

Abschlussbericht

**Entwicklung genetischer Marker zur Gattungs- und
Arterkennung in Holzverbundprodukten mittels *Next
Generation DNA-Barcoding***

Aktenzeichen DBU: 33949/01-31

Hilke Schröder, Bernd Degen, Birgit Kersten

Thünen-Institut für Forstgenetik
Sieker Landstrasse 2
22927 Großhansdorf

Projektkennblatt
der
Deutschen Bundesstiftung Umwelt



Az	33949/01	Referat	31	Fördersumme	398.700,00 €
----	-----------------	---------	-----------	-------------	---------------------

Antragstitel **Entwicklung genetischer Marker zur Gattungs- und Arterkennung in Holzverbundprodukten mittels Next Generation DNA-Barcoding**

Stichworte Analytik, Biomasse/Nachwachsende Rohstoffe, Ökobilanz, Ressource
Biodiversität, Holz, Möbel

Laufzeit	Projektbeginn	Projektende	Projektphase(n)	
54 Monate	01.04.2018	30.09.2022		
Zwischenberichte	31.03.2019	31.03.2020	31.03.2021	30.03.2022

Bewilligungsempfänger	Thünen-Institut für Forstgenetik (TI-FG)	Tel	+49 4102-696105
	Sieker Landstr. 2	Fax	+49 4102-696200
	22927 Großhansdorf	Projektleitung	
		PD Dr. Birgit Kersten	
		Bearbeiter	
		Dr. Hilke Schröder	
		Dr. Thomas Schott	
		Dr. Katrin Schöning-Stierand	
		Annika Eikhof	
		Stefanie Palczewski	
		Björn Klinkhammer	
		Imke Burau	
		Dr. habil. Bernd Degen	

Kooperationspartner Gesamtverband Deutscher Holzhandel e. V.
10117 Berlin

Zielsetzung und Anlass des Vorhabens

Zur Eindämmung des illegalen Holzeinschlages wird sowohl in den USA (FLEGT, Zertifizierungssysteme, Lacey Act) als auch in Europa (EU-Holzhandelsverordnung) seit einigen Jahren vom Gesetzgeber eine genaue Kennzeichnung von importiertem Holz vorgeschrieben. Dazu gehört u.a. die genaue Deklaration der botanischen Art und der geographischen Herkunft des Holzes. Hierfür sind DNA-basierte Nachweismethoden bestens geeignet. Für eine ganze Reihe an gehandelten Baumarten stand die Entwicklung von geeigneten genetischen Markern jedoch noch aus. Insbesondere in Holzverbundprodukten stellt die Markerentwicklung eine große Herausforderung dar, da hier eine Mischung diverser Gattungen und Arten zur Herstellung verwendet werden kann. Holzanatomische Methoden sind in diesem Fall nur begrenzt erfolgreich. Daher war das Ziel dieses Projektes die Entwicklung genetischer Markersysteme zur gleichzeitigen Erkennung von Laub- und Nadelbaumgattungen, die in Holzverbundprodukten am häufigsten eingesetzt werden. Zusätzlich sollten innerhalb der Gattung *Pinus* die Differenzierung auf Artebene für die wichtigsten gehandelten Arten erfolgen.

Darstellung der Arbeitsschritte und der angewandten Methoden

Zur Erreichung der Zielsetzung wurden die folgenden Arbeitsschritte durchgeführt:

- Optimierung von Protokollen zur DNA-Extraktion aus Holzverbundplatten
- Sequenzierung von relevanten Laub- und Nadelbaumarten mittels „Next Generation Sequencing“, bioinformatische Datenanalysen und Selektion von SNPs und InDels, die eine potentielle Differenzierung auf verschiedenen taxonomischen Ebenen ermöglichen
- Validierung selektierter SNPs und InDels in erweiterten Sets von Individuen
- Umsetzung der DNA-Barcodes in ein genetisches Markersystem und Validierung an Holzverbundprodukten
- Testung der praktischen Nutzbarkeit der Gattungs- und Artmarker bei der unabhängigen Prüfung von Artdeklarationen von Holzverbundprodukten in Zusammenarbeit mit GD Holz

