

Technischer Ausschuss

TC/56/13

**Sechshundfünfzigste Tagung
Genf, 26. und 27. Oktober 2020**

**Original: englisch
Datum: 9. Oktober 2020**

**ÜBERARBEITUNG VON DOKUMENT UPOV/INF/17 „RICHTLINIEN FÜR DIE DNS-PROFILIERUNG:
AUSWAHL MOLEKULARER MARKER UND AUFBAU VON DATENBANKEN (‘BMT-RICHTLINIEN’)”**

vom Verbandsbüro erstelltes Dokument

Haftungsausschluss: dieses Dokument gibt nicht die Grundsätze oder eine Anleitung der UPOV wieder

ZUSAMMENFASSUNG

1. Zweck dieses Dokuments ist es, einen Vorschlag für die Überarbeitung von Dokument UPOV/INF/17 „Richtlinien für die DNS-Profilierung: Auswahl molekularer Marker und Aufbau von Datenbanken (‘BMT-Richtlinien’), wie in Dokument UPOV/INF/17/2 Draft 4 und in der Anlage dieses Dokuments dargelegt, zu unterbreiten.

2. Der TC wird ersucht:

a) die vorgeschlagene Überarbeitung von Dokument UPOV/INF/17 auf der Grundlage von Dokument UPOV/INF/17/2 Draft 4 zu prüfen;

b) die TWP zu ersuchen, auf ihren Tagungen im Jahr 2021 den Entwurf einer Überarbeitung von Dokument UPOV/INF/17/1 (Dokument UPOV/INF/17/2 Draft 5) zu prüfen; und

c) zur Kenntnis zu nehmen, dass dem Rat auf seiner fünfundfünfzigsten Tagung vom 29. Oktober 2021, vorbehaltlich der Zustimmung des TC auf seiner siebenundfünfzigsten Tagung und des CAJ auf seiner achtundsiebzigsten Tagung im Jahr 2021, ein Entwurf einer Überarbeitung von Dokument UPOV/INF/17 (UPOV/INF/17/2 Draft 6) zur Annahme vorgeschlagen werde.

3. Der Aufbau dieses Dokuments ist wie folgt:

| | |
|---|-------------------------------------|
| ZUSAMMENFASSUNG..... | ERROR! BOOKMARK NOT DEFINED. |
| HINTERGRUND..... | ERROR! BOOKMARK NOT DEFINED. |
| ENTWICKLUNGEN AUF DER NEUNZEHNTE TAGUNG DER BMT | ERROR! BOOKMARK NOT DEFINED. |
| NÄCHSTE SCHRITTE..... | ERROR! BOOKMARK NOT DEFINED. |

ANLAGE:..... VORSCHLAG DER BMT FÜR EINE ÜBERARBEITUNG VON DOKUMENT UPOV/INF/17

4. In diesem Dokument werden folgende Abkürzungen verwendet:

- BMT: Arbeitsgruppe für biochemische und molekulare Verfahren und insbesondere für DNS-Profilierungsverfahren
- CAJ: Verwaltungs- und Rechtsausschuss
- TC: Technischer Ausschuss
- TWA: Technische Arbeitsgruppe für landwirtschaftliche Arten
- TWC: Technische Arbeitsgruppe für Automatisierung und Computerprogramme
- TWF: Technische Arbeitsgruppe für Obstarten
- TWO: Technische Arbeitsgruppe für Zierpflanzen und forstliche Baumarten
- TWP: Technische Arbeitsgruppe(n)
- TWV: Technische Arbeitsgruppe für Gemüsearten

HINTERGRUND

5. Der Hintergrund zu dieser Angelegenheit ist in Dokument TC/55/7 „Molekulare Verfahren“ dargelegt.
6. Der TC stimmte auf seiner fünfundfünfzigsten Tagung¹ dem Vorschlag der BMT auf ihrer achtzehnten Tagung² zu, dass die Europäische Union, Frankreich und die Niederlande einen neuen Entwurf des Dokuments UPOV/INF/17 (Dokument UPOV/INF/17/2 Draft 3) zur Prüfung auf der neunzehnten Tagung der BMT erstellen sollten (vergleiche Dokument TC/55/25 Corr. „Bericht“, Absatz 181).
7. Die TWP prüften auf ihren Tagungen im Jahr 2020 das Dokument TWP/4/7 „*Molecular Techniques*“. Die TWP nahmen die von der BMT vereinbarten Änderungen an Dokument UPOV/INF/17, wie in Dokument TWP/4/7, Anlage II, wiedergegeben, zur Kenntnis. Die TWP nahmen zur Kenntnis, dass der TC vereinbart habe, die Europäische Union, Frankreich und die Niederlande zu ersuchen, einen neuen Entwurf des Dokuments UPOV/INF/17 zur Prüfung durch die BMT auf ihrer neunzehnten Tagung zu erstellen (vergleiche Dokumente TWV/54/12 „*Report*“, Absätze 13 und 14; TWO/52/11 „*Report*“, Absätze 84 und 85; TWA/49/7 „*Report*“, Absätze 58 und 59; TWF/51/10 „*Report*“, Absätze 13 und 14; und TWC/38/11 „*Report*“, Absätze 66 und 67).
8. Das Programm für die Ausarbeitung von Anleitung und einschlägigem Informationsmaterial sieht die Vorlage eines neuen Entwurfs des Dokuments UPOV/INF/17 zur etwaigen Annahme durch den Rat auf seiner fünfundfünfzigsten Tagung im Jahr 2021 vor, wie in Dokument TC/56/14 „Erstellung von Anleitung und Informationsmaterial“ dargelegt.

