

Molekulargenetische Charakterisierung von Kirschengenotypen der Deutschen Genbank Obst mittels eines SNP Marker Arrays

Abschlussbericht

Auftragnehmer	TraitGenetics GmbH Am Schwabeplan 1b 06466 Stadt Seeland OT Gatersleben Germany
Projektnummer	2819BE001
Laufzeit	15.08.2019 – 10.07.2020
Berichtszeitraum	20.09.19 – 20.06.20
Autoren	Dr. Anja Hohmeyer, Dr. Jörg Plieske, Dr. Martin Ganai

1. Ziele und Aufgabenstellung des Projekts:

Ziel des Projekts ist die molekulargenetische Charakterisierung von 296 Süßkirschen- und 66 Sauerkirschengenotypen der Deutschen Genbank Obst (DGO) sowie von jeweils 50 Vergleichssorten aus dem internationalen Kirschensortiment, einschließlich der dafür notwendigen Isolierung der genomischen DNA aller 462 Genotypen, mittels des hochauflösenden Infinium RosBREED Cherry_V2 Genotyping Arrays sowie die Identifizierung möglicher Duplikate.

2. Planung und Ablauf des Projekts:

Die für die Genotypisierung notwendigen Blattproben wurden TraitGenetics (TG) zu Beginn des Projektes von der Koordinierungsstelle der DGO zur Verfügung gestellt. Gleichzeitig erhielt TG eine Liste (Excel-Format) mit Informationen zu den Proben, in der folgende Daten enthalten sind: Sortenname, Nummer der molekularen Gruppe, Akzessionsnummer, Standort des Baumes und Probennummer. Zunächst erfolgte die Isolierung der genomischen DNA aller 462 Genotypen. Die Genotypisierung der 462 Kirschensorten (zusätzlich 18 Kontrollproben von TG) der DGO durch TG erfolgte mit dem Infinium RosBREED Cherry_V2 Genotyping Array. Die von TG verwendeten Kontrollen dienten der Überprüfung der Robustheit der Marker und der Reproduzierbarkeit dieser und möglichen weiteren Analysen. TG entwickelte basierend auf dem Material zwei Clusterfiles (diploid und tetraploid). Im Anschluss wurden die Rohdaten mit Hilfe der Cluster Files ausgewertet und eine Genotypentabelle im IUB Code erstellt.

Die Genotypentabelle sollte den Sortennamen, die Nummer der molekularen Gruppe, die Akzessionsnummer, die Standortnummer des Baumes sowie die Probennummer und die Analysedaten für die ausgewählten Marker enthalten. Danach sollten alle Proben nach identischen Fingerprints sortiert und jedem einzelnen Fingerprint eine eindeutige Identifikationsnummer ID zugeordnet werden. Diese eindeutige ID muss in einer gesonderten Spalte ebenfalls in der Tabelle aufgeführt werden. Mögliche Duplikate sollten in dieser Tabelle sichtbar gekennzeichnet werden (z.B. farblich hinterlegt). Die Übergabe der Ergebnisse muss auf elektronischem Wege (z. B. als MS-Excel Tabelle) erfolgen.

3. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde:

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 Illumina Infinium Technologie basiert auf bewährter Chemie kombiniert mit der zuverlässigen Bead Array-Plattform und sorgt so für eine hohe Datenqualität, hohe Call-Raten sowie eine hohe Reproduzierbarkeit. Der Infinium RosBREED Cherry_V2 Genotyping Array bzw. sein Vorgängerarray RosBREED cherry 6K SNP array v1 wurde für die diploide Süßkirsche (*Prunus avium*) und allotetraploide Sauerkirsche (*P. cerasus*) entwickelt. Die ausgewählten SNPs basieren auf der Next-Generation Sequenzierung von diversem genetischem Material, die 25 Milliarden Basenpaare (Gb) Kirschensequenzen zur Verfügung stellte. Basierend darauf wurden genomweite SNPs für die Süßkirsche und für die beiden Subgenome der Sauerkirsche (Avium Subgenom und Fruticosa Subgenom) identifiziert. Da der genutzte Array bereits vom Auftraggeber ausgewählt wurde, wird hier nicht weiter im Detail auf die Zusammensetzung und Qualität des Arrays eingegangen.