Ergebnisse und Diskussion

Im Projekt wurden - wie ursprünglich geplant - die Nadelbaumgattungen *Pinus*, *Picea*, *Pseudotsuga* und zusätzlich *Abies*, *Tsuga*, *Cedrus* und *Larix* sowie die Laubbaumgattungen *Betula*, *Eucalyptus*, *Populus*, *Fagus*, *Shorea* und auch hier zusätzlich *Quercus/Castanea* und *Fraxinus* bearbeitet. Von Arten dieser Gattungen (außer *Populus*) sowie anderen Gattungen der bearbeiteten Familien und Ordnungen wurde Blatt/Nadel- bzw. Holzmaterial von verschiedenen Quellen beschafft und die jeweilige DNA extrahiert. Es wurden von uns mitentwickelte DNA-Extraktionsprotokolle für Holz validiert, um DNA aus stark verarbeiteten Holzgemischen zu gewinnen. Es konnte bei der Verwendung von Gemischen aus Laub- und Nadelholz gezeigt werden, dass Chloroplastenmarker deutlich besser als mitochondriale Marker zur Identifizierung der enthaltenen Laub- und Nadelhölzer geeignet sind. Die experimentell beobachtete Tendenz, dass Laubbaum-DNA im Gemisch besser in der PCR amplifiziert wird als Nadelbaum-DNA, insbesondere bei der Verwendung mitochondrialer Marker, konnte mit bioinformatischen Ergebnissen zum Vergleich der zellulären Kopienzahlen von nukleären, mitochondrialen und Chloroplastengenomen sowie Genomgrößenvergleichen zwischen einigen Nadel- und Laubbaumarten untermauert werden.

Als Grundlage für die Markerentwicklung wurden sowohl öffentliche Daten als auch neugenerierte NGS-Daten verschiedener Laub- und Nadelbaumarten genutzt. Wichtige Ergebnisse der bioinformatischen Analysen dieser Daten waren die Assemblierung und Annotierung neuer Chloroplasten- und/oder mitochondrialer Referenzgenome, u.a. von *Fagus sylvatica* und *Pinus cembra*. Diese Genome waren u.a. eine Grundlage für die weitere bioinformatische Auswahl von Taxon-spezifischen SNPs und InDels für die genetische Markerentwicklung. Im Projekt konnten für alle im Antrag geplanten Gattungen sowie die beiden Subgattungen von *Pinus*, vier Sektionen, drei Subsektionen und zahlreiche Arten innerhalb von *Pinus* genetische Marker entwickelt und erfolgreich validiert werden. Über die ursprünglichen Gattungen des Interesses hinaus, konnten auch für einige andere Gattungen und weitere Arten Marker entwickelt werden. Zur zusätzlichen Absicherung der Gattungs- und Artidentifizierung auf höheren taxonomischen Ebenen wurden zusätzlicher Marker entwickelt, die spezifisch für diese Ebenen sind und zum Teil bereits in Marker-sets eingeflossen sind. Neben einem validierten Multiplex-Markerset, das eine gleichzeitige Erkennung vieler Gattungen innerhalb der Ordnung der Fagales erlaubt, befinden sich weitere Multiplex-Sets für Ordnungen sowie für Familien und Gattungen innerhalb der Coniferales in der Optimierungsphase. Im gemeinsamen Praxistest mit dem Kooperationspartner GD-Holz konnten die meisten der Deklarationen mit den im Projekt entwickelten genetischen Markertests für die untersuchten Sperrhölzer bestätigt werden.

Publikationen:

- Akhmetzyanov L, Copini P, Sass-Klaassen U, **Schroeder H**, et al. (2020) DNA of centuries-old timber can reveal its origin. *Scientific Reports* 10:20316. doi: 10.1038/s41598-020-77387-2
- Braga WB, Deklerck V, Espinoza E, Groening M, Koch G, Monteiro Pastore TC, Ramanantoandro T, **Schröder H**, et al. (2020) Scientific methods for taxonomic and origin identification of timber [online]. GTTN (Global Timber Tracking Network), 6 p. doi:10.13140/RG.2.2.28416.46087
- Bruegmann T, Fladung M, **Schroeder H** (2022) Flexible DNA isolation procedure for different tree species as a convenient lab routine. *Silvae Genetica* 71: 20-30. doi: 10.2478/sg-2022-0003
- Bünger K** (2021) Validierung und Anwendung molekularer Marker zur Holzarten-Identifizierung in Holzverbundprodukten. Bachelorarbeit (am TI-FG im Rahmen von Holz-DNA-Barcoding; 23.02.2021)
- Degen B**, Blanc-Jolivet C, Bakhtina S, Yanbaev R, Yanbaev Y, Mader M, Nürnberg S, **Schroeder H** (2021) Applying targeted genotyping by sequencing with a new set of nuclear and plastid SNP and Indel loci for *Quercus robur* and *Quercus petraea*. *Conservation Genetics Resources* 13: 345-347. doi: 10.1007/s12686-021-01207-6
- Kersten B**, Rellstab C, **Schroeder H**, Fladung M, Krutovsky K, Gugerli F (2022) The mitochondrial genome sequence of *Abies alba* Mill. reveals a high structural and combinatorial variation. *BMC Genomics* 23:776. doi: 10.1186/s12864-022-08993-9
- Kersten B**, Rellstab C, Gugerli F (2022) *Abies alba* isolate AA_WSL01 mitochondrion, scaffold sequences: NCBI, accession numbers ON378818→ON378828
- Kersten B**, **Schott T**, Mader M (2020) *Fagus sylvatica* isolate FASYL_29_1 mitochondrion, complete genome: NCBI, accession MT446430
- Mader M, **Schroeder H**, **Schott T**, **Schöning-Stierand K**, Leite Montalvão AP, Liesebach H, Liesebach M, Fussi B, **Kersten B** (2020) Mitochondrial genome of *Fagus sylvatica* L. as a source for taxonomic marker development in the Fagales. *Plants* 9(10), 1274. doi: 10.3390/plants9101274
- Mader M, **Kersten B** (2020) Mitochondrial genome of *Fagus sylvatica* L. as a source for taxonomic marker development in the Fagales: NCBI SRA, accession PRJNA648273
- Mader M, Liesebach H, Liesebach M, **Kersten B** (2019) The complete chloroplast genome of *Fagus sylvatica* L. (Fagaceae). *Mitochondrial DNA Part B* 4(1): 1818-19. doi: 10.1080/23802359.2019.1612712
- Schmitz N, Beeckman H, Cabezas JA, Cervera MT, Espinoza E, Fernandez-Golfín J, Gasson P, Hermanson JC, Arteaga MJ, Koch G, Lens F, Martínez-Jarquín S, Paredes-Villanueva K, Pastore TCM, Ramanantoandro T, Schraml R, **Schröder H**, et al. (2019) The Timber Tracking Tool Infogram: overview of wood identification methods' capacity. GTTN, 5 p. doi: 10.13140/RG.2.2.27920.25603

- Schoening-Stierand K, Schroeder H, Degen B, Kersten B** (2019) Identification of tree species in wood composite products by DNA barcoding. *Genome* 62 (6): 431-432. doi: 10.1139/gen-2019-0083
- Schott T, Schroeder H, Schöning-Stierand K, Kersten B** (2019) The complete chloroplast genome sequence of *Pinus cembra* L. (Pinaceae). *Mitochondrial DNA Part B* 4 (2): 4202-4203. doi: 10.1080/23802359.2019.1693297
- Schroeder H, Kersten B** (2023) A small set of nuclear markers for reliable differentiation of the two closely related oak species *Quercus robur* and *Q. petraea*. Manuscript in preparation

Öffentlichkeitsarbeit und Präsentation

Öffentlichkeitsarbeit:

- Fernsehbericht zum Thema „Holz“ am 28. November 2018, WDR, Planet Wissen; <https://www.planetwissen.de/sendungen/sendung-holz-108.html>; mit Beteiligung des Thünen-Instituts für Forstgenetik, Holzherkunftsidentifizierung
- Besuch der Firma „Formafantasma“ am 19.09.2019 für Filmaufnahmen/Interviews über das Holzkompetenzzentrum für eine Ausstellung der Firma in London (geplant 04. März bis 17. Mai 2020), die Corona-bedingt leider abgesagt wurde.
- Stand in der BMEL-Halle zum Kompetenzzentrum Holzherkünfte auf der Internationalen Grünen Woche, 17.01. bis 26.01.2020, in Berlin.
- Am 17. Juli 2021 wurde der nachfolgende Artikel online veröffentlicht, der über die Arbeit des Thünen-Kompetenzzentrums Holzherkünfte berichtet: „Germany’s timber detectives take on illegal loggers“ <https://atlasofthefuture.org/project/centre-of-competence-on-the-origin-of-timber/>
- Am 10. März 2022 wurde der Artikel "The 'timber detectives' on the front lines of illegal wood trade - Germany's Center of Competence on the Origin of Timber tracks down which EU wood imports come from illegal sources in the world's third-biggest criminal sector" in National Geographic veröffentlicht: <https://www.nationalgeographic.com/environment/article/the-timber-detectives-on-the-front-lines-of-illegal-wood-trade>

