

CHAPITRE VI

MARQUEURS MOLECULAIRES

L'étude de génomes, la détermination et la comparaison de génotypes ne peuvent aujourd'hui être réalisées par séquençage systématique que pour quelques espèces modèles, essentiellement pour une raison de coût. Malgré l'émergence de nouvelles technologies permettant un séquençage plus rapide et moins cher - dont le système « genome sequencer 20 » de la société ROCHE qui permet en un seul cycle de 4h, de séquencer, d'ordonner et de comparer 20 millions de bases avec une exactitude voisine de 100 % ; l'alternative au séquençage choisie par de nombreux laboratoires à travers le monde reste l'utilisation de marqueurs moléculaires.

Les marqueurs moléculaires permettent de caractériser un génome de manière fiable, spécifique et rapide. Le marqueur a une position définie dans le génome et doit idéalement présenter les caractéristiques suivantes :

- 1) Le marqueur doit être polymorphe, c'est-à-dire qu'il doit posséder plus d'un allèle au moins dans la population étudiée.
- 2) Le marqueur idéal est codominant, ce qui signifie qu'un hétérozygote peut être différencié de l'homozygote au locus en question.
- 3) Il est non épistatique, c'est-à-dire que le génotype peut être lu à partir de son phénotype sans influence du génotype des autres locus. Il y a une absence d'interactions intra et inter locus.
- 4) Le marqueur est neutre, une modification des locus marqueurs n'a pas d'autres effets phénotypiques que ceux qui permettent de déterminer son génotype.
- 5) Le génotype peut être inféré à partir du phénotype quel que soit le milieu, un « bon » marqueur moléculaire est donc insensible au milieu.

La notion de polymorphisme est omniprésente dans les marqueurs moléculaires. La différence entre individus d'une même espèce réside dans le polymorphisme du génome, certaines séquences étant conservées, d'autres présentant des variations de séquence.

Par marqueurs moléculaires on peut suivre et contrôler la présence du gène d'intérêt chez n'importe quel individu sans recours à l'observation phénotypique du caractère. Ils peuvent être utilisés tout au long d'une expérimentation et sont observables à n'importe quel stade de développement de la plante et sur n'importe quel organe (l'information génétique de la plante est contenue en totalité dans toutes les cellules).

Ces marqueurs se trouvent à des endroits bien précis du génome. Ils sont utilisés pour localiser un gène particulier ou vérifier si un individu a hérité d'une caractéristique particulière d'un organisme parent. Les marqueurs moléculaires sont identifiables et situés à

des emplacements spécifiques du génome, et associés à la transmission d'une caractéristique ou d'un gène lié.

Sur le plan génétique, on peut considérer d'une part les techniques fournissant des marqueurs codominants et révélés individuellement (les « bons » marqueurs génétiques), d'autre part celles qui fournissent en « masse » des marqueurs dominants (tabl.). Cette séparation est certes simplificatrice (on peut trouver des marqueurs spécifiques de locus et dominants, etc.) mais elle correspond tout de même à deux types majeurs d'utilisation des marqueurs.

Critère génétique	Critère moléculaire			
	Séquence (et insertion-délétion ^a)		Nombre de répétitions dans les ADN répétés	
	Différence recherchée	Technique	Taille de l'unité de répétition	Technique
Codominants et révélés individuellement	Site d'enzyme de restriction (ER)	- RFLP - PCR ciblée puis CAPS	1 à 4 nucléotides (microsatellites)	PCR ciblée puis électrophorèse en acrylamide ou agarose
	Conformation	PCR ciblée puis SSCP		
	Stabilité	PCR ciblée puis D/TGGE		
Dominants et révélés « en masse » (« empreintes génétiques »)	Site d'hybridation d'une amorce arbitraire	MAAP : - RAPD - AP-PCR - DAF	1 à 4 nucléotides (microsatellites)	ISSR (amorce microsatellite + quelques bases arbitraires)
	Sites ER et amorce arbitraire	- AFLP - tec MAAP	5 à > 100 nucléotides (minisatellites)	Southern avec sonde minisatellite

Sur le plan moléculaire, on peut classer le polymorphisme en trois catégories : le polymorphisme de séquence, d'insertion-délétion, et de nombre d'unités de répétitions dans les régions répétées. Il n'existe pas de techniques développées spécifiquement pour le polymorphisme d'insertion-délétion, qui est de toute façon révélé par les techniques de mise en évidence du polymorphisme de séquence. Nous considérerons donc d'une part les variations de séquence, en mentionnant le cas échéant les variations d'insertion-délétion, d'autre part les variations de nombre d'unités de répétitions. En croisant les critères génétique et moléculaire, on définit les quatre familles de marqueurs présentées dans le tableau suivant :

