



Directives BMT (proj.11)

ORIGINAL: anglais

DATE: 17 janvier 2008

**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 de son examen par*

*le Comité technique à sa quarante-quatrième session,  
qui se tiendra à Genève du 7 au 9 avril 2008*

*le Comité administratif et juridique à sa cinquante-septième session,  
qui se tiendra à Genève le 10 avril 2008*

**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 applicables à des types spécifiques de marqueurs moléculaires .....	5
3. Accès à la technologie .....	6
4. Matériel à analyser.....	6
4.1 Source du matériel végétal .....	6
4.2 Type de matériel végétal.....	7
4.3 Taille de l'échantillon.....	7
4.4 Échantillon d'ADN de référence.....	8
5. Normalisation des protocoles analytiques.....	8
5.1 Introduction .....	8
5.2 Critères qualitatifs.....	8
5.3 Phase d'évaluation .....	9
5.4 Notation des données moléculaires .....	10
6. Bases de données .....	11
6.1 Type de base de données.....	11
6.2 Modèle de base de données .....	11
6.3 Dictionnaire de données.....	12
6.4 Liens entre les tableaux.....	12
6.5 Transfert des données dans la base de données .....	13
6.6 Accès aux données et propriété des données .....	13
6.7 Analyse des données .....	14
6.8 Validation des bases de données .....	14
7. Résumé.....	14
<b>GLOSSAIRE</b> .....	<b>15</b>
Microsatellites, ou répétitions de séquences simples (SSR).....	15
Polymorphismes nucléotidiques (SNP).....	15
Séquences polymorphes amplifiées coupées (CAPS) .....	15
Régions amplifiées caractérisées par des séquences (SCARS) .....	16
Valeurs du contenu de l'information sur le polymorphisme (PIC).....	16
Tresses.....	16
Allèle à fréquence nulle.....	16
Bandes à répétition .....	17

## A. INTRODUCTION

Le présent document (Directives BMT) contient 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.~~ Le présent document vise à définir des exigences élevées en ce qui concerne la qualité des marqueurs et l'intérêt de générer des données reproductibles à l'aide de ces marqueurs dans des situations où le matériel ou les réactifs chimiques peuvent varier. En outre, des précautions particulières doivent être prises pour assurer la qualité des données saisies dans la base de données.<sup>a</sup>

[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 dans les documents TC/38/14 – CAJ/45/5 et TC/38/14 Add. – CAJ/45/5 Add. la position actuelle de l'UPOV à ce sujet.]<sup>b</sup>

[proposition<sup>a</sup> : supprimer la phrase]

## B. PRINCIPES GÉNÉRAUX

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

1.1 <sup>a</sup>Les critères ci-après ~~notamment~~ sont ~~les plus~~ importants dans la sélection d'une méthode :

- a) reproductibilité de la production de données entre laboratoires et plates-formes de détection (différents<sup>c</sup> types de matériel);
- b) possibilité de répétition dans le temps;
- c) pouvoir de discrimination;
- d) possibilités de compilation dans des bases de données; et<sup>a</sup>
- e) accessibilité de la méthode.
- [f] coût <sup>c</sup>

1.2 Compte tenu du progrès technique et de l'évolution du matériel, il importe, pour assurer la viabilité des bases de données dans le temps, que l'interprétation des données produites soit indépendante de l'équipement utilisé pour les produire. C'est le cas, par exemple, des 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, alors qu'on peut utiliser aujourd'hui des teintures fluorescentes avant de procéder à une séparation sur des systèmes d'électrophorèse en gel capillaire à débit élevé, largement automatisés.

