

Marqueurs moléculaires pour l'analyse des ressources génétiques et l'amélioration des plantes

Sylvain Santoni, Patricia Faivre-Rampant, Emilce Prado, Daniel Prat

En biologie végétale, le terme marqueur s'applique à plusieurs concepts, parfois très différents. Il est, par exemple, possible de définir des marqueurs physiologiques et des marqueurs génétiques. Les premiers correspondent à tout type de molécules, facilement repérables, dont la présence renseigne sur un stade de développement ou un état physiologique, situations complexes faisant intervenir de nombreux paramètres en interaction. Les marqueurs génétiques sont, quant à eux, toujours synonymes de locus marqueurs. Un locus marqueur est un locus polymorphe qui renseigne sur le génotype de l'individu qui le porte et/ou sur les génotypes des locus voisins. Les marqueurs génétiques sont par définition des caractères héréditaires.

D'après Bretting et Widrechner [1], les marqueurs moléculaires doivent réunir les caractéristiques idéales suivantes : être

des caractères mendéliens à hérédité simple ; avoir plusieurs allèles ; être codominants ; ne pas avoir d'effet pléiotropique ou épistatique ; être dispersés le long du génome ; ne pas être liés entre eux ; être insensibles au milieu ; être stables à tous les stades du développement ; ne pas avoir d'effet sur la croissance ou la reproduction sexuée ; être sélectivement neutres ; être facilement observables et sans ambiguïté.

Les marqueurs des génomes chloroplastiques et mitochondriaux ont des propriétés idéales sensiblement différentes du fait de la transmission et de l'organisation des génomes cytoplasmiques. Aucun de ces marqueurs n'a une hérédité mendélienne et tous les marqueurs d'un même génome chloroplastique ou mitochondrial sont liés, puisque portés par la même molécule d'ADN. La codominance, pas plus que la dominance, ne s'applique à ces génomes dont toutes les copies sont identiques.

Les marqueurs largement utilisés à ce jour concernent directement l'information portée par les acides nucléiques ou les produits de la traduction des gènes. On distingue les marqueurs biochimiques, issus de l'expression des gènes, y compris les produits du métabolisme secondaire, et les marqueurs moléculaires, directement issus du polymorphisme existant au niveau de l'ADN. Dans cet article, nous n'évoquerons que les marqueurs moléculaires concernant l'ADN en insistant sur ceux qui sont le plus couramment utilisés dans le domaine végétal. Nous avons choisi de séparer les marqueurs moléculaires et les tech-

niques qui permettent de les révéler en deux grands groupes, selon des critères génétiques plutôt que technologiques ou historiques. On peut considérer, d'une part, les techniques qui fournissent des marqueurs codominants et révélés individuellement et, d'autre part, celles qui révèlent « en masse » des marqueurs dominants. Cette séparation est certes simplificatrice mais elle correspond à deux types majeurs d'utilisation des marqueurs.

Marqueurs codominants révélés individuellement

Marqueurs révélés par la technique RFLP

Les enzymes de restriction coupent l'ADN en des sites spécifiques comprenant, en général, un nombre pair de bases (4, 6 ou 8). Une enzyme ayant un site de reconnaissance à 6 bases coupe l'ADN toutes les 4 096 bases en moyenne (4⁶). Un génome de 10⁹ bases peut donc produire environ 250 000 fragments de restriction de longueur variable. La spécificité est telle que le remplacement d'une seule base dans un site suffit à empêcher la coupure de l'ADN par l'enzyme utilisée. C'est cette spécificité qui est exploitée pour la mise

S. Santoni : INRA, Station de génétique et d'amélioration des plantes, Centre INRA de Montpellier, 2, place Viala, 34060 Montpellier cedex 01, France.

P. Faivre-Rampant : Université Henri-Poincaré Nancy I, Laboratoire de biologie forestière, BP 239, 54506 Vandœuvre-les-Nancy cedex, France.

E. Prado : Station de génétique et d'amélioration des plantes, Centre INRA de Dijon, Bretenières, 21110 Genlis, France.

D. Prat : INRA, Station d'amélioration des arbres forestiers, Centre INRA d'Orléans, BP 20619 Ardon, 45166 Olivet cedex, France.

Tirés à part : S. Santoni

en évidence du polymorphisme : une présence/absence de site de restriction entraîne un polymorphisme de longueur des fragments. Ce phénomène n'étant pas rare, la digestion de l'ADN génomique de deux individus quelconques dans une espèce donnée produit de très nombreuses différences de longueur de fragments. Mais, s'il est aisé d'obtenir des fragments polymorphes, leur visualisation est délicate lorsqu'il s'agit d'analyser des génomes complexes. En effet, l'électrophorèse sur gel des très nombreux produits de digestion de l'ADN génomique, suivie d'une révélation avec un réactif coloré (de type bromure d'éthidium, par exemple), génère une image de traînée continue (*smear*). Dans ce cas, la technique RFLP (*Restriction Fragment Length Polymorphism*, voir *glossaire*) correspond à l'identification de fragments particuliers, homologues entre plusieurs individus, et éventuellement alléliques grâce à l'utilisation de sondes moléculaires selon la technique de Southern [2].

La mise en œuvre de la technique RFLP implique de réaliser les manipulations suivantes : l'ADN génomique total extrait et purifié en quantité assez importante à partir de chaque génotype est soumis à une digestion par une enzyme de restriction. Les fragments générés sont séparés selon leur taille par une électrophorèse en gel. L'ADN est ensuite transféré du gel sur une membrane de nylon ; la position relative des fragments d'ADN est conservée durant le transfert. La dernière étape consiste en la réalisation de l'hybridation moléculaire avec une sonde d'ADN marquée préalablement soit par la radioactivi-

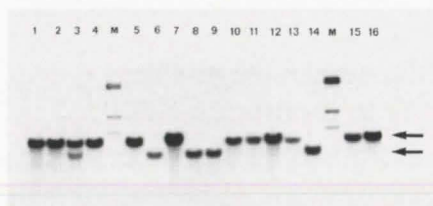


Figure 1. Exemple de résultat d'une analyse RFLP dans une population de lignées recombinantes de maïs. Les pistes 1 à 16 correspondent à différents individus de la descendance. M : marqueur de poids moléculaire. Tous les individus sont homozygotes au locus révélé par cette sonde sauf l'individu 3.

Figure 1. RFLP analysis of maize recombinant inbred lines. Lines 1 to 16 correspond to different genotypes of the segregating population. M: molecular weight marker. All genotypes except number 3 are homozygous at the locus revealed by the probe.

té, soit chimiquement. La sonde va alors s'hybrider avec le ou les fragments de l'ADN avec lesquels elle présente des homologies (*figure 1*). La membrane est un support solide qui peut être soumis à plusieurs cycles d'hybridation/déshybridation. Elle peut donc être utilisée avec plusieurs sondes successivement (jusqu'à une dizaine).

Les sondes moléculaires sont des fragments d'ADN isolés et individualisés par clonage ou par amplification PCR. Deux sources de sondes sont couramment utilisées : les sondes génomiques (ADNg) et les sondes d'ADN complémentaire (ADNc). Les premières peuvent correspondre à tous les types d'ADN, codant ou non, copie unique ou séquence répétée dans le génome, d'origine nucléaire ou cytoplasmique. Les secondes sont les copies ADN d'ARN messagers et correspondent donc uniquement à des séquences transcrites du génome. Pour être utilisée, une sonde doit permettre d'obtenir un profil RFLP lisible (*figure 2*) et éventuellement de révéler du polymorphisme [3].

Le polymorphisme révélé résulte d'événements de mutation au niveau des sites de restriction de l'enzyme utilisée ou d'insertion/délétion dans la séquence ciblée par la sonde ou à proximité de celle-ci. Les deux situations ne sont pas mutuellement exclusives. Une sonde révèle toutes les séquences qui lui sont homologues : elle permet donc, chez les plantes hétérozygotes, de visualiser les deux allèles simultanément,

pour peu que ceux-ci soient distinguables. On dit alors que la technique RFLP, lorsque sont utilisées des sondes qui correspondent à des ADN d'origine nucléaire, produit des marqueurs codominants.

Une sonde est dite homologue lorsqu'elle correspond à un fragment d'ADN isolé chez l'espèce analysée, et hétérologue lorsqu'elle provient du génome d'une espèce différente, plus ou moins proche phylogénétiquement. Le pourcentage de sondes hétérologues utilisables décroît à mesure qu'augmente l'éloignement génétique. À cet égard, les sondes de régions codantes, ADNc ou gènes, sont, *a priori* plus intéressantes car leur séquence évolue moins vite. Elles conservent donc mieux leur capacité d'hybridation que des sondes d'ADNg. Les sondes hétérologues permettent de marquer le génome d'une espèce sans avoir à refaire l'effort important d'isolement et de criblage initial. Elles offrent, de plus, la possibilité de comparer les cartes entre espèces proches.

La mise en œuvre de la technique RFLP demande des compétences en biologie moléculaire ainsi qu'une organisation et des matériels de laboratoire adaptés. Il est très difficile d'automatiser les différentes étapes de cette technique. Les laboratoires spécialisés ont optimisé et standardisé les protocoles et s'appuient sur des personnels très compétents. La quantité d'ADN requise par génotype à

Figure 2. Expérience de criblage de polymorphisme RFLP. Les ADN de trois lignées de maïs ont été digérés par les enzymes *Hind* III (pistes 1 à 3) et *Eco* RI (pistes 4 à 6) (M : marqueur de poids moléculaire). Les profils obtenus avec trois sondes montrent des situations différentes. **A** : le profil est simple, la sonde révèle une séquence unique dans le génome. **B** : le profil est multibandes, la sonde révèle plusieurs locus. **C** : la sonde produit un bruit de fond important car elle correspond à une séquence répétée de nombreuses fois dans le génome.

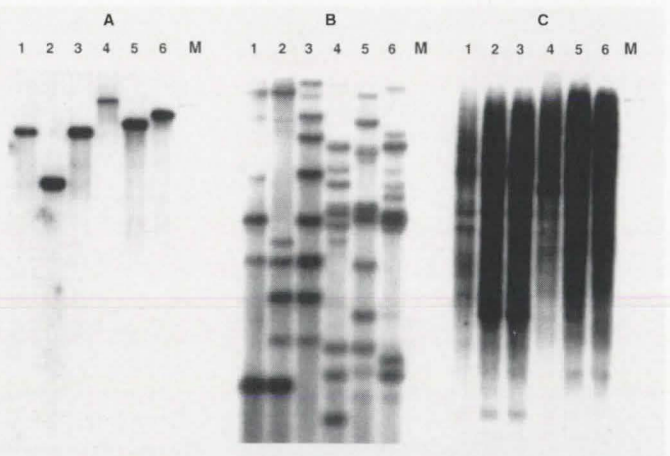


Figure 2. Screening of RFLP polymorphism. DNAs of three maize lines were digested by *Hind* III (lines 1 to 3) and *Eco* RI (lines 4 to 6) (M : molecular weight marker). Patterns revealed with three probes show different situations. **A** : simple pattern : the probe corresponds to a single copy sequence. **B** : complex pattern : the probe reveals several loci. **C** : high background : the probe corresponds to a repeated sequence.

analyser est assez importante (quelques microgrammes). Les délais pour révéler le résultat d'une analyse RFLP sur autoradiographie sont au minimum de quelques jours et peuvent aller jusqu'à quelques semaines. La reproductibilité des résultats dépend du soin apporté à la préparation des ADN, afin d'exclure tout risque de digestion partielle par les enzymes de restriction.

Une multitude de travaux, en amélioration des plantes ou analyse de la diversité génétique, utilise les marqueurs RFLP. En fonction des sondes utilisées, les marqueurs peuvent provenir de tout le génome, nucléaire ou cytoplasmique, codant ou non. Ils sont donc en nombre quasi illimité, locus-spécifiques et codominants. La grande diversité des enzymes de restriction disponibles permet de trouver très souvent du polymorphisme, cela dépend de l'effort de crible réalisé. Des plantes aussi peu polymorphes que le blé tendre ou la pomme de terre disposent désormais d'une carte génétique réalisée, dans des croisements intraspécifiques, grâce à des marqueurs RFLP [4]. Les sondes hétérologues sont largement utilisables. Chez les graminées, certaines d'entre elles, dites « sondes ancres », ont été choisies pour leur capacité à fournir un signal RFLP chez toutes les plantes de cette famille et pour leur bonne répartition chez les représentants les plus connus (maïs, orge, blé). Leur utilisation permet des comparaisons sur les positions et l'ordre des locus et constitue la base des études de cartographie comparée [5].