Präsentationen:

Datum	Veranstaltung	Ort	Autoren	Thema
15.03.2018	Außenhandelstag des GD Holz	Bremen, Deutschland	Hilke Schröder , Birgit Kersten, Bernd Degen	Entwicklung genetischer Marker zur Gattungs- und Arterkennung in Holzverbundprodukten mittels Next Generation DNA-Barcoding (Vortrag)
24.04. bis 26.04.2018	Deutsche Baumpflege-tage	Augsburg, Deutschland	Hilke Schröder , Bernd Degen	Einsatz molekularer Marker zur Art- und Herkunftsbestimmung von Bäumen und Holz (Poster)
29.07.2018	Jahrestreffen der Deutschen Dendrologischen Gesellschaft	Ahrensburg, Deutschland	Bernd Degen	Genetische Identifizierung von Baumart und Baumherkunft (Vortrag)
08.10. bis 09.10.2018	Holzhandel und Walderhaltung – fünf Jahre Erfahrungen mit der EU-Holzhandelsverordnung und dem Thünen-Kompetenzzentrum Holzherkünfte	Hamburg, Deutschland	Hilke Schröder , Céline Blanc-Jolivet, Bernd Degen	Anwendung genetischer Methoden zur Bestimmung von Holzart und -herkunft (Vortrag)
21.01. bis 22.01.2019	„A Bed of Roses – Royal or Revival“ Symposium	London, Großbritannien	Hilke Schröder	The Bed of Roses – DNA analysis (Vortrag)
23.05.2019	Besucherguppe "Rantzau-Haus" Ahrensburg	Großhansdorf, Deutschland	Hilke Schröder , Bernd Degen	Anwendung genetischer Methoden zur Bestimmung von Holzart und -herkunft (Vortrag)
17.06. bis 20.06.2019	Barcode of Life conference	Trondheim, Norwegen	Hilke Schröder , Katrin Schöning-Stierand Bernd Degen, Birgit Kersten	Identification of tree species in wood composite products ("Lightning talk" und Poster)
06.02.2020	Besuch Oberstufen-Schülergruppe aus Hamburg-Niendorf	Großhansdorf, Deutschland	Hilke Schröder , Bernd Degen	Anwendung genetischer Methoden zur Bestimmung von Holzart und -herkunft (Vortrag)
13.09. bis 16.09.2021	Forstwissenschaftliche Tagung 2021	digital	Hilke Schröder , Birgit Kersten, Bernd Degen	Genetische Marker zur Gattungs- und Arterkennung in Holzverbundprodukten (Vortrag)
03.12.2021	Workshop des „Thünen-Kompetenzzentrum Holzherkünfte“	digital	Hilke Schröder , Céline Blanc-Jolivet, Bernd Degen	Entwicklung genetischer Methoden zur Bestimmung von Holzart und -herkunft (Vortrag)

09.03.2022	Thünen Themenfeldgespräch „Buche“	digital	Hilke Schröder	Thünen-Kompetenzzentrum Holzherkünfte – Prüfaufträge bei der Buche (Vortrag)
06.04.2022	Besuch Oberstufen-Schülergruppe aus Hamburg-Nienstedten	Großhansdorf, Deutschland	Hilke Schröder	Anwendung genetischer Methoden zur Bestimmung von Holzart und –herkunft (Vortrag)
19.04. bis 21.04.2022	“From Forests to Heritage” Conference	Amsterdam, Niederlande	Hilke Schröder	Genetic analysis of „The Bed of Roses“ (Vortrag)