<sup>a</sup> [1.3] Malgré ces différences, les données produites à l'aide des diverses techniques restent cohérentes entre elles, indépendamment des techniques utilisées pour les produire. Cela peut s'appliquer également aux données produites, par exemple, par des microsattellites 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. Ces possibilités de répétition et de reproduction sont importantes pour la construction, le fonctionnement et la longévité des bases de données, et notamment pour la création d'une base de données centrale alimentée à l'aide de données vérifiées émanant de différentes sources. Une telle base de données peut être établie de manière économiquement rationnelle, l'investissement important nécessaire au départ n'étant consenti qu'une seule fois.~~<sup>a</sup>

<sup>a</sup>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 être répétées, être reproductibles et être cohérentes. Ainsi, si les diverses techniques de profils d'ADN à locus multiples ont été utilisées avec succès pour la recherche, il est difficile d'enregistrer dans bon nombre d'entre elles une quelconque codominance et la reproductibilité des configurations de bandes entre laboratoires utilisant des équipements différents risque d'être problématique.

[...] <sup>a</sup> ~~On estime que~~ Ces facteurs présentent des difficultés dans le cadre des profils de variétés. C'est pourquoi le présent document contient essentiellement 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, les informations de séquençage (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) **(CAPS)** et les ~~SCARS~~ (régions amplifiées caractérisées par des séquences) **(SCARS)** sont elles aussi susceptibles de satisfaire aux critères susmentionnés, mais leur utilisation dans le cadre des profils d'ADN des variétés végétales n'a pas encore été explorée.

<sup>d</sup> [mentionner également les marqueurs d'isoenzymes et d'autres marqueurs génétiques (RAPD, AFLP, etc.)]

## 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 moléculaires quelle que soit leur utilisation, même si certaines utilisations particulières peuvent imposer l'application de critères supplémentaires :

- a) le niveau utile de polymorphisme (indiqué, par exemple, par la valeur du ~~PIC~~ (contenu de l'information sur le polymorphisme : voir Glossaire) **(PIC)** [ou par la valeur de la fréquence de polymorphisme (FP)<sup>e</sup>] obtenue à partir d'un ensemble de variétés représentatives (voir le glossaire);
- b) la possibilité de répétition à l'intérieur des laboratoires et de reproductibilité d'un laboratoire à l'autre en termes de notation des données;
- c) la répartition connue des marqueurs dans l'ensemble du génome (c'est-à-dire, la position cartographique), information qui, bien que non indispensable, est utile pour éviter de choisir des marqueurs qui pourraient être reliés;

- d) la possibilité d'éviter, ~~dans la mesure du possible,~~<sup>a</sup> des marqueurs dont les allèles sont à fréquence "nulle" (c'est-à-dire des allèles dont l'effet se manifeste par une absence de produit ACP au niveau de la molécule), ce qui, à nouveau, n'est pas indispensable, mais conseillé.
- [e] l'application de marqueurs dont les séquences sont d'usage courant (ou figurant dans une publication ou un journal dont l'index est fréquemment cité)]<sup>c</sup>

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

### <sup>a</sup>2.2.1 Marqueurs microsatellites (~~SSR~~)

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

<sup>a</sup>2.2.1.2 Les marqueurs SSR sont exprimés de façon codominante, ~~ont~~ généralement faciles à notifier (ou à enregistrer) et peuvent être facilement cartographiés. ~~Il a été constaté qu'ils pouvaient~~ peuvent être utilisés et analysés par différents laboratoires, et ~~(sous réserve qu'une attention suffisante soit portée aux, dans des conditions expérimentales particulières),~~ et peuvent se révéler fiables 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 en 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. Il convient notamment de tenir compte 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, ce qui peut être évité grâce à l'utilisation d'amorces "tressées" (voir le glossaire);
- c) taille du produit d'amplification;
- d) séparation réelle entre les divers allèles dans différents des systèmes de détection adaptés<sup>f</sup>;
- e) notation fiable et reproductible des allèles dans des systèmes de détection différents;
- f) niveau de polymorphisme (~~nombre d'allèles détectés~~)<sup>g</sup> entre les variétés (il est à noter que cela 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 des équipements de détection différents, il est indispensable que les allèles de référence (c'est-à-dire les ensembles de variétés) soient définis et pris en considération dans toutes les analyses. Ces allèles de référence sont nécessaires car les normes de pondération moléculaire varient selon les divers systèmes de détection dont on dispose actuellement, de sorte qu'elles ne conviennent pas à l'identification des allèles.

