

**Molekulargenetische Charakterisierung von Apfelgenotypen der Deutschen Genbank  
Obst mittels eines optimierten SNP Marker Arrays**

**Abschlussbericht**

Projektnummer	<b>2819BE004</b>
Auftragnehmer	SGS Institut Fresenius GmbH TraitGenetics Section Am Schwabeplan 1b 06466 Seeland OT Gatersleben Germany
Laufzeit	17.02.2020 - 30.06.2022
Berichtszeitraum	17.02.2020 - 15.07.2022
Autoren	Dr. Jörg Plieske, Dr. Heike Gnad

## **1. Ziele und Aufgabenstellung des Projekts**

Ziel des Projekts war die molekulargenetische Charakterisierung von 954 Apfelgenotypen der Deutschen Genbank Obst (DGO) sowie von 80 Vergleichssorten aus dem internationalen Apfelsortiment mittels eines optimierten SNP Marker Arrays für Apfel. Grundlage hierfür war die Entwicklung eines optimierten und kostengünstigen SNP Marker Arrays für Apfel mit etwa 50.000 Markern für die Analyse von Apfelgenotypen auf der Axiom Plattform von Thermo Fisher.

## **2. Planung und Ablauf des Projekts**

Zunächst wurde die Markerliste für das Design des neuen Arrays erstellt und an Thermo Fisher übergeben. Die Blattproben der Apfelgenotypen wurden der SGS Institut Fresenius GmbH TraitGenetics Section (SGS IF TG) von der Koordinierungsstelle der DGO zur Verfügung gestellt. Gleichzeitig erhielt SGS IF TG eine Liste der Proben im Excel-Format mit ihrer Position in den Mikrotiterplatten. Anschließend wurde die genomische DNS aller Genotypen isoliert. Die Kapazität der Axiom Array Analyse mit drei 384er Arrays lag bei zusammen 1152 Samples über der ursprünglich beauftragten Anzahl von 1034 Samples. Zur vollständigen Ausnutzung der Kapazität wurde mit der DGO daher eine entsprechende kostenneutrale Anpassung der Anzahl vereinbart. Letztendlich wurden 1152 Samples, die sich aus 1128 di- und triploiden Apfelgenotypen sowie 12 biologischen Duplikaten und 12 internen Kontrollen zusammensetzten, molekulargenetisch charakterisiert.

Die Genotypisierung erfolgte mit dem neu entwickelten Axiom Array. Die dabei zwölfmal verwendete Kontrollprobe diente der Überprüfung der Robustheit der Marker und der Reproduzierbarkeit der Daten auch für zukünftige Analysen. Die SNP-Marker wurden dann in den diploiden Apfelgenotypen mit der Axiom Analysis Suite (AxAS) und in den triploiden Genotypen mit einem R-Skript, welches von Thermo Fisher zur Verfügung gestellt wurde, analysiert und abschließend entsprechende Genotypentabellen für die diploiden und triploiden Apfelgenotypen im IUB Code erstellt.

Ein Treffen mit der Koordinierungsstelle der DGO zur Erläuterung der Ergebnistabelle und der Schritte der Datenanalyse fand aufgrund der Corona-Pandemie nicht vor Ort, sondern online am 5. Mai 2022 statt. Die ersten Ergebnistabellen sind der Koordinierungsstelle am 12. Mai 2022 zur Verfügung gestellt worden, überarbeitete Tabellen und Ergebnisse der Duplikatanalyse wurden am 14. Juli 2022 übermittelt. In diesem Zusammenhang wurde auch ein weiteres online-meeting durchgeführt.

## **3. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde**

Die Array-basierte Genotypisierung mit molekularen Markern, wie beispielsweise SNPs (single nucleotide polymorphisms), bietet die Möglichkeit viele tausend Marker im Hochdurchsatz, reproduzierbar und kosteneffizient zu untersuchen. Die hier genutzte Axiom Technologie basiert auf bewährter Chemie kombiniert mit der zuverlässigen und robusten Axiom Solid Phase Array Plattform und sorgt so für eine hohe Datenqualität, hohe Call-Raten sowie eine hohe Reproduzierbarkeit. Der neu entwickelte Axiom Array vereint qualitativ hochwertige Marker des Illumina Infinium INFFruitbreed\_Appl\_20K und des Axiom(®) Apple480K

Genotypisierungsarrays, kombiniert mit zusätzlichen neuen Markern aus noch ungenügend abgedeckten Regionen des Apfelgenoms und Markern, die mit wichtigen Eigenschaften des Apfels gekoppelt sind. Diese Neuentwicklung weist ein erheblich verbessertes Preis/Leistungsverhältnis gegenüber den bisherigen Arrays für Apfel auf, und steht nach Abschluss des Projektes der allgemeinen Forschungs- und Züchtungsgemeinschaft für die Genotypisierung zur Verfügung.

