

UPOV/INF/17/2 Draft 6**Original** : anglais**Date** : 10 juin 2021

pour examen par correspondance

PROJET (RÉVISION)

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**le Comité technique, le Comité administratif et juridique et le Conseil en 2021**Avertissement : le présent document ne représente pas les principes ou les orientations de l'UPOV*

<u>Précisions concernant cette version</u>
--

Le texte **biffé** (surligné en gris) a été supprimé du document [UPOV/INF/17/1](#).

Le texte **souligné** (surligné en gris) a été ajouté au texte proposé du document [UPOV/INF/17/1](#).

TABLE DES MATIÈRES

A. INTRODUCTION.....	3
B. PRINCIPES GÉNÉRAUX.....	3
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 certains types 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 de matériel végétal.....	6
4.3 Taille de l'échantillon.....	6
4.4 Échantillon d'ADN de référence.....	6
5. Normalisation des protocoles analytiques.....	6
5.1 Introduction.....	6
5.2 Critères qualitatifs.....	7
5.3 Phase d'évaluation.....	7
5.4 Notation des données moléculaires.....	8
6. Bases de données.....	8
6.1 Type de base de données.....	8
6.2 Modèle de base de données.....	9
6.3 Dictionnaire de données.....	9
6.4 Liens entre les tableaux.....	10
6.5 Transfert des données dans la base de données.....	11
6.6 Accès aux données et propriété des données.....	11
6.7 Analyse des données.....	11
6.8 Validation des bases de données.....	11
7. Résumé.....	12
GLOSSAIRE.....	13
Microsatellites, ou répétitions de séquences simples (SSR).....	13
Polymorphismes nucléotidiques (SNP).....	13
Séquences polymorphes amplifiées coupées (CAPS).....	13
Régions amplifiées caractérisées par des séquences (SCARS).....	14
Tresses.....	14
Allèle à fréquence nulle.....	14
Bandes à répétition.....	14
A. INTRODUCTION.....	3
B. PRINCIPES GÉNÉRAUX.....	3
1. Choix des marqueurs moléculaires.....	4
1.1 Ensembles de variétés pour le processus de sélection.....	4
1.2 Marqueurs moléculaires – considérations relatives aux performances.....	5
2. Sélection de la méthode de détection.....	6
2.1 Méthodes relatives aux profils d'ADN – considérations générales.....	6
2.2 Accès à la technologie.....	7
3. Validation et harmonisation d'un ensemble de marqueurs et d'une méthode de détection.....	7
3.1 Validation et harmonisation – considérations générales.....	7

3.2	Considérations relatives aux performances – validation des marqueurs et des méthodes	8
3.3	Considérations relatives à la cohérence	8
4.	Construction d'une base de données spécifique à une espèce	8
4.1	Recommandations pour la conception d'une base de données	8
4.2	Conditions requises pour le matériel végétal	9
4.3	Traitements des données relatives aux séquences	11
4.4	Type de base de données	11
4.5	Modèle de base de données	11
4.6	Dictionnaire de données	12
4.7	Accès aux données et propriété des données	14
5.	Échange des données	14
5.1	Scénarios d'échange des données	14
5.2	Méthodes d'échange de données	14
6.	Résumé	14
C.	LISTE DES SIGLES	17
ANNEXE SCÉNARIOS D'ÉCHANGE DES DONNÉES ET MÉTHODES DE TRANSFERT DE DONNÉES		

A. INTRODUCTION

Le présent document (Directives BMT) contient des directives en vue de l'élaboration de sur les méthodes harmonisées principes harmonisés relatifs à l'utilisation de marqueurs moléculaires qui serviront à produire des données moléculaires de haute qualité destinées à diverses applications. Seuls les marqueurs moléculaires d'ADN sont examinés dans le présent document.

Les Directives BMT ont également pour but de permettre l'élaboration de bases de données contenant des profils moléculaires de variétés végétales, qui peuvent être produits dans différents laboratoires à l'aide de diverses techniques. En outre, l'objectif est de 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. Des précautions particulières doivent être prises pour assurer la qualité des données saisies dans la base de données.

B. PRINCIPES GÉNÉRAUX

Aux fins de l'établissement du profil d'ADN d'une variété végétale, un ensemble de marqueurs moléculaires et une méthode de détection des marqueurs sont requis. Deux ensembles distincts de marqueurs moléculaires détectés au moyen d'une même méthode donneront deux profils d'ADN distincts pour une variété donnée. En revanche, deux méthodes distinctes pour détecter les allèles spécifiques d'un marqueur moléculaire donné sont censées donner des profils d'ADN identiques. La normalisation de la méthode de détection et de la technologie n'est pas requise, dès lors que les résultats remplissent les critères de qualité et que les profils d'ADN résultants sont cohérents. Quelle que soit la technologie utilisée pour détecter les ensembles de marqueurs définis, le génotype d'une variété particulière ne devrait pas être affecté.

