

Pflanzen

Identifikation von Obstsorten: Validierung einer Analysemethode

Jürg E. Frey¹, Bernhard Koller², Beatrice Frey¹ und Markus Bünter¹

¹Forschungsanstalt Agroscope Changins-Wädenswil ACW, CH-8820 Wädenswil

²ecogenics GmbH, Wagistrasse 23, CH-8952 Zürich-Schlieren

Auskünfte: Jürg E. Frey, E-Mail: juerg.frey@acw.admin.ch, Tel. +41 44 783 63 32

Zusammenfassung

Die Genotypisierung zur Identifikation von Sorten unter Einsatz von Mikrosatelliten ist heute bei Kulturpflanzen weit verbreitet und hat sich sehr gut bewährt. Dabei kommen viele verschiedene Methoden zum Einsatz, die sich in Bezug auf Arbeitsaufwand und Effizienz teilweise stark unterscheiden. Wir haben eine robuste und effiziente Methode entwickelt, die wir seit mehreren Jahren erfolgreich einsetzen. In diesem Projekt ging es darum, wie gut diese Methode in einem externen Labor einsetzbar ist und wie sich die Resultate von denjenigen unseres Labors unterscheiden. Wir konnten zeigen, dass die Methode ohne grossen Aufwand in externen Labors eingeführt werden kann und Resultate mit hoch signifikanter Korrelation ($R^2 = 0,9993$) zwischen den eingesetzten Analyse-Systemen liefert. Wie erwartet wurden kleine Abweichungen in den Lauflängen einzelner Genfragmente (Allele) beobachtet. Diese Abweichungen sind allerdings nur dann ein Problem, wenn bei der Analyse keine Standardsorte für den direkten Vergleich mit der Testsorte bekannt ist. Wir haben ein einfaches Verfahren entwickelt, um dieses Problem lösen zu können. Bei diesem Verfahren wird die Gesamtabweichung aller Allele von denjenigen von bekannten Sorten verglichen. Sofern die Standardsorte in der Vergleichsdatenbank vorhanden ist, wird sie die geringste Abweichung von der Testsorte aufweisen. In einem direkten Vergleich wird dann die Testsorte mit der so identifizierten Standardsorte analysiert. Damit ermöglicht dieses Verfahren eine eindeutige Identifikation auch bei grösseren systematischen Abweichungen zwischen unterschiedlichen Analysesystemen.

Die Genotypisierung von Pflanzen ist eine wichtige Grundlage für die modernen Methoden der markergestützten Selektion in der Züchtung, die Beschreibung genetischer Ressourcen und die präzise Sortenidentifikation. In den 90er Jahren wurden verschiedene molekulare Marker auf ihre Eignung für die Genotypisierung von Kulturpflanzen untersucht (Morgante und Olivieri 1993; Wang *et al.* 1994; Milbourne *et al.* 1998). Dabei zeigte sich, dass Mikrosatelliten-Marker für diese Aufgabe ausgezeichnet geeignet sind und seit über zehn Jahren sehr erfolgreich bei vielen Kulturpflanzen eingesetzt werden (z.B. McCouch *et al.* 1997). Mikrosatelliten oder einfache repetierte Sequenzen (simple sequence repeats, SSR) sind in den Genomen von Tieren und Pflanzen weit verbreitete Struktu-

ren (Tautz and Ranz 1984; Moran 1993; Estoup *et al.* 1993; Morgante und Olivieri 1993; Wang *et al.* 1994). Beispiele für Mikrosatelliten sind CACA-CACA (= (CA)₄) oder TGATGATGATGA (= (TGA)₅), also einfache Repetitionen eines DNA-Grundmotivs. Die Anzahl Repetitionen bestimmt die Länge eines Allels. Eine der wichtigsten Eigenschaften von Mikrosatelliten ist ihr hoher Grad an allelischer Diversität. Das heisst, generell existieren für jeden Mikrosatelliten viele verschiedene Längenvarianten in der Form einer unterschiedlichen Anzahl von Repetitionen. Zum Beispiel kann ein einzelner Mikrosatellit mit der Repetitionseinheit CA mehrere Dutzend unterschiedlicher Allele aufweisen, wie (CA)₅, (CA)₇, (CA)₁₅, etc. Diese genetische Diversität

ist die Basis für die Fähigkeit der Mikrosatelliten, auch nahe verwandte Gruppen wie Sorten von Kulturpflanzen differenzieren zu können. Für eine präzise Sortenidentifikation müssen die Daten mehrerer Mikrosatelliten kombiniert werden.