1) Les marqueurs RFLP

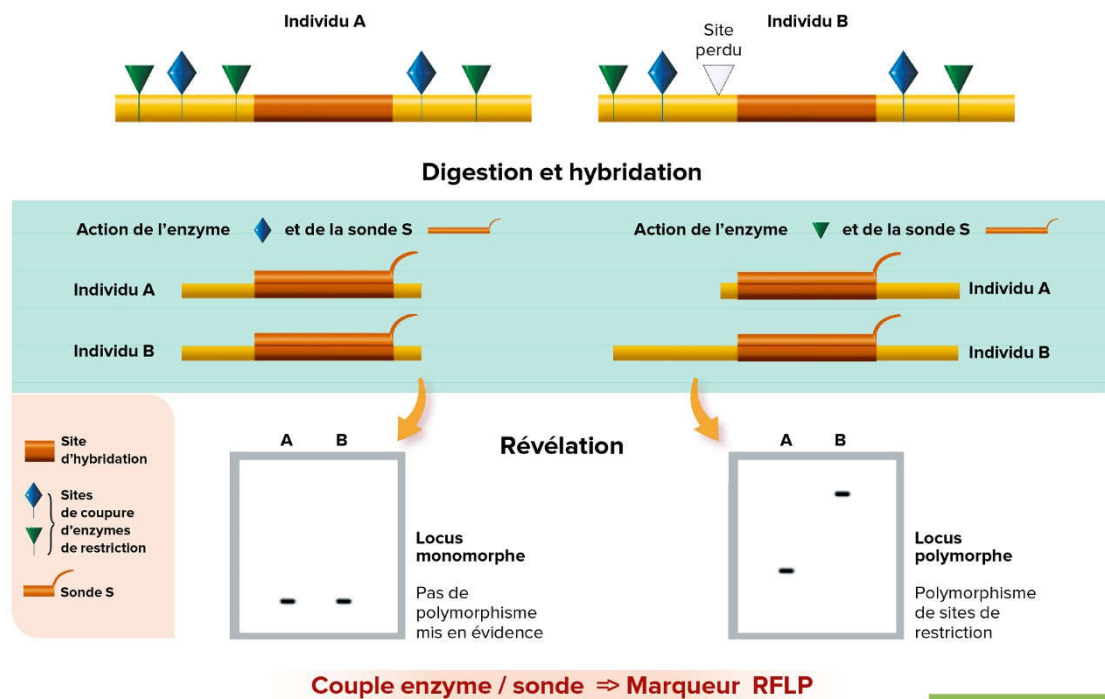
La « RFLP » (développée par Botstein et al. 1980) ou Restriction Fragment Length Polymorphism repose sur des enzymes de restriction et la variabilité de taille des fragments d'ADN obtenus via ces enzymes de restriction. Une enzyme de restriction coupe l'ADN dans des séquences en palindrome appelées sites de restriction.

Le site de restriction est spécifique de l'enzyme qui reconnaît des séquences de 4, 6, 8 ou parfois plus de paires de bases. Un grand nombre de fragments d'ADN peut être obtenu par la digestion d'un génome par une enzyme de restriction : « une enzyme ayant un site de

reconnaissance de 6 bases va couper l'ADN toutes les 4096 bases en moyenne. Un génome de 10^9 bases va donc produire environ 25.000 fragments de longueurs variables ».

Nous observerons donc un polymorphisme de longueur de fragment en fonction de la présence ou l'absence d'un site de restriction.

