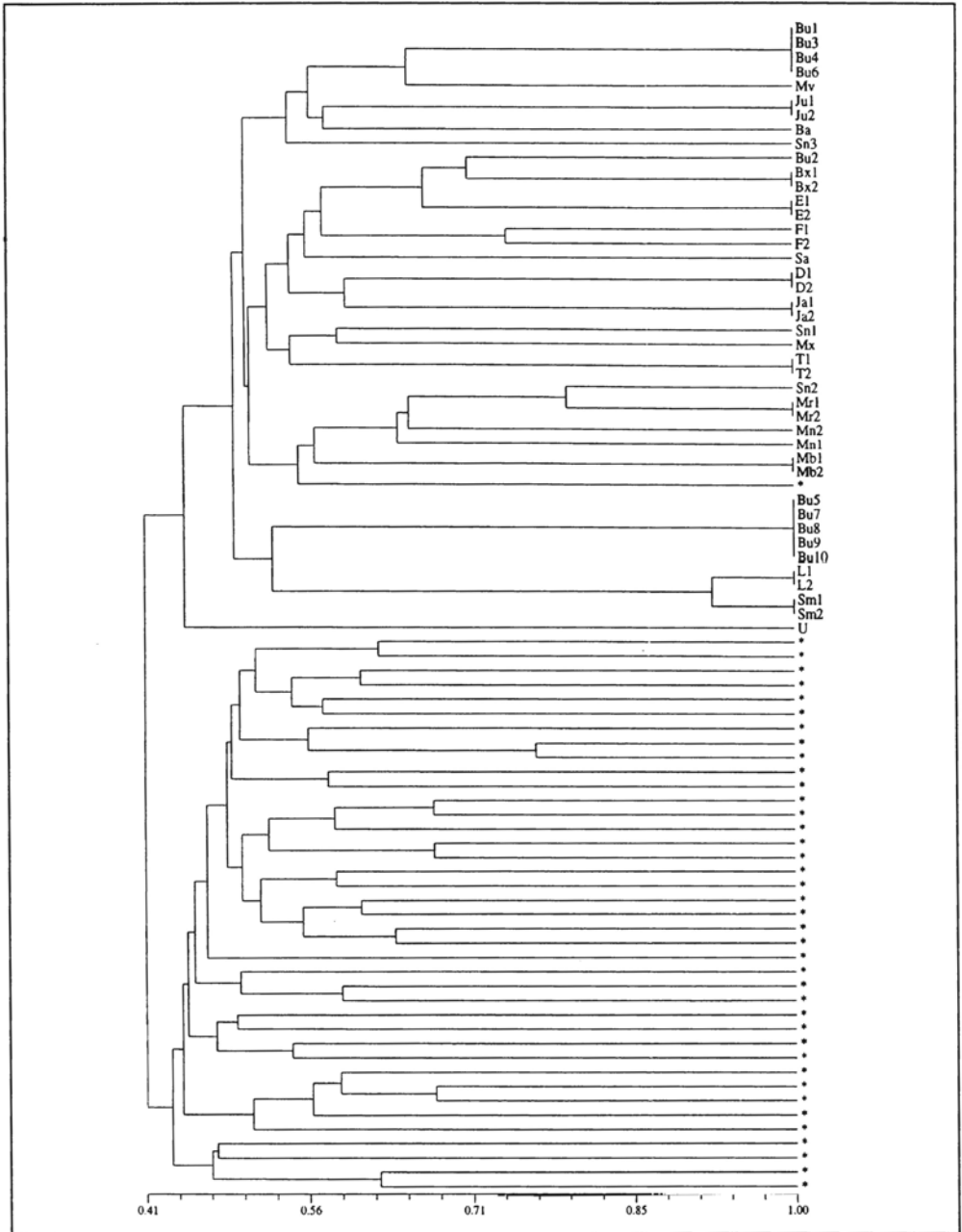


Figure 3. Diversité moléculaire comparée des 42 plantes issues du même champ amazonien (les abréviations représentent les noms de variétés, sauf Sn qui représente des variétés sans nom, les chiffres sont des numéros d'ordre de plantes de même nom) et des 40 plantes représentant les collections mondiales (symbolisées par un astérisque). Dendrogramme UPGMA à partir de la matrice de similarité calculée pour 143 bandes polymorphes (60 AFLP et 83 RAPD).

les résultats obtenus avec les analyses par RAPD, à savoir, l'absence de structuration forte d'une diversité globale pourtant considérable. Compte tenu du regroupement de certains génotypes sur une base géographique, il nous semble cependant qu'une étude comportant un grand nombre de génotypes et portant sur plus de locus pourrait permettre de préciser la faible structuration que l'on peut pressentir.

#### L'ANALYSE DE LA DIVERSITÉ DU GENRE *MANIHOT*

La première caractérisation moléculaire du genre *Manihot* est celle de BERTRAM (1993). Fondée sur une analyse par RFLP (sur *Southern blots*) des ADN chloroplastiques et ribosomiques, elle se limite aux espèces d'Amérique centrale, au manioc et à une espèce d'Amérique du Sud, *M. carthaginensis*. Les résultats montrent un degré relativement faible de divergence (0,1 % au maximum pour l'ADN chloroplastique) et une divergence maximale entre les espèces d'Amérique centrale, d'une part, et *M. carthaginensis* et le manioc, d'autre part. Ces résultats sont en désaccord avec l'hypothèse d'un apparentement maximal du manioc avec *M. aesculifolia*, proposée par ROGERS et APPAN (1973). En revanche, ils suggèrent que l'Amérique du Sud pourrait être la zone d'origine du manioc. L'étude de quelques espèces d'Amérique du Sud et du Centre par RAPD (COLOMBO, 1997) confirme les conclusions de BERTRAM (1993).

Les études par AFLP (ROA *et al.*, 1997) et par microsatellites (ROA *et al.*, 1998) du manioc, de quatre espèces sauvages d'Amérique du Sud — formes sauvages de *M. esculenta*, *M. tristis* (une espèce affine de *M. esculenta* subsp. *flabellifolia*), *M. carthaginensis* et *M. brachyloba* — et de *M. aesculifolia* confirment l'hypothèse de ALLEM (1994) selon laquelle les formes sauvages de *M. esculenta* seraient les ancêtres directs du manioc. De plus, elles montrent que la diversité de ces formes sauvages est plus importante que celle du manioc. Dans le cas des microsatellites, le nombre d'allèles par locus varie de 4 à 22 avec un total de 124 allèles pour les 10 locus analysés (un ou deux des locus n'étaient pas détectés chez *M. aesculifolia*, *M. brachyloba* et *M. carthaginensis*). Le groupe d'espèces formé par *M. aesculifolia*, *M. brachyloba* et *M. carthaginensis* (espèce affine de *M. glaziovii*) présente moins de 20 % de similarité avec le manioc. *M. esculenta* subsp. *flabellifolia* et subsp. *peruviana* et *M. tristis* forment un groupe plus proche du manioc, avec 35 à 50 % de similarité. Il est intéressant de noter, en rapport avec le brassage de la diversité génétique que nous avons observé dans la gestion traditionnelle du manioc, que les espèces sauvages présentent un déficit d'hétérozygotes alors que le manioc cultivé a des taux d'hétérozygotie observée équivalents aux valeurs attendues.

Notre étude, fondée principalement sur les marqueurs AFLP et conduite à la même époque que la précédente — indépendamment mais avec certains échantillons communs — est, à ce jour, la plus exhaustive du genre *Manihot*. Elle inclut toutes les espèces brésiliennes, quatre espèces d'Amérique centrale

et concerne au total 282 formes sauvages et 82 variétés de manioc. Les analyses ont porté sur la présence ou l'absence de 93 bandes polymorphes révélées à partir d'un seul couple d'amorces.

Une analyse par UPGMA des similarités (*simple matching*) montre que le génotype le plus divergent de l'ensemble est celui qui représente l'espèce centraméricaine *M. aesculifolia*. La distribution des 364 accessions dans le plan défini par les deux premiers axes d'une AFTD est représentée sur la figure 4a. Le premier axe isole la plupart des cultivars de toutes les espèces sauvages. Il existe cependant entre ces deux groupes un continuum, où l'on trouve les formes sauvages de *M. esculenta* et *M. pruinosa*. Les autres espèces sauvages se répartissent le long de l'axe 1, tandis que l'axe 2 permet de séparer ces espèces sauvages d'un groupe correspondant à *M. glaziovii* au sens large, tel que A.C. Allem l'a défini, considéré comme synonyme de *M. carthagenensis* (tableau 1 dans SECOND *et al.*, 1997). Deux groupes de formes en position intermédiaire entre *M. carthagenensis* et *M. esculenta* (isolés par des cercles sur la figure 4a) sont mis en évidence. Ils comprennent des formes classées comme *M. glaziovii*, des formes « arbre manioc », fréquentes dans le nord-est et le centre du Brésil, et, dans le groupe le plus proche de l'ensemble du manioc, quatre cultivars, dont SRT1316, originaires du nord-est brésilien. Nous considérons ces formes comme des hybrides entre *M. glaziovii* et *M. esculenta*.

Dans le nord-est du Brésil, on trouve donc une situation similaire à celle qui a été décrite en Afrique de l'Ouest (LEFEVRE, 1989). Des cultivars de manioc ont sans doute été introgressés par *M. glaziovii*, probablement par l'intermédiaire des formes arbre manioc, même si elles sont souvent largement stériles et multipliées par voie végétative.

Notre analyse confirme donc que les formes sauvages de *M. esculenta* sont bien les plus proches du manioc parmi toutes les espèces d'Amérique du Sud. De plus, elle montre que *M. pruinosa* ne se distingue pas sur le plan moléculaire des formes sauvages de *M. esculenta*. Cette proximité se retrouve aussi dans la morphologie (A.C. Allem, comm. pers.). Notre étude suggère aussi que *M. carthagenensis*, et en particulier la forme reconnue comme *M. glaziovii*, a participé à la constitution génétique de certains cultivars. Cette contribution est à mettre en relation avec le succès de l'introduction de *M. glaziovii* dans l'Ancien Monde et avec son utilisation comme ressource génétique dans les travaux modernes d'amélioration variétale du manioc. Enfin, certaines espèces, en particulier *M. fruticulosa*, pourraient être génétiquement proches du manioc, même si leur contribution directe à sa domestication n'est pas prouvée.

L'AFTD restreinte à l'espèce *M. esculenta* et aux espèces brésiliennes considérées *a priori* comme les plus proches du manioc (figure 4b) montre une séparation claire entre *M. esculenta* et *M. pruinosa*, d'une part, et les autres espèces brésiliennes, d'autre part. *M. fruticulosa* se scinde en deux sous-espèces en

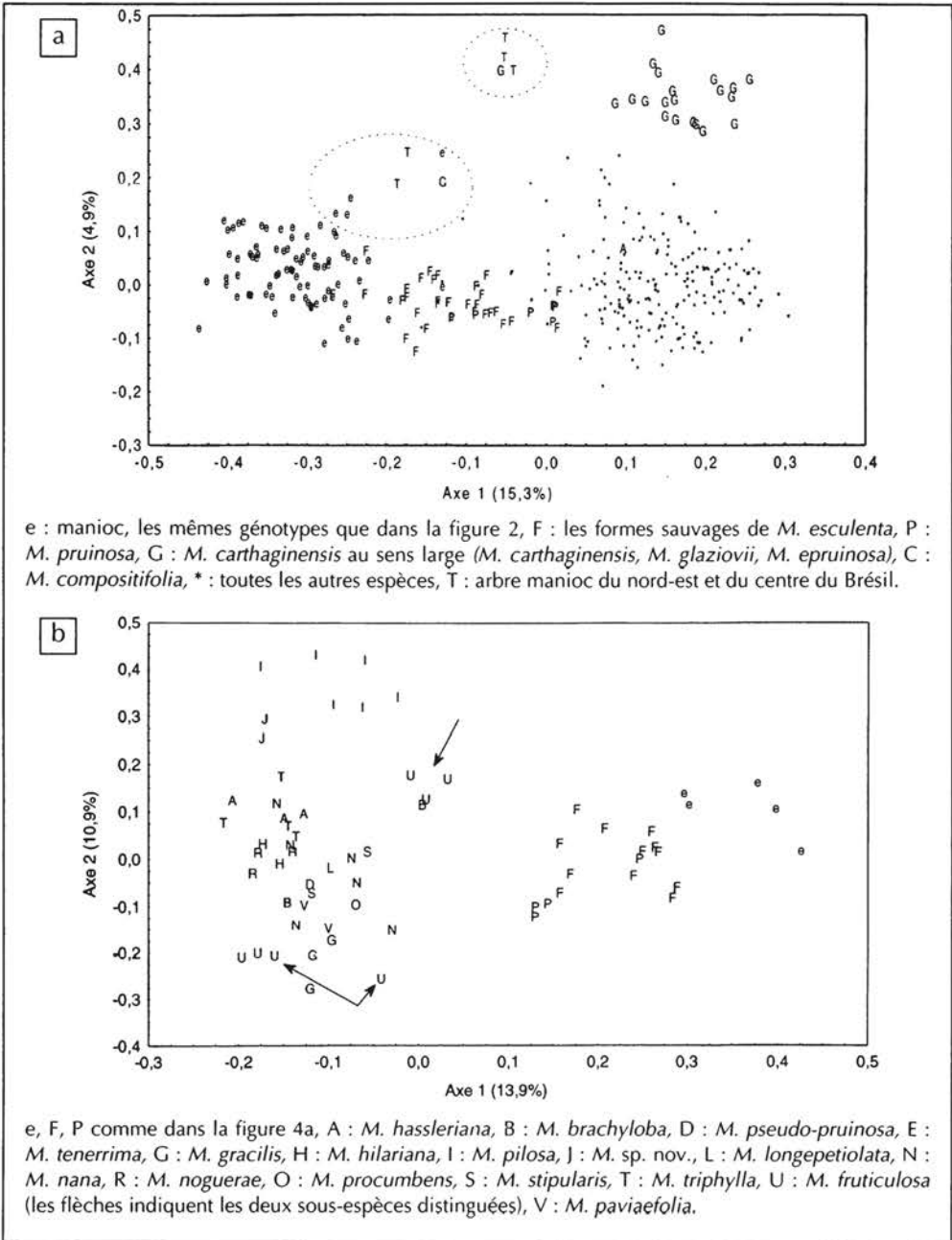


Figure 4. Diversité du genre *Manihot* : analyses en coordonnées principales fondées sur les similarités (simple matching) des génotypes deux à deux dans l'analyse de 93 fragments AFLP polymorphes. a : distribution des 364 plantes représentatives du genre. b : distribution des génotypes de *M. esculenta* (formes sauvages et cultivées), de *M. pruinosa* et des espèces brésiliennes morphologiquement les plus apparentées qui se trouvaient sur un même gel AFLP : données de la figure 4a réduites à ce gel unique.

accord avec l'analyse morphologique qui reconnaît une sous-espèce *caiaponia* (A.C. Allem, comm. pers.). La forme *M. fruticulosa* au sens strict apparaît comme la plus proche de *M. esculenta*. *M. fruticulosa* étant une espèce herbacée, on peut donc suggérer de l'utiliser en amélioration variétale pour modifier l'architecture du manioc.

La structure illustrée par la figure 4a est fiable malgré la dispersion des échantillons sur six gels et la difficulté de transcrire les gels en termes de présence ou d'absence de bandes spécifiques qui leur sont associées. L'analyse par AFLP d'un sous-échantillon des ADN utilisés dans la figure 4a, pour trois combinaisons d'amorces, tous placés sur le même gel pour chaque amorce vient en effet confirmer cette structuration. On retrouve bien les trois groupes suggérés par la figure 4a : *M. esculenta*, avec ses formes sauvages subsp. *flabellifolia* et *peruviana*, et *M. pruinosa* ; *M. carthaginensis* au sens large ; un groupe incluant toutes les autres espèces d'Amérique du Sud représentées, parmi lesquelles seule *M. compositifolia* apparaît comme intermédiaire entre ce groupe et *M. carthaginensis*. On note aussi que les regroupements selon la similarité moléculaire ne recoupent pas toujours la classification taxonomique, ce qui illustre une fois encore la difficulté de la classification de ce genre.

*M. carthaginensis* (incluant *M. glaziovii*) apparaît dans cette étude comme l'espèce d'Amérique du Sud la plus divergente du manioc. Le fait qu'elle a indiscutablement contribué à la domestication du manioc indique que toutes les espèces de *Manihot* peuvent être *a priori* considérées comme des ressources génétiques directes pour le manioc — pool primaire ou secondaire selon la terminologie de Harlan.

## Les relations entre la diversité morphologique et moléculaire

Les relations entre la diversité morphologique et moléculaire n'ont pas fait l'objet d'expérimentations rigoureuses à une échelle significative. Nous avons insisté sur la plasticité agromorphologique du manioc, qui complique cette comparaison. La plasticité des espèces sauvages est également considérable. Il est néanmoins possible, sur la base de ce qui a été rapporté précédemment, de faire les parallèles suivants entre les observations réalisées pour ces deux niveaux de diversité.

Pour 31 des accessions étudiées, toutes brésiliennes, une description a été effectuée à partir de neuf caractères : couleur de trois tissus racinaires et rugosité de la racine, couleur du bourgeon apical, de la tige jeune, du pétiole, largeur et sinuosité du lobe central, soit au total 27 états de caractères. Cette description a permis une comparaison avec la diversité observée avec les RAPD (74 bandes polymorphes). Les AFTD effectuées indépendamment pour les des-



cripteurs morphologiques et les marqueurs moléculaires montrent des structures similaires avec, notamment, un regroupement des variétés présentant des caractéristiques favorables pour l'industrie (COLOMBO, 1997).

Plus généralement, pour les formes cultivées, l'absence de division subs spécifique révélée par l'analyse botanique ou agromorphologique se retrouve sur le plan moléculaire, où la variabilité apparaît continue, sans grand déséquilibre de liaison, la majeure partie de la diversité de l'espèce se retrouvant dans un seul champ amazonien. Les Amérindiens reconnaissent en fait un grand nombre de variétés en accord avec la large diversité moléculaire observée, mais ces variétés sont dynamiques, en perpétuelle évolution, notamment grâce à la recombinaison génétique. Bien qu'il y ait concordance entre les deux échelles d'observation, l'analyse moléculaire semble mieux à même de décrire de manière fiable la diversité génétique du manioc, dans la mesure où elle est peu sensible aux variations de l'environnement.

Pour le genre *Manihot*, le classement en espèces bien délimitées est difficile que ce soit à l'aide des descripteurs morphologiques ou des marqueurs moléculaires. Cependant, pour le genre *Astrocaryum* (palmier), les marqueurs moléculaires AFLP permettent de retrouver sans ambiguïté les grandes subdivisions et en suggèrent une nouvelle qui n'est pas en contradiction avec l'analyse morphologique (KAHN et SECOND, 1998). Utilisés chez le palmier à huile, les marqueurs RFLP et AFLP mettent clairement en évidence des sous-espèces cryptiques (BARCELOS, 1998 ; BARCELOS *et al.*, 1998). La difficulté à classer le genre *Manihot* apparaît donc bien comme résultant de ses caractéristiques biologiques.

L'espèce *M. aesculifolia* — classée du point de vue morphologique comme la plus proche du manioc (ROGERS et APPAN, 1973) mais très divergente sur le plan moléculaire — représente néanmoins un cas notable de discordance entre les deux niveaux d'observation. Elle illustre l'intérêt des marqueurs moléculaires du point de vue phylogénétique : les espèces centraméricaines sont géographiquement isolées, ce qui peut expliquer leur distance phylogénétique malgré des ressemblances morphologiques plus sujettes à la sélection naturelle que la diversité moléculaire.

Par ailleurs, des formes classées comme intermédiaires entre les espèces proches *M. caerulescens* et *M. quinquepartita* selon les critères morphologiques se retrouvent également en position intermédiaire dans l'analyse moléculaire. En outre, plusieurs cas d'hybridation spontanée, soupçonnés lors de l'étude morphologique, sont confirmés par l'analyse moléculaire (SECOND *et al.*, 1997). Ces hybridations expliqueraient, pour partie, les difficultés à délimiter les espèces dans le genre *Manihot*.

De même que pour le manioc, il n'y a généralement pas de contradiction entre les résultats des deux niveaux d'observation pour l'ensemble du genre *Manihot*. C'est par une démarche associant l'observation morphologique et la caractérisation moléculaire que l'on parviendra à une meilleure compréhens-

sion de ce genre difficile. D'ores et déjà, il semble confirmé que les entités considérées comme des espèces selon la botanique classique ont généralement conservé la possibilité d'hybridations fécondes. Dans ces conditions, il est probablement inapproprié de représenter les relations entre ces espèces par un arbre phylogénétique. Un « réseau » de relations donnerait une image plus adéquate de la réalité. Il serait illusoire de vouloir mettre en évidence ce réseau de manière exhaustive, mais il pourrait être illustré au mieux par l'analyse multivariée sur plusieurs axes de variation. Si l'on admet la réalité de cette situation, on comprend qu'un modèle de conservation dynamique du genre soit particulièrement approprié.

## L'application à la gestion des ressources génétiques

Les caractéristiques biologiques du genre *Manihot*, mais aussi des considérations sur la complémentarité avec les approches de conservation *in situ* et *ex situ*, ont conduit à proposer un schéma de conservation dynamique pour les espèces sauvages du genre (SECOND *et al.*, 1998a ; 1998b). Des populations artificielles peuvent être rassemblées dans les biotopes d'origine en regroupant des individus qui appartiennent à une même espèce envisagée au sens strict. Au contraire, dans des biotopes ou des continents nouveaux pour ces espèces, le regroupement peut s'effectuer par espèce au sens large. On aboutit ainsi à un nombre gérable d'espèces qui, par le brassage de la diversité génétique, ont plus de chance de s'adapter aux conditions nouvelles. Un échange périodique de graines entre les populations de la même espèce — en excluant les populations de la zone d'origine qui ne doivent pas être génétiquement « polluées » — assurerait alors le maintien d'une diversité génétique élevée à moindre coût. Cette diversité est directement soumise à la sélection naturelle et disponible pour son intégration dans des schémas de sélection variétale. Ce dispositif de conservation dynamique viendrait en complément de la conservation *in situ*, quand elle est applicable, et de la conservation *ex situ* par graines, qui reste à optimiser.

Quant au manioc cultivé, son statut de plante reproduite généralement par voie végétative a conduit à favoriser un schéma de conservation par clones *in vitro*, qui a au moins deux inconvénients majeurs : son coût élevé et les risques associés de transmission de parasites, en particulier viraux, lors des échanges de matériel génétique.

La confirmation qu'il existe bien une gestion dynamique de la diversité en milieu traditionnel — qui représente probablement le processus de domestication du manioc — conduit au contraire à revaloriser la conservation de la diversité du manioc par graine : les génotypes sont modifiés mais la diversité

est conservée. Pour une conservation dynamique au champ des maniocs cultivés (SECOND *et al*, 1998b), il faut noter une particularité liée à la consommation des racines. Contrairement à la situation des céréales, pour lesquelles le cycle récolte-semis représente dans les conditions du champ une pression de sélection en faveur d'une variété cultivée productive, chez une plante cultivée pour ses racines, un tel schéma suppose qu'une pression de sélection soit imposée directement par le cultivateur. Ici réside probablement l'explication du modèle de champ traditionnel de manioc où, souvent, les clones d'une même variété sont cultivés en groupes dans le champ, ce qui permet de les évaluer jusqu'au stade de la consommation. Il s'agit donc, pour perpétuer ce système de manière efficace et justifiable, de maintenir l'association entre conservation et amélioration. Une grande diversité organisée, dans un même champ, en groupes de clones permet d'évaluer les clones dans le cadre d'une agriculture familiale. Il est alors possible, en fonction des connaissances scientifiques acquises, d'assister ce processus à la ferme en le rendant plus efficace : production d'hybrides, rétrocroisements et transformation génétique, notamment. Les biotechnologies, telles que les microplaques ou les puces à ADN, devraient permettre, dans un tel schéma, de suivre relativement simplement l'évolution de la fréquence de certains locus dans des ADN extraits à partir de mélanges d'individus de la population. Il serait donc possible d'associer la conservation et l'amélioration de la diversité génétique à la ferme avec les avantages qui devraient résulter des technologies de manipulations génétiques.

## Conclusion

L'application des marqueurs moléculaires à l'analyse de l'amplitude et de l'organisation de la diversité génétique du manioc et du genre *Manihot* aboutit à un ensemble de résultats, qui sont généralement en accord avec les données morphologiques, à une exception notable près, mais qui permettent de progresser plus rapidement dans la connaissance de ce genre.

L'analyse moléculaire confirme sans ambiguïté l'apparentement entre le manioc cultivé et ses formes présumées ancestrales. Elle conclut aussi à la proximité entre *M. pruinosa* et *M. esculenta*. *M. pruinosa* serait aussi une forme sauvage de *M. esculenta*, adaptée au *cerrado*, c'est-à-dire à la savane brésilienne.

Cette analyse met en évidence la singularité de l'espèce *M. carthaginensis* au sens large, dont la validité du regroupement avec *M. glaziovii* est vérifiée. Elle suggère l'introgession de *M. glaziovii* dans certaines variétés cultivées, alors que cette espèce est, parmi les espèces brésiliennes, la plus divergente de *M. esculenta*. *M. glaziovii* est donc, depuis sa diffusion dans l'Ancien Monde à la fin du XIX<sup>e</sup> siècle, un exemple de conservation dynamique de la diversité génétique associée à une création variétale à la ferme (SECOND, 1998).

La nature génétique complexe du genre *Manihot* est confirmée par l'analyse de la diversité moléculaire ; elle semble impliquer de nombreuses hybridations spontanées.

A partir de ces résultats, de nouveaux concepts peuvent être avancés. Les variétés distinguées traditionnellement correspondent à des familles de clones résultant d'une gestion dynamique de la diversité, qui doit être encouragée. Ce processus de gestion dynamique pourrait inclure une amélioration variétale à la ferme, qui bénéficierait des apports de la génétique moderne. Dans la mesure où la structure du genre *Manihot* est mieux décrite par un réseau de relations entre les espèces que par un arbre phylogénétique, il semble approprié d'envisager également une gestion dynamique pour la conservation de ce genre.

Le manioc est une plante alimentaire importante pour les régions tropicales, ce qui justifie la poursuite des travaux que nous avons entrepris. Il reste à examiner un échantillonnage plus représentatif de l'ensemble des variétés de manioc, qui inclut, outre les collections déjà étudiées, l'ensemble du bassin amazonien et les continents africain et asiatique. L'analyse du processus traditionnel de gestion de la diversité doit être approfondie, et la nature des variétés traditionnelles doit être examinée à partir d'un échantillonnage plus large, structuré selon le champ, le village, la région... Dans le même temps, la mise au point des marqueurs microsatellites et cytoplasmiques doit être poursuivie. Pour le genre *Manihot*, il serait nécessaire de rassembler l'ensemble des espèces centre et sud-américaines dans une même analyse. Des études de cas doivent être entreprises sur l'organisation de la diversité et des flux géniques dans des complexes d'espèces proches du genre *Manihot* en fonction de leur distribution et de leur écologie. L'analyse des formes de *M. glaziovii* diffusées dans le monde depuis plus d'un siècle et des sélections variétales, cultivées et spontanées, qui en ont résulté permettrait de faire une première évaluation des possibilités de conservation dynamique du genre *Manihot*.

## Annexe

### Matériel végétal

Les analyses ont porté sur 130 variétés de manioc issues des collections du Brésil et du CIAT et sur 278 plantes représentant toutes les espèces sud-américaines et quatre espèces centraméricaines du genre *Manihot*. Une collection vivante de l'ensemble des espèces brésiliennes, telles qu'elles sont considérées par A.C. Allem au CENARGEN (Centro Nacional de Pesquisa de Recursos Genéticos e Biotecnologia, Brésil ; tableau 1 dans SECOND *et al.*, 1997), a été constituée à partir de boutures, de transplants ou de graines collectées directement dans les populations naturelles. Cette collection d'espèces sauvages a été complétée par des échantillons d'herbier de la collection du CENARGEN. La plus grande partie des formes distinguées d'un point de vue morphologique et pour lesquelles il existe actuellement des populations connues a été ainsi représentée. Six hybrides présumés ont été inclus ainsi que leurs parents présumés prélevés dans les mêmes populations. Un tiers des échantillons environ provient d'herbiers, le reste, à quelques exceptions près, est issu de collectes directes dans la nature ou de plantes obtenues à partir de graines collectées dans la nature. Sept plantes, dont les quatre représentants des espèces d'Amérique centrale, proviennent des collections *in vitro* du CIAT, en Colombie, et du laboratoire de l'université de Washington, aux Etats-Unis. Pour l'analyse, des échantillons de jeunes feuilles saines ont été séchés à l'étuve ventilée à 50 °C pendant 20 heures, puis conservées au sec en présence de gel de silice.

### Extraction de l'ADN

Un protocole modifié de la technique au CTAB (cetyl-triméthyl ammonium bromide) a été utilisé (DELLAPORTA *et al.*, 1983 ; COLOMBO, 1997).

### Analyses par RAPD

Pour les analyses par RAPD, 208 amorces décamères ont été testées et 22 ont été retenues pour la qualité des profils de bandes obtenus (COLOMBO, 1997, pour les détails sur les amorces et l'amplification par PCR).

### Analyses par AFLP

Toutes les analyses par AFLP ont été confiées au laboratoire commercial de Linkage Genetics à Salt Lake City, Etats-Unis (actuellement, PE AgGen Inc.). Celui-ci emploie la technique AFLP telle que publiée et le programme Keygene pour une lecture des gels assistée par ordinateur. Pour la partie arbitraire des amorces, les combinaisons AGA/CAG et, éventuellement, AGT/CTC et AGA/CAA, lorsque deux ou trois combinaisons étaient employées parmi les douze combinaisons testées, ont été utilisées.

### Analyses statistiques

Les matrices de présence ou d'absence des bandes sont utilisées pour calculer les coefficients de similarité (coefficients de Jaccard ou *simple matching*) dans les comparaisons de génotypes deux à deux ou pour réaliser une AFC. Les matrices sont comparées à l'aide du test de Mantel. Les matrices de similarités sont utilisées pour construire des dendrogrammes (UPGMA) ou pour effectuer des analyses en coordonnées principales (ACoP, AFTD). Pour construire un dendrogramme UPGMA à partir de la valeur des coordonnées sur les premiers axes, une distance euclidienne est utilisée.

Pour tous les calculs ci-dessus, le programme Ntsys d'Exeter Software (ROHLF, 1998) est utilisé. L'analyse de variance moléculaire est calculée avec le programme Amova (EXCOFFIER *et al.*, 1992). Pour les graphiques on utilise Ntsys et Statistica (Statsoft).

Tableau 1. Caractères de la végétation, de l'inflorescence et des racines de manioc permettant de différencier les variétés, d'après COURS (1951). Les critères signalés en italique proviennent de sources bibliographiques plus récentes. Les flèches renvoient aux illustrations de la figure 5.

Organes	type
<b>Bois (tiges)</b>	
<ul style="list-style-type: none"> <li>couleur dominante des feuilles non entièrement développées sur la pousse terminale (équilibre entre les pigmentations chlorophyllienne et anthocyanique)</li> </ul>	1 - vert 2 - vert foncé avec reflets 3 - violet clair 4 - violet foncé
<ul style="list-style-type: none"> <li>temps d'imprégnation (durée de la coloration exprimée en nombre de feuilles colorées)</li> </ul>	de 0 à 12 (généralement 5 ou 6)
<ul style="list-style-type: none"> <li>couleur de la partie jeune de la tige non aoûtée</li> </ul>	1 - vert 2 - vert jaunâtre 3 - vert et début de pétiole rouge 4 - vert et début de pétiole rouge avec côtes rouges 5 - vert et rouge à surface égale 6 - quelques traces de vert 7 - entièrement rouge clair 8 - rouge foncé à violet-brun
<ul style="list-style-type: none"> <li>couleur de la tige (partie ancienne) à 1 et éventuellement à 2 ans</li> </ul>	1 - gris cendré à vert foncé 2 - vert olive 3 - acajou 4 - brun foncé
<ul style="list-style-type: none"> <li>port de la tige à la fin du cycle en relation avec la ramification liée à la floraison du méristème végétatif terminal</li> </ul>	1 - rampant (plus de 6 floraisons) 2 - étalé (4 à 6 floraisons) ➔ A1 3 - dressé (2 ou 3 floraisons) ➔ A2 4 - érigé (1 floraison tardive) ➔ A3 5 - cylindrique (pas de floraison)
<ul style="list-style-type: none"> <li>ramifications latérales (développement de bourgeons latéraux par levée de dominance apicale)</li> </ul>	1 - absence 2 - présence avec : 2a) port érigé (parallèle à la tige) 2b) port retombant (écarté de la tige puis redressé)
<ul style="list-style-type: none"> <li><i>nombre moyen de branches développées à chaque floraison</i></li> </ul>	1 - 2 branches 2 - 3 branches 3 - 4 branches 4 - plus de 4 branches
<ul style="list-style-type: none"> <li><i>angle d'écartement entre les branches développées par floraison</i></li> </ul>	1 - aucun 2 - de 15° à 30° 3 - de 45° à 60° 4 - de 75° à 90°
<ul style="list-style-type: none"> <li><i>hauteur totale de la tige principale</i></li> </ul>	<i>en cm</i>
<ul style="list-style-type: none"> <li><i>temps mis par la première inflorescence pour apparaître</i></li> </ul>	<i>en jour après plantation</i>

Tableau 1. Suite.