<sup>a</sup>2.2.1.5 Les amorces utilisées dans un laboratoire donné devraient être synthétisées par un fournisseur confirmé, afin de réduire les risques ~~de production~~ d'obtention de profils d'ADN différents à partir d'amorces synthétisées par des sources différentes.

### 2.2.2 Polymorphisme nucléotidique (SNP)

<sup>a</sup>Les polymorphismes nucléotidiques (SNP : voir le glossaire) peuvent être détectés par séquençage d'ADN, technique courante qui offre généralement des niveaux élevés de répétabilité dans le temps et de reproductibilité entre laboratoires. Cela étant, la détection de SNP spécifiques ~~s'effectue actuellement~~ peut être effectuée à l'aide de diverses techniques dont beaucoup ne sont pas encore couramment utilisées. 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 simplicité des SNP fait que leur notation est relativement aisée et fiable, ce qui devrait permettre de réduire ou d'éliminer bon nombre des problèmes susmentionnés.~~ Cela signifie 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 fiables d'un génotype particulier.

<sup>h</sup>[Nombre de locus]

## 3. Accès à la technologie

<sup>a</sup>Certains marqueurs et matériel moléculaires sont accessibles au public. Cela étant, l'obtention de marqueurs SSR de grande qualité, par exemple, suppose vraisemblablement un investissement important, de sorte qu'il est probable que certains marqueurs et autres méthodes ou matériel soient protégés par des 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 tous travaux préliminaires.

## 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

<sup>a</sup>Le matériel végétal à analyser doit être un échantillon authentique et représentatif de la variété et, lorsque c'est possible, être obtenu à partir de l'échantillon de la variété utilisé pour l'examen aux fins de l'octroi des droits d'obtenteur et de l'enregistrement officiel. L'utilisation d'échantillons du matériel remis pour l'examen aux fins des droits d'obtenus ou de l'enregistrement officiel devra faire l'objet, selon le cas, d'une autorisation de l'autorité compétente, de l'obtenteur ou du conservateur. ~~Lorsque cela est approprié,~~ Les plantes sur lesquelles les échantillons sont prélevés devraient pouvoir être retrouvées dans le cas où il apparaîtrait par la suite que certaines d'entre elles ne sont pas représentatives de la variété.

#### 4.2 Type de matériel végétal

Le type de matériel végétal à échantillonner et la procédure d'échantillonnage à suivre en vue de l'extraction d'ADN dépendront, dans une large mesure, de l'espèce végétale concernée. Ainsi, dans le cas des variétés reproduites par voie sexuée, la semence peut servir de source d'ADN, alors que, dans le cas des variétés multipliées par voie végétative, l'ADN peut être extrait à partir des feuilles. Quelle que soit la source du matériel végétal, la méthode d'échantillonnage et d'extraction de l'ADN devrait être normalisée et référencée. En outre, il conviendra de vérifier que les méthodes d'échantillonnage et d'extraction produisent des résultats constants par le biais de l'analyse de l'ADN.

#### 4.3 Taille de l'échantillon

Il est essentiel que les échantillons prélevés pour l'analyse soient représentatifs de la variété.

##### 4.3.1 *Variétés à multiplication végétative*

En principe, un seul échantillon pourrait être analysé dans le cas de variétés à multiplication végétative, dans la mesure où tous les échantillons devraient être identiques. Il est toutefois conseillé dans chaque cas d'analyser au moins des échantillons doubles. En cas de différence, il conviendrait d'analyser des échantillons supplémentaires.