SGS IF TG (vormals die TraitGenetics GmbH) wurde im Jahr 2000 gegründet und arbeitet somit seit über 20 Jahren erfolgreich auf dem Gebiet der Pflanzengenotypisierung. SGS IF TG entwickelt nicht nur SNP-Marker, sondern bietet die routinemäßige Hochdurchsatzgenotypisierung als Dienstleistung an. Unsere Erfahrung und Expertise in der Auswertung derartiger Daten wird regelmäßig durch wissenschaftliche Publikationen demonstriert.

#### **4. Material und Methoden**

##### a) DNS Extraktion

Die DNS Extraktion wurde im 96 Well-Format durchgeführt und beruhte auf einer CTAB-Methode.

CTAB-Puffer (500ml):

- 2% CTAB (Cetyltrimethylammoniumbromid) (10g)
- 200mM Tris (100ml 1M Tris pH8)
- 20mM EDTA (20ml 0,5M Na-EDTA pH8)
- 1,4M NaCl (40,9g)
- 1% PVP (5g 40.000g/mol)

Zu 500 ml CTAB-Puffer wurden frisch 1,9 g Na-Bisulfit hinzugegeben und die Lösung auf 60°C vorgewärmt. Das getrocknete Blattmaterial wurde in 96er deep-well Mikrotiterplatten mit Kugeln vermahlen, dann 300 µl warmer Puffer zugegeben, homogenisiert und 15 min bei 60°C im Wasserbad inkubiert. Im Anschluss fand eine Dichlormethan/Isoamyl (24:1) Aufreinigung statt. Dafür wurden 300 µl Dichlormethan/Isoamyl (24:1) zugegeben, homogenisiert und 20 min bei 3.000g zentrifugiert. 115 µl Überstand wurden in eine neue Platte überführt und mit 85 µl Isopropanol gefällt. Nach der Zentrifugation (45 min, 3.500g) wurde der Überstand abgegossen, das Pellet nochmals in 70% Ethanol gewaschen, getrocknet und in 50µl Wasser gelöst.

##### b) Axiom Analyse

Der komplette Work Flow der Analyse wurde nach dem Hersteller Protokoll durchgeführt. Das entsprechende Equipment war bei SGS IF TG vorhanden (Biomek FX, GeneTitan, Peripherie). Alle Protokolle und Spezifikationen können auf der Herstellerseite bei Thermo Fisher eingesehen werden:

<https://www.thermofisher.com/de/de/home/life-science/microarray-analysis/agrigenomics-solutions-microarrays-gbs/axiom-genotyping-solution-agrigenomics.html>

##### c) Auswertung der Daten

Die Datenanalyse erfolgte für die diploiden Apfelgenotypen mit der Axiom Analysis Suite (AxAS) Software mit den Default Einstellungen des Herstellers und für die polyploiden Apfelgenotypen mit einem R-Skript von Thermo Fisher ebenfalls mit den Default Einstellungen

des Providers (Polyploid dosage estimation in Axiom data using R and Fitpoly). Für die Erstellung der Genotypen Tabelle für die polyploiden Apfelgenotypen wurde ein hausinternes Skript der SGS IF TG Bioinformatik verwendet.

Auch wurde eine Duplikat-Analyse mit einem eigens bei SGS IF TG für die Haplotypen- und Fingerprintanalyse entwickelten Skript durchgeführt. Hier wurden die genetischen Fingerprints der Sorten über alle Marker verglichen und jedem unterschiedlichen Fingerprint eine eindeutige ID zugewiesen. Genotypen (Sorten) mit gleichem Fingerprint tragen die gleiche ID. Im vorliegenden Datensatz ist nur die Auswertung diploider Proben sinnvoll, da die bei den triploiden Proben die heterozygoten Allelzustände teilweise nicht eindeutig zugeordnet werden können. Daher, konnten nur die 950 diploiden Apfelgenotypen entsprechend über das Skript analysiert werden. Die Duplikat-Analyse der triploiden Samples wurde herkömmlich anhand einer Verwandtschaftsanalyse vorgenommen. Die Verwandtschaftsanalyse und eine Hauptkomponentenanalyse wurden für beide Ploidiestufen mit NTSYS durchgeführt.

## 5. Ausführliche Darstellung der Ergebnisse

### Auswahl der SNP Marker für den neuen 50k Axiom Array

In enger Zusammenarbeit mit Dr. Andreas Peil vom Institut für Züchtungsforschung an Obst des Julius Kühn Instituts wurden die Marker nach folgenden Kriterien ausgewählt:

- gute und gleichmäßige Abdeckung des aktuellen Apfelgenoms (HFTH1)
- Vergleichbarkeit mit dem 20K Infinium und dem 480K Axiom SNP-Array
- Integration von merkmalsbezogenen und mütterlich vererbten SNP Markern
- Integration von *Malus*-Wildartspezifischen Markern

Für die gute und gleichmäßige Abdeckung des Apfelgenoms mit SNP Markern und um die in der Genomsequenzversion GDDH13 bestehenden Lücken zu füllen, wurde das neue Referenzgenom HFTH1 verwendet. Zu diesem Zweck wurde ein neues Sample Panel zur Ausrichtung auf HFTH1 definiert. Entsprechende Apfelsorten wurden aus verschiedenen Clustern einer Hauptkomponentenanalyse ausgewählt, um eine möglichst hohe Diversität sicherzustellen. Auch wurden eine Reihe von *Malus*-Wildartenakzessionen von *M. baccata*, *M. orientalis*, *M. sieversii*, *M. sylvestris* und *M. x robusta* für die Neuausrichtung auf HFTH1 ausgewählt um mögliche Introgressionen dieser Wildarten bei domestizierten Arten identifizieren zu können.