Da eine Analyse mit Mikrosatelliten mehrere zeitaufwändige und teure Arbeitsschritte erfordert, konnten trotz ihrer Vorteile bei der Identifikation entsprechende Methoden lange nicht auf breiter Basis eingesetzt werden. An der Forschungsanstalt Agroscope Changins-Wädenswil ACW haben wir daher in den letzten Jahren eine neue Methode zur raschen und ökonomischen Genotypisierung von Pflanzen entwickelt (Frey *et al.* 2004). Diese Methode wurde erfolgreich in der markergestützten Selektion von Sämlingen in der Apfelzüchtung eingesetzt. Ebenso erfolgreich war ihr Einsatz bei der Genotypisierung des Obst-Nuklearstocks, wo gezeigt werden konnte, dass sie im Prinzip auf alle Rosaceen-Obstarten angewendet werden kann. Wie auf der Basis von Literatur- und Erfahrungsdaten zu erwarten war, stellte sich heraus, dass für verschiedene Obstarten unterschiedliche Primer zu verwenden sind. In den letzten beiden Jahren wurden an der ACW mit dieser neuen Methode 410 verschiedene Sorten (Apfel, Birnen, Quitten, Kirschen, Zwetschgen) mit 12 Mikrosatelliten-Markern analysiert. Die Analysen wurden von verschiedenen Personen durchgeführt, was keinerlei Einfluss auf die Ergebnisse hatte.

Bei der Entwicklung analytischer molekularbiologischer Methoden stellt sich generell die Frage nach der Wiederholbarkeit zwischen verschiedenen Laboratorien, denn eine Vielzahl von Faktoren kann zu Abweichungen in den Resultaten führen. Doch genau diese Wiederholbarkeit ist bei Sortenidentifikations-Projekten wichtig, da ansonsten bei geänderten Bedingungen (z.B. das ausführende Labor) kein Vergleich zu bereits erstellten Daten gemacht werden kann. Obwohl die neue Methode in den Labors der ACW mit grossem Erfolg eingesetzt wird, wurde sie bisher noch nie in externen Labors eingesetzt.

Das Ziel dieses Projektes war es, zu zeigen, dass die an der ACW entwickelte Methode für die Genotypisierung von Apfelsorten so robust ist, dass ihre Anwendung in einem externen Labor leicht zu wiederholen ist und identische Ergebnisse liefert. Für diese Untersuchung wurden Blattproben von Apfel- und Kirscharten entnommen und parallel an der ACW und in den Labors der Firma ecogenics analysiert.

Untersuchung von Apfel- und Kirscharten

Die Proben stammten von je drei Apfel- (Golden, Gala und Braeburn) und Kirscharten (Star, Regina und Kordia). Dabei wurden von jeder Sorte Proben von vier Bäumen genommen. Die Probenahme wurde am 12. Juni 2006 durchgeführt. Für die Probenahme wurde jeweils das vierte Blatt hinter der Triebspitze verwendet. Die Proben wurden anonymisiert an die beiden Labors ausgeliefert, wobei folgende Codes verwendet wurden: Apfel Sorte 1: ts10-ts13; Sorte 2: ts21-23; Sorte 3: ts31-34; Kirscharten Sorte 1: ms51-54; Sorte 2: ms61-64; Sorte 3: ms71-74. Die Blattproben wurden in Papiersäckchen gelagert. Alles Material wurde bis zur Weiterverar-

beitung tiefgefroren bei -20 °C aufbewahrt.

Molekularbiologische Analysen

Die molekularbiologischen Analysen wurden sowohl von der ecogenics GmbH als auch vom Labor für molekulare Diagnostik und Epidemiologie der ACW durchgeführt. Die Blattproben wurden gemäss der von Frey *et al.* (2004) entwickelten Methode aufbereitet und mit 12 Mikrosatelliten pro Probe untersucht. Total ergaben sich so 72 zu analysierende Proben (24 Bäume zu jeweils 3 Multiplex-Reaktionen à 4 Loci).

