

*Auftragnehmer:*

Microsynth ecogenics GmbH,  
Schützenstrasse 15  
9436 Balgach  
Schweiz

*Auftraggeber:*



## **Abschlussbericht**

*Projektnummer:*

**2821BE002**

*Thema:*

**Detaillierte Untersuchungen der Sauerkirschensorten der Deutschen Genbank Obst mittels Genotyping-by-Sequencing (GBS)**

*Laufzeit:*

04.11.2021 – 24.03.2023

*Berichtszeitraum:*

04.11.2021 – 21.03.2023

Balgach, 21. März 2023

### **1. ZIEL UND AUFGABENSTELLUNG DES PROJEKTES**

---

Aufgrund von internationalen Verpflichtungen koordiniert der Bund die deutschlandweite systematische Erhaltung, Charakterisierung und Evaluierung des Nationalen Inventars obstgenetischer Ressourcen. Die Deutsche Genbank Obst (DGO), ein dezentrales nationales Netzwerk in Deutschland, ist mandatiert mit der Sammlung und Erhaltung von genetischen Ressourcen bei Obst in wissenschaftlicher, langfristig abgesicherter, nachhaltiger sowie kosteneffizienter Art und Weise. Die Koordinierungsstelle der DGO ist das Institut für Züchtungsforschung an Obst in Dresden-Pillnitz (ZO) des Julius Kühn-Instituts, Bundesforschungsinstitut für Kulturpflanzen (JKI).

Der Erhalt und die nachhaltige Nutzung der pflanzengenetischen Ressourcen sind gleichrangige Ziele des Internationalen Saatgutvertrages. Bisherige genetische Analysen mit Mikrosatelliten und einem SNP Array für die Kirsche, konnten die Sauerkirschen Akzessionen nicht sauber unterscheiden. Aus diesem Grund wurde nach einer alternativen Analyseverfahren gesucht.

Genotyping-by-Sequencing (GBS) ist ein Protokoll, bei welcher eine Genomreduktion erfolgt. Dieses «reduzierte» Genom wird anschliessend sequenziert und auf single nucleotide polymorphisms (SNPs) analysiert. Die Annahme ist dabei, dass mithilfe dieser Technologie auch die Sauerkirschensorten der Deutschen Genbank Obst und damit die Akzessionen genetisch unterschieden werden können. Allerdings ist die allotetraploid genomische Struktur der Sauerkirsche eine Herausforderung bei der Analyse der GBS-Daten und angepasste Analyse-Strategien müssen verwendet werden.

### **2. PLANUNG UND ABLAUF DES PROJEKTES**

---

Pflanzenmaterial von 150 Akzessionen wurde durch das Julius Kühn-Instituts zur Verfügung gestellt. Alle weiteren Arbeiten wie DNA Isolation, Library Herstellung, Sequenzierung und die bioinformatischen Analysen wurden bei Microsynth Ecogenics durchgeführt.

Die GBS Analysen sind Standard-Analysen, welche durch Microsynth Ecogenics angeboten werden. Einzige Besonderheit im vorliegenden Projekt war die hohe Sequenziertiefe, welche angestrebt wurde (>4.0 Mio Reads pro Probe), was etwa der 2- bis 4-fachen Sequenziertiefe bei Standard GBS Projekten entspricht.

Die eigentliche Herausforderung im Projekt war aber eine adäquate Analysestrategie der Sequenzdaten zu definieren. Hier lagen auch die grössten Unsicherheiten, in wieweit die generierten Daten zu verwertbaren Resultaten führen würden und wie die Forschungsfrage beantwortet werden kann.

### **3. WISSENSCHAFTLICHER UND TECHNISCHER STAND**

---

Als Grundlage für den systematischen Erhalt der obstgenetischen Ressourcen wurden bisher pomologische und molekulargenetische Sortenechtheitsbestimmungen durchgeführt. Die für die Sortenechtheitsbestimmung benutzten, international abgestimmten molekularen Markersets für Mikrosatellitenmarker enthalten jedoch i.d.R. keine Merkmalsinformationen, sie ermöglichen keine weiterführende Verwandtschaftsanalyse und erlauben nur eine sehr eingeschränkte Abbildung der