Les signaux d'hybridation sont, dans certaines conditions, quantifiables. Ainsi l'analyse de populations de maïs utilisées en Europe a-t-elle été réalisée sur des plantes dont l'ADN avait été mélangé. Chaque piste RFLP contenait l'ADN de 10 plantes, potentiellement toutes hétérozygotes. Pour chaque piste, la mesure de l'intensité des bandes correspondant aux différents allèles révélés par une sonde permet de reconstituer les fréquences alléliques à chaque marqueur pour chaque population [6]. Ce protocole, tout en économisant le volume des analyses, permet de disposer de données de diversité très informatives.

Les importants programmes de séquençage systématique d'ADNc (programme EST pour *Expressed Sequence Tag*, voir *glossaire*) produisent des informations très utiles aux biologistes moléculaires et aux physiologistes. Cependant, peu d'efforts sont réalisés afin de disposer du locus d'origine des gènes séquencés. Dans le cas

des plantes modèles, riz ou *Arabidopsis*, et en attendant les résultats du séquençage génomique complet, de nombreux gènes séquencés sont localisés, par des techniques de PCR et sans recherche de polymorphisme, sur les grands fragments, YAC (*Yeast Artificial Chromosome*) ou BAC (*Bacterial Artificial Chromosome*), qui constituent la trame de la carte physique [7]. Chez toutes les autres plantes, même celles qui bénéficient de gros programmes de génomique, une localisation génétique est nécessaire. La technique RFLP permet de localiser plusieurs centaines de loci d'ADNc. Des alternatives techniques, puces à ADN ou SNP (*Single Nucleotide Polymorphism*, voir *glossaire*) ne pourront être utilisées que lorsque les connaissances et les outils issus des programmes de génomique seront réellement disponibles.

Les marqueurs RFLP présentent toutes les qualités d'un très bon marqueur génétique. Ils peuvent être mono-locus ou multi-locus, codominants, bi ou multialléliques. Ils permettent de marquer toutes les régions du génome, nucléaire ou cytoplasmique, les régions codantes ou non codantes. Ils présentent un bon potentiel de transfert entre espèces, même assez éloignées, grâce à certaines sondes de séquences très conservées. Toutes ces qualités génétiques doivent être tempérées par les difficultés techniques de mise en œuvre et par l'impossibilité d'automatiser le procédé.

Marqueurs obtenus grâce à une amplification PCR ciblée

Le procédé PCR (*Polymerase Chain Reaction*) [8] permet l'amplification d'un fragment donné d'ADN, dans toute une série d'individus même si celui-ci est en simple copie au milieu d'un génome de grande taille. Ceci impose de connaître la séquence du fragment « ciblé » ou au moins celle de ses extrémités, afin de définir des amorces oligonucléotidiques spécifiques. La technique PCR est simple, rapide et très économe en ADN. Quelques nanogrammes suffisent pour obtenir une bonne amplification. Pour que ces produits deviennent des marqueurs moléculaires, il faut explorer le polymorphisme à l'intérieur du fragment. Selon l'objectif poursuivi, les connaissances préalables et les moyens disponibles, plusieurs techniques sont disponibles :

– le séquençage, qui révélera de manière exhaustive tous les polymorphismes, substitutions de bases ou insertions/délétions mais il est, aujourd'hui, peu utilisable comme technique de routine ;

– les techniques de digestion du produit d'amplification ;

– les techniques d'analyse globales de la conformation du fragment.

Ces deux derniers procédés sont généralement utilisés lors du développement de marqueurs de type STS (*Sequence Tagged Site*, voir *glossaire*). On peut, par exemple, conserver les caractéristiques génétiques de marqueurs RFLP déjà cartographiés en les convertissant en marqueurs « PCR ciblée » ;

– les techniques d'électrophorèse à haute résolution lorsque les fragments d'amplification sont constitués pour partie de répétitions en nombre variable d'une séquence de base répétée en tandem. Ces produits correspondent aux marqueurs VNTR (*Variable Number of Tandem Repeats*, voir *glossaire*) dont les marqueurs microsatellites sont les représentants les plus couramment utilisés.

Techniques de digestion du produit d'amplification : STS en CAPS ou PCR/RFLP

La technique CAPS (*Cleaved Amplified Polymorphic Sequence*, voir *glossaire*) ou PCR/RFLP consiste à digérer le fragment amplifié avec une ou plusieurs enzymes de restriction à site de reconnaissance tétranucléotidique et à révéler le polymorphisme des sites de restriction par électrophorèse classique en gel d'agarose ou d'acrylamide non dénaturant. Le choix des enzymes à 4 bases, qui coupent l'ADN en moyenne toutes les 256 bases (4^4), est imposé par la longueur moyenne des produits de la PCR, entre 0,5 et 3 kb, et permet d'avoir une probabilité élevée de coupure. Comme la RFLP classique, cette technique fournit des marqueurs codominants et spécifiques de locus. Le polymorphisme, contenu uniquement à l'intérieur du fragment d'amplification, est parfois difficile à mettre en évidence. Chez les graminées, par exemple, cette technique a été utilisée pour convertir certaines sondes RFLP en marqueurs STS [9]. La technique de PCR/RFLP est cependant surtout utile lorsque peu de marqueurs doivent être manipulés sur un grand nombre de plantes, comme dans le cas de marqueurs liés à des gènes de résistance aux parasites. Elle autorise un crible rapide sur des populations importantes.

Techniques d'analyse globales de la conformation du fragment : STS en PCR-SSCP

La technique de SSCP (*Single Strand Conformation Polymorphism*, voir *glossaire*) s'applique à l'analyse de produits d'amplification PCR. Lorsqu'un tel fragment d'ADN double brin est dénaturé par chauffage à 95 °C, puis rapidement refroidi, les molécules simple brin n'ont pas le temps de se réassocier entre elles mais forment une structure secondaire stable par des réassociations intramoléculaires au niveau de zones de séquences complémentaires. Les différences de séquences peuvent entraîner des différences de conformation qui sont décelées par une migration en conditions non dénaturantes dans un gel d'acrylamide. Cent pour cent des différences de séquences sont théoriquement décelables dans des fragments d'ADN de taille inférieure à 200 pb. Ce pourcentage diminue lorsque la longueur du fragment augmente, mais la technique reste assez performante pour des fragments de 800 à 1 000 pb. Chez un individu homozygote, deux bandes sont en général observables, car deux molécules d'ADN simple brin complémentaires ont des conformations secondaires différentes. Chez les végétaux, la technique SSCP a été utilisée dans le programme de cartographie de référence du génome du riz [10] ou l'analyse des flux de gènes entre des populations de chênes [11]. Le procédé SSCP est potentiellement très intéressant car il ne nécessite pas de digestion, comme les CAPS, et peut concerner n'importe quelle séquence. Il est utile pour le crible rapide des variants des produits d'amplification lorsque la connaissance précise des substitutions en cause n'est pas nécessaire. Cette technique est cependant encore assez peu utilisée en raison surtout de certaines difficultés de maîtrise des conditions d'électrophorèse.

Il existe d'autres techniques similaires d'analyse de la conformation des fragments d'amplification. Celles-ci utilisent des conditions d'électrophorèse dénaturantes. Il s'agit des techniques de D/TGGE (*Denaturing/Temperature Gradient Gel Electrophoresis*, voir *glossaire*).

Marqueurs microsatellites

Les séquences microsatellites ou SSR (*Simple Sequence Repeats*, voir *glossaire*) sont constituées de répétitions en tandem de motifs mono, di, tri ou tétranucléotidiques. Les plus courantes sont (A)_n,

(AT)_n, (GA)_n, (GT)_n, (TAT)_n, (GATA)_n, etc., la valeur de n pouvant aller de quelques unités à plusieurs dizaines. De tels motifs sont très abondants dans le génome des organismes eucaryotes. Chez les végétaux supérieurs, les premières estimations indiquent qu'il y aurait en moyenne un microsatellite dinucléotidique tous les 30 à 100 kb [12]. Une banque génomique de riz a été criblée par hybridation en utilisant comme sondes des motifs microsatellites. Les motifs poly (GA)_n et poly (GT)_n sont, par exemple, respectivement présents environ 1 360 et 1 230 fois dans le génome [13]. Des séquences microsatellites ont également été identifiées dans le génome chloroplastique de diverses espèces d'Angiospermes et de Gymnospermes [14].

Outre leur distribution sur l'ensemble du génome, l'intérêt en génétique des microsatellites réside dans leur polymorphisme extrêmement élevé. Chez le soja, jusqu'à 26 allèles par locus ont été décrits dans un groupe de 100 génotypes [15]. Le polymorphisme concerne le nombre des unités de répétition qui constituent la séquence microsatellite. Celui-ci varie vraisemblablement surtout à cause des erreurs dues au « glissement » de la polymérase lors de la réplication des chromosomes. La séquence microsatellite est, de plus, soumise aux mécanismes d'évolution rapides des séquences répétées en tandem, tels que les *crossing-over* asymétriques. Les taux de mutation des microsatellites ont été estimés chez l'homme et chez certains animaux. Ils s'établissent en général entre $4 \cdot 10^{-4}$ et $5 \cdot 10^{-6}$ par allèle et par génération [16]. Ce taux élevé, comparé aux autres séquences de l'ADN génomique, explique que les microsatellites soient très peu présents dans les séquences codantes, car trop instables et fortement contre-sélectionnés [17].

Un motif microsatellite donné n'est pas spécifique d'un locus ; en revanche, les séquences bordantes le sont. Une paire

d'amorces oligonucléotidiques spécifiques de ces régions bordantes permettra, grâce à la PCR, d'amplifier ce seul locus microsatellite et de disposer ainsi d'un marqueur microsatellite ou STMS (*Sequence-Tagged Microsatellite Site*, voir *glossaire*). Le polymorphisme de longueur sera révélé en électrophorèse en gel d'acrylamide ou d'agarose à haute résolution (*figure 3*). Le développement des marqueurs microsatellites, c'est-à-dire la définition des amorces bordantes spécifiques, est un travail assez lourd. Selon les protocoles classiques, il faut réaliser une banque génomique et la cribler afin d'y repérer les clones qui portent des motifs microsatellites. Ces clones sont séquencés, des amorces sont définies à partir de leur séquence et les conditions d'amplification et de révélation du polymorphisme sont mises au point. Des protocoles d'élaboration de banques enrichies en séquences microsatellites permettent d'alléger ce travail de clonage et de crible [18]. Les programmes importants de génétique fondés sur les microsatellites (génomes humains, animaux et végétaux) montrent que, pour un à deux microsatellites utilisables, cinq auront dû être séquencés. Il est parfois possible de repérer des séquences qui portent des motifs microsatellites simplement en interrogeant les bases de données qui compilent les séquences d'ADN publiées. Pour chaque plante, 1 à 5 % en moyenne des séquences disponibles de gènes ou d'ADNc contiennent des microsatellites. Ceux-ci sont presque exclusivement localisés dans les régions non codantes : introns, séquences amont du codon *start* ou région 3' non traduite [19].

Le transfert des marqueurs microsatellites à des espèces proches du même genre a été réalisé chez des plantes annuelles (revue *in* [20]) ou des arbres forestiers. À l'intérieur d'un groupe d'espèces très proches, les zones bordantes des microsatellites sont assez conservées, ce qui permet de transférer les amorces d'amplifi-

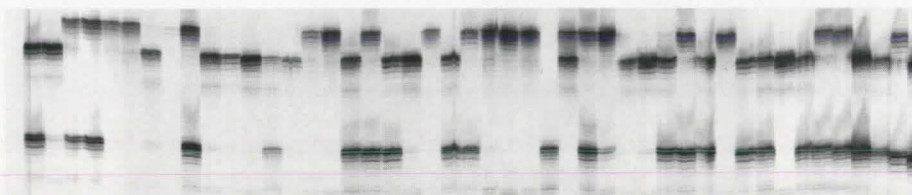


Figure 3. Marqueurs microsatellites en ségrégation dans une descendance de peupliers.

Figure 3. Segregation of microsatellite markers in a *Populus* family.

cation. Le passage à des espèces d'autres genres de la même famille montre un succès réduit dans l'obtention d'amplification de microsatellites et de leur polymorphisme. La conservation de ces séquences disparaît rapidement avec l'éloignement génétique des taxons. Chez les *Eucalyptus*, le transfert de l'amplification et du polymorphisme des quelques locus microsatellites testés était de 100 % en restant dans la même section, et seulement de 50 % ou de 0 % en passant à des espèces d'autres sections [21]. Cependant, des amorces assez généralistes permettent d'amplifier les locus microsatellites chloroplastiques chez diverses espèces de la famille des Pinacées [14].