ENTWICKLUNGEN AUF DER NEUNZEHNTE TAGUNG DER BMT

9. Die BMT prüfte auf ihrer neunzehnten Tagung³ die Dokumente BMT/19/3 Rev. „*Revision of document INF/17*“ und UPOV/INF/17/2 Draft 3 (vergleiche Dokument BMT/19/15 „*Report*“, Absätze 11 und 12).
10. Die BMT vereinbarte, dass der in der Anlage dieses Dokuments wiedergegebene Entwurf einer Anleitung dem Technischen Ausschuss als Grundlage für eine künftige Überarbeitung von Dokument UPOV/INF/17 vorgeschlagen werden sollte.

NÄCHSTE SCHRITTE

11. Der von der BMT vereinbarte Entwurf einer Anleitung, wie in der Anlage dieses Dokuments wiedergegeben, ist in Dokument UPOV/INF/17/2 Draft 4 unter Angabe der Änderungen gegenüber dem Wortlaut des Dokuments UPOV/INF/17/1 wiedergegeben. Der TC wird das Dokument UPOV/INF/17/2 Draft 4 möglicherweise als Grundlage für eine Überarbeitung des Dokuments UPOV/INF/17 prüfen wollen.
12. Der TC wird die TWP möglicherweise ersuchen wollen, auf ihren Tagungen im Jahr 2021 einen Entwurf einer Überarbeitung von Dokument UPOV/INF/17/1 (Dokument UPOV/INF/17/2 Draft 5) aufgrund der vom TC auf seiner sechsfundfünfzigsten Tagung vereinbarten Änderungen von UPOV/INF/17/2 Draft 4 zu prüfen.
13. Auf der Grundlage der Schlussfolgerungen des TC auf seiner sechsfundfünfzigsten Tagung und der TWP auf ihren Tagungen im Jahr 2021 würde dem Rat auf seiner fünfundfünfzigsten Tagung vom 29. Oktober 2021, vorbehaltlich der Zustimmung des TC auf seiner siebenundfünfzigsten Tagung und des CAJ auf seiner achtundsiebzigsten Tagung im Jahre 2021, ein Entwurf einer Überarbeitung von Dokument UPOV/INF/17 (UPOV/INF/17/2 Draft 6) zur Annahme vorgeschlagen.
14. Der TC wird ersucht,
 - a) die Überarbeitung von Dokument TGP/INF/17 auf der Grundlage von Dokument UPOV/INF/17/2 Draft 4 zu prüfen;

¹ abgehalten am 28. und 29. Oktober 2019 in Genf.

² Abgehalten in Hangzhou, China, vom 16. bis 18. Oktober 2019.

³ Veranstaltet von den Vereinigten Staaten von Amerika und auf elektronischem Wege vom 23. bis 25. September 2020 abgehalten.

b) die TWP zu ersuchen, auf ihren Tagungen im Jahr 2021 den Entwurf einer Überarbeitung von Dokument UPOV/INF/17/1 (Dokument UPOV/INF/17/2 Draft 5) zu prüfen; und

c) zur Kenntnis zu nehmen, dass dem Rat auf seiner fünfundfünfzigsten Tagung vom 29. Oktober 2021, vorbehaltlich der Zustimmung des TC auf seiner siebenundfünfzigsten Tagung und des CAJ auf seiner achtundsiebzigsten Tagung im Jahr 2021, ein Entwurf einer Überarbeitung von Dokument UPOV/INF/17 (UPOV/INF/17/2 Draft 6) zur Annahme vorgeschlagen werde.

[Anlage folgt]

VORSCHLAG DER BMT FÜR EINE ÜBERARBEITUNG VON DOKUMENT UPOV/INF/17

Inhalt

| | | |
|------|---|---|
| I. | EINFÜHRUNG | 1 |
| B. | ALLGEMEINE GRUNDSÄTZE | 1 |
| 1. | AUSWAHL MOLEKULARER MARKER | 2 |
| 1.1 | Sortenserien für das Auswahlverfahren | 2 |
| 1.2 | Molekulare Marker - Leistungskriterien | 2 |
| 2. | AUSWAHL DER DETEKTIONSMETHODE | 3 |
| 2.1 | DNS-Profilierungsverfahren – allgemeine Überlegungen | 3 |
| 2.2. | Zugang zur Technik | 3 |
| 3. | VALIDIERUNG UND HARMONISIERUNG VON MARKERSATZ UND DETEKTIONSMETHODE | 3 |
| 3.1 | Validierung und Harmonisierung - allgemeine Überlegungen | 3 |
| 3.2 | Leistungsaspekte – Validierung von Markern und Verfahren | 3 |
| 3.3 | Konsistenzaspekte | 4 |
| 4. | AUFBAU EINER ARTSPEZIFISCHEN DATENBANK | 4 |
| 4.1 | Empfehlungen für die Gestaltung der Datenbank | 4 |
| 4.2 | Anforderungen an das Pflanzenmaterial | 5 |
| 4.3 | Verarbeitung der Sequenzdaten | 5 |
| 4.4 | Ausprägungstypen | 6 |
| 4.5 | Datenbankmodell | 6 |
| 4.6 | Liste der Datenbankfelder | 7 |
| 4.7 | Datenzugriff / -eigentum | 7 |
| 5. | DATENAUSTAUSCH | 7 |
| 5.1 | Datenaustausch-Szenarien | 7 |
| 5.2 | Verfahren für den Datenaustausch | 7 |
| 6. | ZUSAMMENFASSUNG | 8 |
| C. | LISTE DER AKRONYME | 8 |

I. EINFÜHRUNG

Zweck dieses Dokuments (BMT-Richtlinien) ist es, Anleitung zu harmonisierten Grundsätzen für die Verwendung molekularer Marker zu geben, um qualitativ hochwertige molekulare Daten für eine Reihe von Anwendungen zu erzeugen. In diesem Dokument werden nur molekulare DNS-Marker berücksichtigt.