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

4. Material und Methoden:

a) DNS Extraktion

Die DNS Extraktion wird im 96 Well-Format durchgeführt und beruht 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 werden frisch 1,9 g Na-Bisulfit hinzugegeben und die Lösung auf 60°C vorgewärmt. Das Blattmaterial wird in 96er deep-well Mikrotiterplatten mit Kugeln unter Stickstoff vermahlen, dann 400 µl warmer Puffer zugegeben, homogenisiert und 15 min bei 60°C im Wasserbad inkubiert. Im Anschluss findet eine Chloroform/Isoamyl (24:1) Aufreinigung statt. Dafür werden 400 µl Chloroform/Isoamyl (24:1) zugegeben, homogenisiert und 20 min bei 3.000g zentrifugiert. 115 µl Überstand werden in eine neue Platte überführt und mit 85 µl Isopropanol gefällt. Nach der Zentrifugation (45 min, 3.500g) wird der Überstand abgegossen, das Pellet getrocknet und in Wasser gelöst.

b) Genotypisierung mit Infinium RosBREED Cherry V2 Genotyping Array

Die Details der Genotypisierung können dem im Anhang befindlichen Infinium® HD Assay Ultra Protocol (Guideinfinium_hd_ultra_user_guide_11328087_revb Illumina HD Protokolls) von Illumina entnommen werden.

Der Array basiert auf dem Infinium HD-Typ und kann parallel 13.559 Marker analysieren. Diese setzen sich aus Markern speziell für die Süßkirsche und aus sauerkirschspezifischen Markern zusammen.

d) Clusterfileentwicklung und Auswertung der Daten

Die Rohdaten aus dem gescannten Array werden mit Hilfe der Genomstudio Software ausgewertet (User Guide befindet sich im Anhang). Das Clusterfile dient zur Qualitätskontrolle der Marker und als Schablone für eine reproduzierbare Bestimmung der Allele in einzelnen Experimenten. Zu diesem Zweck wird versucht das Clusterfile manuell für alle 13.559 Marker, sowohl für die Süßkirsche als auch für die Sauerkirsche, anzupassen. Auf dieser Basis werden anschließend die Genotypendaten generiert, die mit Bezug zur Quellsequenz im IUB Code ausgegeben werden.

c) NTSYS Analyse

Die Verwandtschaftsanalyse wird mit Hilfe der Software NTSYS, basierend auf den zuvor generierten Genotypdaten, vorgenommen. Hierzu werden ausschließlich die Daten aller polymorphen Marker genutzt und sowohl ein Stammbaum als auch eine Tabelle mit den paarweisen Vergleichen aller Proben ausgegeben. Die Ergebnisse werden in eine Genotypentabelle übernommen und alle Proben nach identischen Fingerprints sortiert. Jedem einzelnen Fingerprint wird eine eindeutige Identifikationsnummer ID zugeordnet.

5. Ausführliche Darstellung der Ergebnisse

Die Probenübergabe an TG erfolgte am 06.12.2019 zusammen mit allen nötigen digitalen Informationen. Im Anschluss erfolgte die DNS Extraktion. Während der Aufreinigung wurden zur Sicherheit zwei Abnahmen gemacht und beide separat aufgereinigt. Eine detaillierte Qualitätskontrolle der DNS-Extraktion ist im Anhang „200116_PCSA19001_ 1. Übergabe Kirsche Gatersleben“

dargestellt. Dort ist ebenso dargestellt, welche Proben aus welcher der beiden Aufreinigungen für die weitere Genotypisierung verwendet wurden.