allelischen Diversität als Grundlage nachhaltiger Nutzungsmöglichkeiten dieser Genotypen. Aus diesem Grund wurden u.a. bei der Sauerkirsche erste SNP-Analysen mittels eines SNP-Arrays durchgeführt. Aus dem Ergebnisbericht geht jedoch für die Sauerkirsche hervor, dass mit dem im Rahmen dieser Untersuchungen eingesetzten Cherry\_V2 Array nur eine eingeschränkte Zuordnung der Allelzustände möglich ist. Die Sauerkirsche ist allotetraploid und hat  $2n=4x=32$  Chromosomen und ist aus der Hybridisierung von *Prunus avium* x *Prunus fruticosa* entstanden [1]. Viele Markerallele kommen auf beiden Subgenomen des Sauerkirschgenoms vor. Eine eindeutige Zuordnung welches Allel welchen Genort repräsentiert ist dadurch nicht möglich. Aufgrund dieses Mangels konnte das Ziel der Identifizierung von Duplikaten im Sauerkirschmaterial nur eingeschränkt erreicht werden. Dadurch ergibt sich als weiterführende Fragestellung die Suche nach besser geeigneten (insbesondere subgenomspezifischen) Markern für die Sauerkirschenproben.

Im Falle eines Genotyping-by-Sequencing (GBS), wird das Genom mittels einem Restriktionsverdau und anschließender Längenselektion der Fragmente reduziert und damit nur ein Teil des Gesamtgenoms pro Probe sequenziert. Dadurch kann eine hohe Sequenzierentiefe für diejenigen Stellen/Loci erreicht werden, welche in der reduzierten Library übrigbleiben. Zudem gilt allgemein bei Sequenzdaten, dass man wesentlich mehr Möglichkeiten hat zu eruieren, wo sich die entsprechenden Sequenzen im Genom befinden, da es für die Sauerkirsche ein Referenzgenom gibt. Daher wurde davon ausgegangen, dass man die gefunden SNPs den Subgenomen zuordnen kann, was ein wesentlicher Vorteil der Methode sein dürfte im Vergleich zu den bisher verwendeten Markern.

## **4. MATERIAL UND METHODEN**

---

### **4.1. DNA Isolation**

Das Blattmaterial jeder Probe wurde in ein 50 mL Falcon-Tube transferiert und während 60 Stunden bei 65 °C getrocknet. Anschliessend wurden die getrockneten Blätter in einer Schüttelmühle während 1 h homogenisiert (350 rpm, 2 x 10 mm Glasbeads).

30 mg der homogenisierten Probe wurden in die DNA-Extraktion eingesetzt und die Isolation erfolgte gemäss dem NucleoSpin 96 Plant II Kit Protokoll (Machery Nagel) unter Verwendung des Lysepuffers PL1 (CTAB-basiert). Die Proben wurden über Nacht lysiert und die folgenden Schritte wurden gemäss dem Protokoll des Herstellers durchgeführt. Die DNA-Isolationen wurden anschliessend mittels PicoGreen® quantifiziert.

### **4.2. Library Herstellung und Sequenzierung**

Für das Genotyping-by-Sequencing wurde ein Protokoll analog dem ddRAD-Protokoll [2] verwendet.

Die DNA-Konzentration der Proben wurde auf 16.6 ng/µL normalisiert. Proben, welche eine tiefere Konzentration hatten, wurden nicht weiter verdünnt und direkt in die Library Preparation eingesetzt. 12 µL der Proben-DNA wurde in einem ersten Schritt mittels der Restriktionsenzyme PstI und MspI verdaut. Passende Adaptoren, welche kompatibel mit dem Illumina NovaSeq unique dual indexing (UDI) Protokoll sind, wurden an die geschnittenen DNA-Fragmente ligiert. Im Anschluss wurden UDI-

Adaptoren mittels PCR an die Libraries angefügt. Die Konzentration der finalen Libraries wurde mit PicoGreen® quantifiziert und die Libraries equimolar gepoolt. Vom aufgereinigten Pool der Libraries wurde eine Size-Selection auf dem PippinPrep Gerät durchgeführt (300-400 bp). Der Pool der grössen-selektierten Libraries wurde dann auf der Illumina NovaSeq Plattform mit 2 x 100 bp Sequenzlänge sequenziert.

#### 4.3. Bioinformatische Analysen (STACKS)

Die Sequenzdaten wurden mit der Software bcl2fastq v2.20.0.422 demultiplexed (0 mismatches) und die Adptoren getrimmt. Die Rohsequenzen wurden weiters einer qualitativen Filterung unterzogen mittels BBTools v38.96 (<https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>) und alle Sequenzen mit Q-value <24 und Reads mit N Basen wurden verworfen. Die qualitativ guten Rohsequenzen wurden anschliessend gegen die Referenzsequenz der Sauerkirsche kartiert mit der Software bwa mem v0.7.17 (default settings; <https://bio-bwa.sourceforge.net/bwa.shtml>). Die Genomsequenz wurde durch das JKI zur Verfügung gestellt. Basierend auf den Mapping-Resultaten wurden 2 Teildatensätze sortiert, welche (i) nur einmalig abgegliche Sequenzen enthalten und (ii) entweder mit dem Fruticosa- oder aber dem Avium-Subgenom abgeglichen wurden. Dazu wurde die Software samtools v.1.15.1 (<http://www.htslib.org/>) verwendet. Alle Reads, welche mehrfach auf der Referenzsequenz abgeglichen wurden, sind für die anschliessenden Analysen verworfen worden.