Die BMT-Richtlinien sollen ferner den Aufbau von Datenbanken mit molekularen Profilen von Pflanzensorten behandeln, die möglicherweise mit verschiedenen Techniken in verschiedenen Labors erzeugt werden. Ziel ist es zudem, hohe Anforderungen zu stellen an die Qualität der Marker und an das Bestreben, reproduzierbare Daten anhand dieser Marker zu erzeugen, wenn sich Ausrüstungen und/oder Reaktionschemikalien ändern. Spezifische Vorsichtsmaßnahmen sind zu treffen, um qualitativ hochwertige Eingaben in eine Datenbank sicherzustellen.

B. ALLGEMEINE GRUNDSÄTZE

Die DNS-Profilierung einer Pflanzensorte erfordert einen Satz molekularer Marker und eine Methode, diese festzustellen. Zwei verschiedene molekulare Markersätze, die mit derselben Methode festgestellt wurden, führen bei einer bestimmten Sorte zu zwei unterschiedlichen DNS-Profilen. Dagegen wird die Feststellung der spezifischen Allele eines gegebenen molekularen Markersatzes mit zwei verschiedenen Methoden voraussichtlich zu identischen DNS-Profilen führen. Eine Standardisierung der Detektionsmethode und -technologie ist nicht erforderlich, solange die Qualitätskriterien erfüllt werden und die gewonnenen DNS-Profile

konsistent sind. Die Technologie, die zur Feststellung gegebener Markersätze eingesetzt wird, sollte den Genotyp einer bestimmten Sorte nicht beeinflussen.

Bei den molekularen Markersätzen, den Methoden zur Markerfeststellung und dem anschließenden Aufbau der Datenbank lassen sich 5 verschiedene Phasen unterscheiden:

1. Auswahl molekularer Marker
2. Auswahl der Detektionsmethode
3. Validierung und Harmonisierung der Detektionsmethode
4. Aufbau der Datenbank
5. Datenaustausch

Diese verschiedenen Phasen werden in diesem Dokument eingehender beschrieben. Es wird davon ausgegangen, dass diese Phasen unabhängig sind vom Entwicklungsstand der Genotypisierungstechnologien und von künftigen Verbesserungen der Hochdurchsatz-Sequenzierung.

1. Auswahl molekularer Marker

1.1 Sortenserien für das Auswahlverfahren

Bei der DNS-Profilierung von Pflanzensorten und dem Aufbau von Datenbanken sollten die molekularen Marker zielorientiert ausgewählt werden. Für die Einleitung des Markerauswahlverfahrens wird eine geeignete Anzahl von Sorten benötigt (Entwicklungsserie), um die Vielfalt widerzuspiegeln, die innerhalb der Gruppe/Pflanze/Art/bzw. des Typs beobachtet wird, für die bzw. den die Marker unterscheidend sein sollen. Eine weitere Auswahl erfolgt mittels Profilierung zusätzlicher Sorten (Validierungsserie), um die Leistung der Marker zu messen. Als Kriterien für die Auswahl der Validierungsserie kommen infrage:

- a) genetisch sehr ähnliche Sorten oder Linien, NILs, RILs
- b) Elternlinien und Nachkommen
- c) genetisch nah verwandte, jedoch morphologisch unterschiedliche Arten (z. B. Mutanten)
- d) einige morphologisch ähnliche Sorten mit unterschiedlicher Zuchtformel
- e) verschiedene Partien derselben Sorte
- f) unterschiedliche Herkunft innerhalb derselben Sorte

1.2 Molekulare Marker - Leistungskriterien

Die folgenden allgemeinen Kriterien für die Auswahl eines bestimmten Markers oder einer Serie von Markern sollen für molekulare Marker ungeachtet der Verwendung der Marker geeignet sein:

- a) Laborintern Wiederholbarkeit und Solidität und laborübergreifend Reproduzierbarkeit bei der Auswertung der Daten;
- b) Mögliche Quellen für molekulare Marker
 - Molekulare Marker, die aus öffentlichen Quellen stammen
 - Molekulare Marker, die aus nicht-öffentlichen Quellen stammen oder durch Aussortieren und Auswählen von handelsüblichen artspezifischen Chips und Arrays gewonnen wurden.
 - Molekulare Marker, die aus neu generierten Sequenzdaten ausgewählt wurden;
- c) nach Möglichkeit Vermeiden von Markern mit „Nullallelen“ (d. h. ein Allel, das das Fehlen eines PCR-Produkts auf molekularer Ebene bewirkt), was ebenfalls nicht wesentlich, jedoch ratsam ist;
- d) Zulassen einer Auswertung von Markerprofilen, die problemlos, objektiv und eindeutig ist. Diese leistungsstarken Marker werden gegenüber komplexen Markerprofilen, die zur Mehrdeutigkeit neigen, bevorzugt. Klare Schwarz-Weiß-Antworten erleichtern zudem die Harmonisierung;
- e) Kodominante Marker werden dominanten Markern gegenüber grundsätzlich bevorzugt, weil sie ein besseres Unterscheidungsvermögen besitzen;
- f) Marker können in kodierenden und/oder nichtkodierenden Regionen lokalisiert sein; und
- g) Die Verwendung molekularer Marker ist artspezifisch und sollte die Besonderheiten der Vermehrung der Art berücksichtigen.