Organes	type
<ul style="list-style-type: none"> <li>• hauteur de la première ramification si floraison</li> </ul>	en cm
<ul style="list-style-type: none"> <li>• nombre de tiges par plant obtenu par bouturage (en notant le mode de plantation)</li> </ul>	nombre
<b>Nœuds et entre-nœuds sur la tige</b>	
<ul style="list-style-type: none"> <li>• couleur de l'œil (bourgeon) à l'état latent</li> </ul>	1 - vert 2 - base colorée et écailles vertes 3 - entièrement coloré
<ul style="list-style-type: none"> <li>• émergence de l'œil (bourgeon) à l'état latent</li> </ul>	1 - encaissée ⇒ B1 2 - saillante ⇒ B2
<ul style="list-style-type: none"> <li>• coussinet (renflement qui porte la cicatrice foliaire à l'endroit où était inséré le pétiole) avec stipules dentées de part et d'autre</li> </ul>	1 - dents éphémères ⇒ B3 - B4 2 - dents persistantes ⇒ B3 - B4
<ul style="list-style-type: none"> <li>• dimension du bourrelet stipulaire (gonflement des stipules et du coussinet)</li> </ul>	en cm ⇒ B3 - B4
<ul style="list-style-type: none"> <li>• alignement des entre-nœuds (partie jeune)</li> </ul>	1 - ligne brisée (en zig-zag) 2 - ligne droite
<ul style="list-style-type: none"> <li>• cannelures sur la partie jeune des tiges et des côtes</li> </ul>	1 - côtes éphémères 2 - côtes caduques 3 - côtes persistantes
<ul style="list-style-type: none"> <li>• vigueur des tiges (mesurée par le diamètre de base entre deux nœuds)</li> </ul>	en cm
<b>Feuille</b>	
<ul style="list-style-type: none"> <li>• forme générale</li> </ul>	1 - palmipartite (forme normale) 2 - palmiséquée (les lobes sont totalement séparés)
<ul style="list-style-type: none"> <li>• dimension du pétiole</li> </ul>	1 - type sessile (absence ou inférieur à 1 cm) 2 - type intermédiaire (associé à la forme palmiséquée) 3 - type long et cylindrique
<ul style="list-style-type: none"> <li>• angle d'attache du pétiole sur la tige</li> </ul>	1 - inférieur à 20° 2 - de 20° à 55° 3 - de 55° à 70° 4 - de 70° à 90° 5 - de 90° à 120° 6 - supérieur à 120°
<ul style="list-style-type: none"> <li>• zone d'apparition de la déviation du pétiole</li> </ul>	1 - submédiane 2 - médiane 3 - subterminale 4 - terminale

Tableau 1. *Suite.*

Organes	type
• coloration du pétiole	1 - entièrement verte 2 - entièrement vert-jaune ou vert clair 3 - crosse rouge à la base et le reste vert 4 - partie submédiane verte et tout le reste rouge 5 - partie submédiane verte (représentant la moitié de la longueur) et le reste coloré 6 - rouge sauf une partie de la zone submédiane verte 7 - rouge vif, plus foncée sur la crosse 8 - rouge foncé ou rouge violacé
• longueur et diamètre du pétiole	de 5 à 60 cm et de 1 à 5 mm
• pubescence de l'ombilic foliaire, partie à l'extrémité du pétiole où convergent principale des lobes la nervure	1 - cils répartis de manière uniforme et dense 2 - cils regroupés et plus clairsemés
• <i>pubescence des jeunes feuilles</i>	1 - <i>absente</i> 2 - <i>légère</i> 3 - <i>modérée</i> 4 - <i>forte</i>
• dimension de l'ombilic foliaire	1 - ombilic rétréci ⇒ C1 2 - ombilic étalé ⇒ C2
• forme de l'ombilic foliaire	1 - concave ⇒ C3 2 - convexe ⇒ C4
• nombre de lobes constituant le limbe, calculé sur une moyenne au cours de 12 mois de croissance de la plante	1 - moins de 3 2 - de 3 à 5,5 3 - de 5,5 à 7 4 - plus de 7
• forme des lobes (rapport L/l entre longueur et largeur, mesurées sur le lobe médian)	1 - très étroite (L/l supérieur à 20) ⇒ C5 2 - à marges parallèles (L/l de 6 à 20) 3 - normale (L/l de 4,5 à 6) 4 - large (L/l de 3 à 4,5) ⇒ C6 5 - arrondie (L/l inférieur à 3) avec deux situations du point d'élargissement
• forme des lobes définie par la position du point d'élargissement maximal le long du lobe médian (pour les types 3 et 4)	1 - proximale (dans la 1 <sup>re</sup> moitié basale du lobe) ⇒ C7 2 - médiane (au milieu du lobe) 3 - submédiane (entre milieu et les trois quarts) 4 - terminale (au-delà du dernier quart vers la pointe) ⇒ C8
• ornementation des lobes	1 - présence ou non d'échancrures médianes 2 - présence ou non d'éperons



Tableau 1. *Suite.*

Organes	type
• ornementation des lobes ( <i>suite</i> )	3 - limbe incurvé 4 - marges gondolées
• couleur de la face supérieure du limbe	1 - blanche sans chlorophylle (albinos partiel ou total) ou avec panachures 2 - vert clair 3 - vert jaunâtre 4 - vert foncé 5 - violacée
• couleur de la face inférieure du limbe	1 - verte 2 - vert blanchâtre 3 - vert jaunâtre 4 - violette ou rouge
• coloration des nervures du limbe	1 - toujours verte 2 - rouge sur la face inférieure des jeunes feuilles 3 - rouge sur la face supérieure des jeunes feuilles 4 - rouge sur la face supérieure des feuilles adultes 5 - rouge sur la face inférieure des feuilles adultes 6 - rouge sur les deux faces des jeunes feuilles 7 - rouge sur la face supérieure des feuilles jeunes ou âgées 8 - rouge sur la face inférieure des feuilles quel que soit l'âge 9 - sur les deux faces des feuilles jeunes et subsistant sur la face inférieure des feuilles adultes 10 - toujours sur les deux faces
• sinus lobaire (espace vide séparant deux lobes)	d'autant plus fermé que le nombre de lobes est élevé ⇒ C9
• sinus basilaire (angle formé entre les nervures centrales des deux lobes extrêmes ou inférieurs pris de part et d'autre du pétiole)	1 - fermé (angle inférieur à 180°) ⇒ C10 2 - ouvert (lobes extrêmes remontants, angle supérieur à 180°) ⇒ C11
• dimension du velum (surface de limbe à la confluence des lobes extrêmes au niveau du sinus basilaire)	1 - bande de 0,5 mm 2 - inférieure au diamètre de l'extrémité du pétiole ⇒ C12 3 - supérieure au diamètre de l'extrémité du pétiole 4 - plus large que la base du pétiole 5 - supérieure à 2 fois le diamètre de la base du pétiole ⇒ C13

Tableau 1. *Suite.*

Organes	type
• port du velum	1 - rectiligne ⇒ C14 2 - redressé 3 - retroussé 4 - plissé ⇒ C15
• ornementation du velum	1 - aucune 2 - dentée 3 - stipulée ⇒ C16 4 - frangée ⇒ C17
<b>Inflorescence ⇒ D1</b>	
• fructification de l'inflorescence	1 - continue (les premières fleurs sont fonctionnelles) 2 - tardive (avortement des premières inflorescences, fructification des plus tardives)
• forme des sépales du calice de la fleur femelle	1 - large 2 - moyenne 3 - étroite
• couleur des sépales de la fleur ⇒ D2	1 - entièrement verte 2 - verte et nervures colorées 3 - rouge et verte 4 - rouge à violette
• torus, bourrelet à nectaires entre calice et pistil	1 - jaune ou jaunâtre 2 - rougeâtre 3 - brun-rouge
• forme des six ailes portées par l'ovaire à trois carpelles (correspondant aux points de suture des carpelles)	1 - rectiligne quel que soit le stade d'évolution du fruit 2 - rectiligne puis sinueuse ⇒ D3 3 - sinueuse puis rectiligne à la maturité ⇒ D4 4 - toujours sinueuse
• couleur des ailes le jour de l'ouverture de la fleur	1 - entièrement verte 2 - rouge
• couleur du corps de l'ovaire	1 - vert 2 - rouge
• couleur du stigmate le jour de l'ouverture de la fleur	1 - absente (blanche ou légèrement rose) 2 - rouge
• présence de pollen dans la fleur mâle avec calice et torus (non utilisés pour la classification) et androcée à 10 étamines	1 - fleur mâle stérile 2 - fleur mâle fertile
• couleur du fruit (ou capsule) avec renflement du pédoncule	1 - verte 2 - claire à rouge violet

Tableau 1. *Suite.*

Organes	type
• couleur du fruit (ou capsule) avec renflement du pédoncule ( <i>suite</i> )	3 - vive dominante 4 - vive totale
• <i>dimension du fruit (longueur et diamètre)</i>	<i>en mm</i>
• couleur du tégument de la graine ⇒ D5	1 - grise 2 - brune
• marbrures du tégument de la graine ⇒ D5	1 - peu nombreuses 2 - denses
• couleur de la caroncule de la graine	1 - blanche ou crème 2 - rose ou rouge 3 - violette
• <i>dimension de la graine (longueur et diamètre)</i>	<i>en mm</i>
<b>Racine</b>	
• point d'attache sur la bouture d'origine	1 - sessile 2 - pédonculé 3 - longuement pédonculé (supérieur à 10 cm)
• longueur de la racine	1 - courte (inférieure à 40 cm) 2 - normale (de 40 à 80 cm) 3 - longue (supérieure à 80 cm)
• diamètre de la racine à son maximum	1 - mince 2 - moyen 3 - gros
• forme de la racine	1 - conique ⇒ E1 2 - fusiforme ⇒ E2 3 - cylindro-conique ⇒ E3 4 - cylindrique ⇒ E4
• <i>constriction sur les racines</i>	1 - <i>absente</i> 2 - <i>présente</i>
• direction de la racine	1 - traçante (horizontale) 2 - plongeante (verticale)
• nombre de racines	nombre
• <i>homogénéité des poids entre racines</i> (effectifs par classe de poids)	<i>en %</i>
• <i>texture de la surface des racines</i>	1 - <i>lisse</i> 2 - <i>moyenne</i> 3 - <i>rugueuse</i>
• aspect de l'écorce externe (liège)	1 - gris et mince 2 - brun et épais

Tableau 1. Suite.

Organes	type
• couleur du phelloderme	1 - blanche 2 - rose 3 - violette
• détachement du phelloderme du cylindre central	1 - facile 2 - difficile
• couleur de la pulpe (ou chair, cylindre central)	1 - blanche 2 - jaune
• rendement utile (rapport entre le poids de racines commercialisables et le poids total des racines produites)	en %
• indice de précocité (rendement en matière sèche des racines à 6-8 mois rapporté au rendement en matière sèche des racines à 12-14 mois)	en %
• index de récolte (poids sec des racines tubérisées rapporté à la biomasse totale du plant)	en %
• taux de fibres	1 - absence 2 - quelques fibres observables à l'œil 3 - nombreuses
• goût (importance de la libération d'acide cyanhydrique)	1 - doux 2 - amer, ou trois classes définies par l'intensité de la coloration à l'acide picrique : basse, moyenne, élevée

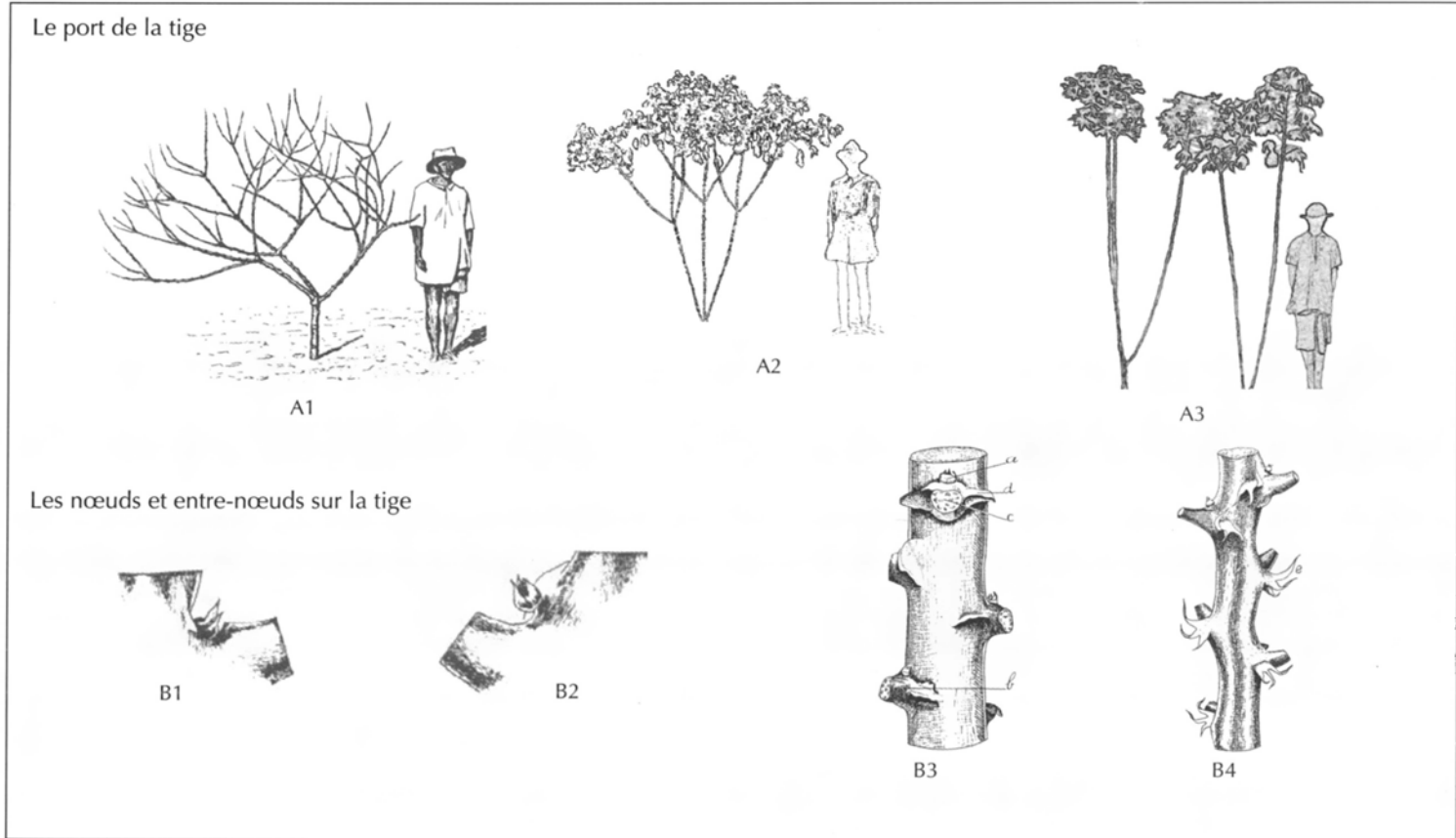


Figure 5. Caractères de la végétation, de l'inflorescence et des racines de manioc permettant de différencier les variétés, d'après COURS (1951).

La feuille

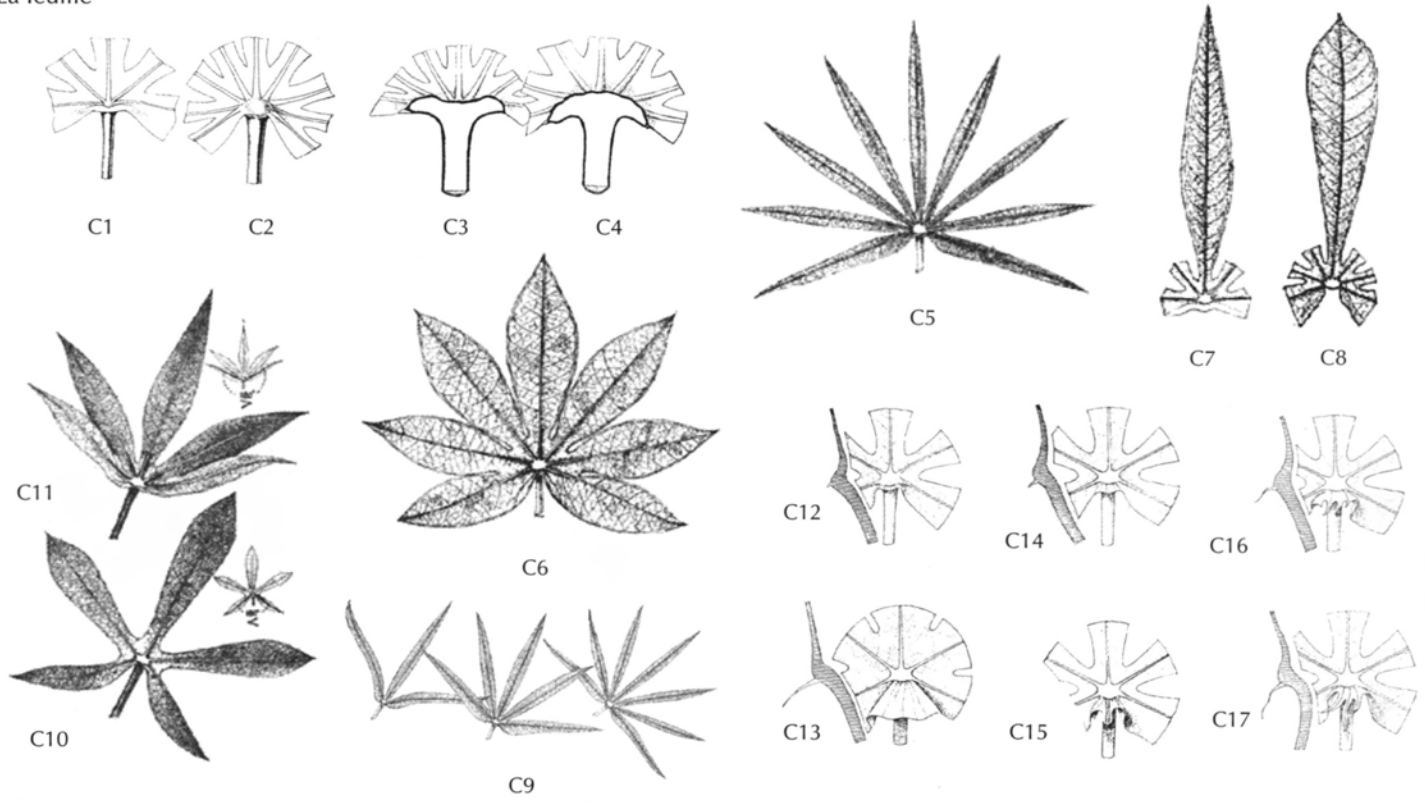


Figure 5. Suite.

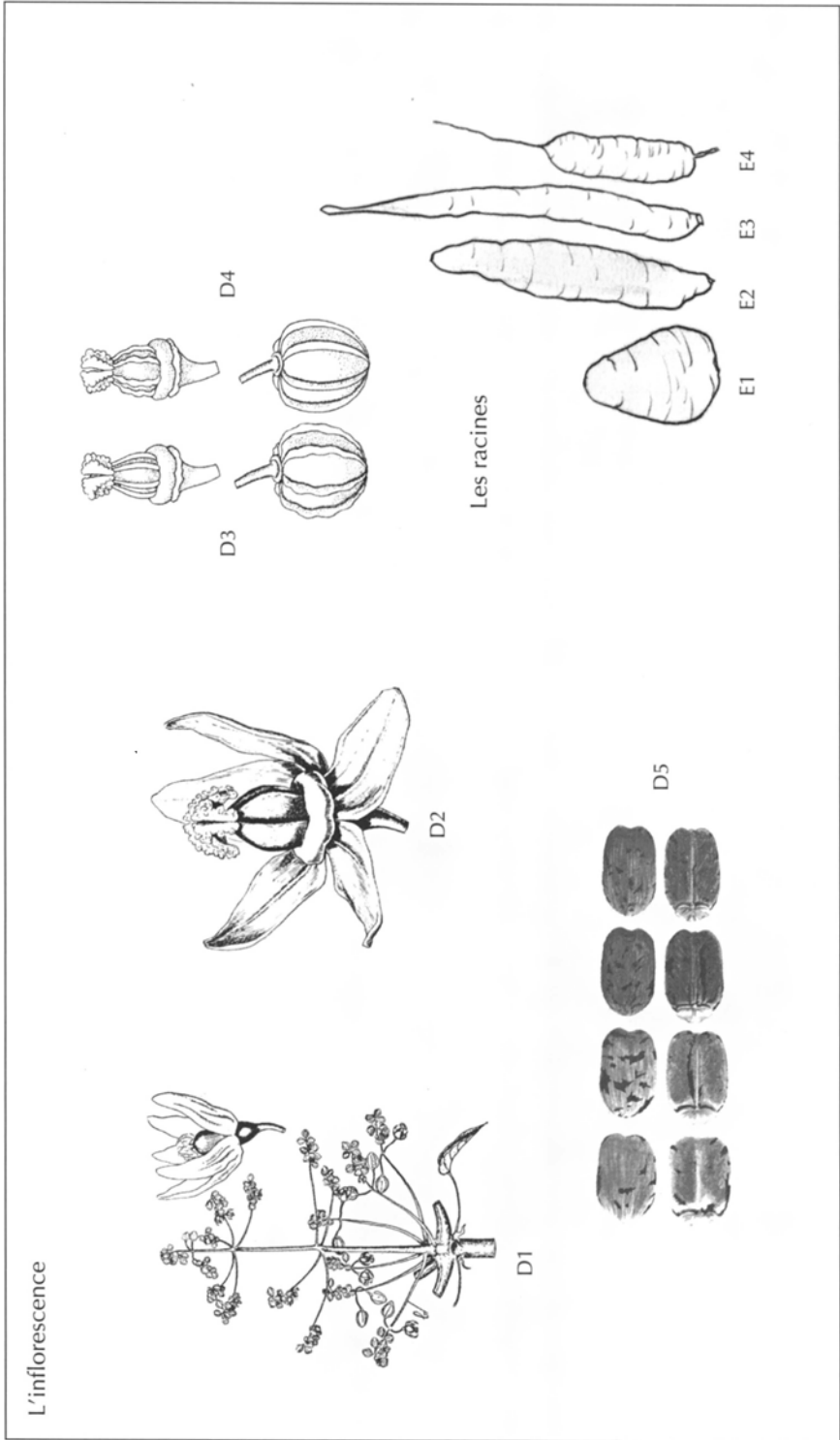


Figure 5. Suite.

## Références bibliographiques

- ALLEM A.C., 1994. The origin of *Manihot esculenta* Crantz (Euphorbiaceae). Genetic Resources and Crop Evolution, 41 : 133-150.
- BAI K.V., ASIEDU R., DIXON A.G.O., 1993. Cytogenetics of *Manihot* species and interspecific hybrids. In : 1st international scientific meeting of the Cassava Biotechnology Network, W.M. Roca et A.M. Thro éd., Cali, Colombie, CIAT, p. 51-55.
- BARCELOS E., 1998. Etude de la diversité du genre *Elaeis* (*E. oleifera* Cortès et *E. guineensis* Jacq.) par marqueurs moléculaires (RFLP et AFLP). Thèse de doctorat, ENSAM, Montpellier, France, 137 p.
- BARCELOS E., SECOND G., KAHN F., AMBLARD P., LEBRUN P., SEGUIN M., 1998. Molecular markers applied to the analysis of the genetic diversity and to the biogeography of *Elaeis*. Memoirs of the New York Botanical Garden (sous presse).
- BERTRAM R.B., 1993. Application of molecular techniques to genetic resources of cassava (*Manihot esculenta* Crantz, Euphorbiaceae): interspecific evolutionary relationships and intraspecific characterization. Thèse PhD, University of Maryland, Etats-Unis, 465 p.
- DE BRUIJN G.H., DHAMAPUTRA T.S., 1974. The Mukibat system, a high-yielding method of cassava production in Indonesia. Netherlands Journal of Agricultural Science, 22 : 89-100.
- CAB, 1974. Cassava (*Manihot esculenta*). Maidenhead, Royaume-Uni, CAB, Annotated Bibliography n° G-405, 27 p.
- CHAVARRIAGA-AGUIRE P., MAYA M.M., BONIERBALE M.W., KRESOVICH S., FREGENE M.A., TOHME J., KOCKERT G., 1998. Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. Theoretical and Applied Genetics, 97(3) : 493-501.
- COLOMBO C., 1997. Etude de la diversité génétique de maniocs américains (*Manihot esculenta* Crantz) par les marqueurs moléculaires (RAPD et AFLP). Thèse de doctorat, ENSAM, Montpellier, France, 144 p.
- COURS G., 1951. Le manioc à Madagascar. Mémoire de l'Institut scientifique de Madagascar, série B, 3(2) : 203-400.
- DELLAPORTA S.L., WOOD J., HICKS J.B., 1983. A plant DNA preparation: version II. Plant Molecular Biology Report, 4 : 19-21.
- EL-SHARKAWY M.A., COCK J.H., 1990. Photosynthesis of cassava (*Manihot esculenta*). Experimental Agriculture, 26 : 325-340.
- EMPERAIRE E., PINTON F., SECOND G., 1998. Gestion dynamique de la diversité variétale du manioc en Amazonie du nord-ouest. Nature, science et société, 6(2) : 27-42.
- EXCOFFIER L., SMOUSE P.E., QUATTRO J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA data. Genetics, 131 : 479-491.
- FAO, 1998. Le manioc. In : Perspectives de l'alimentation n° 2. Rome, Italie, FAO-SMIAR.



FREGENE M., ANGEL F., GOMEZ R., RODRIGUEZ F., CHAVARRIAGA P., ROCA W.M., TOHME J., BONIERBALE M.W., 1997. A molecular genetic map for cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics*, 95(3) : 431-441.

GULICK P., HERSHEY C., ESQUINAR-ALCAZAR J., 1983. Genetic resources of cassava and wild relatives. Rome, Italie, IBPGR, 56 p.

HAHN S.K., 1984. Les plantes à racines et tubercules tropicales : amélioration et utilisation. Ibadan, Nigeria, IITA, Rapport de conférence n° 2, 32 p.

HERSHEY C., IGLESIAS C., IWANAGA M., TOHME J., 1994. Definition of a core collection for cassava. *In* : 1st meeting of the International Network for Cassava Genetic Resources. Rome, Italie, IPGRI, International Crop Network Series n° 10, p. 145-156.

HOWLER R.H., 1990. Long-term effect of cassava cultivation on soil productivity. *Field Crops Research*, 26 : 1-18.

IGLESIAS C., HERSHEY C., CALLE F., BOLANOS A., 1994. Propagating cassava (*Manihot esculenta*) by sexual seed. *Experimental Agriculture*, 30 : 283-290.

JONES D.A., 1998. Why are so many food plants cyanogenic? *Phytochemistry*, 47(2) : 155-162.

KAHN F., SECOND G., 1998. The genus *Astrocaryum* in Amazonia: classical taxonomy and DNA analysis (AFLP). *Memoirs of the New York Botanical Garden* (sous presse).

KEATING B.A., WILSON G.L., EVENSON J.P., 1985. Effect of photoperiod on growth and development of cassava (*Manihot esculenta* Crantz). *Australian Journal of Plant Physiology*, 12 : 621-630.

LEBART L., MORINEAU A., TABARD N., 1977. Techniques de la description statistique : méthodes et logiciels pour l'analyse des grands tableaux. Paris, France, Bordas, p. 217-244.

LEFEVRE F., 1989. Ressources génétiques et amélioration du manioc, *Manihot esculenta* Crantz, en Afrique. Paris, France, Orstom, Travaux et documents microédités n° 57, 175 p.

MAGOON M.L., KRISHNAN R., BAI K.V., 1969. Morphology of the pachytene chromosomes and meiosis in *Manihot esculenta* Crantz. *Cytologia*, 34 : 612-626.

MARIE D., BROWN C., 1993. A cytometric exercise in plant DNA histograms, with 2C values for 70 species. *Biology of the Cell*, 78 : 41-51.

MCKEY D., BECKERMAN S., 1996. Ecologie et évolution des produits secondaires du manioc et relations avec les systèmes traditionnels de culture. *In* : L'alimentation en forêt tropicale. 1. Les ressources alimentaires : production et consommation, C.M. Hladik *et al.* éd., Paris, France, UNESCO, p. 165-202.

MEDARD R., SELL Y., BARNOLA P., 1992. Le développement du bourgeon axillaire du *Manihot esculenta*. *Canadian Journal of Botany*, 70 : 2041-2052.

PURSEGLOVE J.W., 1992. Tropical crops. Londres, Royaume-Uni, Longman, 2 volumes.

RAFFAILLAC J.P., 1985. Pluviométrie et qualité de la production chez le manioc dans le sud de la Côte d'Ivoire. *In* : Eau et développement agricole. Adiopodoumé, Côte d'Ivoire, Orstom, p. 78-81.

- RAFFAILLAC J.P., 1992. Enracinement de la bouture de manioc (*Manihot esculenta* Crantz) au cours des premières semaines de croissance. *L'Agronomie tropicale*, 46(4) : 273-281.
- RAFFAILLAC J.P., 1997. Le manioc : quelles priorités de recherche pour améliorer la production en relation avec la transformation et la commercialisation? *Les Cahiers de la recherche-développement*, 43 : 7-19.
- RAFFAILLAC J.P., 1998. Le manioc et la fertilité du milieu. Montpellier, France, CNEARC-EITARC, 30 p.
- RAFFAILLAC J.P., NEDELEC G., 1988. Comportement du manioc en début de cycle en fonction de la durée de stockage de la bouture. *In* : VII<sup>e</sup> symposium ISTRC, juin 1985. Paris, France, Inra.
- RAFFAILLAC J.P., SECOND G., 1997. Le manioc. *In* : L'amélioration des plantes tropicales, A. Charrier *et al.* éd., Montpellier, France, Cirad-Orstom, collection Repères, p. 429-455.
- ROA A.C., CHAVARRIAGA P., DUQUE M.C., MAYA M.M., BONIERBALE M.W., TOHME J., IGLESIAS C., 1998. Cross-species amplification of cassava (*Manihot esculenta* Crantz) microsatellite primers within the genus: polymorphism, genetic diversity and degree of relationship. *Genome* (sous presse).
- ROA A.C., MAYA M.M., DUQUE M.C., TOHME J., ALLEM A.C., BONIERBALE M.W., 1997. AFLP analysis of relationships among cassava and other *Manihot* species. *Theoretical and Applied Genetics*, 95 : 741-750.
- ROGERS D.J., APPAN M., 1973. *Manihot*, Manihotoides (Euphorbiaceae). New York, Etats-Unis, Hafner Press, Flora Neotropica Monograph n° 13, 274 p.
- ROGERS D.J., FLEMING H., 1973. A monograph of *Manihot esculenta* with an explanation of the taximetrics methods used. *Economic Botany*, 27 : 1-113.
- ROHLF F.J., 1998. Ntsys-pc, numerical taxonomy and multivariate analysis system, version 2.0: user guide. New York, Etats-Unis, Exeter Software, 31 p.
- SECOND G., 1998. *Manihot glaziovii* contributed to the genetic make-up of cassava and represents an example of dynamic conservation and on-farm breeding of genetic resources. *In* : IV<sup>th</sup> international scientific meeting of the Cassava Biotechnology Network, Salvador de Bahia, Brésil.
- SECOND G., ALLEM A.C., EMPERAIRE L., INGRAM C., COLOMBO C., MENDES R.A., CARVALHO J.C.B., 1997. Molecular markers (AFLP) based *Manihot* and cassava genetic structure analysis and numerical taxonomy in progress: implications for their dynamic conservation and genetic mapping. *In* : III<sup>rd</sup> international scientific meeting of the Cassava Biotechnology Network, A.M. Thro et M.O. Akoroda éd., African Journal of Root and Tuber Crops, 2(1-2) : 140-147.
- SECOND G., COLOMBO C., MENDES R. A., BERTHAUD J., 1998a. A scheme for a dynamic conservation of the genetic resources of wild *Manihot* and cultivated cassava in America and Africa and its extension to yam. *In* : Regional workshop for the conservation and utilisation of cassava, sweetpotato and yam germplasm in sub-Saharan Africa. Nairobi, Kenya, ILRI.

SECOND G., IGLESIAS C., DIXON A., 1998b. The state of use of cassava genetic diversity and a proposal to enhance it. *In* : Broadening the genetic bases of crop production. Rome, Italie, FAO-IPGRI.

SERIER J.B., 1989. Historico da disseminação da maniçoba fora do Brasil. *In* : Primeiro encontro nordestino da maniçoba. Recife, Brésil, IPA, p. 89-95.

WEBSTER G.L., 1975. Conspectus of a new classification of the Euphorbiaceae. *Taxon*, 24 : 593-601.

# Le mil

---

Gilles Bezançon

**L**e mil représente, avec le sorgho, la base de l'alimentation pour une part importante de la population des continents africain et indien. C'est la céréale la plus tolérante à la sécheresse ; elle est cultivée, au Sahel, dans des zones où la pluviosité ne dépasse pas 200 millimètres. Pour ces raisons, la conservation, l'évaluation et la valorisation des ressources génétiques des mils représentent un enjeu considérable. La culture du mil est surtout pratiquée dans les zones arides et semi-arides de l'Afrique, où elle couvre 11,5 millions d'hectares, et de l'Inde, avec 14,7 millions d'hectares (FAO, 1996).

Les trois quarts de la production africaine proviennent de l'ouest du continent, les principaux pays producteurs sont par ordre décroissant : le Nigeria, le Niger, le Burkina, le Tchad, le Mali et la Mauritanie. En Afrique de l'Est, le Soudan et l'Ouganda sont les premiers producteurs, alors qu'en Afrique australe cette culture traditionnelle a quasiment disparu. En Afrique, les variétés locales, précoces ou tardives, sont de taille haute et à fort tallage. Elles sont peu productives : mises en place sur des sols souvent pauvres, elles ne bénéficient d'aucun travail du sol et ne font l'objet que de rares apports d'intrants, sous forme de fumure organique par la pratique du parcage. Le semis, effectué à l'arrivée des pluies, doit être renouvelé plusieurs fois en cas de mauvaise pluviométrie.

En Inde, la culture est importante dans les Etats du Rajasthan, du Gujerat et de l'Haryana. L'utilisation de la traction animale permet de pratiquer un labour du sol et les semis sont effectués à l'aide de semoirs traditionnels. Les variétés traditionnelles sont tardives, à paille courte et à faible tallage. Mais là encore, la très rare utilisation des engrais chimiques ne permet pas d'obtenir des rendements en grain performants : ils restent faibles, entre 0,6 et 0,8 tonne par hectare, comparés à ceux des autres céréales cultivées en milieu tropical, telles que le riz et le maïs. Le mil peut être intégré dans des systèmes de cultures associées, avec le niébé par exemple. Enfin, aux Etats-Unis, le mil est cultivé sur plus de 150 000 hectares, pour produire du fourrage et du grain : les rendements dépassent alors 1,2 tonne par hectare (FAO, 1996).

Les grains, d'une valeur nutritive supérieure à celle du blé et du riz, représentent la principale utilisation de la plante. Ils sont consommés sous forme de pâte, de bouillie, de couscous ou encore de galettes et peuvent servir, dans certaines régions, à la fabrication de boissons alcoolisées (bière de mil). La paille peut également être utilisée comme fourrage ou encore dans la construction des habitations traditionnelles pour les toitures et les clôtures.

## La taxonomie et les ressources génétiques

### La botanique et la taxonomie du genre *Pennisetum*

Le mil, *Pennisetum glaucum* (L.) R. Br., appartient au genre *Pennisetum* (famille des poacées, sous-famille des *Panicoideae*, tribu des *Paniceae*), dont les différentes espèces (environ 60) sont distribuées en zone tropicale et subtropicale à travers le monde. Le genre *Pennisetum* est divisé en cinq sections, le mil appartenant à la section *Penicillaria*, qui se caractérise par la présence d'une touffe de poils sur l'apex des étamines. Dans cette section, VAN DER ZON (1992) reconnaît trois sous-espèces au sein de l'espèce *P. glaucum* ( $x = 7$  et  $2n = 2x = 14$ ) : *P. glaucum* subsp. *glaucum*, le mil cultivé, présent en Afrique (ouest, centre, est et sud), en Inde (centre et sud) et en Amérique du Nord ; *P. glaucum* subsp. *violaceum*, la forme sauvage, largement distribuée en Afrique dans la zone sahélienne et subdésertique, de façon discontinue de l'Atlantique à la mer Rouge, dans des situations écologiques variées ; *P. glaucum* subsp. *sieberianum*, qui rassemble les formes intermédiaires issues d'hybridations naturelles entre formes cultivées et formes sauvages.

Sur le modèle de HARLAN et DE WET (1971), le complexe d'espèces du genre *Pennisetum* peut être structuré en trois pools géniques. Le pool primaire, monospécifique, rassemble les trois sous-espèces de *P. glaucum*. Dans certaines régions de culture, l'importance des formes intermédiaires a conduit les populations locales à leur attribuer une appellation particulière : *chibra* en

haoussa au Niger et *n'doul* en oulof au Sénégal. Le pool secondaire se compose des deux espèces *P. purpureum* et *P. squamulatum*, qui peuvent s'hybrider facilement avec *P. glaucum* (HANNA, 1987). *P. purpureum* est une espèce pérenne, allotétraploïde ( $x = 7$  et  $2n = 4x = 28$ ), sexuée, allogame. *P. squamulatum* est une espèce pérenne, tétraploïde ( $x = 9$  et  $2n = 4x = 36$ ), apomictique. Ces deux espèces sont utilisées dans les travaux d'amélioration du mil par l'apomixie (HANNA, 1987 ; 1990). Le pool tertiaire regroupe les autres espèces du genre, dont les espèces de la section *Brevivalvula*. Très représentées en Afrique, où elles occupent des niches écologiques diverses, ces dernières sont soit annuelles, soit pérennes avec des systèmes de reproduction variés (sexué, apomictique, végétatif). Elles pourraient être utilisées pour ces raisons dans l'amélioration génétique du mil cultivé.

## Les ressources génétiques

La Fondation Rockefeller et l'Indian Agricultural Research Institute avaient entrepris vers 1960 la création d'une collection mondiale des cultivars de mils. Cette première tentative a échoué et la disparition d'un nombre relativement élevé de cultivars traditionnels, causée d'une part, par l'extension du front de la sécheresse dans les zones particulièrement exposées aux aléas climatiques et d'autre part, par les transformations socio-économiques de certains pays, a conduit la FAO (Food and Agriculture Organization of the United Nations), puis l'IBPGR (International Board for Plant Genetic Resources), à entreprendre, au début des années 70, la collecte et la conservation des ressources génétiques des mils et des sorghos de la zone sahélienne. L'ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), mandaté par le GCRAI (Groupe consultatif pour la recherche agronomique internationale), a coordonné ces collectes, réalisées en collaboration avec l'IBPGR, l'Orstom (Institut français de recherche scientifique pour le développement en coopération) et les instituts nationaux de recherche agronomique. Une collection mondiale de près de 24 000 échantillons — cultivars traditionnels et formes sauvages confondus — originaires de 44 pays est actuellement conservée, à long et à moyen termes, au centre indien de l'ICRISAT, à Patancheru. D'autres sites de conservation existent, notamment le laboratoire GeneTrop de l'IRD (Institut de recherche pour le développement, anciennement Orstom), à Montpellier, en France — conservation à moyen terme d'environ 3 200 échantillons dont 11 % appartiennent à l'espèce sauvage —, et le centre sahélien de l'ICRISAT de Sadoré, au Niger. Des doubles de ces collections sont implantés au Centre de conservation des ressources génétiques d'Ottawa, au Canada, ainsi qu'au NSSL (National Seed Storage Laboratory) de Fort Collins, aux Etats-Unis, pour une conservation à long terme.

Pour chacune de ces collections les échantillons sont accompagnés d'une fiche « passeport », regroupant 11 descripteurs, qui donnent des informations sur la situation d'origine de l'échantillon.