Die von der ACW an ecogenics gelieferten Informationen über die Methode entsprachen mit einer Ausnahme denjenigen, die in der Arbeit von Frey *et al.* (2004) bereits publiziert worden waren. Die zusätzliche Information betraf die Menge eines der Primerpaare, für welche die doppelte Menge in den Multiplex-Mix empfohlen wurde. Von ecogenics wurde für jede Probe ein Experiment durchgeführt, hinzu kamen entsprechende Extraktions- und PCR-Kontrollen.

Vergleich der Analyse-Systeme

Die beiden in diesem Projekt eingesetzten Analyse-Systeme sind vom gleichen Anbieter (Applied Biosystems; ACW mit 3130xl; ecogenics mit 3130) und verwenden dasselbe Polymer (POP7). Ein wichtiger Unterschied bestand allerdings in der Länge der eingesetzten Kapillaren, die bei ecogenics 22 cm und bei ACW 36 cm betrug. Die Daten beider Systeme wurden einer Korrelationsanalyse unterzogen. Für eine detailliertere Analyse der Unterschiede in Laufdistanzen einzelner Genfragmente wurden neben den oben erwähnten Daten auch solche aus früheren Versuchen der ACW mit dem alten Gerät verwendet (3130, Applied Bio-

systems, Kapillarlänge 36cm, Polymer POP6).

Resultate der ecogenics GmbH

Bereits im ersten Ansatz gelang sowohl die Extraktion der DNA wie auch deren Analyse mittels Multiplex-PCR und Auftrennung auf einem Kapillar-Sequenziergerät. Einzig die Auftrennungsbedingungen mussten bei gewissen Markern leicht verändert werden, um Signale geeigneter Stärke zu erhalten. Solche Anpassungen waren aber zu erwarten und stellen im Hinblick auf die Übertragbarkeit der Methode auf andere Laboratorien kein Problem dar.

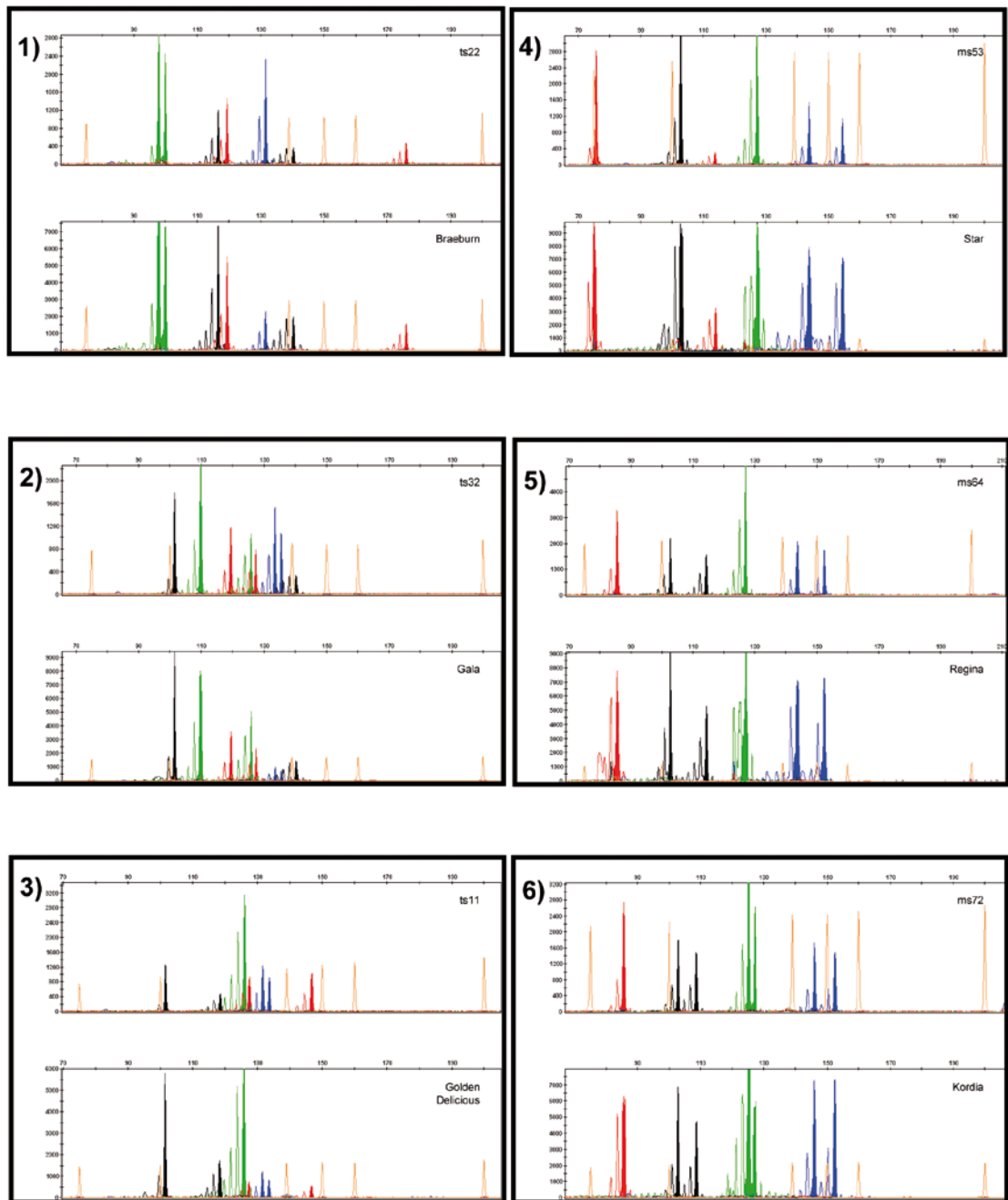
Resultate und Auswertungen der ACW Sortenidentifikation