Es wird eingeräumt, dass spezifische Anwendungen bestimmte zusätzliche Überlegungen erfordern könnten (aber nicht darauf beschränkt sind):

- a) Die Anzahl der Marker sollte im Verhältnis stehen zur jeweils geforderten Genauigkeit des Genotyps. Die Anzahl der zum Erreichen der erforderlichen Auflösung oder Unterscheidungskraft einzusetzenden Marker ist abhängig vom Markertyp (dominant/kodominant, bi-/multiallelisch), der Art und der Qualität der Markerleistung;
- b) Die Abdeckung des Genoms und des Verknüpfungsungleichgewichts sollte die Ziele widerspiegeln. Die Kenntnis der physischen und/oder genetischen Position der ausgewählten Marker auf dem Genom ist nicht unbedingt erforderlich, ermöglicht aber eine gute Auswahl der Marker.

2. Auswahl der Detektionsmethode

2.1 DNS-Profilierungsverfahren – allgemeine Überlegungen

2.1.1 Wichtige Überlegungen zur Wahl eines DNS-Profilierungsverfahrens zur Gewinnung hochwertiger molekularer Daten sind:

- a) Reproduzierbarkeit der Datengenerierung in den Laboren und Detektionsplattformen sowie zwischen diesen (unterschiedliche Arten von Ausrüstung);
- b) Wiederholbarkeit im Zeitablauf;
- c) Unterscheidungskraft des Verfahrens;
- d) Zeit- und Arbeitsintensität des Verfahrens;
- e) Belastbarkeit hinsichtlich der zeitlichen Gegebenheiten und der Bedingungen (Empfindlichkeit gegenüber subtilen Änderungen von Ablauf oder Bedingungen);
- f) Flexibilität des Verfahrens; Möglichkeit, die Anzahl der Proben und/oder der Marker zu variieren;
- g) Auswertung der Daten ist von der Ausrüstung unabhängig;
- h) Nachhaltigkeit der Datenbanken;
- i) Zugänglichkeit der Methodik;
- j) Unabhängigkeit von besonderen Maschinen, Chemikalien, Zulieferern, Partnern oder Produkten;
- k) zur Automatisierung geeignet;
- l) für Multiplexing geeignet; und
- m) kostengünstig; Kosten, Zahl der Proben und Zahl der Marker stehen zueinander im Verhältnis.

2.2. Zugang zur Technik

Einzelne molekulare Marker und Materialien sind öffentlich verfügbar. Es dürften jedoch hohe Investitionen erforderlich sein, um hochqualitative 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. Es ist zu empfehlen, dass Angelegenheiten bezüglich der Rechte des geistigen Eigentums zu Beginn jeder Entwicklungsarbeit behandelt werden.

3. Validierung und Harmonisierung von Markersatz und Detektionsmethode

3.1 Validierung und Harmonisierung - allgemeine Überlegungen

Molekulare Marker und Detektionsverfahren müssen solide sein und zu konsistenten DNS-Profilen führen. Die Leistung der molekularen Marker und der Verfahren zur Genotypisierung wird im Rahmen des Validierungsprozesses evaluiert. Bei gemeinsamen Datenbanken wird die Übereinstimmung der DNS-Profile im Rahmen des Harmonisierungsprozesses in verschiedenen Laboren evaluiert, wobei unterschiedliche Ausrüstungen und Chemikalien verwendet werden. Die Verwendung validierter Marker und Verfahren wird zu harmonisierten Ergebnissen führen.

3.2 Leistungsaspekte – Validierung von Markern und Verfahren

Der ausgewählte Markersatz muss geeignet sein (Zweckmäßigkeit). Die Genauigkeit muss gemessen werden. Um die Eignung eines Verfahrens oder DNS-Markersatzes festzustellen, sind diverse Punkte zu berücksichtigen:

- a) Unterscheidungsvermögen/Informativität;

- b) Wiederholbarkeit; wenn identische Prüfungsergebnisse anhand desselben Verfahrens, an identischen Testgegenständen, im gleichen Labor, durch denselben Prüfer, unter Verwendung der gleichen Ausrüstung innerhalb kurzer Zeitintervalle erzielt werden.
- c) Reproduzierbarkeit; wenn Testergebnisse anhand desselben Verfahrens, an identischen Testgegenständen, im gleichen Labor oder zwischen verschiedenen Laboren, durch verschiedene Prüfer, unter Verwendung unterschiedlicher Ausrüstungen erzielt werden.
- d) Belastbarkeit; ein Maß für seine Fähigkeit, von kleinen, aber absichtlichen Abweichungen von den in den Verfahrensparametern beschriebenen Versuchsbedingungen unbeeinflusst zu bleiben, und das auch einen Hinweis auf seine Zuverlässigkeit bei normaler Verwendung liefert; und
- e) Fehlerrate.

