Les méthodes de marquage biochimique et moléculaire

Laurent Grivet, Jean-Louis Noyer

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

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

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

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

L'organisation du génome

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

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

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

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

Le polymorphisme de l'ADN

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

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

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

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

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

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

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

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

La nature et l'interprétation génétique du polymorphisme

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

LES MARQUEURS MULTIALLÉLIQUES

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

Les isoenzymes et les RFLP

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

❑ Pour les RFLP, un allèle peut correspondre à une combinaison de plusieurs bandes s'il existe un ou des sites de coupure de l'enzyme de restriction dans la séquence d'ADN homologue à celle de la sonde.

❑ Pour les isoenzymes, les enzymes polymériques peuvent engendrer de nouvelles bandes chez les hétérozygotes.

❑ Un substrat (isoenzymes) ou une sonde (RFLP) peuvent révéler simultanément les allèles de plusieurs locus paralogues, c'est-à-dire de locus présentant la même séquence d'ADN ou des séquences très proches.

❑ Pour les RFLP, si plusieurs enzymes de restriction sont utilisées en combinaison avec une même sonde, chaque allèle correspond en principe à une combinaison spécifique de bandes révélées avec les différentes enzymes. En pratique, les reconstructions d'allèles à partir d'analyses de polymorphisme multienzymatique sont rarement effectuées car trop complexes. On se contente dans ce cas de noter la présence ou l'absence de chaque bande ou de chaque profil.

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

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

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

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

Les microsatellites

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

Les empreintes génétiques

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

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

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

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

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

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

La comparaison des cinq techniques

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

Le polymorphisme des techniques multiallélIques

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

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

Espèce	Technique	Nombre d'individus étudiés	Nombre de locus étudiés	Taux de polymorphisme (%)	Diversité de Nei	Nombre moyen d'allèles par locus	Référence
Brassica campestris	Iso	285	5	60	0,41	2,7	Mitchell et Quiros (1992)
	RFLP	277	4	100***	0,60	8,5	
Hordeum vulgare	Iso	268	7	100	0,44	4,9	Zhang *et al.* (1993)
	RFLP	240	13	100***	0,47	4,2	
Populus tremuloides	Iso	118	14	77	0,25	2,8	Liu et Furnier (1993)
	RFLP	91	41	71	0,25	2,7	
Populus grandidenta	Iso	96	14	29	0,08	1,4	Liu et Furnier (1993)
	RFLP	75	37	65	0,13	1,8	
Zea mays	Iso	21	22	68	—	2,1	Messmer *et al.* (1991)
	RFLP	21	144 (79)*	94	—	3,34**	
Zea mays	Iso	31	27	—	—	2,2	Gerdes et Tracy (1994)
	RFLP	43	71	—	—	4,1	
Zea mays	Iso	445	20	75	0,23	2,4	Dubreuil et Charcosset (1998)
	RFLP	285	35	100	0,61	6,3	
Hordeum vulgare	RFLP	18	114 (42)*	—	0,32	2,62**	Russell *et al.* (1997)
	SSR	18	13	100	0,57	5,38	
Glycine max	RFLP	19	114	100***	0,38	2,15	Morgante *et al.* (1994)
	SSR	19	4	100	0,52	4,25	
Glycine sp.	RFLP	12	110	25	0,41	—	Powell *et al.* (1996)
	SSR	12	36	100	0,60	4,28	
Oryza sativa	RFLP	14	18	33	0,10	2,3	Wu et Tanksley (1993)
	SSR	14	8	100	0,70	5,2	
Oryza sativa	RFLP	20	12	83	0,32	2,5	Olufowote *et al.* (1997)
	SSR	20	10	100	0,62	7,4	
Zea mays	RFLP	12	96	—	0,58	—	Taramino et Tingey (1996)
	SSR	12	34	100	0,76	6,56	

* x (y) : x nombre de couples enzyme-sonde, y nombre de sondes.
** Nombre moyen de bandes par couple enzyme-sonde.
*** Sondes RFLP prétriées pour leur polymorphisme.

