

les marqueurs moléculaires en génétique

D. de Vienne, éd.

et biotechnologies végétales



MIEUX COMPRENDRE

 **INRA**
EDITIONS

les marqueurs moléculaires en génétique et biotechnologies végétales

D. de Vienne, éd.

MIEUX COMPRENDRE

Ouvrages parus dans la même collection :

Assimilation de l'azote chez les plantes.

Aspects physiologique, biochimique
et moléculaire

Jean-François MOROT-GAUDRY (éd)
1997, 422 p.

L'eau dans l'espace rural.

Production végétale et qualité de l'eau

C. RIOU, R. BONHOMME, P. CHASSIN, A. NEVEU,
F. PAPY (éds)
1997, 414 p.

La pomme de terre

P. ROUSSELLE, Y. ROBERT et J.-C. CROSNIER (éds)
1996, 640 p.

**Vie microbienne du sol
et production végétale**

Pierre DAVET
1996, 380 p.

Nutrition des ruminants domestiques

R. JARRIGE, Y. RUCKEBUSH,
C. DEMARQUILLY, M.-H. FARCE,
et M. JOURNET (éds)
1995, 921 p.

**Sols caillouteux
et production végétale**

Raymond GRAS
1994, 178 p.

Biologie de la lactation

Jack MARTINET,
et Louis-Marie HOUBEINE
1993, 587 p.

**Amélioration des espèces végétales
cultivées.**

Objectifs et critères de sélection

André GALLAIS et Hubert BANNEROT
1992, 768 p.

**La régression non linéaire :
méthodes et applications en biologie**

Sylvie HUET, Emmanuel JOLIVET,
et Antoine MESSÉAN
1992, 250 p.

**L'épidémiologie en pathologie
végétale : mycoses aériennes**

Frantz RAPILLY
1991, 318 p.

**Principes d'amélioration génétique
des animaux domestiques**

Francis MINVIELLE
1990, 211 p.

**Cytogénétique
des mammifères d'élevage**

Paul C. POPESCU
1989, 114 p.

**Les oligo-éléments
en agriculture et élevage**

Yves COÏC et Marcel COPPENET
1989, 114 p.

Éléments de virologie végétale

(épuisé)
Pierre CORNUET
1987, 208 p.

1^{re} édition © CNED-AUPELF UREF, 1995

2^e édition revue et augmentée © INRA, 1998

ISBN : 2-7380-0776-7

ISSN : 1144-7605

Le code de la propriété intellectuelle du 1^{er} juillet 1992 interdit la photocopie à usage collectif sans autorisation des ayants droit. Le non-respect de cette disposition met en danger l'édition, notamment scientifique. Toute reproduction, partielle ou totale, du présent ouvrage est interdite sans autorisation de l'éditeur ou du Centre français d'exploitation du droit de copie (CFC), 3 rue de Hautefeuille, Paris, 6^e.

Préface

La Science progresse-t-elle essentiellement par l'acquisition de nouveaux outils ou par l'émergence de nouveaux concepts ? Ce débat entre *l'Homo faber* et *l'Homo sapiens* peut se nourrir de nombreux exemples illustrant la pertinence de ces deux points de vue et surtout la dialectique en œuvre entre ces deux sources de renouvellement des approches : à ceux qui mettent en avant l'attraction universelle, la théorie de l'évolution ou la relativité générale répondent ceux qui soulignent le rôle décisif des instruments d'optique (télescopes comme microscopes de tous types) ou des outils analytiques (spectrométrie de masse, électrophorèse, résonance magnétique...).

Dans ce débat, les marqueurs moléculaires fournissent une nouvelle illustration du rôle clé que peuvent avoir de nouveaux outils pour renouveler des disciplines riches, voire surabondantes en concepts théoriques mais qui souffraient d'une carence manifeste dans la capacité d'observation fine de leur objet : la génétique quantitative comme la génétique des populations ont en effet construit progressivement un appareil théorique qui permettait souvent d'expliquer un fait donné par plusieurs hypothèses non exclusives, qu'il était le plus souvent difficile de hiérarchiser. Le « génotype », dont ces disciplines parlaient d'abondance, pouvait être comparé à un puzzle dont on n'aurait disposé que de quelques morceaux et d'une image globale très floue. On comprend à travers cette analogie que les débats sur ce que représentait l'image aient pu être souvent animés !