Definitionen von Leistungsmerkmalen basieren auf: ISO 16 577:2016

3.3 Konsistenzaspekte

Zur Erzielung einer Konsistenz von Ergebnissen, ist der Prozess der Harmonisierung von Markern und Verfahren zwischen verschiedenen Laboren bei gemeinsamer Datenbank (Ringtest) in Betracht zu ziehen:

- a) Zur Prüfung der laborübergreifenden Konsistenz sollen in allen Laboren vorgegebene Referenz-Sortensammlungen, die ein breites Spektrum von Allelen abdecken, verwendet werden
- b) Aufnahme von Duplikaten, Unterproben und individuellen Exemplaren einer Art zur Prüfung der Konsistenz der DNS-Profile und zur Einschätzung der Fehlerquote zwischen den Laboren
- c) Vereinbarungen zur Auswertung molekularer Daten. Die Notwendigkeit der Erstellung eines Protokolls für die Allel-/Bandenauswertung zwischen den Laboren ist abhängig vom verwendeten Markertyp (z. B. wesentlich bei SSR-Markern). Das Protokoll könnte sich mit der Auswertung folgender Daten befassen:
 - i. seltene Allele (d. h. diejenigen an einem spezifischen Locus, die mit einer Häufigkeit unter einem vereinbarten Schwellenwert (in der Regel 5-10 %) in einer Population) auftreten);
 - ii. Nullallele (ein Allel, das das Fehlen eines PCR-Produkts auf Molekularebene bewirkt);
 - iii. „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);
 - iv. fehlende Daten (d. h. Loci, für die aus welchem Grund auch immer für eine oder mehrere Sorten keine Daten erfasst wurden); und
 - v. monomorphe Banden oder nicht-informative Allel-Scorewerte (diejenigen Allele/Banden, die bei jeder analysierten Sorte auftreten, d. h. in einer bestimmten Sortensammlung nicht polymorph sind).

4. Aufbau einer artspezifischen Datenbank

Die in einer Datenbank gespeicherten Daten sowie die Art und Weise der Speicherung sollten das datenproduzierende Verfahren widerspiegeln. Daher sollte der Aufbau einer Datenbank verschiedene Stufen der Datenproduktion berücksichtigen (d. h. Rohdaten, Sequenzdaten...). In der Datenbank sollten die Endergebnisse, d. h. das DNS-Profil ebenso wie die Art seiner Gewinnung im Sinne der Verfahrensbeschreibung des Labors und der Rechenschritte zur Gewinnung eines DNS-Profils gespeichert sein.

4.1 Empfehlungen für die Gestaltung der Datenbank

Bei der Gestaltung der Datenbank sollten folgende Aspekte berücksichtigt werden:

- a) Die Architektur der Datenbank sollte flexibel sein und z. B. sowohl flache Dateien als auch komprimierte Archivformate speichern können.
- b) Für die experimentelle Laborarbeit, die Datenerzeugung und die Allel-Scorewerte sind jeweils eigenständige Tabellen und Einträge erforderlich.

c) Speicherung von Daten auf verschiedenen Stufen, z. B. Allel-Scorewerte und alle Auswertungs-Regeln, die einer Entscheidung zugrunde liegen und Verknüpfungen zu den Rohdaten (tiff-Dateien, bam-Dateien), die generiert wurden.

d) Dateien zum Variantenaufwurf im VCF- oder BCF-Format (entsprechend der Standardversion 4.2 oder höher). Die Header-Einträge sollten Namen und Version der verschiedenen Scripte enthalten, die für Kartierung und Filterung der Sequenzabschnitte sowie für Aufruf und Filterung der Varianten verwendet werden, und zwar dergestalt, dass der Bioinformatiker die Analyse wiederholen kann.

e) Bei Wiederholungsproben, bei denen das DNS-Profil nicht übereinstimmt, muss der Eintrag gekennzeichnet oder gegebenenfalls herausgefiltert werden. Die in solchen Fällen angewandten Regeln sind in einem öffentlich zugänglichen Code Repository zu dokumentieren, das Verweise auf die Variantenaufwurf-Datei enthält. Die Häufigkeiten könnten auch für heterogene Sorten verwendet werden.

f) Validierung der VCF- und/oder BCF-Daten mittels einschlägiger Spezifikationen.

g) Leicht austauschbare Daten (z. B. API).

4.2 Anforderungen an das Pflanzenmaterial

Quelle und Art des Materials und die Anzahl der zu analysierenden Proben, die in der Datenbank gespeichert werden sollen, sollten berücksichtigt werden.

4.2.1 Quelle des Pflanzenmaterials

Das zu analysierende Pflanzenmaterial sollte eine authentische, repräsentative Probe der Sorte sein und, wenn möglich, aus dem Muster der für die Prüfung im Hinblick auf die Erteilung von Züchterrechten oder auf die amtliche Eintragung verwendeten Sorte stammen. Die Verwendung von Mustern erfordert gegebenenfalls die Genehmigung der zuständigen Behörde, des Züchters und/oder des Erhaltungszüchters. Das Pflanzenmaterial, dem die Proben entnommen werden, sollte rückverfolgbar sein, falls sich einige der Pflanzen im Nachhinein als nicht repräsentativ für die Sorte erweisen.

4.2.2 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 werden weitgehend von der betreffenden Pflanze oder Art abhängen. Bei samenvermehrten Sorten beispielsweise kann der Samen 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, es sollte das Verfahren für die Probenentnahme und die DNS-Extraktion genormt und dokumentiert werden. Zudem sollte überprüft werden, dass die Verfahren für die Probeentnahme und die Extraktion bei der DNS-Analyse übereinstimmende Ergebnisse zeigen.

