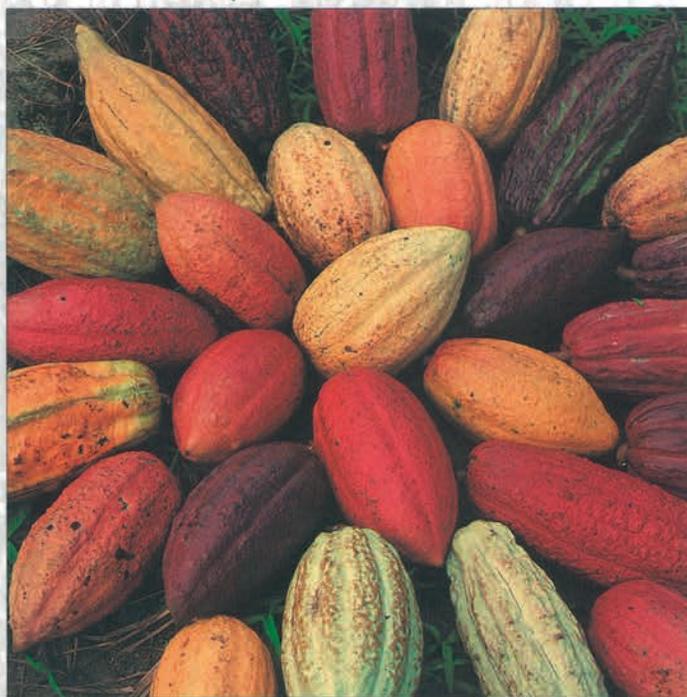


■ REPÈRES

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^e 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'introgession 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^e 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 contrôlées 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 α -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-