Or, en quelques années, le généticien se retrouve potentiellement doté d'un nombre important de nouveaux morceaux du puzzle, cette nouvelle donne permettant à l'évidence de discriminer des hypothèses, voire peut-être de les éliminer toutes pour en construire de nouvelles, illustrant ainsi la dialectique de l'outil et du concept évoquée plus haut.

Cette révolution technologique a été extrêmement rapide : lorsque les « empreintes génétiques » ont été développées chez l'homme au milieu des années quatre-vingt, beaucoup pensaient que le coût et la complexité de cette technique allaient la limiter à la criminologie ou à des expertises très ponctuelles ; dix ans plus tard, il est difficile d'aborder la génétique d'une espèce animale ou végétale, aussi bien à des fins de description de sa biodiversité que d'amélioration génétique, sans s'appuyer sur les marqueurs moléculaires.

L'ouvrage dont Dominique de Vienne a conduit l'élaboration présente de manière extrêmement claire ces différentes approches et leurs champs d'application. Le pari était difficile : ces techniques sont en évolution rapide, présentent souvent des similitudes et la définition de leur « domaine de pertinence » nécessite une connaissance fine des problèmes à aborder. Il nous semble pleinement réussi : au-delà de son caractère technique, je suis sûr que cet ouvrage fera partager au lecteur l'enthousiasme actuel, teinté de vertige, des généticiens, qui voient peu à peu se révéler les traits d'un objet qu'ils manipulent depuis longtemps sans vraiment le connaître : le génome des espèces végétales et animales.

Bernard Chevassus-au-Louis
Jouy-en-Josas, le 10 juin 1997

Remerciements

La rédaction de ce document a été possible grâce au concours de nombreux collègues et étudiants. En particulier, les synthèses bibliographiques de M. Lorieux, S. Gerber, C. Plomion et L. Grivet nous ont été très utiles pour les chapitres 2 (Cartographie) et 4 (QTL). La relecture de tout ou partie du texte par Ph. Brabant, C. Damerval, C. Dillmann, J.-M. Elsen, F. Hospital, A. Leonard, D. Manicacci, A. Maurice, R. Petit, M. Zivy et les étudiants du DEA de Biologie, Diversité et Adaptation des Plantes Cultivées a permis de l'améliorer sur bien des points. B. Piégu et M. Le Guilloux ont assuré la préparation souvent délicate des figures, et D. Létang s'est occupée de la documentation et de la bibliographie. Que tous et toutes en soient ici très sincèrement remerciés.

Avant-propos

Bien qu'elle ne coïncide pas avec l'émergence d'une discipline nouvelle, l'utilisation intensive des marqueurs moléculaires, depuis une dizaine d'années, a d'ores et déjà bouleversé de nombreux secteurs de la biologie. A titre d'exemple, on peut citer le clonage positionnel, qui permet d'isoler un gène dont le produit n'est pas connu à partir de sa seule position dans le génome. Mais cet aspect, que la génétique médicale a largement médiatisé, ne doit pas masquer d'autres applications tout aussi fondamentales. Dans le domaine de l'évolution, la cartographie comparative permet ainsi de répertorier les remaniements génomiques qui sont intervenus au cours de la diversification d'un genre ou d'une famille ; en génétique formelle, les facteurs qui influencent le taux de recombinaison peuvent être analysés systématiquement ; en génétique des populations, la diversité et les flux de gènes peuvent être mesurés, etc. Mais la discipline la plus renouvelée par les marqueurs moléculaires est sans doute la génétique quantitative. Qualifiée naguère de « science d'ingénieur », car davantage utilisée par les sélectionneurs que par les chercheurs, elle connaît aujourd'hui des succès qui vont de l'analyse du déterminisme génétique de la variation morphologique à la sélection assistée par marqueurs, en passant par le développement de stratégies de caractérisation des gènes en cause (les fameux QTL). Un bain de jouvence pour une vieille dame...