4.2.3 Probengröße und -art (Massen- oder Einzelproben)

Es ist wesentlich, dass die für die Analyse entnommenen Proben für die Sorte repräsentativ sind. Die Besonderheiten der Vermehrung sollten beachtet werden (vergleiche Allgemeine Einführung).

4.2.4 DNS-Referenzprobe

Eine DNS-Referenzprobensammlung kann aus dem Pflanzenmaterial angelegt werden, dem Proben entnommen werden. Das Verfahren zur Probeentnahme sollte der empfohlenen Vorgehensweise folgen und es sollten Qualitätskriterien für die DNS-Extraktion festgelegt werden. Beides muss dokumentiert werden.

Die DNS-Proben sollten so gelagert werden, dass ein Zerfall verhindert wird (z. B. durch Lagerung bei -80°C). Der Transport von DNS-Referenzproben ist in Dokument TGP/5: Abschnitt 1. beschrieben.

4.3 Verarbeitung der Sequenzdaten

Ein ausführliches Protokoll der Datenverarbeitungspipeline kann Folgendes beinhalten:

- a) Art und Version der Tools;
- b) die für das Tool verwendete Befehlszeile einschließlich Schwellenwerten;

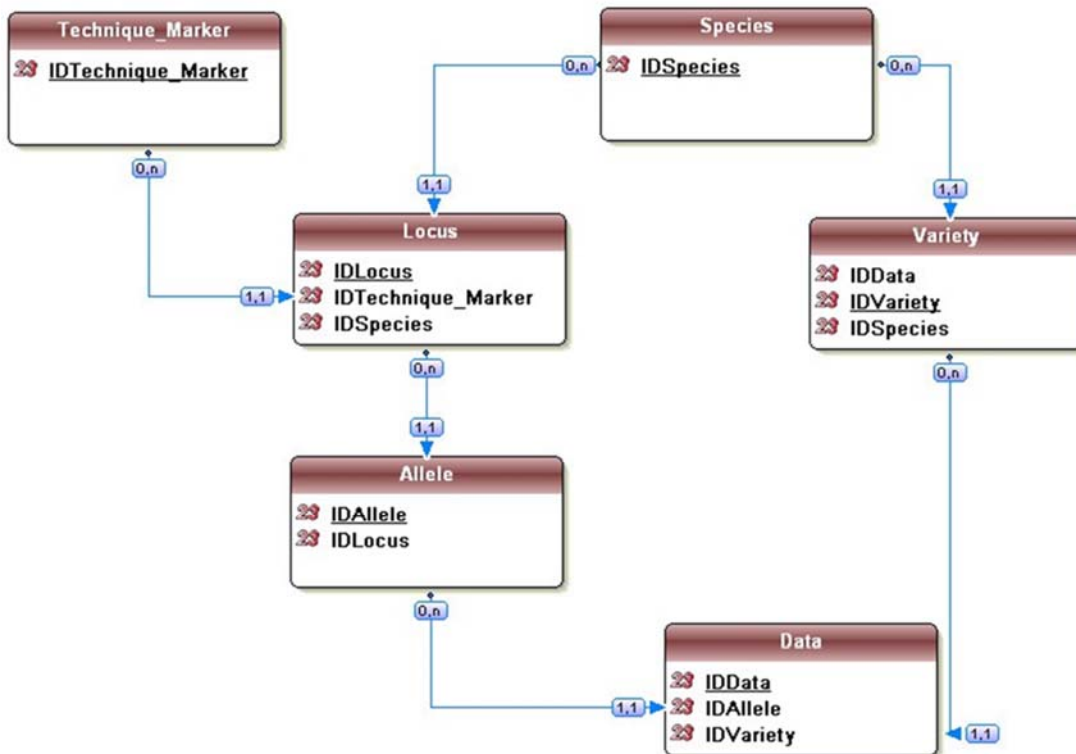
- c) Reproduzierbarkeitszählwerte;
- d) Möglichkeit, die Daten weiterzugeben und gemeinsam zu verarbeiten;
- e) Abgleich-Rohdaten (BAM oder CRAM-Dateien) sollten nach Möglichkeit gespeichert werden;
- f) mehrprobenbasierte VCF-Dateien sind nicht geeignet. Es muss eine VCF-Datei pro Sorte vorhanden sein;
- g) wenn VCF-Dateien gespeichert werden, sollten alle Positionen (sowohl Varianten als auch Nicht-Varianten) sowie deren Tiefe gespeichert werden;
- h) sowohl der heuristische als auch der probabilistische Ansatz sollten erwogen und im Hinblick auf Detektionsmethoden verglichen werden;
- i) die Datenbanken sollten die Ein- und Ausgabe von Variantenaufrufdaten in standardisiertem Format (VCF oder BCF) unterstützen;
- j) die Datenverarbeitungspipeline sollte eine ausführliche Protokolldatei generieren, die zusammen mit den Variantenaufrufdaten zu speichern ist;
- k) nach Möglichkeit sollten Rohdaten gespeichert werden, so dass die Datenverarbeitung mit neuen oder aktualisierten Tools wiederholt werden kann; und
- l) es sollte in Bezug auf ein gegebenes Allel ein p-Wert oder ein Unsicherheitshinweis gespeichert werden.