Die Daten zu allen von beiden Labors untersuchten Proben stimmten vollständig überein. An der ACW wurden dann die Fragmentlängen der beobachteten Genotypen mit denjenigen der Nuklearstock-Datenbank der ACW verglichen, wobei ein eigenes entwickeltes Verfahren eingesetzt wurde (s. unten). Dieser Vergleich ergab eine sehr gute Übereinstimmung mit den Genotypen der Apfelsorten Golden, Gala, und Braeburn und der Kirscharten Star, Regina und Kordia (Abb. 1). Die Validierung mit der uncodierten Liste der Proben bestätigte die Richtigkeit der Zuordnung, womit gleichzeitig bestätigt werden konnte, dass mit Hilfe dieser neuen Methode in beiden Labors eine präzise Diagnose der Sorte erreicht wurde. Die Identifikation wurde verifiziert, indem von allen zugeordneten Sorten Proben aus dem Obst-Nuklearstock in einem direkten Vergleich auf dem ACW-System analysiert wurden.

System-Korrelation

Die ausgeführten Analysen basieren auf dem Prinzip der Kapillar-Gelelektrophorese. Diese

Abb. 1. Fragmentanalysen mit vier Mikrosatelliten-Markern aus Marker-Set 1 der anonymisierten Proben für Apfel- (1-3) und Kirschensorten (4-6), jeweils im direkten Vergleich mit den entsprechenden Obst-Nuklearstocksorten. Die Fragmentlängen stimmen in allen Fällen sehr gut überein und bestätigen damit die Diagnose.



Technologie ist anfällig auf Unterschiede in Bezug auf die Temperatur, die Länge der Kapillare und das verwendete Polymer. Unterschiedliche Bedingungen erzeugen kleine Unterschiede in den Laufdistanzen von Genfragmenten gleicher Länge (Tab. 1 und 2). Diese Unterschiede treten auf, obwohl ein interner Standard mitläuft. Die Abweichungen sind generell klein, weshalb zwischen verschiedenen Analysesystemen eine starke Korrelation existiert. So sind zwar Abweichungen in den Laufdi-

stanzen zwischen den beiden Systemen von ecogenics und ACW zu beobachten, insgesamt ist die Korrelation jedoch linear und hoch signifikant ($y = 0,9957x + 2,2236$, $R^2 = 0,9993$).

Abweichungen in Fragmentlängen

Die leichten Unterschiede in den Laufdistanzen der Genfragmente zwischen den Analysesystemen sind für eine Sortenidentifikation irrelevant, wenn zur Bestätigung eine entsprechende Vergleichsprobe zur Verfügung

steht. Stellt sich also die Frage, ob die Testsorte mit der bekannten Sorte X identisch ist, so kann diese Frage durch eine gemeinsame Analyse beider Sorten beantwortet werden, sofern von Sorte X eine Probe vorhanden ist. Ist jedoch die Testsorte unbekannt oder die Referenzsorte nicht vorhanden, wird die Zuordnung durch den Vergleich der Fragmentlängen mit einer Datenbank ermöglicht.

