



**BMT Guidelines (proj.8)**

**ORIGINAL:** anglais

**DATE:** 18 janvier 2007

**UNION INTERNATIONALE POUR LA PROTECTION DES OBTENTIONS VÉGÉTALES**  
GENÈVE

**PROJET**

**DIRECTIVES CONCERNANT LES PROFILS D'ADN :  
CHOIX DES MARQUEURS MOLÉCULAIRES  
ET  
CONSTRUCTION D'UNE BASE DE DONNÉES Y RELATIVE  
("DIRECTIVES BMT")**

*Document établi par le Bureau de l'Union*

*aux fins d'examen par les*

*Comité technique, à sa quarante-troisième session,  
qui se tiendra à Genève du 26 au 28 mars 2007*

*Comité administratif et juridique, à sa cinquante-cinquième session,  
qui se tiendra à Genève le 29 mars 2007*

## TABLE DES MATIÈRES

|  |           |
|--|-----------|
| <b>A. INTRODUCTION</b> .....   | <b>3</b>  |
| <b>B. PRINCIPES GÉNÉRAUX</b> .....   | <b>3</b>  |
| 1. Sélection de la méthode à appliquer aux marqueurs moléculaires .....      | 3         |
| 2. Choix des marqueurs moléculaires .....                                    | 4         |
| 2.1 Critères généraux .....  | 4         |
| 2.2 Critères destinés à des types spécifiques de marqueurs moléculaires..... | 4         |
| 3. Accès à la technologie.....   | 5         |
| 4. Matériel à analyser .....   | 6         |
| 4.1 Source du matériel végétal .....   | 6         |
| 4.2 Type du matériel végétal.....  | 6         |
| 4.3 Taille de l'échantillon.....   | 6         |
| 4.4 Échantillon d'AND de référence.....                                      | 7         |
| 5. Normalisation des protocoles analytiques.....                             | 7         |
| 5.1 Introduction .....   | 7         |
| 5.2 Critère de qualité.....  | 8         |
| 5.3 Phase d'évaluation .....   | 8         |
| 5.4 Notation des données moléculaires .....                                  | 9         |
| 6. Bases de données .....  | 10        |
| 6.1 Type de base de données.....   | 10        |
| 6.2 Modèle de bases de données.....  | 10        |
| 6.3 Dictionnaire de données .....  | 10        |
| 6.4 Liens entre les tableaux .....   | 11        |
| 6.5 Transfert des données vers la base de données .....                      | 12        |
| 6.6 Accès/appropriation des données .....                                    | 12        |
| 6.7 Analyse des données .....  | 13        |
| 6.8 Validation des bases de données .....                                    | 13        |
| 7. Résumé.....   | 13        |
| <b>GLOSSAIRE</b> .....   | <b>14</b> |
| Microsatellites, ou répétitions de séquences simples (SSR).....              | 14        |
| Polymorphismes nucléotidiques (SNP).....                                     | 14        |
| Séquences polymorphes amplifiées coupées (CAPS) .....                        | 14        |
| Régions amplifiées caractérisées par des séquences (SCARS) .....             | 15        |
| Valeurs du contenu de l'information polymorphe (PIC).....                    | 15        |
| Tresses.....   | 15        |
| Allèle à fréquence nulle.....  | 15        |
| Bandes à répétition .....  | 15        |

## A. INTRODUCTION

Le présent document (Directives BMT) offre des directives en vue de l'élaboration de méthodes harmonisées qui serviront à produire des données moléculaires de haute qualité destinées à diverses applications. Les directives BMT ont également pour but de permettre l'élaboration de bases de données contenant des profils moléculaires de variétés, qui peuvent être produits dans différents laboratoires à l'aide de diverses techniques. Ceci nécessite des marqueurs de haute qualité et oblige à reproduire des données à l'aide de ces marqueurs dans des situations où l'équipement et/ou les produits chimiques réactifs peuvent changer. En outre, il convient de prendre des précautions spécifiques en ce qui concerne la qualité des données entrées dans une base de données.

En ce qui concerne l'utilisation éventuelle de marqueurs moléculaires dans l'examen de la distinction, de l'homogénéité et de la stabilité, on trouvera à l'annexe du présent document la position actuelle de l'UPOV à ce sujet.

## B. PRINCIPES GÉNÉRAUX

### 1. Sélection de la méthode à appliquer aux marqueurs moléculaires

1.1 Les critères les plus importants dans la sélection d'une méthode sont les suivants :

- a) reproductibilité de la production de données entre laboratoires et plates-formes de détection (types d'équipement);
- b) possibilité de renouveler la manœuvre dans le temps;
- c) pouvoir de discrimination;
- d) possibilités de compilation dans des bases de données
- e) accessibilité de la méthode à appliquer.