Im Anschluss wurden die Sequenzen am Start und Ende um 10 bp gekürzt. Dieser Schritt soll sicherstellen, dass keine «falschen» SNPs detektiert werden, welche sich an den Rändern durch die Library Herstellung resp. das Trimmen der Adaptoren ergeben könnten. Im Anschluss wurden die so getrennten Reads gegen ihre eigene Referenz (nur Chromosomen der eigenen Referenz) mittels bowtie2 v2.4.5 (<https://bowtie-bio.sourceforge.net/bowtie2/manual.shtml>; very-sensitive pre-settings) gemappt. Die Referenz-basierte GBS Analyse wurde mit der Software STACKS v2.62 [3] für beide Subgenome durchgeführt. Bei dieser Analysestrategie werden die Positionen der detektierten SNPs auf dem Referenzgenom ermittelt.

Die Hauptparameter bei der STACKS Analyse wurden wie in Tabelle 1 gewählt. Pro Locus wurde nur ein SNP berücksichtigt (flag: Process one single snp per locus).

Tabelle 1: Parametrisierung für die STACKS Analysen

Parameter	Value
Minimum read depth required to create a stack	3
Maximum number of mismatches allowed for loci	2
Maximum gaps for alignment	0
Minimum fraction of samples per population to include a locus	0.09
P-value cutoff	0.05
Process one single snp per locus	Yes

#### 4.4. Weitergehende Analysen

Die resultierende SNP-Tabelle für jedes Subgenom wurde gefiltert, um SNPs mit >20% missing data auszuschliessen und (ii) nur SNPs zu berücksichtigen, bei welchen der SNP in mindestens 12 Proben gefunden wurde (Ausschluss von SNPs mit kleiner Frequenz). Durch den Ausschluss von weniger häufig auftretenden SNPs (low-frequency) wurden durch Sequenzierfehler entstandene SNP aus der Analyse ausgeschlossen.

Die so gefilterten SNPs wurden neu codiert (A=100, T=110, G=120; C=130, Missing=0), so dass die paarweisen Distanzen zwischen den Proben gemäss untenstehender Formel mit einem Python-Skript ermittelt werden konnten. Loci mit missing Data werden von der Berechnung der paarweisen Distanz ausgeschlossen.

$$Distance = 1 - \frac{2 * \sum Shared Alleles}{\sum Allele Sample1 + \sum Allele Sample2}$$

Die berechnete Similarity, welche von 1 subtrahiert wird, entspricht dem Sørensen Index [4]. Die Verteilung der paarweisen Distanzen und eine Clusteranalyse wurde mit der Statistiksoftware R durchgeführt. Zudem wurden Gengruppen basierend auf den Clusteranalysen definiert. Als Cut-Offs wurde beim Fruticosa-Datensatz 0.03 (3%), 0.05 (5%) und 0.10 (10%) gewählt. Im Falle des Avium-Datensatzes wurden als cut-offs 0.05 (5%), 0.10 (10%) und 0.15 (15%) gewählt. Diese Cut-Off Werte wurden anhand der Verteilung der paarweisen Distanzen im vorhanden Datensatz abgeschätzt.

## 5. RESULTATE

---

### 5.1. DNA Isolation, Library Herstellung und Sequenzierung

Die gemessenen Konzentrationen der DNA-Isolation zeigte eine große Streuung und reichte von 4.1 bis 88.0 ng/uL. Bei 74 Proben war die DNA-Konzentration >16.6 ng/uL und diese Proben wurden für die Herstellung der Libraries verdünnt. Alle anderen Proben wurden unverdünnt bei der Herstellung der Libraries eingesetzt.

Alle bis auf 4 Proben resultierten in >4.0 Mio R1- Sequenzen (forward Reads) in der Sequenzierung. Die 4 Proben, welche ausgefallen sind, haben entweder bei der Library Herstellung nicht funktioniert (SNP-KIC0026, SNP-KIC0077 und SNP-KIC0096) oder im Falle der Probe SNP-KIC0020 sehr wenige Sequenzdaten generiert. Dementsprechend wurden diese Proben von den Analysen ausgeschlossen. Die Ausfallrate bei der Library Herstellung und Sequenzierung beträgt damit 2.7% aller Proben.