Les microsatellites sont parmi les marqueurs les plus puissants pour révéler du polymorphisme. Ils sont de plus en plus fréquemment utilisés pour de nombreuses études. Ils constituent d'excellents marqueurs génétiques, spécifiques de locus, codominants et hautement polymorphes. Ils marquent essentiellement les régions non codantes ; cependant, leur répartition est assez uniforme sur le génome. Ils sont manipulables par PCR. Les protocoles de visualisation peuvent être en partie automatisés. Leur faible portabilité entre espèces végétales, même proches, nécessite des efforts de développement, parfois importants, pour chaque espèce.

Marqueurs dominants révélés « en masse »

Par rapport aux techniques décrites précédemment qui, toutes, visent à révéler le polymorphisme dans ou au voisinage d'une séquence particulière (technique spécifique de locus), les techniques décrites ici, regroupées sous l'appellation MAAP (*Multiple Arbitrary Amplicon Profiling*, voir *glossaire*), toutes fondées sur la PCR, ne « ciblent » pas une région particulière du génome fixée à l'avance. Presque toujours, elles révèlent simultanément plusieurs locus (jusqu'à plus d'une centaine), ce qui permet souvent, en peu d'expériences, de caractériser sans ambiguïté un génotype. Elles sont naturellement utilisées pour réaliser du génotypage rapide (ou *fingerprinting*) empreintes génétiques) mais servent également pour la cartographie génétique et chaque fois qu'il faut saturer en marqueurs une région particulière du génome.

Technique RAPD

Le principe des RAPD (*Randomly Amplified Polymorphic DNA*, voir *glossaire*) consiste à réaliser une réaction PCR sur l'ADN de l'individu étudié en utilisant une amorce courte, de 10 nucléotides en général, de séquence arbitraire. L'amorce va s'hybrider chaque fois que se trouvera dans l'ADN une séquence qui lui est complémentaire (ou comportant un nombre limité de mésappariements). Si deux sites d'hybridation sont proches l'un de l'autre (à moins de 3 000 pb) et en direction opposée, c'est-à-dire dans une configuration permettant la PCR, l'amplification aura lieu. Si un de ces deux sites est absent dans un autre individu, il n'y aura pas d'amplification et un polymorphisme de présence/absence sera observé. Les produits d'amplification obtenus en RAPD, dont le nombre dépasse rarement la dizaine, sont en général séparés par électrophorèse en gel d'agarose (*figure 4*).

Dans la majorité des cas, le polymorphisme révélé par la technique RAPD est de type présence/absence. Dans ces conditions, les homozygotes pour l'allèle « absence de bande » sont repérés sans ambiguïté, tandis que la présence de bande ne permet pas de trancher entre l'hétérozygote et l'homozygote pour l'allèle « présence de bande ». Ces marqueurs sont donc qualifiés de « dominants ». Lors

d'analyses de populations en ségrégation, il est cependant possible de repérer des bandes d'amplification allèles qui diffèrent par leur taille. Les marqueurs ainsi générés sont donc codominants. On peut considérer que 10 à 15 % des marqueurs RAPD sont codominants. La cartographie des marqueurs RAPD montre parfois que certaines bandes sont présentes dans les génotypes de la descendance alors qu'elles sont absentes des plantes parentales. Ces marqueurs sont des produits d'amplification issus de la formation de molécules d'ADN hétéroduplex lorsque des fragments allèles de tailles différentes sont co-amplifiés à partir d'un génotype hétérozygote. La formation de ces hétéroduplex au cours de la révélation de marqueurs RAPD codominants est une situation rare. Par exemple, 0,2 % seulement des bandes RAPD produites chez *Brassica napus* sont dans ce cas [22]. Cependant cette situation est évoquée par plusieurs auteurs.

Les marqueurs RAPD, marqueurs anonymes qui ne ciblent pas une séquence donnée, doivent être prudemment transférés à une espèce végétale autre que celle où ils ont été établis. Les marqueurs sont caractérisés par une taille de fragment, ce qui n'implique en aucun cas l'homologie de séquence des fragments amplifiés de même taille qui sont observés chez des individus différents. Les études de diversité qui reposent sur ces marqueurs devraient comporter une étape de vérifi-

Figure 4. Exemple d'une analyse RAPD dans une série de lignées de maïs. Les fragments d'amplification sont révélés au bromure d'éthidium après électrophorèse sur gel d'agarose.

Figure 4. RAPD analysis of maize lines. Detection of amplified fragments by ethidium bromide after agarose gel electrophoresis.



cation pour éviter la confusion d'allèles appartenant à des locus différents. Ces marqueurs ont parfois pu être transférés à des espèces proches. Rieseberg [23] a analysé 220 fragments RAPD chez quelques espèces sauvages d'*Helianthus* et a constaté, par l'hybridation des fragments d'ADN amplifiés de même taille, des homologies de séquences entre espèces pour 90 % des fragments, ce qui laisse présager d'un transfert élevé de ces marqueurs entre espèces proches. Mais cet auteur a aussi montré, en utilisant la cartographie des marqueurs, que l'homologie de séquence ne suffisait pas : la duplication des séquences induit des homologies de séquences entre locus différents. Finalement, seulement 79 % des marqueurs RAPD pouvaient être transférés entre ces espèces d'*Helianthus*. L'obtention de l'amplification d'un marqueur n'est pas assurée en changeant d'espèce, mais le polymorphisme du marqueur est encore moins évident à transférer. Il faut rester très prudent sur les possibilités de transfert de ces marqueurs.

Lorsqu'un marqueur RAPD a été repéré au voisinage d'un gène intéressant, par exemple un gène qu'il est utile de suivre au cours des générations dans un schéma de sélection, il est possible de le convertir en un marqueur révélé individuellement spécifique du locus considéré. Pour cela, la bande d'amplification RAPD intéressante est extraite du gel. L'ADN est alors cloné et séquencé. Des amorces d'une vingtaine de bases sont ensuite définies afin d'amplifier de façon spécifique le locus marqueur. Il devient alors un marqueur SCAR (*Sequence Characterized Amplified Region*, voir *glossaire*). Les diverses techniques de révélation du polymorphisme à l'intérieur des produits d'amplification STS pourront dès lors être mises en œuvre.

La RAPD et les techniques similaires, AP-PCR (*Arbitrarily Primed-PCR*) ou DAF (*DNA Amplification Fingerprinting*, voir *glossaire*) utilisant des amorces arbitraires, jouissent d'une grande popularité auprès des généticiens dans le domaine végétal. La raison principale est leur grande simplicité par rapport aux techniques précédemment décrites : il n'y a ni digestion de l'ADN, ni hybridation et surtout aucune connaissance préalable n'est requise sur les séquences. Elles sont fondées sur la PCR : elles sont rapides, nécessitent peu d'ADN et sont automatisables. Elles présentent cependant parfois des problèmes de reproductibilité.

Technique ISSR

Une exploitation plus simple des marqueurs microsatellites consiste à les révéler en masse, en s'inspirant du principe de la RAPD. Pour cela, on utilise une amorce constituée pour partie d'une séquence de microsatellite et pour partie de bases arbitraires. Deux types d'amorces sont concevables, selon les positions relatives de ces deux parties. La PCR va amplifier des fragments flanqués de motifs microsatellites (figure 5). Dans des conditions d'amplification adaptées, la technique ISSR (*Inter SSR-PCR*, voir *glossaire*) permet de produire plusieurs dizaines de produits qui sont visualisés soit sur des gels d'agarose, soit sur des gels d'acrylamide. Le polymorphisme révélé est essentiellement de type présence/absence, comme pour les RAPD, mais correspond parfois à des différences de longueur de fragment, comme pour les microsatellites. Les marqueurs ISSR se sont révélés très polymorphes. Ils permettent de distinguer des variétés génétiquement très proches [24].

Technique AFLP™

La technique AFLP (*Amplified Fragment Length Polymorphism*, voir *glossaire*) est fondée sur la mise en évidence conjointe de polymorphisme de site de restriction et d'hybridation d'amorces arbitraires. Elle a été mise au point par la société néerlandaise Keygene pour une utilisation initiale en amélioration des plantes et elle est cou-

verte par un brevet. L'ADN génomique purifié est d'abord coupé par deux enzymes de restriction (généralement *EcoRI* et *MseI*) ayant des sites de reconnaissance respectivement de 6 et 4 bases. Des adaptateurs de séquence connue de 10 à 15 bases, adaptés aux extrémités cohésives des sites de restriction, sont ajoutés aux extrémités des fragments d'ADN par une réaction de ligation. Une première amplification, dite présélective, amplifie les fragments d'ADN à l'aide d'amorces correspondant à la séquence des adaptateurs et du site de restriction additionné en 3' d'une base définie arbitrairement. Un mésappariement de cette base 3' terminale de l'amorce empêche totalement l'amplification. Dans cette réaction, seul 1/16 des fragments de restriction initiaux sont donc amplifiés. La deuxième amplification, dite sélective, utilise des amorces identiques aux premières additionnées de 1, 2 ou 3 bases supplémentaires à leur extrémité 3'. Un sous-ensemble de fragments est alors amplifié. Les produits de l'amplification finale sont séparés par électrophorèse en gel d'acrylamide dénaturant (gel de séquence). Ils peuvent être ensuite visualisés par coloration au nitrate d'argent ou révélés grâce à un marquage radioactif ou fluorescent réalisé lors de la seconde amplification. Les profils sont en général complexes (50 à 100 bandes) (figure 6). Le nombre de bases arbitraires des amorces de l'amplification sélective détermine le nombre de bandes produites en fonction de la taille du génome de l'organisme étudié.

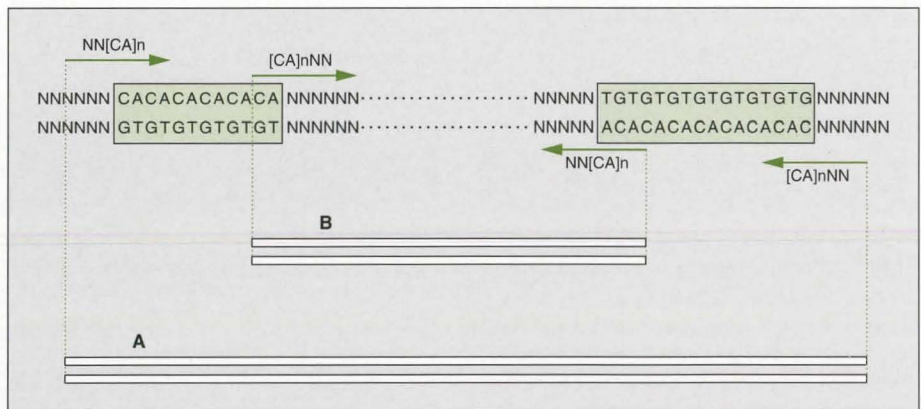


Figure 5. Principe de l'ISSR. Si les bases arbitraires de l'amorce sont du côté 5' du microsatellite, on obtiendra le produit d'amplification A, si elles sont du côté 3', le produit B.

Figure 5. ISSR principle. If arbitrary nucleotides of the primer are located at the 5' end of the microsatellite, product A will be generated. If they are located at the 3' end, product B will be generated.

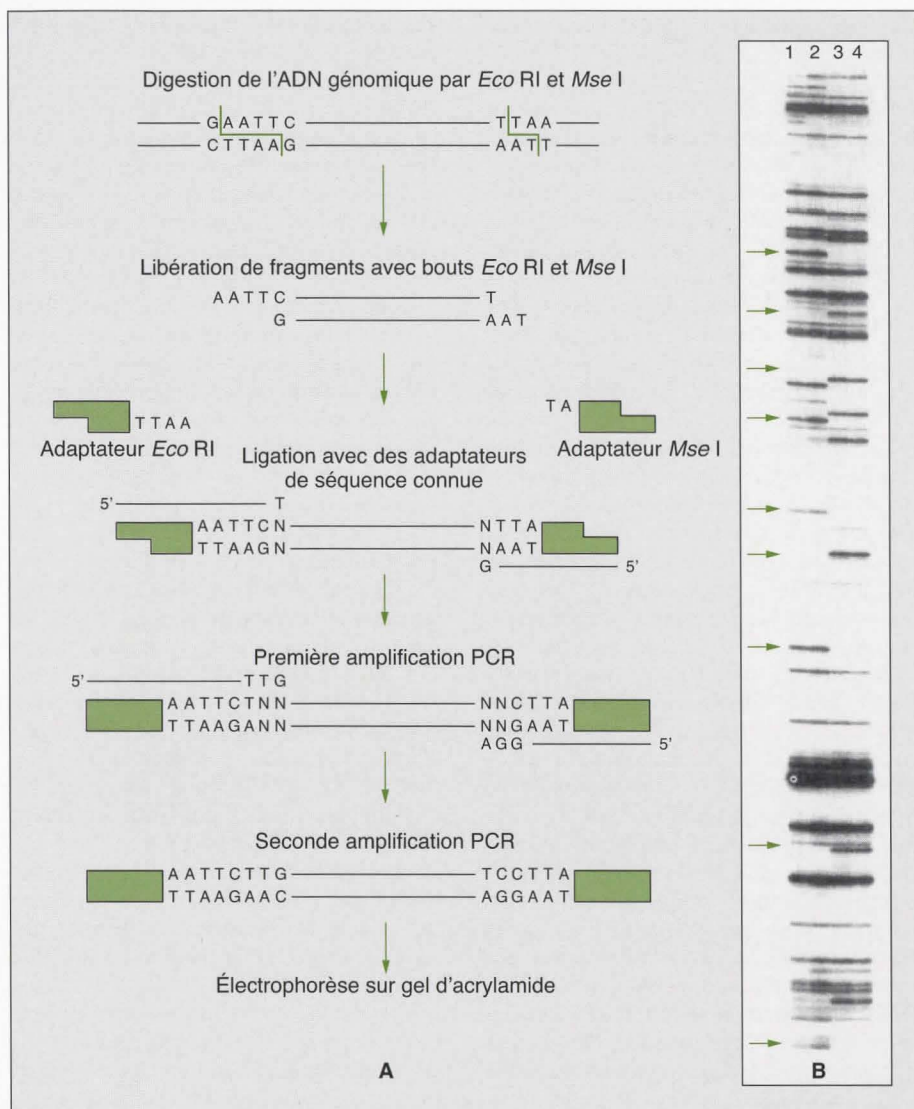


Figure 6. A : Principe de la technique d'AFLP. **B :** Profils de deux génotypes de l'écotype Columbia (pistes 1 et 2) et de deux génotypes de l'écotype Landsberg (pistes 3 et 4) d'*Arabidopsis*.