1.2 À mesure que des progrès technologiques voient le jour et que de nouveaux équipements sont disponibles, il est important, pour assurer la continuité et la stabilité des bases de données, que l'interprétation des<sup>a</sup> données produites soit indépendante de l'équipement utilisé pour les produire. C'est le cas, par exemple, avec les données de séquences d'ADN. À l'origine, pour produire de telles données, on utilisait des amorces marquées de façon radioactive et des gels de séquences d'ADN. Aujourd'hui, cette manœuvre peut se faire à l'aide de teintures fluorescentes. Elle est alors suivie par une séparation sur des systèmes d'électrophorèse sur gel capillaire à débit élevé et à forte automatisation. Malgré ces différences, les données produites par le biais des diverses techniques coïncident les unes avec les autres, indépendamment des techniques utilisées pour les produire. Ceci peut s'appliquer également aux données produites, par exemple, par des microsatellites d'ADN (répétitions de séquences simples – SSR) ou des techniques de Polymorphismes nucléotidiques (SNP). Ces possibilités de répétition et de reproduction sont importantes à l'élaboration, l'utilisation et la longévité des bases de données. Ce n'est que par ce moyen qu'une base de données dont l'entretien est centralisé, dont les données sont vérifiées à partir de sources diverses, peut être élaborée à un prix abordable, de sorte que l'investissement important nécessaire au départ n'est effectué qu'une seule fois.

1.3 Les sortes de techniques moléculaires qui s'appliquent aisément aux profils des variétés sont limitées par le fait que les données doivent pouvoir se répéter, se reproduire et être cohérentes. Ainsi, si les diverses techniques de profils d'ADN à loci multiples ont été utilisées avec succès pour la recherche, il est difficile d'enregistrer dans bon nombre d'entre elles une quelconque codominance et il risque de s'avérer difficile de reproduire des configurations de bandes entre laboratoires utilisant des équipements différents. On estime que ces facteurs présentent des difficultés dans le cadre des profils de variétés. C'est pourquoi le présent document est axé essentiellement sur des considérations et des recommandations concernant des applications de SSR (microsatellites) bien définies et qui ont fait l'objet des recherches appropriées et, pour l'avenir, sur des informations relatives aux séquences (c'est-à-dire les SNP). D'autres techniques reposant sur des informations relatives aux séquences d'ADN, telles que les CAPS (séquences polymorphes amplifiées coupées) et les SCARS (régions amplifiées caractérisées par des séquences) sont elles aussi susceptibles de satisfaire les critères susmentionnés, mais leur utilisation dans le cadre des profils d'ADN des variétés végétales n'a pas encore été explorée.

## 2. Choix des marqueurs moléculaires

### 2.1 Critères généraux

Les critères généraux ci-après utilisés pour le choix d'un marqueur spécifique ou d'un ensemble de marqueurs spécifiques sont conçus pour s'appliquer aux marqueurs qu'importe leur utilisation. Il est cependant reconnu que certaines utilisations spécifiques risquent d'imposer l'utilisation de critères supplémentaires :

- a) valeur du niveau utile de polymorphisme (indiqué, par exemple, par la valeur du PIC (contenu de l'information sur le polymorphisme : voir Glossaire) obtenue sur la base d'un ensemble de variétés représentatives);
- b) possibilité de répétition des manœuvres au sein des laboratoires et de leur reproduction d'un laboratoire à l'autre en termes de notation des données;
- c) répartition connue des marqueurs à travers tout le génome (c'est-à-dire position cartographique), qui, bien que non indispensable, constitue une information utile pour éviter le risque de choisir des marqueurs qui pourraient être reliés;
- d) possibilité d'éviter, dans la mesure du possible, des marqueurs dont les allèles sont "nuls" (c'est-à-dire des allèles dont l'effet se manifeste par une absence d'un produit ACP au niveau de la molécule), ce qui, à nouveau, n'est pas indispensable, mais conseillé.

### 2.2 Critères destinés à des types spécifiques de marqueurs moléculaires

#### 2.2.1 *Marqueurs microsatellites (SSR)*

2.2.1.1 L'analyse des répétitions de séquences simples (SSR ou microsatellites : Voir glossaire) à l'aide de l'ACP (amplification en chaîne par polymérase), qui est aujourd'hui largement utilisée, présente plusieurs avantages

2.2.1.2 Les marqueurs SSR sont exprimés de façon codominante, sont généralement faciles à notifier (ou à enregistrer) et peuvent être facilement cartographiés. Il a été constaté qu'ils pouvaient être analysés par différents laboratoires, et (sous réserve qu'une attention suffisante soit portée aux conditions expérimentales), qu'ils sont robustes et reproductibles. En outre, ils

peuvent être analysés à l'aide de séquenceurs d'ADN automatisés, à débit élevé et non radioactifs, soit sur la base de l'électrophorèse sur gel, soit sur celle de l'électrophorèse capillaire, et plusieurs d'entre eux peuvent être analysés simultanément (multiplexage).

2.2.1.3 Pour une analyse par microsatellite efficace, il est essentiel de choisir des marqueurs de grande qualité. Pour cela, il convient de tenir compte, entre autres, des éléments ci-après :

- a) degré de répétition (production d'une série d'une ou de plusieurs bandes, différenciées en taille par une unité de répétition);
- b) pics (n+1); le Taq-polymérase ajoute souvent 1 bp à la fin d'un fragment. Ceci peut être évité grâce à l'utilisation d'amorces "tressées" (voir Glossaire);
- c) taille du produit d'amplification;
- d) séparation réelle entre les divers allèles des différents systèmes de détection;
- e) notation fiable et reproductible des allèles contenus dans des systèmes de détection différents;
- f) niveau de polymorphisme (nombre d'allèles détectés) entre les variétés (il est à noter que ceci nécessite l'analyse d'un nombre important de variétés);
- g) liaisons à éviter.