Les ensembles de marqueurs moléculaires, les méthodes de détection de marqueurs et le processus ultérieur d'élaboration de bases de données peuvent être divisés en cinq phases distinctes :

1. sélection des marqueurs moléculaires
2. sélection de la méthode de détection
3. validation et harmonisation de la méthode de détection
4. construction de la base de données
5. échange des données

Le présent document décrit ces différentes phases de façon plus détaillée. Il est considéré que ces phases sont indépendantes du stade de développement des technologies pour l'établissement de descriptions de génotypes et des améliorations futures en matière de séquençage à haut débit.

1. Sélection de la méthode à appliquer aux marqueurs moléculaires

1.1 Les critères ci-après sont importants dans la sélection d'une méthode :

- a) reproductibilité de la production de données entre laboratoires et plateformes de détection (différents types de matériel);
- b) possibilité de répétition dans le temps;
- c) pouvoir de discrimination;
- d) possibilités de compilation dans des bases de données; et
- e) accessibilité de la méthode.

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.

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 microsatellites d'ADN (répétitions de séquences simples—SSR) ou des techniques de polymorphismes nucléotidiques (SNP). 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.

1.4 Les 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.

1.5 Ces facteurs sont source de 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 séquences polymorphes amplifiées coupées (CAPS) et les 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.

1. Choix des marqueurs moléculaires

1.1 Ensembles de variétés pour le processus de sélection

Pour les profils d'ADN des variétés végétales et la construction de bases de données, les marqueurs moléculaires devraient être sélectionnés en fonction de l'objectif à atteindre. Pour débiter le processus de sélection des marqueurs, un nombre approprié de variétés (ensemble de développement) est nécessaire pour refléter au mieux la diversité observée au sein du groupe, de la plante, de l'espèce ou du type pour lequel les marqueurs doivent avoir un pouvoir de discrimination. Une nouvelle sélection est réalisée à partir des profils de variétés supplémentaires (ensemble de validation) pour mesurer la performance des marqueurs. Les critères pour le choix de l'ensemble de validation peuvent être :

- a) variétés ou lignées génétiquement très proches, NIL, RIL
- b) lignées parentales et descendance
- c) variétés génétiquement proches mais morphologiquement distinctes (p. ex. mutants)
- d) certaines variétés morphologiquement proches avec des généalogies différentes
- e) différents lots d'une même variété
- f) différentes origines de la même variété

1.2 Marqueurs moléculaires – considérations relatives aux performances

Les critères généraux suivants utilisés pour sélectionner un marqueur spécifique ou un ensemble de marqueurs s'appliquent quelle que soit leur utilisation :

- a) la répétabilité, la reproductibilité et la robustesse au sein d'un laboratoire et d'un laboratoire à l'autre en terme de notation des données;
- b) les sources possibles de marqueurs moléculaires
 - marqueurs moléculaires tirés de ressources publiques
 - marqueurs moléculaires tirés de ressources non publiques, présélection et sélection de puces et de matrices spécifiques d'une espèce disponibles dans le commerce
 - marqueurs moléculaires sélectionnés à partir de données relatives à la séquence nouvellement générées;
- c) éviter, dans la mesure du possible, des marqueurs avec des allèles "null" (c'est-à-dire des allèles dont l'effet se manifeste par une absence de produit PCR au niveau moléculaire), ce qui n'est à nouveau pas indispensable, mais conseillé;
- d) permettre une notation aisée, objective et indiscutable des profils de marqueurs. Ces marqueurs performants sont préférables aux profils de marqueurs complexes qui sont délicats en termes d'interprétation. Des réponses claires et tranchées permettent également de faciliter l'harmonisation;
- e) les marqueurs codominants sont généralement préférables aux marqueurs dominants du fait de leur pouvoir de discrimination plus élevé;
- f) les marqueurs peuvent être situés dans des régions codantes ou non codantes; et
- g) l'utilisation de marqueurs moléculaires est spécifique d'une espèce et devrait tenir compte des particularités de reproduction ou de multiplication de l'espèce.

Il est reconnu que certaines utilisations particulières peuvent imposer l'application de considérations supplémentaires, notamment (mais pas uniquement) les suivantes :

- a) Le nombre de marqueurs devrait être équilibré par rapport à la précision du génotype requise pour atteindre l'objectif. Le nombre de marqueurs pour atteindre la résolution ou le pouvoir de discrimination requis dépend du type de marqueur (dominant/codominant; bi/multiallélique), de l'espèce et de la qualité de la performance du marqueur;
- b) La couverture du génome et du déséquilibre de liaison devrait refléter les objectifs. Connaître la position physique ou génétique des marqueurs sélectionnés sur le génome n'est pas indispensable, mais permet un bon choix des marqueurs.

