

**BMT-Richtlinien (proj.9)****ORIGINAL:** Englisch**DATUM:** 17. August 2007**INTERNATIONALER VERBAND ZUM SCHUTZ VON PFLANZENZÜCHTUNGEN**
GENÈVE**ENTWURF****RICHTLINIEN FÜR DIE DNS-PROFILIERUNG:
AUSWAHL MOLEKULARER MARKER
UND
AUFBAU VON DATENBANKEN
(„BMT-RICHTLINIEN“)***Vom Verbandsbüro erstelltes Dokument**zu prüfen vom**Rat auf seiner einundvierzigsten ordentlichen Tagung vom
25. Oktober 2007 in Genf*

INHALTSVERZEICHNIS

A. EINLEITUNG	3
B. ALLGEMEINE GRUNDSÄTZE.....	3
1. Auswahl einer auf molekularen Markern beruhenden Methode	3
2. Auswahl molekularer Marker	3
2.1 <i>Allgemeine Kriterien</i>	3
2.2 <i>Kriterien für spezifische Typen molekularer Marker</i>	3
3. Zugang zur Technik	3
4. Zu analysierendes Material	3
4.1 <i>Quelle des Pflanzenmaterials</i>	3
4.2 <i>Art des Pflanzenmaterials</i>	3
4.3 <i>Probengröße</i>	3
4.4 <i>DNS-Referenzprobe</i>	3
5. Normung der Analyseprotokolle.....	3
5.1 <i>Einleitung</i>	3
5.2 <i>Qualitätskriterien</i>	3
5.3 <i>Evaluierungsphase</i>	3
5.4 <i>Auswertung der molekularen Daten</i>	3
6. Datenbanken	3
6.1 <i>Typ der Datenbank</i>	3
6.2 <i>Datenbankmodell</i>	3
6.3 <i>Liste der Datenbankfelder</i>	3
6.4 <i>Beziehung zwischen den Tabellen</i>	3
6.5 <i>Datentransfer in die Datenbank</i>	3
6.6 <i>Datenzugriff / -eigentum</i>	3
6.7 <i>Datenanalyse</i>	3
6.8 <i>Validierung der Datenbank</i>	3
7. Zusammenfassung.....	3
GLOSSAR	3
Mikrosatelliten oder einfache Sequenzwiederholungen (<i>Simple Sequence Repeats, SSR</i>)	3
Polymorphismen mit einem einzigen Nukleotid (<i>Single Nucleotide Polymorphisms, SNP</i>).....	3
<i>Cleaved Amplified Polymorphic Sequences (CAPS)</i>	3
Sequenzcharakterisierte amplifizierte Regionen (<i>Sequence-Characterized Amplified Regions (SCAR)</i>).....	3
<i>Polymorphic Information Content (PIC)</i> -Werte	3
Primerverlängerung („ <i>Pig-tailing</i> “)	3
Nullallel	3
„ <i>Stotter</i> “-Banden	3

A. EINLEITUNG

Dieses Dokument (BMT-Richtlinien) soll Anleitung zur Entwicklung harmonisierter Methoden geben, um qualitativ hochwertige molekulare Daten für eine Reihe von Anwendungen zu erzeugen. Die BMT-Richtlinien sollen ferner den Aufbau von Datenbanken mit molekularen Profilen von Sorten behandeln, die möglicherweise mit verschiedenen Techniken in verschiedenen Labors erzeugt werden. Dies stellt hohe Anforderungen an die Qualität der Marker und an die notwendige Reproduzierbarkeit der Daten bei der Verwendung dieser Marker in Situationen, in denen sich Ausrüstungen und/oder Reaktionschemikalien ändern könnten. Zudem sind spezifische Vorsichtsmaßnahmen bezüglich der Qualität der in eine Datenbank eingegebenen Daten zu treffen.

In bezug auf die etwaige Verwendung molekularer Marker bei der Prüfung der Unterscheidbarkeit, der Homogenität und der Beständigkeit (DUS) wird die derzeitige Lage in der UPOV in den Dokumenten TC/38/14-CAJ/45/5 und TC/38/14 Add.-CAJ/45/5 Add. erläutert.^a

B. ALLGEMEINE GRUNDSÄTZE

1. Auswahl einer auf molekularen Markern beruhenden Methode

1.1 Die wichtigsten Kriterien für die Auswahl einer Methode sind:

- a) Reproduzierbarkeit der Datengenerierung zwischen Labors und Detektionsmethoden (Ausrüstungstypen);
- b) Wiederholbarkeit im Laufe der Zeit;
- c) Unterscheidungskraft;
- d) Möglichkeiten für die Entwicklung von Datenbanken;
- e) Zugänglichkeit der Methode.