Dans la mesure où l'ensemble des méthodes que recouvre l'expression « marqueurs moléculaires » pénètre des secteurs très variés de la génétique et des biotechnologies végétales, il n'a pu évidemment être question de présenter les bases des différents domaines concernés. Dans chaque cas nous avons voulu montrer ce que les marqueurs apportaient de spécifique, les questions nouvelles qu'ils permettaient d'aborder, et le cas échéant les problèmes qu'ils soulevaient. Le lecteur devra donc se reporter, si nécessaire, à des ouvrages de base en génétique (formelle, quantitative ou des populations). De bonnes notions de biologie moléculaire et de statistique seront également les bienvenues. Une exception a toutefois été faite pour la construction des cartes génétiques, dont nous avons détaillé la théorie, et ce pour deux raisons. D'une part les cartes sont à la base de la plupart des applications des marqueurs, si bien que des connaissances insuffisantes sur leur nature et leur construction auraient hypothéqué la compréhension de la suite de l'ouvrage. D'autre part, même si

les principes en sont bien connus depuis presque 70 ans, leur enseignement à l'Université s'était étiolé au fil des décennies faute d'applications nouvelles. Enfin l'usage de l'informatique offre des perspectives qui méritaient d'être soulignées.

L'ouvrage a donc été organisé de la manière suivante. Le premier chapitre détaille les différentes techniques de marquage moléculaire, en mettant l'accent sur les aspects génétiques, car ceux-ci déterminent le type d'utilisation que l'on peut en faire. La construction de cartes de liaisons génétiques fait l'objet du second chapitre, où les avantages et inconvénients des descendances les plus courantes sont précisés. Le cas particulier du marquage de gènes majeurs, notamment en vue du clonage positionnel, est abordé dans le troisième chapitre. La détection et les applications des QTL (locus contrôlant les caractères quantitatifs) sont traitées dans le quatrième chapitre, qui aborde également la question complexe de leur identification. Le cinquième chapitre souligne l'apport majeur des marqueurs moléculaires pour l'analyse de la structure et de l'évolution des populations naturelles. Enfin l'intérêt des marqueurs en sélection, tant pour les études de diversité que dans le contexte de la sélection assistée par marqueurs, est traité dans le dernier chapitre. Tout au long de l'ouvrage, nous nous sommes efforcés non seulement d'exposer en détail le principe des nombreuses possibilités offertes par les marqueurs, mais aussi de les illustrer systématiquement par des résultats bibliographiques pertinents. Sur le plan de la forme, nous avons cherché à donner une certaine unité à l'ensemble, tout en respectant l'esprit des différents chapitres, dû à la personnalité des auteurs comme à la spécificité des différents domaines abordés.

Nous espérons qu'ainsi cet ouvrage répondra aux besoins des étudiants comme des enseignants, et qu'il sera un outil précieux pour tous les chercheurs ayant recours aux marqueurs, quelle que soit leur problématique.