Fazit

Die im Projekt erfolgreich entwickelten und validierten Marker und Multiplex-Markersetts erlauben eine effiziente genetische Erkennung vieler Gattungen und Arten von Hölzern, die in Holzverbundprodukten am häufigsten eingesetzt werden und fließen direkt in die praktische Anwendung im Rahmen des Thünen-Kompetenzzentrum Holzherkünfte ein. Wichtige Beiträge für die Praxis sind dabei auch die optimierten Protokolle zur DNA-Extraktion und die gewonnenen Ergebnisse zur Eignung verschiedener Markertypen für Mischhölzer aus Nadel und Laub. Die mittels NGS und bioinformatischer Auswertung gewonnenen neuen Chloroplasten und mitochondrialen Referenzgenome, z.B. von *Fagus sylvatica* bieten eine gute Grundlage für die Entwicklung weiterer genetischer Marker, deren Validierung durch die im Projekt gesammelten Referenzproben unterstützt wird. Außerdem können diese Referenzgenome in zukünftigen Studien für andere wissenschaftliche Fragestellungen verwendet werden. Das im Projekt generierte Wissen zu Taxonspezifischen DNA-Varianten könnte auch in potentielle zukünftige Ansätze zum Metabarcoding einfließen.

Inhaltsverzeichnis

Zusammenfassung	4
1. Einführung / Motivation	5
2. Methodik/Vorgehensweise/Projekttablauf	6
2.1 Material (Beschaffung zusätzlichen Materials)	6
2.2 Methoden und Vorgehensweise	7
2.2.1 Methoden bei der Validierung der DNA-Extraktion (Teilziel 1).....	7
2.2.2 Methoden zu NGS und Bioinformatik (Teilziel 2)	8
2.2.2.1 NGS.....	8
2.2.2.2 Bioinformatik.....	11
2.2.3 Methoden bei der Markerentwicklung (Teilziel 3)	13
2.2.4 Methoden bei der Entwicklung von Multiplex-Markersets (Teilziel 4).....	13
2.2.5 Methoden für Praxistests (Teilziel 5)	14
2.3 Projekttablauf	15
3. Ergebnisse	16
3.1 Ergebnisse zur Validierung der DNA-Extraktion (Teilziel 1)	16
3.2 Ergebnisse zu NGS und Bioinformatik (Teilziel 2)	19
3.3 Ergebnisse zur Markerentwicklung (Teilziel 3)	23
3.3.1 Nadel-Laub Differenzierung	24
3.3.2 Differenzierung von Ordnungen	25
3.3.3 Differenzierung von Familien.....	28
3.3.4 Identifizierung von Gattungsgruppen und Gattungen/Arten.....	29
3.3.5 Identifizierung von Subgattungen, Sektionen und Arten innerhalb der Gattung <i>Pinus</i>	33
3.4. Ergebnisse zu Multiplex-Markersets (Teilziel 4)	35
3.5. Ergebnisse zu den Praxistests (Teilziel 5)	37
4. Öffentlichkeitsarbeit / Veröffentlichungen / Vorträge	38
5. Fazit / Ausblick	41
6. Weitere Literaturangaben	41

Abbildungsverzeichnis

Abbildung 1: Phylogenetischer Baum basierend auf einem „whole plastome“-Alignment verschiedener Arten der Familie der Dipterocarpaceae.....	12
Abbildung 2: Gelelektrophorese der Restriktion des Mischtestversuchs (Mischen vor Extraktion) nach PCR mit den Primern NL_petA_F1/R1b.....	17
Abbildung 3: Genkarte der assemblierten Chloroplasten-Genomsequenz von <i>Pinus cembra</i> L.....	20
Abbildung 4: Zirkuläre grafische Darstellung der assemblierten mitochondrialen Genomsequenz von <i>Fagus sylvatica</i> L. und abgeleitete genetische Marker.....	21
Abbildung 5: Stand der Markerentwicklung für taxonomische Ebenen des Interesses und übergeordnete Ebenen.....	24
Abbildung 6: Experimentelle Validierung des Malpighiales-Markers 5_MP_matR/HaeIII.....	26
Abbildung 7: Darstellung eines Ordnungsmarkers für Myrtales (5_MY_matR/BsaWI).....	28
Abbildung 8: Darstellung eines Gattungsmarkers für <i>Pinus</i> (3_PI_rpoC2/Msel).....	30
Abbildung 9: Validierung eines Gattungsmarkers für <i>Populus</i> (3_PO_rpl33/Stul).....	32
Abbildung 10: Taxonomie und Markerentwicklung innerhalb der Gattung <i>Pinus</i>	34
Abbildung 11: Ausschnitt aus dem Multiplex-Markerset zur Identifizierung von Vertretern der Ordnung Fagales.....	36