Weil ein Vergleich mit einer Datenbank auf den Laufdistanzen

Tab. 1. Unterschiede in den Laufdistanzen zwischen verschiedenen Kapillarelektrophorese-Systemen

| Sorte | Analyse- bedingungen | M1A1 | M1A2 | M2A1 | M2A2 | M3A1 | M3A2 | M4A1 | M4A2 |
|------------------|-------------------------|------|------|------|------|------|------|------|------|
| Golden Delicious | A | 164 | 170 | 184 | 190 | 177 | 187 | 181 | 201 |
| | B | 165 | 171 | 185 | 190 | 177 | 187 | 183 | 203 |
| | C | 166 | 172 | 186 | 192 | 178 | 188 | 186 | 206 |
| Braeburn | A | 168 | 170 | 196 | 196 | 187 | 215 | 201 | 209 |
| | B | 169 | 171 | 196 | 196 | 187 | 213 | 203 | 211 |
| | C | 170 | 172 | 198 | 196 | 188 | 214 | 206 | 214 |
| Gala | A | 168 | 170 | 178 | 190 | 177 | 213 | 177 | 181 |
| | B | 169 | 171 | 179 | 190 | 177 | 211 | 179 | 183 |
| | C | 170 | 172 | 180 | 192 | 178 | 212 | 182 | 186 |

Bedingungen der Fragmentanalyse: A) Modell 3130, Polymer POP6, Kapillarlänge 36 cm; B) Modell 3130xl, Polymer POP7, Kapillarlänge 36 cm; C) Modell 3130, Polymer POP7, Kapillarlänge 22 cm. M1A1 = Marker 1, Allel 1; M1A2 = Marker 1, Allel 2; M2A1 = Marker 2, Allel 1, etc.

Tab. 2. Differenz der beiden Allele von vier Mikrosatelliten-Markern

| Sorte | Analyse- bedingungen | M1A2-M1A1 | M2A2-M2A1 | M3A2-M3A1 | M4A2-M4A1 |
|------------------|-------------------------|-----------|-----------|-----------|-----------|
| Golden Delicious | A | 6 | 6 | 10 | 20 |
| | B | 6 | 5 | 10 | 20 |
| | C | 6 | 6 | 10 | 20 |
| Braeburn | A | 2 | 0 | 28 | 8 |
| | B | 2 | 0 | 26 | 8 |
| | C | 2 | 0 | 26 | 8 |
| Gala | A | 2 | 12 | 36 | 4 |
| | B | 2 | 11 | 34 | 4 |
| | C | 2 | 12 | 34 | 4 |

Die Differenz ist bei Marker 1 und 4 bei allen drei Geräten identisch, dagegen zeigen sich beim Marker 2 kleine Unterschiede, die auf die Länge der Kapillaren zurück geführt werden kann. Grössere Unterschiede zeigen sich beim Marker 3, wenn unterschiedliche Polymere verwendet werden. Bedingungen der Fragmentanalyse: A) Modell 3130, Polymer POP6, Kapillarlänge 36 cm; B) Modell 3130xl, Polymer POP7, Kapillarlänge 36 cm; C) Modell 3130, Polymer POP7, Kapillarlänge 22 cm.