##### 4.3.2 *Variétés autogames ou principalement autogames ou lignées endogames<sup>a</sup>*

Il n'est pas toujours approprié de supposer que des variétés autogames ou principalement autogames ou les lignées endogames<sup>a</sup> sont homozygotes dans tous les locus, y compris les locus SSR ou SNP. L'hétérogénéité peut provenir, par exemple, d'une hétérozygotie résiduelle, de la pollinisation croisée ou d'un mélange physique. C'est pourquoi il est généralement recommandé que plusieurs graines séparées, ou des échantillons provenant de plusieurs plantes, soient analysés, car c'est ainsi que l'on pourra constater afin de mettre en évidence une quelconque hétérozygotie. ~~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.~~<sup>b</sup>

[Proposition : élaborer des méthodes normalisées de constitution d'échantillons globaux pour certaines variétés et certains groupes de végétaux afin d'harmoniser les méthodes d'échantillonnage des semences et des plantes aux fins de l'analyse moléculaire.]<sup>c</sup>

##### 4.3.3 *Variétés à pollinisation croisée*

Des considérations semblables s'appliquent en principe aux variétés d'espèces à pollinisation croisée. On recommande en général que chaque graine ou 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 de certains allèles ou de certaines bandes. 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.

#### 4.3.4 Hybrides

La méthode qu'il convient d'appliquer pour veiller à ce que les échantillons prélevés pour l'analyse d'hybrides soient représentatifs de la variété dépendra du type d'hybride. Il faut tenir compte du fait que les variétés d'hybrides seront hétérozygotes dans les locus dont le codage sert aux marqueurs d'ADN, tout en restant uniformes phénotypiquement. Le nombre d'échantillons choisis aux fins d'analyse dépend de ce que l'on veut savoir 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'ADN 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. Cela a pour avantage de permettre de stocker les échantillons d'ADN de référence et de les transmettre à d'autres laboratoires. Les échantillons d'ADN doivent être conservés dans des conditions empêchant leur dégradation.

### 5. Normalisation des protocoles analytiques

#### 5.1 Introduction

Le présent document n'a pas pour objectif de fixer des protocoles techniques détaillés pour la production de profils ADN des variétés. En principe, toute méthodologie analytique appropriée peut être utilisée, mais il importe qu'elle soit dûment validée. On peut soit appliquer une méthode de validation reconnue à l'échelle internationale, soit élaborer une méthode fondée sur les résultats. Dans un cas comme dans l'autre, il est utile de prendre en considération un certain nombre de principes généraux.

Toute méthode utilisée pour l'établissement de génotypes et la construction de bases de données doit être techniquement simple à mettre en oeuvre, fiable et sûre, et permettre une notation aisée et indiscutable des profils de marqueurs dans les différents laboratoires. Cela suppose un certain degré de normalisation, par exemple dans le choix des marqueurs et des allèles de référence et dans la désignation et la notation des allèles.

#### 5.2 Critères qualitatifs

5.2.1 Il importe de s'accorder sur certains critères de qualité concernant, par exemple :

- a) la qualité de l'ADN;  
[(nouveau) la normalisation des méthodes d'extraction de l'ADN]<sup>k</sup>
- b) les séquences d'amorces utilisées;
- c) le polymérase à utiliser dans les méthodes fondées sur l'ACP;
- d) en ce qui concerne les méthodes fondées sur l'ACP, la quantité ou la concentration de chaque composante d'ACP et des autres composantes;
- e) les conditions de cycles d'ACP.  
[(nouveau) la machine ACP]<sup>l</sup>



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

[Proposition : indiquer le logiciel utilisé pour l'analyse des bandes amplifiées.]<sup>m</sup>

### 5.3 Phase d'évaluation

#### 5.3.1 *Introduction*

Afin de choisir les marqueurs qui conviennent et de produire des protocoles de laboratoire acceptables pour une espèce donnée, il est recommandé de prévoir une phase d'évaluation préliminaire impliquant l'intervention de plusieurs laboratoires (méthode de validation reconnue à l'échelle internationale, par exemple un test d'étalonnage effectué selon des normes internationalement reconnues). Cette phase doit être principalement consacrée au choix d'une série de marqueurs, ce qui suppose habituellement l'évaluation des marqueurs existants, qu'ils soient publiés ou disponibles par d'autres moyens. Le nombre de marqueurs à évaluer est variable et dépend des possibilités offertes par les différentes espèces. Les marqueurs doivent être tirés de sources fiables (par exemple, des publications examinées collégialement) et provenir de fournisseurs confirmés. Le choix définitif du nombre de marqueurs sera fonction d'un arbitrage entre le coût à supporter et la nécessité d'obtenir en fin de compte un ensemble satisfaisant de marqueurs agréés. L'objectif consiste à établir un ensemble agréé de marqueurs qui peuvent être analysés, notifiés et enregistrés de façon fiable et reproductible dans différents laboratoires, avec la possibilité d'utiliser différents types d'équipements et différentes sources de réactifs chimiques, etc.