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

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

Le polymorphisme révélé par l'ensemble des techniques

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

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

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

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

Les similarités génétiques

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

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

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

Référence	Espèce	N*	Test**	Comparaison***	Correlation *(r)*	Seuil *(P)*
Lu *et al.* (1996)	pois	10	Tm	RFLP - RAPD	0,5	< 0,05
				RFLP - AFLP	0,7	< 0,05
				RAPD - AFLP	0,6	< 0,05
Powell *et al.* (1996)	soja	10	Tm	RFLP - RAPD	0,24	ns
				RFLP - AFLP	0,42	< 0,01
				RFLP - SSR	0,18	ns
				RAPD - AFLP	0,45	< 0,01
				RAPD - SSR	0,15	ns
				AFLP - SSR	0,14	ns
Beer *et al.* (1993)	*Avena sterilis*	177	Tm	RFLP - Iso	0,26	< 0,005
Heun *et al.* (1994)	*Avena sterilis*	24	Tm	RFLP - Iso	0,36	< 0,01
Peakall *et al.* (1995)	*Buchloe dactyloides*	48	Tm	RAPD - Iso	0,63	< 0,001
Messmer *et al.* (1991)	maïs	21	Cp	RFLP - Iso	0,23	< 0,01
Gerdes et Tracy (1994)	maïs	31	Cp	RFLP - Iso	0,26	< 0,01
Prabhu *et al.* (1997)	soja	10	Cp	RFLP - DAF	0,70	< 0,01
Russell *et al.* (1997)	orge	18	Cr	RFLP - RAPD	0,20	< 0,01
				RFLP - AFLP	0,71	< 0,01
				RFLP - SSR	0,51	< 0,01
				RAPD - AFLP	0,11	ns
				RAPD - SSR	0,24	< 0,01
				AFLP - SSR	0,52	< 0,01
Dos Santos *et al.* (1994)	choux	45	Cr	RFLP - RAPD	0,75	< 0,01
Engquist et Beeker (1994)	colza	17	Cr	RFLP - RAPD	0,76	< 0,01
				RFLP - Iso	0,53	< 0,01
				RAPD - Iso	0,67	< 0,01
Thormann *et al.* (1994)	*Brassica* spp.	18	Tm	RFLP - RAPD	0,93	—

* Nombre d'individus utilisés pour calculer la corrélation.
** Tm : test de Mantel ; Cp : coefficient de corrélation de Pearson ; Cr : coefficient de corrélation de rang de Spearman.
*** Iso : isoenzymes ; SSR : microsatellites ; DAF : *DNA amplification fingerprinting*.