Les marqueurs RFLP



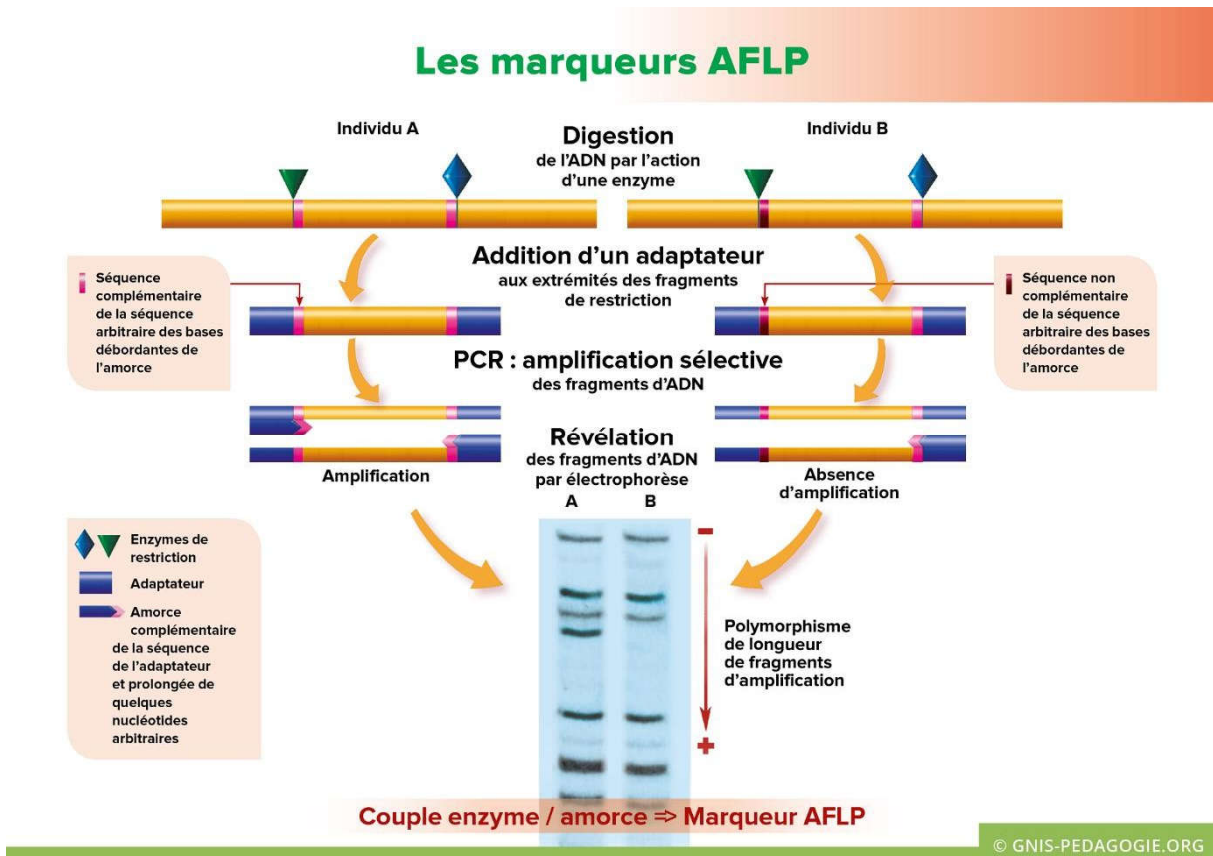
© GNIS-PEDAGOGIE.ORG

2) L'AFLP ou Amplified Fragment Length Polymorphism, (Vos et al, 1995)

Cette technique combine la technique de la RFLP à la PCR. L'AFLP utilise ainsi les enzymes de restriction et l'hybridation d'amorce PCR. L'ADN est clivé par deux enzymes de restriction (exemple Eco RI et Mse I). Des adaptateurs de séquences connues (20pb) sont ajoutés de part et d'autre des fragments d'ADN. Les bouts ajoutés constituent un template pour la fixation d'amorces PCR.

On amplifie ainsi les fragments d'ADN via PCR. Dans une deuxième étape dite sélective, on utilisera les mêmes amorces mais prolongées à l'extrémité 3' de quelques nucléotides arbitraires (**1 à 3**).

Cette étape permet de n'amplifier que les fragments complémentaires des bases arbitraires entraînant une réduction du nombre de fragments amplifiés. Une centaine de fragments sont finalement révélés en gel d'acrylamide après électrophorèse. Ceux-ci correspondent aux fragments dont le polymorphisme provient des sites de restriction mais également des sites d'hybridation des bases arbitraires.