1.2 Da verbesserte Techniken und neue Ausrüstungen verfügbar werden, ist es für die weitere Nachhaltigkeit der Datenbanken wichtig, daß die Interpretation der erzeugten Daten unabhängig von den für ihre Generierung genutzten Ausrüstungen ist. Dies ist beispielsweise der Fall bei DNS-Sequenzierungsdaten. Anfänglich wurden radioaktiv markierte Primer und Sequenzierungsgel für die Generierung dieser Daten verwendet, während dies heute mit fluoreszierenden Farbstoffen und anschließender Auftrennung mit weitgehend automatisierten, Kapillargel-Elektrophoresesystemen mit hoher Durchsatzleistung erfolgt. Trotz dieser Unterschiede sind die mit den verschiedenen Verfahren erzeugten Daten übereinstimmend und unabhängig von den für ihre Generierung genutzten Verfahren. Dies kann auch für Daten gelten, die beispielsweise anhand von DNS-Mikrosatelliten (einfache Sequenzwiederholungen (*Simple Sequence Repeats*, SSR)) oder Polymorphismen in einem einzigen Nukleotid (Single Nucleotide Polymorphisms (SNP)) erzeugt werden. Diese Wiederholbarkeit und Reproduzierbarkeit sind beim Aufbau und Betrieb und für die Lebensdauer der Datenbanken wichtig. Nur so kann eine zentral betriebene Datenbank, die mit verifizierten Daten aus einer Reihe von Quellen bestückt ist, auf kosteneffiziente Weise entwickelt werden, so daß die erhebliche Investition, die für ihre Einrichtung erforderlich ist, nur einmal getätigt wird.

1.3 Die Arten von molekularen Verfahren, die für die Sortenprofilierung problemlos anwendbar sind, werden durch die Anforderung eingeschränkt, daß die Daten wiederholbar,

reproduzierbar und übereinstimmend sein müssen. Obwohl verschiedene Multilocus-DNS-Profilierungsverfahren erfolgreich für die Forschung genutzt wurden, kann bei vielen davon Kodominanz nicht problemlos erfasst werden und die Reproduzierbarkeit von komplexeren Bandenmustern zwischen Labors, die verschiedene Ausrüstungen benutzen, kann problematisch sein. Es besteht die Ansicht, daß diese Faktoren im Kontext der Sortenprofilierung Schwierigkeiten verursachen. Infolgedessen befaßt sich dieses Dokument insbesondere mit Überlegungen und Empfehlungen zu angemessen definierten und erforschten Anwendungen von SSR (Mikrosatelliten) und, im Hinblick auf die Zukunft, zu Sequenzierungsinformationen (d. h. *Single Nucleotide Polymorphisms*, SNP). Weitere Verfahren, die sich auf DNS-Sequenzinformationen stützen, wie CAPS (*Cleaved Amplified Polymorphic Sequences*) und SCAR (*Sequence-Characterized Amplified Regions*) können die obigen Kriterien ebenfalls erfüllen, doch ist ihre Anwendung bei der DNS-Profilierung von Pflanzensorten noch nicht erforscht.

2. Auswahl molekularer Marker

2.1 Allgemeine Kriterien

Folgende allgemeinen Kriterien für die Wahl eines spezifischen Markers oder Markersets sollen für molekulare Marker, ungeachtet der Verwendung der Marker, geeignet sein, obwohl eingeräumt wird, daß spezifische Anwendungen bestimmte zusätzliche Kriterien aufzwingen könnten:

- a) nutzbares Polymorphismusniveau (angegeben beispielsweise durch den an einer Serie repräsentativer Sorten erzielten PIC-Wert (Informationsinhalt der Polymorphismen (*Polymorphism Information Content*: siehe Glossar));
- b) Wiederholbarkeit innerhalb und Reproduzierbarkeit zwischen Labors in bezug auf die Auswertung der Daten;
- c) bekannte Verteilung der Marker im ganzen Genom (d.h. Kartenposition), was zwar nicht wesentlich, jedoch eine wertvolle Information ist und hilft, die Auswahl von gekoppelten Markern zu vermeiden;
- d) nach Möglichkeit Vermeiden von Markern mit „Nullallelen“ (d. h. ein Allel, dessen Wirkung das Fehlen eines PCR-Produkts auf Molekularebene ist), was ebenfalls nicht wesentlich, jedoch ratsam ist.

2.2 Kriterien für spezifische Typen molekularer Marker

2.2.1 Mikrosatellitenmarker (SSR)

2.2.1.1 Die Analyse der einfachen Sequenzwiederholungen (SSR oder Mikrosatelliten: siehe Glossar) anhand der Polymerase-Kettenreaktion (PCR) wird heute bereits genutzt und hat verschiedene Vorteile.

2.2.1.2 SSR-Marker sind kodominant ausgeprägt, in der Regel einfach zu erfassen und können problemlos kartiert werden. Es wurde nachgewiesen, daß sie in verschiedenen Labors analysiert werden können und (bei ausreichender Beachtung der Bedingungen des Experiments) robust und wiederholbar sind. Zudem können sie mit automatisierten nichtradioaktiven DNS-Sequenzierern mit hoher Durchsatzleistung mittels Gel-Elektrophorese oder der Kapillar-Elektrophorese analysiert werden, und mehrere können gleichzeitig analysiert werden (Multiplexing).

2.2.1.3 Für eine effektive Mikrosatellitenanalyse ist die Auswahl hochqualitativer Marker von wesentlicher Bedeutung. Hierzu ist u. a. zu beachten:

- a) Grad des „Stotterns“ (Generierung einer Serie von einer oder mehreren Banden, die sich in der Größe durch 1 Wiederholungseinheit unterscheiden);
- b) (n+1)-Peaks; die Taq-Polymerase fügt häufig 1 bp zum Ende eines Fragments hinzu. Dies kann durch Verwendung von verlängerten („pigtailed“) Primern verhindert werden (siehe Glossar);
- c) Größe des Amplifikationsprodukts;
- d) effektive Auftrennung zwischen den verschiedenen Allelen mit verschiedenen Detektionsmethoden;
- e) zuverlässige und reproduzierbare Auswertung der Allele mit verschiedenen Detektionsmethoden;
- f) Polymorphismusniveau (Anzahl erfaßter Allele) zwischen Sorten (bitte beachten, daß dies eine Analyse einer erheblichen Anzahl Sorten voraussetzt);
- g) Vermeiden von Kopplung.