Dominique de Vienne,
Gif-sur-Yvette, le 3 février 1997

Table des matières

Introduction	13
Définitions	13
Qu'est-ce qu'un « bon » marqueur génétique ?	13
1. Les principales sources de marqueurs moléculaires	15
D. de Vienne et S. Santoni	
Critères de classification	15
Marqueurs codominants révélés individuellement	16
Polymorphisme de séquence	16
<i>Différences au niveau de sites d'enzymes de restriction :</i>	
<i>la technique de RFLP</i>	17
<i>Un cas particulier de RFLP : les CAPS</i>	27
<i>Différences de conformation : la SSCP</i>	27
<i>Différences de stabilité : la D/TGGE</i>	30
Polymorphisme de nombre d'unités de répétitions : les microsatellites .	31
Marqueurs dominants révélés « en masse » : les empreintes génétiques ..	35
Polymorphisme de séquence	35
<i>Différences de sites d'hybridation d'une amorce arbitraire, ou techniques de MAAP : RAPD, AP-PCR et DAF</i>	35
<i>Différences de sites de restriction et de sites d'hybridation d'une amorce arbitraire : AFLP™ et tecMAAP</i>	37
Polymorphisme de nombre d'unités de répétitions	39
<i>La technique d'ISSR</i>	39
<i>Les minisatellites</i>	40
La technique de RDA	40
Les marqueurs géniques : ADNc et protéines	41
Quels marqueurs pour quoi faire ?	43
Glossaire	44
2. Etablissement de cartes de liaisons génétiques	49
D. de Vienne	
La notion de distance génétique	50
La distance de Haldane	51

La distance de Kosambi et autres distances	52
Estimation du taux de recombinaison et tests de liaison	56
Comparaison des descendance les plus courantes	60
Populations donnant accès à la phase haploïde	60
<i>Haploïdes doublés</i>	60
<i>Mégagamétophyte des Gymnospermes</i>	60
<i>Populations issues de rétrocroisements (backcross)</i>	60
Populations F_2	61
Lignées recombinantes	63
Populations issues de parents non fixés	66
<i>Cas général</i>	67
<i>Le double pseudo-testcross</i>	68
Comparaison des divers types de descendance	68
<i>Pérennité</i>	68
<i>Estimation de la dominance</i>	69
<i>Précision</i>	69
Aspects fondamentaux des cartes génétiques	72
La saturation des cartes	72
Relativité des longueurs des cartes	73
Relations entre distances génétiques et distances physiques	74
Analyse de l'organisation des génomes	76
La cartographie comparée	76
3. Le marquage des gènes majeurs	81
D. de Vienne	
Les dispositifs utilisables pour le marquage de gènes majeurs	81
Les lignées quasi isogéniques	81
Les mélanges d'individus de même génotype : la BSA	83
Intérêt des marqueurs pour le clonage de gènes majeurs	85
Principe du clonage positionnel	85
La cartographie à haute résolution	86
Recherche du gène-cible dans les clones génomiques	86
4. La cartographie et la caractérisation des locus contrôlant la variation des caractères quantitatifs	89
D. de Vienne et Mathilde Causse	
Principe de la cartographie des locus à effets quantitatifs	92
Détection de QTL en considérant les marqueurs individuellement ...	95
<i>Recherche de QTL par analyse de variance dans une descendance F_2</i>	95
<i>Mesures de l'effet d'un QTL</i>	98
<i>Intérêts et limites de la détection de QTL sur marqueurs individuels</i> ..	100

Détection de QTL à partir de deux ou plusieurs marqueurs	100
<i>Les méthodes de cartographie d'intervalle</i>	100
<i>Les méthodes multimarqueurs</i>	102
Facteurs influençant la détection des QTL	103
<i>Puissance de détection et variance intraclasse</i>	103
<i>Effectif de population et densité de marqueurs</i>	104
<i>Choix des risques statistiques</i>	105
<i>L'écart à la normalité</i>	105
Avantages et inconvénients des descendance couramment utilisées pour la détection de QTL	105
<i>Comparaison F_2, backcross, haploïdes doublés et lignées recombinantes</i>	105
<i>Descendance issues de parents hétérozygotes</i>	107
<i>Descendance dont le fonds génétique a été partiellement fixé</i>	108
Le marquage des extrêmes	108
Bases génétiques et moléculaires de la variation des caractères quantitatifs	109
Données génétiques sur les QTL	109
<i>Des QTL sont (presque) toujours trouvés</i>	109
<i>Effets de dominance</i>	110
<i>Transgressions</i>	111
<i>Epistasie</i>	111
<i>Analyse génétique des corrélations entre caractères</i>	112
<i>Décomposition des caractères</i>	113
<i>Effet de l'environnement</i>	114
<i>Cartographie comparée de QTL</i>	114
Caractérisation des QTL	115
<i>Le clonage</i>	115
<i>Les gènes candidats</i>	116
Conclusion	118
5. Les marqueurs moléculaires en génétique des populations	119
A. Kremer	
Apports des marqueurs moléculaires relativement aux marqueurs enzyma- tiques	120
Le multiallélisme	120
L'augmentation du nombre de locus	121
L'accès au polymorphisme de l'ADN cytoplasmique	121
L'ordonnement des allèles	122
Analyse de la diversité moléculaire	123
Mesure de la divergence nucléotidique à partir de séquences	124
Mesure de la divergence nucléotidique à l'aide de techniques de marquage	125
Polymorphisme à l'intérieur d'une population	127
Expression du polymorphisme à l'intérieur d'une population	127
<i>Mesures prenant en compte le nombre de variants</i>	127
<i>Mesures prenant en compte la fréquence des variants</i>	127