2. Sélection des marqueurs moléculaires

2.1 Critères généraux

Les critères généraux suivants utilisés pour le choix d'un marqueur spécifique ou d'un ensemble de marqueurs s'appliquent aux marqueurs moléculaires quelle que soit leur utilisation, même si certaines utilisations particulières peuvent imposer l'application de critères supplémentaires :

- a) le niveau nécessaire de polymorphisme;
- b) la répétabilité au sein d'un laboratoire et la reproductibilité d'un laboratoire à l'autre en terme de notation des données;
- c) la répartition connue des marqueurs dans l'ensemble du génome (c'est-à-dire, la cartographie), information qui, bien que non indispensable, est utile pour éviter de choisir des marqueurs qui pourraient être liés; et
- d) éviter, dans la mesure du possible, des marqueurs avec des allèles "null" (c'est-à-dire des allèles dont l'effet se manifeste par une absence de produit PCR au niveau moléculaire), ce qui, à nouveau, n'est pas indispensable, mais conseillé.

2.2 Critères applicables à certains types de marqueurs moléculaires

2.2.1 Marqueurs microsatellites

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 (PCR), largement utilisée aujourd'hui, présente plusieurs avantages.

2.2.1.2 Les marqueurs SSR sont codominants, et sont généralement faciles à noter (ou à enregistrer) et à cartographier. Utilisés et analysés par différents laboratoires, ils se sont révélés, dans certaines conditions expérimentales, en général robustes et reproductibles. En outre, ils peuvent être analysés à l'aide de séquenceurs à ADN automatiques, à haut débit et non radioactifs, basés sur un système d'électrophorèse sur gel, ou par capillaires, 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 suivants :

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

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 séries de variétés) soient définis et inclus dans toutes les analyses. Ces allèles de référence sont nécessaires car les standards de poids moléculaire varient selon les divers systèmes de détection disponibles actuellement, de sorte qu'ils ne conviennent pas à l'identification des allèles.

2.2.1.5 Les amorces utilisées dans un laboratoire donné devront être synthétisées par un fournisseur attitré, afin de réduire les risques d'obtention de profils d'ADN différents à partir d'amorces synthétisées par plusieurs sources.

2.2.2 Polymorphisme nucléotidique simple (SNP)

Les polymorphismes nucléotidiques simples (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 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 il y aura des effets de dosage. La simplicité des SNP fait que leur notation est relativement aisée et fiable. 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.

2. Sélection de la méthode de détection

2.1 Méthodes relatives aux profils d'ADN – considérations générales

2.1.1 Les critères ci-après sont importants dans la sélection des méthodes relatives aux profils d'ADN pour générer des données moléculaires de qualité :

- a) reproductibilité de la production de données entre laboratoires et plateformes de détection (différents types de matériel);
- b) possibilité de répétition dans le temps;
- c) pouvoir de discrimination;
- d) temps et main d'œuvre requis;

- e) robustesse des performances dans le temps et les conditions (sensibilité aux changements subtils du protocole ou des conditions);
- f) flexibilité de la méthode, possibilité de faire varier le nombre d'échantillons ou le nombre de marqueurs;
- g) l'interprétation des données produites est indépendante du matériel;
- h) pérennité des bases de données;
- i) accessibilité de la méthode;
- j) indépendante d'un appareil, d'une composition chimique, d'un fournisseur, de partenaires ou de produits spécifiques;
- k) possibilité d'automatisation;
- l) possibilité de multiplexage; et
- m) rentabilité (équilibre des coûts, du nombre d'échantillons et du nombre de marqueurs).

3.2.2 Accès à la technologie

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/ou 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 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.

3. Validation et harmonisation d'un ensemble de marqueurs et d'une méthode de détection

3.1 Validation et harmonisation – considérations générales

Les marqueurs moléculaires et les méthodes de détection doivent être robustes et donner lieu à des profils d'ADN cohérents. Les performances des marqueurs moléculaires et des méthodes utilisées pour l'établissement de descriptions de génotypes sont évaluées dans un processus de validation. Dans le cas de bases de données partagées, la cohérence des profils d'ADN dans les différents laboratoires est évaluée dans le processus d'harmonisation à l'aide de différents matériels et compositions chimiques. L'utilisation de marqueurs et de méthodes validés aboutira à des résultats harmonisés.

3.2 Considérations relatives aux performances – validation des marqueurs et des méthodes

L'ensemble de marqueurs sélectionné devrait être adapté à son objet. Sa précision devrait être mesurée. Pour déterminer si une méthode et un ensemble de marqueurs d'ADN conviennent, plusieurs éléments doivent être pris en considération :

- a) pouvoir de discrimination/pouvoir informatif;
- b) possibilité de répétition : lorsque des résultats d'essai identiques sont obtenus avec la même méthode, sur des objets identiques, dans le même laboratoire, par le même opérateur, en utilisant le même équipement dans de courts intervalles de temps;
- c) reproductibilité : lorsque les résultats des essais sont obtenus avec la même méthode, sur des objets identiques, dans le même laboratoire ou entre différents laboratoires, avec différents opérateurs, en utilisant des équipements différents;
- d) robustesse : mesure de sa capacité à ne pas être affecté par des écarts faibles mais délibérés par rapport aux conditions expérimentales décrites dans les paramètres de la procédure, donne une indication de sa fiabilité en utilisation normale; et
- e) taux d'erreur.