2.2.1.4 Für die Auswertung von SSR in verschiedenen Labors und die Nutzung verschiedener Detektionssysteme ist es entscheidend, daß Referenzallele (d. h. Sortenserien) festgelegt und in alle Analysen einbezogen werden. Diese Referenzallele sind notwendig, weil sich die Molekulargewichtsstandards in den verschiedenen derzeit verfügbaren Detektionsmethoden unterschiedlich verhalten und daher für die Allelidentifikation nicht geeignet sind.

2.2.1.5 Die in einem bestimmten Labor verwendeten Primer sollten durch einen gesicherten Lieferanten synthetisiert werden, um die Möglichkeit zu reduzieren, daß verschiedene DNS-Profile aus Primern erzeugt werden, die durch verschiedene Quellen synthetisiert wurden.

2.2.2 *Polymorphismen mit einem einzigen Nukleotid (Single Nucleotide Polymorphism, SNP)*

Polymorphismen mit einem einzigen Nukleotid (*Single Nucleotide Polymorphisms*, (SNP): (siehe Glossar) können anhand der DNS-Sequenzierung nachgewiesen werden. Dies ist ein Routineverfahren, das im allgemeinen eine sehr hohe Wiederholbarkeit über die Zeit sowie Reproduzierbarkeit zwischen den Labors zeigt. Der Nachweis spezifischer SNP wird jedoch zur Zeit anhand einer Reihe von Verfahren vorgenommen, von denen viele noch nicht Routine sind. Die SNP haben bei diploiden Pflanzen naturgemäß lediglich zwei allelische Stufen, während dies bei Polyploiden variieren kann, bei denen es Dosiereffekte geben wird. Dies macht die Auswertung der SNP verhältnismäßig unkompliziert und zuverlässig und dürfte viele der obenerwähnten Probleme reduzieren oder beseitigen. Das bedeutet auch, daß möglicherweise zahlreiche Marker entweder einzeln oder in Multiplexen analysiert werden müssen, um die effiziente und effektive Profilierung eines bestimmten Genotyps zu ermöglichen.

3. **Zugang zur Technik**

Einzelne molekulare Marker und Materialien sind öffentlich verfügbar. Es sind jedoch hohe Investitionen erforderlich, um beispielsweise hochqualitative SSR-Marker zu erzielen. Infolgedessen können Marker und andere Verfahren und sonstiges Material durch Rechte des geistigen Eigentums geschützt sein. Die UPOV erarbeitete eine Anleitung zur Nutzung der Produkte oder Methodiken, die Gegenstand von Rechten des geistigen Eigentums sind

(vergleiche Dokument TGP/7/1 Anlage 3, GN 14), und diese Anleitung sollte für diese Richtlinien befolgt werden. Es ist zu empfehlen, daß Angelegenheiten bezüglich der Rechte des geistigen Eigentums zu Beginn jeder Entwicklungsarbeit behandelt werden.

4. Zu analysierendes Material

Quelle und Art des Materials und die Anzahl der zu analysierenden Proben sind die Hauptaspekte hinsichtlich des zu analysierenden Materials.

4.1 Quelle des Pflanzenmaterials

Das zu analysierende Pflanzenmaterial sollte eine authentische, repräsentative Probe der Sorte sein und nach Möglichkeit aus der Probe der für die Prüfung im Hinblick auf die Erteilung von Züchterrechten oder auf die amtliche Eintragung verwendeten Sorte stammen. Die Verwendung von Proben des für die Prüfung im Hinblick auf die Erteilung von Züchterrechten oder auf die amtliche Eintragung eingereichten Materials erfordert gegebenenfalls die Genehmigung der zuständigen Behörde, des Züchters und/oder des Erhaltungszüchters. Die einzelnen Pflanzen, von denen Proben entnommen wurden, sollten rückverfolgbar sein, falls sich einige der Pflanzen im Nachhinein als nicht repräsentativ für die Sorte erweisen.

4.2 Art des Pflanzenmaterials

Die Art des Pflanzenmaterials, dem Proben zu entnehmen sind, und das Verfahren für die Entnahme von Proben des Materials für die DNS-Extraktion wird weitgehend von der betreffenden Pflanze oder Art abhängen. Bei samenvermehrten Sorten beispielsweise kann das Saatgut als Quelle der DNS verwendet werden, während die DNS bei vegetativ vermehrten Sorten aus dem Blattmaterial extrahiert werden kann. Welches auch immer die Quelle des Materials ist, sollte das Verfahren für die Probenentnahme und die DNS-Extraktion genormt und dokumentiert werden. Zudem sollte überprüft werden, daß die Verfahren für die Probenahme und die Extraktion bei der DNS-Analyse übereinstimmende Ergebnisse zeitigen.

4.3 Probengröße

Es ist wesentlich, daß die für die Analyse entnommenen Proben für die Sorte repräsentativ sind.