### 5.2. Qualitatives Filtern und Subgenom-Abgleich

Abhängig von der Probe erfüllten 75% bis 92% der Rohsequenzen das qualitative Filtern und 7.5 Mio – 42.7 Mio Sequenzen (R1 & R2) wurden in den Abgleichgen die Subgenome eingeschlossen. Von

diesen Sequenzen konnten rund 80% eindeutig auf eines der Subgenome zugeordnet werden. Die restlichen Sequenzen wurden verworfen. Von allen eindeutig auf ein Subgenom kartierten Sequenzen entfielen im Durchschnitt 59% auf das Avium-Subgenom und 41% auf das Fruticosa-Subgenom.

### 5.3. STACKS-Analysen und Auswahl der SNPs

Mit der gewählten Parametrisierung für die STACKS Analysen wurden 135'125 SNPs auf dem Avium-Subgenom und 105'494 SNPs auf dem Fruticosa-Subgenom detektiert. Wie bei GBS-Analysen üblich, sind die meisten detektierten SNPs nur für einen kleinen Teil der Proben detektiert worden und alle anderen Proben weisen keine Daten für den entsprechenden Locus auf. Daher wurden die SNPs weiter gefiltert, damit nur gute Kandidaten-SNPs in die weiteren Analysen eingeschlossen wurden (siehe auch Kapitel 4 zu den Kriterien). Dies resultierte in einem finalen Set von **1'189 SNPs** auf dem Avium-Subgenom und **630 SNPs** auf dem Fruticosa-Subgenom. Die Verteilung der SNPs auf den verschiedenen Genomen der beiden Subgenome wird in Tabelle 2 gelistet.

Tabelle 2: Verteilung der SNPs auf den 16 Chromosomen der beiden Subgenome, welche in die weitere Analyse eingeschlossen wurden.

<i>Accession</i>	<i>Chromosome size</i>	<i>Number of SNPs in dataset</i>
PCE_Avium_Chro1	52824182	287
PCE_Avium_Chro2	38212775	142
PCE_Avium_Chro3	30675746	134
PCE_Avium_Chro4	28226313	104
PCE_Avium_Chro5	23885644	93
PCE_Avium_Chro6	35906979	170
PCE_Avium_Chro7	31537404	113
PCE_Avium_Chro8	27735405	146
PCE_Fruticosa_Chro1	53495658	178
PCE_Fruticosa_Chro2	43438225	76
PCE_Fruticosa_Chro3	34378064	51
PCE_Fruticosa_Chro4	32374258	59
PCE_Fruticosa_Chro5	22600031	45
PCE_Fruticosa_Chro6	39437034	72
PCE_Fruticosa_Chro7	41603557	62
PCE_Fruticosa_Chro8	32145936	87

### 5.4. Definition von Gengruppen

Die Verteilungen der paarweisen Distanzen zwischen den Proben sind in Abbildung 1 und Abbildung 2 dargestellt. In beiden Fällen sieht man eine zweigipflige Verteilung mit einem klaren Hauptpeak und einem zweiten kleineren Peak im Bereich kleiner paarweiser Unterschiede. Der Grund für diese zweigipflige Verteilung liegt darin, dass es Gruppen von Proben gibt, welche untereinander ähnlich sind, aber klar von allen anderen Proben abgegrenzt sind.

Im Gegensatz zu anderen Obstbäumen (wie z.B. Apfel), bei welchen es einfacher ist eine Limite für die Definition von Gengruppen zu setzen, da es eine klare Abgrenzung zwischen Eigenvergleichen

und Fremdvergleichen ersichtlich ist, ist dies bei der Sauerkirsche schwieriger. Aus diesem Grund wurden für die für die weiteren Analysen unterschiedliche Limiten gewählt.

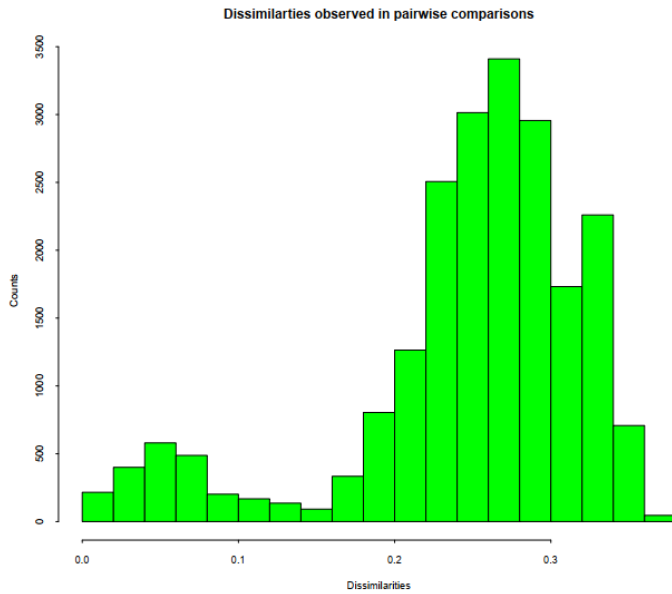


Abbildung 1: Verteilung der Distanzen aus den paarweisen Vergleichen des SNP-Datensatzes für das Avium-Subgenom (inkl. Selbstvergleich)