[Proposition : ajouter des renseignements sur l'organisation de tests d'étalonnage pour les laboratoires qui ont l'intention de faire figurer les résultats dans une base de données sur les marqueurs moléculaires, ainsi que sur les méthodes d'obtention de ces résultats ; par exemple, les tests organisés à l'ISTA dans le cadre de l'examen des modifications génétiques.]<sup>o</sup>

#### 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, doit être choisi comme point de départ pour la phase d'évaluation. Le choix des variétés doit rendre compte de leur diversité et, chaque fois que cela est possible, inclure certaines variétés apparentées et d'autres variétés morphologiquement similaires, afin de pouvoir évaluer le niveau de discrimination dans de tels cas.

#### 5.3.3 *Interprétation des résultats*

<sup>a</sup>L'étape d'évaluation suivante doit, si possible, comprendre une méthode de validation reconnue internationalement, qui permette d'évaluer objectivement l'ensemble de la méthodologie. Tout marqueur posant des problèmes dans l'un des laboratoires participant à cette phase d'évaluation ne devrait plus être utilisé par la suite. ~~Dans l'idéal,~~ Dans la mesure où il ressort de l'expérience empirique que 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-échantillons de semences de la même variété), ces derniers étant analysés ~~dans différents laboratoires~~ par plusieurs laboratoires. Les sous-échantillons (ou extraits d'ADN provenant de ces sous-échantillons) pouvant être échangés en cas de divergence, cette méthode est très efficace pour mettre en évidence 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

Un protocole de notation des allèles/bandes devrait être mis au point avec la phase d'évaluation, afin de déterminer comment noter les éléments suivants :

- a) allèles rares (c'est-à-dire allèles à locus spécifique apparaissant dans une population à une fréquence inférieure à un seuil convenu (habituellement 5 à 10%);
- b) allèles à fréquence nulle (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, pour quelque raison que ce soit, dans une ou plusieurs 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).