4.3.1 Vegetativ vermehrte Sorten

Grundsätzlich könnte für vegetativ vermehrte Sorten eine einzige Probe analysiert werden, da alle Individuen identisch sein sollten. Es ist jedoch in allen Fällen ratsam, mindestens Doppelproben zu analysieren. Werden Unterschiede festgestellt, sollten weitere Proben analysiert werden.

4.3.2 Selbstbefruchtende und überwiegend selbstbefruchtende Sorten

Die Annahme, daß selbstbefruchtende und überwiegend selbstbefruchtende Sorten an allen Loci, einschließlich der SSR- oder SNP-Loci, homozygot sind, ist nicht immer angebracht. Eine Heterogenität kann sich beispielsweise aus Rest-Heterozygotie, Fremdbefruchtung oder physischer Beimischung ergeben. Aus diesem Grunde wird in der Regel empfohlen, eine

Reihe einzelner Samen zu analysieren, da dadurch eine etwaige derartige Heterozygotie nachgewiesen werden kann. Es kann jedoch Gründe, einschließlich Kostengründen, dafür geben, eine Mischprobe einer vereinbarten Anzahl Individuen zu analysieren, um das DNS-Profil einer Sorte darzustellen.

4.3.3 *Fremdbefruchtende Sorten*

Ähnliche Überlegungen gelten grundsätzlich für Sorten fremdbefruchtender Arten. Es wird allgemein empfohlen, daß Einzelsamen/-pflanzen analysiert werden, weil die Unterschiede zwischen den Sorten das Ergebnis von Unterschieden bei den Allel- (oder Banden-) Frequenzen sowie des Vorhandenseins oder Fehlens bestimmter Allele/Banden sein können. Es könnte jedoch Gründe, u. a. die Kosten, dafür geben, eine Mischprobe einer vereinbarten Anzahl Individuen zu analysieren, um ein Gesamt-DNS-Profil einer Sorte darzustellen.

4.3.4 *Hybriden*

Das geeignete Verfahren um sicher zu stellen, daß die für die Analyse von Hybriden entnommenen Proben für die Sorte repräsentativ sind, wird vom Typ der Hybride abhängen. Es sollte beachtet werden, daß Hybridsorten an den Loci, die für DNS-Marker kodieren, heterozygot sein werden, jedoch noch immer phänotypisch homogen sein könnten. Die Zahl der für die Analyse gewählten Proben wird von dem genauen Aspekt, der behandelt wird, und dem Typ der geprüften Hybride abhängen. Diesbezüglich sollten die Informationen über verschiedene Typen von Hybridsorten in Dokument TG/1/3, „Allgemeine Einführung Zur Prüfung auf Unterscheidbarkeit, Homogenität und Beständigkeit und Erarbeitung harmonisierter Beschreibungen von neuen Pflanzensorten“ (vergleiche Kapitel 6.4.3), berücksichtigt werden.

4.4 DNS-Referenzprobe

Es wird empfohlen, gemäß den Abschnitten 4.1 bis 4.3 eine DNS-Referenzprobensammlung aus dem Pflanzenmaterial anzulegen, dem Proben entnommen werden. Dies hat den Vorteil, daß die DNS-Referenzprobe gelagert und anderen zur Verfügung gestellt werden kann. Die DNS-Probe sollte so gelagert werden, daß ein Zerfall verhindert wird.

5. Normung der Analyseprotokolle

5.1 Einleitung

Dieses Dokument soll nicht detaillierte technische Protokolle für die Erstellung von DNS-Profilen von Sorten bereitstellen. Grundsätzlich kann jede geeignete Analysemethodik angewandt werden, doch ist es wichtig, daß die Methodik angemessen validiert ist. Dies kann über ein international anerkanntes Validierungsverfahren oder durch Entwicklung eines methodenspezifischen Vorgehens erfolgen. In beiden Fällen gibt es einige zweckdienliche allgemeine Überlegungen vorhanden.

Jedes für die Bestimmung eines Genotyps und den Aufbau von Datenbanken angewandte Verfahren sollte aus technischer Sicht einfach durchzuführen, zuverlässig und robust sein und eine problemlose und eindeutige Auswertung der Markerprofile in verschiedenen Labors ermöglichen. Dies setzt ein Normungsniveau voraus, beispielsweise bei der Auswahl der Marker, der Referenzallele und der Allelbenennung/-auswertung

5.2 Qualitätskriterien

5.2.1 Es ist wichtig, sich auf bestimmte Qualitätskriterien zu einigen, die beispielsweise folgendes betreffen:

- a) die Qualität der DNS;
- b) die verwendeten Primersequenzen;
- c) die bei PCR-basierten Methoden zu verwendende Polymerase;
- d) für PCR-basierte Methoden die Menge/Konzentration jeder PCR-Komponente und anderer Komponenten;
- e) PCR-Durchlaufbedingungen.