Tabellenverzeichnis

Tabelle 1: Übersicht der in NCBI verfügbaren kompletten Sequenzen.....	10
Tabelle 2: Ergebnis der bioinformatischen Analyse von Nadel- und Laubgenomen.....	18
Tabelle 3: Screening des Malvales-Markers 5_MA_BR_ccmFc auf dem ABI.....	27
Tabelle 4: Zusammenstellung der zwei alternativen Fagales-Multiplex-Sets mit den Farbstoffen für Anwendung auf dem ABI (oben) bzw. den Beckmann (unten).....	36
Tabelle 5: Deklaration und Ergebnis der Markertests für 17 Sperrhölzer von GD-Holz.....	37
Tabelle 6: Vorträge/Poster von Wissenschaftlern zum Thema und zu Projektergebnissen.....	40

Anhangsverzeichnis

Anhang 1: Übersicht der verfügbaren Marker auf verschiedenen taxonomischen Ebenen.....	43
--	----

Abkürzungsverzeichnis:

bp	Basenpaare
cp	Chloroplast
CAPS	„Cleaved amplified polymorphic site“
EUTR	European Timber Regulation (EU Holzhandelsverordnung)
FLEGT	„Forest Law Enforcement, Governance and Trade“: EU-Aktionsplan zur Bekämpfung des illegalen Holzeinschlags
InDel	“Insertion/Deletion“: Polymorphismus, der auf einer Insertions- bzw. Deletions-Mutation beruht
mt, mito	Mitochondrium
NCBI	National Center for Biotechnology Information
NGS	„Next Generation Sequencing“: Methoden zur Hochdurchsatz-Sequenzierung
PCR	„Polymerase Chain Reaction“: Methode zur Vervielfältigung einzelner DNA-Abschnitte
PCR-RFLP	„Restriction Fragment Length Polymorphism“: Auf einer PCR und anschließendem Verdau mit einem Restriktionsenzym basierende Methode zur Analyse von Fragmentlängenunterschieden
Plastom	Komplette DNA-Sequenz des Genoms des Chloroplasten
SNP	„Single Nucleotide polymorphism“: Vererbare Variation eines einzelnen Nukleotides (Base) im DNA-Strang
SRA	„Short Read Archive“ (bei NCBI)

Zusammenfassung

Zur Eindämmung des illegalen Holzeinschlages wird sowohl in den USA als auch in Europa seit einigen Jahren vom Gesetzgeber eine genaue Kennzeichnung von importiertem Holz vorgeschrieben. Dazu gehört u.a. die genaue Deklaration der botanischen Art und der geographischen Herkunft des Holzes. Hierfür sind DNA-basierte Nachweismethoden bestens geeignet. Für eine ganze Reihe an gehandelten Baumarten stand die Entwicklung von geeigneten Markern noch aus. Insbesondere für Holzverbundprodukte stellt die Markerentwicklung und deren Anwendung eine große Herausforderung dar, da hier eine Mischung diverser Gattungen und Arten zur Herstellung verwendet werden kann. Holzanatomische Methoden sind in diesem Fall nur begrenzt erfolgreich. Daher war das Ziel dieses Projektes die Entwicklung genetischer Markersysteme – mit dem Fokus auf Chloroplasten- und mitochondrialen Markern - zur gleichzeitigen Erkennung vieler Gattungen und Arten, die in Holzverbundprodukten am häufigsten eingesetzt werden.