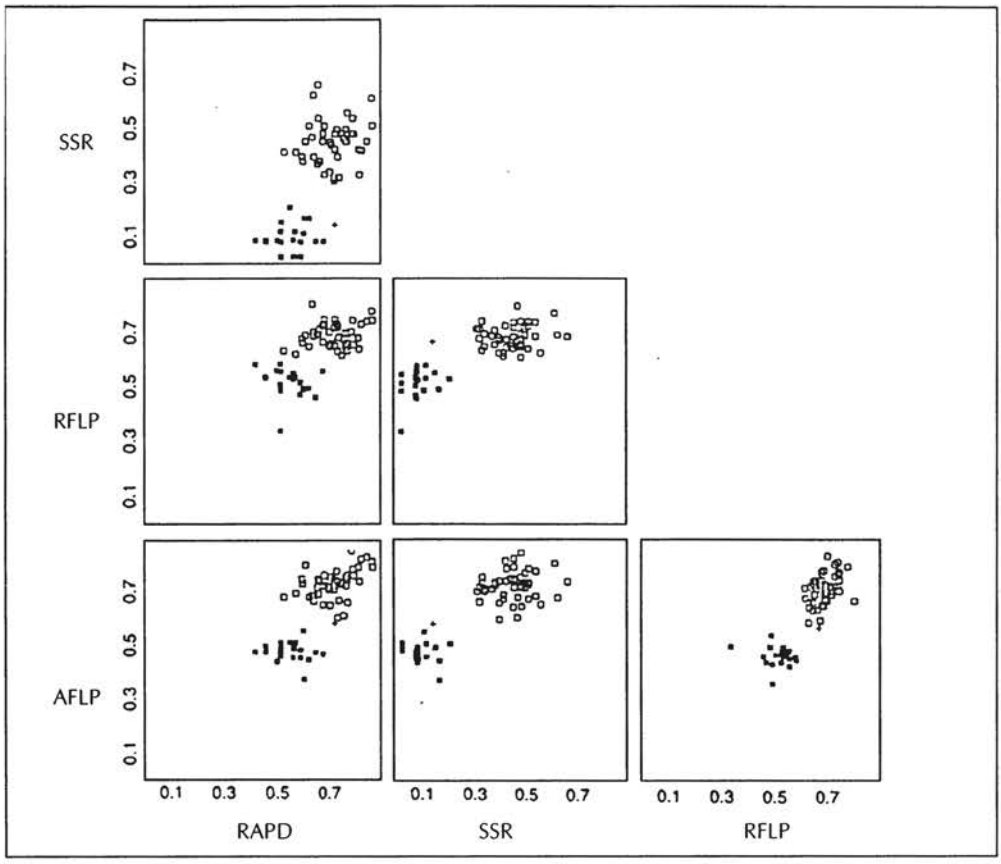


Figure 1. Relation à deux dimensions entre les similarités génétiques calculées à partir des RFLP, des AFLP, des microsatellites (SSR) et des RAPD chez le soja. Les carrés blancs représentent les similarités entre accessions de l'espèce Glycine max, *les croix représentent les similarités entre les accessions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

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

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

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

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

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

LA RELATION AVEC LA GÉNÉALOGIE ET LES CARACTÈRES MORPHOLOGIQUES

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

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

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

Référence	Espèce	Technique	Test*	N**	Correlation	Seuil *(P)*
Cox *et al.* (1985)	soja	Isoenzymes	Cr	27	0,15	< 0,01
				32	0,40	< 0,01
				39	0,45	< 0,01
O'Donoughue *et al.* (1994)	avoine	RFLP	Tm	55	0,32	< 0,001
Graner *et al.* (1994)	orge	RFLP	Cr	21	0,21	< 0,05
				17	0,42	< 0,01
Gerdes et Tracy (1994)	maïs	RFLP	Cp	42	0,54	< 0,01
		Isoenzymes	Cp	31	0,32	< 0,01
Tinker *et al.* (1993)	orge	RAPD	Cr	27	0,51	< 0,01
Plaschke *et al.* (1995)	blé	Microsatellites	Cr	40	0,55	< 0,001

* Tm : test de Mantel ; Cp : coefficient de corrélation de Pearson ; Cr : coefficient de corrélation de rang.
** Nombre d'individus utilisés pour calculer la corrélation.