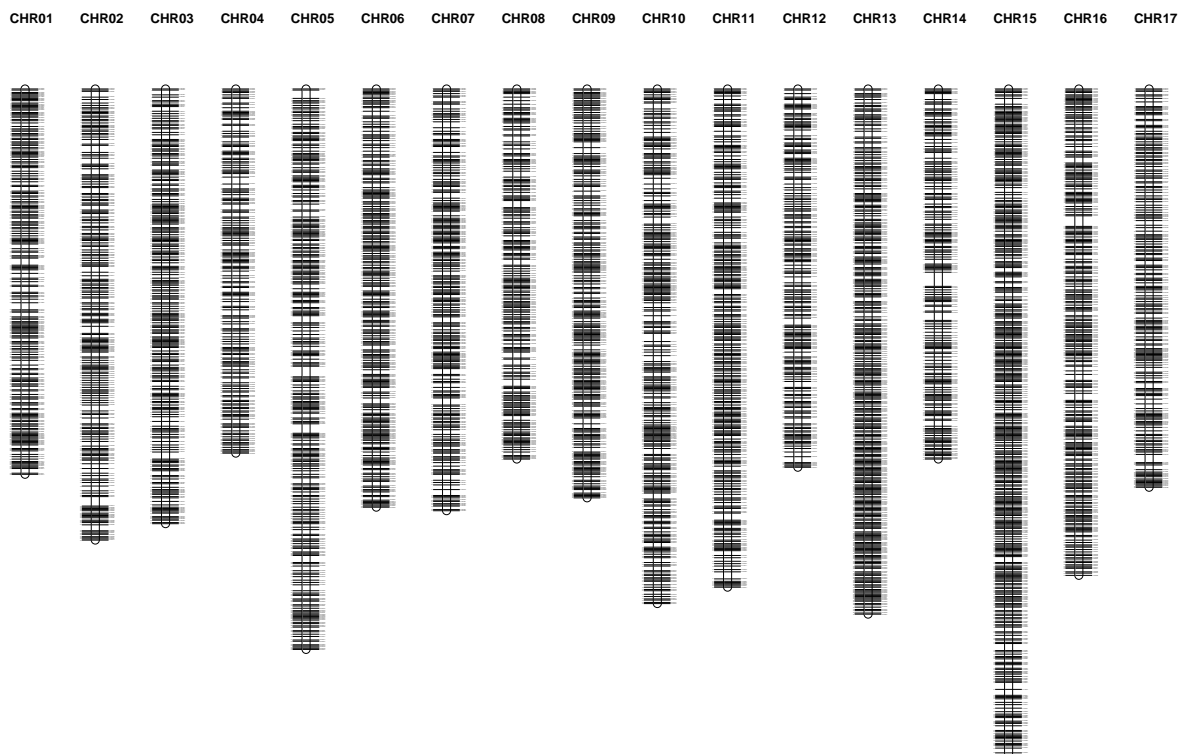
Für die Abdeckung des neuen Apfelgenoms HFTH1 wurden sogenannte Haploblöcke definiert, wobei jeder Haploblock sich über 125 Kilobasen erstreckte. Zusätzliche Blöcke wurden an den Chromosomenenden, die eine höhere Rekombination aufweisen, definiert. Jeder Block sollte durch acht SNP Marker in einem Fenster von 10 Kilobasen repräsentiert werden. Bei einer geschätzten Größe des Apfelgenoms von 651 Megabasen wurden dafür mindestens 41.800 SNPs berechnet. Die verbliebenen SNPs wurden für *Malus*-Wildart-spezifische SNPs und für SNPs im Zusammenhang mit züchterisch interessanten Merkmalen (z. B. Resistenz, Qualität, Selbstinkompatibilität) sowie für maternal vererbte SNPs aus den Plastidengenomen reserviert (s. Tab. 1). Die physikalische Verteilung der Marker ist in Abb. 1 zu sehen.

Die finale Liste der ausgewählten SNP Marker wurde dann in der entsprechenden Vorlage für das Assay Design an Thermo Fisher übergeben. Für die Analyse von 1128 unterschiedlichen Genotypen wurden drei 384er Arrays bestellt, die zusammen mit allen notwendigen

Informationen (Annotation und Analyse-Files), Verbrauchsmittel und Reaktionschemie von Thermo Fisher geliefert wurden.