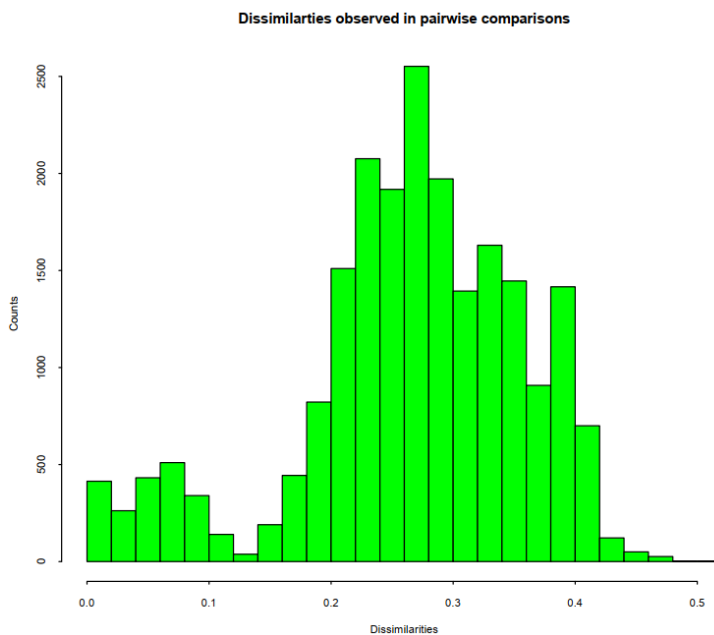


Abbildung 2: Verteilung der Distanzen aus den paarweisen Vergleichen des SNP-Datensatzes für das Fruticosa-Subgenom (inkl. Selbstvergleich)

Die Anzahl gefundener Gruppen für die beiden Subgenome und für die kombinierten Gengruppen sind in Tabelle 3 zusammengefasst. Dabei sind die gefundenen Gengruppen zwischen den beiden Subgenomen erstaunlich gleich verteilt, was die Anzahl der Gengruppen für ein bestimmtes Limit betrifft, wie auch die Proben, welche darin gruppiert werden. Wenn die Gengruppen beider Subgenome kombiniert werden, so findet man nur wenige Gengruppen mehr, was auf eine gemeinsame Historie der beiden Subgenome hinweisen könnte und auf praktisch keine Rekombination hinweist. So werden für beide Subgenome bei einer Limite von 5% für beide Subgenome total 86 Gengruppen gefunden.

Tabelle 3: Anzahl Gengruppen pro Subgenom für die verschiedenen eingesetzten Limiten und die Kombination davon.

Subgenom	Limite	Anzahl Gengruppen
Fruticosa	3%	107
	5%	79
	10%	53
Avium	5%	79
	10%	58
	15%	48
Combined	5% + 10%	81
	5% + 5%	86

Die Anzahl Proben in den Gengruppen ist in Tabelle 4 dargestellt. Während 65 Proben eine eigene Gengruppe bildeten, wurden mehrere Gengruppen erkannt, welche 2 oder mehr Proben einschlossen. Die grösste Gengruppe enthielt 17 Proben.

Tabelle 4: Verteilung der Anzahl Proben pro Gengruppe.

Anzahl Proben in Gengruppe	Anzahl Gengruppen	Totale Anzahl Proben
1	65	65
2	12	24
3	2	6
4	3	12
5	1	5
8	1	8
9	1	9
17	1	17



Für die 86 definierten Gruppen im Datensatz 5% + 5% wurde jeweils pro Gruppe ein repräsentativer Genotyp mit möglichst kleinem Anteil missing data ausgewählt und die paarweisen Differenzen über alle SNPs gerechnet und ein Dendrogramm konstruiert. Die Resultate für die Verteilung der paarweisen Unterschiede und das Dendrogramm sind in Abbildung 3 und Abbildung 4 abgebildet.