# L'organisation de la diversité génétique

## La variabilité agromorphologique

La taille imposante des collections n'est pas sans soulever des problèmes pour l'évaluation de la diversité qu'elles renferment. De nombreux travaux d'évaluation conduits par l'ICRISAT ont porté sur les collections spécifiques de certains pays tels que le Ghana, le Malawi, le Cameroun, la République centrafricaine et différents Etats de l'Inde. Les résultats concernant 1 938 et 2 458 échantillons de la collection mondiale, évalués pour 20 et 18 caractères respectivement, sont réunis dans un catalogue édité par le NBPGR (National Bureau of Plant Genetic Resources, Inde) et l'ICRISAT (NBPGR et ICRISAT, 1993). Ces données ne se rapportent qu'aux formes cultivées. Il n'existe pas d'études incluant les formes sauvages, qui permettraient de décrire la diversité génétique de l'ensemble des collections et rendraient alors possible la mise en parallèle avec les données de l'électrophorèse enzymatique obtenues par ailleurs.

En marge de ces études purement descriptives réalisées en Inde, différentes analyses ont été menées. Elles ont abouti à des classifications, notamment pour les formes cultivées, qui concordent avec le classement des cultivars en grands groupes régionaux établi à partir d'observations effectuées au cours des différentes missions de prospection (CLEMENT, 1985). PORTERES (1950 ; 1976) reconnaît seize espèces réparties en quatre groupes sur la base de leur répartition géographique : groupe extrême ouest-africain, groupe ouest et centre-africain, groupe nilo-soudanais et groupe est-africain et angolais. Pour l'Afrique de l'Ouest francophone, BONO (1973) utilise essentiellement les caractères du faux épis, ce qui le conduit à constituer deux groupes : le groupe I, divisé en deux sous-groupes (les formes rencontrées au Mali, en Côte d'Ivoire, en Mauritanie et quelques formes du Niger d'une part, les formes rencontrées au Burkina, d'autre part), et le groupe II, lui aussi scindé en deux sous-groupes (mils du Niger d'une part, et mils du Sénégal d'autre part). Sur la base de la forme du grain, BRUNKEN *et al.* (1977) définissent quatre races : *typhoides*, *nigritarum*, *globosum* et *leonis*. Les notations réalisées sur la longueur du cycle ainsi que sur une dizaine de caractères de la chandelle et du grain pour quelque 1 500 chandelles collectées au Mali et au Sénégal ont mis en évidence une différenciation régionale dans chacun de ces deux pays (MARCHAIS, 1982).

MARCHAIS *et al.* (1993) ont étudié 267 accessions de mil cultivé et 118 de mil sauvage à l'aide de 14 caractères morphologiques. Les résultats obtenus (figure 1) montrent une distinction très marquée entre mils cultivés et mils sauvages sur la base de quelques caractères seulement : la longueur des chandelles, la caducité des épillets, la longueur du pédicelle de l'involucre, la taille des graines, le diamètre des tiges et la largeur des feuilles.

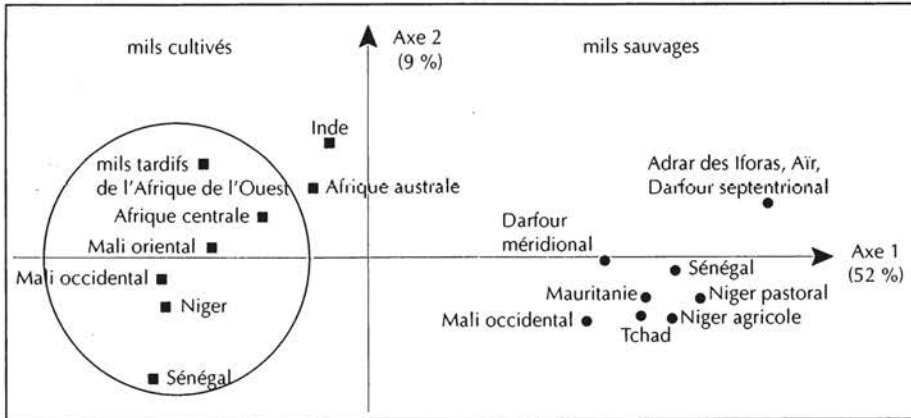


Figure 1. Position des moyennes des grandes familles morphologiques de mil dans le plan 1-2 d'une analyse en composantes principales de 14 caractères botaniques à forte héritabilité, d'après MARCHAIS et al. (1993). Les mils cultivés en Afrique de l'Ouest au sens large sont entourés d'un cercle.

Au sein des mils cultivés, il ressort cinq groupes pour l'Afrique de l'Ouest, un groupe pour l'Afrique centrale, un groupe pour l'Inde et un groupe pour l'Afrique australe, ces deux derniers étant plus homogènes. La discrimination au sein des mils sauvages est moins marquée du fait de leur plus grande homogénéité morphologique.

## La diversité phénologique des mils cultivés

Les données sur la phénologie de la floraison de plus de 12 000 cultivars obtenues par l'ICRISAT (L. Marchais *et al.*, comm. pers.) ont permis de mettre en évidence deux phénomènes qui concourent au déclenchement de la floraison du mil : d'une part, l'exigence d'un nombre minimal de degrés-jours et d'autre part, la sensibilité à la longueur du jour. La figure 2 regroupe les résultats observés pour deux cycles de culture, l'un réalisé en jours longs et l'autre en jours courts. On constate que, pour chacune des régions étudiées, il existe une gamme de variétés qui en jours longs fleurissent entre 45 et 140 jours. Cette diversité peut être imputable aux différences de longueur de la saison des pluies dans chacune des régions.

D'autre part, en jours courts, les cycles ne dépassent pas 90 jours, excepté pour les cultivars du sud de l'Inde. Les cultivars qui fleurissent entre 40 et 60 jours pour une culture en jours longs fleurissent sur une période équivalente en somme de degrés-jours s'ils sont cultivés en jours courts. Ces cultivars peuvent donc être considérés comme thermosensibles, mais non photosensibles. Pour les cultivars qui fleurissent au-delà de 60 jours pour une culture en jours longs, les comportements en jours courts sont différents selon les régions.



Globalement, trois cas peuvent se présenter. Le premier concerne les cultivars du groupe Sud, pour lesquels, jusqu'à environ 80 jours, l'augmentation du cycle en jours longs s'accompagne d'une augmentation du cycle en jours courts puis la longueur du cycle en jours courts n'augmente plus, ce qui peut s'expliquer par une exigence thermopériodique maximale avec l'apparition d'une photosensibilité. Le deuxième est représenté par les cultivars de l'Afrique de l'Ouest, au sein desquels on distingue, d'une part, les mils qui fleurissent en moins de 80 jours et que l'on peut qualifier de thermopériodiques (mais avec une exigence moindre en degrés-jours que les mils du

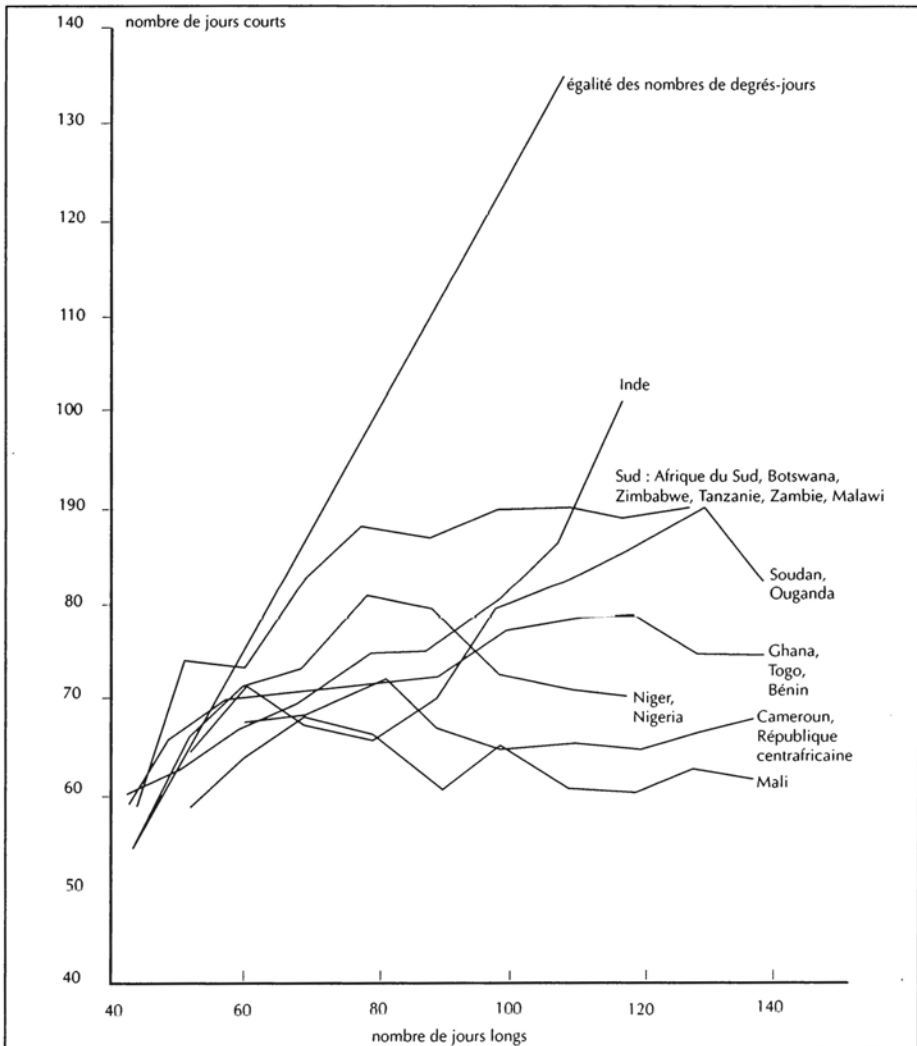


Figure 2. Evolution comparée des floraisons en jours longs et en jours courts suivant l'origine géographique des mils observés, d'après L. Marchais et al. (comm. pers.).

groupe Sud) — ce sont les mils traditionnels précoces — et, d'autre part, les mils qui fleurissent après 80 jours, que l'on peut qualifier de photosensibles et dont les exigences thermopériodiques sont minimales — ce sont les mils appelés traditionnellement mils tardifs. Le troisième cas concerne les variétés de l'Inde, qui représentent une situation intermédiaire, pour lesquelles l'allongement du cycle en jours longs s'accompagne d'un allongement du cycle en jours courts, de façon continue : il s'agit là en fait d'un gradient nord-sud, les cycles les plus longs caractérisant les cultivars du sud de l'Inde. Dans cette situation, il est difficile de parler de mils photosensibles et non photosensibles.

Les autres groupes notés sur la figure se rapprochent plus ou moins de l'un des grands groupes que nous venons de décrire.

Globalement, on retrouve dans cette analyse phénologique de la floraison des mils cultivés des regroupements proches de ceux qui ont été observés avec les autres descripteurs de la diversité. La distinction entre mils précoces et mils tardifs n'apparaît de façon évidente qu'en Afrique de l'Ouest. Les cultivars les plus caractéristiques sont les mils très précoces et non photosensibles de la région du Ghana, du Togo et du Bénin, les mils à cycle long mais non photosensibles du Botswana, les mils photosensibles et à cycle long du Mali et les mils particuliers de Sierra Leone à cycle très long.

## La variabilité biochimique et moléculaire

L'essentiel des résultats sur le polymorphisme enzymatique des mils provient des travaux de TOSTAIN (1994), qui portent sur 549 accessions, dont 361 mils cultivés et 188 mils sauvages. Pour les huit systèmes enzymatiques étudiés douze locus se sont révélés polymorphes, pour un total de 46 allèles, soit une moyenne de 3,8 allèles par locus.

Les résultats obtenus montrent que, dans leur zone d'origine, mil sauvage et mil cultivé ne sont pas différenciables aux locus étudiés par des allèles qui seraient fixés dans un groupe et absents de l'autre. Les deux groupes diffèrent uniquement par les fréquences alléliques à ces locus.

Les analyses effectuées sur les mils sauvages mettent en évidence cinq groupes, qui correspondent à des entités géographiques bien distinctes les unes des autres : le groupe « ouest » (I) comprend des populations du Sénégal, de Mau-ritanie et de l'ouest du Mali ; le groupe « centre » (II) correspond à des populations de l'Adrar des Iforas et du Gourma au Mali, de l'Oudalan au Burkina, des vallées de l'Azawak et de l'Ader Doutchi au Niger ; le groupe « Aïr » (III) est constitué de la population du massif de l'Aïr au Niger ; le groupe « Tchad-ouest » (IV) regroupe des populations situées au sud du lac Tchad ; le groupe « Darfour » (V) rassemble des populations de l'est du lac Tchad et du Darfour au Soudan.

Pour les mils cultivés, différentes analyses enzymatiques partielles ont conduit l'auteur à répartir l'ensemble des échantillons en sept groupes, sur une base à la fois géographique et physiologique (longueur du cycle) : le groupe A (mils précoces présents du Sénégal au Niger occidental), le groupe B (mils précoces du Mali occidental), le groupe C (mils précoces rencontrés du Niger au Soudan), le groupe D (mils précoces du Togo et du Ghana), le groupe E (mils tardifs d'Afrique de l'Ouest), le groupe F (mils précoces et tardifs d'Afrique de l'Est et d'Afrique australe), le groupe G (mils précoces et tardifs d'Inde). Les analyses multivariées sur l'ensemble ne confirment pas dans son intégralité cette classification, certains groupes, A, C et E notamment, se chevauchant en partie. L'analyse discriminante donne 72 à 76 % d'individus bien classés pour ces trois groupes, contre 83 à 94 % pour les autres groupes.

Une analyse globale a réuni 188 échantillons de la forme sauvage répartis sur l'ensemble de son aire de distribution et 123 échantillons de la forme cultivée appartenant aux quatre groupes A, B, C et D. Elle montre, pour les mils sauvages, une diversité génétique comparable à celle des mils cultivés : l'indice de diversité de Nei est égal à 0,249 pour les mils sauvages et à 0,256 pour les mils cultivés. En revanche, la structuration des mils sauvages est moins forte que celle des mils cultivés ( $G_{ST} = 0,13$  contre 0,17). Ce sont les mils cultivés des groupes A et B (précoces de l'Afrique de l'Ouest) qui sont les plus proches des mils sauvages (figure 3a), alors que les mils du groupe C (précoces présents du Niger jusqu'au Soudan) sont en position plus excentrée. En résumé, deux grands ensembles peuvent être distingués : le premier rassemble les formes sauvages et les mils cultivés des groupes A, B et D ; le second renferme le reste des mils cultivés (mils tardifs d'Afrique de l'Ouest, mils d'Afrique de l'Est et d'Afrique australe, mils d'Inde). Un cas particulier est représenté par la variété Tiotandé, originaire du Sénégal et cultivée en contre-saison, qui se détache de l'ensemble (figure 3b).

Des analyses portant sur des échantillons qui ne représentent qu'une part réduite de la diversité génétique des mils sauvages et des mils cultivés ont été réalisées avec l'aide d'autres descripteurs moléculaires du polymorphisme (séquençage, RFLP, RAPD). La comparaison de quelques échantillons de mil par RFLP a montré que l'ADN chloroplastique sondé est très peu polymorphe et que l'ADN ribosomique l'est uniquement chez le mil sauvage (GEPTS et CLEGG, 1989). Une analyse du polymorphisme RFLP dans la région *ADH-1* a mis en évidence une variabilité génétique intra et interpopulations ainsi que la spécificité de certains profils de restriction des formes sauvages (PILATE-ANDRE *et al.*, 1993). Le séquençage des allèles du gène *ADH-1* de ces échantillons n'a montré aucune différence significative entre mil sauvage et mil cultivé (GAUT et CLEGG, 1993). TOSTAIN (1996) a étudié la diversité de 14 populations de mil (4 sauvages, 9 cultivées et 1 accession de Tanzanie intermédiaire entre formes sauvages et cultivées) à l'aide des marqueurs RAPD. Les résultats confirment globalement les classifications obtenues à l'aide des marqueurs enzymatiques, mais, là encore, les marqueurs RAPD révèlent une très faible diversité

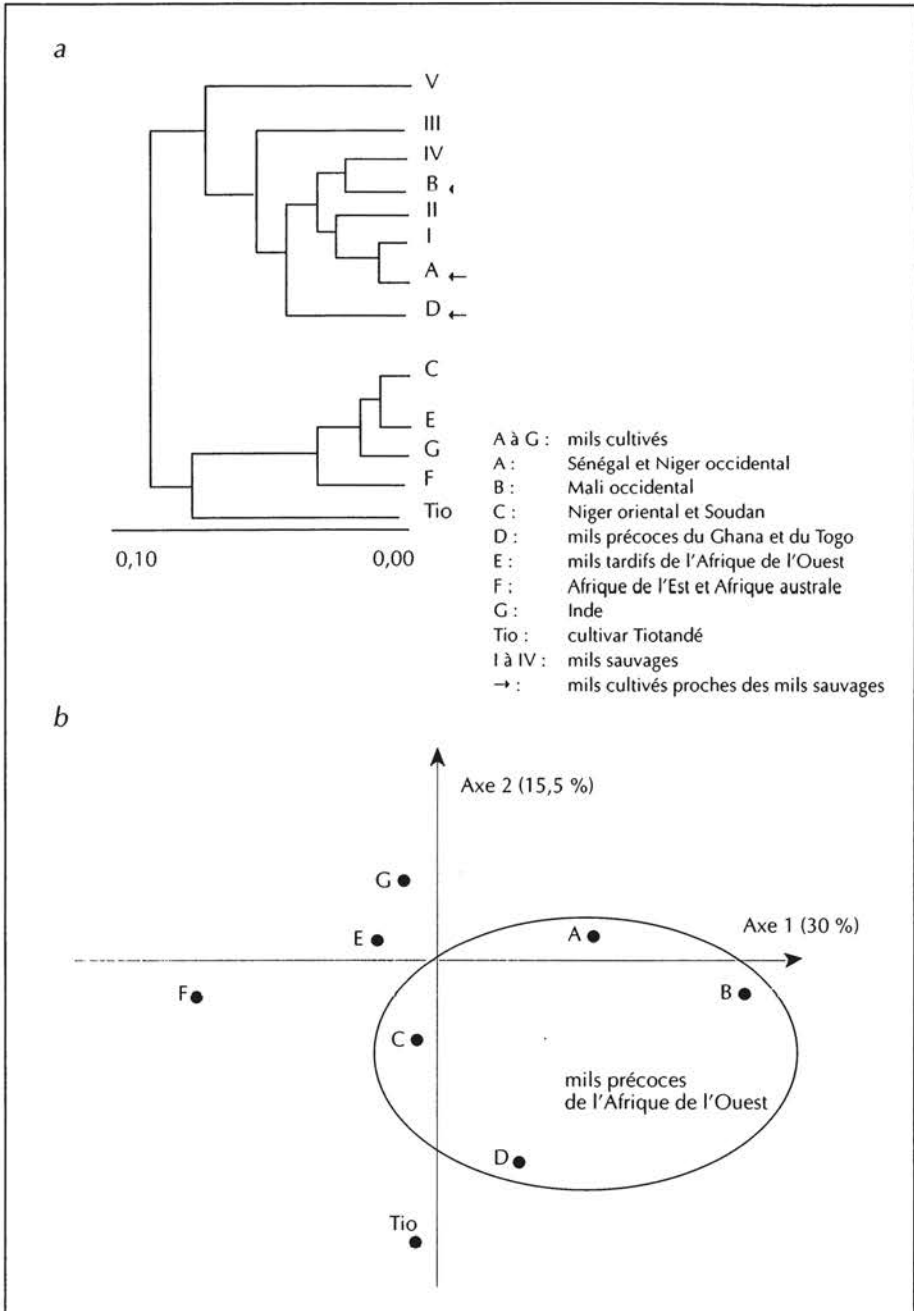


Figure 3. a. Dendrogramme des distances de Nei entre groupes de mils cultivés et de mils sauvages, d'après TOSTAIN (1998). b. Analyse en composantes principales de 361 accessions de mils cultivés. Seules les projections des centres de gravité des 7 groupes de mils cultivés et du cultivar Tiotandé sont représentées, d'après TOSTAIN (1998).

intrapopulation et ne permettent pas de séparer nettement les formes sauvages des formes cultivées. L'ensemble des résultats obtenus à l'aide des différents descripteurs du polymorphisme, qu'ils soient biochimiques (isoenzymes) ou moléculaires (RFLP, séquençage RAPD), convergent quant aux difficultés à différencier le mil sauvage et le mil cultivé.

### Les relations entre les différents niveaux de variabilité

La classification des mils sauvages et des mils cultivés obtenue à partir des données du polymorphisme enzymatique a été rapprochée des classifications qui avaient été établies sur la base des caractères morphologiques. On observe ainsi une assez bonne correspondance avec la classification de MARCHAIS *et al.* (1993).

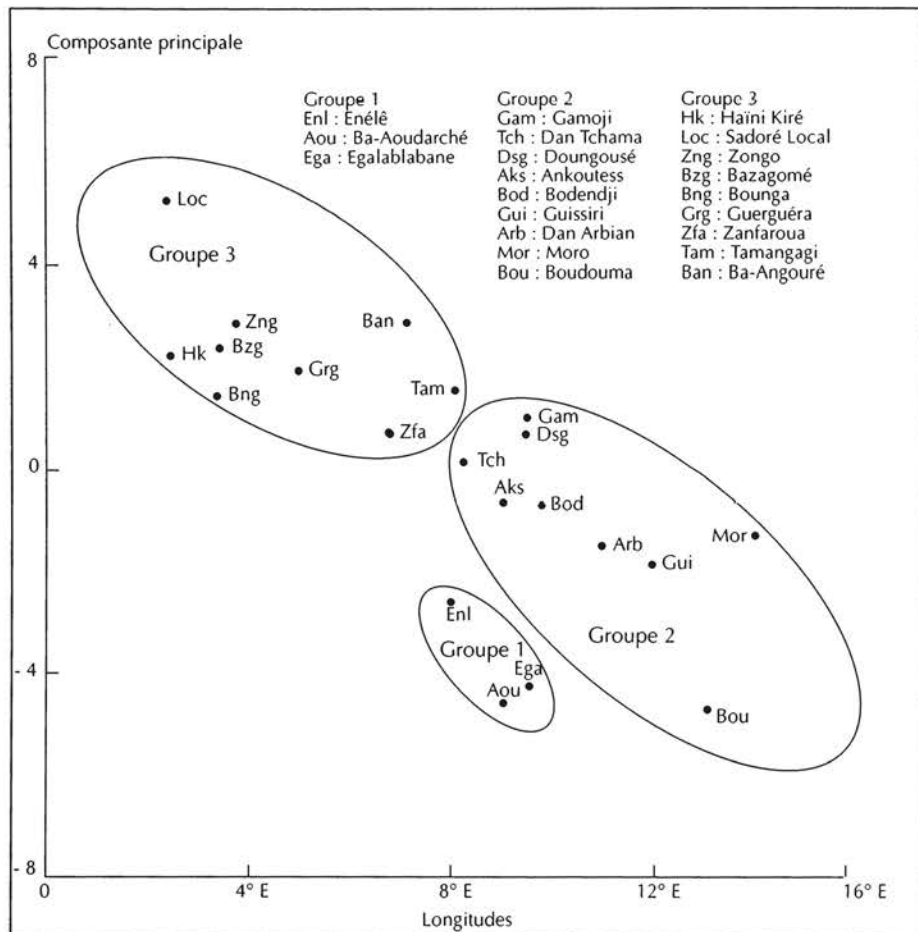


Figure 4a. Projection suivant la longitude et le premier axe de l'analyse en composantes principales des 21 cultivars locaux du Niger, d'après SIAKA *et al.* (1996).

Dans le cas particulier des mils du Niger, une analyse comparative a été réalisée sur un échantillon de 21 cultivars représentatifs des mils cultivés au Niger à l'aide de douze caractères quantitatifs et de trois caractères qualitatifs (SIAGA *et al.*, 1996). Les résultats (figure 4a) montrent une structuration de la diversité en trois groupes : les cultivars du groupe 1 sont originaires de la zone désertique, les cultivars du groupe 2 sont cultivés dans l'est du pays entre les longitudes 8° E et 13° E, les cultivars du groupe 3 sont cultivés dans l'ouest du pays entre les longitudes 1° E et 8° E. Cette structuration correspond aux trois groupes (figure 4b) mis en évidence par TOSTAIN (1994) sur la base du polymorphisme enzymatique : mils tardifs (groupe 1), mils précoces à épis courts (groupe 2) et mils précoces à épis longs (groupe 3).

D'autres critères ont été également utilisés pour décrire la diversité des mils. Ainsi, L. Marchais (comm. pers.) a testé près de 200 cultivars pour leur capacité à restaurer la fertilité mâle sur le cytoplasme mâle-stérile décrit par MARCHAIS et PERNES (1985). Les résultats mettent en évidence une différenciation géographique : les mils de l'Inde et de l'Afrique australe constituent deux groupes homogènes, qui présentent un très faible taux de restauration, alors

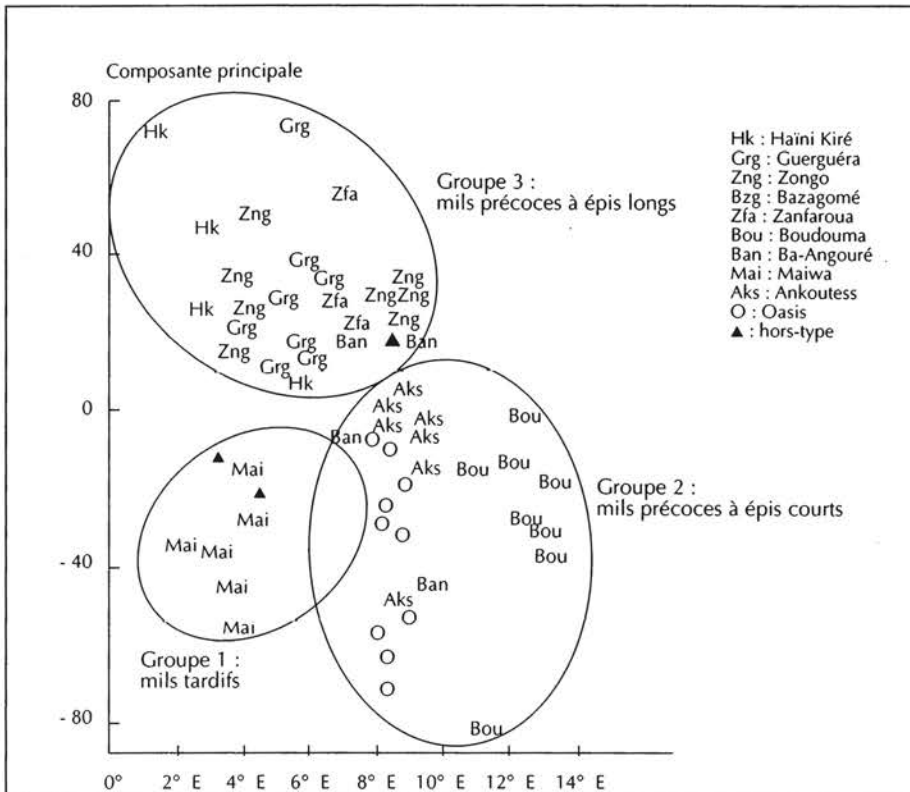


Figure 4b. Projection suivant la longitude et le premier axe de l'analyse en composantes principales de 66 accessions du Niger, d'après TOSTAIN (1994).

qu'en Afrique de l'Ouest on retrouve la distinction entre les mils précoces, qui montrent une grande variabilité du pouvoir restaurateur, et les mils tardifs, qui sont en majorité des mainteneurs de stérilité.

L'ensemble de ces résultats — données morphologiques, phénologiques, biochimiques et cytoplasmiques — tend à confirmer les informations recueillies par les prospecteurs en ce qui concerne la différenciation géographique des mils. Cette différenciation résulte, d'une part, de la sélection naturelle et de l'adaptation des mils sauvages et des cultivars à des environnements divers et, d'autre part, de la domestication, processus dans lequel le rôle du paysan est déterminant. Cette structuration géographique de la diversité des mils sauvages et cultivés doit donc être prise en compte prioritairement dans l'élaboration de nouvelles méthodes de conservation, de gestion et de mise en valeur des ressources génétiques.

Pour préciser l'origine de la domestication du mil, des études complémentaires, fondées sur des descripteurs du polymorphisme génétique moins sensibles que les isoenzymes aux flux de gènes actuels, comme l'ADN chloroplastique, devraient être une source d'informations nouvelles. De plus, il serait souhaitable d'étendre l'échantillonnage à l'ensemble de l'aire de distribution du mil sauvage, jusqu'au Soudan et à l'Éthiopie. En effet, cette région est extrêmement importante pour comprendre la domestication du mil. Elle correspond au centre d'origine du niébé et du sorgho, qui, comme le mil, sont aussi cultivés en Inde, contrairement aux espèces, telles que le riz *Oryza glaberrima* et les ignames du complexe *Dioscorea cayenensis-D. rotundata*, dont la domestication aurait eu lieu en Afrique de l'Ouest.

## La domestication et la gestion des ressources génétiques

La distribution géographique du mil sauvage, limitée à l'Afrique sahélienne, conduit à penser que c'est dans cette région qu'il a été domestiqué pour donner la céréale que nous connaissons aujourd'hui. Les plus anciens vestiges de mil cultivé en contact avec du mil sauvage ont été trouvés en Mauritanie et remonteraient à plus de 3 000 ans (AMBLARD et PERNES, 1989), alors que des empreintes de mil sauvage ont été découvertes sur des poteries datant d'environ 5 000 ans dans le centre du Soudan (STEMLER, 1990). Différentes analyses génétiques ont montré que le syndrome de domestication dépend d'un petit nombre de gènes liés — longueur du pédicelle de l'involucre, caducité de l'épillet à maturité, taille des enveloppes de la graine et présence ou absence des soies —, dont les allèles à l'état récessif permettent l'expression du phénotype cultivé (PERNES *et al.*, 1984), ce qui pourrait expliquer une domestication rapide.

Dans le cas du mil, espèce allogame, il existe des flux géniques entre formes cultivées et formes sauvages, qui sont plus ou moins contrôlés par des barrières à la reproduction pré et postzygotiques. Ils donnent naissance à des formes hybrides viables, que l'agriculteur a dû éliminer au cours de la domestication, comme il le fait encore actuellement, pour maintenir le phénotype cultivé. Mais cette sélection reste douce, les hybrides pouvant être récoltés en cas de disette. Ces brassages permanents brouillent les traces d'une éventuelle structuration génétique issue directement de la domestication. Dans ces conditions, les calculs de distances génétiques à partir des données du polymorphisme enzymatique sont difficilement utilisables pour situer le centre de domestication de la forme cultivée. En effet, les différences de distances génétiques pourraient principalement refléter des variations dans l'efficacité des barrières à la reproduction entre formes sauvages et formes cultivées. L'hypothèse d'une domestication accompagnée du renforcement de ces barrières aux échanges géniques entre la forme ancestrale et la forme cultivée ne peut pas être écartée et va alors à l'encontre de la démarche associant un centre de domestication à la distance génétique la plus faible entre populations de la forme sauvage actuelle et populations de la forme cultivée (BEZANÇON *et al.*, 1997). C'est pourquoi, plutôt que de rechercher un centre de domestication du mil dans une région délimitée par la Mauritanie, le Sénégal et l'ouest du Mali (TOSTAIN, 1992), il est préférable de conserver l'idée d'un non-centre au sens de HARLAN (1971). La domestication du mil, associée à des phénomènes migratoires, se serait produite dans différentes zones géographiques, simultanément ou à des époques différentes, chaque civilisation agraire sélectionnant un morphotype spécifique particulièrement bien adapté aux conditions locales — climat, sols et techniques culturales. Les migrations sur de longues distances, à l'échelle continentale pour l'Afrique australe et intercontinentale pour l'Inde, ont abouti à la constitution de groupes plus homogènes et bien différenciés des formes d'origine.

À l'ICRISAT, plusieurs techniques ont été préconisées pour maintenir la diversité de la collection mondiale. La technique d'intercroisements entre un nombre défini de plantes (100) d'un même cultivar (APPA-RAO, 1980) présente des contraintes liées aux dispositifs nécessaires à l'isolement des parcelles de multiplication. Ces problèmes peuvent être surmontés en réalisant des croisements contrôlés avec des mélanges de pollen, qui engendrent alors un surcroît de travail. La technique d'autofécondation (BURTON, 1983) représente une économie de temps et de main-d'oeuvre : l'autofécondation est pratiquée sur les plantes qui servent à la caractérisation des cultivars et permet de conserver les gènes, dominants et récessifs, qui sont utilisés pour classer les cultivars. Mais dans ce cas, les plantes produites sont moins vigoureuses, de taille réduite et leurs graines sont plus petites et moins nombreuses. La technique des pools favorise les échanges géniques et permet d'augmenter l'adaptabilité. Elle repose sur la constitution de pools à partir d'un mélange de graines de cultivars proches — cultivars d'une même région ou cultivars présentant des caractéristiques identiques —, le mélange étant cultivé dans des conditions d'isolement.



La technique d'intercroisements et celle des pools nous paraissent particulièrement bien adaptées au cas du mil. En revanche, la technique d'autofécondation a des inconvénients majeurs, elle pose en particulier des problèmes de dérive génétique.

Pour l'ICRISAT, l'objectif est de produire suffisamment de graines pour l'évaluation, pour l'utilisation en sélection, pour les échanges et pour la conservation. Il existe deux types de conservation des graines : la première est une conservation à long terme (température de  $-18^{\circ}\text{C}$  et 5 % d'humidité relative), destinée aux collections de base ; la seconde, à moyen terme (température de  $+5^{\circ}\text{C}$  et 20 % d'humidité relative), concerne les collections pour l'évaluation, la multiplication et la distribution. On conçoit aisément que de telles méthodes appliquées aux collections de mils représentent un travail énorme et des coûts élevés.

La structuration de la diversité génétique des mils, telle qu'elle a été mise en évidence, doit servir de base pour d'autres projets de conservation des ressources génétiques. Dans ce contexte, la constitution de *core collections* peut être conseillée.

Habituellement, pour constituer une *core collection*, il est recommandé, après une stratification sur la base des données taxonomiques, géographiques, quelquefois morphologiques, enzymatiques ou agronomiques, de procéder à un échantillonnage d'environ 10 % de la collection initiale par la sélection aléatoire dans chaque groupe d'un nombre d'échantillons proportionnel au logarithme de la taille du groupe (BROWN, 1989). Dans le cas où les données intra-groupes sont disponibles, MARCHAIS (1996), dans le cadre d'une étude sur les contributions parentales relatives qui maximisent la valeur moyenne de certains caractères comme le rendement en grains pour la fabrication d'une variété synthétique, propose une méthode qui permet d'identifier exactement — et non plus au hasard — les échantillons qui constitueront la *core collection* ainsi que leurs proportions. Cette sélection s'effectue en maximisant la diversité totale de  $N_{Ei}$  (1973) du *bulk* pondéré.

L'algorithme fait appel au raisonnement utilisé en génétique des populations pour des populations panmictiques sur plusieurs générations, avec une valeur sélective constante (CROW et KIMURA, 1970). Il établit une relation entre les fréquences alléliques aux générations  $n$  et  $n + 1$  et la valeur sélective des différents génotypes. Cette méthode a été appliquée à une collection de 95 échantillons de mil sauvage collectés au Niger, qui a été décrite pour huit systèmes enzymatiques — 12 locus et 46 allèles — par TOSTAIN (1992). Cet ensemble présente une diversité enzymatique relativement homogène par rapport aux autres groupes géographiques. Un état d'équilibre a été atteint après 700 itérations et les résultats montrent que la diversité de  $N_{Ei}$  pour la totalité de la collection est égale à 3,58 dans le cas où tous les échantillons sont représentés en proportions égales, alors qu'elle est de 6,48 après application de l'algorithme. Ce niveau de diversité est obtenu avec seulement huit accessions.

Cette méthode s'est donc révélée très efficace pour accroître la diversité tout en éliminant les redondances dans ce groupe des mils sauvages du Niger. Le même procédé pourrait être appliqué à la moyenne des fréquences alléliques de chacun des groupes, afin de maximiser la diversité d'une *core collection* pour l'ensemble de la collection. Enfin, cette procédure de création de *core collections* peut être utilisée pour les caractères qualitatifs comme pour les caractères quantitatifs, ceux-ci devant être répartis en classes.

La principale objection que l'on peut émettre à l'encontre de cette méthode est que, dans la plupart des cas, les accessions ne sont pas suffisamment analysées avant la constitution d'une *core collection*. La procédure pourrait alors être mise en place sur des échantillons prélevés aléatoirement selon les principes de BROWN (1989) et après l'analyse de ces échantillons.