Die Genotypisierung erfolgte in insgesamt 5 96-Well Mikrotiterplatten nach dem im Anhang befindlichen Protokoll des Illumina Infinium HD Workflows „Infinium® HD Assay Ultra Protocol“. Die Rohdaten wurden mit Hilfe der Software Genome Studio 2.0 ausgewertet. Zunächst wurde ein optimiertes Clusterfile für die Auswertung der Süßkirschenproben erstellt. Da die Süßkirsche ein diploider Organismus ist, konnten sehr zuverlässig drei Cluster für die drei unterschiedlichen allelischen Zustände gefunden und ausgewertet werden (Abbildung 1).

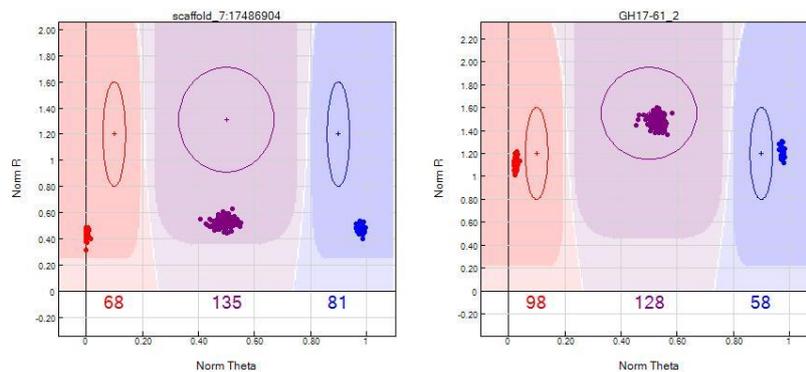


Abbildung 1: Drei Cluster für die drei unterschiedlichen allelischen Zustände der diploiden Süßkirsche exemplarisch dargestellt für zwei verschiedene Marker.

Die drei klar unterscheidbaren Cluster entsprechen den Genotypen AA (homozygot Allele 1); AB (heterozygot) und BB (homozygot Allele 2) und erlaubt die Erstellung einer Genotypentabelle. Es konnten nur 3.855 der 13.559 Marker als polymorph identifiziert werden. Diese niedrige Anzahl polymorpher Marker deutet entweder auf eine enge genetische Basis des untersuchten Materials oder auf einen sogenannten Ascertainment-Bias (die Marker wurden in anderem genetischen Material als dem untersuchten ausgewählt) hin. Die Genotypen wurden im IUB Code und in Bezug auf die Quellsequenz ‚forward‘ ausgegeben. Die Quellsequenz wurde aus dem von Illumina zur Verfügung gestellten Manifest entnommen. Die ausführliche Genotypentabelle der Süßkirschenproben ist im Anhang unter „200127_PCS19001_Süßkirsche_GT“ dargestellt. Die Verwandtschaftsanalyse wurde mit Hilfe der Software NTSYS durchgeführt und hat zuverlässig sowohl die technischen als auch die biologischen Replikate identifiziert. Außerdem konnten etliche weitere Duplikate im Material gefunden und entsprechend der Vorgaben in der Genotypentabelle kenntlich gemacht werden. Das Ergebnis ist eine Reduktion der 353 verschiedenen Genotypen auf nur noch 303. In Abbildung 2 ist ein Ausschnitt der Genotypentabelle mit Kennzeichnung der entsprechenden Duplikate gezeigt. Das Clusterfile steht dem Kunden nun für die standardisierte und reproduzierbare Auswertung weiterer Analysen mit diesem Array zur Verfügung.

dies praktisch nur bei sehr wenigen Markern der Fall. Dazu kommt noch, dass aufgrund der geringen Probenzahl bei der Sauerkirsche in vielen Fällen nicht alle 5 möglichen Cluster beobachtet werden und damit die klare Zuordnung der bis zu 5 allelischen Zustände schwierig ist (Abbildung 3b-c)

Weiterhin verhalten sich knapp 10% der Marker monomorph, weil sie nur zwischen den Genomen polymorph (Abbildung 4a) sind. Diese Marker sind ebenfalls nicht nutzbar für die Verwandtschaftsanalysen. Viele polymorphe Marker zeigen auch keine klar definierten Cluster, was bei fünf Clustern zu sehr unscharfen und ineinander verlaufende Cluster (Schmier) führt, welche somit nicht eindeutig auswertbar sind (Abbildung 4b). Derartige nicht klar differenzierte Cluster entstehen oft, wenn noch zusätzlich zu dem zu untersuchenden SNP, flankierende Polymorphismen auftreten bzw. viele der verschiedenen möglichen Allelzustände gleichzeitig auftreten.