5.2.2 Die detaillierte Methode sollte in einem Protokoll erläutert werden.

5.3 Evaluierungsphase

5.3.1 *Einleitung*

Zur Auswahl geeigneter Marker und Erstellung zulässiger Laborprotokolle für eine gegebene Art wird eine vorläufige Evaluierungsphase empfohlen, an der mehr als ein Labor beteiligt ist (d. h. ein international anerkanntes Validierungsverfahren, z. B. ein Ringtest gemäß international vereinbarten Normen). Diese Phase sollte sich hauptsächlich mit der Auswahl eines Markersets befassen und wird in der Regel die Evaluierung bestehender Marker umfassen, die entweder veröffentlicht wurden oder über andere Quellen verfügbar sind. Die Zahl der zu bewertenden Marker wird variieren und hängt von den Möglichkeiten der jeweiligen Art ab. Die Marker sollten aus zuverlässigen Quellen (z. B. „peer-reviewed“ Veröffentlichungen) und von gesicherten Lieferanten stammen. Die endgültige Wahl einer zu evaluierenden Anzahl wird am Ende des Prozesses ein Gleichgewicht zwischen Kosten und der Anforderung sein, ein zufriedenstellendes Set vereinbarter Marker zu erzeugen. Ziel ist es, ein vereinbartes Markersets zu erzeugen, das in verschiedenen Labors, die potentiell unterschiedliche Arten von Ausrüstungen und verschiedene Quellen chemischer Reagenzien usw. verwenden, zuverlässig und reproduzierbar analysiert, ausgewertet und erfaßt werden kann.

5.3.2 *Wahl der Sorte*

Als Grundlage für die Evaluierungsphase sollte eine angemessene Anzahl Sorten aufgrund der genetischen Variabilität innerhalb der Arten und des Typs der betreffenden Sorten ausgewählt werden. Die Wahl der Sorten sollte eine angemessene Diversitätsspanne reflektieren und nach Möglichkeit einige verwandte und einige morphologisch ähnliche Sorten einschließen, um die Beurteilung des Unterscheidungsniveaus in diesen Fällen zu ermöglichen.

5.3.3 *Interpretation der Ergebnisse*

Die nächste Evaluierungsphase sollte nach Möglichkeit ein international anerkanntes Validierungsverfahren einschließen, damit die gesamte Methode objektiv beurteilt werden kann. Marker, die in einem der an dieser Evaluierungsphase beteiligten Labors Schwierigkeiten verursachen, sollten für die spätere Verwendung zurückgewiesen werden. Da sich, wie die empirische Erfahrung zeigt, die meisten Fehler bei der Analyse großer Sortensammlungen aus Auswertungsfehlern zu ergeben scheinen, sollte der Aufbau von Datenbanken idealerweise auf Doppelproben gestützt werden (z. B. verschiedene Unterproben

von Saatgut derselben Sorte), die in verschiedenen Labors analysiert werden. Da die Unterproben (oder DNS-Extrakte aus diesen) im Falle einer Diskrepanz ausgetauscht werden können, ist dieses Verfahren bei der Feststellung von Fehlern bei der Probenentnahme oder von Fehlern aufgrund der Heterogenität innerhalb der Proben äußerst wirksam und eliminiert etwaige Laborartefakte (systemabhängige Fehler).

5.4 Auswertung der molekularen Daten

In Verbindung mit der Evaluierungsphase sollte ein Protokoll für die Allel-/Bandenauswertung erstellt werden. Das Protokoll sollte sich mit der Art und Weise der Auswertung folgender Daten befassen:

- a) seltene Allele (d. h. diejenigen an einem spezifische Locus, die mit einer Häufigkeit unter einem vereinbarten Schwellenwert (in der Regel 5-10 %) in einer Population) auftreten;
- b) Nullallele (ein Allel, dessen Wirkung das Fehlen eines PCR-Produkts auf Molekularebene ist);
- c) „schwache“ Banden (d. h. Banden, bei denen die Intensität unter einen vereinbarten Schwellenwert für die Erfassung fällt, der entweder empirisch oder automatisch festgelegt wird und dessen Auswertung anfechtbar sein kann);
- d) fehlende Daten (d. h. Loci, für die aus welchem Grund auch immer für eine Sorte oder Sorten keine Daten erfaßt wurden);
- e) monomorphe Banden (diejenigen Allele/Banden, die bei jeder analysierten Sorte auftreten, d. h. in einer bestimmten Sortensammlung nicht polymorph sind).