L'application de cette méthodologie à la collection mondiale de mil décrite sur la base de marqueurs enzymatiques et moléculaires permettrait de supprimer les nombreuses redondances, notamment en ce qui concerne les cultivars, tout en accroissant le niveau global de diversité.

A l'échelle régionale, la mise au point d'une méthodologie de conservation *in situ*, associée à la gestion et à l'amélioration participative des ressources génétiques des mils, serait également une solution. L'IRD conduit actuellement des travaux de ce type dans la région sud-ouest du Niger. L'aspect dynamique d'une telle conservation *in situ* l'oppose à toutes les autres méthodologies, qui permettent la conservation de structures figées. Ici, l'évolution des cultivars traditionnels n'est pas bloquée : les flux de gènes avec les formes sauvages, si elles sont présentes, continuent, les brassages se poursuivent également entre cultivars traditionnels (échanges entre paysans, migrations) et avec les cultivars génétiquement améliorés fournis dans certains cas par les instituts de recherche agricole. L'ensemble de la diversité poursuit son évolution sous l'influence des facteurs de l'environnement, physiques et humains, et l'on se rapproche au maximum des méthodes utilisées par les paysans pour l'entretien de leurs cultivars traditionnels. La connaissance et le contrôle des mécanismes qui gouvernent ces différents facteurs pourront permettre une meilleure gestion des ressources génétiques.

## Annexe

### Etudes agromorphologiques

Les évaluations les plus récentes réalisées sur les mils cultivés ont porté sur 1 938 échantillons originaires de 33 pays observés pour 20 descripteurs, dans un premier temps, puis sur 2 458 échantillons originaires de 38 pays observés pour 18 descripteurs, dans un deuxième temps (NBPGR et ICRISAT, 1993).

L'étude de MARCHAIS *et al.* (1993) a été réalisée sur 267 plantes cultivées et 118 plantes sauvages issues d'échantillons représentatifs de l'ensemble des collections. La culture a été effectuée au Niger pendant la saison des pluies (semis le 25 juin). Chaque échantillon était représenté par une ligne de 10 pieds. Les notations ont porté sur 14 caractères morphologiques : 12 caractères quantitatifs (hauteur de la tige principale, longueur et largeur de la troisième feuille sous l'épi de la tige principale, longueur de l'épi de la tige principale, diamètre du rachis de l'épi de la tige principale, longueur du pédicelle d'involucre, longueur des soies de l'involucre, longueur des arêtes, longueur des glumelles des soies de l'involucre, nombre d'épillets par involucre, longueur et largeur du grain) et 2 caractères qualitatifs (pilosité du limbe foliaire et vitrosité de l'albumen).

### Analyses phénologiques

Plus de 12 000 cultivars ont été évalués sur leur floraison (durée semis-floraison) dans deux conditions de culture : saison des pluies (semis le 25 juin, en jours longs) et contre-saison (semis le 25 novembre, en jours courts). Chaque échantillon était cultivé sur une ligne de 4 mètres (environ 20 pieds) et la floraison a été notée lorsque 50 % des pieds avaient produit un épi en floraison femelle (L. MARCHAIS *et al.*, comm. pers.). Cette expérimentation a été réalisée sur la station de l'ICRISAT à Patancheru, en Inde.

### Analyses biochimiques et moléculaires

Au total, 549 échantillons ont été analysés : 361 représentaient les formes cultivées originaires de 29 pays d'Afrique et de l'Inde et 188, collectés dans 8 pays d'Afrique subsaharienne, représentaient les populations de la forme sauvage dans la quasi-totalité de son aire de distribution géographique. Les systèmes enzymatiques utilisés sont au nombre de 8 : ADH, CAT, EST, GOT, MDH, PGD, PGI et PGM (TOSTAIN, 1994).

### Analyse des données

Les méthodes d'analyse statistique utilisées ont été l'analyse en composantes principales et l'analyse discriminante, réalisées avec le logiciel Stat-ITCF. Les paramètres du polymorphisme biochimique ont été calculés à l'aide du logiciel Biosys 1-7. Différentes distances génétiques ont également été calculées : Rogers, Cavalli-Sforza, Wright.

## Références bibliographiques

- AMBLARD S., PERNES J., 1989. The identification of cultivated pearl millet (*Pennisetum*) amongst plant impressions on pottery from Oued Chebbi (Dhar Oualata, Mauritania). *African Archeological Review*, 7 : 117-126.
- APPA-RAO S., 1980. Progress and problems of pearl millet germplasm maintenance. *In* : Trends in genetical research on *Pennisetum*, V.P. Gupta et J.L. Minocha éd., Ludhiana, Inde, Punjab Agricultural University, p. 279-282.
- BEZANÇON G., RENNO J.F., ANAND KUMAR K., 1997. Le mil. *In* : L'amélioration des plantes tropicales, A. Charrier et al. éd., Montpellier, France, Cirad-Orstom, collection Repères, p. 457-482.
- BONO M., 1973. Contribution à la morphosystématique des *Pennisetum* annuels cultivés pour leur grain en Afrique occidentale francophone. *L'Agronomie tropicale*, 28(3) : 229-356.
- BROWN A.H.D., 1989. Core collections: a practical approach to genetic resources management. *Genome*, 31 : 818-824.
- BRUNKEN J.N., DE WET J.M.J., HARLAN J.R., 1977. The morphology and domestication of pearl millet. *Economic Botany*, 31 : 163-174.
- BURTON G.W., 1983. Breeding pearl millet. *In* : Plant breeding reviews: volume 1, J. Janik éd., Westport, Etats-Unis, AVI Publishing, p. 162-182.
- CLEMENT J.C., 1985. Les mils pénicillaires de l'Afrique de l'Ouest : prospections et collectes. Rome, Italie, IBPGR, 231 p.
- CROW J.F., KIMURA M., 1970. An introduction to population genetics theory. New York, Etats-Unis, Harper and Row, 591 p.
- FAO, 1996. Annuaire production : 1995. Rome, Italie, FAO.
- GAUT B.S., CLEGG M.T., 1993. Nucleotide polymorphism in the *ADH-1* locus of pearl millet (*Pennisetum glaucum*, Poaceae). *Genetics*, 135(4) : 1091-1097.
- GEPTS P., CLEGG M.T., 1989. Genetic diversity in pearl millet, *Pennisetum glaucum* (L.) R. Br., at the DNA sequence level. *Journal of Heredity*, 80 : 203-208.
- HANNA W.W., 1987. Utilization of wild relatives of pearl millet. *In* : International pearl millet workshop, J.R. Witcombe éd., Patancheru, Inde, ICRISAT, p. 33-42.
- HANNA W.W., 1990. Transfer of germplasm from the secondary to the primary gene pool in *Pennisetum*. *Theoretical and Applied Genetics*, 80 : 303-308.
- HARLAN J.R., 1971. Agricultural origins: centers and non-centers. *Science*, 14 : 468-474.
- HARLAN J.R., DE WET J.M.J., 1971. Toward a rational classification of cultivated plants. *Taxon*, 20(4) : 509-517.
- MARCHAIS L., 1982. La diversité phénotypique des mils pénicillaires cultivés au Sénégal et au Mali. *L'Agronomie tropicale*, 37 : 68-80.
- MARCHAIS L., 1996. Parental proportion maximizing the mean value of a parameter in a panmictic population can be useful in plant breeding. *Agronomie*, 16 : 257-264.

MARCHAIS L., PERNES J., 1985. Genetic divergence between wild and cultivated pearl millets (*Pennisetum typhoides*). 1. Male sterility. *Zeitschrift für Pflanzenzüchtung*, 95 : 103-112.

MARCHAIS L., TOSTAIN S., AMOUKOU I., 1993. Signification taxonomique et évolutive de la structure génétique des mils pénicillaires. *In* : Le mil en Afrique, S. Hamon éd., Paris, France, Orstom, Colloques et séminaires, p. 119-128.

NBPGR, ICRISAT, 1993. Evaluation of pearl millet germplasm : part 1 and part 2.

NEI M., 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, 70 : 3321-3323.

PERNES J., COMBES D., LEBLANC J.M., 1984. Le mil. *In* : Gestion des ressources génétiques des plantes. 1. Monographies, J. Pernès éd., Paris, France, ACCT, p. 159-210.

PILATE-ANDRE S., LAMY F., SARR A., 1993. Diversité génétique des mils détectable par RFLP au niveau de la région du gène *ADH-1*. *In* : Le mil en Afrique, S. Hamon éd., Paris, France, Orstom, Colloques et séminaires, p. 67-75.

PORTERES R., 1950. Vieilles agricultures de l'Afrique intertropicale : centres d'origine et de diversification variétale primaire et berceaux de l'agriculture antérieurs au XVI<sup>e</sup> siècle. *L'Agronomie tropicale*, 5(9-10) : 489-507.

PORTERES R., 1976. African cereals: *Eleusine*, fonio, black fonio, teff, *Bracchiaria*, *Paspalum*, *Pennisetum* and African rice. *In* : Origins of African plant domestication, J.R. Harlan et al. éd., La Haye, Pays-Bas, Mouton, p. 409-452.

SIAGA S., OUENDEBA B., ANAND KUMAR K., 1996. Caractérisation des cultivars locaux du Niger. *In* : Premières journées biologiques et agronomiques du Niger, 20-25 septembre 1996.

STEMLER A., 1990. A scanning electron microscopic analysis of plant impressions in pottery from the sites of Kadero, El Zakiab, Um Direiwa and El Kadada. *Archéologie du Nil moyen*, 4 : 87-105.

TOSTAIN S., 1992. Enzyme diversity in pearl millet (*Pennisetum glaucum* L.). 3. Wild millet. *Theoretical and Applied Genetics*, 83 : 733-742.

TOSTAIN S., 1994. Evaluation de la diversité génétique des mils pénicillaires diploïdes, *Pennisetum glaucum* (L.) R. Br., au moyen de marqueurs enzymatiques : étude des relations entre formes sauvages et cultivées. Paris, France, Orstom, Travaux et documents microédités n° 124, 331 p.

TOSTAIN S., 1996. Genetic diversity of pearl millet (*Pennisetum glaucum*) estimated by random amplified DNA markers and compared with isoenzymatic diversity. *In* : Réunion sur les plantes tropicales. Montpellier, France, EUCARPIA-Cirad, p. 255.

TOSTAIN S., 1998. Le mil, une longue histoire : hypothèses sur sa domestication et ses migrations. *In* : Plantes et paysages d'Afrique : une histoire à explorer, M. Chastanet éd., Paris, France, Karthala-CRA, p. 461-490.

VAN DER ZON A.P.M., 1992. Graminées du Cameroun. 2. Flore. Wageningen, Pays-Bas, Agricultural University, Papers n° 92-1, 557 p.

# Le riz asiatique

---

Jean Christophe Glaszmann, Laurent Grivet,  
Brigitte Courtois, Jean-Louis Noyer, Claude Luce,  
Michel Jacquot, Laurence Albar,  
Alain Ghesquière, Gérard Second

**H**éritage des plus anciennes civilisations — en Chine, près du fleuve Yang-Tseu, des traces de riziculture remontent à environ sept millénaires —, le riz représente aujourd’hui dans les pays en développement près de 50 % de la production de céréales, largement devant le blé et le maïs. La production mondiale de riz paddy, riz non décortiqué, avoisine 540 millions de tonnes, récoltées sur environ 150 millions d’hectares, dont 90 % en Asie. Dans cette région, la consommation annuelle de riz usiné par habitant est fréquemment supérieure à 100 kilos, et peut atteindre près de 200 kilos, comme au Myanmar ou au Laos. Sur d’autres continents, la consommation est plus variable mais peut s’élever à 100 kilos, en particulier à Madagascar, au Sierra Leone ou au Surinam. La consommation annuelle dans l’Union européenne est, quant à elle, de l’ordre de 5 kilos par habitant. La plus grande partie du riz est consommée localement, les échanges commerciaux mondiaux portant sur moins de 5 % de la production. Cependant, avec le progrès économique, les populations fortement consommatrices de riz diversifient peu à peu leur nourriture, tandis que les populations peu consommatrices l’intègrent progressivement dans leurs habitudes alimentaires. Selon les prévisions démographiques, les besoins en riz augmenteront de 70 % d’ici 2025.

Le riz est une plante semi-aquatique : elle tolère des conditions de culture aquatique, mais n’en dépend pas absolument. Dès l’origine, le développement

de la riziculture a été conditionné par une fertilité naturelle suffisante des sols et par une concurrence faible de l'enherbement. Les situations les plus favorables se rencontraient sur terres inondables et sur terrains défrichés en forêt.

La culture sur terres inondables a connu le développement le plus important, en particulier grâce aux aménagements pour l'irrigation et le drainage et à la possibilité de cultiver riz sur riz. La culture est mise en place soit par semis direct, soit par repiquage de plants obtenus en pépinières. Les bas-fonds humides des vallées et les larges plaines ont été systématiquement exploités. Une riziculture dite flottante a pu être pratiquée dans les zones sujettes à des crues pouvant atteindre 4 à 5 mètres, grâce à des variétés capables d'allonger leurs tiges de plusieurs centimètres par jour, comme au Bangladesh et au Mali. Au voisinage des côtes, des techniques facilitant le dessalement du sol par l'eau douce ont été élaborées pour dégager une saison de culture du riz, par exemple en Indonésie et en Guinée. A l'intérieur des terres, sur les flancs des collines, ce sont des terrasses qui ont été aménagées pour pratiquer la culture « en casiers », comme au Népal, aux Philippines et à Bali. La riziculture aquatique se rencontre ainsi de l'équateur à des latitudes de 40°, voire 50° nord, en Chine, et, dans les régions tropicales, du niveau de la mer à plus de 2 000 mètres d'altitude, au Népal et à Madagascar, entre autres.

Dans les régions tropicales humides, en dehors des conditions de culture aquatique, une riziculture dite pluviale, car elle dépend exclusivement des pluies pour son alimentation en eau, est aussi pratiquée. Elle s'est d'abord développée sur défriche de forêt. Après abattage et brûlis, le riz est semé à la volée ou en poquets. L'entretien de la culture exige peu d'interventions pendant deux ou trois ans, puis la fertilité naturelle du sol diminue tandis que les mauvaises herbes deviennent envahissantes. On déplace alors la culture, que l'on dit donc itinérante. Cette pratique disparaît peu à peu avec l'augmentation de la pression démographique. Des variantes intéressantes se rencontrent au Brésil, où sur des millions d'hectares le riz pluvial est cultivé avant l'installation de pâturages, mais aussi dans de nombreux pays, où le riz pluvial est cultivé en intercalaire dans de jeunes plantations d'hévéa, de caféier...

## La taxonomie et les ressources génétiques

### Les riz cultivés au sein du genre *Oryza*

Le riz cultivé appartient au genre *Oryza*, très divers, qui comprend plusieurs génomes reconnus sur une base cytogénétique, ainsi que des formes diploïdes et allotétraploïdes. On reconnaît plus de vingt espèces sauvages, dont neuf sont tétraploïdes. Elles se distribuent aujourd'hui sur tous les continents mais il est probable que l'origine du genre soit eurasiatique (SECOND, 1985).

Deux espèces sont cultivées : l'une est d'origine africaine et cantonnée presque uniquement à l'Afrique de l'Ouest, l'autre est d'origine asiatique. Toutes deux sont diploïdes et autogames. L'espèce cultivée africaine, *Oryza glaberrima*, est issue de la domestication en Afrique de l'espèce *Oryza brevili-gulata*. L'espèce cultivée asiatique, *Oryza sativa*, a une structure génétique tout à fait particulière, objet des travaux décrits dans ce chapitre. Sa domestication, à partir de l'espèce *O. rufipogon*, est très ancienne puisqu'on retrouve des restes de grains associés à des activités humaines vieux de près de 4 500 ans au Pakistan, de plus de 8 500 ans dans le nord de l'Inde, d'au moins 6 000 ans en Thaïlande et 7 000 ans en Chine. Les migrations dans un cercle plus large ont été constantes vers l'ouest, au Pakistan et en Iran, vers l'est — l'introduction du riz au Japon date du début du 1<sup>er</sup> siècle — et vers les îles du sud, Sri Lanka, Malaisie, Philippines, Indonésie, Taïwan. Les navigateurs malais ont probablement introduit le riz d'Indonésie à Madagascar au v<sup>e</sup> ou au vi<sup>e</sup> siècle. Les Européens ont bien plus tard introduit le riz asiatique en Afrique tropicale puis en Amérique.

Les riz cultivés présentent une diversité de formes considérable. Les généticiens, notamment au Japon et en Inde, ont cherché très tôt à classer les variétés pour résumer la variabilité globale en quelques sous-ensembles. Pour ce faire, ils ont utilisé les caractères morphologiques et physiologiques et observé le comportement en croisement. On a ainsi rapidement identifié deux sous-espèces, *indica* et *japonica*, avec des types « tempéré » et « tropical » pour la seconde (OKA, 1958). Certains sélectionneurs ont distingué trois types morphologiques *Indica*, *Japonica* et *Javanica* (CHANG, 1976), la situation précise de ce dernier type étant diversement interprétée.

## Les ressources génétiques

Les ressources génétiques du riz sont l'objet d'une grande attention, notamment du fait de l'existence d'un Centre international de recherche agronomique, l'IRRI, International Rice Research Institute, consacré à cette seule filière (CHANG *et al.*, 1984). Dans les années 30 à 50, les principaux pays producteurs de riz en Asie ont constitué des collections nationales importantes : celle de l'Inde rassemblait plus de 20 000 accessions, celle de Chine, près de 40 000. L'IRRI, créé en 1961, a rapidement acquis un rôle prépondérant, servant à la fois de centre d'échange international de semences et de centre de conservation à moyen terme grâce à des installations de qualité. La diffusion croissante des variétés semi-naines à haut rendement a par la suite stimulé la collaboration internationale en matière de collection. Les financements apportés par un autre membre du système international, l'IBPGR (International Board for Plant Genetic Resources), ont facilité cette collaboration. Entre 1971 et 1982, plus de 25 000 échantillons ont été collectés. Les institutions françaises sont intervenues en apportant un appui aux collectes en Afrique. Depuis 1994, un programme de collecte complémentaire dans les régions peu repré-



sentées dans les collections est financé par l'Agence suisse pour le développement et la coopération. La collection internationale compte aujourd'hui plus de 80 000 accessions, dont 95 % appartiennent à l'espèce *O. sativa* (JACKSON, 1997).

La collection internationale comprend la totalité des échantillons conservés dans certaines collections nationales, mais pas toutes. Un duplicata des échantillons est systématiquement déposé au centre de conservation de Fort Collins, aux Etats-Unis, après la première culture des accessions, la *rejuvenation*. Les accessions africaines font l'objet d'une duplication à l'IITA (International Institute for Tropical Agriculture), au Nigeria, et à l'ADRAO (Association pour le développement de la riziculture en Afrique de l'Ouest), en Côte d'Ivoire. A côté de recherches sur la physiologie et la conservation des semences, de nombreux travaux sont conduits pour caractériser et évaluer les accessions pour des descripteurs botaniques et des caractères d'intérêt agronomique (JACKSON, 1997).

## L'organisation de la diversité génétique : quel apport des marqueurs moléculaires ?

Les marqueurs moléculaires apportent beaucoup à la compréhension de la diversité génétique du riz. Les isoenzymes, malgré le nombre limité de locus et d'allèles accessibles, ont fourni les premiers éléments de quantification de la différenciation intravariétale. En Asie, l'espèce cultivée apparaît fondamentalement bipolaire, ce qui est en accord remarquable avec les travaux de OKA (1958), mais avec un continuum de formes intermédiaires (SECOND, 1982). La majorité des variétés traditionnelles peut être divisée en deux groupes majeurs, correspondant aux sous-espèces *indica* et *japonica*. Seules 20 % des variétés n'y sont pas clairement rattachées (GLASZMANN, 1987 ; 1988). Les isoenzymes ont également montré une équidistance triangulaire entre l'espèce cultivée africaine, les formes typiques *indica* et les formes typiques *japonica* (SECOND, 1982). L'application du concept de l'horloge moléculaire en regard du scénario des modifications paléoenvironnementales (SECOND, 1985) a permis une interprétation globale de la structure génétique du genre *Oryza*, avec un scénario de triple domestication à partir d'espèces sauvages formées il y a environ deux millions d'années, à la suite de l'isolement de l'Afrique tropicale par le refroidissement du climat et de l'émergence en Asie de l'Himalaya comme barrière géographique. Les 20 % de variétés d'*Oryza sativa* atypiques pourraient provenir d'introgessions avec des formes sauvages locales des contreforts de l'Himalaya ou de recombinaisons entre formes *indica* et *japonica*. Une forme particulière a été également découverte dans l'ouest de l'Inde ; elle présente des traces d'introgession d'*O. glaberrima* (LOLO et SECOND, 1988). On

peut considérer que l'inventaire de ces formes échappant aux sous-espèces *indica* et *japonica* au sens strict n'est pas terminé.

La classification établie sur la base des isoenzymes concorde globalement avec des classifications locales, empiriques, ainsi qu'avec les classifications synthétiques, comme celles de OKA (1958) et de CHENG *et al.* (1984), établies sur des critères multiples associant composantes morphologiques — largeur du grain, pilosité, par exemple — et physiologiques — tolérance des plantules au froid (ZHOU *et al.*, 1988).

Avec l'avènement des nouvelles techniques moléculaires se sont succédé des études fondées sur plusieurs types de marqueurs moléculaires, qui ont permis de construire différentes classifications infraspécifiques. Chaque schéma nouvellement obtenu a généralement pu être comparé directement à un, parfois deux ou trois, des schémas existants sur la base d'une partie du matériel commune aux diverses études, les comparaisons plus globales s'effectuant de proche en proche. Des images remarquablement similaires, qui traduisent une structure bipolaire très marquée, ont été obtenues avec les différents outils utilisés : isoenzymes (SECOND, 1982 ; 1985 ; GLASZMANN *et al.*, 1984 ; GLASZMANN, 1985 ; 1987 ; 1988), RFLP (WANG et TANKSLEY, 1989 ; ZHANG *et al.*, 1992 ; SHII *et al.*, 1995 ; SECOND et GHESQUIERE, 1995), RAPD (VIRK *et al.*, 1995 ; MACKILL, 1995 ; PARSONS *et al.*, 1997), ISSR-PCR (PARSONS *et al.*, 1997 ; BLAIR *et al.*, 1999), AFLP (ZHU *et al.*, 1998) et microsatellites (YANG *et al.*, 1994), ces derniers donnant cependant une image plus floue du fait d'un polymorphisme plus important et de nombreuses formes alléliques rares.

Par ailleurs, VIRK *et al.* (1996) ont signalé la valeur des marqueurs moléculaires pour prédire certains caractères d'intérêt agronomique à expression quantitative lorsque l'on considère un échantillon de variétés très diverses. Cette observation porte en elle la question de l'organisation de la diversité génétique et de son déterminisme : quelle structuration, pour quels caractères ? quels mécanismes de maintien ou d'évolution ? quelle utilisation ?

Nous proposons ici de comparer plusieurs schémas de classification reposant sur différents types de caractères, en concentrant les analyses sur des échantillons largement partagés (voir annexe). Ensuite, nous examinons quelques situations particulières permettant d'aborder certains aspects de la dynamique qui conditionne la diversité génétique des riz cultivés.

## La diversité révélée par différents types de marqueurs

### LES ISOENZYMES

L'étude la plus vaste a porté sur un échantillon de 1688 variétés traditionnelles asiatiques caractérisées pour 15 locus polymorphes (GLASZMANN, 1987). Elle a permis de distinguer six groupes : deux majeurs (I et VI), à distribution panasia-

tique, deux mineurs (II et V), localisés essentiellement sur les contreforts himalayens d'Asie du Sud, et deux très particuliers (III et IV), restreints aux zones de submersion profonde du Bangladesh. Le groupe I représente les formes typiques *indica* et le groupe VI, les formes typiques *japonica*. Le groupe II contient notamment les riz de l'écotype « aus » du nord de l'Inde et du Bangladesh. Le groupe V est représenté par les riz de qualité particulière (arôme, élongation à la cuisson) du groupe « basmati » d'Inde et du Pakistan. Les groupes III et IV correspondent à des riz flottants, dont certains, comme le « rayada » du groupe IV, sont capables d'allonger leurs tiges sur plus de 5 mètres.

Un échantillon A de 270 variétés, constitué pour représenter la diversité géographique, écotypique et enzymatique des riz cultivés d'origine asiatique (voir annexe), a été utilisé pour une analyse similaire. Les 15 locus permettent de révéler 57 formes alléliques. Ces données ont été soumises à une analyse factorielle des correspondances (AFC). La part de variation expliquée par les axes de l'AFC décroît très rapidement, puisqu'elle est de 18,7 %, 9,2 %, 7,9 %, 6,3 % et 5,1 % pour les axes 1 à 5. Cette décroissance rapide illustre une structure très marquée. La figure 1 montre la distribution des accessions dans le plan 1-2 de l'AFC. Les six groupes, I à VI, se retrouvent, en conformité avec les résultats obtenus sur des échantillons plus vastes.

## LES RFLP

Un échantillon B de 147 variétés a été analysé grâce à 202 sondes (voir annexe). Parmi elles, 148, soit 74 %, ont révélé du polymorphisme et produit au total 482 bandes polymorphes. Ces données ont été soumises à une AFC. La part de variation expliquée par les axes de l'AFC décroît très vite, puisqu'elle est de 16,5 %, 6,2 %, 4,4 %, 3,8 % et 3,1 % pour les axes 1 à 5, dénotant, comme pour les isoenzymes, une structure générale marquée.

La figure 2a montre la distribution des 147 variétés étudiées dans le plan 1-2. La position des variétés sur le premier axe est de loin celle qui les caractérise le mieux. Il est particulièrement intéressant de remarquer que cet axe représente la variation entre les variétés *indica* de référence (groupe isoenzymatique I) et *japonica* de référence (groupe isoenzymatique VI). Les groupes II, III, IV, V et « intermédiaires » apparaissent globalement en position intermédiaire entre *indica* et *japonica* de référence, avec cependant un faible pourcentage d'allèles originaux. A cette remarque près, *O. sativa* apparaît donc bien comme organisé selon un axe principal de variation *indica-japonica*.

Dans l'analyse fine de la diversité d'*O. sativa*, l'axe 3 laisse apparaître un résultat nouveau : la différenciation de la sous-espèce *japonica* en deux groupes, l'un « tempéré », qui comprend les variétés de régions tempérées et tropicales d'altitude, l'autre « tropical », qui inclut les variétés de riz pluvial. Cette différenciation se fait sur la base de quelques locus seulement et recoupe la différenciation morphologique de OKA (1958).

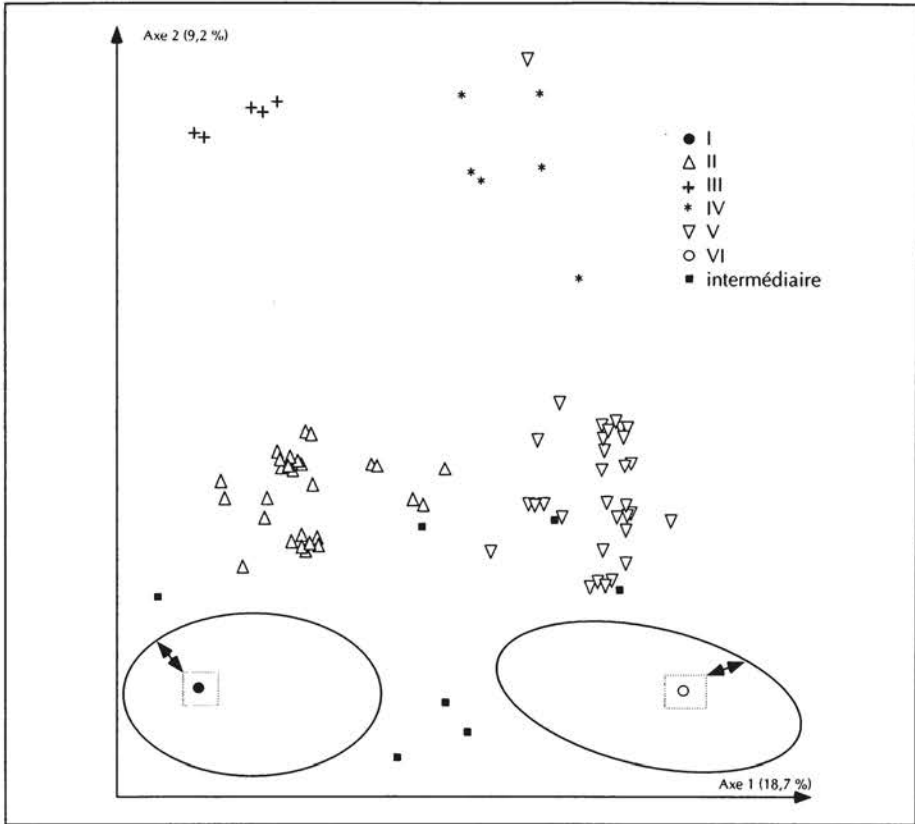


Figure 1. Distribution dans le plan 1-2 de l'AFC sur les données isoenzymatiques (15 locus et 57 allèles) caractérisant l'échantillon A.

Enfin, deux sondes RFLP supplémentaires caractérisent la diversité du cytoplasme en fonction de données antérieures qui avaient permis d'identifier deux types principaux, tant au niveau chloroplastique (DALLY ET SECOND, 1990) qu'au niveau mitochondrial (SECOND ET WANG, 1992). La figure 2b montre la distribution des deux types cytoplasmiques parmi 141 des 147 variétés. Il est clair que l'un des types caractérise la sous-espèce *indica* et certains intermédiaires, alors que l'autre type se rencontre dans tous les groupes variétaux.

#### LES MICROSATELLITES

La diversité des microsattellites a été étudiée pour 12 locus (*RM7*, *RM12*, *RM13*, *RM19*, *RM122*, *RM148*, *RM164*, *RM167*, *RM168*, *OSR4*, *OSR7*, *OSR35*) au sein d'un sous-ensemble de 54 variétés représentatives des éléments de structure connus (voir annexe). Les locus ont révélé entre 4 et 19 allèles par locus, avec une moyenne de 10,8. Une AFC a été effectuée sur les données obtenues, soit 130 allèles. Les premiers axes expliquent 6,0 %, 5,8 %, 4,8 %, 3,3 %, 3,2 %, 3,1 %, 3,0 %, 2,9 %, 2,8 %, 2,7 %, 2,6 %, 2,5 %, 2,4 %, 2,3 %, 2,2 %, 2,1 %, 2,0 %, 1,9 %, 1,8 %, 1,7 %, 1,6 %, 1,5 %, 1,4 %, 1,3 %, 1,2 %, 1,1 %, 1,0 %, 0,9 %, 0,8 %, 0,7 %, 0,6 %, 0,5 %, 0,4 %, 0,3 %, 0,2 %, 0,1 %.

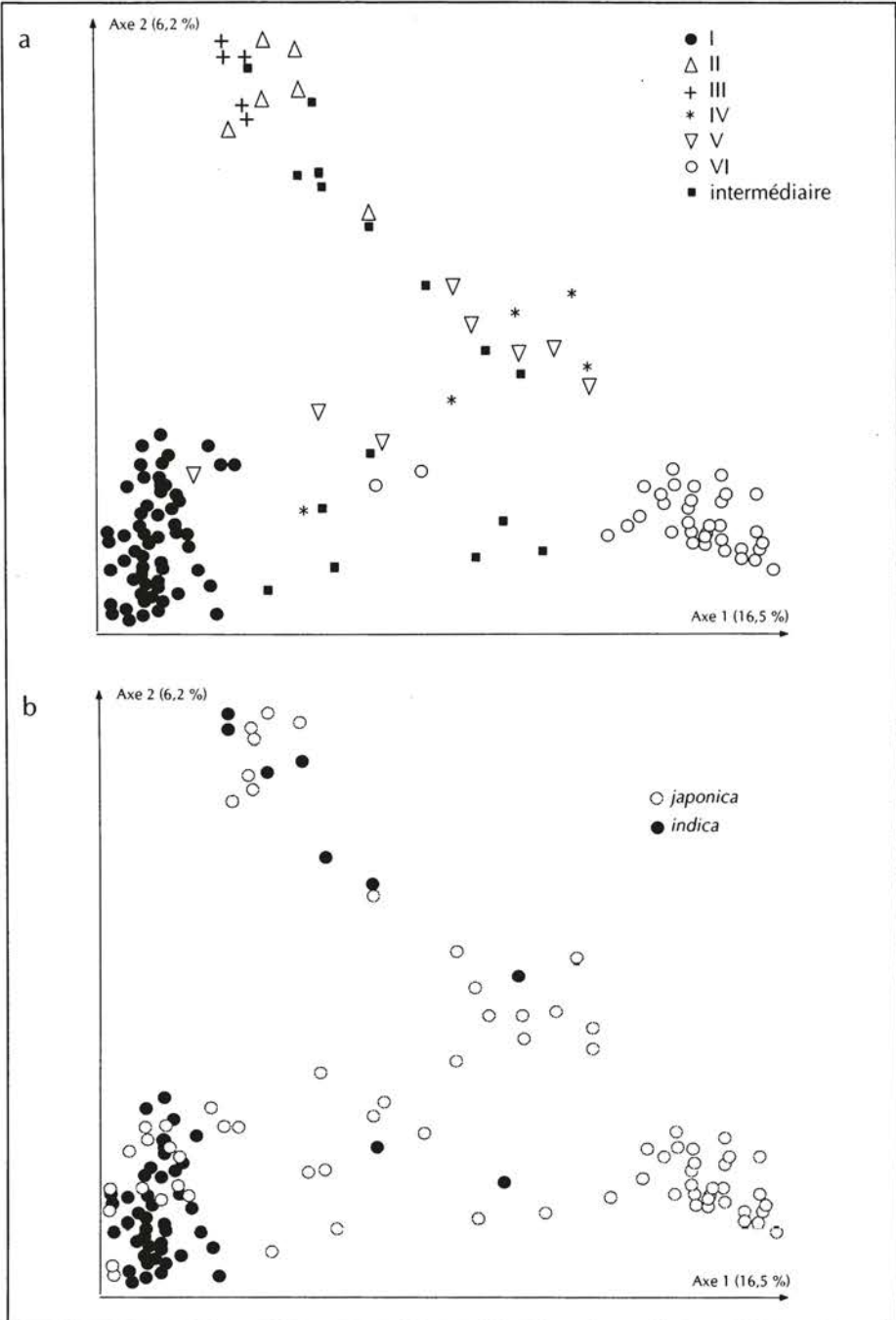


Figure 2. Distribution dans le plan 1-2 de l'AFC sur les données RFLP (202 sondes nucléaires et 482 bandes) caractérisant l'échantillon B, avec a) indication de la classification enzymatique et b) indication du type cytoplasmique, d'après DALLY et SECOND (1990).

4,7 % et 3,9 % de la variation ; cette décroissance lente révèle une structure relativement peu marquée. Le plan 1-2 de l'AFC (figure 3) met en évidence trois ensembles : le premier correspond au groupe I de la classification enzymatique, le deuxième aux groupes II et III, le troisième aux groupes IV, V et VI. Les axes 3 et 4 séparent les groupes IV, V et VI. En revanche, les groupes II et III restent confondus. On trouve une grande diversité à l'intérieur des groupes.

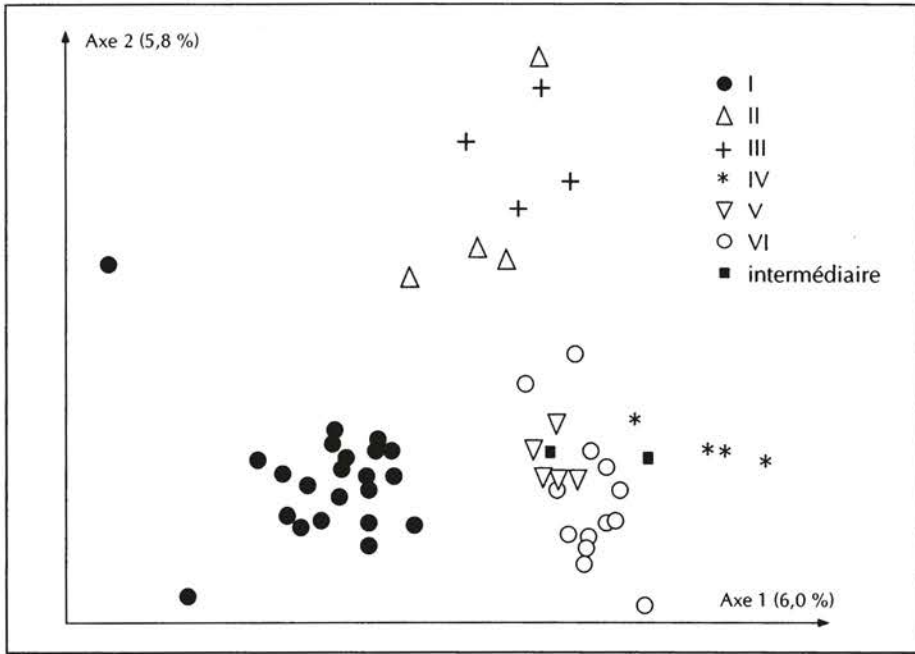


Figure 3. Distribution dans le plan 1-2 de l'AFC sur les données microsatellites (12 locus et 130 allèles) caractérisant un échantillon de 54 variétés communes à A et B.