Mesure de la diversité génétique à l'aide de différents marqueurs	129
<i>Au niveau allélique</i>	129
<i>Au niveau nucléotidique</i>	131
Différenciation entre populations	132
Expression générale de la différenciation	132
Mesure de la différenciation	134
<i>Niveau allélique</i>	134
<i>Niveau nucléotidique</i>	134
Flux de gènes	135
Taux de migration comparés entre graines et pollen	135
Origine des fondateurs lors de la colonisation	136
Conclusion	138
6. Intérêt des marqueurs en sélection	139
A. Charcosset et A. Gallais	
Apport des études de diversité à la sélection	139
Principales méthodes d'analyse de données	140
Relations entre divergence moléculaire, divergence phénotypique et parenté	141
Application des méthodes de classification : groupes d'aptitude à la combinaison et gestion des ressources génétiques	143
Relation entre divergence moléculaire et hétérosis	145
Apport des marqueurs pour la protection des obtentions végétales . . .	147
Perspectives	148
La sélection assistée par marqueurs	149
La construction de génotypes	149
<i>Le rétrocroisement assisté par marqueurs pour un caractère monogénique</i>	149
<i>Le rétrocroisement assisté par marqueurs pour un caractère polygénique</i>	151
<i>La sélection généalogique assistée par marqueurs</i>	152
La sélection assistée par marqueurs au niveau de populations	153
<i>La prédiction des valeurs génétiques additives et la sélection avec les marqueurs seuls</i>	154
<i>La sélection combinée phénotype + marqueurs</i>	155
<i>Comparaison avec la construction de génotypes</i>	158
<i>Choix entre utilisation des marqueurs et augmentation du nombre de répétitions</i>	158
Conclusion	159
Annexes	161
Références bibliographiques	169
Index	195

Introduction

Définitions

Dans cet ouvrage, le terme *marqueur* sera pris dans le sens de *marqueur génétique*, c'est-à-dire qu'il sera toujours synonyme de *locus marqueur*. Un locus marqueur est un locus polymorphe qui renseigne :

- sur le génotype de l'individu qui le porte ; c'est à ce titre que les marqueurs sont utilisés en génétique des populations ;
- sur le génotype d'un (de) locus voisin(s) ; les applications vont ici du clonage positionnel à la sélection assistée par marqueurs.

Les plus courants de ces marqueurs génétiques sont, selon une terminologie consacrée, les marqueurs *morphologiques*, les marqueurs *moléculaires* (au niveau de l'ADN) et les marqueurs *biochimiques* (isozymes, protéines). Ce dernier terme est malheureusement ambigu, puisqu'il désigne dans d'autres contextes des molécules dont la présence indique un stade de différenciation ou un état physiologique. Nous ne l'emploierons ici que dans l'acception génétique. Nous supposerons connus les grands principes de l'analyse des isozymes, et traiterons essentiellement des marqueurs moléculaires. Les marqueurs issus de l'électrophorèse bidimensionnelle des protéines (EBD) seront toutefois évoqués.