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

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

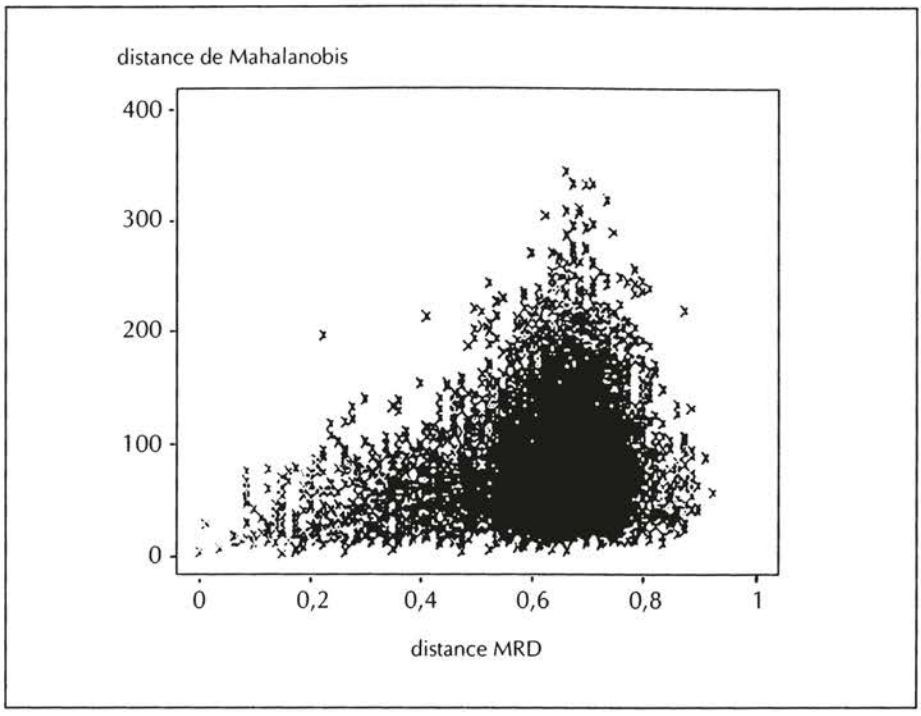


Figure 2. Relation entre la similarité (distance de Rogers, MRD) calculée à partir des marqueurs RFLP et la distance calculée à partir de données morphologiques (distance de Mahalanobis) chez le maïs, d'après DILLMANN et al. *(1997).*

La mise en œuvre pratique des techniques de marquage

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

La facilité de mise en œuvre

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

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

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

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

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

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

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

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

Le coût

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

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

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

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

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

Conclusion

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

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

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

Annexe

Description des techniques moléculaires

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

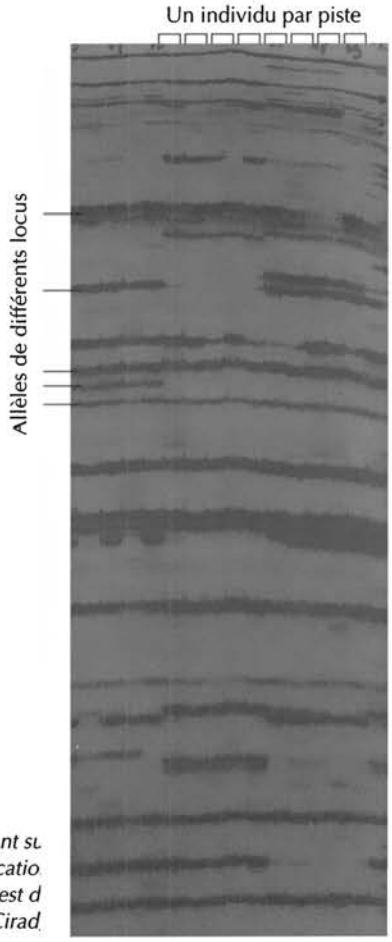
Les textes qui accompagnent la description des techniques comportent un certain nombre de mots signalés par un astérisque. Ces termes sont expliqués en utilisant un vocabulaire courant (quoique...) dans le glossaire (p. 35). Le site Internet de l'Institut Pasteur (http://www.pasteur.fr/other/biology/francais/bio-docs-fr.html)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

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

AFLP *(amplified fragment length polymorphism)*

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

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

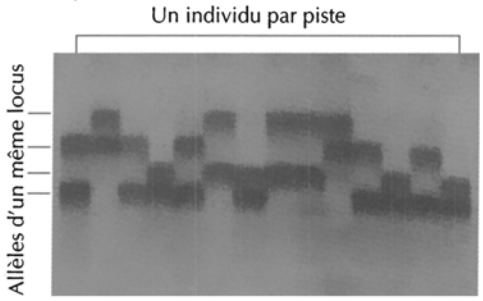


Autoradiographie d'un gel d'acrylamide dénaturant su lequel ont été séparés des fragments issus d'une amplificatio radioactive AFLP. La longueur totale du gel est d 40 centimètres. (photo R. Purba-Cirad

Microsatellites

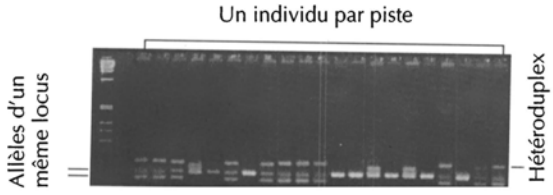
Après extraction*, l'ADN est utilisé tel quel lors d'une amplification PCR*, qui utilise deux amorces* définies après séquençage* dans les régions flanquantes* d'un site microsatellite (répétition de séquences simples, SSR). Les fragments produits par cette amplification sont séparés par électrophorèse*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