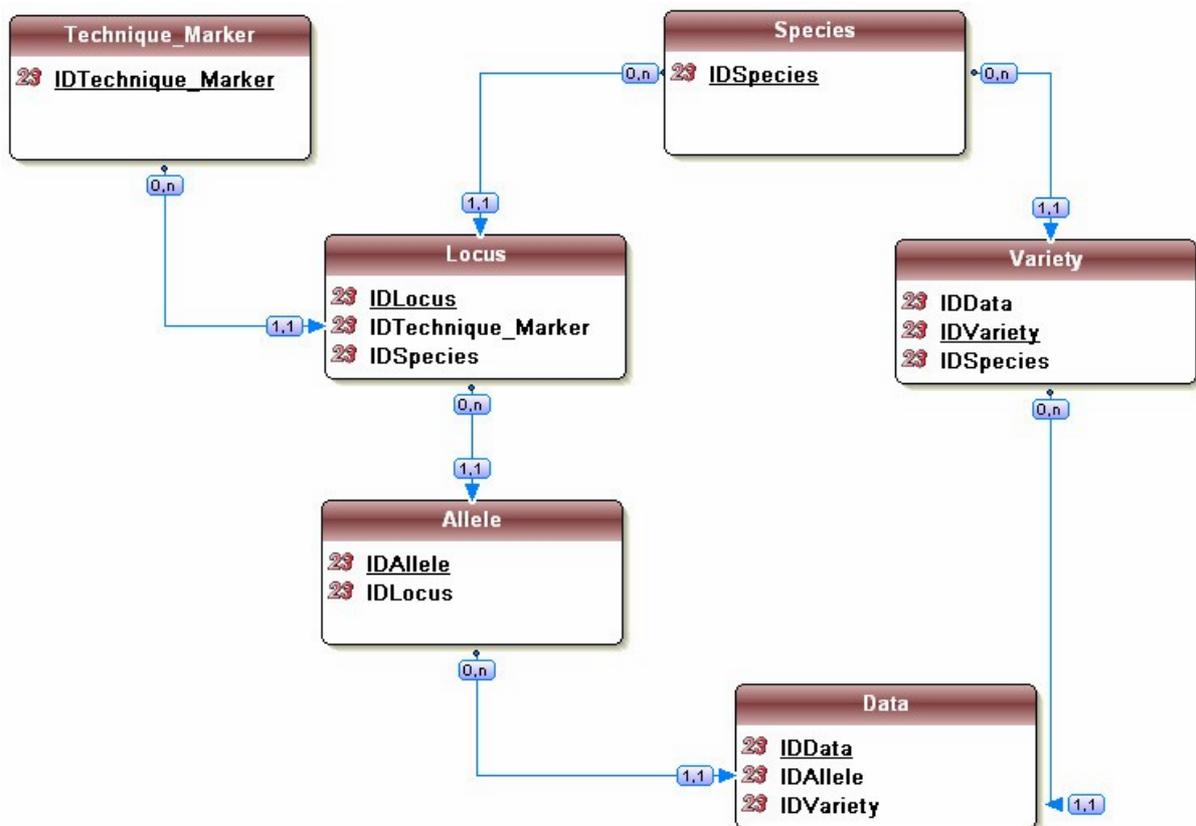
6. Datenbanken

6.1 Typ der Datenbank

Molekulare Daten können auf zahlreiche Arten gespeichert werden. Deshalb ist es wichtig, daß die Struktur der Datenbank so entwickelt wird, daß sie mit allen beabsichtigten Verwendungen von Daten kompatibel ist.

6.2 Datenbankmodell

Das Datenbankmodell sollte von IT-Datenbankexperten zusammen mit den Benutzern der Datenbank festgelegt werden. Das Datenbankmodell sollte mindestens sechs Kernobjekte enthalten: Art, Sorte, Verfahren, Marker, Locus und Allel.



Verfahren_Marker
IDVerfahren_Marker

Art
IDArt

Locus
IDLocus
IDVerfahren_Marker
IDArt

Sorte
IDDaten
IDSorte
IDArt

Allele
IDAllele
IDLocus

Daten
IDDaten
IDAllele
IDSorte

6.3 Liste der Datenbankfelder

6.3.1 In einer Datenbank wird jedes der Objekte zu einer Tabelle, in der Felder festgelegt sind, beispielsweise:

a) Verfahrens-/Markercode:

Gibt den Code oder den Namen des Verfahrens oder den Typ des verwendeten Markers an, z. B. *SSR, SNP usw.*

b) Locuscode:

Gibt den Namen oder den Code des Locus für die betreffende Art an, z. B. *gwm 149, A2 usw.*

c) Allelcode:

Gibt den Namen oder den Code eines Allels eines gegebenen Locus für die betreffende Art an, z. B. *1, 123 usw.*

d) Datenwert:

Gibt einen Datenwert für eine gegebene Probe auf einem gegebenen Locus-Allel an, z. B. *0 (Fehlen), 1 (Vorhandensein), 0,25 (Häufigkeit) usw.*

e) Sorte:

Die Sorte ist das Objekt, für das die Daten erlangt wurden.

f) Art:

Die Art wird durch den botanischen Namen oder den landesüblichen Namen angegeben, der sich mitunter auch auf den Sortentyp bezieht (z. B. Verwendung, Winter/Sommertyp usw.). Die Verwendung des UPOV-Codes würde die Probleme mit Synonymen lösen und wäre daher für die Koordinierung von Vorteil.

6.3.2 In jeder Tabelle müssen die Zahl der Felder, ihr Name und ihre Definition, die möglichen Werte und die zu befolgenden Regeln in der „Liste der Datenbankfelder“ festgelegt werden.

6.4 Beziehung zwischen den Tabellen

6.4.1 Die Verknüpfungen zwischen den Tabellen sind ein wichtiger Aspekt der Datenbankkonstruktion. Die Verknüpfungen zwischen den Tabellen lassen sich wie folgt veranschaulichen:

Tabelle	Verknüpfung	Tabelle	Beschreibung
Frau	0 oder 1 bis n (0, n)	Kind	0: Eine Frau hat möglicherweise kein Kind 1 bis n: Eine Frau kann 1 bis n Kinder haben (sie ist dann eine Mutter)
Kind	1 bis 1 (1,1)	Frau	Ein bestimmtes Kind hat nur eine biologische Mutter

6.4.2 Die nachstehende Tabelle gibt die Beziehung zwischen den sechs minimalen Kernobjekten an, wie im Datenbankmodell in Abschnitt 6.2 vorgeschlagen:

Tabelle	Verknüpfung	Tabelle	Beschreibung
Verfahren/Marker	0 oder 1 bis n	Locus	0: Ein Verfahren/Marker kann in ‚Verfahren/Marker‘ vorhanden sein, selbst wenn in der Datenbank noch kein Locus/Allel verwendet wird 1 bis n: Ein gegebener Markertyp kann 1 bis n nützliche Loci ergeben
Locus	1 bis 1	Verfahren/Marker	Ein gegebener Locus wird im Rahmen eines gegebenen Verfahrens/Markers definiert
Locus	1 bis n	Allel	Für jeden Locus kann 1 oder mehr als 1 Allel beschrieben werden
Allel	1 bis 1	Locus	Ein gegebenes Allel wird im Rahmen eines gegebenen Locus definiert
Allel	0 oder 1 bis n	Daten	0: Ein gegebenes Allel kann definiert werden, jedoch ohne Daten 1 bis n: Ein gegebenes Allel kann in 1 bis n Daten gefunden werden
Daten	1 bis 1	Allel	Die Daten entsprechen einem gegebenen Allel
Sorte	0 oder 1 bis n	Daten	0: Die Sorte hat keine Daten 1 bis n: Die Sorte hat Daten
Daten	1 bis 1	Sorte	Die Daten entsprechen einer gegebenen Sorte
Daten	1 bis 1	Art	Die Daten werden für eine gegebene Sorte und dann für die Art der Sorte erlangt
Art	0 oder 1 bis n	Daten	0: Eine Art hat möglicherweise keine Daten 1 bis n: Eine Art kann 1 bis n Daten haben.

6.5 Datentransfer in die Datenbank

Zur Reduzierung der Fehlerzahl beim Datentransfer und der Transkription ist es ratsam, den Datentransfer in die Datenbanken möglichst weitgehend zu automatisieren.

6.6 Datenzugriff / -eigentum

Es wird empfohlen, daß alle Angelegenheiten bezüglich des Eigentums der Daten und des Zugriffs zu den Daten in der Datenbank zu Beginn der Arbeit behandelt werden.

6.7 Datenanalyse

Das Analyseverfahren wird durch den Zweck bestimmt, zu dem die Daten analysiert werden. In diesen Richtlinien werden daher keine ausdrücklichen Empfehlungen abgegeben.

6.8 Validierung der Datenbank

Nach Abschluß der ersten Phase der Datenbank wird empfohlen, einen „Blindtest“ durchzuführen, d. h. eine Reihe von Proben an verschiedene Labors zu schicken und diese zu ersuchen, das vereinbarte Protokoll zusammen mit der Datenbank anzuwenden, um die Proben zu identifizieren.

7. Zusammenfassung

Nachstehend ist eine Zusammenfassung des für die Auswahl und die Verwendung molekularer Marker empfohlenen Vorgehens im Hinblick auf den Aufbau zentraler und nachhaltiger Datenbanken für DNS-Profile von Sorten wiedergegeben (d. h. Datenbanken, die künftig aus einer Reihe von Quellen, unabhängig von der angewandten Technik, bestückt werden können).

- a) Prüfung des Vorgehens nach Pflanzen;
- b) Einigung auf einen akzeptierten Markertyp und die Quelle;
- c) Einigung auf zulässige Detektionsmethoden/-ausrüstungen;
- d) Einigung auf die an der Prüfung zu beteiligenden Labors;
- e) Einigung auf Qualitätsaspekte (vergleiche Abschnitt 5.2);
- f) Überprüfung der Quelle des verwendeten Pflanzenmaterials (vergleiche Abschnitt 4);
- g) Einigung auf die Marker, die in einer vorläufigen kollaborativen Evaluierungsphase verwendet werden sollen, in die mehr als ein Labor und verschiedene Detektionsmethoden einbezogen werden (vergleiche Abschnitt 2);
- h) Durchführung einer Evaluierung (vergleiche Abschnitt 5.3);
- i) Erstellung eines Protokolls für die Auswertung der molekularen Daten (vergleiche Abschnitt 5.4);
- j) Einigung auf das Pflanzenmaterial/Referenzset, das zu analysieren ist, und auf die Quelle(n);
- k) Analyse der vereinbarten Sortensammlung in verschiedenen Labors/verschiedenen Detektionsmethoden anhand von Doppelproben und Austausch von Proben/DNS-Extrakten, wenn Probleme auftreten;
- l) Verwendung von Vergleichssorten/DNS-Proben/Allelen bei allen Analysen;
- m) Überprüfung aller Stadien (einschließlich der Dateneingabe) – möglichst weitreichende Automatisierung;
- n) Durchführung eines ‚Blindtests‘ in verschiedenen Labors anhand der Datenbank;
- o) Annahme der Verfahren zur Hinzufügung neuer Daten.