**Tab. 1:** Anzahl Marker je Gruppe, die letztendlich geeignet waren für das Assay Design und im Annotation File zu finden sind

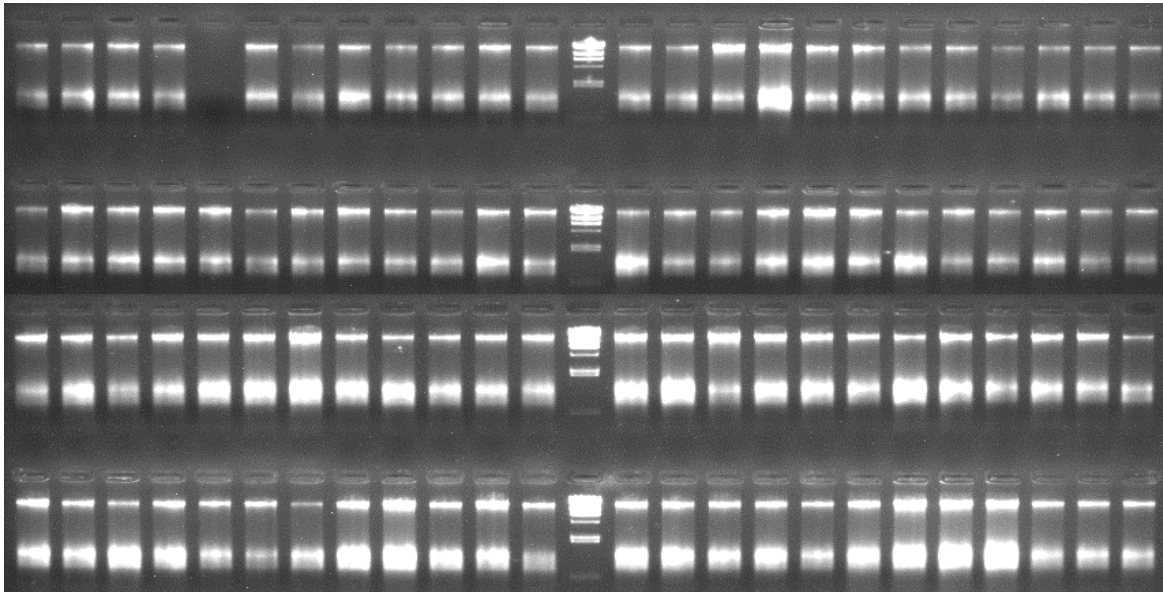
Gruppe	Anzahl
20K/480K kompatible Marker	37450
neue Haploblock spezifische Marker	7573
Wild Arten spezifische Marker	2945
Marker für S-Allele	112
Merkmals-spezifische Marker	54
Maternal vererbte Marker	4
Summe	48138



**Abb. 1:** Physikalische Karte des Apfelgenoms. Auflösung 1 Marker / 100.000 Basen, Gesamtlänge 651.861.844 Basen

## DNS Extraktion

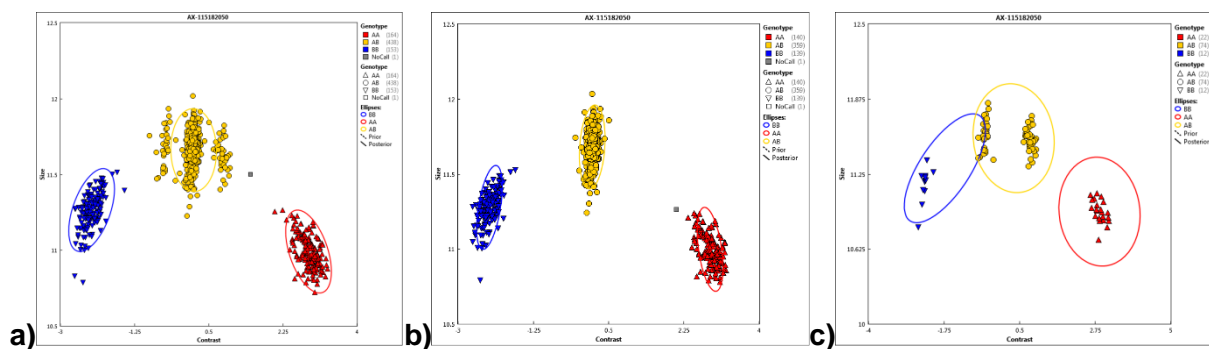
Aus den von der DGO gelieferten Blattproben wurde nach einem optimierten Protokoll die DNS der Genotypen erfolgreich extrahiert und auf Agarosegelen auf Qualität und Quantität überprüft (Abb. 2).