Figure 6. A : General AFLP procedure. **B :** Amplification pattern for two *Arabidopsis* plants of the Columbia ecotype (lines 1 and 2) and two *Arabidopsis* plants of the Landsberg ecotype (lines 3 and 4).

Il est théoriquement possible d'utiliser plusieurs dizaines de combinaisons d'enzymes de restriction, à site de 4 et 6 bases, et de très nombreuses combinaisons de bases sélectives sur les amorces d'amplification. Les combinaisons restriction/amplification sont presque infinies. La technique AFLP permet de révéler un polymorphisme important chez toute les plantes, essentiellement de type présence/absence, et peut produire des marqueurs issus de toutes les régions du génome. Elle constitue une technique de choix pour marquer rapidement un grand nombre de locus dans un génome.

Cette technique permet donc de réaliser des empreintes génétiques hautement informatives afin de comparer des génotypes génétiquement très proches, de réaliser facilement des cartes génétiques ou de saturer une région particulière d'un génome.

Comme pour la technique RAPD, les marqueurs AFLP sont essentiellement dominants. Cependant, des protocoles techniques optimisés et des logiciels d'analyse d'image proposés par les concepteurs de l'AFLP (<http://www.keygene.com>) permettent, dans certains cas, de doser les fragments amplifiés. Il est

alors possible de faire la différence entre l'état homozygote et hétérozygote d'un marqueur AFLP dont les séquences d'origine dans le génome sont respectivement à deux et à une copie. Des SCAR peuvent être dérivés de marqueurs AFLP.

La technique AFLP nécessite plus de technicité que la RAPD mais génère plus de marqueurs, plus polymorphes et avec plus de fiabilité. De nombreuses études de diversité ou de cartographie utilisent désormais le procédé AFLP.

Plusieurs techniques de génotypage dérivent de la technique AFLP. L'une d'elles, la S-SAP (*Sequence-Specific Amplification Polymorphism*, voir glossaire), se focalise sur le polymorphisme au niveau des gènes. Les premières étapes sont identiques à celles de l'AFLP classique. Cependant, l'amplification sélective est réalisée en utilisant une amorce AFLP aléatoire à 2 ou 3 bases débordantes en 3' couplée à une amorce définie dans une séquence connue. Celle-ci peut être une séquence de rétrotransposon ou un motif conservé d'une famille multigénique. Dans ces cas, tous les fragments d'amplification correspondent aux parties flanquantes des séquences qui sont ciblées. La S-SAP avec rétrotransposons permet de générer de nombreux marqueurs à partir de séquences d'ADN considérées comme relativement instables. Les profils S-SAP permettent de visualiser des modifications des sites d'insertion des rétrotransposons. Ils constituent un outil intéressant de recherche de polymorphisme entre des plantes très proches, éventuellement pour des caractérisations intraclonales. La S-SAP avec des amorces définies dans les séquences de gènes répétés (analogues de gènes de résistance, protéines kinases, facteurs de transcription, etc.) permet de générer, aussi facilement que pour une analyse AFLP classique, des marqueurs qui sont tous issus de séquences potentiellement codantes. L'élaboration de cartes génétiques avec ces marqueurs permettra, par exemple, de tester des hypothèses de type « gènes candidats » au cours de recherche de QTL (*Quantitative Trait Loci*). De la même manière, la technique SAMPL (*Selective Amplification of Microsatellite Polymorphic Loci*, voir glossaire) associe des séquences microsatellites à une amorce AFLP. L'usage de ce type de marqueurs semi-aléatoires devrait se généraliser.

Un certain nombre d'autres techniques de marquage moléculaire MAAP (ASAP, EPIC, ISTR, RAHM, RAMP, RAMPO, TecMAAP) sont décrites dans le glossaire final.

Reproductibilité : comparaison entre les différentes techniques de marquage moléculaire

Les génotypes déterminés pour les individus analysés doivent être reproductibles, sinon ils n'ont aucune signification génétique. Le manque de reproductibilité provient du prélèvement, de l'extraction, de problèmes techniques, de l'adaptation du protocole ou d'erreurs humaines. Certains types de marqueurs sont plus sensibles que d'autres aux petites fluctuations expérimentales difficiles à éviter.

Effet du prélèvement ou du tissu analysé

Le stade de développement, la date de prélèvement et la nature du tissu prélevé ne doivent pas avoir d'effet pour les marqueurs portant directement sur l'ADN, pour peu que l'ADN extrait soit de qualité suffisante. Cependant, Donini *et al.* [25] ont rapporté des variations dans les profils de marqueurs AFLP en fonction de l'organe de prélèvement (feuilles et grains de blé). Ces auteurs excluent la présence de contaminations dans les extraits, les effets de variation éventuelle du niveau de ploïdie, les réarrangements du génome, les variations en quantité suffisante des génomes cytoplasmiques. Le plus vraisemblable reste alors des variations spécifiques de la méthylation de l'ADN en fonction du tissu qui se traduiraient par des fragments de restriction et donc d'amplification différents. Les séquences d'ADN sont naturellement méthylées et cette méthylation peut donc induire des variations non contrôlées lorsque les enzymes de restriction utilisées sont sensibles à la méthylation (RFLP, AFLP).

Manipulations et réactifs

Comme pour toutes les expériences de biologie moléculaire et de biochimie, une grande minutie dans la réalisation des manipulations élémentaires est nécessaire pour avoir des résultats de marquage moléculaire reproductibles. De plus, de nombreux échantillons sont analysés. Les risques d'erreur dans la ges-

tion des expériences (génotypes et données) ne sont pas négligeables. L'automatisation de la manipulation des échantillons et de la collecte des résultats permet de diminuer ces risques d'erreur. La plupart des marqueurs développés ces dernières années reposent sur l'utilisation de la réaction de polymérisation en chaîne. Outre la qualité des manipulations, divers facteurs liés à ces réactions sont des sources de non-reproductibilité des résultats.

La *Taq* polymérase est une enzyme commercialisée par plusieurs sociétés. Durant plusieurs années, les différents fournisseurs n'ont pas tous délivré un produit de même qualité ni respectant les mêmes standards de dosage d'activité enzymatique. Les protocoles de marquage, surtout de type RAPD, mis au point avec une qualité d'enzyme ne pouvaient pas être appliqués sans adaptation avec une autre qualité d'enzyme. La généralisation des enzymes respectant les spécifications et le paiement d'une redevance imposées par le brevet Hoffman-Laroche permet désormais de disposer de produits très comparables. Les thermocycleurs contrôlent les températures et les changements de température pour la réaction de polymérisation en chaîne. La dynamique des changements de température est propre à chaque machine, ce qui conduit à un contrôle de la réaction qui peut varier selon la machine. Une vérification des résultats d'amplification obtenus avec différents thermocycleurs est toujours intéressante et des ajustements sont parfois nécessaires. La qualité et les concentrations des amorces utilisées pour les réactions de polymérisation en chaîne sont des facteurs importants qui interviennent sur la reproductibilité des profils d'amplification multi-bandes des techniques MAAP. Il est recommandé de disposer d'amorces hautement purifiées pour éviter de manipuler en mélange des traces d'amorces incomplètement synthétisées. Ces éléments polluants peuvent être source de variations en raison d'une moindre spécificité de l'amorce pour ses cibles et des risques de mauvais appariements en conditions de PCR peu spécifiques.

Transfert entre laboratoires

Un projet européen a permis de tester la reproductibilité de marqueurs RAPD, AFLP et microsatellites dans plusieurs laboratoires [26]. Les marqueurs RAPD sont les plus difficiles à transférer d'une équipe à une autre. Un protocole unique mis au point

sur peuplier a été transféré et appliqué par huit laboratoires sur deux clones de peuplier. Deux amorces ont servi à cette étude. La source de *Taq* polymérase est la même pour toutes les équipes, mais chaque équipe a son propre équipement en thermocycleurs. L'analyse des profils montre que chaque équipe a obtenu un profil qui lui est propre si on considère toutes les bandes publiées. Même en ne tenant compte que des bandes les plus fortes, les profils fluctuent entre laboratoires. Plusieurs bandes sont toutefois notées par toutes les équipes. Le transfert des marqueurs RAPD entre laboratoires était déjà mis en doute par d'autres travaux. Le facteur thermocycleur mentionné plus haut peut expliquer une partie de ces difficultés.

Les marqueurs AFLP ont été testés par six équipes différentes qui ont travaillé sur deux clones de betterave à sucre. Les deux couples d'amorces utilisés ont permis la lecture de 172 bandes au total. Les profils sont ici très reproductibles entre les six laboratoires, une seule bande n'a pas été observée par l'une des équipes. Les publications concernant les marqueurs AFLP n'ont pas signalé de difficultés de reproductibilité.

La reproductibilité des marqueurs microsatellites a été testée par six laboratoires qui ont suivi le protocole établi par une équipe pilote. Les génotypes établis pour les deux échantillons sont les mêmes pour tous les laboratoires, cependant la taille des produits d'amplification varie en fonction des équipes (écart d'une paire de base). Cet écart est faible et correspond aux erreurs rencontrées dans l'estimation de la taille des fragments.

Ces études confirment les difficultés d'utilisation des marqueurs RAPD par une autre équipe, alors que les marqueurs AFLP et microsatellites sont beaucoup plus fiables et peuvent être transférés d'un laboratoire à un autre. Les marqueurs RFLP développés par un laboratoire peuvent également être utilisés dans d'autres laboratoires, sous réserve d'utiliser le même protocole.

Utilisations des marqueurs

Les marqueurs sont utilisés pour décrire la variabilité génétique et sa répartition au sein de populations et d'espèces, ils servent aussi à préciser les mécanismes

évolutifs des populations qui rendent compte de cette description. Le choix des marqueurs dépend de l'objectif précis fixé et des moyens des utilisateurs (tableau).

Phylogénie

Les relations de proximité entre espèces ou populations sont abordées de deux façons différentes : la phénétiqne est fondée sur les distances génétiques entre les unités taxonomiques et aboutit à la construction de dendrogrammes ; la cladistique est fondée sur l'identification des mutations successives et repose sur le principe de parcimonie qui minimise le nombre d'événements pour décrire un arbre phylogénétique.

Les marqueurs utilisés doivent être neutres pour retracer correctement l'évolution des taxons. Plus les taxons sont proches génétiquement, plus les marqueurs utilisés doivent avoir un taux de mutation élevé pour révéler suffisamment de polymorphisme exploitable. Les relations entre les taxons sont fondées sur leur divergence qui est croissante au cours du temps.