2.2.1.4 Pour la notation des SSR dans différents laboratoires et avec différents équipements de détection, il est indispensable que les allèles de référence (c'est-à-dire les ensembles de variétés) soient définis et inscrits dans toutes les analyses. Ces allèles de référence sont nécessaires car les normes pondérales moléculaires varient selon les divers systèmes de détection dont on dispose actuellement, en conséquence de quoi elles ne conviennent pas à l'identification des allèles.

2.2.1.5 Les amorces utilisées dans un laboratoire spécifique devraient être synthétisées par un fournisseur assuré, afin de réduire les risques de production de différents profils d'ADN à partir d'amorces synthétisées à l'aide de sources différentes.

## 2.2.2 *Polymorphisme nucléotidique (SNP)*

Les polymorphismes nucléotidiques (SNP : voir Glossaire) peuvent être détectés par séquençage d'ADN, technique courante qui révèle généralement des niveaux élevés de répétabilité dans le temps et de reproduction possible entre laboratoires. Ceci étant dit, la détection de SNP spécifiques s'effectue actuellement grâce à diverses techniques dont beaucoup ne sont pas encore courantes. De par leur nature, les SNP n'ont que deux états alléliques dans les plantes diploïdes, mais il peut en être autrement dans le cas des plantes polyploïdes dans lesquelles on constate des effets de dosage. Ainsi, la notation des SNP est relativement simple et fiable, ce qui devrait permettre de réduire ou d'éliminer bon nombre des problèmes susmentionnés. Ceci veut dire également qu'il faudra peut-être analyser un nombre important de marqueurs, soit indépendamment, soit sous forme de multiplexes, afin d'obtenir des profils efficaces et réels d'un génotype particulier.

## 3. Accès à la technologie

Certains marqueurs et matériel moléculaires sont disponibles au grand public. Cela dit, un investissement est important pour obtenir, par exemple, des marqueurs SSR de grande qualité et, en conséquence de quoi, il est probable que certains marqueurs et autres méthodes ou matériel fassent l'objet de droits de propriété intellectuelle. L'UPOV a mis au point des

directives en vue de l'utilisation de produits ou de méthodes qui font l'objet de droits de propriété intellectuelle (voir le document TPG/7/1, Annexe 3, GN 14) et il convient de les suivre aux fins des présentes directives. Il est recommandé de régler les questions concernant les droits de propriété intellectuelle avant d'entreprendre tout travail de développement.

#### **4. Matériel à analyser**

La source et le type de matériel ainsi que le nombre d'échantillons à analyser sont les questions essentielles à se poser en ce qui concerne le matériel à analyser.

##### **4.1 Source du matériel végétal**

Le matériel végétal à analyser devrait être un échantillon authentique et représentatif de la variété et, lorsque cela est possible, devrait être obtenu à partir de l'échantillon de la variété utilisée pour examen aux fins des Droits d'obtenteur et de l'enregistrement officiel. L'utilisation d'échantillons de matières soumis pour examen aux fins des droits d'obteneurs ou d'enregistrement officiel devra faire l'objet, selon le cas, d'une autorisation de l'autorité compétente, de l'obtenteur et/ou du conservateur. Lorsque cela est approprié, chaque plante à partir de laquelle est tiré un échantillon devrait pouvoir être retrouvée au cas où il s'avère a posteriori que certaines d'entre elles ne sont pas représentatives de la variété.

##### **4.2 Type du matériel végétal**

Le type du matériel végétal à échantillonner et la procédure d'échantillonnage à suivre en vue de l'extraction d'ADN dépendront, dans une large mesure, de la récolte ou de l'espèce végétale concernée. Par exemple, dans les variétés dont la propagation se fait par graines, celles-ci peuvent servir de sources d'ADN, alors que, dans le cas des variétés à propagation végétale, l'ADN peut être extrait à partir du matériel qui constitue la feuille. Quelle que soit la source du matériel, la méthode d'échantillonnage et l'extraction de l'ADN devraient être normalisées et référencées. En outre, il convient de vérifier si les méthodes d'échantillonnage et d'extraction produisent des résultats consistants par le biais de l'analyse de l'ADN.

##### **4.3 Taille de l'échantillon**

Il est indispensable que les échantillons pris pour analyse soient représentatifs de la variété.

###### *4.3.1 Variétés à propagation végétale*

En principe, un seul échantillon pourrait être analysé dans le cas de variétés à propagation végétale, dans la mesure où tous les échantillons devaient être identiques. Il est toutefois conseillé dans chaque cas d'analyser au moins des doubles d'échantillons. Si la moindre différence est constatée, il serait bon d'analyser des échantillons supplémentaires.