#### LA COMPARAISON DES DIFFÉRENTS MARQUEURS

Les études les plus vastes (isoenzymes) et les plus approfondies (RFLP) donnent la même image de l'espèce : une structure fortement bipolaire avec néanmoins quelques formes particulières, partiellement intermédiaires. Avec les microsatellites, on conserve la discrimination entre les groupes les plus différenciés, mais on révèle un polymorphisme élevé au sein des groupes. La figure 4 montre les relations entre les distances génétiques intergroupes évaluées à partir des différents types de marqueurs. Les disparités entre les nombres d'accessions et de locus pris en compte pour chacun de ces types limitent les conclusions que l'on peut tirer de cette comparaison. On note cependant que, par rapport aux RFLP et aux isoenzymes, les microsatellites amplifient considérablement les distances génétiques entre groupes et que, parmi les comparai-

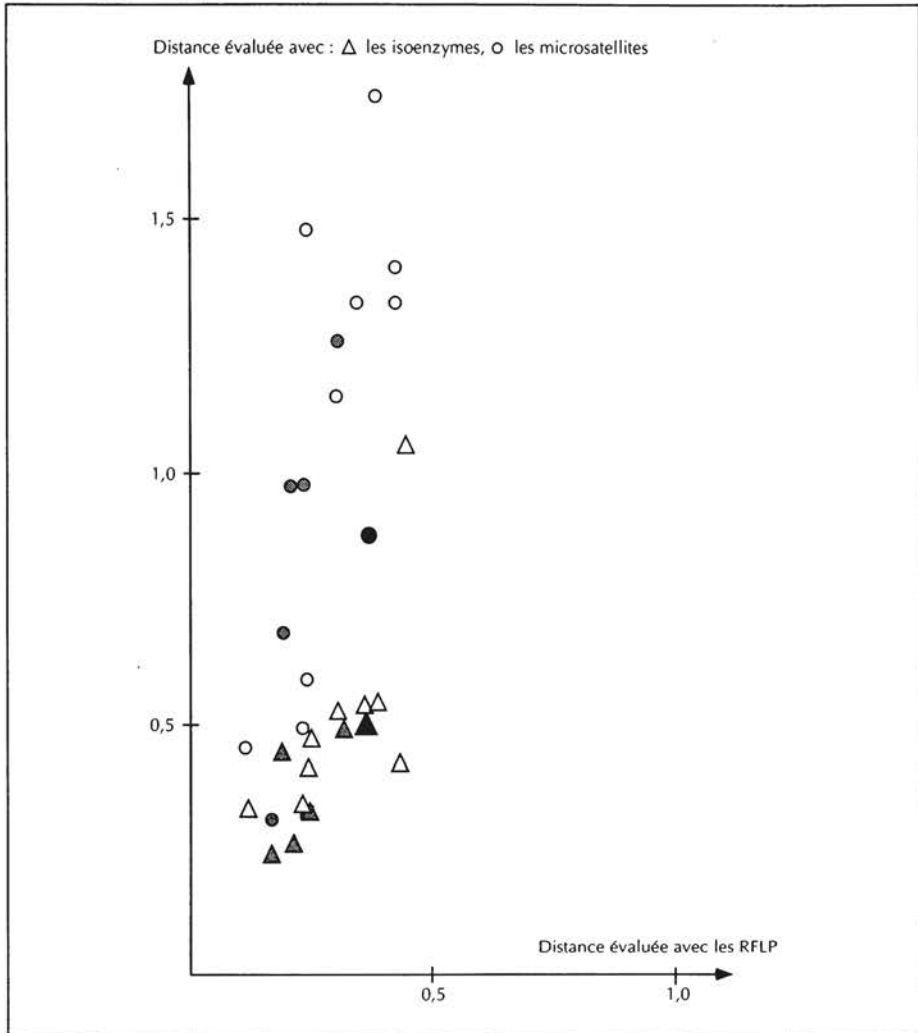


Figure 4. Comparaison entre les distances évaluées avec les RFLP, les isoenzymes et les microsatellites.

Les symboles en noir caractérisent les distances entre les deux groupes principaux I et VI ; les symboles en gris celles entre les groupes de grande taille, I, II, V et VI ; les symboles en blanc celles qui impliquent un groupe de petite taille, III ou IV.

sons entre les groupes de plus grande taille (groupes I, II, V, VI), la distance entre les groupes I et VI est la plus grande avec les RFLP comme avec les isoenzymes, alors qu'elle est loin de l'être avec les microsatellites.

Ainsi, les microsatellites révèlent un polymorphisme plus important, qui affine la résolution dans les groupes généralement les moins polymorphes. Ce polymorphisme préserve l'empreinte générale de la structure de l'espèce mais jette un flou sur les relations entre les groupes.



## La classification moléculaire et la diversité pour des caractères d'intérêt agronomique

### LA DIVERSITÉ MORPHOLOGIQUE

Les données de caractérisation de l'échantillon A disponibles à l'IRRI ont été récupérées. Un total de 248 variétés présentait des données complètes pour 11 variables ; parmi ces variétés tous les groupes enzymatiques étaient représentés à l'exception du groupe IV, qui fleurit difficilement dans les conditions de l'IRRI. Ces données ont été soumises à une analyse en composantes principales (ACP). La part de variation expliquée par les axes est de 31,3 %, 19,1 %, 12,0 %, 7,9 % et 7,0 % pour les axes 1 à 5. La figure 5 montre la distribution des accessions dans le plan 1-2. L'axe 1 sépare les variétés à longs organes (ligules, feuilles, tiges et, dans une moindre mesure, panicules) et à grain plutôt étroit, qui se trouvent du côté des valeurs négatives, des variétés complémentaires, qui se situent du côté des valeurs positives. L'axe 2 attire vers les valeurs négatives les variétés à faible tallage, à feuille large, à tige plutôt épaisse et à grain large et lourd. On reconnaît ainsi la structure tripolaire avec les types Indica dans la partie supérieure gauche du plan, les Japonica dans la partie extrême droite du plan et les Javanica dans partie extrême inférieure du plan.

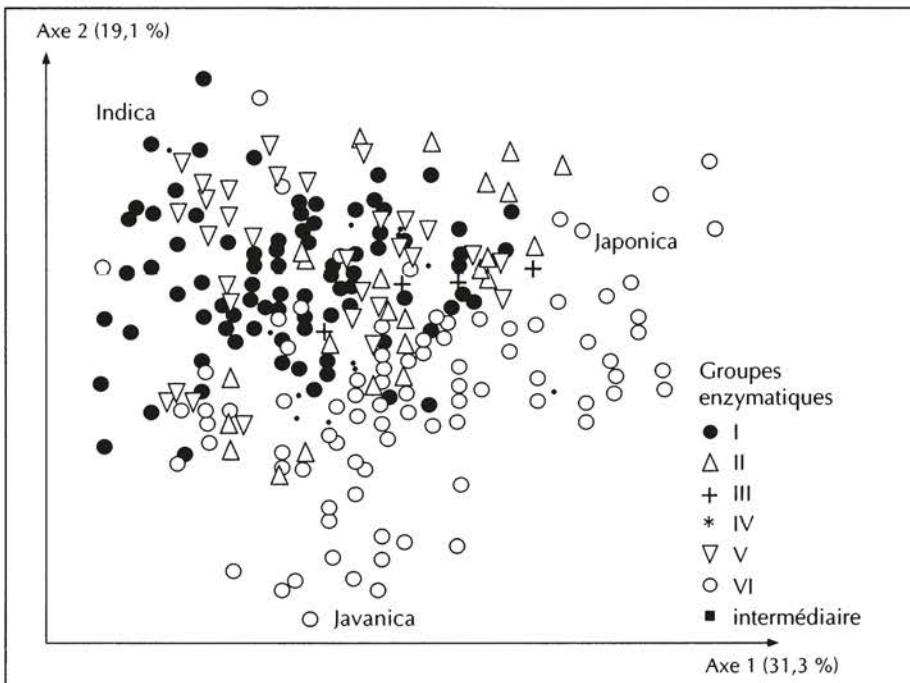


Figure 5. Distribution dans le plan 1-2 de l'ACP de la variation parmi 248 variétés pour 11 caractères morphologiques.



Le type morphologique *Indica* recouvre des groupes très diversifiés sur le plan des marqueurs moléculaires, avec surtout des représentants des groupes I et V, mais aussi des groupes II et VI. Les formes morphologiques typiques des *Japonica* et des *Javanica* correspondent au seul groupe VI. Alors que les groupes I et VI occupent des zones du plan quasiment séparées, les accessions des groupes II à V sont plus fréquentes à la frontière entre ces deux zones principales.

Au sein du groupe VI, très dispersé sur le plan, on note une différenciation entre les origines géographiques, qui oppose les formes tempérées (pôle *Japonica*) aux formes tropicales (pôle *Javanica*). Entre ces deux pôles, on trouve des variétés d'origine tempérée, subtropicale ou tropicale, mais surtout des variétés tropicales d'altitude issues des régions montagneuses d'Asie du Sud-Est et d'Asie du Sud. Cette répartition illustre les observations réalisées par GLASZMANN et ARRAUDEAU (1986) : on a en fait, avec les types morphologiques *Japonica* et *Javanica*, deux formes qui se différencient sur une base génétique étroite et représentent les extrêmes d'une variation continue qui suit un cline géographique selon la latitude et l'altitude. La distinction classique entre ces deux types est probablement due au fait que les premières études de diversité morphologique ont été effectuées sur des échantillons dans lesquels les types originaires du Japon et de Java, zones plus accessibles aux collecteurs que les montagnes d'Asie du Sud-Est ou du Sud, étaient surreprésentés. Les données RFLP montrent un cline similaire avec la dispersion des variétés sur l'axe 3 de l'AFC.

#### LA DIVERSITÉ DE RÉACTION À DIVERS PATHOGÈNES

Les 270 variétés de l'échantillon A ont été testées pour leur réaction à l'égard des principales maladies fongiques, bactériennes et virales (BONMAN *et al.*, 1990 ; GLASZMANN *et al.*, 1995). Cette collection a permis de révéler une très grande diversité de réponses. Deux exemples en sont donnés ici.

La *pyriculariose*, due au champignon *Magnaporthe grisea*, est la principale maladie fongique du riz. Il existe une grande variabilité chez le champignon et de nombreux gènes de résistance sont connus chez la plante. Le repérage de ces gènes de résistance représente un enjeu majeur. Les accessions de l'échantillon A ont été inoculées avec 13 souches du champignon dans le cadre d'un réseau de collaborations. Les symptômes ont été notés d'après la taille des lésions relevée sur plusieurs plantes par accession, selon une échelle de 0 à 5. La moyenne des notes d'une accession sert à caractériser sa tolérance à l'égard de la souche utilisée. Une ACP a été réalisée sur les données complètes obtenues pour 257 accessions (figure 6a). L'axe 1 sépare les variétés généralement résistantes (valeurs négatives) des variétés généralement sensibles. En Asie du Sud, où tous les groupes enzymatiques sont présents, les groupes II à V se différencient des groupes majeurs I et VI par une résistance nettement plus marquée. Parmi les variétés les plus souvent sensibles, l'axe 2 tend à différencier les variétés *indica* de référence (groupe I) des variétés *japonica* de référence (groupe VI), du fait de la sensibilité de chaque groupe à plusieurs souches dif-

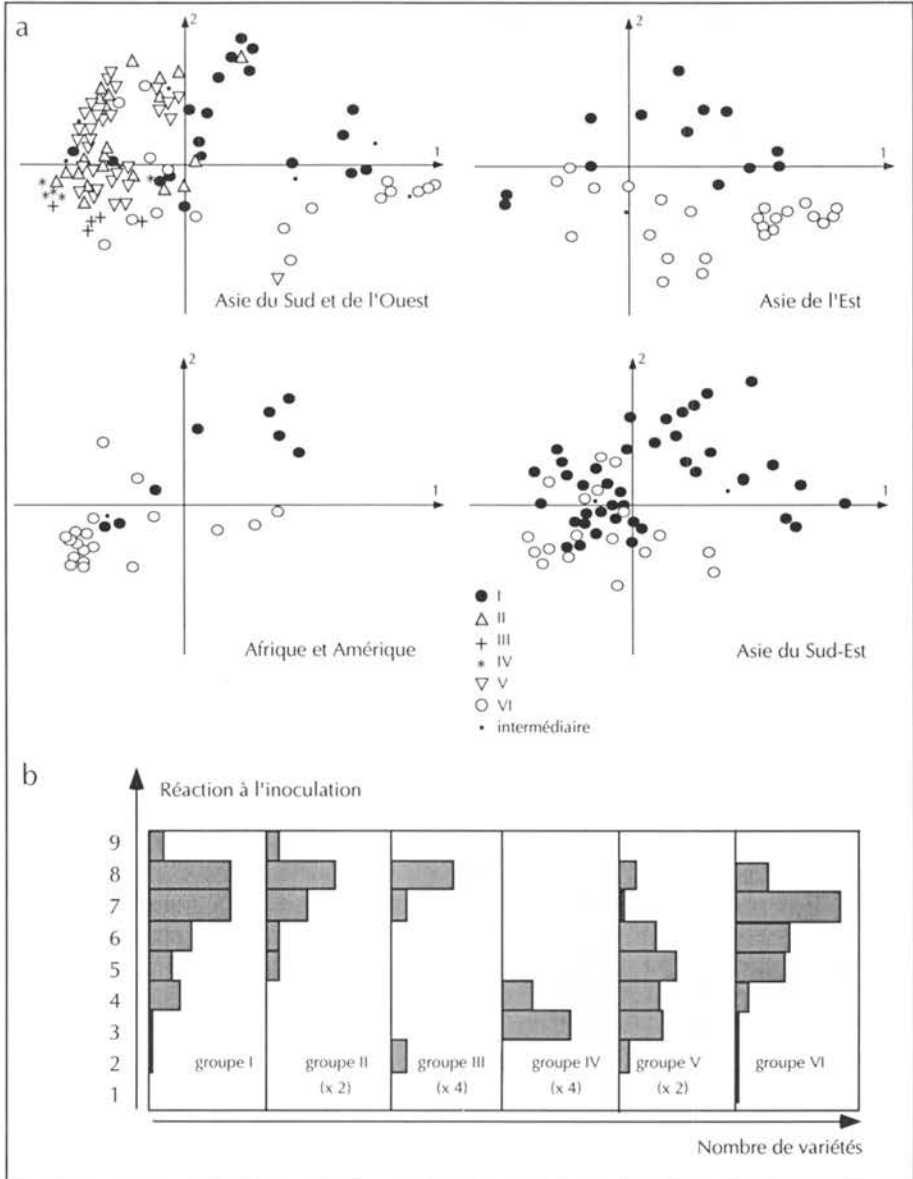


Figure 6. Réaction des variétés représentant différents groupes enzymatiques à l'égard de plusieurs parasites.

a. Résistance à la pyriculariose des variétés représentant différents groupes enzymatiques : distribution de 257 variétés dans le plan 1-2 de l'ACP sur la taille des lésions produites par l'inoculation de 13 souches de *Magnaporthe grisea*.

b. Tolérance relative des 6 groupes enzymatiques au virus du tungro : distribution des réactions à l'inoculation par vecteur (notation de 1 pour l'absence complète de symptômes à 9 pour la sensibilité maximale) dans un ensemble de 261 variétés représentant les différents groupes enzymatiques.

férentielles. Au sein du groupe VI, on retrouve une différenciation entre les formes tempérées (Asie du Sud et de l'Est) et les formes tropicales (Asie du Sud-Est, Afrique, Amérique). Il est difficile d'attribuer ces différences à une différenciation générale qui marque l'isolement ancien entre ces formes ou à une coévolution entre l'hôte et le pathogène. Le fait le plus marquant à retenir concerne la diversité des réponses obtenues pour des variétés d'origine géographique voisine, donc *a priori* soumises à des populations du pathogène peu ou non différenciées, en fonction de leur appartenance aux différents groupes moléculaires. Ainsi, les groupes de la classification moléculaire présentent des différences de comportement très marquées, avec des fréquences de résistance contrastées et des spécialisations fortes des souches de référence.

Le tungro est une maladie virale de première importance. Pour évaluer la résistance portée par chaque génotype, on a procédé à l'inoculation par vecteur puis noté les réactions sur une échelle de 1, pour l'absence complète de symptômes, à 9, pour la sensibilité maximale. La figure 6b montre la distribution des réactions dans un ensemble de 261 variétés selon leur appartenance aux différents groupes moléculaires. La tolérance est nettement plus fréquente dans les groupes IV et V.

Une étude pionnière avait montré que les souches du parasite fongique de la pyriculariose se classent en deux ensembles en fonction de leur réaction différentielle face aux variétés *indica* et *japonica* (MORISHIMA, 1969). Une étude plus récente sur le flétrissement bactérien a montré que certains gènes de résistance totale se répartissent largement suivant l'appartenance des variétés aux groupes de la classification enzymatique (BUSTO *et al.*, 1990). Nos résultats confortent ces observations et permettent de conclure que la structuration des riz en groupes variétaux est essentielle pour comprendre les interactions hôte-pathogène et pour rechercher de nouvelles sources de résistance.

## L'association entre caractères et la dynamique du génome

### LA DIVERSITÉ À MADAGASCAR, SITE D'ÉVOLUTION INSULAIRE

Madagascar représente un cas très particulier : l'introduction du riz asiatique y est très ancienne et l'insularité en fait un laboratoire d'évolution des riz cultivés, en limitant les influences extérieures à quelques grandes introductions en provenance vraisemblablement d'Inde, de Malaisie et d'Indonésie. L'observation par les sélectionneurs de types variétaux particuliers à Madagascar a justifié l'analyse détaillée de la diversité présente dans l'île (RABARY *et al.*, 1989 ; AHMADI *et al.*, 1991).

Une collection de variétés malgaches a été étudiée sur les plans morphologique et isoenzymatique. Des types morphologiques tout à fait particuliers, intermédiaires entre les types *Indica* et *Javanica*, ont été trouvés avec une fréquence élevée (figure 7). La diversité enzymatique révélée est forte, mais très

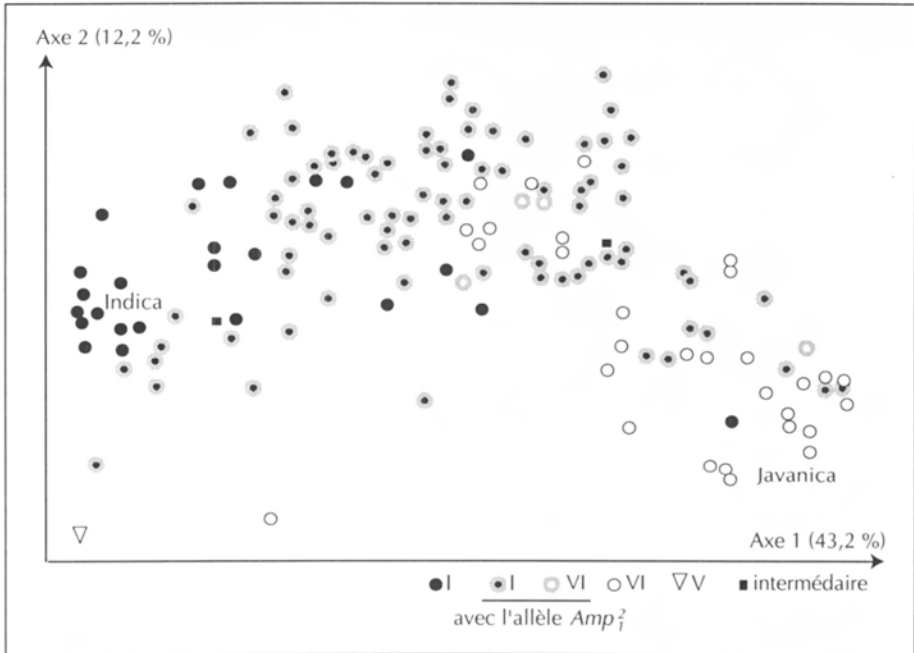


Figure 7. Distribution de 144 variétés dans le plan 1-2 de l'AFC sur 39 caractères morphologiques au sein du matériel de Madagascar. Les mêmes symboles que dans les figures précédentes indiquent l'appartenance aux groupes enzymatiques ; une distinction supplémentaire est faite pour les variétés possédant l'allèle  $Amp_1^2$ .

inférieure à celle des riz cultivés en Asie — l'indice de diversité H de Nei est de 0,29 à Madagascar contre 0,62 en Asie. La structure génotypique est nettement bipolaire et rappelle l'opposition *indica-japonica*. Cependant, certains types particuliers présentent un allèle codant pour une aminopeptidase,  $Amp_1^2$ , qui est très rarement observé dans les groupes enzymatiques I et VI en Asie. Parmi ces types, certaines associations classiques entre isoenzymes et caractères morphologiques sont faibles, voire inexistantes.

Ces particularités font supposer que de nombreuses recombinaisons entre groupes variétaux ont eu lieu depuis les premières introductions de riz à Madagascar. Malgré ces recombinaisons, les associations entre les isoenzymes marqueurs de la différenciation *indica-japonica* se sont maintenues, alors qu'elles impliquent des gènes situés sur de nombreux chromosomes. Cette situation démontre qu'il est possible de recombiner des gènes participant au contrôle de l'architecture de la plante, mais témoigne aussi des pressions de sélection, naturelles et humaines, qui tendent à maintenir le déséquilibre gamétique impliqué dans la différenciation *indica-japonica*. De nombreux marqueurs enzymatiques semblent situés dans des zones du génome soumises à ces pressions, ce qui confirme l'intérêt de la classification fondée sur l'analyse enzymatique.

## LE GÉNOME CYTOPLASMIQUE ET L'ORIENTATION DES FLUX GÉNIQUES

La situation particulière de Madagascar a permis de documenter un phénomène probablement généralisé en Asie, celui de l'introgression entre *indica* et *japonica*, source de diversité des riz cultivés (SECOND, 1982).

SECOND et GHESQUIERE (1995) ont confronté les caractéristiques nucléaires et cytoplasmiques au sein de l'échantillon B. Deux types cytoplasmiques se distinguent, en accord avec l'interprétation globale de l'analyse du genre *Oryza* pour les marqueurs moléculaires : un type « *indica* », I, et un type « *japonica* », J (DALLY et SECOND, 1990 ; SECOND et WANG, 1992). Le type cytoplasmique *japonica* se rencontre seul dans les variétés du groupe VI, mais en alternance avec le type *indica* (distribution J:I voisine de 3:1) dans les variétés des groupes intermédiaires II à V. Il est également présent avec une fréquence significative parmi les variétés du groupe I (distribution J:I voisine de 1:3). Le cytoplasme *japonica* aurait donc été introgressé par nombre de variétés du groupe I, ce qui indique, étant donné l'héritage maternelle du cytoplasme, que les hybrides à l'origine des introgressions de gènes nucléaires entre sous-espèces *indica* et *japonica* ont tous été formés par l'hybridation d'un parent femelle *japonica* pollinisé par un *indica*.

## LA DISTRIBUTION DE LA DIVERSITÉ MOLÉCULAIRE LE LONG DU GÉNOME

La distribution du polymorphisme le long du génome présente un biais par rapport à ce que pourrait être une distribution aléatoire (SECOND et GHESQUIERE, 1995 ; ZHUANG *et al.*, 1998). On constate une tendance significative au regroupement des locus monomorphes dans l'espèce. De même, on observe pour les variétés une tendance au regroupement des marqueurs homogènes *indica* ou *japonica* (SECOND *et al.*, 1995). Cela indique bien que les chromosomes du riz peuvent s'interpréter comme une mosaïque de régions issues soit du type ancestral *indica*, soit du type ancestral *japonica* et de régions monomorphes, que la dérive génétique a pu fixer au sein de l'espèce dans son état actuel à partir de l'un ou de l'autre des types ancestraux. Le cas du chromosome 12 met bien en évidence ce phénomène (ALBAR, 1998) : il existe des associations fortes, vérifiées à partir de plusieurs couples enzymes-sondes, entre des allèles de même origine *indica* ou *japonica* (figure 8). Ces déséquilibres de liaison s'observent entre des locus très proches (notés b), situés dans la zone centromérique du chromosome 12, ou entre ces locus et des locus nettement plus éloignés (notés a). Dans cette trame, peuvent s'intercaler des locus monomorphes ou quasi monomorphes et des locus pour lesquels le polymorphisme est indépendant de la structure *indica-japonica* (notés c, d et e). Ces derniers représentent vraisemblablement des zones soumises aux pressions de sélection de manière différente, comme cela a été observé autour de certains locus isoenzymatiques dans les variétés de Madagascar.

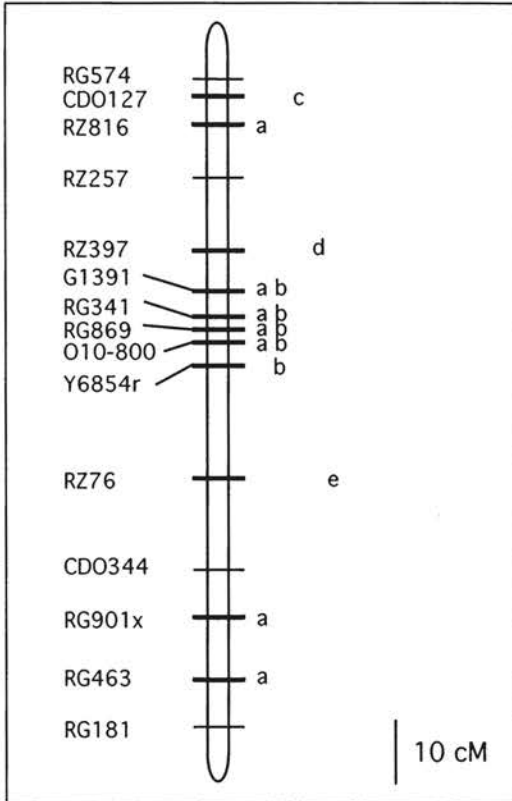


Figure 8. Déséquilibre de liaison le long du chromosome 12. Sur la base de l'analyse du polymorphisme au sein de l'échantillon B, 11 locus sur les 15 testés ont été trouvés polymorphes et sont indiqués par une barre horizontale plus épaisse. Les déséquilibres de liaison entre locus ont été testés sur la base des deux allèles les plus fréquents. Les locus en déséquilibre significatif ( $p < 0,01$ ) entre eux sont indiqués par la présence d'une même lettre à côté de la barre horizontale.

## Conclusion

Le cas des riz cultivés asiatiques présente la particularité d'inclure une diversité très large, puisque les groupes indica et japonica représentent presque deux espèces différentes si l'on accepte l'hypothèse d'une origine double à partir de populations sauvages déjà différenciées (SECOND, 1982 ; 1984). On a donc une structure très marquée, que l'on retrouve avec les divers types de marqueurs utilisés. On voit bien les apports complémentaires de marqueurs comme les isoenzymes et les RFLP, d'un côté, les microsatellites, de l'autre. Les isoenzymes et les RFLP donnent des images semblables et des indices de différenciation comparables ; les RFLP, du fait qu'ils sont plus nombreux, permettent de mettre en évidence un élément supplémentaire : la différenciation écogéographique entre les formes Japonica tempérées et tropicales. Les microsatellites mémorisent moins bien une structuration ancienne (*indica-japonica*), mais révèlent une variabilité plus étendue à l'intérieur de chaque composante.

Les marqueurs moléculaires apportent des éléments importants, inaccessibles avec la seule diversité morphologique : plusieurs groupes (I à V) sont identifiés à l'intérieur d'un ensemble considéré comme homogène sur le plan morpholo-

gique. En revanche, la continuité entre deux types morphologiques différents, Japonica et Javanica, apparaît clairement, ce qui vient corroborer les conclusions tirées de l'examen du comportement des descendances hybrides : elles sont généralement fertiles et se fixent rapidement au fil des autofécondations successives.

Par rapport aux travaux de classification antérieurs, les marqueurs moléculaires attirent l'attention sur des formes particulières, représentées par les groupes II à V de la classification enzymatique. Ces résultats recoupent différentes observations réalisées par les sélectionneurs, comme le comportement particulier des types « aus » (groupe II) en croisement avec des variétés *indica* et *japonica* typiques (WAN et IKEHASHI, 1997) ou celui des variétés « basmati » (groupe V), de morphologie Indica mais très difficiles à exploiter en combinaison avec les variétés à haut rendement du groupe I (G.S. Khush, comm. pers.). Nos résultats sur le comportement des variétés face aux principales maladies illustrent également la particularité de ces groupes, qui viennent élargir les sources de résistance présentes dans les groupes majeurs. L'identification de ces groupes a été rendue possible par une approche d'échantillonnage structurée. Un premier lot de 1 688 variétés asiatiques a été constitué en diversifiant au maximum l'origine géographique et l'environnement de culture et en tenant compte des classifications locales (GLASZMANN, 1987). Cet échantillon a été élargi ensuite à d'autres continents. La structuration en groupes enzymatiques a été utilisée pour surreprésenter les sous-ensembles riches en diversité particulière. Cette démarche a permis de constituer une collection « noyau » de taille acceptable pour des évaluations fines.

Les résultats de ces évaluations confortent la démarche générale. On constate, par exemple, que les groupes mineurs recèlent des sources de résistance originales à la pyriculariose ou au tungro et que les variétés japonica d'Afrique sont généralement plus résistantes à la pyriculariose que celles d'Asie. Ces associations internes à l'échantillon mettent en évidence des paramètres simples à appréhender — origine géographique, type de culture, classification établie à l'aide de quelques marqueurs choisis... — qu'il convient de diversifier pour enrichir la collection. *A contrario*, on peut se servir des résultats pour rechercher rapidement des variétés qui font exception aux associations ou aux corrélations générales afin de progresser dans une direction de sélection précise. On peut par exemple, si l'on veut produire des variétés à faible tallage et à longue panicule fournie (caractéristiques générales du groupe VI), mais aussi à grain fin (caractéristique des autres groupes), sélectionner parmi les variétés du groupe VI, qui ont généralement un grain large, celles qui ont le grain le plus fin. Cette démarche sera plus efficace que celle qui vise à recombinaison intergroupes, dont tout indique qu'elles perturbent des agencements de facteurs coadaptés ; on tirera profit de l'accumulation de facteurs de finesse du grain dans certaines variétés possédant le fond génétique recherché, que cette accumulation soit le résultat d'une juxtaposition de facteurs dispersés

dans les populations du groupe VI ou de l'introggression réussie de facteurs spécifiques des autres groupes.

Nos observations mettent en évidence une multitude d'associations entre caractères moléculaire et caractères quantitatifs, du fait de l'existence de groupes variétaux différenciés pour les deux types de caractères, comme l'ont observé VIRK *et al.* (1996). De telles associations ont schématiquement deux déterminants principaux.

Le premier est un effet de fondation au moment de la domestication, venant éventuellement renforcer l'action conjuguée de la sélection naturelle et de la dérive génétique qui ont différencié les populations fondatrices. L'origine probablement séparée des types *indica* et *japonica* à partir de formes ayant divergé depuis 2 à 3 millions d'années (SECOND, 1985) donne une importance particulière à ce phénomène chez le riz asiatique et explique l'omniprésence d'associations. L'origine des groupes mineurs reste à établir. Il est néanmoins clair à la lumière de nos résultats que, même en l'absence d'une interprétation globale satisfaisante, les données moléculaires peuvent constituer une source d'information précieuse pour diversifier les sources de gènes utiles. Ce raisonnement a une efficacité déjà éprouvée dans la pratique (KAMESWARO-RAO et JACKSON, 1997).

Le deuxième déterminant est l'introggression entre types ancestraux. De nombreuses études, dont la nôtre sur les riz de Madagascar, montrent que les échanges géniques entre variétés de types différents sont fréquents. Néanmoins, s'il y avait recombinaison généralisée, les associations disparaîtraient rapidement au cours du temps. La possibilité d'évaluer la diversité moléculaire avec des marqueurs cartographiés sur le génome présente dans ce contexte beaucoup d'intérêt. Les résultats disponibles suggèrent que le polymorphisme moléculaire est distribué sur le génome par plages non aléatoires (SECOND et GHESQUIERE, 1995 ; ZHUANG *et al.*, 1998). Il pourrait avoir un potentiel discriminant différencié, en particulier pour le groupe variétal II, selon la position de ces marqueurs par rapport aux centromères (PARSONS *et al.*, 1997). Ces phénomènes méritent d'être étudiés plus en détail, comme cela a été le cas pour le chromosome 12 (ALBAR, 1998). Dans la recherche de QTL pour la résistance partielle au virus de la panachure jaune du riz (RYMV), la mise en évidence d'interactions entre marqueurs — ou entre marqueurs et QTL — dispersés sur le génome mais appartenant chaque fois à des zones à fort déséquilibre de liaison *indica-japonica* peut illustrer l'importance des phénomènes de coadaptation chez une plante comme le riz cultivé d'origine asiatique (PRESSOIR *et al.*, 1998). Dans l'état actuel des connaissances, nous retiendrons que les associations apparentes sont le résultat de deux phénomènes : d'une part, des déséquilibres de liaison qui impliquent des gènes et des marqueurs étroitement liés sur le génome, d'autre part, des associations entre gènes coadaptés qui peuvent être dispersés sur le génome.



La mise en place d'une base de données de caractérisation moléculaire et d'évaluation détaillée pour des caractères agronomiques sur des lignées pures, issues de croisements contrôlés, mais aussi non apparentées, devrait permettre d'accumuler les résultats et de mieux connaître, comprendre et utiliser la dynamique d'évolution de la diversité génétique des riz.

## Annexe

### Matériel végétal

Les résultats présentés correspondent à une mise en perspectives de travaux parfois anciens et déjà partiellement décrits et de données nouvelles.

Sur la base d'informations rassemblées avant 1985, un premier échantillon de 270 variétés, l'échantillon A, a été constitué pour représenter la diversité géographique, écotypique et enzymatique des riz cultivés d'origine asiatique. Les échantillons ont été purifiés — passage par une plante et autofécondations contrôlées —, afin de servir de collection de référence pour étudier la diversité pour des caractères utiles (BONMAN *et al.*, 1990 ; GLASZMANN *et al.*, 1995). Ces variétés sont ici utilisées pour confronter la classification obtenue à partir des isoenzymes avec la diversité morphologique et la diversité de réponse à plusieurs parasites. Les données morphologiques sont extraites de la base entretenue par l'IRRI. Les données sur les réactions à divers pathogènes sont celles qui ont été rapportées par GLASZMANN *et al.* (1995).

Un deuxième échantillon de 147 accessions, l'échantillon B, a servi à analyser la diversité au moyen des RFLP pour 202 sondes nucléaires cartographiées couvrant l'ensemble du génome du riz. Parmi ces accessions, 141 ont été analysées à l'aide de deux sondes de l'ADN cytoplasmique : une sonde située au niveau de la mutation 28 de l'ADN chloroplastique (DALLY et SECOND, 1990) et la sonde mitochondriale Col (SECOND et WANG, 1992).

Le sous-ensemble commun à A et B, qui comprend 54 variétés, a été utilisé pour étudier la diversité révélée par les microsatellites ; 12 locus ont été pris en compte.

Un autre échantillon de 144 variétés de Madagascar a fait l'objet d'une étude particulière qui portait sur la comparaison des classifications moléculaire et morphologique dans un milieu insulaire isolé de l'aire d'origine (AHMADI *et al.*, 1991). Vingt-quatre caractères quantitatifs codés selon une distribution en grandes classes ainsi que 15 caractères qualitatifs ont été utilisés.

### Analyse des données

L'analyse des données repose essentiellement sur l'utilisation de l'analyse en composantes principales (ACP) pour des données de type quantitatif et de l'analyse factorielle des correspondances (AFC) pour des données de type qualitatif.

Une comparaison de différents marqueurs a été réalisée en quantifiant la diversité génétique entre groupes variétaux à l'aide de l'indice de distance de Nei (1978).