GLOSSAR

Mikrosatelliten oder einfache Sequenzwiederholungen (*Simple Sequence Repeats, SSR*)

Mikrosatelliten oder einfache Sequenzwiederholungen (SSR) sind doppelt wiederholte DNS-Sequenzen, in der Regel mit einer Wiederholungseinheit von 2 bis 8 Basenpaaren (z. B. GA, CTT und GATA). Bei vielen Arten wurde nachgewiesen, daß für einige Mikrosatelliten multiple Allele vorhanden sind, die sich aus Unterschieden in der Wiederholungszahl dieser Wiederholungseinheit ergeben. Mikrosatelliten können mit PCR anhand spezifischer Primer analysiert werden. Dieses Verfahren wird als Mikrosatelliten mit sequenzmarkierten Loci (*Sequence-tagged-site Microsatellites, STMS*) bezeichnet. Die Allele (PCR-Produkte) können dann mittels Agarosegel- oder Polyacrylamidgel-Elektrophorese aufgetrennt werden. Zur Entwicklung von Mikrosatelliten mit sequenzmarkierten Loci werden Informationen über die Sequenz der DNS benötigt, die den Mikrosatelliten seitlich begrenzt. Diese Informationen können mitunter aus bestehenden Datenbanken für DNS-Sequenzen beschafft werden. Andernfalls müssen sie empirisch erzielt werden.

Polymorphismen mit einem einzigen Nukleotid (*Single Nucleotide Polymorphisms, SNP*)

Polymorphismen mit einem einzigen Nukleotid (SNP) (ausgesprochen wie „snips“) sind Sequenzvariationen, die auftreten wenn ein einziges Nukleotid (A, T, C oder G) in der Genomsequenz verändert ist. Ein SNP könnte beispielsweise die DNS-Sequenz A**A**GGCTAA in A**T**GGCTAA ändern. Damit eine Variation als SNP angesehen wird, muß sie im allgemeinen in mindestens 1 % der Population auftreten. Die potentielle Zahl der SNP-Marker ist sehr hoch, was bedeutet, daß es möglich sein sollte, sie in allen Teilen des Genoms nachzuweisen. SNP können sowohl in kodierenden (Gen-) Regionen als auch nichtkodierenden Regionen des Genoms auftreten. Der Nachweis der SNP beinhaltet eine vergleichende Sequenzierung einer Anzahl Individuen aus einer Population. Potentielle SNP werden in der Regel eher aus verfügbaren Sequenz-Datenbanken durch Vergleich aneinander ausgerichteter Sequenzen identifiziert. Sie können zwar durch verhältnismäßig einfache PCR- und Gel-Elektrophorese nachgewiesen werden, doch werden zur Zeit Verfahren mit hohem Datendurchlauf und Mikroanordnungen für die gleichzeitige automatische Auswertung Hunderter von SNP-Loci entwickelt.

Cleaved Amplified Polymorphic Sequences (CAPS)

Cleaved Amplified Polymorphic Sequences (CAPS) sind DNS-Fragmente, die durch PCR anhand spezifischer Primer mit 20-25 bp amplifiziert werden, gefolgt von Verdau mit einer Restriktions-Endonuklease. Dann werden die Längen-Polymorphismen, die sich aus der Variation beim Auftreten an den Restriktions-Loci ergeben, durch Gel-Elektrophorese der extrahierten Produkte identifiziert. Im Vergleich zu Markern wie RFLP sind die Polymorphismen wegen der geringen Größe der amplifizierten Fragmente (300-1800 bp) schwieriger zu identifizieren. Die CAPS-Analyse erfordert jedoch keine Southern-blot-Hybridisierung und keinen radioaktiven Nachweis. CAPS wurden bisher in der Regel vorwiegend für Genkartierungsstudien angewandt.

Sequenzcharakterisierte amplifizierte Regionen (*Sequence-Characterized Amplified Regions* (SCAR))

Sequenzcharakterisierte amplifizierte Regionen (*Sequence-Characterized Amplified Regions*, SCAR) sind DNS-Fragmente, die durch PCR anhand spezifischer Primer mit 15-30 bp amplifiziert werden, die aus zuvor identifizierten polymorphen Sequenzen erzeugt wurden. Durch die Verwendung längerer PCR-Primer umgehen die SCAR-Marker das Problem der geringen Reproduzierbarkeit. Sie sind in der Regel auch kodominante Marker. SCAR sind locuspezifisch und wurden für Genkartierungsstudien und die markerunterstützte Selektion angewandt.

***Polymorphic Information Content* (PIC)-Werte**

Polymorphic Information Content (PIC)-Werte sind ein Maßstab für die Alleldiversität an einem Locus, der für die Schätzung und den Vergleich der Unterscheidungskraft molekularer Marker angewandt wird. Der PIC-Wert eines PCR-basierten Markers kann wie folgt berechnet werden:

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

wobei P_{ij} die Häufigkeit des j . PCR-Musters für Genotyp i ist.

Primerverlängerung („Pig-tailing“)

Bei der SSR-Analyse ist das „pig-tailing“ die Addition einer kurzen spezifischen Oligonukleotid-Sequenz zu den beim PCR-Verfahren verwendeten Primern als Mittel zur Verbesserung der Klarheit der Amplifikationsprodukte und zur Reduzierung der Artefakte.

Nullallel

Bei der SSR-Analyse ist ein „Nullallel“ ein Allel an einem bestimmten Locus, dessen Wirkung das Fehlen eines PCR-Produkts ist.

„Stotter“-Banden

Bei der SSR-Analyse bedeuten „Stotterbanden“ das Auftreten einer Serie von einer oder mehreren Banden, die sich nach der PCR in der Größe durch 1 Wiederholungseinheit unterscheiden.

^a Änderung vom Technischen Ausschuß nach dessen dreiundvierzigster Tagung auf dem Schriftweg vereinbart.