Qu'est-ce qu'un « bon » marqueur génétique ?

Un marqueur génétique « idéal » est :

- polymorphe : la « matière première » du généticien est la variabilité ;
- multiallélique ;
- codominant : l'hétérozygote présente simultanément les caractères des deux parents homozygotes ; il peut donc être distingué de chacun des homozygotes parentaux ;
- non épistatique : son génotype peut être « lu » à partir de son phénotype quel que soit le génotype aux autres locus. La codominance et la non-épistasie peuvent être respectivement définies comme l'absence d'interactions intra et interlocus ;

- « neutre » : les substitutions alléliques au locus marqueur n'ont pas d'autres effets phénotypiques (et donc éventuellement sélectifs) que ceux qui permettent de déterminer son génotype. Dans leur très grande majorité, les polymorphismes moléculaires sont neutres ;
- insensible au milieu : le génotype peut être inféré à partir du phénotype quel que soit le milieu.

Les marqueurs morphologiques répondent mal à ces critères. Peu polymorphes, en général dominants, ils interfèrent souvent avec d'autres caractères, et peuvent être influencés par le milieu. En outre, même s'ils sont très nombreux chez certaines espèces (plusieurs centaines chez le riz ou le maïs), peu d'entre eux peuvent être conjointement polymorphes dans une descendance donnée.

En revanche les marqueurs biochimiques ou moléculaires ont, pour la plupart, toutes ces qualités. Les limitations majeures des isozymes sont le faible nombre de locus susceptibles d'être révélés (rarement plus de 30 à 40 locus chez le riz et le maïs, et tous ne sont pas polymorphes dans un fonds génétique donné), et le fait qu'il y ait une certaine spécificité d'organe : tous les enzymes ne sont pas présents ou actifs dans tous les organes. Les protéines polymorphes révélées en EBD peuvent être plus nombreuses, mais dépendent également de l'organe considéré. Au contraire, les marqueurs au niveau de l'ADN sont en nombre quasiment illimité et sont indépendants du stade ou de l'organe analysé, puisque l'ADN est le même dans tous les tissus. De plus ils ont l'avantage d'être plus directement utilisables pour les applications ultérieures en biologie moléculaire.

Les principales sources de marqueurs moléculaires

D. de Vienne et S. Santoni

Critères de classification

De nombreuses techniques de marquage moléculaire sont aujourd'hui disponibles, et de nouvelles sont très régulièrement publiées (voir glossaire p. 44). Le profane s'y perd d'autant plus que les sigles sont nombreux, parfois redondants, et la plupart du temps indéchiffrables (même traduits en français). L'ambition de ce chapitre est non seulement de décrire le principe de ces techniques, mais aussi d'y mettre un peu d'ordre. Plusieurs critères de classification sont possibles, mais nous avons choisi de privilégier les critères génétique et moléculaire, plutôt que technique ou historique.

Sur le plan génétique, on peut considérer d'une part les techniques fournissant des marqueurs codominants et révélés individuellement (les « bons » marqueurs génétiques, selon les critères définis p. 13), d'autre part celles qui fournissent en « masse » des marqueurs dominants (tabl. 1). 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.

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 1.

Tableau 1. Classification des techniques de marquage moléculaire (pour la signification des sigles, voir le glossaire p. 44).

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

a. En dehors du cas particulier des ADN répétés (colonne de droite), il n'y a pas de technique *spécifique* pour révéler le polymorphisme d'insertion-délétion : celui-ci est mis en évidence par les techniques de révélation des différences de séquences. Lorsqu'un polymorphisme est observé, on ne peut donc pas savoir quelle est son origine sans expériences complémentaires. Sur le plan génétique, cette ambiguïté n'a aucune importance, l'essentiel étant d'avoir accès à des locus polymorphes, quelle que soit l'origine de ce polymorphisme.