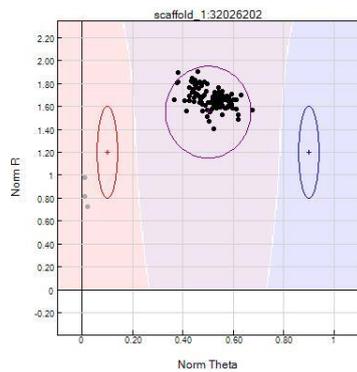


Abbildung 4a

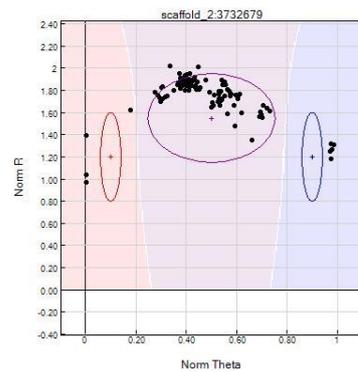


Abbildung 4b

Abbildung 4a:

Beispielbild für einen Marker der sehr wahrscheinlich nur polymorph zwischen den Genomen ist.

Abbildung 4b:

Beispielbild für einen Marker dessen verschiedenen allelischen Zuständen (und vermutlich flankierende SNPs) zu einem Schmier führen, in dem keine klar separierten Cluster auftreten.

Zunächst wurde ein Clusterfile erstellt, das nur die 16 eindeutig auswertbaren Marker berücksichtigt. Ein Ausschnitt der im Anhang befindlichen Genotypentabelle „200224_PCS19001_Sauerkirsche_GT_nurHomUndHet“ ist in Abbildung 5 dargestellt.

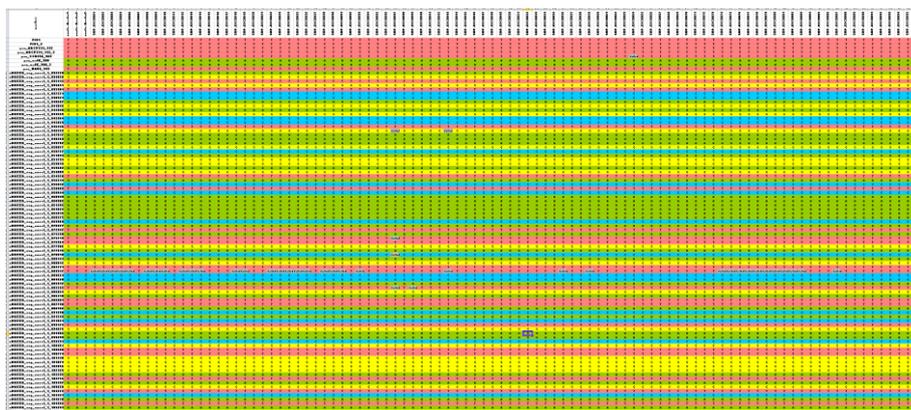


Abbildung 5: Ausschnitt der Genotypentabelle für die Sauerkirschenproben ausgewertet für nur eindeutige Marker.

Da für die Verwandtschaftsanalysen nur polymorphe Marker einen Informationsgehalt besitzen, ist eine Identifikation der Duplikate an der geringen Anzahl dieser polymorphen und eindeutig auswertbaren Marker gescheitert. Auch weitere Versuche wie die Einbeziehung von Markern, die durch Anwesenheit und Abwesenheit polymorph erscheinen, haben keine weiteren Erkenntnisse geliefert. Die geringe

Anzahl der eindeutig auswertbaren und polymorphen Marker lässt damit keine Verwandtschaftsanalyse zu.