Les définitions des caractères relatifs à la performance sont fondées sur la norme ISO 16 577:2016

3.3 Considérations relatives à la cohérence

Pour obtenir une cohérence des résultats, le processus d'harmonisation des marqueurs et des méthodes entre les différents laboratoires dans le cas d'une base de données partagée (test d'étalonnage) doit inclure ce qui suit :

- a) Utilisation d'une collection de variétés définie représentant un large éventail d'allèles comme référence dans tous les laboratoires pour vérifier la cohérence entre les laboratoires;

- b) Inclusion d'échantillons doubles, sous-échantillons, plantes individuelles d'une variété pour vérifier la cohérence des profils d'ADN et estimer le taux d'erreur entre les laboratoires;
- c) Accords sur la notation des données moléculaires. La nécessité de mettre au point un protocole de notation des allèles/bandes entre les laboratoires dépend du type de marqueur utilisé (essentiel pour les SSR). Le protocole pourrait permettre de déterminer comment noter les éléments suivants :
 - i) 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%));
 - ii) allèles à fréquence nulle (un allèle dont l'effet consiste en l'absence d'un produit PCR à l'échelle moléculaire);
 - iii) 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);
 - iv) 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); et
 - v) bandes monomorphes ou notations d'allèles non informatives (allèles/bandes apparaissant dans chaque variété analysée, c'est-à-dire qui ne sont pas monomorphes dans une collection particulière de variétés).

4. Construction d'une base de données spécifique à une espèce

Les données qui sont stockées dans une base de données et la façon dont elles sont stockées doivent fournir des indications sur le processus de production des données. Par conséquent, la construction d'une base de données devrait tenir compte des différents niveaux de traitement des données (c.-à-d. données brutes, données séquentielles, etc.). La base de données devrait contenir les résultats finaux, p. ex. le profil d'ADN ainsi que des indications sur la façon dont il a été établi avec une description de la méthode employée par le laboratoire et les étapes de calcul.

4.1 Recommandations pour la conception d'une base de données

Dans la conception de bases de données, il convient de tenir compte des éléments suivants :

- a) L'architecture de la base de données devrait être flexible, par exemple permettre de stocker à la fois des fichiers plats et des fichiers d'archives compressés.
- b) Des tableaux et des entrées distincts sont nécessaires pour le travail expérimental en laboratoire, le traitement des données et la notation des allèles.
- c) Le stockage d'informations à différents niveaux, par exemple notations des allèles et toute règle d'interprétation ayant présidé à la décision, et les liens vers les données brutes (fichiers tiff, fichiers bam) qui ont été produites.
- d) Pour les données de séquençage, des fichiers de détection des variants au format VCF ou BCF correspondant à la version standard 4.2 ou à une version supérieure. Les entrées d'en-tête doivent contenir le nom et la version des différents scripts utilisés pour la cartographie des lectures, le filtrage des lectures, la détection des variants et le filtrage des variants, de sorte qu'un bioinformaticien puisse répéter l'analyse.
- e) Dans le cas d'échantillons répétés, lorsque le profil d'ADN ne correspond pas, l'enregistrement doit être marqué ou filtré, le cas échéant. Les règles appliquées dans ces cas doivent être documentées dans un référentiel de codes accessible au public établi à partir du fichier de détection des variants. Des fréquences pourraient également être utilisées pour des variétés hétérogènes.
- f) Validation des données VCF et BCF par rapport aux exigences pertinentes.
- g) Données faciles à partager (p. ex. API).

4.2.2 Matériel à analyser-Conditions requises pour le matériel végétal

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 restaurer et à partager dans la base de données devraient être pris en considération.

4.2.1 Source du matériel végétal

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~~de ces échantillons du matériel remis pour l'examen aux fins des droits d'obtent~~eurs 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. 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.

4.2.2 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 permettent d'obtenir des résultats d'analyse ADN stables.

4.2.3 Taille et type de l'échantillon (échantillons globaux ou individuels)

Il est essentiel que les échantillons prélevés pour l'analyse soient représentatifs de la variété. ~~En ce qui concerne la représentativité de la variété, il~~ convient de prendre en considération les particularités de la reproduction ou multiplication de la variété (voir l'Introduction générale). ~~La taille de l'échantillon doit être déterminée conformément à des procédures statistiques appropriées.~~

4.2.4 Échantillon d'ADN de référence

~~Il est recommandé d'établir, conformément aux sections 4.1 à 4.3, une~~ Une collection d'échantillons d'ADN de référence peut être établie 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. La méthode d'échantillonnage devrait suivre les procédures recommandées et les critères de qualité pour l'extraction d'ADN doivent être établis. Tous deux doivent être documentés.~~

Les échantillons d'ADN doivent être conservés dans des conditions empêchant leur dégradation (p. ex. stockage à -80 °C). Le transfert d'échantillons d'ADN de référence est décrit dans la section 1 du document TGP/5.