Marqueurs codominants révélés individuellement

Polymorphisme de séquence

Le polymorphisme de séquence peut naturellement être mis en évidence par séquençage de fragments homologues. Mais en dépit des progrès de l'automatisation, cette technique reste lourde, notamment lorsque les taux d'erreurs acceptables sont très faibles. Le séquençage d'allèles ne constitue donc pas aujourd'hui une méthode de routine lorsque l'on s'intéresse conjointement à un grand nombre de locus. On a donc recours à des méthodes indirectes, non exhaustives mais beaucoup plus rapides, fondées sur la détection de différences :

- de sites de restriction ;
- de conformation ;
- de stabilité ;
- de sites d'hybridation d'amorces oligonucléotidiques.

Les méthodes développées pour la détection des trois premiers types de différences fournissent des marqueurs codominants qui sont révélés individuellement.

Différences au niveau de sites d'enzymes de restriction : la technique de RFLP *

Les enzymes de restriction

Les enzymes de restriction (ou endonucléases de restriction) sont des enzymes qui coupent (on dit aussi qui « digèrent ») l'ADN en des sites spécifiques, appelés *sites de restriction*, comprenant le plus souvent un nombre pair de bases (4, 6, ou 8, parfois plus), arrangées en général en palindrome (tabl. 2). Ainsi, en supposant que les bases A, T, G, et C de l'ADN sont équiprobables, un enzyme ayant un site de reconnaissance de 6 bases va couper l'ADN toutes les 4096 bases *en moyenne* (4⁶). Un génome de 10⁹ bases va donc produire environ 250 000 fragments de longueurs variables. 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 utilisé. C'est cette spécificité qui est exploitée pour la mise en évidence du polymorphisme : une présence-absence de site de restriction entraîne un *polymorphisme de longueur de fragments* (fig. 1). Ce phénomène n'étant pas rare (le polymorphisme est une propriété fondamentale du vivant), la digestion de l'ADN

Tableau 2. Enzymes de restriction couramment utilisées pour la technique de RFLP.

Enzyme	Site de restriction
<i>Apa</i> I	5'... GGG CC C ... 3' 3'... C CC GGG ... 5'
<i>Cla</i> I	5'... AT CG AT ... 3' 3'... TA GC TA ... 5'
<i>Bam</i> HI	5'... G GATC C ... 3' 3'... C CTAG G ... 5'
<i>Dra</i> I ⁽¹⁾	5'... TTT AAA ... 3' 3'... AAA TTT ... 5'
<i>Eco</i> RI	5'... G AATTC ... 3' 3'... CTTAA G ... 5'
<i>Eco</i> RV ⁽¹⁾	5'... GAT ATC ... 3' 3'... CTA TAG ... 5'
<i>Hind</i> III	5'... A AGCTT ... 3' 3'... TTCGA A ... 5'
<i>Pst</i> I ⁽²⁾	5'... CTGCA G ... 3' 3'... G ACGTC ... 5'
<i>Pvu</i> II ⁽¹⁾	5'... CAG CTG ... 3' 3'... GTC GAC ... 5'
<i>Sma</i> I ⁽¹⁾	5'... CCC GGG ... 3' 3'... GGG CCC ... 5'
<i>Xba</i> I	5'... T CTAG A ... 3' 3'... A GATC T ... 5'
<i>Xho</i> I	5'... G AGCTC ... 3' 3'... CTC GA G ... 5'

(1) Contrairement aux autres enzymes, qui génèrent par coupure des « bouts cohésifs », *Dra* I, *Eco* RV, *Pvu* II et *Sma* I génèrent des « bouts francs ».

(2) Lorsque les cytosines en 5' sont méthylées, *Pst* I ne coupe pas ce site de restriction. Cette propriété est mise à profit pour la construction de banques génomiques enrichies en séquences peu ou pas répétées.

* Voir glossaire p. 44.