###### *4.3.2 Variétés autogames ou principalement autogames*

Il n'est pas toujours approprié de supposer que des variétés autogames ou principalement autogames sont homozygotes dans tous les loci, y compris les loci SSR ou SNP. L'hétérogénéité peut provenir, par exemple, d'une hétérozygote résiduelle, de la pollinisation croisée ou d'un mélange physique. C'est pourquoi il est généralement recommandé que plusieurs graines séparées soient analysées, car c'est ainsi que l'on pourra

constater une hétérozygotie quelconque. Cependant, il peut y avoir des raisons, notamment liées au coût, pour analyser un ensemble d'échantillons d'un nombre convenu d'individus pour représenter le profil ADN d'une variété donnée.

#### 4.3.3 Variétés à pollinisation croisée<sup>b</sup>

Des considérations semblables s'appliquent en principe aux variétés d'espèces à pollinisation croisée. On recommande en général que chaque graine/plante soit analysée individuellement car des différences entre des variétés peuvent résulter de différences de fréquences d'allèles (ou de bandes), ainsi que de la présence ou de l'absence d'allèles/bandes particuliers(ières). Il peut toutefois être justifié, du point de vue du coût notamment, d'analyser un échantillon global constitué d'un nombre convenu d'individus pour représenter le profil ADN d'une variété<sup>a</sup>.

#### 4.3.4 Hybrides

La méthode qu'il convient d'appliquer pour veiller à ce que les échantillons pris pour l'analyse d'hybrides soient représentatifs de la variété dépendra du type d'hybride. Il convient d'admettre que les variétés d'hybrides sont hétérozygotes dans les loci dont le codage sert aux marqueurs d'ADN, tout en restant uniformes phénotypiquement. Le nombre d'échantillons choisis aux fins d'analyse dépend précisément de la question qui est posée et du type de variété d'hybride évalué. À ce sujet, il convient d'étudier les renseignements concernant différents types de variétés d'hybrides, contenus dans le document TG/1/3 : "Introduction générale à l'examen de la distinction, de l'homogénéité et de la stabilité et à l'harmonisation des descriptions des obtentions végétales" (voir le chapitre 6.4.3).

#### 4.4 Échantillon d'AND de référence

Il est recommandé d'établir, conformément aux sections 4.1 à 4.3 une collection d'échantillons d'ADN de référence à partir du matériel végétal échantillonné. Ceci a pour avantage de permettre le stockage de l'échantillon d'ADN de référence de pouvoir le transmettre à d'autres. L'échantillon d'ADN doit être conservé dans des conditions empêchant sa dégradation<sup>a</sup>.

### 5. Normalisation des protocoles analytiques

#### 5.1 Introduction

Le présent document n'a pas pour objectif de fournir des protocoles techniques détaillés pour la production de profils ADN de variétés. En principe, toute méthodologie analytique appropriée peut être utilisée, mais il est important que la méthode utilisée soit soumise à un processus de validation approprié. Ceci peut se faire soit par l'intermédiaire d'une méthode de validation reconnue à l'échelle internationale, ou par l'élaboration d'une approche fondée sur la performance. Pour un cas comme pour l'autre, l'on dispose de considérations générales utiles.

Toute méthode utilisée pour l'élaboration de génotypes et la construction de bases de données devrait être techniquement simple à réaliser, fiable et robuste, permettre dans les différents laboratoires une notation facile et indiscutable des profils de marqueurs. Ceci exige un certain

niveau de normalisation, par exemple dans le choix des marqueurs, des allèles de référence et dans la désignation/notation des allèles.

## 5.2 Critère de qualité

5.2.1 Il est important de s'accorder sur certains critères de qualité concernant, par exemple<sup>c</sup> :

- a) la qualité de l'ADN ~~(qui peut être évaluée en mesurant, par exemple, le taux d'absorption de l'extrait à 260/280 nm; certains recommandent également le taux de 260/230 nm);~~
- b) les séquences d'amorces utilisées ~~(avec ou sans tresses, position de l'amorce, type d'étiquette utilisée, etc.);~~
- c) le polymérase à utiliser dans les méthodes fondées sur l'ACP ~~(il peut être avantageux d'élaborer une liste de produits assurés);~~
- d) en ce qui concerne les méthodologies fondées sur l'ACP, la quantité/concentration de chaque composante d'ACP et autres composantes ~~(par exemple mémoire tampon d'ACP, MgCl<sub>2</sub>, dNTP, amorce, Taq polymérase, modèle d'ADN);~~
- (e) conditions de cycles d'ACP ~~(y compris la longueur et la température de la dénaturation initiale; nombre de cycles; longueur et température de la dénaturation, laminage pour chaque paire d'amorces et extension; et longueur et température de l'extension finale).~~