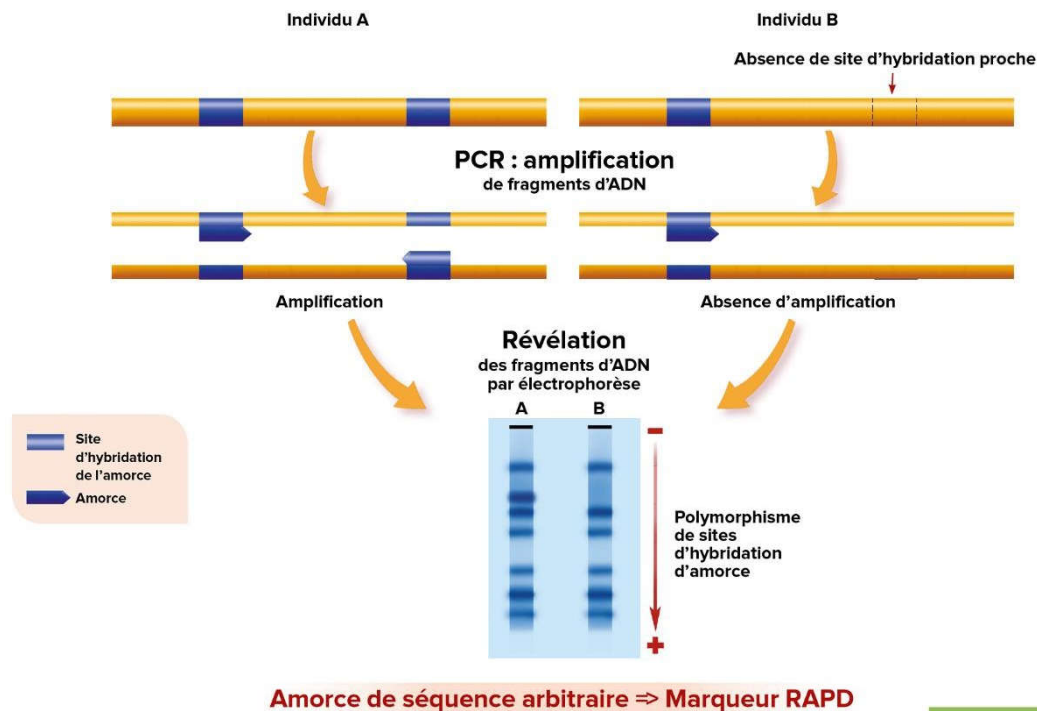


3) Les marqueurs RAPD (Williams et al., 1990).

Elle consiste en l'amplification par PCR de fragments de l'ADN génomique en utilisant des amorces arbitraires de taille courte (10 pb). Une amorce RAPD permet généralement l'amplification d'une dizaine de fragments correspondant à des locus dominants. Les produits d'amplification sont généralement visualisés par électrophorèse sur gel d'agarose. Le polymorphisme décelé est dû à des mutations soit dans les régions amplifiées soit au niveau des sites de fixation des amorces.

Cette technique est simple, rapide et ne nécessite ni un marquage radioactif ni une connaissance préalable de la séquence nucléotidique. Néanmoins, la RAPD manque de reproductibilité puisqu'elle est très sensible à la concentration de l'ADN et aux conditions d'amplification.