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 descriptions de génotypes et la construction de bases de données doit être techniquement simple à mettre en œuvre, 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 prendre en considération certains critères de qualité concernant, par exemple :

- a) — la qualité de l'ADN;
- b) — les méthodes d'extraction de l'ADN
- c) — les séquences d'amorces utilisées;
- d) — la polymérase à utiliser dans les méthodes fondées sur la PCR;
- e) — en ce qui concerne les méthodes fondées sur la PCR, la quantité ou la concentration de chaque composante de PCR et des autres composantes; et
- f) — les conditions de cycles de PCR.

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 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.

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

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 doit plus être utilisé par la suite. 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 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.~~

~~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 PCR 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).~~

~~4.3 Traitements des données relatives aux séquences~~

~~Un journal détaillé du pipeline de traitement de données pourrait indiquer :~~

- ~~a) le type et les versions des outils;~~
- ~~b) la ligne de commande utilisée pour l'outil, y compris les seuils;~~
- ~~c) la reproductibilité (comptage);~~
- ~~d) les possibilités de partage des données et processus;~~
- ~~e) les données d'alignement brutes (fichiers BAM ou CRAM) devraient être stockées si possible;~~
- ~~f) les fichiers VCF multiéchantillons ne conviennent pas, un fichier VCF par variété doit être présent;~~
- ~~g) si des fichiers VCF sont stockés, toutes les positions (variants et non variants) et leur profondeur devraient être stockées;~~
- ~~h) des approches à la fois heuristiques et probabilistes devraient être envisagées et comparées pour les méthodes de détection;~~
- ~~i) les bases de données devraient faciliter l'entrée et la sortie de données de détection des variants dans un format normalisé (VCF ou BCF);~~
- ~~j) le pipeline de traitement des données devrait aboutir à un fichier journal détaillé qui devrait être stocké conjointement avec les données de détection des variants;~~
- ~~k) si possible, les données brutes devraient être stockées de sorte que le traitement des données puisse être répété avec de nouveaux outils ou avec des outils mis à jour; et~~
- ~~l) la valeur p ou l'incertitude pour un allèle donné devrait être stockée.~~

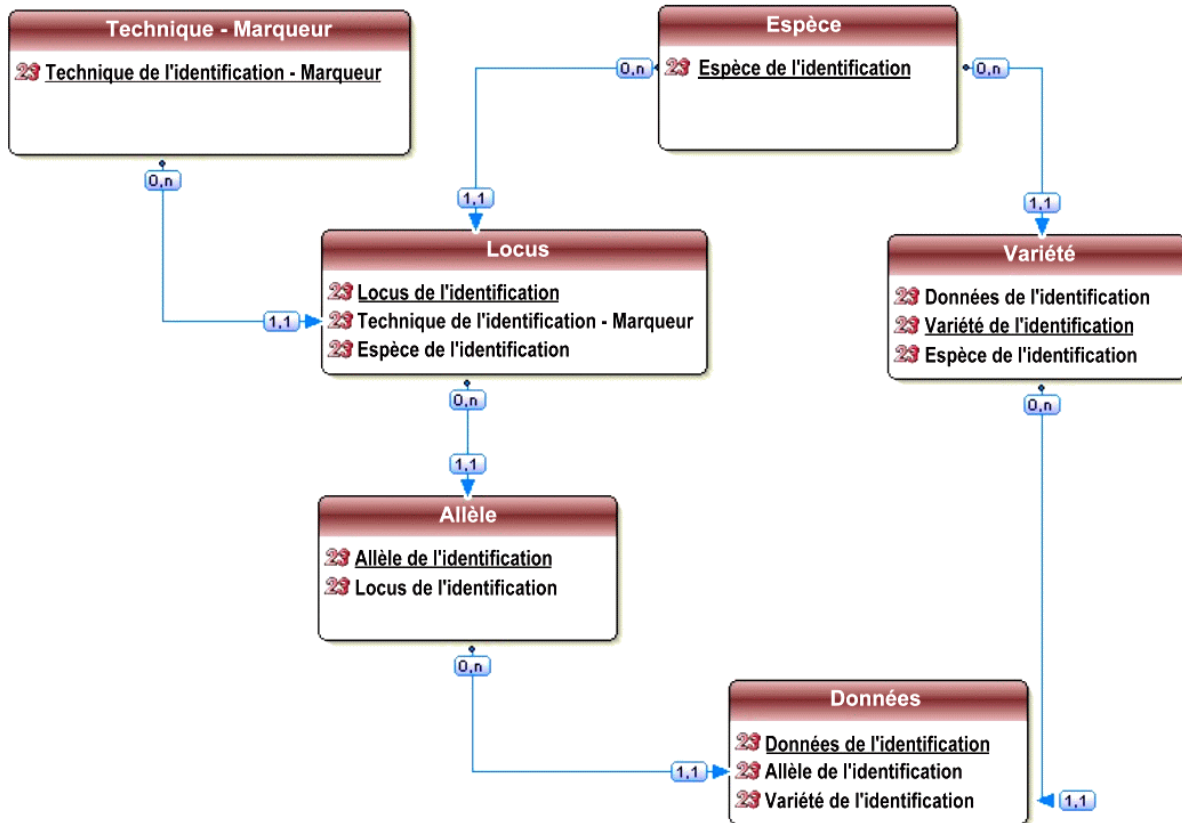
~~6. Bases de données~~

~~6.14.4 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 s'appliquer à toutes les utilisations prévues des données.~~