## Références bibliographiques

- AHMADI N., GLASZMANN J.C., RABARY E., 1991. Traditional highland rices originating from intersubspecific recombination in Madagascar. *In* : Rice genetics II. Manille, Philippines, IRRI, p. 67-79.
- ALBAR L., 1998. Etude des bases génétiques de la résistance partielle du riz au virus de la panachure jaune (RYMV) : cartographie moléculaire et relation avec la diversité génétique. Thèse de doctorat en sciences, université Paris XI, Paris, France, 187 p.
- BLAIR N.W., PANAUD O., MCCOUCH S.R., 1999. Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, 98 : 780-792.
- BONMAN J.M., MACKILL A.O., GLASZMANN J.C., 1990. Resistance to *Gerlachia oryzae* in rice. *Plant Disease*, 74 : 306-309.
- BUSTO G.A., OGAWA T., ENDO N., TABIEN R.E., IKEDA R., 1990. Distribution of genes for resistance to bacterial blight of rice in Asian countries. *Rice Genetics Newsletter*, 7 : 127-128.
- CHANG T.T., 1976. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. *Euphytica*, 25 : 435-441.
- CHANG T.T., ADAIR C.R., JOHNSTON T.H., 1984. The conservation and use of rice genetic resources. *Advances in Agronomy*, 35 : 37-90.
- CHENG K.S., WANG X.K., ZHOU J.W., LU Y.X., LOU J., HUANG H.W., 1984. Studies on indigenous rices in Yunnan and their utilization. 2. A revised classification of Asian cultivated rice. *Acta Agronomica Sinica*, 10 : 271-280 (en chinois).
- DALLY A.M., SECOND G., 1990. Chloroplast DNA diversity in wild and cultivated species of rice (genus *Oryza*, section *Oryza*): cladistic-mutation and genetic-distance analysis. *Theoretical and Applied Genetics*, 80 : 209-222.
- GLASZMANN J.C., 1985. A varietal classification of Asian cultivated rice (*Oryza sativa* L.) based on isozyme polymorphism. *In* : Rice genetics I. Manille, Philippines, IRRI, p. 83-90.
- GLASZMANN J.C., 1987. Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics*, 74 : 21-30.
- GLASZMANN J.C., 1988. Geographic pattern of variation among Asian native cultivars (*Oryza sativa* L.) based on fifteen isozyme loci. *Genome*, 30 : 782-792.
- GLASZMANN J.C., ARRAUDEAU M., 1986. Rice plant type variation: Japonica-Javanica relationships. *Rice Genetics Newsletter*, 3 : 41-43.
- GLASZMANN J.C., BENOIT H., ARNAUD M., 1984. Classification des riz cultivés (*Oryza sativa* L.) : utilisation de la variabilité isoenzymatique. *L'Agronomie tropicale*, 39 : 51-66.
- GLASZMANN J.C., MEW T., HIBINO H., KIM C.K., VERGEL DE DIOS-MEW T.I., VERA CRUZ C.H., NOTTEGHEM J.L., BONMAN J.M., 1995. Molecular variation as a diverse source of disease resistance in cultivated rice. *In* : Rice genetics III. Manille, Philippines, IRRI, p. 460-465.

- ISHII T., BRAR D.S., SECOND G., TSUNEWAKI K., KHUSH G.S., 1995. Nuclear differentiation in Asian cultivated rice as revealed by RFLP analysis. *Japanese Journal of Genetics*, 70 : 643-652.
- JACKSON M.T., 1997. Conservation of rice genetic resources: the role of the international rice genebank at IRRI. *Plant Molecular Biology*, 35 : 61-67.
- KAMESWARO-RAO N., JACKSON M.T., 1997. Variation of seed longevity of rice cultivars belonging to different isozyme groups. *Genetic Research and Crop Evolution*, 44 : 159-164.
- LOLO O.M., SECOND G., 1988. Peculiar genetic characteristics of *Oryza rufipogon*. *Rice Genetics Newsletter*, 5 : 67-70.
- MACKILL D.J., 1995. Classifying japonica rice cultivars with RAPD markers. *Crop Science*, 35 : 889-894.
- MORISHIMA H., 1969. Differentiation of pathogenic races of *Pyricularia oryzae* into two groups, indica and japonica. *SABRAO Newsletter*, 1 : 81-94.
- NEI M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89 : 583-590.
- OKA H.I., 1958. Intravarietal variation and classification of cultivated rice. *Indian Journal of Genetics and Plant Breeding*, 18 : 78-89.
- PARSONS J.B., NEWBURY H.J., JACKSON M.T., FORD-LLOYD B.V., 1997. Contrasting genetic diversity relationships are revealed in rice (*Oryza sativa* L.) using different marker types. *Molecular Breeding*, 3 : 115-125.
- PRESSOIR G., ALBAR L., AHMADI N., RIMBAULT I., LORIEUX M., FARGETTE D., GHESQUIERE A., 1998. Genetic basis and mapping of the resistance to rice yellow mottle virus. 2. Evidence of a complementary epistasis between two QTLs. *Theoretical and Applied Genetics*, 97 : 1155 -1161.
- RABARY E., NOYER J.L., BENYAYER P., ARNAUD M., GLASZMANN J.C., 1989. Variabilité génétique du riz (*Oryza sativa* L.) à Madagascar : origine de types nouveaux. *L'Agronomie tropicale*, 44 : 305-312.
- SECOND G., 1982. Origin of the genic diversity of cultivated rice (*Oryza sativa* L.): study of the polymorphism scored at 40 isozyme loci. *Japanese Journal of Genetics*, 57 : 25-57.
- SECOND G., 1985. Evolutionary relationships in the *Sativa* group of *Oryza* based on isozyme data. *Génétique, sélection, évolution*, 17 : 89-114.
- SECOND G., GHESQUIERE A., 1995. Cartographie des introgressions réciproques entre les sous-espèces indica et japonica de riz cultivé (*Oryza sativa* L.). *In* : Colloque techniques et utilisations des marqueurs moléculaires. Paris, France, Inra, les Colloques de l'Inra n° 72, p. 83-93.
- SECOND G., PARCO A., CIAIOLE A., 1995. The hybrid origin hypothesis of cultivated rice (*Oryza sativa*) explains some of the gaps in its RFLP maps and suggests an efficient mapping population for useful genes and QTLs. *In* : Plant genome and plastome: their structure and evolution, K. Tsunewaki éd., Tokyo, Japon, Kodansha Scientific, p. 129-136.
- SECOND G., WANG Z.Y., 1992. Mitochondrial DNA RFLP in genus *Oryza* and cultivated rice. *Genetic Research and Crop Evolution*, 39 : 125-140.

- VIRK P.S., FORD-LLOYD B.V., JACKSON M.T., NEWBURY H.J., 1995. Use of RAPD for the study of diversity within plant germplasm collections. *Heredity*, 74 : 170-179.
- VIRK P.S., FORD-LLOYD B.V., JACKSON M.T., POONI H., CLEMENO T.P., NEWBURY H.J., 1996. Predicting quantitative traits in rice using molecular markers and diverse germplasm. *Heredity*, 76 : 296-304.
- WAN J., IKEHASHI H., 1997. Identification of two types of differentiation in cultivated rice (*Oryza sativa* L.) detected by polymorphism of isozymes and hybrid sterility. *Euphytica*, 94 : 151-161.
- WANG Z.Y., TANKSLEY S.D., 1989. Restriction fragment length polymorphism in *Oryza sativa* L. *Genome*, 32 : 1113-1118.
- YANG G.P., SAGHAI MAROOF M.A., XU C.G., ZHANG Q., BIYASHEV R.M., 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Molecular and General Genetics*, 245 : 187-194.
- ZHANG Q., SAGHAI MAROOF M.A., LU T.Y., SHEN B.Z., 1992. Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. *Theoretical and Applied Genetics*, 83 : 495-499.
- ZHOU H., GLASZMANN J.C., CIENG K.S, SHI X., 1988. A comparison of methods in classification of cultivated rice. *Chinese Journal of Rice Science*, 2 : 1-7 (en chinois).
- ZHU J., GALE M.D., QUARRIE S., JACKSON M.T., BRYAN G.J., 1998. AFLP markers for the study of rice biodiversity. *Theoretical and Applied Genetics*, 96 : 602-611.
- ZHUANG J.Y., QIAN H.R., ZHENG K.L., 1998. Screening of highly-polymorphic RFLP probes in *Oryza sativa* L. *Journal of Genetics and Breeding*, 52 : 39-48.

# Le sorgho

---

Monique Deu, Perla Hamon,  
François Bonnot, Jacques Chantereau

Le sorgho, *Sorghum bicolor* (L.) Moench, est cultivé aussi bien dans les pays tropicaux que tempérés. En 1995, plus de 43 millions d'hectares ont été consacrés à sa culture pour une production globale de plus de 54 millions de tonnes (FAO, 1995). C'est la cinquième céréale mondiale après le blé, le riz, le maïs et l'orge. Malgré l'importance croissante du riz et du maïs, le sorgho demeure un élément essentiel dans l'alimentation humaine pour de nombreux pays d'Afrique (Soudan, Botswana, Burkina, Rwanda, Tchad et Cameroun) et d'Asie (Inde et Chine). La production de sorgho grain s'établit, en 1995, à plus de 16 millions de tonnes en Afrique et à 15 millions de tonnes en Asie. Les grains sont consommés entiers ou sous forme de bouillie ou de galettes après avoir été transformés en farine. Certains types variétaux mieux adaptés sont utilisés pour la fabrication de bière, de grains sucrés ou de pop-corn.

D'autres utilisations du sorgho sont possibles. En Argentine, en Australie, en Afrique du Sud, au Mexique et surtout aux Etats-Unis, dont la production en 1995 a dépassé 17 millions de tonnes, le sorgho est principalement destiné à l'alimentation animale. Enfin, dans certaines régions d'Afrique et d'Asie, la panicule peut être utilisée pour la confection de balais tandis que les tiges sont exploitées sous forme de fourrage, de combustible ou de matériau de construction ou peuvent servir à la teinture des cuirs et à la fabrication de papier. La moelle peut donner du sucre, du sirop, des colles et des alcools.

Le sorgho est une plante très rustique, adaptée à des environnements difficiles, qui supporte les sols pauvres, la sécheresse, les températures élevées et même l'inondation. Le sorgho peut se développer là où d'autres cultures plus prestigieuses ne peuvent le faire.

## La taxonomie et les ressources génétiques

### La taxonomie du genre *Sorghum* et sa répartition géographique

Le sorgho (genre *Sorghum*, famille des poacées) est une céréale proche du maïs et de la canne à sucre ; tous trois appartiennent à la tribu des andropogonées. La grande diversité morphologique du genre *Sorghum* a conduit les botanistes à multiplier les taxons — 712 ont été décrits par SNOWDEN (1936). Une classification simplifiée, prenant en compte les échanges géniques, est largement utilisée de nos jours (DE WET, 1978). Préconisée par l'IPGRI (International Plant Genetic Resources Institute), elle divise le genre *Sorghum* en cinq sections. La section *Sorghum* inclut tous les sorghos à grains cultivés (*S. bicolor* subsp. *bicolor*, diploïde à  $2x = 20$ ), les sorghos sauvages, diploïdes, annuels, originaires d'Afrique (*S. bicolor* subsp. *verticilliflorum*), les sorghos sauvages, diploïdes, à rhizomes, pérennes, présents en Inde, au Sri Lanka et en Asie du Sud-Est (*S. propinquum*), et les sorghos sauvages, tétraploïdes, également à rhizomes et pérennes, rencontrés en Asie du Sud-Est, en Inde, au Moyen-Orient et sur le pourtour méditerranéen (*S. halepense*). Selon MANN *et al.* (1983), la domestication du sorgho se serait produite aux environs de 3000 avant notre ère dans la partie nord-est de l'Afrique. Cependant, des données archéologiques plus récentes (WENDORF *et al.*, 1992) permettraient de situer les premières utilisations du sorgho à plus de 6 000 ans avant notre ère. Selon HARLAN et STEMLER (1976) et DOGGETT (1988), les sorghos cultivés dériveraient des sorghos sauvages africains *S. bicolor* subsp. *verticilliflorum*.

Les sorghos cultivés présentent une grande diversité phénotypique. La forme de l'inflorescence et surtout la structure de l'épillet ont permis de les classer en cinq races principales (*bicolor*, *caudatum*, *durra*, *guinea* et *kafir*) et dix races intermédiaires obtenues par la combinaison deux à deux des précédentes (*guinea-bicolor*, *durra-caudatum* ; HARLAN et DE WET, 1972). Cependant, selon DOGGETT (1988), les sorghos *bicolor*, sélectionnés sur des caractères non liés à la structure de l'épillet, tels que la tige sucrée et le fourrage, ne constituent pas une race mais forment plutôt un ensemble assez hétérogène. Ces sorghos seraient proches des sorghos primitifs à partir desquels s'est faite la domestication.

Les différentes races de sorgho occupent actuellement des aires de répartition non disjointes bien qu'une race puisse être prédominante dans une région. Les sorghos cultivés en Inde sont principalement des durra, que l'on retrouve aussi en Afrique de l'Est (Ethiopie) et au Moyen-Orient (Turquie, Syrie...). Le type bicolor est très répandu en Afrique et en Asie, tandis que les kafir se rencontrent essentiellement en Afrique australe, de la Tanzanie jusqu'en Afrique du Sud. La race caudatum, prédominante en Afrique centrale, est devenue une source importante de matériel génétique dans les programmes de sélection pour les pays tempérés. Les guinea, sorghos typiques de l'Afrique de l'Ouest, sont aussi présents en Afrique de l'Est et en Afrique australe.

La dispersion et la répartition géographique actuelle des sorghos cultivés (figures 1 et 2) sont intimement liées aux migrations humaines passées, aux traditions culturelles et aux usages alimentaires des ethnies qui les cultivent. En Afrique, la race durra est ainsi associée à l'Islam par opposition aux kafir, sorghos cultivés par les « infidèles ». La large répartition géographique des sorghos cultivés montre aussi que les races se sont adaptées à des zones écologiques différentes. La race guinea — grands sorghos à panicule très lâche facilitant le séchage et à grains vitreux, résistants aux moisissures — est très souvent cultivée dans les zones humides, tandis que les durra — sorghos courts à panicule généralement compacte et à gros grains — sont bien adaptés à la sécheresse.

Les sorghos cultivés sont monoïques et préférentiellement autogames bien que des taux d'allogamie relativement élevés (10 à 30 %) aient été mis en évidence dans des populations de guinea (OLLITRAULT, 1987 ; CHANTEREAU et KONDOMBO, 1994). Toutes les races de sorghos cultivés peuvent se croiser entre elles et s'hybrider avec les sorghos sauvages diploïdes apparentés (*S. bicolor* subsp. *verticilliflorum*). De même, des flux de gènes sont possibles entre *S. bicolor* et *S. propinquum*, sauvage diploïde (DE WET et al., 1976), et entre *S. bicolor* et *S. halepense*, sauvage tétraploïde (ARRIOLA, 1995).

## Les ressources génétiques

Dans les années 70, la communauté scientifique internationale prend conscience de la disparition inquiétante des variétés traditionnelles de plusieurs plantes cultivées ainsi que de celle de nombreuses populations sauvages. Cette disparition est liée à l'abandon des variétés traditionnelles au profit des nouvelles variétés améliorées et à la destruction rapide des habitats naturels, souvent provoquée par une urbanisation non maîtrisée. A l'échelle mondiale, l'ICRISAT (International Crops Research Institute for the Semi-Arid Tropics) est alors mandaté pour assurer la collecte et la conservation des ressources génétiques du sorgho. Dépositaire en 1972 d'environ 10 000 accessions collectées auparavant par la fondation Rockefeller, l'Unité de ressources génétiques de l'ICRISAT, en Inde, dispose au début des années 90 d'environ 33 100 accessions collectées dans 86 pays. Cette collection devrait atteindre



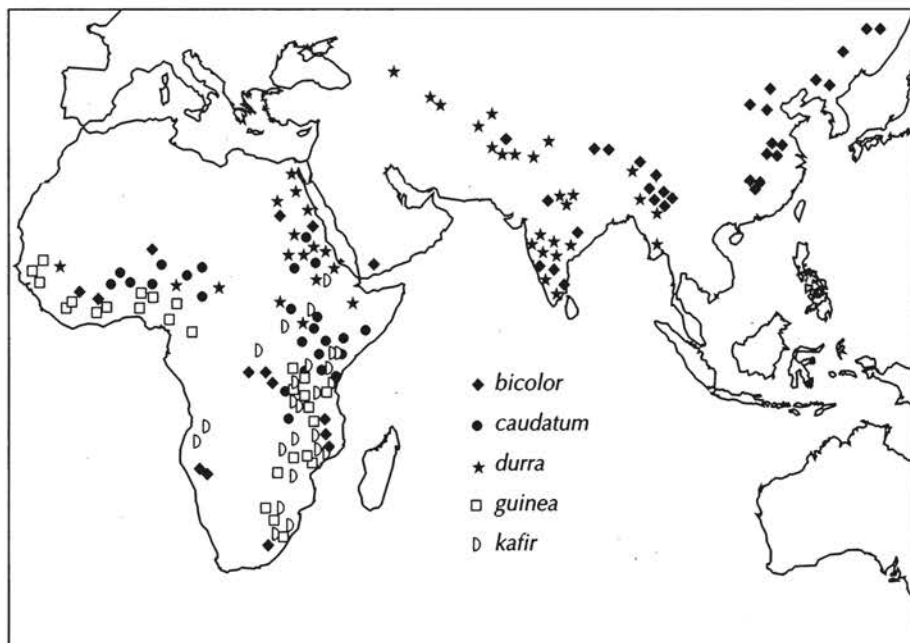


Figure 1. Répartition des races de sorghos cultivés.



Figure 2. Les races de sorghos cultivés et leurs hybrides : durra, à gauche, guinea, en haut, caudatum, en bas, bicolor, au centre (photo J.L. Noyer-Cirad).

45 000 accessions à la fin de cette décennie (PRASADA RAO et RAMANATHA RAO, 1995). Les accessions sont conservées et multipliées sous la responsabilité de l'ICRISAT à Patancheru, en Inde. Elles ont été pour la plupart dupliquées aux Etats-Unis, où plus de 42 000 accessions sont recensées par le NPGS (National Plant Germplasm System). Leur conservation à long terme est assurée dans trois sites principaux : à Fort Collins, au Colorado, par le NSSL (National Seed Storage Laboratory) et à Griffin, en Georgie, et Mayaguez, à Porto Rico, par l'USDA (United States Department of Agriculture). Parallèlement, des collections de taille plus modeste ont été constituées dans différents pays. Par exemple, en Chine, 10 386 accessions collectées à partir des années 50 sont conservées par la banque nationale de gènes à Pékin. La collection de sorgho d'Afrique du Sud compte environ 4 000 accessions, celle d'Australie en comprend 3 800. En France, 5 850 accessions sont conservées, dont 3 850 par l'IRD (Institut de recherche pour le développement, anciennement Orstom) et 2 000 par le Cirad. Toutes ces collections ont en commun deux points faibles : leur faible représentativité en sorghos sauvages — c'est ainsi qu'ils ne représentent que 1,2 % de la collection des Etats-Unis et 1,1 % de celle de l'ICRISAT — et une représentativité inégale des races et des origines géographiques — la collection de l'ICRISAT, par exemple, comporte 17,4 % de guinea, 20,2 % de durra, 21,8 % de caudatum, contre 2,3 % de kafir et 3,2 % de bicolor, et provient essentiellement de cinq pays, l'Inde (17,3 %), le Cameroun (10,2 %), le Yémen (8,5 %), le Soudan (8,1 %) et l'Ethiopie (8,1 %).

Au centre indien de l'ICRISAT, à Patancheru, les accessions sont évaluées à l'aide d'un ensemble de 20 descripteurs morphologiques, dont 19 appartiennent à la liste recommandée par l'IBPGR (IBPGR, 1993). Les données « passeport » et d'évaluation sont complètes pour environ 28 000 accessions. En revanche, aux Etats-Unis, seulement 13 % des accessions de la collection globale ont été complètement évaluées.

La gestion et l'évaluation de telles collections posent des problèmes de coût. Néanmoins, l'ICRISAT s'est efforcé de procéder à une évaluation plus complète. Selon DAHLBERG et SPINKS (1995), des efforts devraient être aussi consentis par les Etats-Unis, où l'évaluation en champ (40 descripteurs) de toutes les accessions en collection devrait être entreprise.

La constitution d'une banque de données facilement accessible, permettant l'utilisation effective de ces ressources génétiques par les programmes nationaux de recherche, devient donc une étape essentielle. Aux Etats-Unis, les données sont régulièrement mises à disposition via la base de données du réseau GRIN (Germplasm Resources Information Network), accessible sur Internet (<http://www.ars-grin.gov> ; DAHLBERG et SPINKS, 1995). A l'ICRISAT, la base de données est aussi consultable par Internet (<http://noc1.cgiar.org/search-type.htm>).

La distribution du matériel génétique se fait naturellement sous forme de graines. Entre 1985 et 1995, 118 381 échantillons ont ainsi été distribués dans

le monde à partir des Etats-Unis (DAHLBERG et SPINKS, 1995) et 237 265 par l'ICRISAT (MENGESHA et APPA RAO, 1994). D'autres pays détenteurs de collections plus restreintes participent aussi aux échanges de matériel végétal.

Dans ce chapitre, nous passons en revue les différentes études portant sur la diversité génétique des sorghos cultivés. Puis nous comparons les structurations obtenues par l'analyse d'un échantillon à l'aide de trois types de marqueurs (morphologiques, enzymatiques et moléculaires) et ce, en utilisant plusieurs méthodologies. Les résultats sont considérés dans une optique d'application à la constitution de *core collections*. En conclusion, nous examinons la stratégie de conservation des ressources génétiques et l'intérêt des marqueurs moléculaires pour la compréhension de la structuration génétique des sorghos cultivés.

## L'organisation de la diversité génétique

### La diversité génétique révélée par les descripteurs morphologiques

Les travaux portant sur l'organisation de la diversité génétique révélée par les caractères morphologiques sont peu nombreux. La première étude, réalisée par CHANTEREAU *et al.* (1989), a porté sur 157 écotypes de race et d'origine géographique très diverses. Ces auteurs ont montré que les 25 caractères agromorphologiques étudiés, dont 14 figurent dans la liste de l'IBPGR, permettent de classer les sorghos en trois groupes : le groupe des guinea et bicolor, le groupe des caudatum et kafir et le groupe des durra. Ces groupes se distinguent principalement par leur comportement en culture. Ces résultats traduisent certainement l'adaptation des races aux modes de culture, aux utilisations particulières et aux contraintes biotiques et abiotiques. Ainsi, les guinea, tout comme les bicolor, sont des sorghos rustiques de zones humides, adaptés à la culture extensive. Les durra sont des sorghos rustiques de zones sèches et, occasionnellement, des sorghos de décrue, tandis que les caudatum et les kafir se présentent comme les sorghos les plus modernes, les mieux adaptés à une culture semi-intensive, voire intensive.

Plus récemment, APPA RAO *et al.* (1996) ont étudié près de 4 000 accessions de sorghos originaires de différents Etats de l'Inde et présents dans la collection de l'ICRISAT à l'aide de 14 descripteurs morphologiques et agronomiques. Dans ce travail, les données n'ont pas été traitées par les analyses multivariées. En revanche, les analyses descriptives univariées indiquent une grande diversité morphophysologique avec une diversité inter-Etats plus forte que la diversité intra-Etat. Toutes les races sont présentes en Inde, mais les durra et leurs inter-

médiaires sont largement prédominants. Aucune information n'est donnée quant à la répartition des races en fonction des différents caractères.

En Ethiopie, TESHOME *et al.* (1997) se sont intéressés à une région particulière, constituée par le Nord-Shewa et le Sud-Welo. Dans cette aire géographique restreinte, l'étude de 14 caractères morphologiques, dont 7 figurent dans la liste de l'IBPGR, révèle une grande diversité phénotypique. Quatre races pures — bicolor, caudatum, durra et guinea — et une race intermédiaire — durra-bicolor — sont maintenues par la sélection humaine. Les dendrogrammes obtenus ne révèlent pas une structuration taxonomique claire. En revanche, les analyses multivariées permettent de distinguer trois groupes, essentiellement selon deux critères : la présence ou l'absence de sucre dans la tige et la rondeur des grains. Cette classification, différente de celle de CHANTEREAU *et al.* (1989), repose en fait sur l'utilisation d'un jeu de descripteurs différent.

Dans notre travail, l'analyse a porté sur un échantillon de 230 accessions comportant une plus forte proportion de guinea et de caudatum mais moins de durra, de kafir et d'intermédiaires que celui de CHANTEREAU *et al.* (1989). Vingt et un des 25 caractères précédemment étudiés par CHANTEREAU *et al.* (1989) ont été pris en compte. Soixante-dix modalités ont été définies à partir des 21 variables puis une analyse factorielle des correspondances (AFC) a été conduite sur tableau disjonctif (tous les individus et modalités actifs). La projection des variables sur les premiers plans de l'AFC montre que l'axe 1 (12,7 %) sépare les sorghos de taille réduite, à tige épaisse et à glumes plus courtes que le grain, des sorghos de grande taille, à panicule longue et lâche. L'axe 2 (7,2 %) isole les sorghos à pédoncule crossé et à panicule compacte. L'axe 3 (5,4 %) oppose les sorghos précoces, présentant peu d'entre-nœuds, aux variétés plus tardives et de grande taille. La représentation de l'appartenance raciale sur le plan 1-2 montre des résultats similaires à ceux de CHANTEREAU *et al.* (1989) avec la différenciation des trois ensembles : les guinea et bicolor, les caudatum et kafir et les durra.

La classification ascendante hiérarchique (CAH ; figure 3) donne une image plus globale de la structuration et fournit des indications sur les relations entre les groupes en terme de proximité et d'éloignement. Les guinea (G) et les bicolor (B) forment bien un seul groupe très variable ; les kafir (K) constituent un ensemble relativement homogène ; les caudatum (C), en revanche, se séparent en deux sous-groupes, C1 et C2. Le sous-groupe C1 est caractérisé par des variétés présentant un cycle court et peu d'entre-nœuds (moins de dix), alors que le sous-groupe C2 renferme des sorghos de cycle moyen et à nombre d'entre-nœuds plus élevé. Les deux sous-groupes de caudatum apparaissent assez éloignés l'un de l'autre avec des distances entre les sous-groupes comparables aux distances entre les races ; le sous-groupe C1 est plus proche des kafir que du sous-groupe C2. La séparation en deux sous-groupes pourrait être le reflet d'une différence de comportement à l'égard de la photopériode. Le sous-groupe C1, proche des kafir, devrait donc être peu sensible ou insensible à la photopériode, tandis que le sous-groupe C2, proche des durra (D), devrait

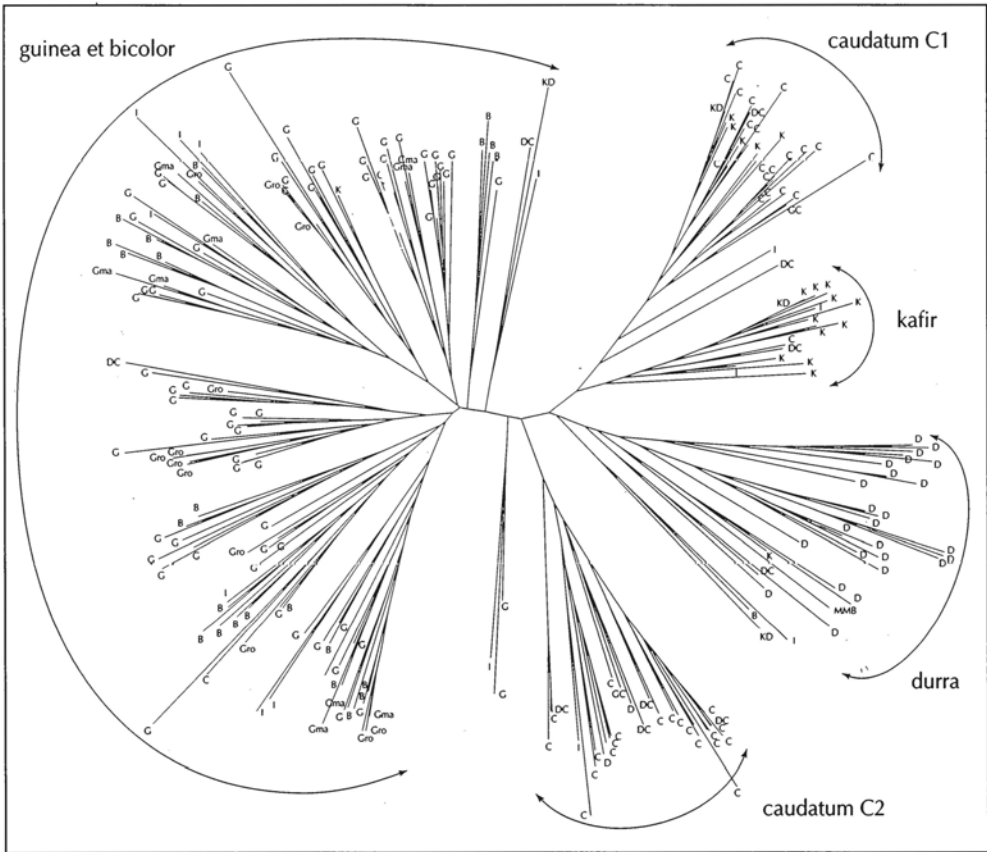


Figure 3. Diversité génétique des sorghos cultivés révélée par les marqueurs morphologiques. L'arbre est construit à partir de l'indice de Sokal et Michener.

être sensible à la photopériode (CHANTEREAU *et al.*, 1997). Cette sous-structuration des caudatum est passée inaperçue dans l'étude de CHANTEREAU *et al.* (1989) et dans l'AFC présentée ci-dessus, ce qui s'explique par le fait que les deux sous-groupes, situés de part et d'autre de l'axe 1 dans le plan 1-3 de notre AFC (figure 4), forment un ensemble apparemment continu.

Aucun groupe n'est totalement homogène du point de vue racial. En règle générale, les individus appartenant à des races intermédiaires se répartissent dans les différents groupes en fonction de leur degré de ressemblance avec les sorghos de race plus pure. Cette observation pourrait traduire une dynamique dans l'évolution des sorghos liée aux nombreux brassages génétiques favorisés par des pratiques culturelles traditionnelles (association de plusieurs variétés, voire de plusieurs races, dans un même champ) et aux différentes pressions de sélection, naturelle et humaine, exercées. Il serait intéressant de voir la position relative de ces intermédiaires par rapport à d'autres critères de classification — comme les critères utilisés par TESHOME *et al.* (1997), ou des critères de

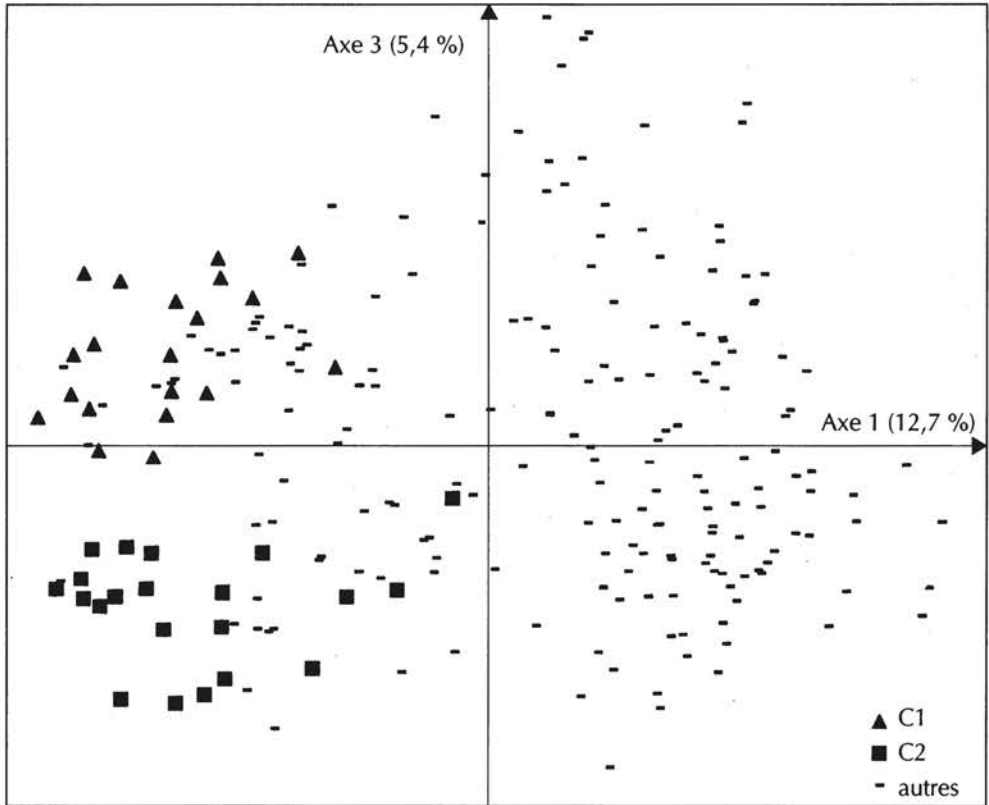


Figure 4. Analyse factorielle des correspondances sur les caractères morphologiques.

qualité technologique du grain, de résistance à des stress biotiques et abiotiques — afin de mieux comprendre leur rôle dans l'évolution des sorghos. Contribuent-ils à l'élargissement de la variabilité génétique tout en évoluant vers un type parental ? Sont-ils maintenus ou en permanence produits et éliminés ? Leur compétitivité face aux races pures serait très forte au Tchad si on en juge par la nature des races qui sont traditionnellement en culture dans cette région (YAGOUA, 1995). En revanche, la situation semble différente en Ethiopie, où TESHOME *et al.* (1997) indiquent la présence de cinq races en culture, dont quatre sont des races « pures ».

### La diversité génétique révélée par les marqueurs enzymatiques

L'analyse du polymorphisme enzymatique des sorghos cultivés a fait l'objet de travaux réalisés aux Etats-Unis et en France au cours des dix dernières années (MORDEN *et al.*, 1989 ; OLLITRAULT *et al.*, 1989b ; ALDRICH *et al.*, 1992 ; DEGREMONT, 1992).

Les travaux de MORDEN *et al.* (1989), complétés par ALDRICH *et al.* (1992), montrent que la plus grande partie de la diversité génétique totale est due bien plus à des différences d'origine géographique qu'à des différences raciales. Ils indiquent aussi que les régions ouest et est-africaines possèdent le niveau d'hétérozygotie le plus élevé et que l'Afrique australe présente la diversité génétique la plus faible.

Pour OLLITRAULT *et al.* (1989b), cette différenciation géographique apparaît plus marquée. Ces auteurs distinguent trois pôles : un pôle ouest-africain, un pôle est et centre-africain et un pôle sud-africain. De plus, une structuration selon les zones géographiques est observée pour deux races, bicolor et guinea. La différenciation à l'intérieur des guinea de trois groupes bien distincts — les guinea d'Afrique de l'Ouest, les guinea d'Afrique australe et les guinea de type margaritiférum — a été par la suite largement démontrée par les travaux de DEGREMONT (1992).

Notre étude a porté sur 230 accessions, représentatives de la diversité géographique et raciale. Elle permet de constater que cet échantillon présente une bonne diversité génétique, avec 11 locus polymorphes (au seuil de 99 %) et un nombre moyen d'allèles par locus polymorphe de 2,8. La structuration de la diversité mise en évidence par la CAH (figure 5) est en accord avec la différenciation géographique Afrique de l'Ouest - Afrique australe notée par OLLITRAULT *et al.* (1989b). De plus, cette étude aboutit aux conclusions suivantes. Les accessions d'Afrique de l'Est et du Centre présentent une large gamme de variabilité et ne forment pas un groupe bien différencié. Aucun groupe n'est totalement homogène, ni du point de vue racial ni sur le plan géographique. Les guinea d'Afrique australe présentent une diversité génétique moindre que ceux d'Afrique de l'Ouest. Les guinea d'Afrique australe sont plus proches des kafir que de tout autre groupe. Les guinea d'Afrique du Centre et de l'Est se retrouvent préférentiellement dans le groupe d'Afrique de l'Ouest. Les guinea margaritiférum (Gma) forment un ensemble très homogène et bien distinct de tous les autres guinea. Très curieusement, ils sont relativement proches d'un petit groupe de 12 accessions, dont les deux tiers sont des caudatum appartenant pour part égale aux sous-groupes C1 et C2 précédemment décrits.