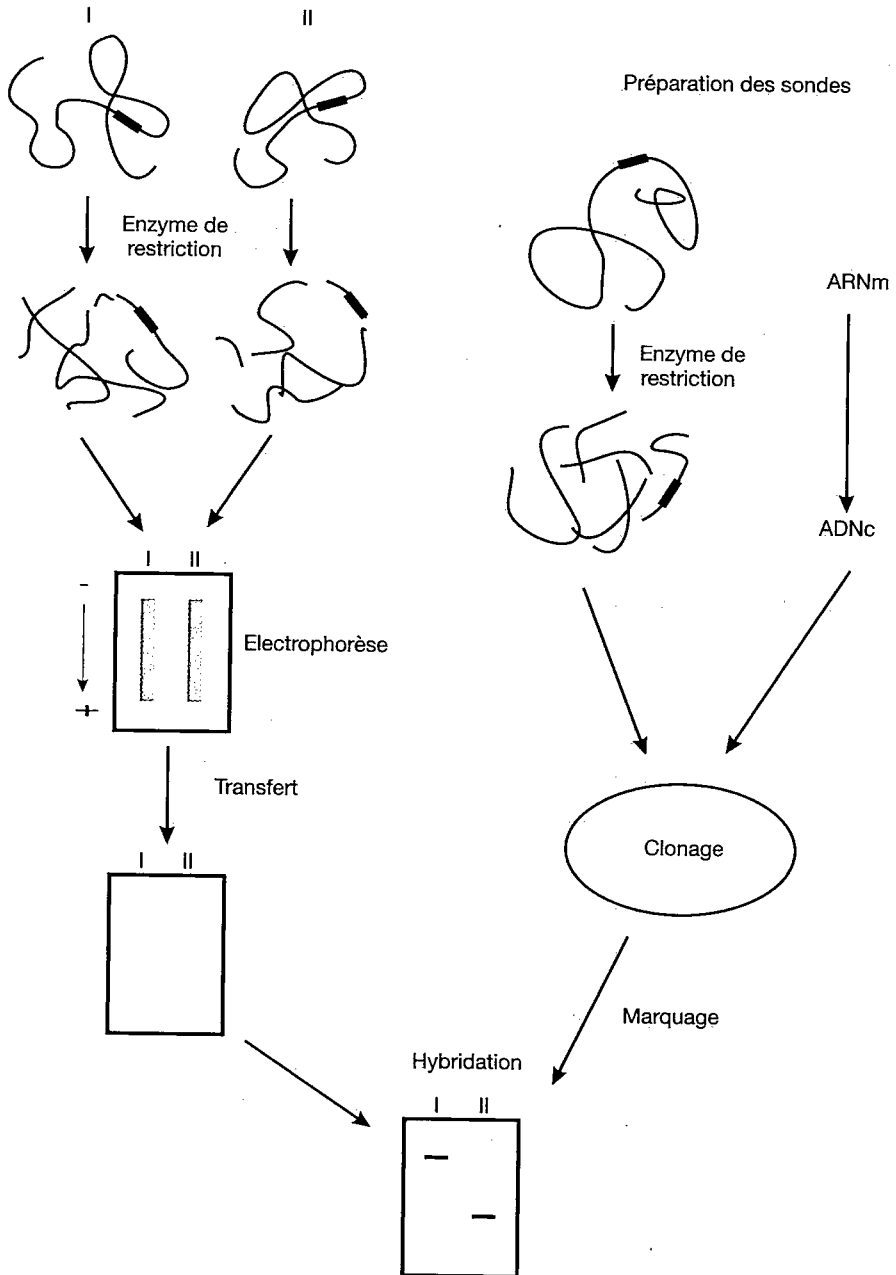


Figure 1. Principe de la révélation du polymorphisme de longueur des fragments de restriction (RFLP). I et II représentent deux individus. La molécule d'ADN est schématisée par un filament. Le rectangle gris sur l'ADN représente une région particulière reconnue par une sonde. Après digestion, le fragment portant cette région est plus long chez l'individu I que chez l'individu II. Cette différence de longueur sera révélée par hybridation de la sonde marquée sur l'ADN après migration (voir texte pour plus de détails). Cette sonde a donc révélé un locus biallélique (Botstein *et al.*, 1980).