Les marqueurs RAPD



© GNIS-PEDAGOGIE.ORG

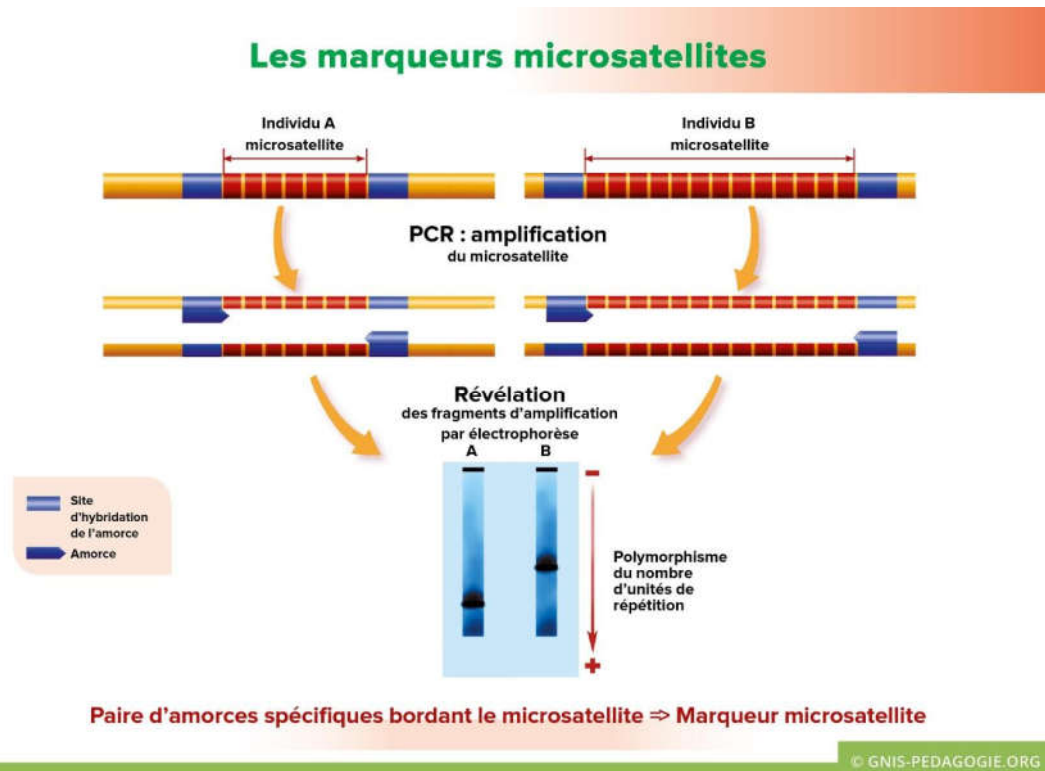
4) Les microsatellites

Les microsatellites ou SSR (simple séquence repeat) appelés aussi STR (short tandem repeat) sont abondamment dispersées dans le génome et sont très variable dans leur longueur entre les individus polymorphe, par exemple le microsatellite le plus connu est la séquence (CA)_n, où n varie de 5 à plus de 50, en d'autres termes cette séquence peut être répétée en tandem de 5 à 50 fois. Le microsatellite (CA)_n peut être retrouvé dans le génome humain chaque 10,000 bases, les chercheurs ont trouvés des milliers de microsatellite (CA)_n dans le génome.

On utilisant des amorces complémentaires aux séquences flanquant les microsatellites (CA)_n spécifiques, des microsatellites particuliers peuvent être amplifiés par PCR. En d'autres termes, les amorces ne vont amplifier seulement un microsatellite particulier mais pas les milliers d'autres qui sont dispersées dans le génome.

Quand la paire d'amorce de la PCR amplifie un site unique dans un groupe de chromosome, la région amplifiée est appelée **sequence tagged site (STS)**. Les **mini-satellites** ou **VNTR** (Variable Number of Tandem Repeat) sont constitués d'une unité de répétition (6 à 80 bp) de longueur et la taille variant de 1 kbp à 20 kbp. Un exemple des mini-satellites est l'ADN du télomère humain dans les cellules spermatiques l'unité de répétition est de 6 pb et la taille du télomère est de 15kpb.

Les séquences répétées en tandem sont appelées satellite parce qu'elles séparent loin du reste de l'ADN chromosomique durant la centrifugation.



5) Les SNP

Les SNP (prononcés « snip ») ou Single Nucléotide Polymorphism constituent la majorité du polymorphisme du génome (90% du polymorphisme du génome humain). Les SNP sont des mutants naturels qui ne concernent qu'un seul nucléotide. On observe un SNP lors du changement d'une base en une autre : transition ou transversion (les insertions et délétions d'une seule base sont également reconnues comme SNP).

En comparant deux génomes humains pris au hasard, on trouve approximativement un SNP toutes les **1000pb** dans les régions non codantes et un SNP toutes les **3000pb** dans les régions codantes.

Comparés aux autres marqueurs moléculaires, les SNP présentent l'avantage d'une répartition homogène dans tout le génome et d'être excessivement nombreux. Ce nombre élevé de SNP permet la création de cartes génétiques de haute densité.

Au-delà de l'intérêt pour le marqueur moléculaire, les SNP permettent l'identification de gènes et la mise en évidence de liens entre la présence de SNP et un phénotype donné. Les SNP s'inscrivent ainsi en génomique fonctionnelle pour l'étude de gènes candidats. Le gène candidat étant un gène susceptible d'être impliqué dans une fonction d'intérêt que l'on va pouvoir étudier grâce aux SNP qui le caractérisent.

On a pu ainsi montrer que la présence de certains SNP caractérisait une maladie ou une disposition à développer une maladie (maladie d'Alzheimer, Martin et al, 2000 ou le diabète type II, Horikawa et al, 2002). La présence d'un SNP dans une région régulatrice,

promotrice, peut aussi affecter la transcription d'un gène. La majorité des SNP n'ont cependant pas d'effet direct sur le phénotype.