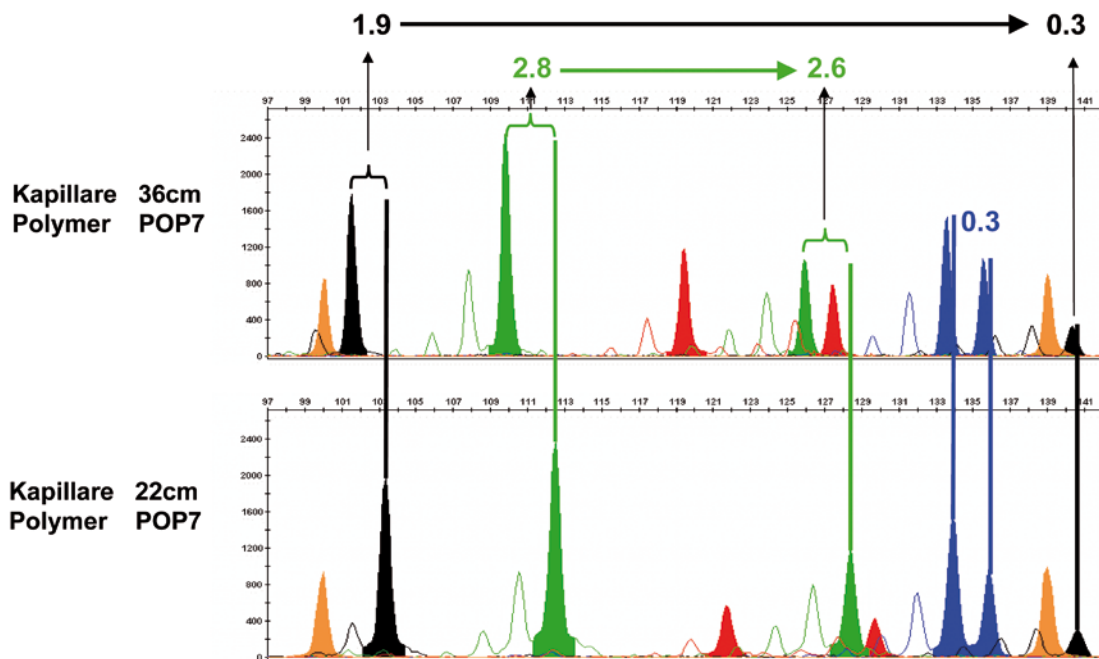
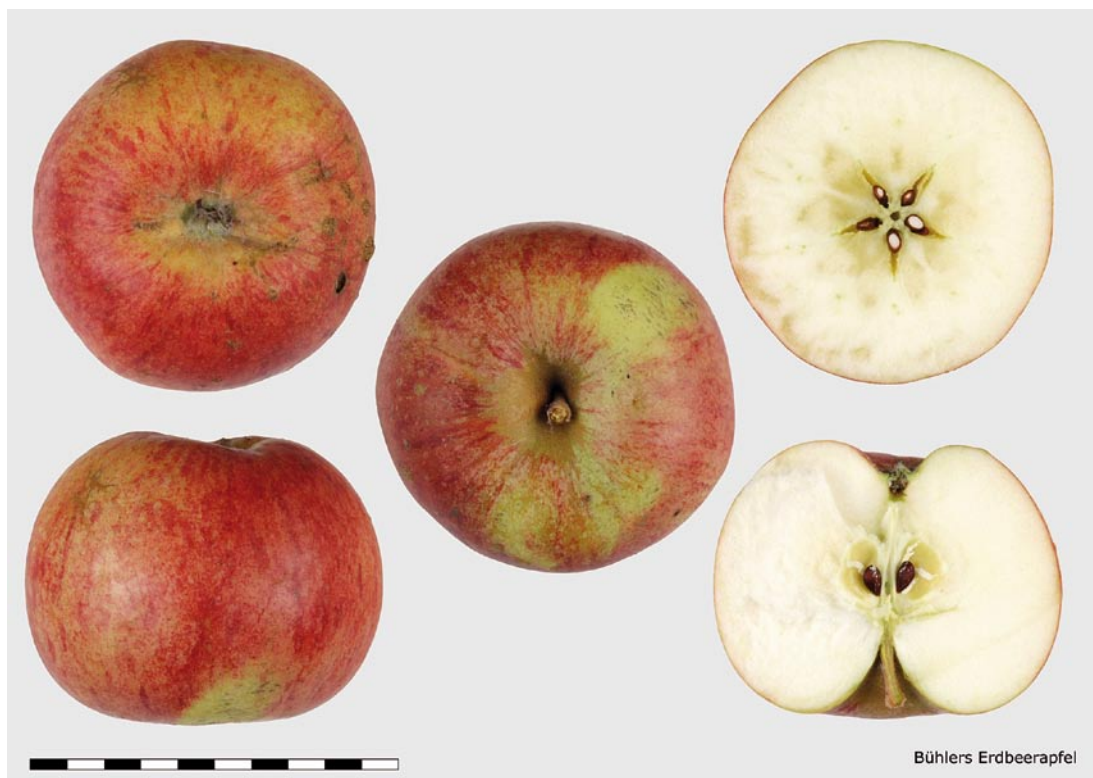