5.2.2 La description détaillée de la méthode appliquée doit figurer dans un protocole.

## 5.3 Phase d'évaluation

### 5.3.1 *Introduction*

Afin de choisir les marqueurs qui conviennent et de produire des protocoles de laboratoires acceptables pour une espèce donnée, il est recommandé de passer par une phase d'évaluation préliminaire donnée qui implique l'intervention de plus d'un laboratoire (ce qui correspond à une méthode de validation reconnue à l'échelle internationale, comme, par exemple, un test d'étalonnage effectué selon des normes reconnues à l'échelle internationale). Pendant cette phase, on se consacrera principalement au processus de sélection d'un ensemble de marqueurs, ce qui incluse habituellement l'évaluation des marqueurs existants, qu'ils soient publiés ou disponibles par un tout autre support. Le nombre de marqueurs à évaluer sera variable et dépendra des possibilités présentées par différentes espèces. Les marqueurs devraient être tirés de sources fiables (par exemple des publications qui auront été examinées préalablement) et fournies par des fournisseurs assurés. Le choix final du nombre de marqueurs sera l'équilibre entre la notion de coûts et la nécessité de produire à la fin du processus un ensemble satisfaisant de marqueurs adoptés. L'objectif consiste à produire un ensemble convenu de marqueurs qui peuvent être analysés, notifiés et enregistrés de façon fiable pour être reproduits dans différents laboratoires, avec la possibilité d'utiliser différents types d'équipements et différentes sources d'agents réactifs chimiques, etc.

### 5.3.2 *Choix des variétés*

Un nombre approprié de variétés, fondé sur la variabilité génétique à l'intérieur de l'espèce et sur le type de variété concerné, devrait être choisi pour servir de point de départ à la phase

d'évaluation. Le choix des variétés devrait refléter leur diversité et, chaque fois que cela est possible, inclure certaines variétés étroitement liées et d'autres variétés morphologiquement similaires, afin de pouvoir évaluer dans de tels cas le niveau de discrimination.

### 5.3.3 *Interprétation des résultats*

La prochaine étape d'évaluation devrait, si possible, inclure une méthode de validation reconnue internationalement, qui puisse permettre d'évaluer objectivement l'ensemble de la méthodologie. On devrait éviter toute nouvelle utilisation d'un marqueur qui aurait posé des difficultés dans l'un quelconque des laboratoires ayant pris part à cette phase d'évaluation. Dans l'idéal, dans la mesure où, à partir de l'expérience empirique, la plupart des erreurs d'analyse des vastes collections de variétés semblent provenir d'erreurs de notation, la construction de bases de données devrait être fondée sur des échantillons doubles (par exemple, différents sous-ensembles de graines à partir d'une même variété), ces derniers étant analysés dans différents laboratoires. Les sous-échantillons (ou extraits d'ADN provenant de ces sous-échantillons) pouvant être échangés en cas de divergence, cette approche est très efficace pour mettre en exergue les erreurs d'échantillonnage, ou les erreurs dues à l'hétérogénéité à l'intérieur des échantillons. De plus, elle permet d'éliminer les éventuels produits de laboratoire (erreurs systématiques).

## 5.4 Notation des données moléculaires

5.4.1 Un protocole de notation d'allèle/bande devrait être mis au point en même temps que la phase d'évaluation, afin de voir comment notifier les éléments suivants :

- a) allèles rares (c'est-à-dire allèles d'un locus spécifique apparaissant dans une population à une fréquence inférieure à un seuil convenu (habituellement 5-10%);
- b) allèles de faible fréquence (un allèle dont l'effet consiste en l'absence d'un produit ACP à l'échelle moléculaire);
- c) bandes "faibles" (c'est-à-dire des bandes dont l'intensité est inférieure à un seuil de détection convenu, fixé soit empiriquement, soit automatiquement, et dont la notation peut être mise en question);
- d) données manquantes (c'est-à-dire tout locus pour lequel aucune donnée n'est enregistrée, qu'importe le motif, dans une ou des variétés);
- e) bandes monomorphes (allèles/bandes apparaissant dans chaque variété analysée, c'est-à-dire qui ne sont pas monomorphes dans une collection particulière de variétés).

~~5.4.2 — De plus, dans le cas des marqueurs tels que les SSR, il est utile de fixer une gamme de tailles minimales et maximales pour permettre la notation des marqueurs pour chaque paire d'amorces (locus). Également, dans les cas où l'on utilise un procédé à base de gel pour visualiser les bandes de marqueurs, il est recommandé d'utiliser une "échelle" de taille convenable (poids moléculaire), afin de simplifier l'interprétation des résultats au sein d'un même laboratoire ou d'un laboratoire à l'autre. Cela ne devrait toutefois pas être considéré comme un substitut d'un ensemble d'allèles de référence<sup>d</sup>.~~