**Abb. 2:** Agarosegelbild der DNS Qualität. Platte 5 mit 95 Genotypen und der laufenden Wasserkontrolle an Position A5 (obere Reihe von links nach rechts: die Reihen A und B in der Mikrotiterplatte, darunter fortlaufend bis Reihe H)

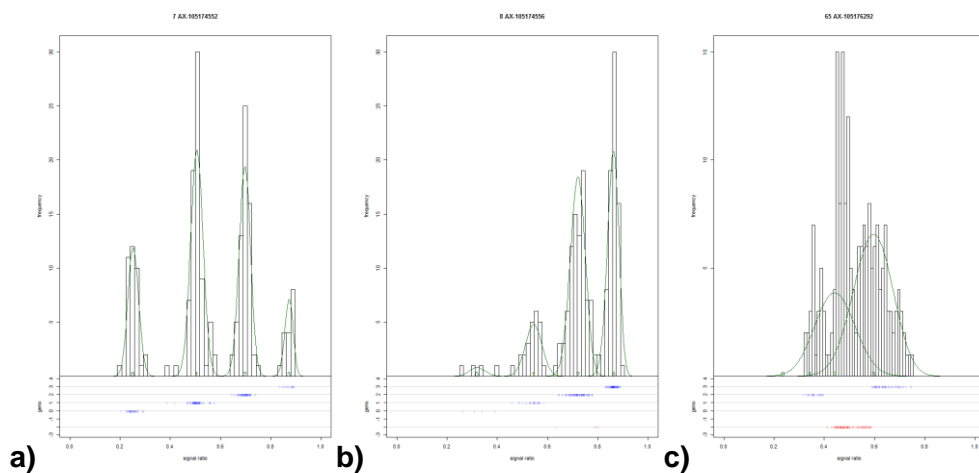
## Axiom Analyse

Die DNS der Genotypen wurde anschließend für die Axiom Analyse entsprechend des Herstellerprotokolls auf 20ng/ $\mu$ l verdünnt und auf dem Biomek FX weiterbearbeitet. Die Kontrolle der Hybridisierung zeigte höchste Qualität, so dass die Proben für die eigentliche Arrayanalyse auf dem GeneTitan fortgeführt und abschließend die CEL Files in der AxAS analysiert werden konnten.



**Abb. 3:** Axiom Cluster Analyse mit der AxAS Software. **a)** diploide und triploide Genotypen zusammen, **b)** nur diploide Genotypen, **c)** nur triploide Genotypen

Zunächst wurden alle Genotypen ungeachtet ihrer Ploidiestufe zusammen analysiert (Abb. 3a). Anhand der Clusteranalyse ausgewählter SNP Marker konnten die Angaben der DGO bezüglich der Ploidie überprüft und die Proben anschließend entsprechend geteilt und separat analysiert werden (Abb. 3b und c). Die diploiden Proben wurden in der AxAS Software und die triploiden Proben mit dem R-Skript abschließend analysiert und die entsprechenden Genotypentabellen erstellt.



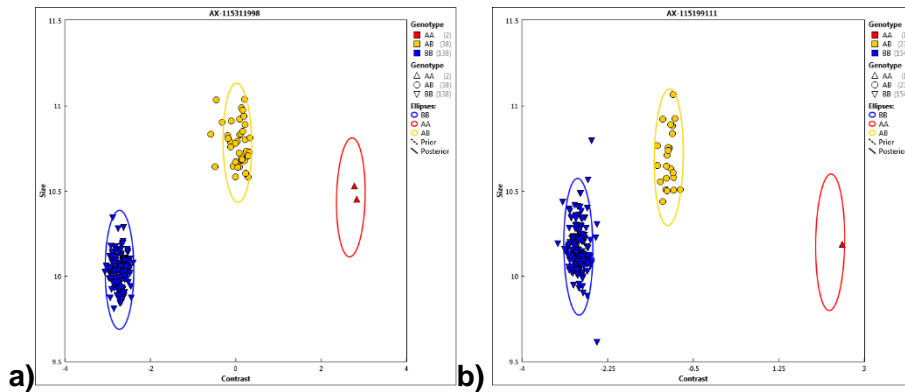
**Abb. 4:** Axiom Cluster Analyse mit dem R-Skript. **a)** qualitativ hochwertiger Marker, **b)** Marker mit ungleichmäßiger Allel-Verteilung, **c)** Marker mit ungenügender Trennung der Klassen

Alle 1128 unterschiedlichen Apfelgenotypen konnten erfolgreich analysiert werden. Es wurden 950 Genotypen als diploid und 178 Genotypen als triploid identifiziert. Die entsprechenden Call Rates für die qualitative Einordnung der Ergebnisse für jeden Genotypen sind in den jeweiligen Genotypen-Tabellen enthalten. Die qualitative Zuordnung der Marker zu sogenannten „Conversion Types“ können nur der Genotypen-Tabelle für die diploiden Samples entnommen werden. Für die triploiden Samples ist eine solche Einordnung der Marker nicht möglich, da die Analyse mit einem R-Skript durchgeführt wurde, welches diesen Parameter so nicht ausgibt. Hier muss im Einzelfall die Qualität des Markers anhand einer Balkengrafik abgeschätzt werden, die nach Bedarf aus dem Skript ausgegeben werden kann (Abb. 4).