Abb. 2. Laufunterschiede bei der Fragmentanalyse mit vier Mikrosatelliten-Markern. Der blaue Marker zeigt keine wesentlichen Unterschiede in den relativen Laufdistanzen. Der grüne Marker dagegen zeigt eine Abweichung von fast 3 Basenpaaren, die aber für beide Allele praktisch gleich gross ist (2.8 vs. 2.6 Basenpaare). Der schwarze Marker zeigt eine Abweichung von 1.9 Basenpaaren im kurzen Allel, dagegen mit 0.3 Basenpaaren praktisch keine Abweichung im langen Allel. Der rote Marker zeigt dasselbe Verhalten wie der grüne Marker, bei einer kleineren Abweichung von 2.4 vs. 2.2 Basenpaaren (nicht angezeigt).

Abb. 3. Büblers Erdbeerapfel. (Foto: Forschungsanstalt Agroscope Changins-Wädenswil, ACW)



der Genfragmente basiert, sind in diesem Fall die kleinen Abweichungen problematisch, da sie eine eindeutige Zuordnung unterschiedlicher Klassierungen oft nicht zulassen (Tab. 1 und 2). So kann zum Beispiel die Fragmentlänge eines Allels als 156 klassiert werden, die Datenbank aber nur die Klassen 155 und 157 enthalten, was die Klassifizierung stark erschwert. Ein Grund für diese Schwierigkeit liegt darin, dass die vom Analysegerät gelieferten Daten nicht per se absolute Werte sind, sondern einer Interpretation bedürfen. Dazu kommt, dass die Abweichungen nicht systematisch sind. Unsere Daten zeigen, dass sowohl die Länge der verwendeten Kapillare (22 cm vs. 36 cm), der Typ des Polymers (POP6 vs. POP7) als auch die Temperatur (frühere Versuche) die Laufdistanz der Genfragmente beeinflussen. Folgende Abweichungen sind besonders schwerwiegend für die Analyse (Abb. 2):

1. Unterschiede in den Abweichungen der Laufdistanz zwi-

schen einzelnen Mikrosatelliten; einzelne Mikrosatelliten zeigen auf den beiden Systemen ein langsames oder schnelleres Laufverhalten als andere, während die Übrigen ein nahezu identisches Laufverhalten aufweisen.

2. Unterschiede im Ausmass der Abweichungen eines einzelnen Mikrosatelliten in Abhängigkeit der Fragmentgrösse; Abweichungen können bei kleinen Fragmenten grösser sein als bei grossen und umgekehrt.

Verfahren für die Sortenzuordnung

Wir entwickelten ein Programm (ein Excel-Makro), das es unter Berücksichtigung der oben erwähnten Problematik ermöglicht, die beobachteten Genotypen (die Gesamtheit der Fragmentlängen beziehungsweise Allel-Längen aller untersuchten Mikrosatelliten) mit denjenigen aus dem gesamten Obst-Nuklearstock (oder anderen Datenbanken von Mikrosatelliten) zu vergleichen und eine Rangordnung

der Ähnlichkeit zu erhalten. Wir berechnen in einem direkten Vergleich jeder Testsorte mit jedem einzelnen Kandidaten aus der Datenbank der Nuklearstock-Standardsorten die Summe der absoluten Differenz aller Genfragmente (Allele). Dieser Wert war in unserem Fall für den besten Kandidaten um mindestens den Faktor 3 tiefer als für den jeweils zweitbesten Kandidaten. Mit dieser einfachen Methode kann für jede Testsorte die wahrscheinlichste übereinstimmende Standardsorte ermittelt werden. Die Identifikation erfolgt dann im Labor durch einen direkten Vergleich der Testsorte mit dieser Standardsorte (Abb. 1).

Ausblick

Die aufgezeigten Methoden eröffnen die Möglichkeit, die Identifikations-Methode zu verfeinern und, mit der Genotypen-Datenbank im Hintergrund, via interaktive Website zu einem öffentlich zugänglichen Genotypisierung-Werkzeug (vorerst für Apfel und Kirschenarten) weiter zu entwickeln.