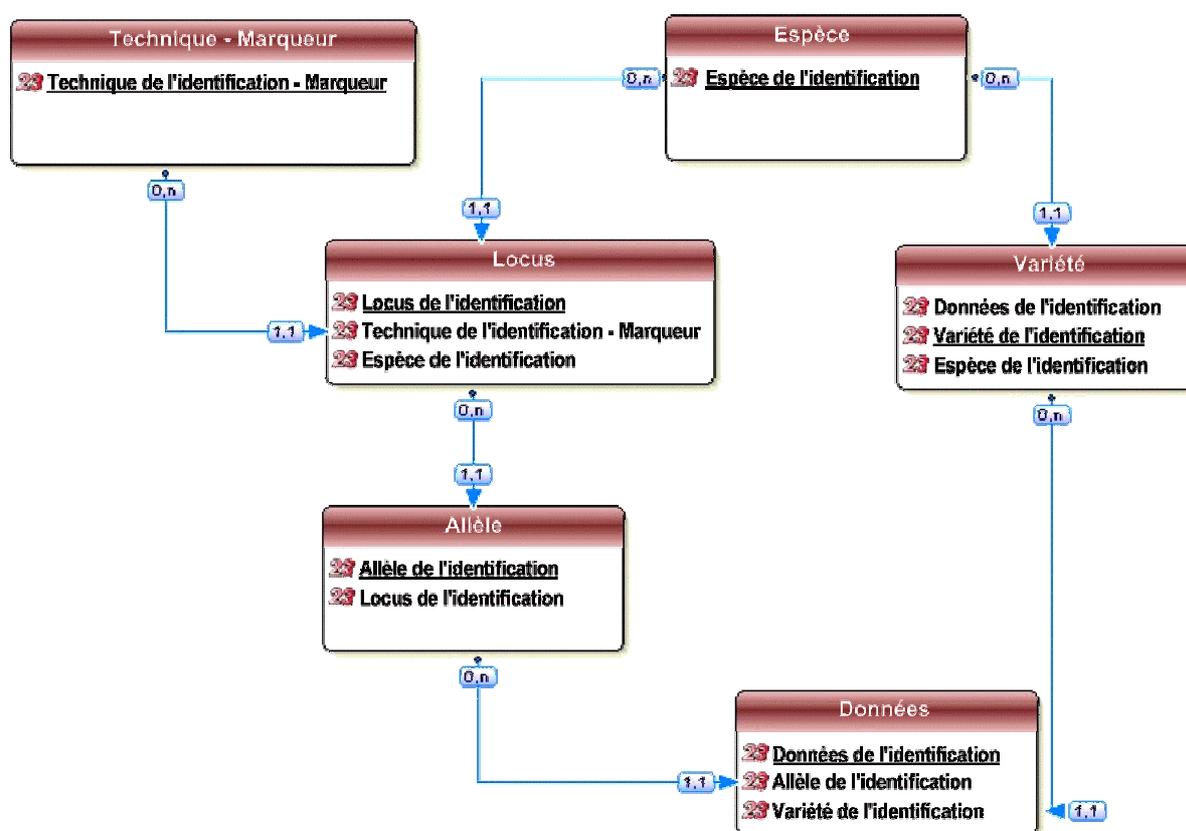
## 6. Bases de données

### 6.1 Type de base de données

existe de nombreux moyens de stockage des données moléculaires. C'est pourquoi il est important que la structure de la base de données soit élaborée de façon à ce qu'elle puisse s'appliquer à toutes les utilisations prévues des données.

### 6.2 Modèle de bases de données

Le modèle de la base de données devrait être défini par des experts en bases de données informatiques, en liaison avec les utilisateurs desdites bases. Chaque modèle devrait contenir au minimum six objets principaux : espèce, variété, technique, marqueur, locus, et allèle.



### 6.3 Dictionnaire de données

6.3.1 Dans une base de données, chacun des objets devient un tableau à l'intérieur duquel des champs sont définis. Par exemple :

a) Code technique/marqueur :

indique le code ou le nom de la technique ou le type de marqueur utilisé, *p.* ex. *SSR*, *SNP*, etc.

b) Code locus :

indique le nom ou le code du locus pour les espèces concernées, *p. ex.* *gwm 149, A2, etc.*

c) Code allèle :

indique le nom ou le code du locus pour les espèces concernées, *p. ex.* *1, 123, etc.*

d) Valeur des données :

indique une valeur de données pour l'échantillon sur un locus-allèle déterminé, *p. ex.* *0 (absence), 1 (présence), 0,25 (fréquence), etc.*

e) Variété :

la variété est l'objet pour lequel les données ont été obtenues.

f) Espèce :

l'espèce est indiquée par le nom botanique ou le nom commun national, qui peut également renvoyer au type de variété (*p. ex.* utilisation, type hivernal/printanier, etc.). L'utilisation du code UPOV pourrait éviter les problèmes liés aux synonymes et serait avantageuse du point de vue de la coordination.

6.3.2 Dans chaque tableau, le nombre de champs, leur nom et leur définition, les valeurs possibles et les règles à suivre doivent être définis dans le "dictionnaire de données".

#### 6.4 Liens entre les tableaux

6.4.1 Les liens entre les tableaux sont un aspect important de la conception de la base de données. Les liens entre les tableaux peuvent être illustrés de la manière suivante :

| Tableau | Lien                    | Tableau | Description   |
|---------|-------------------------|---------|---|
| Femme   | 0 ou<br>1 à n<br>(0, n) | Enfant  | 0 : une femme peut n'avoir aucun enfant<br>1 à n : une femme peut avoir de un à n enfants (elle devient alors une mère) |
| Enfant  | 1 à 1<br>(1,1)          | Femme   | Un enfant n'a qu'une seule mère biologique  |