Im Projekt wurden - wie ursprünglich geplant - die Nadelbaumgattungen *Pinus*, *Picea*, *Pseudotsuga* und zusätzlich *Abies*, *Tsuga*, *Cedrus* und *Larix* sowie die Laubbaumgattungen *Betula*, *Eucalyptus*, *Populus*, *Fagus*, *Shorea* und auch hier zusätzlich *Quercus/Castanea* und *Fraxinus* bearbeitet. Von Arten dieser Gattungen (außer *Populus*) sowie anderen Gattungen der bearbeiteten Familien und Ordnungen wurde Blatt/Nadel- bzw. Holzmaterial von verschiedenen Quellen beschafft und die jeweilige DNA extrahiert. Es wurden von uns mitentwickelte DNA-Extraktionsprotokolle für Holz validiert, um DNA aus stark verarbeiteten Holzgemischen zu gewinnen. Es konnte bei der Verwendung von Gemischen aus Laub- und Nadelholz gezeigt werden, dass Chloroplastenmarker deutlich besser als mitochondriale Marker zur Identifizierung der enthaltenen Laub- und Nadelhölzer geeignet sind. Die experimentell beobachtete Tendenz, dass Laubbaum-DNA im Gemisch besser in der PCR amplifiziert wird als Nadelbaum-DNA, insbesondere bei der Verwendung mitochondrialer Marker, konnte mit bioinformatischen Ergebnissen zum Vergleich der zellulären Kopienzahlen von nukleären, mitochondrialen und Chloroplastengenomen sowie Genomgrößenvergleichen zwischen einigen Nadel- und Laubbaumarten untermauert werden.

Als Grundlage für die Markerentwicklung wurden sowohl öffentliche Daten als auch neugenerierte NGS-Daten verschiedener Laub- und Nadelbaumarten genutzt. Wichtigste Ergebnisse der bioinformatischen Analysen dieser Daten waren die Assemblierung und Annotierung neuer Chloroplasten- und/oder mitochondrialer Referenzgenome, u.a. von *Fagus sylvatica* und *Pinus cembra*. Diese Genome waren eine wichtige Grundlage für die weitere bioinformatische Auswahl von Taxon-spezifischen SNPs und InDels für die genetische Markerentwicklung. Im Projekt konnten für alle im Antrag geplanten Gattungen sowie die beiden Subgattungen von *Pinus*, vier Sektionen, drei Subsektionen und zahlreiche Arten innerhalb von *Pinus* genetische Marker entwickelt und erfolgreich validiert werden. Über die ursprünglichen Gattungen des Interesses hinaus, konnten auch für einige der zusätzlichen Gattungen und weitere Arten Marker entwickelt werden. Zur zusätzlichen Absicherung der Gattungs- und Artidentifizierung auf höheren taxonomischen Ebenen wurden zusätzliche Marker entwickelt, die spezifisch für diese Ebenen sind und zum Teil bereits in Markersets eingeflossen sind. Neben einem validierten Multiplex-Markerset, das eine gleichzeitige Erkennung vieler Gattungen innerhalb der Ordnung der Fagales erlaubt, befinden sich weitere Multiplex-Sets für Ordnungen sowie für Familien und Gattungen innerhalb der Coniferales in

der Optimierungsphase. Im gemeinsamen Praxistest mit dem Kooperationspartner GD-Holz konnten die meisten der Deklarationen mit den im Projekt entwickelten genetischen Markertests für die untersuchten Sperrhölzer bestätigt werden.

1. Einführung / Motivation

Laut WWF (World Wide Fund for Nature) gehen jährlich 13 Millionen Hektar Wald durch illegalen Holzeinschlag verloren. Seit Einführung verschiedener Kontrollmöglichkeiten laufen international agierende Holzhandelsunternehmen Gefahr, gegen internationale Gesetze (FLEGT Abkommen, Lacey Act) zu verstoßen, die sie zur genauen Deklaration der botanischen Art und Herkunft der gehandelten Hölzer verpflichten (Degen & Höltken 2011). Diese Gesetze gelten auch für die Einfuhr von Holzverbundprodukten auf den europäischen Markt. Diese sind aber aufgrund diverser Handelswege und Verarbeitung, vor allem in Fernost, besonders schwer zu verfolgen (Xiufang & Canby 2010).

Spätestens mit dem Inkrafttreten der EU-Holzhandelsverordnung (EUTR) im März 2013 wuchs daher die Nachfrage nach Methoden zum Art- und Herkunftsnachweis. Mit holzanatomischen Methoden ist häufig nur eine Bestimmung der Gattung, aber nicht der Art möglich. Bei Holzverbundprodukten müssen diese bisher in ihre Bestandteile zerlegt werden, um eine Bestimmung vornehmen zu können; d.h. holzanatomische Methoden sind in diesem Fall nur sehr begrenzt anwendbar. Daher benötigen wir effiziente Methoden zur Identifizierung möglichst vieler in Holzverbundprodukten verarbeiteten Gattungen und Arten, um dem Holzhandel weitere Methoden an die Hand zu geben, ihre Produkte überprüfen zu lassen. Genetische Marker sind hier eine wünschenswerte Alternative bzw. Ergänzung. Für eine ganze Reihe an gehandelten Baumarten und –gattungen steht die Entwicklung von geeigneten genetischen Markern jedoch noch aus. Insbesondere für Holzverbundprodukte stellt die Entwicklung solcher Marker und deren Anwendung eine Herausforderung dar, da hier eine Mischung diverser Gattungen und Arten zur Herstellung verwendet werden kann.