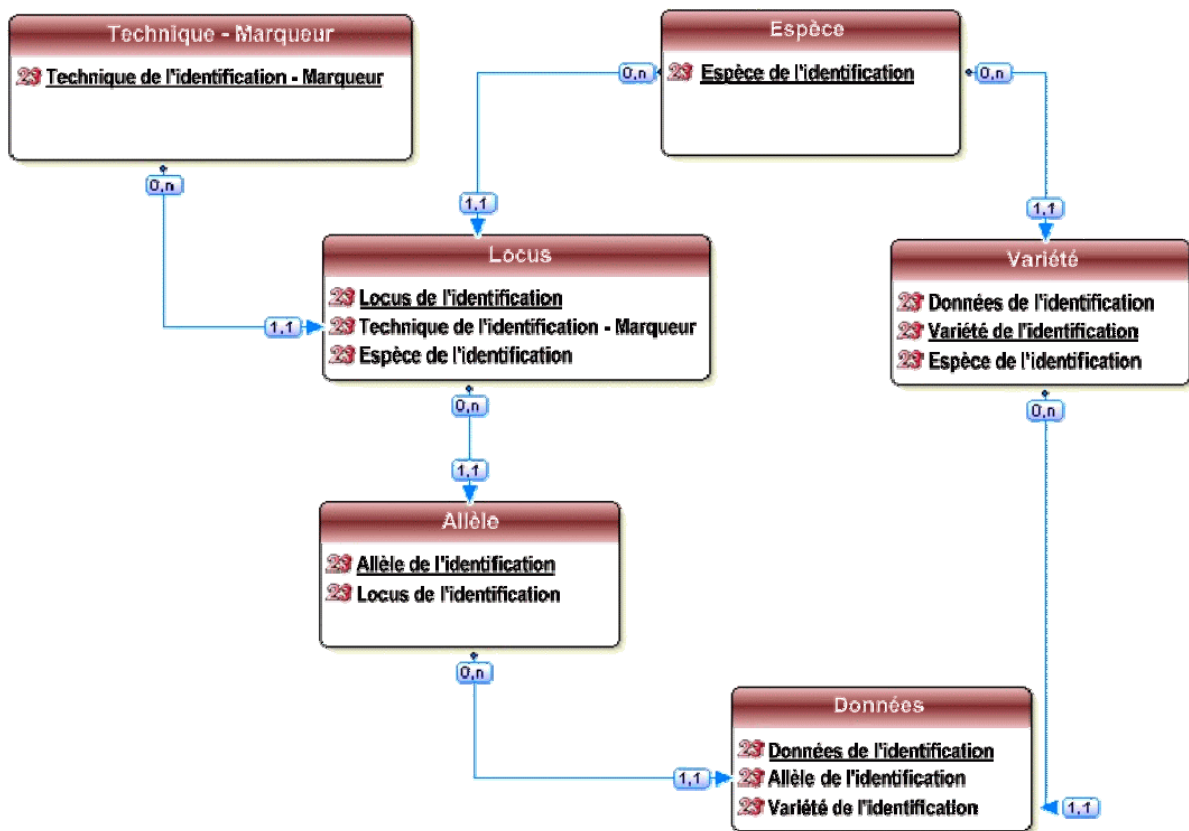
## 6. Bases de données

### 6.1 Type de base de données

Il existe de nombreux moyens de stockage des données moléculaires. C'est pourquoi il importe 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 base de données

Le modèle de base de données devrait être défini par des experts en bases de données informatiques, en liaison avec les utilisateurs. 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 de l'allèle d'un locus donné, 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 1 à n (0, n)	Enfant	0 : une femme peut n'avoir aucun enfant 1 à n : une femme peut avoir de 1 à n enfants (elle devient alors une mère)
Enfant	1 à 1 (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/ marqueur	0 ou 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 1 à n : un type déterminé de marqueur peut donner 1 à n données utiles
Locus	1 à 1	Technique/ 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'un locus déterminé
Allèle	0 ou 1 à n	Donnée	0 : un allèle peut être défini, mais sans données 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 1 à n	Donnée	0 : la variété n'a pas de données 1 à n : la variété a des données
Donnée	1 à 1	Variété	La donnée correspond à une variété déterminée
Donnée	1 à 1	Espèce	La donnée est obtenue pour une variété déterminée, puis pour les espèces de cette variété
Espèce	0 ou 1 à n	Donnée	0 : une espèce peut n'avoir aucune donnée. 1 à n : une espèce peut avoir 1 à n données.

### 6.5 Transfert des données dans 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 dans les bases de données.

### 6.6 Accès aux données et propriété des données

Avant toute chose, il est recommandé de régler les questions concernant la propriété des données et l'accès à la base de données.

## 6.7 Analyse des données

La méthode d'analyse à appliquer sera fonction de la finalité de l'analyse. 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 leur identification le protocole adopté en liaison avec la base de données.

<sup>n</sup>[Bases de données sur les plantes cultivées]

## 7. **Résumé**

On trouvera ci-après un résumé de la méthode qu'il est recommandé de suivre 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 alimentées à l'avenir par des données provenant de plusieurs sources, indépendamment de la technique utilisée) :

- a) envisager une méthode plante par plante;
- b) déterminer les types et les sources de marqueurs acceptables;
- c) déterminer les plates-formes et l'équipement de détection acceptables;
- d) convenir des laboratoires qui participeront à l'essai;
- e) se mettre d'accord sur les questions de qualité (voir la section 5.2);
- f) vérifier la source du matériel végétal utilisé (voir la section 4);
- g) déterminer les marqueurs à utiliser lors de la phase préliminaire d'évaluation en commun, qui doit impliquer plusieurs laboratoires et des équipements de détection différents (voir la section 2);
- h) réaliser 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 retenue, dans différents laboratoires et au moyen d'équipements de détection différents, en utilisant des échantillons doubles et échangeant les échantillons/extraits d'ADN en cas de problème;
- l) utiliser dans toutes les analyses des variétés/échantillons d'ADN/allèles de référence;
- m) vérifier toutes les étapes (y compris la saisie des données) – automatiser les opérations au maximum;
- n) mener un essai "en aveugle" dans différents laboratoires à l'aide de la base de données;
- o) adopter les procédures relatives à l'adjonction de nouvelles données.