4.4 Ausprägungstypen

Molekulare Daten können auf zahlreiche Arten gespeichert werden. Deshalb ist es wichtig, dass eine Datenbankstruktur entwickelt wird, die mit allen beabsichtigten Verwendungen der Daten kompatibel ist.

4.5 Datenbankmodell

Das Datenbankmodell sollte von IT-Datenbankexperten zusammen mit den Nutzern der Datenbank festgelegt werden. Das Datenbankmodell sollte mindestens sechs Kernobjekte enthalten: Art, Sorte, Marker-Detektionsmethode, Marker, Locus und Allel. Bei mittels Sequenzierungsdaten gewonnenen Varianten können die VCF-Dateien in einer relationalen oder einer Nicht-SQL-Datenbank gespeichert werden. In diesem Fall gibt jeder Datenbanksatz in Bezug auf eine Variante eine definierte Genomversion sowie Chromosom, Position und Referenzallel an.



4.6 *Liste der Datenbankfelder*

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

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

b) **Position im Referenzgenom oder Locuscode:** Sofern für die betreffende Art ein Referenzgenom verfügbar ist, sollten vorzugsweise eine Genomaufbauversion, ein Chromosom und eine Position angegeben werden, z. B. SL2.50ch05:63309763 für Tomate *Solanum lycopersicum*, Aufbauversion 2.50 auf Chromosom 5, Position 63309763. Falls kein Referenzgenom verfügbar oder der Standort unbekannt ist, kann für die betreffende Art ein Name oder Locuscode angegeben werden, z. B. gwm 149, A2 usw.

c) **Genotyp:** Für SNP-Profile sollte die Allelzusammensetzung von SNP oder MNP angegeben werden, z. B. A/T oder A/A. Für andere Verfahren gibt der Genotyp den Namen oder den Code des Allels eines gegebenen Locus für die betreffende Art an, z. B. 1, 123 usw.

d) **Alleltiefen / Datenwert:** Für SNP, die aus den Sequenzierungsdaten der nächsten Generation gewonnen werden, sollte die Tiefe der Verteilung für Allele angegeben werden, z. B. 10/20 für ein A/T-Allel, bei dem A durch 10 Lesungen und T durch 20 abgedeckt ist. Andernfalls sollte ein Datenwert für eine gegebene Probe auf einem gegebenen Locus-Allel angegeben werden, z. B. 0 (Fehlen), 1 (Vorhandensein), 0.25 (Häufigkeit) usw.

e) **Sorte:** Sortenbezeichnung oder Angabe des Züchters: Die Sorte ist das Ziel, für das die Daten gewonnen wurden.

f) **Art der Sorte:** z. B. Inzuchtlinie oder Hybrid

g) **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.). Um Probleme zu vermeiden, wird die Verwendung des UPOV-Codes empfohlen.

4.6.2 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.

4.7 *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.

5. Datenaustausch

5.1 *Datenaustausch-Szenarien*

Zu Zwecken der Zusammenarbeit sollte das Datenmodell verschiedene Arten von Szenarien ermöglichen, einschließlich des Austauschs von Daten, die aus einem standardisierten Satz von Markern für eine bestimmte Pflanze erzeugt wurden (Szenario 1), und der Suche und Ansicht von Daten ausgewählter Sorten, die aus derselben standardisierten Serie von Markern erzeugt wurden (Szenario 2). Technische Einzelheiten zu beiden Szenarien sind in der Anlage beschrieben: Datenaustauschszenarien und Datenübertragungsverfahren.

5.2 *Verfahren für den Datenaustausch*

5.2.1 Die Übermittlung von Fingerabdruckdaten kann eine Reihe von Informationen enthalten, z. B. Loci, Proben, DNS, Fingerabdruckdaten und Fingerabdruckprofile. Das Verfahren der Datenübertragung muss je nach dem zu übertragenden Inhalt festgelegt werden, wobei Folgendes berücksichtigt werden sollte:

- a) Datenmenge
- b) Komplexität der Daten
- c) Anforderungen an Abfrage- oder Suchfunktionen

Technische Einzelheiten zu den Datenübertragungsverfahren sind in der Anlage beschrieben: Datenaustauschszenarien und Datenübertragungsverfahren.

5.2.2 Zu den üblicherweise verwendeten Datenformaten gehören: zip, csv, json und xml. Ihre jeweiligen Merkmale sind wie folgt:

1) Das Zip-Format ermöglicht eine Vielzahl von Dateninformationsdateien im Originalformat und eignet sich aufgrund seines großen Datenkompressionsverhältnisses und der einfachen Übertragung für große und komplexe Daten.

2) Das csv-Format eignet sich besser für Dateninformationen in einem einfachen Datenformat, das den Vorteil hat, dass es weniger ungültige Daten enthält und eine höhere Verarbeitungsgeschwindigkeit aufweist.

3) Die Formate json und xml können Dateninformationen betreffend komplexere Zeichen und redundantere Informationen enthalten, aber beide bieten eine gute Lesbarkeit.

6. Zusammenfassung

Nachstehend ist eine Zusammenfassung des Vorgehens wiedergegeben, das für hochwertige DNS-Profilierungsverfahren einschließlich der Auswahl und Verwendung molekularer Marker sowie zum Aufbau gemeinsamer und nachhaltiger molekularer Datenbanken für DNS-Profile von Sorten empfohlen wird (d. h. Datenbanken, die künftig aus einer Reihe von Quellen, unabhängig von der angewandten Technik, bestückt werden können).