~~6.24.5 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-méthode de détection des marqueurs, marqueur, locus et allèle. Pour les variants obtenus à partir de données de séquençage, les fichiers VCF peuvent être stockés dans une base de données relationnelle ou non-SQL. Dans ce cas, chaque enregistrement de la base de données correspondant à un variant présente une version génomique, un chromosome, une position et un allèle de référence définis.~~



6-34.6 Dictionnaire de données

4.6-3.1 Dans une base de données, chacun des objets devient un tableau à l'intérieur duquel des champs sont définis. Par exemple :

- Code technique/Type de marqueur** : indique le code ou le nom de la technique ou le type de marqueur utilisé, p. ex. SSR, SNP, etc.
- Position de génome de référence ou Code code locus** : de préférence, une version d'assemblage de génome, un chromosome et une position devraient être fournis si un génome de référence est disponible pour l'espèce concernée, p. ex. SL2.50ch05:63309763 pour la tomate *Solanum lycopersicum*, version d'assemblage 2.50, en position 63309763 sur le chromosome 5. En l'absence de génome de référence ou si l'emplacement est inconnu, on peut utiliser le nom ou le code du locus pour les espèces concernées, p. ex. gwm 149, A2, etc.
- Code allèle** : indique **Génotype** : pour les profils SNP, la composition allélique du SNP ou du MNP devrait être donnée, p. ex. A/T ou A/A. Pour d'autres techniques, le génotype indique le nom ou le code de l'allèle d'un locus donné, pour les espèces concernées, p. ex. 1, 123, etc.
- Valeur des données** : Profondeurs de l'allèle ou valeur des données : pour les SNP obtenus à partir de données de séquençage de nouvelle génération, cela devrait indiquer la profondeur de la couverture des allèles, p. ex. 10/20 pour un allèle A/T, A étant couvert par 10 lectures et T par 20. Dans les autres cas, elle indique une valeur de données pour l'échantillon sur un locus-allèle déterminé, p. ex. 0 (absence), 1 (présence), 0,25 (fréquence), etc.
- Variété** : **dénomination de la variété ou référence de l'obteneur** : la variété est l'objet pour lequel les données sont obtenues.
- Type de variété** : p. ex. lignée endogame ou hybride
- 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 est recommandée pour~~ éviter les problèmes liés aux synonymes ~~et serait~~ ~~avantaguse du point de vue de la coordination.~~

4.6.3.2 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.

6.6.4.7 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.

6.7.5 Analyse Échange 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.

5.1 Scénarios d'échange des données

À des fins de coopération, le modèle de données devrait permettre différents types de scénarios, notamment l'échange de données produites à partir d'un ensemble normalisé de marqueurs pour une plante spécifique (scénario 1), et la recherche et la consultation de données relatives à des variétés sélectionnées, générées à partir du même ensemble normalisé de marqueurs (scénario 2). Les détails techniques des deux scénarios sont décrits dans l'annexe "Scénarios d'échange de données et méthodes de transfert de données".

5.2 Méthodes d'échange de données

5.2.1 La transmission de données relatives à l'empreinte génétique peut inclure toute une série d'informations, comme des locus, des échantillons, l'ADN, des données relatives à l'empreinte et des profils d'empreinte génétique. La méthode de transmission des données doit être déterminée par le contenu à transférer et doit tenir compte des éléments suivants :

- a) le volume de données,
- b) la complexité des données,
- c) les conditions relatives aux fonctions de requête ou de recherche.

Les détails techniques des méthodes de transfert des données sont décrits dans l'annexe "Scénarios d'échange de données et méthodes de transfert de données".

5.2.2 Les formats de données les plus utilisés sont les suivants zip, csv, json et xml. Leurs caractéristiques respectives sont les suivantes :

- 1) Le format zip permet d'obtenir divers fichiers d'information dans le format d'origine et, grâce à son taux de compression élevé et à sa facilité de transmission, il convient aux données volumineuses et complexes.
- 2) Le format csv est plus adapté pour les informations en format de données simple, qui a l'avantage de réduire le nombre de données non valides et d'accélérer la vitesse de traitement.
- 3) Les formats json et xml peuvent contenir des informations relatives aux données de caractères plus complexes et des informations plus redondantes, mais tous deux offrent une bonne lisibilité.