<sup>o</sup>[(nouveau) organiser le réseau de laboratoires UPOV pour la recherche sur l'identification moléculaire des variétés]<sup>o</sup>

## 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). Dans 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 électrophorèse en gel d'agarose ou de polyacrylamide. Pour élaborer 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 A**A**GGCTAA en A**T**GGCTAA. 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 locus 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 en 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.

### **8 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 sur le polymorphisme (PIC)**

Les valeurs du contenu de l'information polymorphisme (PIC) consistent en la mesure de la diversité des allèles dans un locus et servent à estimer et à comparer le pouvoir 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.

### **9 Valeurs de la fréquence de polymorphisme (FP)**

Les valeurs de la fréquence de polymorphisme (FP) sont une mesure du polymorphisme dans un locus qui sert à estimer et à comparer le pouvoir de discrimination des marqueurs moléculaires. Il s'agit de la proportion de variétés polymorphes par paires dans le nombre total de paires de variétés. La valeur de la FP d'un marqueur fondé sur ACP peut être calculée comme suit :

$$FP = 1 - \sum (m_i \times (m_i - 1)) / (m \times (m - 1))$$

où  $m$  est le nombre total de variétés testées et  $m_i$  est le nombre de variétés présentant le modèle ACP  $i$ .

### **Tresses**

Dans l'analyse SSR, le processus de "tresses" est l'addition d'une séquence d'oligonucléotide courte spécifique aux amorces utilisées dans l'ACP, comme 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> Proposition des États-Unis d'Amérique.

<sup>b</sup> À la soixante-quatorzième session du Comité consultatif, tenue à Genève le 24 octobre 2007, la délégation des États-Unis d'Amérique s'est enquis du statut des documents TC/38/14 – CAJ/45/5 et TC/38/14 Add. – CAJ/45/5 Add. une fois les directives BMT adoptées. Elle a fait observer en particulier que les directives BMT seraient adoptées par le Conseil, alors que les documents TC/38/14 – CAJ/45/5 et TC/38/14 Add. – CAJ/45/5 Add. ne l'ont pas été. Le Comité consultatif a recommandé d'étudier le statut des documents TC/38/14 – CAJ/45/5 et TC/38/14 Add. – CAJ/45/5 Add. compte tenu du fait qu'ils sont mentionnés dans l'introduction des directives BMT (projet 9).

<sup>c</sup> Proposition de l'Ukraine.

<sup>d</sup> Proposition de la Chine : explication : "En ce qui concerne les types de marqueurs, les directives UPOV utilisent seulement les SSR et (à l'avenir) les SNP. L'utilisation des marqueurs SSR suppose généralement que l'on dispose d'études génétiques et moléculaires antérieures sur les espèces ou variétés concernées (il faut un nombre suffisant d'amorces connues pour permettre le tri). Toutefois, pour les espèces ou variétés n'ayant pas fait l'objet d'un grand nombre d'études génétiques (en particulier, des études moléculaires, comme c'est le cas de nombreuses essences ligneuses), l'utilisation des SSR peut s'avérer difficile. C'est pourquoi, il est souhaitable, en l'état actuel des choses, d'ouvrir le choix à un éventail plus large de marqueurs BMT, tels que les marqueurs d'isoenzymes, qui ont été très bien étudiés (p. ex., en Allemagne), et d'autres marqueurs de l'ADN (RAPD, AFLP, etc.) peuvent être choisis dès lors qu'il existe des techniques suffisamment éprouvées permettant de remplir les mêmes fonctions que les SSR".