Die Qualität der Analyse der diploiden Genotypen ist sehr hoch. Über 78% der Marker wurden dem hochwertigsten Conversion Type „PolyHighResolution“ zugeordnet. Zusammen mit ausgewählten Markern aus den anderen Gruppen können über 87% der Marker für die molekulargenetische Charakterisierung der Apfelgenotypen empfohlen werden (Parameter und Grenzwerte: FLD Score über 5, HomFLD Score über 10, Call Rate über 90). Die Call Rate über alle Samples lag bei 99,0 die Übereinstimmung der 12 Duplikate der Kontrollprobe SNP0042 lag bei 99,7 über alle Marker – wobei fast 90% der Abweichungen auf nur eine der 12 Kontrollproben zurückzuführen war. Möglicherweise war bei dieser Probe die DNS Qualität nicht optimal, was zu dieser ungewöhnlich hohen Abweichung beigetragen haben könnte.

Die Qualität der Analyse der triploiden Genotypen ist leider nicht entsprechend genau zu erfassen. Anders als bei den diploiden Genotypen gibt es hier vier Allel-Klassen - AAA, AAB,

ABB, BBB - wobei die beiden heterozygoten Allel-Klassen nicht immer gut zu trennen waren bzw. es nicht immer eindeutig war, ob die heterozygoten Genotypen beide Klassen oder nur eine der beiden Klassen und wenn ja, welche der beiden repräsentierten (Abb. 5).



**Abb. 5:** Axiom Cluster Analyse der triploiden Genotypen mit der AxAS Software. **a)** Marker mit möglicherweise ungenügender Trennung der beiden heterozygoten Allel-Klassen (AAB und ABB), **b)** Marker mit möglicherweise nur einer heterozygoten Allel-Klasse (AAB)

Da die AxAS Software nicht in der Lage war, die vier Allel-Klassen zu handhaben, musste für die Analyse der triploiden Genotypen wie beschrieben das R-Skript verwendet werden. Die Analyse mit diesem Skript ließ eine mit der Analyse der diploiden Proben vergleichbare Qualitätseinschätzung der Marker nicht zu. Auch kann mittels des Skriptes keine Genotyp-Tabelle des gewohnten SGS IF TG Formats erstellt werden. Das Ergebnis des Skriptes war eine Langtabelle, die zunächst mit Hilfe einer exklusiv für diese Anwendung entwickelten Software der Bioinformatischen Arbeitsgruppe der SGS IF TG in eine entsprechende Genotypentabelle umgewandelt werden musste (Abb. 6).

marker	MarkerName	SampleName	ratio	P0	P1	P2	P3	margin	mapP	genP
1	AK-10174206	220302_Axiom_HDS04M-384_SNP001-01-AB1.CEL	0.26509913	0.99999999	6.11E-10	1.18E-44	1.96E-114	0	0.99999999	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP001-01-AB1.CEL	0.42819777	1	9.30E-17	1.02E-63	3	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP001-01-AB1.CEL	0.22017832	1	3.45E-16	1.34E-56	1.17E-151	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP001-01-AB1.CEL	0.47203503	1	4.66E-12	1.27E-35	1	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP002-01-021.CEL	0.16790727	1	3.65E-24	4.69E-71	6.65E-156	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP005-01-011.CEL	0.45507002	2.39E-15	1	7.94E-14	1.63E-59	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP007-01-011.CEL	0.49115659	1.16E-14	1	1.72E-14	7.01E-61	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP009-01-021.CEL	0.20202558	1	7.16E-19	4.52E-61	1.79E-149	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP011-01-021.CEL	0.21937955	1	6.82E-17	2.32E-57	9.77E-143	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP013-01-021.CEL	0.19211359	1	2.32E-20	6.99E-64	8.14E-143	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP013-01-021.CEL	0.89487561	9.99E-131	2.65E-54	1.56E-164	1	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP041-01-041.CEL	0.66354482	5.53E-10	5.00E-10	0.99999999	2.43E-20	0	0.99999999	2
1	AK-10174206	220302_Axiom_HDS04M-384_SNP068-01-021.CEL	0.89912487	1.95E-133	3.96E-58	3.79E-177	1	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP069-01-021.CEL	0.47324251	1.44E-17	1	6.42E-12	2.52E-35	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP069-01-021.CEL	0.61941412	2.46E-43	6.66E-67	0.99999934	5.74E-24	2	0.99999934	2
1	AK-10174206	220302_Axiom_HDS04M-384_SNP080-01-021.CEL	0.23441207	1	3.52E-14	3.19E-52	8.51E-127	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP090-01-011.CEL	0.90819455	7.24E-139	1.05E-61	2.09E-192	1	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP092-01-011.CEL	0.37284677	2.21E-05	0.99997787	1.02E-22	1.17E-78	0	0.99997787	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP096-01-021.CEL	0.19746181	1	1.49E-19	2.34E-62	1.83E-142	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP092-01-021.CEL	0.21501478	1	3.85E-17	1.89E-57	6.70E-143	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP101-01-021.CEL	0.51061114	5.07E-22	0.99999943	3.74E-68	8.73E-47	0	0.99999943	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP120-01-021.CEL	0.26202977	1	2.94E-10	3.19E-40	1.19E-101	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP125-01-021.CEL	0.70363195	1.83E-66	8.54E-18	1	2.53E-11	2	1	2
1	AK-10174206	220302_Axiom_HDS04M-384_SNP130-01-021.CEL	0.47279944	8.54E-18	1	1.05E-11	7.02E-55	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP134-01-021.CEL	0.67309014	6.70E-52	6.31E-11	1	2.86E-19	1	1	2
1	AK-10174206	220302_Axiom_HDS04M-384_SNP141-01-021.CEL	0.88870473	1.75E-127	3.24E-54	2.18E-15	1	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP143-01-021.CEL	0.41930053	5.63E-11	1	1.03E-17	7.58E-68	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP153-01-021.CEL	0.42004883	5.86E-13	1	5.32E-36	3.82E-64	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP159-01-021.CEL	0.48121784	2.38E-16	1	5.40E-13	1.20E-37	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP170-01-021.CEL	0.19042814	1	1.12E-20	2.40E-64	1.55E-140	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP184-01-021.CEL	0.70401586	6.70E-59	3.23E-14	1	1.77E-15	2	1	2
1	AK-10174206	220302_Axiom_HDS04M-384_SNP185-01-021.CEL	0.68240114	1.47E-54	3.54E-12	1	7.60E-18	2	1	2
1	AK-10174206	220302_Axiom_HDS04M-384_SNP190-01-021.CEL	0.20247352	1	1.16E-18	1.12E-60	7.23E-146	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP205-01-021.CEL	0.19380804	1	8.37E-20	7.87E-63	3.44E-141	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP210-01-021.CEL	0.23656294	1	7.05E-14	1.18E-51	6.66E-124	0	1	0
1	AK-10174206	220302_Axiom_HDS04M-384_SNP248-01-021.CEL	0.48304473	1.00E-18	1	8.96E-11	4.32E-31	1	1	1
1	AK-10174206	220302_Axiom_HDS04M-384_SNP248-01-021.CEL	0.51312285	4.06E-143	1.73E-64	6.61E-21	1	1	1	1