Um dennoch eine Verwandtschaftsanalyse und Duplikatidentifizierung durchführen zu können, wurde ein zweites Clusterfile erstellt, das ausschließlich Marker berücksichtigt, die 4 oder 5 klar separierte Cluster aufweisen und somit mit einigermaßen hoher Wahrscheinlichkeit den richtigen Allelstatus zugewiesen bekommen können (Abbildung 6a-c). Insgesamt konnten damit 6.519 polymorphe Marker gefunden und für die NTSYS Analyse genutzt werden. Die Analyse erfolgte mit polyploiden Modul von GenomStudio. Die Resultate sind für eine Verwandtschaftsanalyse ausreichend, sollten aber nur unter Berücksichtigung der Fehlbarkeit für andere Analysen zu Hilfe genommen werden.

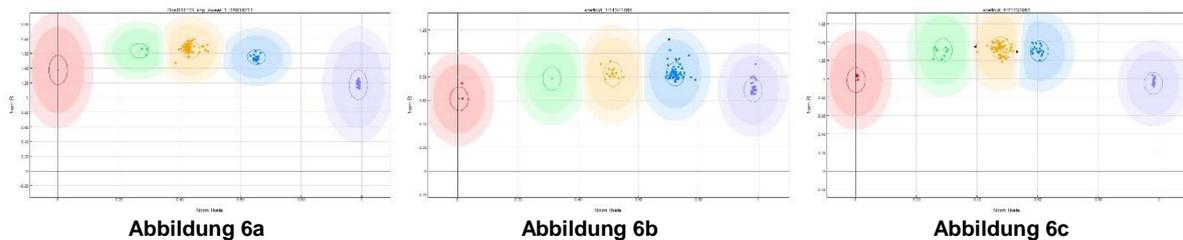


Abbildung 6a:

Beispielbild für einen Marker der Sauerkirsche der mit hoher Wahrscheinlichkeit die Cluster die 4 Cluster AAAB, AABB, ABBB und BBBB besetzt.

Abbildung 6b:

Beispielbild für einen Marker der Sauerkirsche der mit hoher Wahrscheinlichkeit die Cluster die 4 Cluster AAAA, AABB, ABBB und BBBB besetzt.

Abbildung 6c:

Beispielbild für einen Marker der Sauerkirsche der alle 5 Cluster AAAA, AAAB, AABB, ABBB und BBBB besetzt.

Die Genotypen wurden auch hier im IUB Code und mit Bezug zur Quellsequenz ‚forward‘ ausgegeben (allerdings im tetraploiden Modus). Die ausführliche Genotypentabelle der Sauerkirschenproben ist im Anhang unter „200415_PCS19001_sour_polyploid“ dargestellt. Die Verwandtschaftsanalyse wurde auch hier mit Hilfe der Software NTSYS durchgeführt und hat zuverlässig sowohl die technischen als auch die biologischen Replike identifiziert. Außerdem konnten etliche weitere Duplikate im Material gefunden und entsprechend der Vorgaben in der Genotypentabelle kenntlich gemacht werden. Nach Erstellung des Stammbaumes, konnten sogar Multiplike von bis zu 23 Sauerkirschproben identifiziert werden. Das Ergebnis ist eine Reduktion der 108 verschiedenen Genotypen auf nur 58. Abbildung 7 zeigt als Beispiel einen Ausschnitt der Genotypentabelle mit Kennzeichnung der entsprechenden Duplikate. Dieses Clusterfile steht dem Kunden nun für alle weiteren Analysen mit diesem Array zur Verfügung.

Abbildung 7: Ausschnitt der Genotypentabelle Sauerkirsche mit Kennzeichnung der entsprechenden Duplikate.

6. Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Insgesamt konnten die Arbeiten erfolgreich durchgeführt werden. Für die Süßkirsche wurden qualitativ sehr hochwertige und reproduzierbare Ergebnisse erzielt und die Duplikate zuverlässig identifiziert. Für die Sauerkirsche konnte der Array mit eingeschränkter Zuverlässigkeit bezüglich der allelischen Zuordnung verwendet werden. Trotzdem konnten vermutlich auch hier die Duplikate identifiziert werden. Den voraussichtlichen Nutzen und die genaue Verwertbarkeit der Ergebnisse kann nur vom Auftraggeber beurteilt werden, da die genauen Eigenschaften des zu untersuchenden Materials TG unbekannt sind.