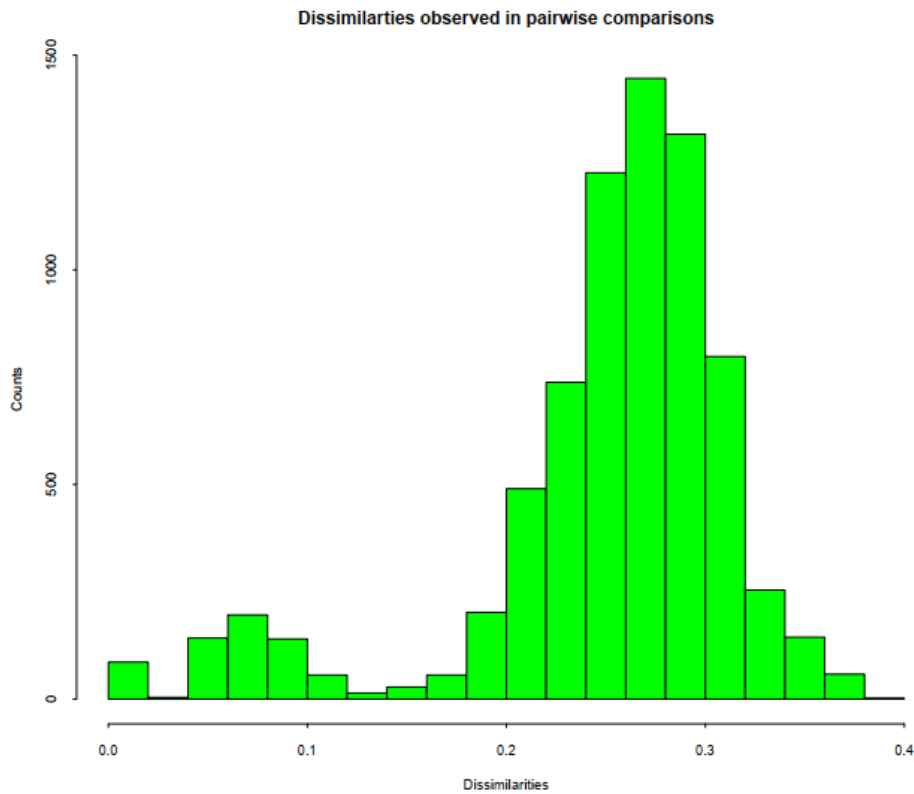


Abbildung 3: Verteilung der Distanzen aus den paarweisen Vergleichen des kombinierten Datensatzes, mit einer Referenzprobe pro Gengruppe (inkl. Selbstvergleich)