**Abb. 6:** Auswertung der triploiden Proben. **a)** Kopf der Langtabelle des R-Skriptes von Thermo Fisher und **b)** gewohntes Format der Genotyp-Tabelle der SGS IF TG



Die Identifizierung der Duplikate bei den diploiden Samples wurde wie oben beschrieben mit dem von SGS IF TG entwickelten Skript zur Haplotypenidentifizierung durchgeführt. Dafür wurden anhand der beschriebenen Qualitätsparameter (Call Rate, FLD Score) 16.197 beste Marker aus dem gesamten Marker set von 48.138 ausgewählt, um eine möglichst vertrauenswürdige Analyse zu gewährleisten und den technischen Fehler bedingt durch Fehlklassifizierungen bei Markern geringerer Qualität so klein wie möglich zu halten. Das Skript wurde für diese Analysen eigens modifiziert und optimiert. Es identifiziert über alle ausgewählten Marker den entsprechenden genetischen Fingerabdruck jeder Akzession. Dieser Fingerabdruck wird anschließend zwischen allen Akzessionen abgeglichen, identische Muster identifiziert und einem Haplotypen zugeordnet. Jedem unterscheidbaren Haplotyp wird eine eindeutige Bezeichnung – Haplotypen ID – zugewiesen. Die Haplotypen ID jeder Akzession wurde dann in der Genotypentabelle vermerkt und die Daten entsprechend sortiert. Akzessionen mit gleicher Haplotypen ID wurden farblich markiert (Abb. 7). Da das Skript aufgrund der teils unklaren Allelzustände nur für diploide Spezies sinnvoll angewendet werden kann, konnten nur die 950 diploiden Apfelgenotypen entsprechend analysiert werden. Insgesamt wurden bei den diploiden Apfelgenotypen 717 unterschiedliche Haplotypen identifiziert.

The image shows a detailed genotype table with multiple columns. The first few columns include 'Sample\_ID', 'Conversion Type empfahen', 'Minor Allel Frequency', 'CR', and 'FLD'. The rest of the table is a dense grid of data points, likely representing marker genotypes. Several rows are highlighted in yellow and orange, representing duplicate samples. The table is very large and contains a lot of repetitive data.

**Abb. 7:** Ausschnitt aus der Genotypentabelle. Die gelbe und orange Markierung der Haplotypen zeigt die Samples mit jeweils identischem Fingerabdruck. Die Kontrollproben sind grün markiert.

Für die Identifizierung der Duplikate bei den triploiden Samples wurde eine Verwandtschaftsanalyse mit letztendlich 9.248 ausgewählten Markern durchgeführt. Sollte aufgrund zusätzlicher Ergebnisse zukünftig auch eine Duplikate-Analyse für triploide Proben mittels des bioinformatischen Skriptes möglich sein, reichen wir diese Daten nach. Insgesamt wurden bei den triploiden Apfelgenotypen 126 unterschiedliche Haplotypen identifiziert.