Dans le cas de la phénétiqne, tout marqueur adapté à la construction d'une matrice de distances entre les unités taxonomiques analysées est utilisable pour la construction des dendrogrammes. Les marqueurs RFLP permettent de calculer des distances génétiques assez fiables entre groupes d'individus et sont largement utilisés pour la construction de dendrogrammes. Étant potentiellement très nombreux, les marqueurs RFLP fournissent des distances peu sensibles aux effets d'un locus particulier. Les marqueurs PCR spécifiques, STS et SCAR sont également préconisés [27]. Les marqueurs cytoplasmiques sont également utilisés et apportent une information différente du fait de leur génétique particulière. Les marqueurs microsatellites sont mal adaptés en raison de leur fort taux de mutation et de la possibilité des mutations réverses [17, 27]. L'utilisation des marqueurs dominants comme les AFLP et RAPD devient fréquente. Les dendrogrammes construits avec différents types de marqueurs montrent parfois des regroupements différents : c'est le cas pour 22 lignées de blé analysées avec les marqueurs AFLP, RFLP et microsatellites [28]. Dans une autre étude, les dendrogrammes ont été construits pour 33 lignées de maïs avec des marqueurs RAPD, AFLP, RFLP et

microsatellites [29] : les marqueurs RAPD fournissent un dendrogramme différent des autres.

La construction d'arbres phylogénétiques suppose que la séquence des événements de mutation puisse être retracée. La technique de choix pour de telles analyses est sans conteste le séquençage d'allèles. Celle-ci ne constitue cependant pas une technique de marquage moléculaire qui entre dans le cadre de cette revue. L'état de chaque site polymorphe doit pouvoir être caractérisé. Les marqueurs préconisés par Karp *et al.* [30] sont les marqueurs RFLP et PCR-RFLP pour lesquels les sites de restriction sont cartographiés de façon à individualiser chaque mutation. Les marqueurs STS et SCAR sont également utilisables si leur polymorphisme est révélé par RFLP et que les sites de restriction sont cartographiés. La mise en évidence du polymorphisme par SSCP ou D/TGGE n'est applicable que pour la phénétiqne. Les marqueurs RFLP des gènes ribosomiques dont les sites de restriction sont cartographiés sont de bons marqueurs pour la cladistique. Leur polymorphisme est analysé chez de nombreuses espèces d'arbres (par exemple les peupliers) [31] ou de plantes annuelles (par exemple les betteraves) [32]. L'information apportée par les rétrotransposons (marqueurs S-SAP) est utilisable pour des constructions de phylogénie du fait de l'accumulation progressive des éléments transposables au cours de l'évolution. Ces éléments intégrés ne sont pas éliminés, ils peuvent évoluer seulement par mutation. Ces marqueurs fournissent non une phylogénie chronologique, car l'activité des éléments transposables n'est pas homogène dans le temps, mais une phylogénie ordonnée.

Les génomes cytoplasmiques ont l'avantage de ne pas recombiner en raison de leur transmission uniparentale. L'évolution des séquences de ces génomes reste progressive et continue, même dans le cas d'introgessions ou de croisements entre taxons déjà différenciés. L'histoire des espèces peut donc être retracée. Les marqueurs cytoplasmiques analysés par RFLP avec cartographie des sites de restriction sont largement utilisés [30]. Les réorganisations fréquentes du génome mitochondrial le rendent cependant difficile à utiliser.

Les marqueurs dominants et anonymes comme les RAPD sont inadaptés à la construction d'arbres phylogénétiques fondés sur le principe de parcimonie, sauf situations très particulières [33].

Analyse de la diversité génétique

Phylogéographie

La phylogéographie analyse simultanément la diversité géographique et les relations phylogénétiques des populations. Comme pour la construction d'arbres phylogénétiques, la succession des mutations conduisant aux allèles observés doit pouvoir être retracée. Les marqueurs RFLP, PCR-RFLP, mtEPIC et cpPCR-RFLP de façon générale sont adaptés à ces études. Citons en exemple les études conduites sur les arbres forestiers dans ce domaine qui ont exploité le polymorphisme du génome chloroplastique avec des marqueurs cpPCR-RFLP. Les données paléontologiques, lorsqu'elles existent, apportent un éclairage complémentaire sur l'évolution et les migrations passées des espèces étudiées [34].

Structure et différenciation

La variabilité génétique peut se mesurer à l'échelle de la sous-population et de la population totale. La distribution des génotypes dans une sous-population est liée à sa sensibilité à la dérive génétique et à son mode de reproduction. Au niveau de la population totale, ce sont les flux de gènes entre unités qui déterminent l'essentiel de leur niveau de différenciation. La structure d'une sous-population est analysée par référence à la situation attendue dans une population panmictique (équilibre de Hardy-Weinberg). Pour tester d'éventuels écarts à cet équilibre, le génotype de tous les individus doit pouvoir être caractérisé aux locus analysés. Seuls les marqueurs codominants sont pertinents. En général, les marqueurs dominants ne sont utilisables que dans des espèces à fort taux de fixation, où les individus hétérozygotes sont presque inexistantes, ou dans des populations dont l'écart à l'équilibre de Hardy-Weinberg peut être connu [35], de préférence pour les marqueurs analysés.

Les taux d'hétérozygotie sont estimés de la même façon, en utilisant des marqueurs codominants, ou en connaissant l'écart à l'équilibre de Hardy-Weinberg. Les valeurs d'hétérozygotie et les paramètres du polymorphisme varient en fonction du marqueur utilisé. Russell *et al.* [36] ont comparé des marqueurs RFLP, AFLP, RAPD et microsatellites chez l'orge : l'indice de diversité (hétérozygotie attendue sous l'hypothèse de la

Tableau

Les principales techniques de marquage moléculaire : coûts et domaines d'application génétiques

Coût de l'analyse (par locus révélé)	Coût et difficulté du développement	Marqueurs	Structure et différenciation génétique	Cladistique et phylogéographie	Identification clonale	Recherche de parenté	Régime de reproduction	Flux de gènes interspécifique	Cartographie et détection de QTL	Clonage positionnel	Cartographie comparée
		Codominants et révélés individuellement									
Moyen	Moyens	EPIC	+++	++ a	++	+	++	++	+		++
Moyen	Moyens	mtEPIC	+++	+++ a	+	++		+			
Moyen	Moyens	ESTP	+++	++ a	+	+	+	+	+		++
Moyen	Moyens	PCR-RFLP	+++	+++ a	++	+	++	+++	+		++
Important	Importants	PCR-SSCP	+++		++	+	+++	+++	+		++
Moyen	Moyens	cpPCR-RFLP	+++	+++ a		+		+++			
Important	Importants	RFLP (ADNc)	+++	+++ a	++	+	++	++	+++		+++
Important	Importants	RFLP (génomique)	+++	+++ a	++	++	++	++	+++		++
Moyen	Importants	SCAR	++		+	+	+	+	+		+
Faible	Importants	STMS microsatellites	+++		+++	+++	+++	++	+++		+
Faible	Moyens	cpSSR	++		++	+++		++			
Moyen	Moyens	STS	++	++ a	+	+	+	+	+		++
		Dominants et révélés « en masse »									
Faible	Faibles	AFLP	+		+++	+	+	+	+++	+++	
Faible	Faibles	AP-PCR	+		++	+			+	++	
Faible	Faibles	ASAP			++	+			+	+	
Faible	Faibles	DAF			++	+			+	++	
Faible	Faibles	ISSR	+		+++	+	+		++	++	
Faible	Moyens	ISTR	+		++	+	+		++	++	
Faible	Faibles	RAMP	+		++	+			+	+	
Faible	Faibles	RAPD	+		++	+			+	+	
Faible	Faibles	SAMPL	+		+++	+		+	+++	++	
Faible	Moyens	S-SAP	+	+	+++	+		+	+++	++	

Les difficultés du développement des marqueurs et le coût de leur utilisation ont été estimés à partir de l'expérience acquise dans nos laboratoires.
 + : utilisable ; ++ : bien adapté ; +++ : très efficace ; a : les sites de restriction révélant le polymorphisme sont cartographiés.

Main molecular marker technology : cost and genetic applications. The problem of developing molecular markers and the cost of their implementation were estimated on the basis of our laboratory experience.

+ : usable ; ++ : well adapted ; +++ : highly efficient ; a : the restriction sites revealing polymorphism are mapped.

panmixie) est plus élevé avec les marqueurs AFLP (0,94) qu'avec les marqueurs microsatellites (0,57), pourtant tous polymorphes, ou les marqueurs RAPD (0,52), les marqueurs RFLP présentant la plus faible diversité (0,32). Les mêmes types de marqueurs ont été utilisés chez *Glycine*, la plus forte hétérozygotie étant observée avec les marqueurs microsatellites [37]. Cependant, les distances génétiques entre lignées de *Glycine max* obtenues avec les marqueurs microsatellites étaient mal corrélées avec celles obtenues par les autres marqueurs. Les auteurs concluent que l'utilisation des marqueurs microsatellites n'est pertinente que dans une gamme de distance génétique donnée. Les marqueurs RFLP conservent l'intérêt des marqueurs codominants.

Les marqueurs dérivés de la PCR tels que des marqueurs PCR-SSCP [11] permettent aussi l'analyse de la structure des sous-populations.

Chez le conifère *Pseudotsuga menziesii*, dont les génomes chloroplastique, mitochondrial et nucléaire ont une hérédité différente, Hong *et al.* [38] ont comparé la différenciation entre provenances dans les différentes régions de l'aire naturelle et pour ces trois génomes. Le génome nucléaire est analysé avec 95 marqueurs DAF, le génome chloroplastique avec 30 fragments RFLP polymorphes et le génome mitochondrial avec 25 fragments RFLP polymorphes. La différenciation entre sous-populations (*Gst*) est plus élevée pour le génome mitochondrial (0,45) que pour le génome chloroplastique (0,20) et le génome nucléaire (0,14). La situation est différente pour la variété côtière qui est moins différenciée pour le génome mitochondrial (0,11) que pour les autres génomes (0,13-0,14). La forte structuration du génome mitochondrial, à transmission maternelle, reflète la faible migration des graines alors que les flux de pollen plus élevés induisent une structuration plus faible du polymorphisme chloroplastique. La transmission biparentale du génome nucléaire et les flux de gènes par le pollen conduisent également à une faible structuration de ce génome. L'analyse du polymorphisme des trois génomes procure donc des informations complémentaires.

Les marqueurs nucléaires codominants étant les plus informatifs, les marqueurs RFLP restent parmi les plus utilisés. Mais l'utilisation des marqueurs microsatellites tend à se généraliser en raison de

Glossaire

Les marqueurs moléculaires ADN : quelques définitions

AFLP *Amplified Fragment Length Polymorphism* [53]

Polymorphisme de longueur des fragments d'amplification
Amplification PCR d'ADN génomique après digestion avec 2 enzymes de restriction et ligation d'adaptateurs d'environ 20 paires de bases. Les amorces PCR correspondent aux adaptateurs plus 2 ou 3 bases aléatoires en 3'. Les profils obtenus sont multibandes.

AP-PCR *Arbitrarily Primed Polymerase Chain Reaction* [54]

PCR avec des amorces arbitraires
Amplification au hasard de type RAPD utilisant des amorces de 20 bases à des températures d'hybridation faibles.

ASAP *Arbitrary Signatures from Amplification Profiles* [55]

Signature arbitraire de profil d'amplification
Amplification PCR de type RAPD en utilisant plusieurs amorces en une ou plusieurs amplifications successives.

CAPS *Cleaved Amplified Polymorphic Sequence* (PCR/RFLP) [56]

Digestion des produits d'amplification polymorphes
Un fragment d'ADN génomique amplifié (STS, SCAR) est digéré par des enzymes de restriction afin de détecter les profils électrophorétiques polymorphes des fragments de restriction.

DAF *DNA Amplification Fingerprinting* [57]

Empreinte génétique de produits d'amplification
Même principe que les RAPD en utilisant des amorces plus courtes (5 à 8 bases). Les profils électrophorétiques des fragments d'ADN amplifiés sont souvent complexes.

DGGE *Denaturing Gradient Gel Electrophoresis* [58]

Électrophorèse en gel à gradient de dénaturation
Électrophorèse de fragments d'ADN amplifiés sur un gel d'acrylamide contenant un gradient linéaire de produit dénaturant de l'ADN (urée, formamide). Le fragment d'ADN, initialement double brin, est partiellement dénaturé pour former une structure « branchée » (en Y) à un endroit précis du gel (donc du gradient) au cours de l'électrophorèse. Sa mobilité électrophorétique est alors modifiée. La séquence de l'ADN détermine la zone du gradient où il sera dénaturé.

EPIC *Exon Priming, Intron Crossing* [59]

Amplification d'intron par amorçage sur exons
Amplification PCR d'un intron d'un gène grâce à des amorces définies dans les exons flanquants. Permet d'amplifier une région potentiellement polymorphe dans un gène.

EST *Expressed Sequence Tags* [60]

Étiquettes de séquences exprimées
Séquences anonymes ou de fonction connue issues de programmes de séquençage systématique d'ADNc.

ESTP *Expressed Sequence Tag Polymorphisms* [27]

Polymorphisme des EST
Amplification PCR à l'aide d'amorces définies à partir de la séquence d'un EST. Marqueur de type STS.