Literatur

- Estoup A., Presa P., Krieg F., Vaiman D. & Guyomard R., 1993. (CT)-n and (GT)-n microsatellites: A new class of genetic markers for *Salmo trutta* L. (brown trout). *Heredity* **71**, 488-496.
- Frey J.E., Frey B., Sauer C. & Kellerhals M., 2004. Efficient low-cost DNA extraction and multiplex fluorescent PCR method for marker-assisted selection (MAS) in breeding. *Plant Breeding* **123**, 554-557.
- Moran C., 1993. Microsatellite repeats in pig *Sus domestica* and chicken *Gallus domesticus* genomes. *Journal of Heredity* **84**, 274-280.
- McCouch S.R., Chen X., Panaud O., Temnykh S., Xu Y., Cho Y.G., Huang N., Ishii T. & Blair M., 1997. Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Molecular Biology* **35**, 89-99.
- Milbourne D., Russel J. & Waugh R., 1998. Comparison of molecular marker assays in inbreeding (Barley) and outbreeding (Potato) species. In: *Molecular Tools for Screening Biodiversity – Plants and animals* (Eds. A. Karp, P.G. Isaac & D.S. Ingram). Chapman and Hall, London, U.K., 371-381.
- Morgante M. & Olivieri A.M., 1993. PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal* **3**, 175-182.
- Tautz D. & Ranz M., 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* **12**, 4127-4138.
- Wang Z., Weber J.L., Zhong G. & Tanksley S.D., 1994. Survey of plant short tandem DNA repeats. *Theoretical and Applied Genetics* **88**, 1-6.

RÉSUMÉ

Validation d'une méthode de génotypage moléculaire pour les variétés fruitières.

Le génotypage assisté par microsatellite pour l'identification de cultivars est une méthode désormais couramment employée et validée chez les plantes cultivées. Il existe différentes variations autour de cette technique qui peuvent parfois différer fortement au niveau de la charge de travail et du rendement obtenu. Nous avons développé une méthode robuste et efficace que nous utilisons avec succès depuis plusieurs années dans notre laboratoire. Le but de ce projet était de tester la performance de notre méthode dans un autre laboratoire et de comparer les résultats obtenus dans les deux conditions. Nous avons pu démontrer que cette méthode pouvait être facilement mise en place dans un autre laboratoire et qu'elle délivrait des résultats identiques. Nous avons obtenu une corrélation générale hautement significative entre les deux systèmes d'analyse ($R^2 = 0.9993$). Comme attendu et malgré la forte corrélation, nous avons obtenu des différences au niveau des temps de migration pour les fragments de gènes individuels (allèles). Cependant, ces différences ne sont importantes que si aucun cultivar standard n'est connu/utilisé lors d'une analyse comparative. Nous avons donc développé une méthode afin de résoudre ce problème. Nous comparons la somme des différences de migration absolues entre le cultivar test et des cultivars présents dans une base de données pour tous les allèles. Si le cultivar standard est présent dans cette base de données, sa déviation par rapport au cultivar testé sera minimale. Le cultivar standard candidat ainsi identifié sera alors utilisé en comparaison directe. Cette méthode permet ainsi une identification non-ambigüe d'un cultivar test même en présence de déviations de résultats systématiques mais modérées entre différents systèmes d'analyse.

SUMMARY

Validation of a method for molecular genotyping of fruit varieties

Microsatellite based genotyping for the identification of cultivar varieties in crop plants is nowadays a widely used and proven method. Many variations of this method are used that differ sometimes strongly with respect to work load and efficiency. We have developed a robust and efficient method that we use with great success since several years. The aim of this project was to test how well this method performs in another laboratory, and to compare the results elaborated in this and the other laboratory. We could show that the method can easily be implemented in another laboratory and that it delivers identical results. We found a highly significant overall correlation between both analysis systems ($R^2 = 0.9993$). As expected, despite this excellent correlation, we found differences in running distances of individual gene fragments (alleles). However, these differences are only relevant if no standard cultivar sample is known/available for a comparative analysis. We developed a simple method to solve this problem. We compare the sum of the absolute differences of all alleles between the test cultivar and those of known cultivars in a database. Provided the correct standard cultivar is present in that database, its deviation from the test cultivar will be least. The candidate standard cultivar thus identified will then be used in a direct comparison. This method thus allows an unambiguous identification of a test cultivar even with moderate systematic deviations between different analysis systems.

Key words: fruit varieties, identification, molecular genotyping, microsatellite, analytical method, validation