6.4.2 Le tableau suivant indique les liens entre les six objets principaux au minimum comme proposé dans le modèle de la section 6.2 :

| Tableau                | Lien          | Tableau                | Description  |
|------------------------|---------------|------------------------|--|
| Technique/<br>marqueur | 0 ou<br>1 à n | Locus                  | 0 : une technique ou un marqueur peut figurer dans le champ technique/marqueur même si aucun locus/allèle n'est encore utilisé dans la base de données<br>1 à n : un type déterminé de marqueur peut donner 1 à n données utiles |
| Locus                  | 1 à 1         | Technique/<br>marqueur | Un locus est défini dans le champ d'application d'une technique ou d'un marqueur déterminé   |
| Locus                  | 1 à n         | Allèle                 | Pour chaque locus 1, ou supérieur à 1, l'allèle peut être décrit   |
| Allèle                 | 1 à 1         | Locus                  | Un allèle est défini dans le champ d'application d'un locus déterminé  |
| Allèle                 | 0 ou<br>1 à n | Donnée                 | 0 : un allèle peut être défini, mais sans données<br>1 à n : un allèle peut être trouvé dans 1 à n données   |
| Donnée                 | 1 à 1         | Allèle                 | La donnée correspond à un allèle déterminé   |
| Variété                | 0 ou<br>1 à n | Donnée                 | 0 : la variété n'a pas de données<br>1 à n : la variété a des données  |
| Donnée                 | 1 à 1         | Variété                | La donnée correspond à une variété déterminée  |
| Donnée                 | 1 à 1         | Espèces                | La donnée est obtenue pour une variété déterminée, puis pour les espèces de cette variété  |
| Espèces                | 0 ou<br>1 à n | Donnée                 | 0 : une espèce peut n'avoir aucune donnée<br>1 à n : une espèce peut avoir 1 à n données.  |

### 6.5 Transfert des données vers la base de données

Pour réduire le nombre d'erreurs dans le transfert et la transcription des données, il est conseillé d'automatiser au maximum le transfert des données vers les bases de données.

### 6.6 Accès/appropriation des données

Avant d'entreprendre tout travail, il est recommandé de régler les questions concernant l'appropriation des données et leur accès dans la base de données.

## 6.7 Analyse des données

L'objectif justifiant l'analyse des données sera l'élément utilisé pour déterminer la méthode d'analyse à appliquer. C'est pourquoi les présentes directives ne contiennent aucune recommandation spécifique.

## 6.8 Validation des bases de données

Une fois la phase d'élaboration de la base de données achevée, il est recommandé de mener un essai "en aveugle", c'est-à-dire de distribuer un nombre d'échantillons aux différents laboratoires et de leur demander d'appliquer pour les identifier le protocole adopté en liaison avec la base de données.

## 7. **Résumé**

On trouvera ci-après un résumé de l'approche qu'il est recommandé d'adopter en vue du choix et de l'utilisation des marqueurs moléculaires aux fins de l'élaboration de bases de données centrales et durables des profils ADN des variétés (c'est-à-dire de bases de données pouvant être peuplées à l'avenir de données provenant de plusieurs sources, indépendamment de la technique utilisée) :

- a) prévoir une approche récolte par récolte;
- b) convenir d'un type et d'une source de marqueur acceptable;
- c) convenir de plates-formes/équipement de détection acceptable;
- d) convenir des laboratoires qui feront partie de l'essai;
- e) se mettre d'accord sur les questions de qualité (voir la section 5.2);
- f) vérifier la source du matériel végétal (voir la section 4);
- g) décider quels marqueurs doivent être utilisés lors de la phase d'évaluation et de collaboration préliminaires, qui doit impliquer plus d'un laboratoire et des équipements de détection différents (voir la section 2);
- h) mener une évaluation (voir la section 5.3);
- i) mettre au point un protocole de notation des données moléculaires (voir la section 5.4);
- j) convenir de l'ensemble matériel/référence végétal(e) à analyser; et de la (des) source(s);
- k) analyser la collection de variétés adoptée, dans différents laboratoires/différents équipements de détection, à l'aide de doubles d'échantillons, et échanger en cas de problème échantillons/extraits d'ADN;
- l) utiliser dans toutes les analyses des variétés/échantillons d'ADN/allèles de référence;
- m) vérifier toutes les étapes (y compris l'entrée des données) – automatiser le plus possible les manœuvres;
- n) mener un essai "en aveugle" dans différents laboratoires à l'aide de la base de données;
- o) adopter les procédures d'adoption des nouvelles données.

## GLOSSAIRE

### **Microsatellites, ou répétitions de séquences simples (SSR)**

Les microsatellites, ou répétitions de séquences simples (SSR) sont des séquences d'ADN répétées en tandem, en général avec une unité de répétition de 2-4 paires de bases (par exemple, GA, CTT et GATA). Chez de nombreuses espèces, il a été mis en évidence que des allèles multiples existent pour certains microsatellites en raison de variations dans le nombre de copies de cette unité de répétition. Les microsatellites peuvent être analysés par la méthode d'ACP (amplification en chaîne par polymérase) au moyen d'amorces spécifiques. Il s'agit d'une procédure connue sous le nom de méthode des sites microsatellites étiquetés par une séquence (STMS). Les allèles (produits ACP) peuvent alors être séparés par agarose ou électrophorèse sur gel polyacrylamide. Pour développer des sites de microsatellites par séquences répertoriées, il est nécessaire d'avoir des informations sur la séquence d'ADN flanquant le microsatellite. Ces informations peuvent parfois être obtenues dans des bases de données existantes sur les séquences d'ADN; à défaut, elles devront être obtenues de manière empirique.