Alors que la structuration à l'aide de descripteurs morphologiques met en évidence des groupes de comportement en culture, la structuration géographique obtenue à l'aide de marqueurs enzymatiques traduit dans ses grandes lignes les aires de distribution privilégiées des différentes races : guinea en Afrique de l'Ouest, guinea et kafir en Afrique australe, durra et caudatum en Afrique de l'Est et du Centre. Cependant, le fait de trouver, par exemple, des durra d'Afrique de l'Ouest dans le groupe des guinea d'Afrique de l'Ouest pourrait indiquer que ces sorghos (en dehors des erreurs d'étiquetage ou d'identification) sont d'introduction ancienne en Afrique de l'Ouest. Ils ont ainsi pu suivre une évolution différente, sous les contraintes de l'Afrique de l'Ouest, de celle des autres durra et s'intégrer totalement dans le groupe d'Afrique de l'Ouest

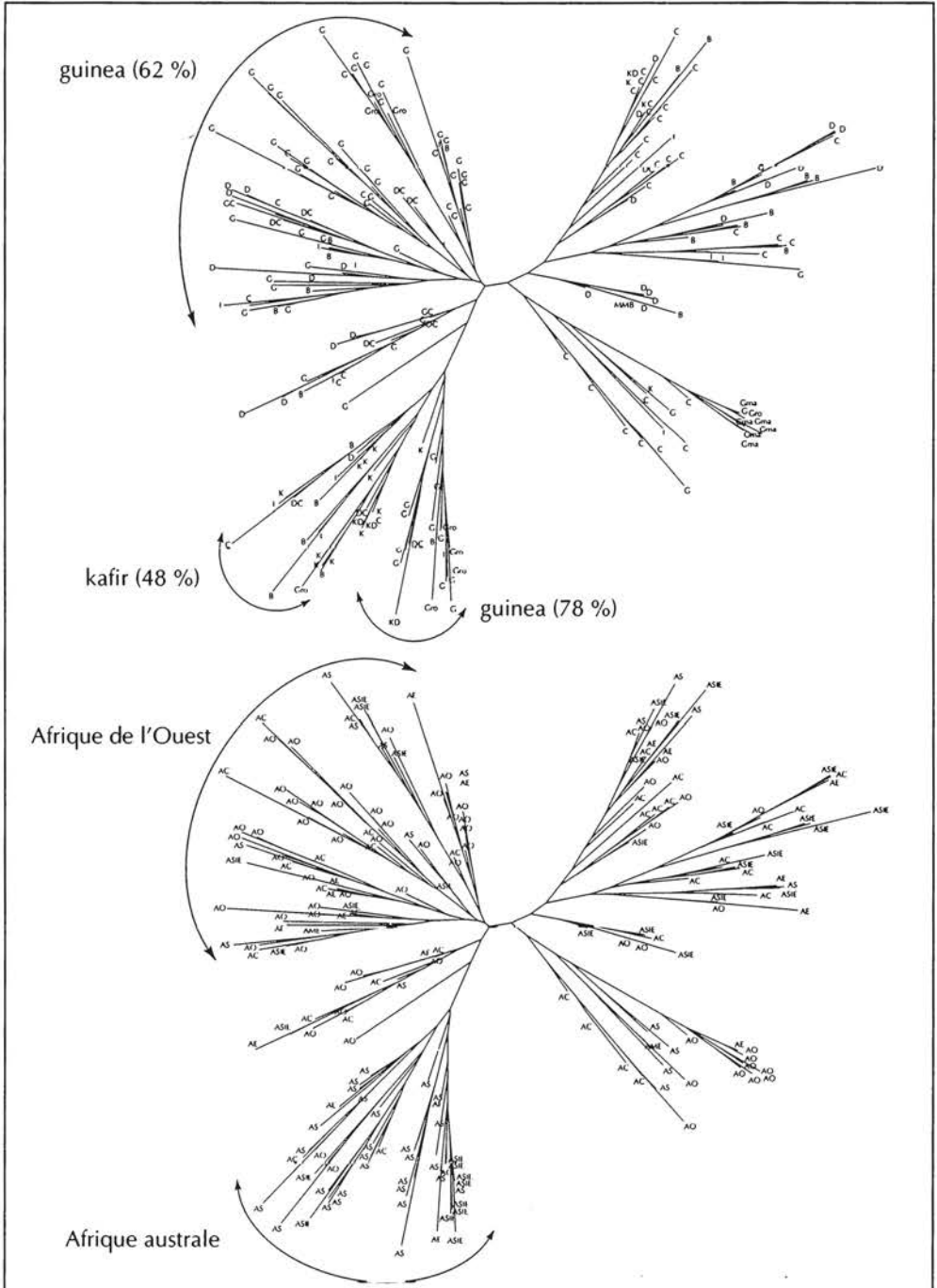


Figure 5. Diversité génétique des sorghos cultivés révélée par les marqueurs enzymatiques. L'appartenance raciale est indiquée sur l'arbre du haut et l'origine géographique sur l'arbre du bas. L'arbre a été construit en utilisant l'indice de Dice.



grâce aux échanges géniques. A l'opposé, les *caudatum* récemment introduits en Afrique de l'Ouest gardent encore des structures génotypiques liées à leur origine centre et est-africaine et ne se retrouvent pas dans le groupe d'Afrique de l'Ouest.

## La diversité génétique révélée par les marqueurs moléculaires

La diversité génétique a tout d'abord été étudiée par ALDRICH et DOEBLEY (1992) à l'aide de marqueurs RFLP (38 sondes) et par TAO *et al.* (1993) en utilisant des marqueurs RFLP (16 sondes) et RAPD (29 amorces). Ces travaux portant sur des échantillons restreints, de moins de 50 accessions, ne permettent pas d'observer une structuration raciale ou géographique marquée. En revanche, leurs auteurs concluent que les marqueurs RFLP et RAPD mettent en évidence une plus grande diversité allélique que les marqueurs enzymatiques. Cependant, la capacité des marqueurs à révéler une structuration dépend du fait qu'il existe ou non une structuration, de la représentativité de l'échantillon analysé, ainsi que du nombre et du type de marqueurs utilisés. Les travaux de ALDRICH et DOEBLEY (1992) portent sur 31 accessions représentant équitablement les cinq races principales mais échantillonnées dans 10 pays répartis inégalement entre les quatre grandes régions d'Asie et d'Afrique. De même, les 36 accessions étudiées par TAO *et al.* (1993) ne sont représentatives ni de la diversité raciale ni des grandes aires de distribution des sorghos.

L'étude à l'aide de 33 sondes RFLP de DEU *et al.* (1994) a porté sur un échantillon de 94 accessions, qui tenait compte de la diversité raciale et géographique des sorghos. Elle met en évidence un pôle Afrique australe, dont la variabilité génétique est plus restreinte que celle de l'Afrique de l'Ouest ou de l'Afrique de l'Est et du Centre. Ces travaux ont aussi révélé une différenciation raciale. Hormis les accessions de type *bicolor*, qui ne constituent pas un ensemble homogène, les races *caudatum*, *durra* et *kafir* forment trois groupes distincts, tandis que les *guinea* sont scindés en trois sous-groupes : les *guinea* d'Afrique de l'Ouest, les *guinea* d'Afrique australe et les *guinea margaritifera*. Cette étude confirme la structuration interne aux *guinea* observée avec les isoenzymes et permet la distinction de races non révélée par les caractères morphologiques (*caudatum* et *kafir*) ou enzymatiques (*caudatum* et *durra*).

L'étude réalisée par CUI *et al.* (1995) avec 61 sondes RFLP sur 41 accessions montre une structuration raciale moins nette puisque les *kafir* et les *guinea* d'Afrique australe et d'Asie forment un seul groupe. Cette étude met aussi en évidence l'originalité d'un *guinea margaritifera*, qui se trouve plus proche des sorghos sauvages que des sorghos cultivés. Pour les auteurs, ce regroupement est en accord avec le phénotype plutôt « sauvage » de cette accession.

La singularité des guinea margaritifera est aussi mentionnée dans l'étude réalisée grâce à 20 sondes RFLP, 13 amorces RAPD et 4 ISSR par DE OLIVEIRA *et al.* (1996) : les trois accessions de ce type analysées diffèrent des autres sorghos cultivés bien plus parce qu'elles ne partagent pas les allèles les plus communs que parce qu'elles possèdent des allèles spécifiques.

Une importante étude a été réalisée récemment par MENKIR *et al.* (1997) sur 190 accessions représentatives des cinq races principales, dont les origines géographiques sont très variées — 13 pays pour les kafir et de 28 à 32 pays pour les autres races. Les analyses effectuées à l'aide de 82 amorces RAPD montrent que la diversité génétique est bien plus forte au sein des bicolor et des guinea que des autres races. Les kafir sont les moins diversifiés. De plus, la variabilité totale s'explique pour 86 % par la variabilité intrarace et pour 14 % par les différences entre races. Seulement 13 % de la variabilité totale est liée à une différenciation géographique. Dans ces conditions, les auteurs n'observent pas de structuration raciale ou géographique. Ces résultats, apparemment en contradiction avec ceux de DEU *et al.* (1994), pourraient s'expliquer soit par le mode d'échantillonnage, soit par la nature des marqueurs utilisés.

En premier lieu, l'échantillonnage des origines géographiques pour chaque race pourrait être trop artificiel et ne refléterait pas la distribution traditionnelle des races — les pays d'origine ne sont pas précisés dans l'étude. En effet, les migrations humaines et les échanges de matériel favorisent les contacts et les flux de gènes entre races, ce qui contribue à élargir la base génétique de chacune des races. Néanmoins, bien qu'il existe de nombreux intermédiaires, à des degrés divers, les types raciaux majeurs persistent en culture. Ainsi, dans le Nord-Shewa et le Sud-Welo, en Ethiopie, quatre des cinq races majeures — bicolor, caudatum, durra et guinea — coexistent (TESHOME *et al.*, 1997). Quoique génétiquement proches, les kafir et les guinea d'Afrique australe maintiennent des caractéristiques morphologiques différentes. Cela pourrait témoigner de l'existence de pressions de sélection très fortes sur les caractères morphologiques et donc du fait que les races majeures correspondent à des types à haute spécialisation écogéographique. Les migrations récentes contribuent, pour leur part, à élargir artificiellement la variabilité intrarégionale et à diminuer d'autant la différenciation interrégionale.

En second lieu, les marqueurs RAPD utilisés dans cette étude sont surtout multilocus, avec une moyenne de 4,2 bandes révélées par amorce, alors que près de 75 % des sondes testées en RFLP par DEU *et al.* (1994) correspondent à des locus simple-copie. Dans ces conditions, les RAPD pourraient correspondre à des régions non codantes ou à des séquences répétées dans le génome, tandis que les RFLP correspondraient davantage à des séquences codantes, uniques. Les taux d'évolution de ces régions sont probablement différents et, de ce fait, ces deux types de marqueurs ne traduiraient pas les mêmes processus évolutifs. Par ailleurs, les phénomènes d'homoplasie, plus fréquents avec les RAPD qu'avec les RFLP, pourraient rendre compte des différences de structures observées (POWELL *et al.*, 1996).

## Comparaison des structurations génétiques révélées par les trois types de marqueurs

Les données RFLP étant disponibles pour 92 accessions déjà caractérisées par les marqueurs morphologiques et enzymatiques, nous avons tout d'abord comparé les structurations obtenues sur ce sous-ensemble avec celles qui ont été révélées par l'étude de l'échantillon initial de 230 accessions.

Pour la diversité morphologique, toutes les modalités, y compris celles de faible fréquence, rencontrées dans la collection initiale sont présentes dans le sous-ensemble. Les représentations radiales obtenues dans les deux cas (230 et 92) sont très semblables. Le groupe présentant la plus grande diversité est constitué des guinea et des bicolor. Quatre autres groupes sont observés : un groupe de durra, un groupe de kafir et deux groupes de caudatum. Dix-huit accessions occupent des positions intermédiaires entre ces groupes.

Les 31 allèles détectés à partir des 11 locus enzymatiques dans les 230 accessions ont été conservés dans le sous-ensemble. La structuration observée fait apparaître trois groupes majeurs identifiables : les guinea margaritifera, un groupe d'Afrique australe incluant les kafir et des guinea et un groupe comportant les caudatum et les durra. Entre ces trois groupes, on note la présence de nombreuses accessions occupant des positions intermédiaires. Le groupe d'Afrique de l'Ouest observable dans l'échantillon initial n'apparaît plus clairement dans le sous-ensemble.

La comparaison des structurations génétiques révélées par les trois types de marqueurs a été réalisée sur ce sous-ensemble de 92 accessions.

Les nombres moyens d'allèles par locus polymorphe apparaissent équivalents quelle que soit la nature du marqueur considéré : 2,8 pour les isoenzymes contre 2,9 pour les RFLP (si l'enzyme de restriction est *Hind*III) et 3,0 (avec *Xba*I). Dans ce cas particulier, les RFLP ne permettent pas d'accéder à un polymorphisme allélique plus important que les isoenzymes.

Nous avons pu identifier 11 locus polymorphes pour les isoenzymes et 33 pour les RFLP. Bien plus que le pourcentage de locus polymorphes, ce qui semble intéressant c'est le nombre de locus polymorphes auxquels il est possible d'accéder relativement facilement. Dans l'étude d'OLLITRAULT *et al.* (1989b) portant sur 348 accessions, 18 locus sur les 25 testés se sont révélés polymorphes. Il paraît donc difficile d'accéder à plus de 20 locus enzymatiques polymorphes alors que CUI *et al.* (1995) ont utilisé 61 locus polymorphes RFLP.

Dans notre étude, les 11 locus enzymatiques permettent d'identifier 83,6 % des génotypes (soit 77 sur 92), tandis que les 33 locus RFLP conduisent à 94,5 % d'identification génotypique. Toutefois, les isoenzymes discriminent, dans certains cas, des accessions non différenciées par les marqueurs RFLP. Ainsi, le locus enzymatique *EST-D* permet de différencier les accessions de chacun des deux couples d'accessions qui apparaissent identiques selon les RFLP. Deux

autres accessions identiques avec les RFLP portent des allèles différents à trois locus enzymatiques (*END*, *HEX* et *LAP*). La combinaison des marqueurs isoenzymatiques et RFLP permet de discriminer toutes les accessions à l'exception de deux couples formés par deux caudatum, d'une part, et deux guinea roxburghii (Gro), d'autre part. Dans ces conditions, il serait intéressant d'identifier la combinaison minimale de marqueurs isoenzymatiques et RFLP qui permette d'identifier la totalité des génotypes, puis de tester son efficacité sur d'autres ensembles d'accessions.

La comparaison des structures révélées par les différents marqueurs montre que la forte diversité morphologique des sorghos guinea s'accompagne d'une grande diversité génétique, avec trois groupes de différenciation nettement marqués par les RFLP. En outre, elle indique que les sorghos caudatum et durra, bien différenciés sur le plan morphologique, constituent deux groupes dont la proximité génétique est mise en évidence par les marqueurs isoenzymatiques et RFLP. De plus, les kafir, relativement homogènes du point de vue morphologique, sont génétiquement plus proches des guinea d'Afrique australe que des autres sorghos.

En fait, bien que les structurations obtenues avec les trois types de marqueurs ne soient pas parfaitement superposables, celles-ci concordent avec les connaissances acquises sur le sorgho.

❑ Les sorghos sont préférentiellement autogames, mais l'allogamie naturelle existante, bien que faible parfois, est favorisée par les pratiques culturelles traditionnelles. De ce fait, les brassages génétiques augmentent la diversité morphologique (dans la limite de la sélection pratiquée) et surtout la diversité génétique pour les caractères, isoenzymatiques et moléculaires, sélectivement neutres *a priori*.

❑ Les guinea margaritifera possèdent la structure génétique la plus différenciée au sein des sorghos cultivés. DEU *et al.* (1995) ont montré que ces guinea présentent un polymorphisme mitochondrial qui les distingue de tous les autres sorghos, cultivés ou sauvages, appartenant à l'espèce *S. bicolor*. Du fait de leurs caractéristiques génétiques et agromorphologiques, ces guinea pourraient susciter l'intérêt des sélectionneurs.

❑ Les guinea d'Afrique australe sont plus proches génétiquement des kafir que des autres guinea. En fait, ces deux groupes de sorgho partageant la même aire de distribution, les flux de gènes se produisent naturellement. La diversité intragroupe augmente donc au détriment de la diversité intergroupe.

Il existe d'autres méthodes pour comparer les informations fournies par différents marqueurs. L'utilisation du test de Mantel et la construction d'arbres consensus ou d'arbres communs minimaux en sont des exemples. Dans le premier cas, des coefficients de corrélation,  $r$ , entre les matrices d'indices de similarité produites à partir des différents marqueurs sont calculés. L'accent est alors mis sur la valeur des distances entre individus. La signification du  $r$  cal-

culé est totalement subjective. Selon ROHLF (1990), la corrélation peut être considérée comme très forte si  $r$  est supérieur ou égal à 0,9, forte si  $r$  est compris entre 0,8 et 0,9, faible si  $r$  est compris entre 0,7 et 0,8 et très faible si  $r$  est inférieur à 0,7. Plusieurs études font référence à l'usage de ce test (MESSMER *et al.*, 1991 ; ENGQVIST et BECKER, 1994 ; THORMANN *et al.*, 1994 ; POWELL *et al.*, 1996). Cependant, peu d'études prennent en compte les marqueurs enzymatiques. ENGQVIST et BECKER (1994) ont calculé les coefficients de corrélation obtenus chez *Brassica napus* à partir des marqueurs RFLP, RAPD et enzymatiques. Le coefficient le plus bas est obtenu avec le couple RFLP-isoenzymes (0,53 contre 0,67 pour le couple RAPD-isoenzymes et 0,76 pour le couple RFLP-RAPD). De même, chez le maïs, MESSMER *et al.* (1991) obtiennent une faible corrélation ( $r = 0,23$ ) pour le couple RFLP-isoenzymes.

Dans notre cas, en accord avec les structurations observées, les corrélations sont très faibles :  $r$  est égal à 0,19 pour le couple isoenzyme-morphologie, à 0,2 pour le couple RFLP-morphologie et à 0,3 pour le couple RFLP-isoenzymes.

La comparaison par construction d'arbres considère les structures obtenues sans tenir compte de la longueur des arêtes.

Dans notre cas, les arbres consensus obtenus sont du type étoile ou « parapluie ouvert », ce qui indique l'absence de regroupements strictement identiques (structures communes) entre les arbres produits à partir des marqueurs différents. Cette contrainte paraît trop forte dans notre cas et non appropriée à la réalité biologique. En effet, il suffit que deux regroupements ne diffèrent que par la présence ou l'absence d'un individu (mal classé quelle qu'en soit la raison) pour que l'arête ne soit pas conservée dans l'arbre consensus.

A l'opposé, la construction des arbres communs prend en compte le fait que tous les individus puissent ne pas être correctement représentés dans tous les arbres (différentes sources d'erreur possibles). Cette méthode permet de rechercher le plus grand sous-ensemble d'individus formant la même structure dans les arbres comparés.

Selon PERRIER *et al.* (1999), au seuil de 5 % et avec un effectif d'environ cent individus, un maximum de 19 points communs entre deux arbres peut être obtenu par le fait du hasard. Dans notre cas, les trois arbres ainsi construits ont entre 19 et 24 points communs (figures 6 et 7). Ce résultat est en accord avec les faibles coefficients de corrélation calculés entre les matrices de distances. Néanmoins, en considérant les structures communes aux arbres construits à partir des marqueurs morphologiques et RFLP (figure 6), on note que les arêtes menant aux caudatum (C2) et aux durra sont présentes dans les deux arbres ; il en va de même pour celles qui mènent à une partie des guinea. Il peut paraître étonnant de ne pas voir figurer le groupe kafir dans les structures communes. En fait, ce groupe est plus homogène sur le plan génétique que morphologique. Les individus les plus proches morphologiquement ne sont pas les plus semblables génétiquement.

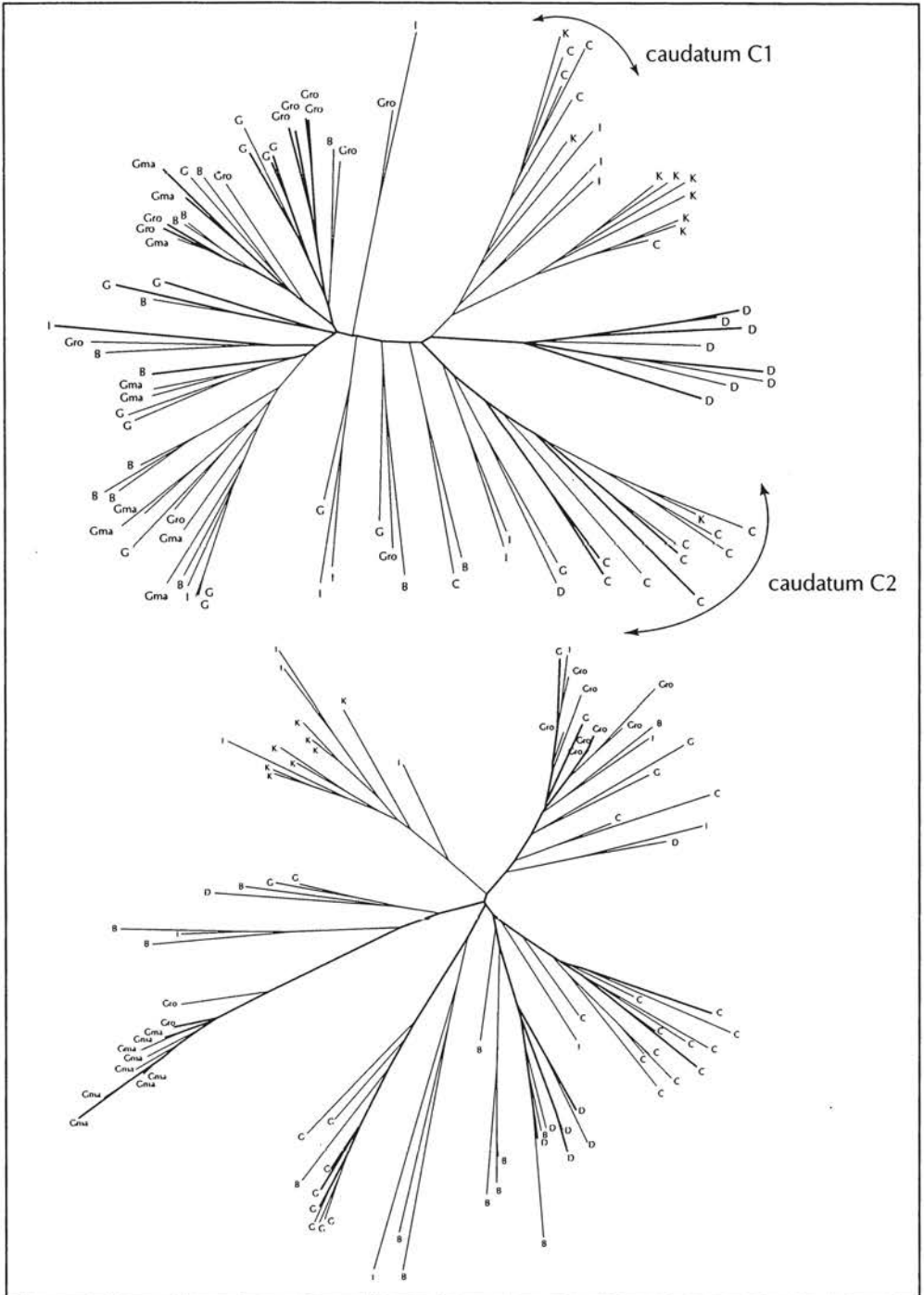


Figure 6. Représentation des arêtes communes (en gras) aux arbres construits à partir des données morphologiques (en haut) et moléculaires (en bas).

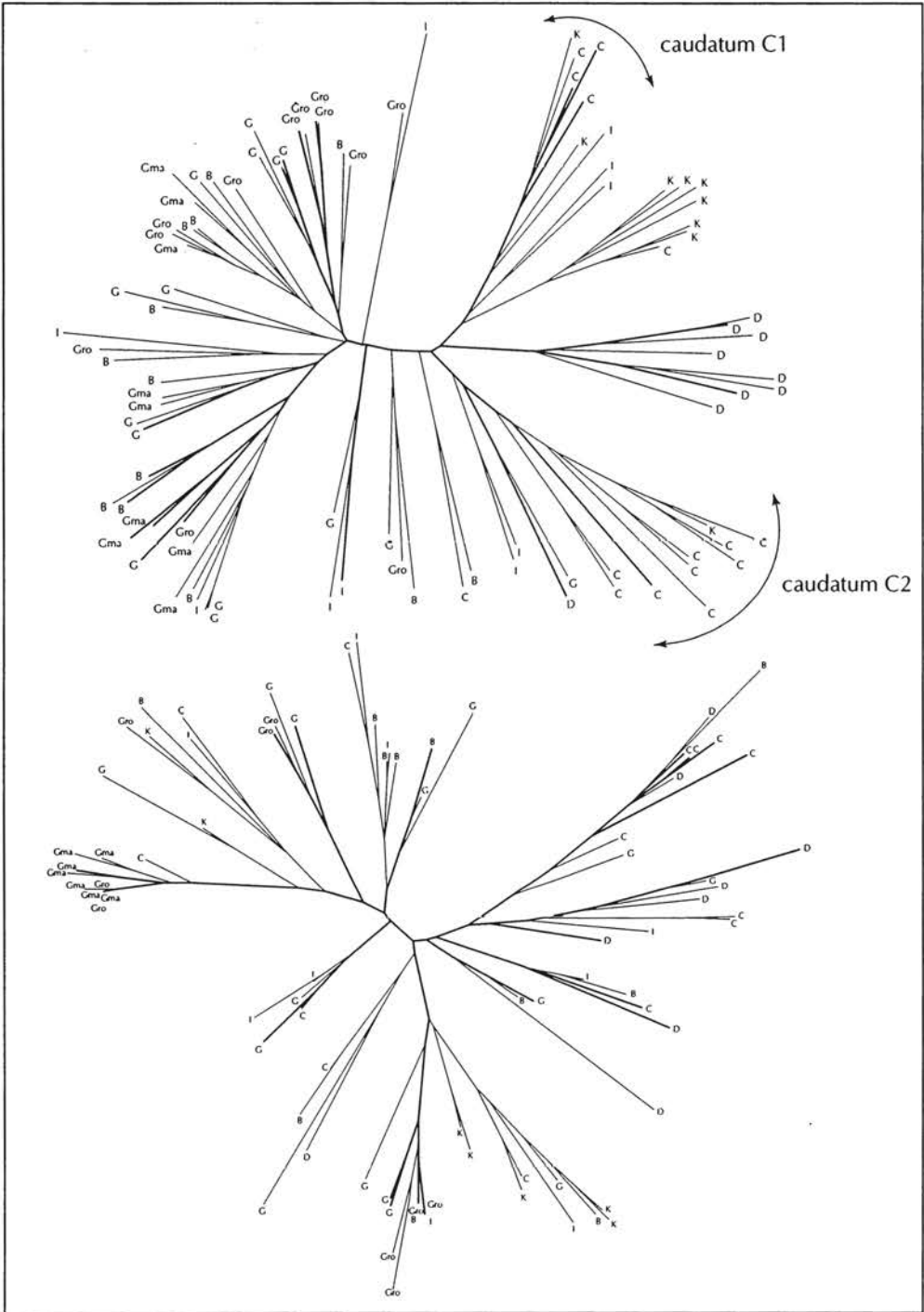


Figure 7. Représentation des arêtes communes (en gras) aux arbres construits à partir des données morphologiques (en haut) et enzymatiques (en bas).

Des observations similaires peuvent être faites lors de la comparaison des arbres construits à partir des marqueurs morphologiques et enzymatiques (figure 7). Cependant, si quelques grandes structures sont conservées (opposition entre guinea et caudatum, durra), ce ne sont pas les mêmes individus dans les deux arbres communs qui constituent l'ossature des structures communes. Cette observation rend compte du fait qu'il n'est pas possible de construire un arbre consensus.

## L'application à la constitution de *core collections*

Afin de faciliter l'utilisation des grandes collections, le concept de *core collection* a été élaboré par FRANKEL et BROWN (1984) puis par BROWN (1989). Différentes stratégies d'échantillonnage ont été proposées (HAMON *et al.*, 1995, pour revue). Elles ont été appliquées à plusieurs poacées, comme l'orge (VAN HINTUM *et al.*, 1990), le ray-grass (CHARMET *et al.*, 1993) et le blé dur (SPAGNOLETTI-ZEULLI et QUALSET, 1993).

Pour le sorgho, des collections de taille limitée ont été développées dès le début des années 90. Une collection de base a ainsi été constituée en Inde par l'ICRISAT. Elle comprend 1 400 accessions choisies selon des critères de taxonomie, de distribution géographique et d'adaptation écologique, tels que HARLAN (1972) les a définis. Très rapidement, cette collection a posé des problèmes d'utilisation en raison du photopériodisme, caractère qui n'avait pas été pris en compte lors de la constitution de la collection. En effet, dans le site de Patancheru, de nombreux écotypes tropicaux en provenance notamment du Cameroun, d'Ethiopie, du Nigeria et du Soudan ne fleurissent pas ou trop tardivement. Cette collection s'est donc avérée spécifique d'une localisation et peu représentative de la diversité phénotypique de la collection mondiale. Il a été alors décidé de créer d'autres collections de taille réduite.

Une *core collection* comportant uniquement des sorghos cultivés a donc été mise en place à l'ICRISAT par PRASADA RAO et RAMANATHA RAO (1995) avec deux objectifs : faciliter l'accès à la collection mondiale et représenter au mieux la diversité génétique des sorghos cultivés.

Pour cela, une stratification de la collection mondiale a été effectuée en sous-groupes définis selon des critères géographiques, taxonomiques et agronomiques. Puis, une analyse en composantes principales sur sept caractères quantitatifs — cycle semis-épiaison, hauteur de la plante, exsersion de l'inflorescence, longueur et largeur de l'inflorescence, recouvrement du grain et poids de 100 grains — a permis de définir d'autres groupes. En tenant compte des deux modes de structuration de la collection initiale, on a procédé à un échantillonnage proportionnel aux effectifs des groupes constitués. Cette *core collection* comporte 3 475 accessions, soit environ 10 % de la collection de l'ICRISAT.



En fait, le processus de stratification précédant la sélection vise à prendre en compte une éventuelle structuration (géographique ou taxonomique, par exemple) de la diversité génétique. La connaissance préalable de l'importance et de la structuration de la diversité génétique dans le pool génique considéré est indispensable pour maximiser la diversité génétique à préserver dans la *core collection*. Dans ce contexte, les sept caractères morphologiques pris en compte pour la constitution de la *core collection* sont-ils suffisants et pertinents pour structurer la diversité globale des sorghos ? Les différents travaux réalisés sur la diversité génétique des sorghos, y compris notre étude, indiquent que divergence phénotypique et divergence génétique ne sont pas totalement concordantes.

A l'aide de la stratégie PCS (*principal component score*) mise au point par NOIROT *et al.* (1996) et HAMON *et al.* (1998), nous avons constitué, à partir de l'échantillon initial de 230 accessions, deux *cores* en utilisant soit les données morphologiques, soit les données enzymatiques. Ces deux *core collections* ont été comparées à l'échantillon initial afin d'appréhender les modifications induites par le processus de sélection.

Les courbes représentant le nombre cumulé d'individus contribuant à maximiser la diversité en fonction du pourcentage cumulé de la diversité totale montrent que 50 % de la diversité enzymatique totale est représentée par 50 accessions, tandis que 77 accessions sont nécessaires pour obtenir 50 % de la diversité phénotypique. Ces deux lots d'accessions ont constitué respectivement la *coreL* et la *coreM*.

Sur le plan de la composition en races botaniques (tableau 1), le test du  $\chi^2$  indique des différences significatives entre l'échantillon initial et les deux *core collections*. On note en particulier que la race kafir n'est pas représentée dans la *coreL*. Cette observation est en accord avec la faible diversité génétique de la race kafir, notée par différents auteurs. En revanche, les grandes régions géographiques sont toutes représentées dans les deux *core collections*, sans qu'il y ait de modifications significatives par rapport à l'échantillon initial (tableau 2).

Pour les fréquences alléliques, comme attendu, des différences très significatives (test du  $\chi^2$ ,  $P < 0,01$ ) sont observées pour la *coreL*, alors qu'elles ne sont pas significatives pour la *coreM*. Cependant, dans la *coreL*, la divergence globale est due à deux locus — endopeptidase et phosphatase acide C — affectés par la sélection opérée. Dans la *coreM*, la sélection sur les caractères morphologiques n'a pas permis de conserver tous les allèles rares : trois allèles sur les cinq de fréquence inférieure à 5 % sont absents (tableau 3). Toutefois, on peut se demander quel serait le devenir de ces allèles rares soumis à la sélection naturelle.

Pour la diversité phénotypique, la situation est logiquement inversée : les différences ne sont pas significatives avec la *coreL*, mais sont très significatives avec la *coreM*. Dans ce cas, la divergence globale est due à des différences significatives (au moins au seuil de 5 %) pour 12 caractères sur les 21 étudiés.

Tableau 1. Répartition raciale dans l'échantillon initial et dans les deux *core collections* constituées par la méthode de PCS (effectif et pourcentage des différentes races).

	bicolor	caudatum	durra	guinea	kafir	autres
Echantillon initial	25 (10,9)	41 (17,8)	27 (11,7)	88 (38,3)	18 (7,8)	31 (13,5)
Corel	4 (8)	5 (10)	10 (20)	25 (50)	0	6 (12)
CoreM	13 (16,9)	6 (7,8)	16 (20,7)	32 (41,6)	2 (2,6)	8 (10,4)

$\chi^2_{\text{coreM}} = 12,4$  (ddl = 4), significatif à 5 %.

$\chi^2_{\text{corel}} = 9,53$  (ddl = 3), significatif à 5 %.

Tableau 2. Distribution géographique des accessions dans l'échantillon initial et dans les deux *core collections* (effectif et pourcentage des accessions).

	Afrique Ouest	Afrique Est	Afrique Centre	Afrique australe	Asie	Amérique
Echantillon initial	74 (32,2)	19 (8,3)	39 (16,9)	55 (23,9)	39 (16,9)	4 (0,02)
Corel	21 (42)	5 (10)	7 (14)	9 (18)	7 (14)	1 (2)
CoreM	26 (33,8)	7 (9)	10 (13)	14 (18,2)	18 (23,4)	2 (2,6)

$\chi^2_{\text{coreM}} = 3,17$  (ddl = 4), ns.

$\chi^2_{\text{corel}} = 9,5$  (ddl = 3), ns.

Tableau 3. Nombre total d'allèles des différentes catégories de fréquences dans l'échantillon initial et dans les deux *core collections*.

Fréquence des allèles (x)	Nombre total d'allèles des différentes catégories de fréquences présents dans		
	l'échantillon initial	la corel	la coreM
$x < 0,05$	5	5	2
$0,05 \leq x < 0,1$	2	2	2
$0,1 \leq x < 0,2$	3	3	3
$0,2 \leq x < 0,4$	9	9	9
$x > 0,4$	12	12	12

Ces comparaisons indiquent que la sélection sur marqueurs neutres, enzymatiques, (corel) est intéressante car elle permet de conserver la diversité allélique tout en préservant la diversité phénotypique initiale. Cependant, les grandes collections ont fait et font l'objet d'évaluations morphoagronomiques plus ou moins systématiques. Il est donc aussi important de considérer ces descripteurs pour constituer des *core collections*. Dans cette optique, on peut se demander si toutes les informations disponibles dans la collection initiale sont utiles à prendre en compte. En d'autres termes, n'y a-t-il pas redondance d'information ? Est-il possible de définir des associations entre marqueurs ? Un élément de réponse peut être fourni par l'utilisation du coefficient d'association  $V$  de Cramer, calculé pour chaque couple de variables (BISHOP *et al.*, 1975). Les coefficients calculés à partir des données de l'échantillon initial (230 accessions) ont des valeurs comprises entre  $-0,3$  et  $+0,65$  (ce coefficient varie théoriquement entre  $-1$  et  $+1$ ). En prenant pour limite inférieure significative d'association la valeur de  $0,5$ , des associations sont relevées pour cinq ensembles de variables : couche brune (Cbr) et vitrosité (Vit) ; largeur de la feuille (Laf) et diamètre de la tige (Dtp) ; délai d'épiaison (Nje) et nombre d'entre-nœuds (Nen) ; compacité de la panicule (Cpa) et longueur de la panicule (Lpa) et du pédoncule (Lpe) ; forme du grain (Fgr) et couche brune (Cbr) ainsi que longueur et forme du pédoncule (Lpe, Fpe).

Etant donné que cet échantillon initial ne correspond pas à un prélèvement aléatoire dans la collection mondiale, on peut se demander si les associations mises en évidence sont fortuites (biais d'échantillonnage), si elles sont le reflet de déséquilibres de liaison (effet fondateur, dérive génétique), si elles sont liées à l'organisation génétique du génome ou si elles représentent des complexes de gènes coadaptés.

La cartographie génomique des différents caractères peut fournir des éléments de réflexion bien que l'analyse d'une descendance ne soit pas comparable à celle d'une population.