ISSR *Inter Simple Sequence Repeats* [61]

Amplification intermicrosatellites
Amplification sur de l'ADN génomique avec des amorces constituées d'un motif répété dinucléotidique, de type microsatellite, associé à des bases (2 à 6) définies au hasard placées en 3' ou en 5'. Les fragments amplifiés sont donc situés dans le génome entre deux locus microsatellites.

Autres appellations : **I-SSR PCR** *Inter Simple Sequence Repeats* PCR, **IMA** *Inter Microsatellites Amplification*, **IRA** *Inter-Repeat Amplification*, **ISA** *Inter-SSR Amplification*.

ISTR *Inverse Sequence-Tagged Repeat* [62]

Amplification PCR à l'aide d'amorces définies dans les séquences d'un élément hautement répété dans le génome, rétrotransposons de type *copia*. Les produits d'amplifications proviennent de régions situées en

leur polymorphisme important, de leur codominance, de leur abondance dans le génome et de la relative simplicité technique de leur manipulation. Les autres marqueurs dérivés de la PCR ciblée (STS, PCR-RFLP/SSCP, SCAR) sont également tout à fait utilisables et les études de différenciation peuvent être menées aussi avec des marqueurs dominants.

Flux de gènes, recherche de parenté et introgression

Les flux de gènes entre les sous-populations sont souvent déduits de leur différenciation, les marqueurs utilisés sont donc alors ceux qui servent à décrire la structuration de la diversité entre sous-populations. Les flux de gènes au niveau d'une sous-population sont en partie décrits par les régimes de reproduction et donc par l'estimation du taux d'allofécondation dans le cas des modèles mixtes [39]. L'hybridation entre espèces sympatriques est une source de flux de gènes et éventuellement d'introgression entre espèces. Les outils les plus appropriés ne sont pas les mêmes dans les différents cas.

Les taux d'allofécondation ou d'autofécondation sont correctement estimés à partir des génotypes notés à une dizaine de locus polymorphes, neutres et indépendants. Ces locus sont de préférence codominants. Des marqueurs cartographiés comme des RFLP, des STS et des microsatellites sont donc intéressants.

La recherche de paternité permet de retrouver les sources de pollen et de décrire le flux de pollen effectif. Les marqueurs utilisés doivent être codominants et posséder de nombreuses formes alléliques pour différencier un maximum d'individus. Les locus microsatellites sont ceux qui ont le plus fort pouvoir d'exclusion grâce au nombre élevé d'allèles par locus. Ainsi, 32 locus microsatellites ont permis de retrouver les espèces parentales de plusieurs cépages prestigieux de vigne. Les auteurs [40] ont montré que le Chardonnay, le Gamay noir, l'Aligoté et 13 autres cépages traditionnels cultivés dans le Nord-Est de la France, en Bourgogne et en Champagne, sont des cultivars mélangés dont les parents sont les variétés Pinot et Gouais blanc. Ce dernier, considéré comme une variété médiocre et réservé un temps aux terroirs pauvres, a presque complètement disparu du vignoble Français. Cependant, l'éloignement génétique assez important entre les deux variétés parentales, établi grâce

deux locus *copia* ou *copia-like*. Similaire à l'amplification inter-séquences Alu chez les mammifères.

MAAP *Multiple Arbitrary Amplicon Profiling* [63]

Profils d'amplification arbitraires multiples

Toute amplification PCR sur l'ADN génomique qui utilise des amorces définies arbitrairement et qui génère des profils électrophorétiques de fragments amplifiés complexes et polymorphes (appellation générique pour AP-PCR, DAF, RAPD, ISSR)

PCR-RFLP *PCR-Restriction Fragment Length Polymorphism*

Voir CAPS

RAHM *Random Amplified Hybridization Microsatellites* [64]

Recherche de microsatellites par hybridation sur amplification au hasard

Des profils d'amplification RAPD sont analysés par hybridation à l'aide de sondes oligonucléotidiques qui correspondent à des motifs microsatellites. Ce procédé permet de repérer des fragments qui portent des séquences microsatellites pour, après clonage et séquençage, en faire des marqueurs microsatellites.

RAMP *Randomly Amplified Microsatellite Polymorphism* [65]

Polymorphisme de microsatellites amplifiés au hasard

Amplification PCR en utilisant, en mélange, une amorce de type RAPD et une amorce correspondant à une séquence microsatellite. Les fragments amplifiés correspondent à l'environnement de plusieurs locus microsatellites.

RAMPO *Randomly Amplified Microsatellite Polymorphisms* [66]

Polymorphisme des microsatellites contenus dans les profils d'amplification au hasard

Utilisation des profils RAHM pour l'analyse du polymorphisme entre génotypes.

RAPD *Randomly Amplified Polymorphic DNA* [67]

ADN polymorphe amplifié au hasard

Amplification PCR de séquences inconnues utilisant une seule amorce de 10 bases définie arbitrairement. Les profils électrophorétiques des fragments d'ADN amplifiés contiennent une dizaine de bandes.

RFLP *Restriction Fragment Length Polymorphism* [68]

Polymorphisme de longueur des fragments de restriction

Polymorphisme de la longueur de fragments d'ADN générés par une hydrolyse par une enzyme de restriction. Celui-ci est révélé après hybridation avec une sonde marquée sur les fragments séparés par électrophorèse sur gel et transférés sur membrane.

SAMPL *Selective Amplification of Microsatellite Polymorphic Loci* [69]

Amplification sélective des locus microsatellites polymorphes

Analyse de type AFLP en utilisant une amorce arbitraire AFLP couplée à une amorce qui correspond à un motif microsatellite. Les fragments amplifiés correspondent à l'environnement de plusieurs locus microsatellites.

SCAR *Sequence-Characterized Amplified Region* [70]

Région amplifiée de séquence caractérisée

Fragment d'ADN génomique amplifié par PCR avec des amorces spécifiques (14 à 20 bases). Ces amorces sont définies grâce à la connaissance de la séquence d'un fragment MAAP (RAPD, AFLP...) intéressant isolé d'un gel, cloné et séquencé.

SNP *Single Nucleotide Polymorphism* [71]

Polymorphisme d'un seul nucléotide

La source principale de polymorphisme dans un génome est l'apparition de mutations ponctuelles qui modifient une seule paire de bases. Lorsque cette mutation atteint la fréquence de 1 % dans une population, elle peut être exploitée comme marqueur génétique et est appelée SNP. Les SNP sont repérés en compilant des données de séquençage. Ils sont présents dans les parties codantes et non codantes d'un génome. Ils sont révélés par des techniques rapides de marquage moléculaire, PCR et spectrométrie de masse, PCR et puces à ADN ou PCR et électrophorèse rapide, en perpétuelle évolution. Les SNP commencent à être utilisés chez l'homme et les génomes modèles.

S-SAP *Sequence-Specific Amplification Polymorphisms* [72]

Polymorphisme d'amplification de séquence spécifique

Analyse de type AFLP en utilisant une amorce arbitraire AFLP couplée à

aux données de typage microsatellite, génère un fort hétérosis parmi les descendants et a participé à l'émergence de variétés performantes.

Les marqueurs microsatellites chloroplastiques (cpSSR) présentent aussi un intérêt pour la recherche de paternité chez les espèces à transmission paternelle du génome chloroplastique et pour la recherche de maternité chez les autres espèces.

Les techniques d'empreintes génétiques fournissent de nombreux marqueurs dominants dont l'analyse simultanée permet également de mener des recherches de paternité. Chez *Persoonia mollis*, le polymorphisme des marqueurs AFLP est suffisant pour identifier 99 % des plantes pollinisatrices avec seulement trois combinaisons d'amorces [41].

L'introgression interspécifique est observée dans de nombreux groupes d'espèces végétales. La mise en évidence des hybrides F1 est possible à partir de marqueurs diagnostiques caractéristiques des espèces parentales. Keim *et al.* [42] ont étudié avec des marqueurs RFLP la structure génétique d'une population de peupliers, dans une zone où *Populus fremontii* et *P. angustifolia* sont sympatriques. Ils ont mis en évidence un polymorphisme intraspécifique et interspécifique, qui leur a permis d'identifier des hybrides F1 et des individus issus du rétrocroisement avec l'espèce parentale *P. angustifolia*. L'introgression est plus difficile à déceler dans les générations plus avancées. Le nombre de marqueurs diagnostiques doit être plus élevé. Des marqueurs nucléaires dominants très polymorphes comme les AFLP ont été aussi utilisés pour rechercher d'éventuels hybrides obtenus à la suite de l'introduction de formes cultivées dans l'aire naturelle de *Populus nigra* [43]. L'analyse des introgressions entre deux espèces d'*Helianthus* a été menée à l'aide de marqueurs RAPD cartographiés et répartis sur les divers groupes de liaison : ceci a permis de déceler les nombreuses régions du génome qui recombinent peu chez ces hybrides et qui sont impliquées dans la baisse de fertilité pollinique des hybrides [44]. Les marqueurs moléculaires révèlent fréquemment des distorsions de ségrégation importantes chez les hybrides interspécifiques [45].

Identification

L'identification de clones ou de lignées est plus facile que celle de population ou de variétés synthétiques. L'efficacité des

une amorce définie dans une séquence connue d'un rétrotransposon ou d'une famille multigénique.

SSCP *Single Strand Conformation Polymorphism (Profil)* [73]

Polymorphisme (ou Profil) de conformation simple brin

Analyse par électrophorèse sur gel d'acrylamide de la conformation en simple brin, après dénaturation et renaturation en boucle des simples brins de fragments d'ADN amplifiés.

SSR *Simple Sequence Repeats* [74]

Répétitions de séquences simples

Séquences microsatellites.

STMS *Sequence-Tagged Microsatellite Site* [75]

Microsatellite caractérisé par deux séquences bordantes connues.

Marqueur microsatellite révélé par amplification PCR en utilisant deux amorces spécifiques encadrant un motif microsatellite.

STS *Sequence-Tagged Site* [76]

Site caractérisée à partir d'une séquence connue. PCR locus-spécifique

Fragment d'ADN génomique amplifié par PCR avec des amorces spécifiques d'une séquence cartographiée. Parfois sert à définir la conversion de l'utilisation d'une sonde RFLP en un test PCR locus-spécifique.

Tec MAAP *Template Endonuclease Cleavage MAAP* [63]

Profils d'amplifications arbitraires multiples après digestion de l'ADN cible

Amplification PCR de type MAAP sur de l'ADN génomique préalablement digéré par des enzymes de restriction.

TGGE *Temperature Gradient Gel Electrophoresis*

Électrophorèse en gel à gradient de température

Même principe d'analyse que la DGGE, cependant le gradient de dénaturation dans le gel est généré par un gradient de température dans la cuve d'électrophorèse.

VNTR *Variable Number of Tandem Repeats* [77]

Nombre variable de répétitions en tandem

Type particulier de polymorphisme d'insertion/délétion dû à la variation du nombre de répétitions dans les séquences micro ou minisatellites. Terme générique désignant parfois à la fois les microsatellites et les minisatellites. Cependant, souvent réservé à la désignation des séquences minisatellites.

Glossary : DNA molecular markers : a few definitions

marqueurs AFLP, RAPD, RFLP et microsatellites pour caractériser les variétés a été comparée chez l'orge [37]. Chaque type de marqueur a permis l'identification des 18 variétés, mais avec un effort plus important pour les marqueurs RFLP. Les regroupements des variétés étaient différents en fonction du marqueur. L'identification de clones ou de lignées peut être menée avec les différentes techniques d'empreinte génétique bien reproductibles et qui fournissent de nombreux marqueurs sur un seul gel : AFLP, RAPD, S-SAP, ISSR.

L'identification de 18 populations de maïs a été conduite à l'aide de marqueurs RFLP [46] grâce à leur nombre d'allèles assez élevé. Cependant, Smith *et al.* [47] utilisent des marqueurs codominants microsatellites pour la caractérisation de variétés de maïs : ces marqueurs se sont avérés plus efficaces que les marqueurs RFLP, tant pour leur reproductibilité que pour leur pouvoir de discrimination.

Les marqueurs cytoplasmiques, cpPCR-RFLP/SSCP, mtEPIC, dont la variabilité est très structurée géographiquement et/ou dont le polymorphisme caractérise une lignée maternelle, peuvent également servir à identifier des cultivars.

Cartographie génétique et applications

Tous les types de marqueurs permettent, potentiellement, de construire des cartes génétiques, d'établir des associations entre marqueurs et gènes et de localiser dans le génome les régions qui sont impliquées dans le déterminisme des caractères quantitatifs (QTL).