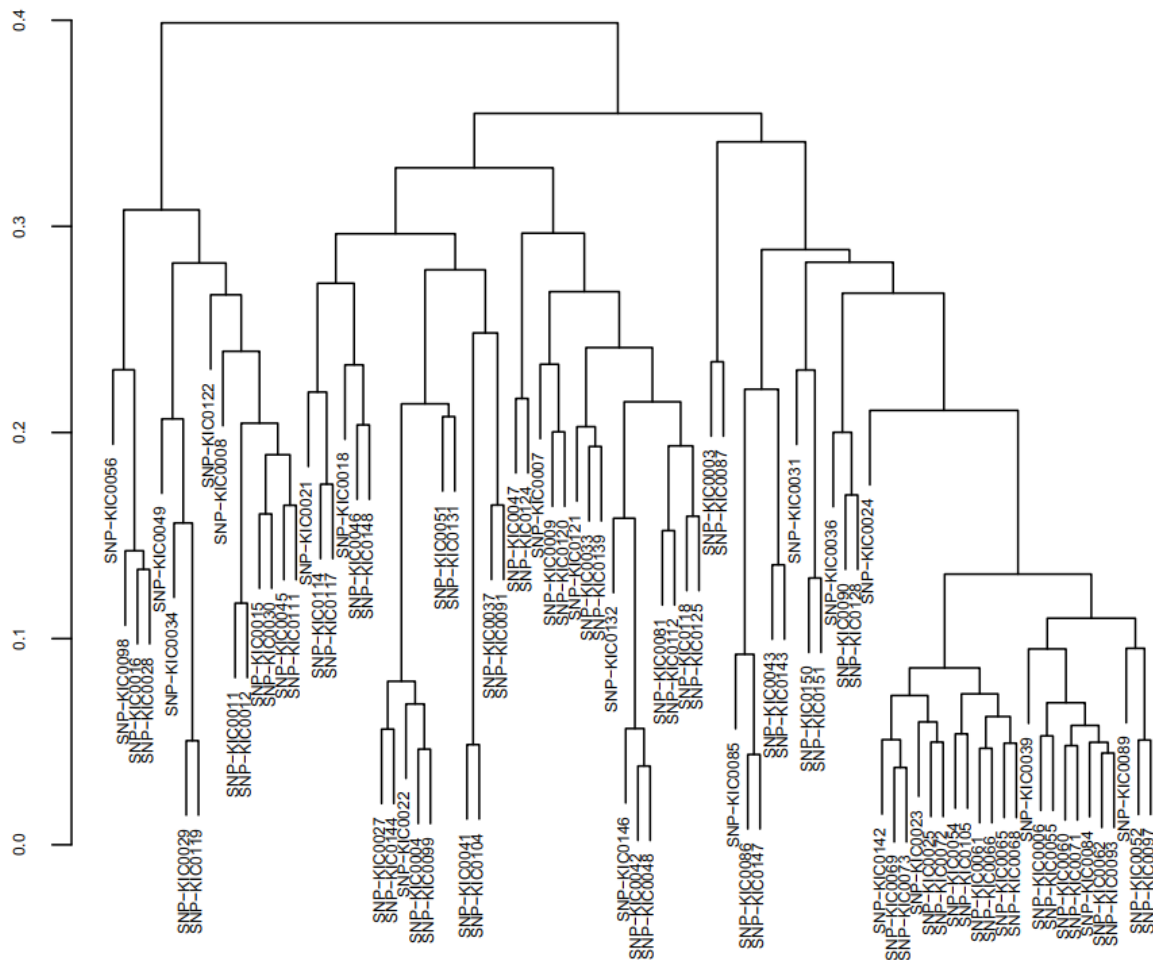


Abbildung 4: Dendrogramm basierend auf den paarweisen Unterschieden für die selektierten Referenzen basierend auf der 5%+5% Gengruppen-Definition.

### 5.5. Vergleich Gengruppen und Sortennamen

Wenn die Sortennamen mit der Einteilung nach Gengruppen verglichen werden, so findet man einige Abweichungen. Ein Beispiel sind die Sorten `Jade` und `Achat`, welche jeweils durch 2 Proben repräsentiert werden und auch 2 Gengruppen bilden. Allerdings werden die Sorten je nach Herkunft unterschiedlich den Gengruppen zugewiesen, was vermutlich auf eine Verwechslung bei der Probenentnahme schliessen lässt (Tabelle 5). Dies muss aber durch das JKI noch genauer abgeklärt werden.

Tabelle 5: Zuweisung der 4 Proben der Sorten `Achat` und `Jade` zu den Gengruppen.

Proben Nr.	Herkunft / Sammlung	Baum Nr.	Sortenname	GenGroup_Combined [5% + 5%]
SNP-KIC0001	BSA (Marquardt)		Achat	1_1
SNP-KIC0002	BSA (Marquardt)		Jade	2_2
SNP-KIC0118	JKI_CG05	KIC0156	Achat	2_2
SNP-KIC0125	JKI_CG05	KIC0155	Jade	1_1

Andererseits gibt es Gengruppen, welche verschiedene Sortennamen enthalten (Tabelle 6), sofern man die Gengruppen 5% + 5% verwendet. Durch eine stringenter Definition der Gengruppe können in vielen Fällen die Gruppen noch weiter unterteilt werden.

Tabelle 6: Gengruppe 15\_15, welche alle Proben mit dem Sortennamen Schattenmorelle enthält. Neben den Schattenmorellen finden sich noch weitere Sorten in derselben Gengruppe.

Proben Nr.	Herkunft / Sammlung	Baum Nr.	Sortenname	GenGroup_Combined [5% + 5%]
SNP-KIC0017	JKI_CG05	KIC0139	Pc 59	15_15
SNP-KIC0035	JKI_CG05	KIC0061	Schattenmorelle	15_15
SNP-KIC0050	Kyffhäuser_		Pc 64	15_15
SNP-KIC0130	JKI_CG05	KIC0098	Nabella	15_15
SNP-KIC0133	JKI_CG05	KIC0058	Schattenmorelle Typ Schamo	15_15
SNP-KIC0137	JKI_CG05	KIC0104	Vowi	15_15
SNP-KIC0140	JKI_CG05	KIC0003	Cerella	15_15
SNP-KIC0141	JKI_CG05	KIC0060	Schattenmorelle Typ Scharö	15_15
SNP-KIC0143	BSA_Schlag 7	DEU610.054.SO333	Schattenmorelle Typ Boscha	15_15

Es gibt aber auch Gengruppen, welche gut mit den Sortennamen übereinstimmen, und z.T. auch weiter mit den definierten Gengruppen unterschieden werden können. Das Beispiel der Sorte `Stevnsbaer` ist in Tabelle 7 dargestellt.

Tabelle 7: Zuweisung der 4 Proben der Sortengruppe `Stevnsbaer` zu 2 Gengruppen.