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

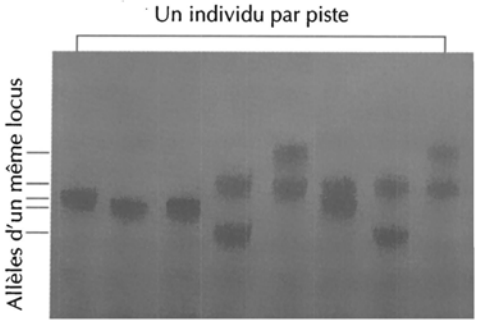


Autoradiographie d'un gel long (40 centimètres) d'acrylamide dénaturant après amplification par PCR radioactive d'un locus microsatellite d'hévéa. (photo M. Rodier-Goud et M. Seguin-Cirad).

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



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



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

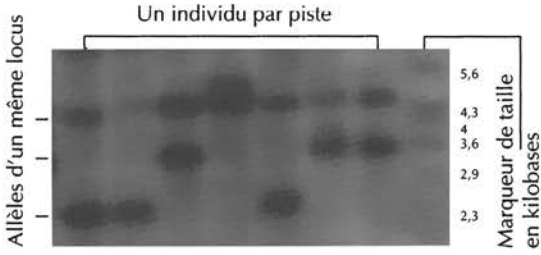
RAPD *(random amplified polymorphic DNA)*

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

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

RFLP *(restriction fragment length polymorphism)*

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



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

Glossaire

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Homologue : voir séquence homologue.

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

Hydrolyse : voir endonucléase.

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

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

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

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

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

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

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

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

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

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

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

Références bibliographiques

BEER S.C., GOFFREDA J., PHILLIPS T.D., MURPHY J.P., SORRELLS M.E., 1993. Assessment of genetic variation in *Avena sterilis* using morphological traits, isozymes and RFLPs. Crop Science, 33 : 1386-1393.

BENNETT M.D., LEITCH I., 1995. Nuclear DNA amounts in angiosperms. Annals of Botany, 76 : 113-176.

BENNETT M.D., SMITH J.B., 1991. Nuclear DNA amounts in angiosperms. Philosophical Transactions of the Royal Society of London, Series B, 334 : 309-345.

CAETANO-ANOLLES G.B., GRESSHOFF P.M., 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotides primers. Bio/Technology, 9 : 553-557.

CHO Y.G., BLAIR M.W., PANAUD O., MCCOUCH S., 1996. Cloning and mapping of variety-specific rice genomic DNA sequences: amplified fragment length polymorphisms (AFLP) from silver-stained polyacrylamide gels. Genome, 39 : 373-378.

COX T.S., KIANG Y.T., GORMAN M.B., RODGERS D.M., 1985. Relationships between coefficient of parentage and genetic similarity indices in the soybean. Crop Science, 25 : 529-532.

VAN DEYNE A.E., SORELLO M.E., PARK W.D., AYRES N.M., FU H., CARTINHOUR S.W., PAUL E., MCCOUCH S.R., 1998. Anchor probes for comparative mapping of grass genera. Theoretical and Applied Genetics, 97 : 356-369.

DILLMANN C., BAR-HEN A., GUERIN D., CHARCOSSET A., MURIGNEUX A., 1997. Comparison of RFLP and morphological distances between maize, *Zea mays* L., inbred lines: consequences for germplasm protection purposes. Theoretical and Applied Genetics, 95 : 92-102.

DOLEZEL J., 1991. Flow cytometric analysis of nuclear DNA content in higher plants. Phytochemical Analysis, 2 : 143-154.

DOS SANTOS J.B., NIENHUIS J., SKROCH P., TIVANG J., SLOCUM M.K., 1994. Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica oleracea* L. genotypes. Theoretical and Applied Genetics, 87 : 909-915.

DUBREUIL P., CHARCOSSET A., 1998. Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. Theoretical and Applied Genetics, 96 : 577-587.