In Unkenntnis des Materials und der Situation in der Obstbaumzucht, können die Ergebnisse unsererseits nicht seriös geprüft werden. Es wurden die Kontrollen und einige gleichlautende Apfelgenotypen jeweils einem Haplotypen zusammen zugeordnet. Die weitere Einschätzung der jeweiligen Zuordnung kann nur vom JKI vorgenommen werden.

## **6. Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse**

Die Markerzusammenstellung für den optimierten Array ist erfolgreich durchgeführt worden. Der neue Array bietet jetzt ein qualitativ hochwertiges Werkzeug für die reproduzierbare Charakterisierung von Apfelgenotypen in den kommenden Jahren. Durch die Einbindung von Markern von den beiden Vorgänger Arrays 20K Infinium und 480K Axiom können auch bereits vorhandene Datensätze, die mit diesen beiden Arrays erzeugt worden sind, in die aktuellen Analysen mit eingebunden werden. Die öffentliche Verfügbarkeit des neuen Arrays bietet die Chance für eine vielfältige Kooperation bei Forschungs- und Entwicklungsprojekten weltweit und der gemeinsamen Verwertung der Ergebnisse.

Die Genotypisierungsarbeiten der 1128 Apfelgenotypen wurden ebenfalls erfolgreich durchgeführt. Es wurden qualitativ sehr hochwertige und reproduzierbare Ergebnisse erzielt und die Duplikate der Kontrolle zuverlässig identifiziert. Den voraussichtlichen Nutzen und die genaue Verwertbarkeit der Ergebnisse kann nur vom Auftraggeber beurteilt werden, da die genauen Eigenschaften des zu untersuchenden Materials SGS IF TG unbekannt sind.

## **7. Zusammenfassung**

Ziel des Projekts war die Entwicklung eines optimierten, kostengünstigen Apfel SNP-Marker Arrays mit etwa 50.000 Markern für die Analyse von Apfelgenotypen auf der Axiom Plattform von Thermo Fisher und die molekulargenetische Charakterisierung von 1128 di- und triploiden Apfelgenotypen, von denen 954 aus der Deutschen Genbank Obst (DGO) stammten. Alle 1128 Genotypen wurden mit dem Array erfolgreich analysiert, es wurde ein qualitativ hochwertiger Datensatz sowohl der di- als auch der triploiden Apfel-Genotypen erzeugt. Über 87% der Marker können für die Analyse zumindest von diploidem Material empfohlen werden. Die Analyse des triploiden Materials lässt nur eine eingeschränkte Empfehlung bezüglich Markerqualität zu, sie muss anhand von Balkendiagrammen individuell für jeden Marker bestimmt werden. Die Duplikaten-Analyse des Materials ergab bei den diploiden Genotypen insgesamt 717 unterschiedliche Haplotypen und bei den triploiden Genotypen 126.

## **8. Gegenüberstellung der ursprünglich geplanten zu den tatsächlich erreichten Zielen; ggf. mit Hinweisen auf weiterführende Fragestellungen**

Ziel des Projekts war die Entwicklung eines optimierten SNP Marker Arrays für die Analyse von Apfel Genotypen auf der Axiom Plattform und die molekulargenetische Charakterisierung von di- und triploiden Apfelgenotypen. Erreicht wurde sowohl die Entwicklung des Arrays als auch die vertrauenswürdige Charakterisierung der Apfelgenotypen mittels dieses Arrays anhand von qualitativ hochwertig isolierter DNS.

## 9. Literaturverzeichnis

Polyploid dosage estimation in Axiom data using R and Fitpoly. Thermo Fisher (PDF File)

## 10. Liste und kurze Beschreibung der Daten auf der DVD

- **Axiom Library Files:** alle relevanten Files, die von Thermo Fisher für die Analyse der Axiom Arrays zur Verfügung gestellt wurden - inklusive des Annotation Files mit allen Informationen zu den SNP-Markern wie Sequenz und Allel-Zuordnung
- **Berichte:** Abschlussbericht und Kurzfassungen (deutsch und englisch)
- **CEL-Files:** Roh-Daten für die Analyse/Auswertung mit der AxAS Software
- **Genotyp-Tabellen:** Genotyp-Tabellen für die 950 diploiden und 178 triploiden Apfel-Genotypen
- **Platten-Layout:** Layout und Bilder der Platten bzw. der DNS QC (Agarosegele)
- **R-Skript:** R-Skript, Packages und Anleitungen für die Analyse von Axiom Daten polyploider Genotypen (von Thermo Fisher)
- **R-Skript Apfel Plots Triploide:** Balkengrafiken der SNP-Marker für die triploiden Apfel-Genotypen aus dem R-Skript von Thermo Fisher
- **Link AxAS.docx:** Link zum Download der AxAS Software von Thermo Fisher