Proben Nr.	Herkunft / Sammlung	Baum Nr.	Sortenname	GenGroup_Combined [5% + 5%]
SNP-KIC0004	BSA (Marquardt)		(Stevnsbaer)	4_4
SNP-KIC0038	JKI_CG05	KIC0066	Stevnsbaer	4_4
SNP-KIC0135	JKI_CG05	KIC0151	Stevnsbaer Birgitte	4_4
SNP-KIC0144	BSA_Schlag 7	DEU610.SO.279	Stevnsbaer Viki	75_76

## 6. NUTZEN DER RESULTATE UND VERWERTBARKEIT

---

Die Struktur der GBS-Daten mit der zweigipfligen Verteilung der paarweisen Unterschiede entspricht den Erwartungen, welche man basierend auf der Geschichte der Sauerkirsche erwarten dürfte. Es wird davon ausgegangen, dass die Sauerkirsche durch wenige unabhängige Hybridisierungen entstanden ist. In den GBS-Daten müssten Vergleiche zwischen solch unterschiedlichen Hybridisierungen zu deutlichen Differenzen führen. Durch weitere Kreuzungen und spontane Mutationen sind dann innerhalb der Hybridisierungsgruppen verschiedenen Sorten entstanden, welche sich zum Teil nur durch geringe Unterschiede auszeichnen. Dies dürfte zur zweigipfligen Verteilung, wie in der Analyse beobachtet, geführt haben. Dabei zeigten beide Subgenome eine ähnliche Verteilung. Auch die Anzahl Gengruppen je nach Definition der Limite ist sehr ähnlich.

Die vorliegenden Daten/Analysen müssen nun mit den pomologischen Daten abgeglichen werden. Basierend auf diesen weiteren Abgleichen lassen sich dann Kriterien definieren, mit welchen die eindeutige Gengruppen definiert werden können.

Erst nach diesen weiteren Arbeiten kann der definitive Nutzen des GBS Datensatzes eingeschätzt werden.

## 7. ZUSAMMENFASSUNG

---

Der Erhalt und die nachhaltige Nutzung der pflanzengenetischen Ressourcen sind gleichrangige Ziele des Internationalen Saatgutvertrages. In bisherigen genetischen Analysen mit Mikrosatelliten bzw. mit einem SNP-Array für die Kirsche konnten die Sauerkirschen-Akzessionen nicht eindeutig unterschieden werden. Im vorliegenden Projekt wurden daher 150 Sauerkirschen-Akzessionen mit einem Gentypping-by-Sequencing-Ansatz untersucht. Die Resultate der Untersuchung zeigen, dass basierend auf den SNP Daten, welche aus der Analyse gewonnen wurden, genetische Gruppen identifiziert werden konnten. Dabei konnten die SNPs auf den beiden Subgenomen der Sauerkirsche verankert werden. Für eine routinemässige Überprüfung von Sauerkirschen-Proben eignet sich der GBS-Ansatz aber allein noch nicht.

## 8. ZIELERREICHUNG

---

Alle bisherigen genetischen Analysen von Sauerkirschen-Sorten resultierten in nicht brauchbaren Resultaten. Technisch hat die GBS-Analyse der Sauerkirschen für 146 der 150 Proben zu auswertbaren Resultaten geführt und ein kombinierter Datensatz von 1'890 SNPs konnte in die Definition von Gengruppen eingeschlossen werden. Alle 1'890 SNPs sind über die Genomkoordinaten auf den beiden Subgenomen verankert und sind damit subgenom-spezifisch.

Der vorliegende Datensatz erkennt 86 Gengruppen, wenn die Limite für die Definition der Gengruppen bei einer Distanz von 5% angesetzt wurde. Damit dürfte der GBS-Datensatz deutlich besser für die Analyse von Sauerkirschen-Sorten geeignet sein als alle bisher vorhandenen Datensätze. In diesem Sinne wurden die wesentlichen Ziele der Studie erreicht.

Allerdings müssen noch weitere Arbeiten, welche die pomologischen Informationen mit den genetischen Daten abgleichen, zeigen, inwieweit die GBS-Daten bei der Bestimmung der Sorten helfen können.

### **9. LITERATURVERZEICHNIS**

---

1. Bird KA, Jacobs MK, Sebolt A, Rhoades K. Parental origins of the cultivated tetraploid sour cherry (*Prunus cerasus* L.). *Wiley Online Library*. 2022;4: 444–450.
2. Elshire RJ, Glaubitz JC, Sun Q, Poland JA. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*. 2011;6: e19379.
3. Catchen JM, Amores A, Hohenlohe P, Cresko W. Stacks: Building and genotyping loci de novo from short-read sequences. *G3 Genes Genom. Genet.* 1, 171–182. *G3 Genes, Genomes, Genetics*. 2011;1: 171–182.
4. Bady P, Diserens AC, Castella V, Kalt S. DNA fingerprinting of glioma cell lines and considerations on similarity measurements. *Neuro-Oncology*. 2012;14: 701–711.