7.6 Résumé

On trouvera ci-après un résumé de la méthode qu'il est recommandé de suivre en vue de l'obtention de profils d'ADN de qualité des variétés, y compris du choix et de l'utilisation des marqueurs moléculaires aux fins de l'élaboration de bases de données centrales de la construction de bases de données moléculaires partagées 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) choisir une méthode plante par plante;
- b) déterminer les types et les sources de marqueurs acceptables;
- c) déterminer les plateformes 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 et approuver 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 éléments de référence (variétés, échantillons d'ADN, allèles de référence, le cas échéant);
- 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 des procédures relatives à l'adjonction de 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). 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 de PCR (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 PCR) 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 AAGGCTAA en ATGGCTAA. 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 PCR.

Séquences polymorphes amplifiées coupées (CAPS)

Les séquences polymorphes amplifiées coupées (CAPS) sont des fragments d'ADN amplifiés par PCR à l'aide d'amorces à 20-25 pb, 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 pb). 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 PCR à l'aide d'amorces spécifiques à 15-30 pb, conçus à partir des séquences polymorphes identifiées préalablement. En faisant appel à des amorces avec PCR 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.

Tresses

Dans l'analyse SSR, le processus de "tresses" est l'addition d'une séquence d'oligonucléotide courte spécifique aux amorces utilisées dans la PCR, 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 PCR.

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 la PCR.

C. LISTE DES SIGLES

<u>API</u>	<u>Application Programming Interface (interface de programmation)</u>
<u>BAM</u>	<u>Binary Alignment Map</u>
<u>BCF</u>	<u>Binary Call Format</u>
<u>CRAM</u>	<u>Compressed Reference-oriented Alignment Map</u>
<u>MNP</u>	<u>Multiple Nucleotide Polymorphism</u>
<u>NGS</u>	<u>Next Generation Sequencing</u>
<u>NIL</u>	<u>Near Isogenic Line</u>
<u>RIL</u>	<u>Recombinant Inbred Line</u>
<u>SAM</u>	<u>Sequence Alignment Map</u>
<u>SNP</u>	<u>Single Nucleotide Polymorphism</u>
<u>SQL</u>	<u>Structured Query Language</u>
<u>SSR</u>	<u>Simple Sequence Repeats</u>
<u>TIFF</u>	<u>Tagged Image File Format</u>
<u>VCF</u>	<u>Variant Call Format</u>

[L'annexe suit]

ANNEXE

SCÉNARIOS D'ÉCHANGE DES DONNÉES ET MÉTHODES DE TRANSFERT DE DONNÉES**A : Scénarios d'échange des données**

Scénario 1 : échange de données produites à partir d'un ensemble normalisé de marqueurs pour une plante spécifique

Afin d'échanger des données sur l'ensemble de marqueurs utilisés pour une plante spécifique, le service Web suivant peut être utilisé :

https://office.org/locus?upov_code={upovcode}&type={marker type}&method={observation method}

Par exemple, pour obtenir des informations sur l'ensemble de marqueurs pour le maïs en utilisant la méthode SSR et CE, il faut accéder à l'URL suivante :

https://office.org/locus?upov_code=ZEAAA_MAY&type=SSR&method=CE

Le résultat serait :