Récemment, certains de ces caractères ont été cartographiés (RAMI *et al.*, 1998). Il est intéressant de noter que le gène *B2/b2* de la couche brune et un QTL majeur de la vitrosité (Vit,  $r^2 = 54\%$ ) ont été colocalisés sur le groupe de liaison F. De même, deux QTL pour la compacité et la longueur de la panicule ont une colocalisation sur le groupe de liaison A (Cpa,  $r^2 = 22\%$ , et Lpa,  $r^2 = 35\%$ ) et sur le groupe de liaison F (Cpa,  $r^2 = 13,5\%$ , et Lpa,  $r^2 = 20\%$ ). Notons que ces QTL situés sur le groupe de liaison F sont indépendants génétiquement du couple *B2/b2*-Vit.

En accord avec les résultats obtenus sur nos deux *core collections*, aucune association (coefficient  $V \geq 0,5$ ) n'a été notée entre marqueurs morphologiques et marqueurs enzymatiques. De même, aucune association n'a été relevée entre les marqueurs enzymatiques au seuil fixé. Malgré l'absence de grands groupes de linkats entre les locus enzymatiques (OLLITRAULT *et al.*, 1989a), deux situations sont intéressantes à considérer : les locus *LAP-2* et *PA-B*, dis-

tants de 12 centimorgans, qui présentent un coefficient  $V$  de Cramer de 0,24 ; les couples de locus *HEX* et *LAP*, *DIA* et *LAP*, *AMY* et *EST-C*, avec un coefficient d'association de 0,45, 0,36 et 0,34 respectivement, et indépendants génétiquement. En considérant la coreL et la coreM, on constate que, pour les couples *AMY* - *EST-C* et *HEX* - *LAP*, les associations sont maintenues et ne sont pas dues au hasard (test exact de Fisher significatif pour coreL, très significatif pour coreM, dans les deux cas). Pour les couples *DIA* - *LAP* et *PAB* - *LAP*, à l'inverse, les associations sont rompues dans les deux *core collections* (test exact de Fisher non significatif).

La cartographie de l'ensemble des caractères morphologiques et enzymatiques ainsi que la recherche d'associations incluant les marqueurs moléculaires sont à approfondir. Elles devraient fournir des informations intéressantes pour mieux préciser le jeu de caractères à prendre en compte dans la constitution de *core collections* et pour faire la part entre structures coadaptées à inclure dans la *core collection* et déséquilibres de liaisons.

## Conclusion

Les études portant sur la diversité génétique montrent tout d'abord qu'il existe une forte variabilité chez les sorghos cultivés. Onze locus enzymatiques sont polymorphes dans notre étude, 13 dans celle de MORDEN *et al.* (1989) et 18 dans celle de OLLITRAULT (1989b). Près de 75 % des sondes hétérologues de maïs, en combinaison avec deux enzymes de restriction, s'hybrident avec de l'ADN de sorgho et révèlent du polymorphisme (DEU *et al.*, 1994). Dans cette étude, chaque locus (enzymatique ou révélé par RFLP) est représenté en moyenne par 2,8 à 3 allèles.

Les marqueurs moléculaires RFLP permettent de discriminer quatre des cinq races botaniques majeures décrites par HARLAN et DE WET (1972). Ces races présentent une variabilité génétique plus ou moins importante et structurée. Ainsi, la race kafir apparaît très homogène, alors qu'on observe une forte structuration à l'intérieur des guinea, en partie reflet de centres de diversification secondaires.

Les structurations observées à l'aide des marqueurs enzymatiques et morphologiques ne coïncident pas parfaitement avec la classification raciale. Certaines races constituent des ensembles qui traduisent des comportements en culture semblables (bicolor et guinea d'une part, caudatum et kafir d'autre part, avec les descripteurs morphologiques) ou une aire d'origine commune (caudatum et durra, avec les marqueurs enzymatiques).

De manière générale, on considère que les marqueurs moléculaires et enzymatiques sont neutres à l'égard de la sélection. Les différences dans la structuration de la diversité génétique révélées par les deux types de marqueurs

résident-elles dans le nombre de locus considérés, dans leur localisation génomique ou dans leur spécificité propre ? L'étude de la diversité génétique révélée par l'analyse d'un grand nombre de locus RFLP très régulièrement répartis sur le génome du sorgho devrait apporter des réponses à ces questions.

Il n'a pas été possible de construire un arbre consensus résultant de la prise en compte des différents caractères morphologiques et enzymatiques et ce, bien que l'on puisse observer quelques éléments de structure communs. Ce résultat indique l'absence d'une compartimentation forte au sein des sorghos cultivés. Ni l'appartenance raciale, ni l'aire d'origine, ni le comportement en culture n'ont conduit à la mise en place de fortes barrières à la reproduction. La présence d'individus en position intermédiaire dans les représentations arborées indique l'existence de flux de gènes. Ainsi, dans les conditions naturelles, des brassages génétiques se produisent et contribuent, même si leur taux est faible, à élargir localement la diversité génétique. Dans le même temps, de fortes pressions de sélection sont exercées ; elles permettent de maintenir des races majeures là où celles-ci offrent un intérêt particulier. Cette gestion dynamique traditionnelle est en faveur de la mise en place de sites privilégiés pour la conservation *in situ* et participative des ressources génétiques du sorgho. Cet aspect est d'autant plus important à considérer que, comme pour d'autres plantes, lors de la multiplication du matériel végétal, se pose le problème de la dérive génétique liée à l'autofécondation de populations pas toujours fixées, bien que préférentiellement autogames. Cette absence de compartimentation nous permet aussi de considérer les sorghos cultivés comme un seul pool génétique.

La stratégie PCS a donc été appliquée directement sur l'échantillon initial (sans stratification préalable) pour constituer des *core collections*, en considérant les caractères morphologiques puis les caractères enzymatiques. L'examen des deux *core collections* ainsi constituées montre que 22 % des accessions, pour la coreI, et 33,4 %, pour la coreM, sont nécessaires pour représenter 50 % de la diversité totale initiale. La diversité morphologique apparaît donc plus éclatée que la diversité enzymatique. Cependant, cette comparaison reste très limitée dans la mesure où le nombre de locus considérés est différent dans chaque cas. Il aurait été particulièrement intéressant de disposer des données RFLP afin d'avoir une autre mesure de la répartition de la diversité génétique dans l'échantillon initial.

En outre, l'examen des deux *core collections* indique qu'il n'y a pas d'association entre les deux types de caractères. Dans notre étude, la sélection sur les caractères enzymatiques n'a pas affecté la variabilité morphologique initiale. La sélection sur les caractères morphologiques conduit à la perte de quelques allèles rares à certains locus enzymatiques. Il serait donc intéressant de pouvoir considérer simultanément les deux types de caractères pour constituer des *core collections*. Dans ces conditions, on pourrait envisager une sélection morphologique sur les groupes enzymatiques préalablement obtenus ou, à l'inverse, une sélection enzymatique sur les groupes morphologiques. En

revanche, quelques associations au sein des caractères morphologiques et des caractères enzymatiques ont été détectées, ce qui pourrait indiquer une certaine redondance des informations. Les investigations devraient être poursuivies pour mieux comprendre l'origine de ces associations et définir des jeux de caractères informatifs. La cartographie génomique fournit des pistes de réflexion et devrait permettre d'optimiser les outils d'évaluation de la diversité. Les marqueurs moléculaires sont d'un grand intérêt chez le sorgho car ils sont susceptibles d'être cartographiés (les locus enzymatiques sont presque tous indépendants génétiquement), ils permettent, en outre, d'étudier le déterminisme des différents caractères quel que soit leur degré de complexité. Les marqueurs moléculaires et la cartographie génomique sont des outils particulièrement prometteurs pour la gestion des ressources génétiques.

Dans le cas du sorgho, la cartographie d'un bon nombre de caractères agromorphologiques a été réalisée ou est en cours (LIN *et al.*, 1995 ; PEREIRA *et al.*, 1995 ; PEREIRA et LEE, 1995 ; RAMI *et al.*, 1998). L'accès à des informations d'ordre génomique devrait favoriser la « lisibilité » des ressources génétiques.

## Annexe

### Matériel végétal

L'échantillon de sorghos cultivés (*S. bicolor* subsp. *bicolor*) initialement utilisé comprend 230 accessions, dont 136 ont été étudiées par CHANTEREAU *et al.* (1989) et OLLITRAULT *et al.* (1989b) et 63, principalement de race guinea, par DEGREMONT (1992). Ces sorghos sont des variétés traditionnelles issues des collections de l'ICRISAT ou du Cirad. Les données morphologiques et enzymatiques sont disponibles pour les 230 accessions.

L'analyse par les RFLP a porté sur 92 accessions, choisies parmi les 230 selon deux critères : l'origine géographique et la classification raciale. Soixante-quatorze font partie de l'échantillon analysé par DEU *et al.* (1994) ; les 18 autres appartiennent en majorité à la race guinea.

### Etude enzymatique

L'analyse a porté sur 8 systèmes enzymatiques révélant 11 locus polymorphes : alcool déshydrogénase (ADH), amylase (AMY), diaphorase (DIA), endopeptidase (END), estérase (EST), hexokinase (HEX), leucine aminopeptidase (LAP) et phosphatase acide (PA).

Les protocoles expérimentaux ainsi que les interprétations génétiques des zymogrammes sont décrits par OLLITRAULT *et al.* (1989a) et DEGREMONT (1992).

### Etude morphologique

Les dispositifs expérimentaux sont décrits par CHANTEREAU *et al.* (1989) et DEGREMONT (1992).

L'analyse a porté sur 21 caractères morphologiques communs à ces deux études. Les descripteurs signalés par un astérisque appartiennent à la liste recommandée par l'IBPGR. Dix variables qualitatives ont été retenues : anthocyane des feuilles (Ant)\*, aristation (Ari)\*, couche brune du grain (Cbr)\*, couleur du grain (Cgr)\*, compacité de la panicule (Cpa)\*, forme du grain (Fgr)\*, forme du pédoncule (Fpe), longueur des glumes (Log)\*, ouverture des glumes (Oug) et vitrosité du grain (Vit)\*. Onze variables quantitatives ont été mesurées sur la tige principale : diamètre de la tige (Dtp), hauteur de la tige (Htp)\*, longueur de la panicule (Lpa)\*, largeur et longueur de la 3<sup>e</sup> feuille sous-panículaire (Laf et Lof), longueur du pédoncule (Lpe), nombre d'entre-nœuds (Nen), nombre de jours entre le semis et 50 % d'épiaison (Nje)\*, nombre de talles utiles (Ntu)\*, poids de grains par panicule (Pgp) et poids de 500 grains (P5g)\*.

### Etude par RFLP

Trente-et-une sondes génomiques de maïs, qui correspondent à 50 combinaisons sonde-enzyme révélant du polymorphisme, ont été utilisées. Ces sondes, réparties sur l'ensemble du génome, représentent des séquences uniques ou très faiblement répétées. Les combinaisons sonde-enzyme sont identiques à celles qui ont été décrites par DEU *et al.* (1994), à deux exceptions près : la combinaison Umc 38-*Hind*III a été éliminée, tandis que le couple Bnl 7.49-*Xba*I a été ajouté.

## Traitement des données

Pour les données enzymatiques et RFLP, l'analyse factorielle des correspondances (BENZECRI, 1973) a été réalisée sur un tableau disjonctif complet dans lequel chaque allèle ou chaque bande est codé par deux variables : présence et absence. Pour les caractères morphologiques, afin de traiter conjointement les caractères qualitatifs et les caractères quantitatifs, ces derniers ont été transformés en variables qualitatives par codage en 2 à 4 classes, selon la distribution de chaque caractère. Ces analyses multivariées ont été effectuées à l'aide du logiciel Addad.

Les distances entre individus ont été calculées en utilisant l'indice de similarité de Dice (1945), pour les données enzymatiques et RFLP, et l'indice de Sokal et Michener (ou *simple matching*), pour les données morphologiques qualitatives. Les matrices de similarités ont été calculées et comparées selon le test de Mantel à l'aide du logiciel Ntsys-pc version 1.80. Des classifications ascendantes hiérarchiques ont été réalisées pour les trois types de données en utilisant comme critère d'agrégation le *neighbor-joining*, développé par SAITOU et NEI (1987) et mis en œuvre dans le logiciel Darwin, créé par le Cirad. Les constructions d'arbres consensus et d'arbres communs minimaux (PERRIER *et al.*, 1999) ont été effectuées grâce au logiciel Darwin.

Les coefficients d'association  $V$  de Cramer (BISHOP *et al.*, 1975) ont été calculés à l'aide du logiciel Sas.

La méthode PCS (NOIROT *et al.*, 1996) a été appliquée à la collection de 230 accessions en utilisant séparément les données enzymatiques et les données morphologiques. La sélection sur les données enzymatiques a permis de définir la corel, celle sur les données morphologiques, la coreM.



## Références bibliographiques

- ALDRICH P.R., DOEBLEY J., 1992. Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. *Theoretical and Applied Genetics*, 85 : 293-302.
- ALDRICH P.R., DOEBLEY J., SCHERTZ K.F., STEC A., 1992. Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. *Theoretical and Applied Genetics*, 85 : 451-460.
- APPA RAO S., PRASADA RAO K.E., MENGESHA M.H., GOPAL-REDDY V., 1996. Morphological diversity in sorghum germplasm from India. *Genetic Resources and Crop Evolution*, 43 : 559-567.
- ARRIOLA P.E., 1995. Crop to weed gene flow in *Sorghum*: implications for transgenic release in Africa. *African Crop Science Journal*, 3(2) : 153-160.
- BENZECRI J.P., 1973. L'analyse des données. Tome II. L'analyse des correspondances. Paris, France, Dunod, 616 p.
- BISHOP Y.M.M., FIENBERG S.E., HOLLAND P.W., 1975. Discrete multivariate analysis: Theory and practice. Cambridge, Etats-Unis, Massachusetts Institute of Technology Press.
- BROWN A.H.D., 1989. Core collections: a practical approach to genetic resources management. *Genome*, 31 : 817-824.
- CHANTEREAU J., ARNAUD M., OLLITRAULT P., NABAYAOGO P., NOYER J.L., 1989. Etude de la diversité morphophysio-logique et classification des sorghos cultivés. *L'Agronomie tropicale*, 44(3) : 223-232.
- CHANTEREAU J., KONDOMBO C., 1994. Estimation des taux d'allogamie chez les sorghos de la race guinea. *In* : Progress in food grain research and production in semi-arid Africa, J.M. Menyonga *et al.* éd., Ouagadougou, Burkina, SAFGRAD, p. 309-314.
- CHANTEREAU J., VAKSMANN M., BAHMANI I., AG-HAMADA M., CHARTIER M., BONHOMME R., 1997. Characterization of different temperature and photoperiod responses in African sorghum cultivars. *In* : Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre, A. Ratnadas *et al.* éd., Montpellier, France, Cirad, p. 29-35.
- CHARMET G., BALFOURIER F., RAVEL C., DENIS J.B., 1993. Genotype  $\times$  environment interaction in a core collection of French perennial ryegrass populations. *Theoretical and Applied Genetics*, 86 : 731-736.
- CUI Y.X., XU G.W., MAGILL C.W., SCHERTZ K.F., HART G.E., 1995. RFLP-based assay of *Sorghum bicolor* (L.) Moench genetic diversity. *Theoretical and Applied Genetics*, 90 : 787-796.
- DAHLBERG J.A., SPINKS M.S., 1995. Current status of the US sorghum germplasm collection. *International Sorghum and Millets Newsletter*, 36 : 4-12.
- DEGREMONT I., 1992. Evaluation de la diversité génétique et du comportement en croisement des sorghos (*Sorghum bicolor* L. Moench) de race guinea au moyen de marqueurs enzymatiques et morphophysio-logiques. Thèse de doctorat, université Paris XI, Orsay, France, 191 p.

- DEU M., GONZALEZ DE LEON D., GLASZMANN J.C., DEGREMONT I., CHANTEREAU J., LANAUD C., HAMON P., 1994. RFLP diversity in cultivated sorghum in relation to racial differentiation. *Theoretical and Applied Genetics*, 88 : 838-844.
- DEU M., HAMON P., CHANTEREAU J., DUFOUR P., D'HONT A., LANAUD C., 1995. Mitochondrial DNA diversity in wild and cultivated sorghum. *Genome*, 38 : 635-645.
- DICE L.R., 1945. Measures of the amount of ecologic association between species. *Ecology*, 26 : 297-302.
- DOGGETT H., 1988. *Sorghum* (2nd ed.). Londres, Royaume-Uni, Longman, 512 p.
- ENGQVIST G.M., BECKER H.C., 1994. Genetic diversity for allozymes, RFLPs and RAPDs in resynthesized rape. *In* : *Biometrics in plant breeding: applications of molecular markers: IXth meeting of the EUCARPIA section biometrics in plant breeding*, J. Jansen éd., Wageningen, Pays-Bas.
- FAO, 1995. *Production yearbook: 1994* (volume 47). Rome, Italie, FAO, 243 p.
- FRANKEL O.H., BROWN A.H.D., 1984. Current plant genetic resources: a critical appraisal. *In* : *Genetics, new frontiers* (volume IV). New Delhi, Inde, Oxford and IBH, p. 1-11.
- HAMON S., DUSSERT J., DEU M., HAMON P., SEGUIN M., GLASZMANN J.C., GRIVET L., CHANTEREAU J., CHEVALLIER M.H., FLORI A., LASHERMES P., LEGNATE H., NOIROT M., 1998. Effects of quantitative and qualitative principal component score strategies on the structure of coffee, rice, rubber tree and sorghum core collections. *Genetics, Selection, Evolution*, 30 (suppl. 1) : 237-258.
- HAMON S., HODGKIN T., DUSSERT S., NOIROT M., 1995. Core collection: accomplishments and challenges. *Plant Breeding Abstracts*, 65(8) : 1125-1133.
- HARLAN J.R., 1972. Genetic resources in sorghum. *In* : *Sorghum in the seventies*, N.G.P. Rao et L.R. House éd., New Delhi, Inde, Oxford and IBH.
- HARLAN J.R., STEMLER A., 1976. The races of sorghum in Africa. *In* : *Origins of African plant domestication*, J.R. Harlan *et al.* éd., La Haye, Pays-Bas, Mouton, p. 465-478.
- HARLAN J.R., DE WET J.M.J., 1972. A simplified classification of cultivated sorghum. *Crop Science*, 12 : 172-176.
- VAN HINTUM T., VON BOTHMER R., FISCHBECK G., KNÜPFER H., 1990. The establishment of the barley core collection. *Barley Newsletter*, 34 : 41-42.
- IBPGR, 1993. *Descriptors for sorghum, Sorghum bicolor* (L.) Moench. Rome, Italie, IBPGR-ICRISAT, 38 p.
- LIN Y.R., SCHERTZ K.F., PATERSON A.H., 1995. Comparative analysis of QTL affecting plant height and maturity across the Poaceae in reference to an interspecific sorghum population. *Genetics*, 141 : 391-411.
- MANN J.A., KIMBER C.T., MILLER F.R., 1983. *The origin and early cultivation of sorghums in Africa*. College Station, Etats-Unis, Texas A and M University Press, Texas Agricultural Experiment Station Bulletin n° 1454.

MENGESHA M.H., APPA RAO S., 1994. Management of plant genetic resources at ICRISAT. *In* : Evaluating ICRISAT research impact (IAC). Patancheru, Inde, ICRISAT, p. 11-14.

MENKIR A., GOLDSBROUGH P., EJETA G., 1997. RAPD-based assessment of genetic diversity in cultivated races of sorghum. *Crop Science*, 37(2) : 564-569.

MESSMER M.M., MELCHINGER A.E., LEE M., WOODMAN W.L., LEE E.A., LAMKEY K.R., 1991. Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. *Theoretical and Applied Genetics*, 83 : 97-107.

MORDEN C.W., DOEBLEY J., SCHERTZ K.F., 1989. Allozyme variation in Old World races of *Sorghum bicolor* (Poaceae). *American Journal of Botany*, 76(2) : 247-255.

NOIROT M., HAMON S., ANTHONY F., 1996. The principal component scoring: a new method of constituting a core collection using quantitative data. *Genetic Resources and Crop Evolution*, 43 : 1-6.

DE OLIVEIRA A.C., RICHTER T., BENNETZEN J.L., 1996. Regional and racial specificities in sorghum germplasm assessed with DNA markers. *Genome*, 39(3) : 579-587.

OLLITRAULT P., 1987. Evaluation génétique des sorghos cultivés, *Sorghum bicolor* (L.) Moench, par l'analyse conjointe des diversités enzymatique et morphophysiologique : relations avec les sorghos sauvages. Thèse de doctorat, université Paris XI, Orsay, France, 187 p.

OLLITRAULT P., ARNAUD M., CHANTEREAU J., 1989b. Polymorphisme enzymatique des sorghos. 2. Organisation génétique et évolutive des sorghos cultivés. *L'Agronomie tropicale*, 44(3) : 211-222.

OLLITRAULT P., ESCOUTE J., NOYER J.L., 1989a. Polymorphisme enzymatique des sorghos. 1. Description de 11 systèmes enzymatiques, déterminisme et liaisons génétiques. *L'Agronomie tropicale*, 44(3) : 203-210.

PEREIRA M.G., AHNERT D., LEE M., KLIER K., 1995. Genetic mapping of quantitative trait loci for panicle characteristics and seed weight in sorghum. *Revista Brasileira de Genética*, 18(2) : 249-257.

PEREIRA M.G., LEE M., 1995. Identification of genomic regions affecting plant height in sorghum and maize. *Theoretical and Applied Genetics*, 90 : 380-388.

PERRIER X., FLORI A., BONNOT F., 1999. Les méthodes d'analyse des données. *In* : Diversité génétique des plantes tropicales cultivées, P. Hamon *et al.* éd., Montpellier, France, Cirad, collection Repères, p. 43-46.

POWELL W., MORGANTE M., ANDRF C., HANAFEY M., VOGEL J., TINGEY S., RAFALSKI A., 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, 2 : 225-238.

PRASADA RAO K.E., RAMANATHA RAO V., 1995. The use of characterisation data in developing a core collection of sorghum. *In* : Core collections of plant genetic resources, T. Hodgkin *et al.* éd., Chichester, Royaume-Uni, Wiley, p. 109-116.

RAMI J.F., DUFOUR P., FLIEDEL G., MESTRES C., DAVRIEUX F., BLANCHARD P., HAMON P., 1998. Quantitative trait loci for grain quality, productivity, morphological and agrono-

mical traits in sorghum, *Sorghum bicolor* (L.) Moench. Theoretical and Applied Genetics, 97(4) : 605-616.

ROHLF F.J., 1990. Fitting curves to outlines. In : Proceedings of the Michigan morphometrics workshop, F.J. Rohlf et F.L. Bookstein éd., Ann Arbor, Etats-Unis, University of Michigan, Museum of Zoology Special Publication n° 2, p. 167-177.

SAITOU N., NEI M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4 : 406-425.

SNOWDEN J.D., 1936. The cultivated races of sorghum. Londres, Royaume-Uni, Adlard, 274 p.

SPAGNOLETTI-ZEULLI P., QUALSET C.O., 1993. Evaluation of five strategies for obtaining a core collection of durum wheat. Theoretical and Applied Genetics, 78 : 295-304.

TAO Y., MANNERS J.M., LUDLOW M.M., HENZELL R.G., 1993. DNA polymorphisms in grain sorghum, *Sorghum bicolor* (L.) Moench. Theoretical and Applied Genetics, 86(6) : 679-688.

TESHOME A., BAUM B.R., FAHRIG L., TORRANCE J.K., ARNASON T.J., LAMBERT J.D., 1997. Sorghum, *Sorghum bicolor* (L.) Moench, landrace variation and classification in North Shewa and South Welo, Ethiopia. Euphytica, 97 : 255-263.

THORMANN C.E., FERREIRA M.E., CAMARGO L.E.A., TIVANG J.G., OSBORN T.C., 1994. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. Theoretical and Applied Genetics, 88 : 973-980.

WENDORF F., CLOSE A.E., SCHILD R., WASYLIKOWA K., HOUSLEY R.A., HARLAN J.R., KROLIK H., 1992. Saharan exploitation of plants 8000 years BP. Nature, 359 : 721-724.

DE WET J.M.J., 1978. Systematics and evolution of *Sorghum* sect. *Sorghum* (Graminae). American Journal of Botany, 65(4) : 477-484.

DE WET J.M.J., HARLAN J.R., PRICE E.G., 1976. Variability in *Sorghum bicolor*. In : Origins of African plant domestication, J.R. Harlan et al. éd., La Haye, Pays-Bas, Mouton, p. 453-463.

YAGOUA N.D., 1995. Caractérisation du sorgho pluvial, *Sorghum bicolor* (L.) Moench, de la zone soudanienne du Tchad. Montpellier, France, Cirad (document interne).



# Annexes

---



# Liste des abréviations

AAT	aspartate aminotransférase
ACM	analyse des correspondances multiples
ACoP	analyse en coordonnées principales
ACP	analyse en composantes principales
ADH	alcool déshydrogénase
ADN	acide désoxyribonucléique
ADNc	ADN complémentaire
ADNcp	ADN chloroplastique
ADNmt	ADN mitochondrial
ADNr	ADN ribosomique
AFLP™	<i>amplified fragment length polymorphism</i>
AFC	analyse factorielle des correspondances
AFM	analyse factorielle multiple
AMP	aminopeptidase
TDC	tableau disjonctif complet
AFTD	analyse factorielle sur tableau de distances
ARN	acide ribonucléique
CAH	classification ascendante hiérarchique
CAT	catalase
DAF	<i>DNA amplification fingerprinting</i>
EST	estérase
FISH	<i>fluorescent in situ hybridization</i>
GISH	<i>genomic in situ hybridization</i>
GOT	glutamate oxaloacétate transaminase
ICD/IDH	isocitrate déshydrogénase
ISSR	<i>inter-simple sequence repeat</i>
LAP	leucine aminopeptidase
MDH	malate déshydrogénase
PCR	<i>polymerase chain reaction</i>
PER	peroxydase
PGD	6-phosphogluconate déshydrogénase
PGI	phosphogluco-isomérase
PGM	phosphoglucomutase
PIC	<i>polymorphism information content</i>
QTL	<i>quantitative trait loci</i>
RAPD	<i>randomly amplified polymorphic DNA</i>
RFLP	<i>restriction fragment length polymorphism</i>
SKDH	shikimate déshydrogénase
SSR	<i>simple sequence repeat</i>
UPGMA	<i>unweighted pair group method with average</i>
VNRT	<i>variable number of tandem repeat</i>
WPGMA	<i>weighted pair group method using average</i>



## Adresses des auteurs

- Albar Laurence, IRD, laboratoire GeneTrop, BP 5045, 34032 Montpellier Cedex 1, France
- Anthony François, CATIE, Apartado 59, 7170 Turrialba, Costa Rica
- Baudouin Luc, Cirad, département des cultures pérennes, BP 5035, 34032 Montpellier Cedex 1, France
- Berthaud Julien, IRD, laboratoire GeneTrop, BP 5045, 34032 Montpellier Cedex 1, France
- Bezançon Gilles, IRD, laboratoire de génétique des plantes, BP 11416, Niamey, Niger
- Bonnot François, Cirad, département des cultures pérennes, BP 5035, 34032 Montpellier Cedex 1, France
- Bourdeix Roland, Cirad, département des cultures pérennes, CNRA, station Marc Delorme, 07 BP 13, Abidjan 07, Côte d'Ivoire
- Carreel Françoise, Cirad, département des productions fruitières et horticoles, station de Neufchâteau, Sainte-Marie, 97130 Capesterre-Belle-Eau, Guadeloupe
- Chantereau Jacques, Cirad, département des cultures annuelles, BP 5035, 34032 Montpellier Cedex 1, France
- Clément-Demange André, Cirad, département des cultures pérennes, BP 5035, 34032 Montpellier Cedex 1, France
- Colombo Carlos, Instituto Agrônomico, Caixa postal 28, 13001-970 Campinas, SP, Brésil
- Combes Marie-Christine, IRD, laboratoire GeneTrop, BP 5045, 34032 Montpellier Cedex 1, France
- Courtois Brigitte, IRRI, PO Box 933, Manille, Philippines
- Deu Monique, Cirad, département des cultures annuelles, BP 5035, 34032 Montpellier Cedex 1, France
- D'Hont Angélique, Cirad, département des cultures annuelles, BP 5035, 34032 Montpellier Cedex 1, France
- Dubois Cécile, Cirad, département des productions fruitières et horticoles, BP 5035, 34032 Montpellier Cedex 1, France
- Dussert Stéphane, IRD, laboratoire GeneTrop, BP 5045, 34032 Montpellier Cedex 1, France
- Flori Albert, Cirad, département des cultures pérennes, BP 5035, 34032 Montpellier Cedex 1, France
- Ghesquière Alain, IRD, laboratoire GeneTrop, BP 5045, 34032 Montpellier Cedex 1, France
- Glaszmann Jean Christophe, Cirad, département d'amélioration des méthodes pour l'innovation scientifique, BP 5035, 34032 Montpellier Cedex 1, France
- Grivet Laurent, Cirad, département d'amélioration des méthodes pour l'innovation scientifique, BP 5035, 34032 Montpellier Cedex 1, France
- Hamon Perla, université Montpellier III, Route de Mende, 34199 Montpellier Cedex 5, France
- Hamon Serge, IRD, laboratoire GeneTrop, BP 5045, 34032 Montpellier Cedex 1, France
- Horry Jean-Pierre, INIBAP, Parc scientifique Agropolis II, 34397 Montpellier Cedex 5, France
- Jacquemond Camille, INRA, station de recherche agronomique de San Giuliano, 20230 San Nicolao, France
- Jacquot Michel, Cirad, département des cultures annuelles, BP 5035, 34032 Montpellier Cedex 1, France

- Jannoo Nazeema, MSIRI, Mauritius Sugar Industry Research Institute, Réduit, Maurice
- Jenny Christophe, Cirad, département des productions fruitières et horticoles, station de Neufchâteau, Sainte-Marie, 97130 Capesterre-Belle-Eau, Guadeloupe
- Lanaud Claire, Cirad, département d'amélioration des méthodes pour l'innovation scientifique, BP 5035, 34032 Montpellier Cedex 1, France
- Lashermes Philippe, IRD, laboratoire GeneTrop, BP 5045, 34032 Montpellier Cedex 1, France
- Lebrun Patricia, Cirad, département des cultures pérennes, BP 5035, 34032 Montpellier Cedex 1, France
- Legnaté Hyacinthe, IDEFOR, département des plantes à latex, 01 BP 1536, Abidjan 01, Côte d'Ivoire
- Luce Claude, Cirad, département des cultures annuelles, BP 5035, 34032 Montpellier Cedex 1, France
- Luro François, INRA, station de recherche agronomique de San Giuliano, 20230 San Nicolao, France
- Montagnon Christophe, Cirad, département des cultures pérennes, 01 BP 6483, Abidjan 01, Côte d'Ivoire
- Motamayor Juan-Carlos, Cirad, département d'amélioration des méthodes pour l'innovation scientifique, BP 5035, 34032 Montpellier Cedex 1, France
- N'cho Yavo-Pierre, CNRA, station Marc Delorme, 07 BP 13, Abidjan 07, Côte d'Ivoire
- Noirot Michel, IRD, laboratoire GeneTrop, BP 5045, 34032 Montpellier Cedex 1, France
- Noyer Jean-Louis, Cirad, département d'amélioration des méthodes pour l'innovation scientifique, BP 5035, 34032 Montpellier Cedex 1, France
- Ollitrault Patrick, Cirad, département des productions fruitières et horticoles, station de Neufchâteau, Sainte-Marie, 97130 Capesterre-Belle-Eau, Guadeloupe
- Perrier Xavier, Cirad, département des productions fruitières et horticoles, BP 5035, 34032 Montpellier Cedex 1, France
- Raffaillac Jean-Pierre, IRD, Cirad, département d'amélioration des méthodes pour l'innovation scientifique, laboratoire Ecotrop, BP 5035, 34032 Montpellier Cedex 1, France
- Second Gérard, IRD, laboratoire GeneTrop, BP 5045, 34032 Montpellier Cedex 1, France
- Seguin Marc, Cirad, département des cultures pérennes, University of Kasetsart, Golden Jubilee Building, Bangkok 10900, Thaïlande
- Sounigo Olivier, Cirad, département des cultures pérennes, CRU, Sainte-Augustine, Trinité-et-Tobago
- Tézenas du Montcel Hugues, Cirad, département des productions fruitières et horticoles, BP 5035, 34032 Montpellier Cedex 1, France
- Tomekpe Kodjo, Cirad, département des productions fruitières et horticoles, CRBP, BP 832, Douala, Cameroun
- Trouslot Pierre, IRD, laboratoire GeneTrop, BP 5045, 34032 Montpellier Cedex 1, France

Fichier préparé par Nicolas Perrier, société 4P  
Imprimé pour vous par Books on Demand (Allemagne)

## LA COLLECTION REPÈRES

---

La collection Repères publie des résultats de recherche obtenus dans les domaines d'activité du Cirad : agriculture, forêt et élevage. Elle s'adresse aux chercheurs, aux enseignants et aux étudiants.

## DÉJÀ PARUS

---

Les ignames au Cameroun

*Roland Dumont, Perla Hamon,  
Christian Seignobos*

La culture *in vitro* de plantes tropicales

*Claude Teisson, éditeur scientifique*  
*In vitro culture of tropical plants*  
*Claude Teisson, editor*

L'élevage bovin en Guyane

*Michel Vivier, Bertrand Vissac,  
Gérard Matheron, éditeurs scientifiques*

Recherche agricole et innovation  
en Afrique tropicale

*Pierre-Marie Bosc, Ellen Hanak Freud*

Agricultural research and innovation  
in tropical Africa

*Pierre-Marie Bosc, Ellen Hanak Freud*

Forêts tropicales et cycle du carbone

*Bruno Locatelli*

Lutte intégrée contre les ravageurs  
des cultures pérennes tropicales

*Dominique Mariau, éditeur scientifique*

Le café de Coatepec

*Benoît Goud, Lyne Moriaux-Sallée,  
Bertrand Sallée*

Politiques publiques et agriculture

*Bruno Losch, Nancy Laudié,  
Frédéric Varlet, François Ruf*

L'amélioration des plantes tropicales

*André Charrier, Michel Jacquot,  
Serge Hamon, Dominique Nicolas,  
éditeurs scientifiques*

Filière et dialogues pour l'action,  
la méthode Cadiac

*Robin Bourgeois, Danilo Herrera*

Les agro-industries rurales

en Amérique latine  
*François Boucher, José Muchnik,  
éditeurs scientifiques*

Du Sahel à la forêt tropicale

*Roger Bertrand*

Les maladies des cultures pérennes tropicales

*Dominique Mariau, éditeur scientifique*

Integrated pest management of tropical  
perennial crops

*Dominique Mariau, editor*





Centre  
de coopération  
internationale  
en recherche  
agronomique  
pour le  
développement

Dans les années 60, la communauté scientifique prenait conscience des menaces que les perturbations des milieux naturels et l'expansion des variétés modernes faisaient peser sur les ressources génétiques de nombre d'espèces cultivées. Elle s'est alors mobilisée pour collecter ces ressources. C'est ainsi qu'une multitude de collections ont été rassemblées à travers le monde. Aujourd'hui, ces collections ont atteint une taille qui rend difficiles leur entretien et leur caractérisation. La question de leur gestion se pose de manière aiguë. Conservation, évaluation, utilisation des ressources génétiques doivent être repensées.

Pour répondre à ces préoccupations, Frankel et Brown ont introduit dans les années 80 le concept de *core collection* : un échantillon d'accessions issues d'une collection plus vaste et choisies pour représenter au mieux le spectre de diversité existant. Mais sur quels critères et avec quels outils constituer cet échantillon ?

Les caractères agronomiques, prioritaires pour le sélectionneur, sont parfois difficiles à évaluer et leur déterminisme génétique est souvent complexe. Les marqueurs génétiques moléculaires, qui n'ont aucune utilité directe, révèlent une structuration de la diversité, qui peut servir de base pour construire une *core collection*. Les relations entre ces deux niveaux de variabilité sont mal connues : les différents types de marqueurs moléculaires sont-ils équivalents ? de fortes structurations à l'échelle moléculaire sont-elles systématiquement associées à de fortes structurations pour les caractères agronomiques ? les structures à ces deux niveaux sont-elles alors concordantes ?

Les outils statistiques, qui permettent d'analyser les ressemblances entre les individus ou les populations, sont indispensables pour repérer une éventuelle structuration de la diversité. Mais quelle méthode est la mieux adaptée à chaque type de marqueur, quelle est la fiabilité de l'image qu'elle donne de la diversité et quelle signification biologique peut-on lui accorder ?

Cet ouvrage apporte des éléments de réponse à ces questions en partant de l'étude de la diversité génétique de onze plantes tropicales. Trois chapitres méthodologiques — sur le marquage biochimique et moléculaire, l'analyse des données et la constitution de *core collections* — viennent compléter ces études.

35,00 €

**Diffusion :** La Librairie du Cirad  
BP 5035  
34032 Montpellier Cedex 1  
France