- a) Prüfung des Vorgehens nach Pflanzenart;
- 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;
- f) Überprüfung der Quelle des verwendeten Pflanzenmaterials;
- 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;
- h) Durchführung einer Evaluierung;
- i) Erstellung eines Protokolls für die Auswertung der molekularen Daten;
- 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 Referenzen bei allen Analysen (Sorten, DNS-Proben und gegebenenfalls Allele);
- 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.

C. LISTE DER AKRONYME

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

ANHANG DER ANLAGE

DATENAUSTAUSCHSZENARIEN UND ÜBERTRAGUNGSVERFAHREN

A: Szenarien für den Datenaustausch

Szenario 1: Austausch von Daten, die aus einem standardisierten Markersatz für eine bestimmte Pflanze erzeugt wurden

Zum Austausch von Daten über den Markersatz, der für eine bestimmte Pflanze verwendet wird, kann der folgende Webdienst genutzt werden:

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

Um z. B. Informationen über Markerserien für Mais unter Verwendung des SSR- und CE-Verfahrens zu erhalten, sollte die folgende URL aufgerufen werden:

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

Das Ergebnis wäre:

```

{"techniqueid":
"CN_SSR_ZEAA_MAY_CE_V
_1",
"description": "Laboratory
method description"
["locusid": "M01",
"alleles":
["alleleid": "238/256",
"examplevariety":
],
["alleleid": "238/271",
"examplevariety":
],
["alleleid": "246/246",
"examplevariety":
],
["alleleid": "246/248",
"examplevariety":
],
["alleleid": "246/250",
"examplevariety":
],
["alleleid": "246/254",
"examplevariety":
],
["alleleid": "246/256",
"examplevariety":
],
["alleleid": "246/260",
"examplevariety":
],
["alleleid": "246/277",
"examplevariety":
],
["alleleid": "246/284",
"examplevariety":
],
["alleleid": "246/288",
"examplevariety":
],
["alleleid": "248/250",
"examplevariety":
],
["alleleid": "248/256",
"examplevariety":
],
],
["alleleid": "248/271",
"examplevariety":
],
["alleleid": "248/290",
"examplevariety":
],
],
["alleleid": "250/250",
"examplevariety":
],
["alleleid": "250/252",
"examplevariety":
],
["alleleid": "250/256",
"examplevariety":
],
["alleleid": "250/275",
"examplevariety":
],
["alleleid": "252/256",
"examplevariety":
],
["alleleid": "252/260",
"examplevariety":
],
["alleleid": "252/271",
"examplevariety":
],
["alleleid": "252/273",
"examplevariety":
],
["alleleid": "252/282",
"examplevariety":
],
["alleleid": "254/254",
"examplevariety":
],
["alleleid": "254/271",
"examplevariety":
],
],
["alleleid": "254/284",
"examplevariety":
],
["alleleid": "254/286",
"examplevariety":
],
],
["alleleid": "256/256",
"examplevariety":
],
["alleleid": "256/264",
"examplevariety":
],
["alleleid": "256/266",
"examplevariety":
],
["alleleid": "256/271",
"examplevariety":
],
["alleleid": "256/284",
"examplevariety":
],
["alleleid": "256/286",
"examplevariety":
],
["alleleid": "258/258",
"examplevariety":
],
["alleleid": "264/284",
"examplevariety":
],
["alleleid": "271/292",
"examplevariety":
]
["locusid"="M02".
"alleles": [...]
]} vi

```

Szenario 2: Suche und Ansicht von Daten ausgewählter Sorten, die aus demselben standardisierten Markersatz erzeugt wurden

Um molekulare Daten einer Sorte zu suchen und anzuzeigen, kann folgender Webdienst genutzt werden:
https://office.org/variety?id={irn}&techniqueid={technique_code} vi

Zum Beispiel:

https://office.org/variety?id=XU_30201800000140 &techniqueid= CN_SSR_ZEAA_MAY_CE_V_1 vi

Das Ergebnis wäre:

```
{ "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: Verfahren zur Datenübertragung

Im Folgenden wird ein Beispiel für die Erstellung eines Fingerabdruckpakets im Zip-Format für die Datenübertragung gegeben. Bei diesem Verfahren müssen zunächst unabhängige IDs zur Identifizierung von Proben, DNS, Fingerabdruckdaten und die Liste der Datenbankfelder des Fingerabdrucks verwendet werden. Danach enthält die Datendatei im json-Format alle Loci-, Proben- und DNS-Informationen. Alle Fingerabdruck-Daten werden unabhängig voneinander in einer eigenen Datei im json-Format gespeichert. Die Fingerabdruck-ID wird an den entsprechenden Locus der Fingerabdruckdaten gebunden sein, und alle Fingerabdruckdatendateien und Fingerabdruckspektrumsdateien werden unabhängig voneinander im entsprechenden Verzeichnis gespeichert. Die Formatstruktur des Fingerabdruckdatenpakets sieht also wie folgt aus:

```
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
```

Das Fingerabdruckpaket im Zip-Format kann um weitere Informationen erweitert werden. Herzstück des Pakets ist die Datei mit den Fingerabdruckdaten, die den Kern der Korrelation bildet, so dass die Korrelation

zwischen Teilen korrekt geparkt werden kann, was die Datenübertragung über verschiedene Systeme hinweg ermöglicht.

[Ende des Anhangs der Anlage und des Dokuments]