<code>{“techniqueid”:</code>	<code>“examplevariety”:</code>	<code>[“alleleid”: “254/286”,</code>
<code>“CN SSR ZEAA MAY CE V</code>	<code>].</code>	<code>“examplevariety”:</code>
<code>1”,</code>	<code>[“alleleid”: “248/271”,</code>	<code>].</code>
<code>“description”: “Laboratory</code>	<code>“examplevariety”:</code>	<code>[“alleleid”: “256/256”,</code>
<code>method description”</code>	<code>].</code>	<code>“examplevariety”:</code>
<code>[“locusid”: “M01”,</code>	<code>[“alleleid”: “248/290”,</code>	<code>].</code>
<code>“alleles”:</code>	<code>“examplevariety”:</code>	<code>[“alleleid”: “256/264”,</code>
<code>[“alleleid”: “238/256”,</code>	<code>].</code>	<code>“examplevariety”:</code>
<code>“examplevariety”:</code>	<code>[“alleleid”: “250/250”,</code>	<code>].</code>
<code>].</code>	<code>“examplevariety”:</code>	<code>[“alleleid”: “256/266”,</code>
<code>[“alleleid”: “238/271”,</code>	<code>].</code>	<code>“examplevariety”:</code>
<code>“examplevariety”:</code>	<code>[“alleleid”: “250/252”,</code>	<code>].</code>
<code>].</code>	<code>“examplevariety”:</code>	<code>[“alleleid”: “256/271”,</code>
<code>[“alleleid”: “246/246”,</code>	<code>].</code>	<code>“examplevariety”:</code>
<code>“examplevariety”:</code>	<code>[“alleleid”: “250/256”,</code>	<code>].</code>
<code>].</code>	<code>“examplevariety”:</code>	<code>[“alleleid”: “256/284”,</code>
<code>[“alleleid”: “246/248”,</code>	<code>].</code>	<code>“examplevariety”:</code>
<code>“examplevariety”:</code>	<code>[“alleleid”: “250/275”,</code>	<code>].</code>
<code>].</code>	<code>“examplevariety”:</code>	<code>[“alleleid”: “256/286”,</code>
<code>[“alleleid”: “246/250”,</code>	<code>].</code>	<code>“examplevariety”:</code>
<code>“examplevariety”:</code>	<code>[“alleleid”: “252/256”,</code>	<code>].</code>
<code>].</code>	<code>“examplevariety”:</code>	<code>[“alleleid”: “258/258”,</code>
<code>[“alleleid”: “246/254”,</code>	<code>].</code>	<code>“examplevariety”:</code>
<code>“examplevariety”:</code>	<code>[“alleleid”: “252/260”,</code>	<code>].</code>
<code>].</code>	<code>“examplevariety”:</code>	<code>[“alleleid”: “264/284”,</code>
<code>[“alleleid”: “246/256”,</code>	<code>].</code>	<code>“examplevariety”:</code>
<code>“examplevariety”:</code>	<code>[“alleleid”: “252/271”,</code>	<code>].</code>
<code>].</code>	<code>“examplevariety”:</code>	<code>[“alleleid”: “271/292”,</code>
<code>[“alleleid”: “246/260”,</code>	<code>].</code>	<code>“examplevariety”:</code>
<code>“examplevariety”:</code>	<code>[“alleleid”: “252/273”,</code>	<code>].</code>
<code>].</code>	<code>“examplevariety”:</code>	<code>].</code>
<code>[“alleleid”: “246/277”,</code>	<code>].</code>	<code>[“locusid”=“M02”.</code>
<code>“examplevariety”:</code>	<code>[“alleleid”: “252/282”,</code>	<code>“alleles”: [...]</code>
<code>].</code>	<code>“examplevariety”:</code>	<code>}] vi</code>
<code>[“alleleid”: “246/284”,</code>	<code>].</code>	
<code>“examplevariety”:</code>	<code>[“alleleid”: “254/254”,</code>	
<code>].</code>	<code>“examplevariety”:</code>	
<code>[“alleleid”: “246/288”,</code>	<code>].</code>	
<code>“examplevariety”:</code>	<code>[“alleleid”: “254/271”,</code>	
<code>].</code>	<code>“examplevariety”:</code>	
<code>[“alleleid”: “248/250”,</code>	<code>].</code>	
<code>“examplevariety”:</code>	<code>[“alleleid”: “254/284”,</code>	
<code>].</code>	<code>“examplevariety”:</code>	
<code>[“alleleid”: “248/256”,</code>	<code>].</code>	

Scénario 2 : recherche et consultation de données relatives à des variétés sélectionnées, générées à partir du même ensemble normalisé de marqueurs

Afin de rechercher et de consulter les données moléculaires d'une variété, le service Web suivant peut être utilisé :

https://office.org/variety?id={irn}&techniqueid={technique_code} vi

Par exemple,

https://office.org/variety?id=XU_30201800000140 &techniqueid= CN SSR ZEAA MAY CE V 1 vi

Le résultat serait :

```
{ "techniqueid": "CN SSR ZEAA MAY PAGE ",  
  "varietyid": " XU 30201800000140 ",  
  "computationalsteps": "xxxxxxxxxxxx"  
  "data":  
  [  
    { "id": "M01",  
      "value": "254/254"  
    },  
    { "id": "M02",  
      "value": "347/347"  
    },  
    { "id": "M03",  
      "value": "292/292"  
    },  
    { "id": "M04",  
      "value": "361/361"  
    },  
    ...  
  ]  
} vi
```

B : Méthodes de transfert de données

Voici un exemple de création d'un paquet d'empreintes génétiques au format zip pour la transmission de données. Cette méthode doit d'abord utiliser des identifiants indépendants pour identifier les échantillons, l'ADN, les données d'empreinte génétique et le registre des empreintes génétiques. Ensuite, le fichier de données au format json contient tous les locus, tous les échantillons et toutes les informations relatives à l'ADN. Chaque donnée d'empreinte génétique est stockée séparément dans son propre fichier au format json. L'identifiant de l'empreinte génétique sera lié au locus correspondant des données d'empreinte génétique, et tous les fichiers de données d'empreinte génétique et les fichiers de spectre d'empreinte génétique seront stockés séparément dans le répertoire correspondant. La structure du paquet de données d'empreinte génétique est donc la suivante :

```
zip/markers.json  
zip/samples.json  
zip/dnas.json  
zip/genes/gene_id_1.json  
zip/genes/gene_id_2.json  
.....  
zip/genes/gene_id_n.json  
zip/maps/map_id_1.png  
zip/maps/map_id_2.png  
.....  
zip/maps/map_id_m.png
```

Le paquet d'empreintes génétiques en format zip peut être étendu pour inclure plus d'informations. Le cœur du paquet est le fichier de données d'empreinte génétique, qui constitue le cœur de la corrélation, de sorte

que la corrélation entre les parties peut être correctement analysée, ce qui permet la transmission des données entre des systèmes différents.

[Fin de l'annexe et du document]