<sup>e</sup> Proposition de la Chine : explication : "Dans la recherche sur les ressources phylogénétiques, d'une manière générale, on utilise le contenu de l'information sur le polymorphisme (PIC) comme indiqué dans les directives. Toutefois, nous estimons que les applications directes du PIC sont insuffisantes. C'est pourquoi, nous adoptons un paramètre intitulé "fréquence du polymorphisme" (FP), qui a une application plus directe pour la mesure du ratio de polymorphisme, moyennant l'indication de la fréquence de différence (polymorphisme) détectée dans les paires de variétés échantillonnées au moyen des marqueurs moléculaires analysés. Si le matériel détecté a une représentation suffisante, la valeur de la FP sera équivalente à la probabilité de différence détectée dans les paires de matériel testées au moyen des marqueurs correspondants. Nous suggérons donc d'ajouter le paramètre FP.

<sup>f</sup> Proposition de la Chine : explication : "Les différents systèmes de détection ont des pouvoirs résolvent différents pour le système PAGE dénaturé; une différence de 1 bp devrait être détectée en théorie".

<sup>g</sup> Proposition de la Chine : explication : "Par rapport au nombre d'allèles détectés, le PIC et la FP rendront mieux compte du degré de polymorphisme. C'est la raison principale pour laquelle il convient de mentionner le PIC dans les directives. Nous ne savons pas pourquoi le "nombre d'allèles détectés" devrait figurer ici. La valeur du PIC est un paramètre beaucoup plus intéressant que le nombre d'allèles pour rendre compte du polymorphisme de l'amorce".

<sup>h</sup> Proposition de la Chine : explication : "Il est souhaitable d'indiquer le nombre de locus afin d'éviter d'éventuelles erreurs d'échantillonnage (erreurs statistiques). Un nombre minimal de locus au moins devrait être fixé".

<sup>i</sup> Selon les États-Unis d'Amérique, "Le document TGP/7/1 n'a pas été adopté par le Conseil; la délégation des États-Unis d'Amérique n'est pas favorable à ce qu'il soit cité comme "directive élaborée par l'UPOV".

<sup>j</sup> Proposition de la Chine : explication : "Il est préférable d'analyser plusieurs individus standard respectivement plutôt que d'utiliser des échantillons globaux".

<sup>k</sup> Proposition de l'Ukraine : explication : "La normalisation des méthodes d'extraction de l'ADN est particulièrement importante pour certaines plantes ayant un niveau élevé de métabolites secondaires".

<sup>l</sup> Proposition de l'Ukraine : explication : "Proposer la machine ACP car, très souvent, les conditions de cycle diffèrent en fonction des machines ACP".

<sup>m</sup> Proposition de l'Ukraine : "Il importe d'indiquer le logiciel utilisé pour l'analyse des bandes amplifiées, notamment pour l'estimation de la taille de bande appropriée en cas d'utilisation de gel d'acrylamide ou d'agarose".

<sup>n</sup> Commentaire de l'Ukraine : "D'une manière générale, il sera très utile de mettre à la disposition de tous les membres de l'UPOV la base de données sur les marqueurs moléculaires pour certaines plantes, notamment

pour les séquences d'amorces de certains marqueurs SSR ou CAPS (dans ce dernier cas, il conviendra d'indiquer l'enzyme utilisée pour la digestion). Par exemple, la partie de la base de données du Ministère de l'agriculture des États-Unis consacrée aux marqueurs ([wheat.pw.usda.gov](http://wheat.pw.usda.gov)) permet de trouver des informations sur les séquences d'amorces du blé et des espèces apparentées, les conditions de l'amplification et les références des articles ou les coordonnées des auteurs avec leur adresse électronique".

- ° Proposition de l'Ukraine : "Ajouter un point consacré à l'organisation du réseau de laboratoires UPOV pour la recherche sur l'identification moléculaire des variétés, aux fins de l'harmonisation et de l'adoption de la méthode d'analyse moléculaire des variétés, du partage des connaissances et des résultats de l'identification des allèles (notamment en ce qui concerne leur taille) et de la participation aux tests d'étalonnage".

[Fin du document]