EFRON B., TIBSHIRANI R., 1986. Bootstrap methods for standart errors, confidence intervals, and other measures of statistical accuracy. Statistical Science, 1 : 54-77.

ENGQVIST G.M., BECKER H.C., 1994. Genetic diversity for allozymes, RFLPs and RAPDs in resynthesized rape. *In :* Biometrics in plant breeding: applications of molecular markers ; IXth meeting of the EUCARPIA section biometrics in plant breeding, J. Jansen éd., Wageningen, Pays-Bas, EUCARPIA.

GERDES J.T., TRACY W.F., 1994. Diversity of historically important sweet corn inbreds as estimated by RFLPs, morphology, isozymes and pedigree. Crop Science, 34 : 26-33.

GRANER A., LUDWIG W.F., MELCHINGER A.E., 1994. Relationships among European barley germplasm. 2. Comparison of RFLP and pedigree data. Crop Science, 34 : 1199-1205.

HALLDEN C., HANSEN M., NILSSON N.O., HJERDIN A., SALL T., 1996. Competition as a source of errors in RAPD analysis. Theoretical and Applied Genetics, 93 : 1185-1192.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

MITCHELL M.J., QUIROS C., 1992. Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. Theoretical and Applied Genetics, 83 : 783-790.

MOHAN M., NAIR S., BHAGWAT A., KRISHNA T.G., YANO M., BHATIA C.R., SASAKI T., 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding, 3 : 87-103.

MORGANTE M., RAFALSKI A., BIDDLE P., TINGEY S., OLIVIERI A.M., 1994. Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome, 37 : 763-769.

N'GORAN J.A.K., LAURENT V., RISTERUCCI A.M., LANAUD C., 1994. Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. Heredity, 73 : 589-597.

O'DONOUGHUE L.S., SOUZA E., TANKSLEY S.D., SORRELLS M.E., 1994. Relationships among North American oat cultivars based on restriction fragment length polymorphisms. Crop Science, 34 : 1251-1258.

OLUFOWOTE J.O., XU Y., CHEN X., PARK W.D., BEACHELL H.M., DILDAY R.H., GOTO M., MCCOUCH S.R., 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome, 40 : 370-378.

ORTI G., PEARSE D.E., AVISE J.C., 1997. Phylogenetic assessment of length variation at a microsatellite locus. Proceedings of the National Academy of Sciences of the United States of America, 94 : 10745-10749.

PEAKALL R., SMOUSE P.E., HUFF D.R., 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. Molecular Ecology, 4 : 135-147.

PERRIER X., FLORI A., BONNOT F., 1999. Les méthodes d'analyse des données. *In :* Diversité génétique des plantes tropicales cultivées, P. HAMON *et al.* éd., Montpellier, France, Cirad, collection Repères, p. 43-76.

PLASCHKE J., GANAL M.W., RÖDER M.S., 1995. Detection of genetic diversity in closely related bread wheat using microsatellites markers. Theoretical and Applied Genetics, 91 : 1001-1007.

POWELL W., MORGANTE M., ANDRE C., HANAFEY M., VOGEL J., TINGEY S., RAFALSKI A., 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding, 2 : 225-238.

PRABHU R.R., WEBB D., JESSEN H., LUK S., SMITH S., GRESSHOFF P.M., 1997. Genetic relatedness among soybean genotypes using DNA amplification fingerprinting (DAF), RFLP, and pedigree. Crop Science, 37 : 1590-1595.

RAGOT M., HOISINGTON D.A., 1993. Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. Theoretical and Applied Genetics, 86 ; 975-984.

RISTERUCCI A.M., 1997. Coût des techniques de marquage moléculaire. *In :* Utilisation des marquers moléculaires en génétique et en amélioration des plantes ; séminaire de formation. Montpellier, France, Cirad-Biotrop.

RUS-KORTEKAAS W., SMULDERS M.J.M., ARENS P., VOSMAN B., 1994. Direct comparison of levels of genetic variation in tomato detected by a gaca-containing microsatellite probe and by random amplified polymorphic DNA. Genome, 37 : 375-381.