Carte référence, portabilité dans l'espèce

Tous les marqueurs peuvent être utilisés pour élaborer la carte génétique de réf-

rence d'une espèce. La carte doit contenir un maximum de marqueurs et être au moins saturée. La difficulté pour utiliser les informations de la carte de référence en vue d'établir une carte sur une autre population est de retrouver du polymorphisme à chaque locus pour chaque marqueur. La présence de marqueurs microsatellites, très polymorphes, dans la carte de référence en fait une carte consensus utilisable dans un grand nombre de croisements. Les marqueurs RFLP sont aussi exploitables. L'effort de recherche de polymorphisme peut cependant être important. Intra-espèce, l'homologie des marqueurs est la règle générale. Les marqueurs MAAP ne sont caractérisés que par une longueur de fragment d'amplification ; cependant, leur homologie peut être vérifiée, si besoin, par hybridation ou séquençage. Dans le cas de l'orge, l'homologie de nombreux marqueurs AFLP est montrée par comparaison de leur localisation sur trois cartes génétiques obtenues avec trois populations différentes [48].

En général, à chaque transfert de la carte, seulement une partie des marqueurs de référence montre du polymorphisme et devient utilisable. Les groupes de liaisons risquent d'être mal définis avec de nombreux marqueurs indépendants. Cette trame de carte génétique doit être alors rapidement complétée avec des marqueurs anonymes propres au croisement analysé (MAAP) très utiles pour saturer une carte génétique.

Recherche de QTL, gènes candidats

Les marqueurs dominants n'apportent pas une information complète sur l'effet du QTL en raison de la confusion des individus hétérozygotes avec certains individus homozygotes. Les effets de dominance du QTL (écart à l'additivité) ne peuvent pas être abordés. L'analyse fonctionnelle du génome, menée essentiellement chez les espèces modèles, donne accès à des séquences codantes, de fonction identifiée et intervenant dans divers caractères. Ces séquences constituent des « gènes candidats » qui peuvent rendre compte de l'effet de certains QTL et remplacer, après validation, le marqueur lié au QTL. Les marqueurs de fonction connue (RFLP avec cDNA, STS) ont l'avantage de permettre de manipuler ces gènes candidats. La technique semi-aléatoire S-SAP présente le grand intérêt de permettre

de révéler de nombreux marqueurs issus de séquences codantes qui peuvent devenir éventuellement des gènes candidats s'ils sont supposés intervenir dans le caractère étudié et s'ils sont co-localisés avec un QTL.

Comparaison

carte génétique/carte physique

L'utilisation des marqueurs locus spécifiques manipulables par PCR (STS, microsatellites) permet de réaliser facilement des correspondances entre les

cartes génétiques et les cartes physiques élaborées grâce à l'organisation en contigs de grands fragments d'ADN génomique clonés en BAC ou en YAC. Un bon exemple nous est fourni par la comparaison des cartes génétiques et physique du blé tendre à l'aide de marqueurs microsatellites [49, 50].

Saturation rapide, marquage de gènes majeurs

Les marqueurs MAAP sont parfaitement adaptés pour saturer une carte

Summary

Molecular markers for genetic resource analysis and plant breeding

S. Santoni, P. Faivre-Rampant, E. Prado, D. Prat

Molecular markers are common tools that can reveal polymorphism directly at the DNA level and are used for genetic resource assessment and breeding programmes. We present various molecular genotyping techniques and their principles. There are two types of molecular markers that differ according to the information provided, i.e. codominant markers and fingerprinting patterns. Codominant markers mainly include RFLPs and PCR-derived markers (including microsatellites). Polymorphism of PCR amplified fragments can be obtained using restriction enzymes or via conformational analysis. RFLPs and microsatellites are the most commonly used codominant molecular markers. Microsatellite loci usually involve many alleles. The reproducibility of RFLPs and PCR-derived markers is well established. Between-species transfer of markers is often possible for RFLPs and, for PCR-derived markers, depends on the phylogenetic proximity of species. Fingerprinting techniques give complex patterns: RAPD, AFLP and ISSR markers are described with AFLP-derived techniques. Markers generally have only two forms (presence, absence) with dominant inheritance. The transfer of fingerprint patterns between populations or species is a critical feature that should be interpreted with caution. Several sources of variation such as the physiological state, reagents, and manipulation can alter marker reproducibility. RFLPs, AFLPs and microsatellites are some of the most reproducible markers.

Technical considerations alone are not sufficient for choosing a marker type. Required marker characteristics vary according to the application. Several applications of molecular markers are considered and discussed: phylogeny, diversity, material identification, and genetic mapping. Genetic distances between populations can be assessed from various types of markers for phenogram construction, while phylogenetic trees require markers whose consecutive mutations can be identified. Most molecular markers enable assessment of gene diversity and genetic differentiation of populations, which is a major application for genetic resource assessment. Gene flux, including introgression and paternity analysis, can be analysed with various molecular markers. Plant breeding applications mainly involve variety identification and marker-assisted selection, which requires genetic mapping and quantitative trait locus localization. Identification of lines or clones is an important practical molecular marker application: fingerprint patterns are fully adapted for this purpose. The choice of markers for genetic mapping and QTL detection will require extension of the study to other pedigrees or species, along with more informative markers. The quick construction of a genetic map is obtained with fingerprinting markers such as AFLPs. Marker characteristics are finally compared and their main application fields are presented and discussed according to technical aspects and information obtained.

Cahiers Agricultures 2000 ; 9 : 311-27.

génétique, surtout pour des espèces peu étudiées, où les sondes et les marqueurs PCR n'ont pas été développés. Pour des objectifs de clonage positionnel, ils permettent, dans des dispositifs génétiques adaptés, d'obtenir une forte densité de marqueurs dans une région précise du génome, au voisinage du gène à cloner (revue in [51]).

Cartographie comparée entre espèces

La cartographie comparée d'espèces voisines présente un intérêt fondamental pour la connaissance de l'évolution des génomes, mais aussi un intérêt appliqué pour la recherche de gène ou de QTL d'un caractère donné dans une espèce mal connue, en exploitant les données disponibles sur une espèce proche. Les marqueurs utilisés doivent être retrouvés dans des espèces voisines. Les séquences d'évolution rapide, séquences non codantes comme des microsatellites, ont peu de chances d'être retrouvées surtout si les espèces sont un peu trop éloignées. En revanche, les séquences codantes sont très utiles pour la cartographie comparée car leur divergence est assez faible. La technique doit aussi permettre de révéler un polymorphisme dans différents fonds génétiques. En pratique, les marqueurs RFLP, surtout avec des sondes ADNc, sont les mieux adaptés à la cartographie comparée.

Conclusion

Les domaines d'application des marqueurs génétiques font apparaître des objectifs variés pour lesquels des marqueurs sont préconisés. Cette analyse conduit à définir 9 applications du marquage génétique qui diffèrent par la nature des marqueurs susceptibles d'apporter l'information recherchée. Le tableau propose une comparaison de l'adéquation des marqueurs décrits dans ce document par rapport à ces finalités. Des objectifs proches sont parfois atteints avec des marqueurs différents (flux de gènes intrapopulation) et des objectifs différents peuvent faire appel aux mêmes marqueurs (identification de variétés/populations et étude de diversité génétique sur la structure et la différenciation) car les informations utilisées sont semblables. L'identification de clones repose sur les techniques d'empreintes génétiques qui lui sont

propres. La phylogénie fondée sur la cladistique et celle fondée sur la phylogéographie partagent les mêmes outils assurant l'identification des mutations à chaque site polymorphe. Les études de structure des populations et de leur différenciation utilisent les mêmes marqueurs que la phénétique et l'identification des variétés ou populations. L'analyse des flux de gènes à l'échelle interspécifique ou intrapopulation exploite surtout le polymorphisme des marqueurs spécifiques de locus.

Les cartes génétiques peuvent être construites avec tous les types de marqueurs. Selon le besoin de transfert des cartes génétiques et des QTL pour des utilisations dans un autre fond génétique ou pour des travaux de cartographie comparée, deux ensembles de marqueurs génétiques peuvent être définis : les marqueurs ancres qui permettent des comparaisons et les marqueurs spécifiques du croisement, facilement et massivement manipulables. Certains gènes candidats sont ainsi impliqués dans des caractères adaptatifs. Le polymorphisme de ces séquences codantes intervenant dans un caractère adaptatif permet de révéler une variabilité non neutre des populations. La recherche de QTL est engagée sur *Arabidopsis* pour des caractères adaptatifs [52] et la connaissance de ce génome permettra prochainement d'en connaître les gènes responsables. La mise en évidence d'un polymorphisme ayant une signification adaptative est très importante pour la gestion des ressources génétiques qui a comme objectif le maintien du potentiel adaptatif des populations et des espèces. L'étude de l'expression des gènes est donc utile pour identifier les marqueurs révélant un avantage sélectif. La recherche de QTL et l'identification des gènes responsables ont des applications à la fois dans l'amélioration des plantes et dans la gestion de leurs ressources génétiques.

Sur le plan technologique, l'avenir des marqueurs moléculaires passe nécessairement par une automatisation. À ce titre, les techniques qui s'appuient sur la PCR peuvent être rapidement automatisées. Les quantités d'ADN à extraire des plantes sont faibles et des robots manipulateurs développés pour des applications biomédicales (Biomek Bekman, Zymate Zymark, Qiagen...) peuvent s'en charger. Les étapes d'amplification PCR peuvent être réalisées grâce à des automates dédiés. Enfin, toutes les analyses de petits fragments d'ADN, de type

microsatellite ou AFLP par exemple, sont réalisables sur des séquenceurs automatiques, en électrophorèse classique sur gel ou en tube capillaire, ce qui permet une très forte productivité et une gestion des données optimisée.

Le concept de génomique, apparu depuis peu, consiste à dresser un inventaire de l'ensemble des gènes d'un organisme pour étudier leur organisation, leurs fonctions, leurs régulations et leurs interactions. Les connaissances accumulées sur la structure des gènes et le développement d'outils d'analyse massive ouvrent la voie à la mise au point et à l'utilisation en très grandes séries de nouveaux types de marqueurs moléculaires. La notion de SNP, recherche et exploitation du polymorphisme à la base près, permet de connaître précisément de nombreux allèles au même locus chez un organisme. De là sont dérivées des technologies, puce à ADN (*DNA chip*) ou spectromètre de masse, associées à des outils informatiques de plus en plus performants, qui réalisent des génotypages sur un grand nombre de locus et sur un grand nombre d'individus, rapidement et très précisément. Ces techniques sont, pour l'instant, le plus souvent réservées à la génétique humaine et à celle de certains microorganismes. Dans le domaine végétal, deux plantes modèles, *Arabidopsis* et le riz, font l'objet d'un séquençage génomique complet et seront étudiées avec tous les outils de la génomique fonctionnelle. D'autres plantes, représentatives d'une grande famille végétale, modèles pour une grande fonction physiologique ou économiquement très importante, profiteront d'études importantes de génomique structurale et fonctionnelle. La génomique comparative permettra d'atteindre une synergie de connaissances entre les études menées sur différentes espèces. Dès lors, les nouveaux types de marqueurs moléculaires issus des connaissances en génomique végétale pourront être mis en œuvre sur de très nombreuses plantes et ouvriront certainement de nouvelles voies à la génétique végétale et à ses grandes applications, l'analyse des ressources génétiques et l'amélioration des plantes ■

Références

1. Bretting PK, Widrechner MP. Genetic markers and plant genetic resource management. In : Janick J, ed. *Plant Breeding Reviews*, Vol 13. New York : John Wiley & Sons Inc, 1995 : 11-86.

2. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975 ; 98 : 503-17.
3. De Vienne D, Santoni S. Les principales sources de marqueurs moléculaires. In : De Vienne D, ed. *Les marqueurs moléculaires en génétique et biotechnologies végétales*. Paris : INRA Éditions, 1998 : 15-47.
4. Nelson JC, Sorrells ME, Van deyne AE, et al. Molecular mapping of wheat : major genes and rearrangements in momoelogenous groups 4, 5 and 7. *Genetics* 1995 ; 141 : 721-31.
5. Devos KM, Gale MD. Comparative genetics in the grasses. *Plant Mol Biol* 1997 ; 35 : 3-15.
6. Dubreuil P, Rebourg C, Merlino M, Charcosset A. Evaluation of a DNA pooled sampling strategy for estimating the RFLP diversity of maize populations. *Plant Mol Biol Rep* 1999 ; 0 : 1-16.
7. Wu J, Shimokawa T, Maehara T, Koike K, Umehara Y, Matsumoto T. Mapping rice EST's onto YAC clones. *Rice Genome* 1997 ; 6 : 2-3.
8. Saiki RK, Gelfand DH, Stoffel S, et al. Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science* 1988 ; 239 : 487-91.
9. Williams MNV, Pande N, Nair S, Mohan M, Bennett J. Restriction fragment length polymorphism analysis of polymerase chain reaction products amplified from mapped loci of rice (*Oryza sativa* L.) genomic DNA. *Theor Appl Genet* 1991 ; 82 : 489-98.
10. Fukuoka S, Inoue T, Miyao A, Monna L, Zhong HS, Sasaki T, Minobe Y. Mapping of sequence-tagged sites in rice by single-strand conformation polymorphism. *DNA Research* 1994 ; 1 : 271-7.
11. Bodenes C, Laigret F, Kremer A. Inheritance and molecular variations of PCR-SSCP fragments in pedunculate oak (*Quercus robur* L.). *Theor Appl Genet* 1996 ; 93 : 348-54.
12. Morgante M, Olivieri AM. PCR-amplified microsatellites as markers in plant genetics. *Plant J* 1993 ; 3 : 175-82.
13. Panaud O, Chen X, McCouch SR. Frequency of microsatellite sequences in rice (*Oryza sativa* L.). *Genome* 1995 ; 38 : 1170-6.
14. Vendramin GG, Lelli L, Rossi P, Morgante M. A set of primers for the amplification of 20 chloroplast microsatellites in pinaceae. *Mol Ecol* 1996 ; 5 : 595-8.
15. Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB. The use of microsatellite DNA markers for soybean genotypes identification. *Theor Appl Genet* 1995 ; 90 : 43-8.
16. Primmer CR, Ellengren H, Saino N, Moller AP. Directional evolution in germline microsatellite mutations. *Nat Genet* 1996 ; 13 : 391-3.
17. Jarne P, Lagoda P. Microsatellites, from molecules to populations and back. *Trends Ecol Evol* 1996 ; 11 : 424-9.
18. Kijas JMH, Fowler JCS, Garbett CA, Thomas MR. Enrichment of microsatellites from citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *BioTechniques* 1994 ; 16 : 656-62.
19. Milbourne D, Meyer RC, Collins AJ, Ramsay LD, Gebhardt C, Waugh R. Isolation, characterization and mapping of simple sequence repeat loci in potato. *Mol Gen Genet* 1998 ; 259 : 233-45.
20. Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A. Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSR) within the genus and other legume genera : implications for the transferability of SSRs in plants. *Mol Biol Evol* 1998 ; 15 : 1275-87.
21. Byrne M, Marquez-Garcia MI, Uren T, Smith DS, Moran GF. Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Aus J Botany* 1996 ; 44 : 331-41.
22. Hallden C, Hansen M, Nilsson NO, Hjerdin A, Säll T. Competition as a source of errors in RAPD analysis. *Theor Appl Genet* 1996 ; 93 : 1185-92.
23. Rieseberg LH. Homology among RAPD fragments in interspecific comparisons. *Mol Ecol* 1996 ; 5 : 99-105.
24. Prevost A, Wilkinson MJ. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor Appl Genet* 1999 ; 98 : 107-12.
25. Donini P, Elias ML, Bougourd SM, Koebner RMD. AFLP fingerprinting reveals pattern differences between template DNA extracted from different plant organs. *Genome* 1997 ; 40 : 521-6.
26. Jones CJ, Edwards KJ, Castiglione S, et al. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding* 1997 ; 3 : 381-90.
27. Echt CS, Harry DE, Nelson CD. PCR-based co-dominant markers in forest genetics. In : *Novel applications of molecular markers in forest trees*, 1996, SRIEG Workshop, Houston, Texas, June 23-26, 1996 ; 4 p.
28. Bohn M, Utz HF, Melchinger AE. Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. *Crop Sci* 1999 ; 39 : 228-37.
29. Pejic I, Ajmonemarsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. *Theor Appl Genet* 1998 ; 97 : 1248-55.
30. Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T. Molecular tools in plant genetic resources conservation : a guide to the technologies. *IPGRI Technical Bulletin* 1997 ; 2 : 47.
31. Faivre-Rampant P, Jeandroz S, Lefèvre F, Lemoine M, Villar M, Bervillé A. Ribosomal DNA studies in poplars : *Populus deltoides*, *P nigra*, *P trichocarpa*, *P maximowiczii*, and *P alba*. *Genome* 1992 ; 35 : 733-40.
32. Santoni S, Berville A. Characterization of the nuclear ribosomal DNA units and phylogeny of *Beta* L. wild forms and cultivated beets. *Theor Appl Genet* 1992 ; 83 : 533-42.
33. Backeljau T, De Bruyn L, De Wolf H, Jordaens K, Van Dongen S, Verhagen R, Winnepeninckx B. Random amplified polymorphic DNA (RAPD) and parsimony methods. *Cladistics* 1995 ; 11 : 119-30.
34. Demesure B, Comps B, Petit RJ. Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution* 1996 ; 50 : 2515-20.
35. Kremer A. Les marqueurs moléculaires en génétique des populations. In : De Vienne D, ed. *Les marqueurs moléculaires en génétique et biotechnologies végétales*. Paris : INRA Éditions, 1998 : 119-38.
36. Russell JR, Fuller JD, Macaulay M, et al. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theor Appl Genet* 1997 ; 95 : 714-22.
37. Powell W, Morgante M, Andre C, et al. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 1996 ; 2 : 225-38.
38. Hong YP, Ponoy B, Carlson JE. Genetic diversity and phylogeny in Douglas-fir based on RFLP and RAPD (DAF) analyses of nuclear, chloroplast, and mitochondrial genomes. In : Baradat P, Adams WT, Müller-Starck G, eds. *Population genetics and genetic conservation of forest trees*. Amsterdam : SPB Academic Publishing, 1995 : 247-66.
39. Ritland K, Jain S. A model for the estimation of outcrossing rate and gene frequencies using n independent loci. *Heredity* 1981 ; 47 : 35-52.
40. Bowers J, Boursiquot JM, This P, Chu K, Johansson H, Meredith C. Historical genetics : the parentage of Chardonnay, Gamay and other wines grapes of northeastern France. *Science* 1999 ; 285 : 1562-5.
41. Krauss SL, Peakall R. An evaluation of the AFLP fingerprinting technique for the analysis of paternity in natural populations of *Persoonia mollis* (*Proteaceae*). *Aust J Bot* 1998 ; 46 : 533-46.
42. Keim P, Paige KN, Whitham TG, Lark KG. Genetic analysis of an interspecific hybrid swarm of *Populus* : occurrence of unidirectional introgression. *Genetics* 1989 ; 123 : 557-65.
43. Winfield MO, Arnold GM, Cooper F, et al. A study of genetic diversity in *Populus nigra* subsp. *betulifolia* in the upper Severn area of the UK using AFLP markers. *Mol Ecol* 1998 ; 7 : 3-10.
44. Rieseberg LH, Whitton J, Gardner K. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 1999 ; 152 : 713-27.
45. Rieseberg LH, Carney SE. Plant hybridization. *New Phytol* 1998 ; 140 : 599-624.
46. Dubreuil P, Charcosset A. Genetic diversity within and among maize populations : a comparison between isozyme and nuclear RFLP loci. *Theor Appl Genet* 1998 ; 96 : 577-87.
47. Smith JSC, Chin ECL, Shu H, et al. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.) : comparisons with data from RFLPs and pedigree. *Theor Appl Genet* 1998 ; 95 : 163-73.
48. Waugh R, Bonar N, Baird B, et al. Homology of AFLP products in three mapping populations of barley. *Mol Gen Genet* 1997 ; 255 : 311-21.
49. Röder MS, Korzun V, Wendehake K, et al. A microsatellite map of wheat. *Genetics* 1998 ; 149 : 2007-23.
50. Röder MS, Korzun V, Gill BS, Ganai MW. The physical mapping of microsatellite markers in wheat. *Genome* 1998 ; 41 : 278-83.
51. De Vienne D. Marquage des gènes majeurs. In : De Vienne D, ed. *Les marqueurs moléculaires en génétique et biotechnologies végétales*. Paris : INRA Éditions, 1998 : 81-7.
52. Kuitinen H, Sillanpää MJ, Savolainen O. Genetic basis of adaptation : flowering time in *Arabidopsis thaliana*. *Theor Appl Genet* 1997 ; 95 : 573-83.

53. Vos P, Hogers R, Bleeker M, *et al.* AFLP : a new technique for DNA fingerprinting. *Nucl Acids Res* 1995 ; 23 : 4407-14.
54. Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. *Nucl Acids Res* 1990 ; 19 : 861-6.
55. Caetano-Anollés G, Gresshoff PM. *DNA markers*. New York : Wiley-VCH Inc, 1997 : 364.
56. Konieczny A, Ausubel FM. A procedure for mapping *Arabidopsis* mutations using codominant ecotype-specific PCR-based markers. *Plant J* 1993 ; 4 : 403-10.
57. Caetano-Anollés G, Bassam BJ, Gresshoff PM. DNA fingerprinting using very short arbitrary oligonucleotides. *Biotechnology* 1991 ; 9 : 553-7.
58. Myers RM, Maniatis T, Lerman S. Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol* 1987 ; 155 : 501-27.
59. Palumbi SR, Baker CS. Contrasting population structure from nuclear intron sequences and mtDNA of Humpback Whales. *Mol Biol Evol* 1994 ; 11 : 426-35.
60. Adams MD, Kelley JM, Gocayne JD, *et al.* Complementary DNA sequencing : expressed sequence tags and human genome project. *Science* 1991 ; 252 : 1651-6.
61. Zietkiewicz E, Rafalski A, Labuda D. Genome fingerprinting by simple-sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 1994 ; 20 : 176-83.
62. Rohde W. Inverse Sequence-Tagged Repeat (ISTR) analysis, a novel and universal PCR based technique for genome analysis in the plant and animal kingdom. *J Genet Breeding* 1996 ; 50 : 249-61.
63. Caetano-Anollés G. MAAP : a versatile and universal tool for genome analysis. *Plant Mol Biol* 1994 ; 25 : 1011-26.
64. Cifarelli RA, gallitelli M, Cellini F. Random amplified hybridization microsatellites (RAHM) : isolation of a new class of microsatellite-containing DNA clones. *Nucl Acids Res* 1995 ; 23 : 3802-3.
65. Matsumoto C, Nabika T, Mashimo T, Kato N, Yamori Y, Masuda J. Construction of a rat genetic map by randomly amplified microsatellite polymorphism (RAMP) markers. *Mamm Genome* 1998 ; 9 : 531-5.
66. Richardson T, Cato S, Ramser J, Kahl G, Weising K. Hybridization of microsatellites to RAPD : a new source of polymorphic markers. *Nucl Acids Res* 1995 ; 23 : 3798-9.
67. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl Acids Res* 1990 ; 18 : 6531-5.
68. Bostein D, White RL, Skolnic M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am J Hum Genet* 1980 ; 32 : 314-31.
69. Morgante M, Vogel J. Compound microsatellite primers for the detection of genetic polymorphism. *US Patent Application* 1994 ; 08/326456.
70. Paran I, Michelmore RW. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor Appl Genet* 1993 ; 85 : 985-93.
71. Wang DG, Fan JB, Siao CJ, *et al.* Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998 ; 280 : 1077-82.
72. Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol Gen Genet* 1997 ; 253 : 687-94.
73. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. Detection of polymorphism of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci USA* 1989 ; 86 : 2766-70.
74. Tautz D. Hypervariability of simple sequence as a general source for polymorphic DNA markers. *Nucl Acids Res* 1989 ; 17 : 6463-71.
75. Beckman JS, Soller M. Toward a unified approach to gene mapping of eukaryotes based on sequence tagged microsatellite sites. *Bio-Technology* 1990 ; 8 : 930-2.
76. Olson M, Hood L, Cantor C, Doststein D. A common language for physical mapping of the human genome. *Science* 1989 ; 254 : 1434-5.
77. Nakamura Y, Leppert M, O'Connell P, *et al.* Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 1987 ; 235 : 1616-22.

Résumé

Les marqueurs moléculaires, directement issus du polymorphisme existant au niveau de l'ADN, sont désormais utilisés fréquemment pour l'analyse des ressources génétiques et dans les programmes d'amélioration des plantes. Leurs caractéristiques permettent de les séparer en deux familles. La première est constituée par les marqueurs codominants et révélés individuellement dont les plus courants sont les marqueurs RFLP et microsatellites. La seconde famille est constituée par les marqueurs dominants et révélés « en masse ». Ils sont générés par des techniques de type RAPD ou AFLP qui permettent de réaliser facilement des empreintes génétiques. Chaque type de marqueur est décrit et la technique de laboratoire qui permet de le révéler est expliquée. En fonction de leurs caractéristiques propres, génétiques et techniques, les domaines d'application les plus pertinents de chacun (analyse de phylogénie, de structuration de la diversité génétique ou établissement de cartes génétiques) sont proposés.