7. Zusammenfassung

Ziel des Projekts war die molekulargenetische Charakterisierung von 296 Süßkirschen- und 66 Sauerkirschengenotypen der Deutschen Genbank Obst (DGO) sowie von jeweils 50 Vergleichssorten aus dem internationalen Kirschensortiment mittels des Infinium RosBREED Cherry_V2 Genotyping Arrays. Die Süß- und Sauerkirschengenotypdaten sollten mit Hilfe eines jeweils für die entsprechenden Ploidiestatus optimierten Clusterfiles generiert werden.

Der Array war für die Charakterisierung der insgesamt 352 Süßkirschenproben gut nutzbar. Zwar konnten nur 3.855 der 13.559 Marker als polymorph identifiziert werden. Trotzdem konnte mit Hilfe der Genotypendaten der polymorphen Marker eine NTSYS–Verwandtschaftsanalyse durchgeführt werden. Alle biologischen und technischen Kontrollen wiesen einen eindeutigen Genotyp auf und zeigen wie stabil und reproduzierbar die arraybasierte Genotypisierung ist. Nach Erstellung des Stammbaumes konnten etliche Duplikate, Triplikate bis hin zu Quadruplikaten identifiziert werden. Das Ergebnis führte zu einer Reduktion der 353 untersuchten Genotypen auf nur noch 303.

Im Rahmen dieser Untersuchungen ist eine eingeschränkte Nutzungsmöglichkeit des RosBREED Cherry_V2 Arrays für die Zuordnung der Allelzustände in der allotetraploide Sauerkirsche aufgefallen. Unter Berücksichtigung aller polymorphen Marker die 4 bzw. 5 gut separierte Cluster aufweisen, konnten trotzdem 6.519 Marker für die Verwandtschaftsanalyse genutzt werden. Nach Erstellung des Stammbaumes, konnten etliche Duplikate, Triplikate bis hin zu Multiplikaten von 23 Sauerkirschenproben identifiziert werden. Das Ergebnis führte zu einer Reduktion der 108 untersuchten Genotypen, einschließlich Kontrollen, auf nur 58 unterschiedliche Akzessionen.

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

Ziel des Projekts war die molekulargenetische Charakterisierung von 296 Süßkirschen- und 66 Sauerkirschengenotypen der Deutschen Genbank Obst (DGO) sowie von jeweils 50 Vergleichssorten aus dem internationalen Kirschensortiment und Identifizierung möglicher Duplikate. Erreicht wurden die molekulargenetische Charakterisierung der Süßkirschenproben und die Identifizierung der Duplikate im Züchtungsmaterial der Süß- und Sauerkirschen. Das Ziel der Identifizierung der Duplikate im Sauerkirschenmaterial inklusive eindeutiger Allelzuordnung konnte nur eingeschränkt erreicht werden. Dadurch ergibt sich als weiterführende Fragestellung die Suche nach besser geeigneten (insbesondere genom-spezifischen) Markern für die Sauerkirschenproben. Dies kann zu einer Neuentwicklung eines Arrays mit ausschließlich genom-spezifischen Markern führen.

9. Literaturverzeichnis

- a) Peace, C; Bassil, N; Main, D; Ficklin, S; Rosyara, U R; Stegmeir, T; Sebolt, A; Gilmore, B; Lawley, C; Mockler, T C; Bryant, D W; Wilhelm, L; Lezzoni A (2012). Development and Evaluation of a Genome-Wide 6K SNP Array for Diploid Sweet Cherry and Tetraploid Sour Cherry. PLOS ONE; December 2012; Volume 7; Issue 12
- b) GenomeStudio™ Genotyping Module v1.0 User Guide
- c) Infinium® HD Assay Ultra Protocol Guide