### **Polymorphismes nucléotidiques (SNP)**

Les polymorphismes nucléotidiques (SNP) (que l'on prononce "snips") sont des variations de séquences d'ADN qui se produisent lorsqu'un seul nucléotide (A, T, C ou G) de la séquence de génomes est altéré. Un SNP peut par exemple transformer la séquence d'ADN AAGGCTAA en ACGGCTAA. D'une manière générale, pour qu'une variation soit considérée comme un SNP, il faut qu'elle se produise dans au moins 1 % de la population. Le nombre possible de marqueurs SNP est très élevé, ce qui veut dire que l'on devrait pouvoir les rencontrer dans toutes les parties du génome. Les SNP se produisent à la fois dans les régions de codage (gène) et de non-codage du génome. La découverte de SNP implique un séquençage comparatif des nombres d'individus tirés d'une population. Plus communément, les SNP potentiels sont identifiés par comparaison des séquences alignées à partir des bases de données de séquences disponibles. Bien qu'elles puissent être détectées par procédés ACP + électrophorèse en gel assez simples, des procédures à micromatrices à débit élevé sont en train d'être élaborées en vue de la notation automatique et simultanée de centaines de loci ACP.

### **Séquences polymorphes amplifiées coupées (CAPS)**

Les séquences polymorphes amplifiées coupées (CAPS) sont des fragments d'ADN amplifiés par ACP à l'aide d'amorces à 20-25 bp, suivis, après digestion, d'une phase d'endonucléase par restriction. À la suite de cela, les polymorphismes de taille résultant de la variation dans l'apparition des sites de restriction sont identifiés au moyen de l'électrophorèse sur gel des produits digérés. Comparés aux marqueurs tels que les RFLP (polymorphismes de taille des fragments de restriction), les polymorphismes sont plus difficiles à identifier en raison de la taille limitée des fragments amplifiés (300-1800 bp). Cela dit, l'analyse des CAPS ne nécessite pas d'hybridation par tache de Southern, ni de détection radioactive. Jusqu'à ce jour et d'une manière générale, les séquences CAPS ont servi avant tout à la cartographie des gènes.

### **Régions amplifiées caractérisées par des séquences (SCARS)**

Les régions amplifiées caractérisées par des séquences (SCARS) sont des fragments d'ADN amplifiés par ACP à l'aide d'amorces spécifiques à 15-30bp, conçus à partir des séquences polymorphes identifiées préalablement. En faisant appel à des amorces avec ACP plus longs, les SCARS permettent d'éviter le problème de la faible capacité de reproduction. De plus il s'agit habituellement de marqueurs codominants. Les SCARS, qui sont propres au locus, ont été appliquées aux études de cartographie génétique ainsi que dans la sélection assistée de marqueurs.

### **Valeurs du contenu de l'information polymorphe (PIC)**

Les valeurs du contenu de l'information polymorphe (PIC) consistent en la mesure de la diversité des allèles dans un locus et servent à estimer et à comparer la puissance de discrimination des marqueurs moléculaires. La valeur du PIC d'un marqueur fondé sur ACP peut être calculée comme suit :

$$1 - \sum_{j=1}^n P_{ij}^2$$

où  $P_{ij}$  est la fréquence du modèle ACP  $j$  pour un génotype  $i$  donné.

### **Tresses**

Dans l'analyse SSR, le processus de "tresses" est l'addition d'une séquence d'oligonucléotide courte **spécifique<sup>a</sup>** avec les amorces utilisées dans l'ACP, comme un moyen d'améliorer la clarté des produits d'amplification et de réduire les produits.

### **Allèle à fréquence nulle**

Dans l'analyse SSR, un "allèle à fréquence nulle" est un allèle placé dans un locus particulier dont l'effet se manifeste sous forme d'absence d'un produit ACP.

### **Bandes à répétition**

Dans l'analyse SSR, les "bandes à répétition" consistent en une série d'une ou de plusieurs bandes, se distinguant l'une de l'autre en taille par une unité de répétition, suite à l'ACP.

---

<sup>a</sup> Texte proposé par le Groupe de travail sur les techniques biochimiques et moléculaires, notamment les profils d'ADN (BMT) à sa dixième session, tenue à Séoul (République de Corée) du 21 au 23 novembre 2006.

<sup>b</sup> Le BMT a noté que, à la vingt-quatrième session du Groupe de travail technique sur les systèmes d'automatisation et les programmes d'ordinateur (TWC), plusieurs experts avaient estimé que la section 4.3 "Taille de l'échantillon" devrait comporter davantage d'indications sur le choix de la taille de l'échantillon, notamment dans le cas de variétés allogames. Toutefois, le BMT a considéré qu'il n'y avait pas lieu d'élaborer des indications détaillées sur l'échantillonnage dans le cadre des Directives BMT.

<sup>c</sup> Le BMT a proposé de supprimer l'ensemble du texte entre crochets.

<sup>d</sup> Le BMT a proposé de supprimer la section 5.4.2.

[Fin du document]