Im Zuge der „Barcode of Life“-Initiative (www.barcodeoflife.org, Hollingsworth et al. 2009) sind bereits verschiedene Regionen des Chloroplastengenoms als universelle Marker ausgewählt worden (z. B. Kress & Erickson 2008, Taberlet et al. 2007, Lahaye et al. 2008). Diese Marker reichen jedoch nicht, um alle Gattungen unter Berücksichtigung verschiedener taxonomischer Ebenen identifizieren zu können. Daher sollten in diesem Projekt anhand der bioinformatischen Analyse kompletter Chloroplastengenome und auch mitochondrialer Genome (insbesondere bei der Gattung *Pinus*) ausreichend Marker entwickelt werden, die eben diese Identifizierung auf verschiedenen taxonomischen Ebenen ermöglicht.

Damit also das Holzhandels-Sicherungs-Gesetz als deutsche Umsetzung der EU-Holzhandelsverordnung in der Praxis greift, müssen fälschungssichere Methoden zur Überprüfung der Art und Herkunft auch für Holzverbundprojekte zur Verfügung gestellt werden. Das vorliegende Projekt hatte deshalb zum Ziel, mit der Entwicklung von genetischen Markern zur Identifizierung häufig verwendeter Gattungen und Arten praxistaugliche Testverfahren bereitzustellen.

2. Methodik/Vorgehensweise/Projektablauf

2.1 Material (Beschaffung zusätzlichen Materials)

Im Projekt werden die Nadelbaumgattungen *Pinus*, *Picea* und *Pseudotsuga* sowie die Laubbaumgattungen *Betula*, *Eucalyptus*, *Populus*, *Fagus* und *Shorea* bearbeitet (Tab. 1 im Antrag). Eine wichtige Grundlage für die erfolgreiche Umsetzung der Teilziele des Projektes war die Beschaffung von zusätzlichem Pflanzenmaterial dieser Gattungen sowie von anderen Gattungen der bearbeiteten Familien und Ordnungen („outgroups“).

Von Baumarten der o.g. Gattungen (außer *Populus*) wurde Blatt/Nadel-Material bzw. Holzmaterial aus verschiedenen Quellen beschafft, u.a. aus dem Arboretum in Großhansdorf und dem Botanischen Garten in München, und die jeweilige DNA extrahiert.

Für die Gattung *Populus* kann auf eine umfangreiche DNA-Probensammlung aus verschiedenen Projekten, u.a. FastWood zurückgegriffen werden. Ausserdem konnten über eine Kooperation mit einer Kollegin aus Krasnojarsk 200 Individuen von *Pinus sibirica* aus verschiedenen Populationen eingesammelt werden. Im Rahmen einer Kooperation mit Felix Gugerli (WSL, Birmesdorf, Schweiz) erhielten wir Proben von 19 Individuen, die drei verschiedene Pinusarten repräsentieren. Diese waren vor allem für die Artdifferenzierung innerhalb der Gattung *Pinus* sehr hilfreich.

Für einige Gattungen konnte zusätzlich auf eine große Anzahl Proben aus anderen Projekten zurückgegriffen werden (u.a. *Fagus*-Proben aus dem GenMon-Projekt; *Picea*-Proben aus dem Projekt Fichte-Trockenheit, *Pseudotsuga*-Proben aus FitForClim, *Betula*-Proben aus einer Doktorarbeit) bzw. aus internationalen Kooperationen des Institutes genutzt werden (u.a., *Eucalyptus*-Proben aus einer Kooperation mit Prof. Christian Wehenkel in Mexico; *Betula*-, *Populus*-, *Pinus*- und *Picea*-Proben aus einer Kooperation mit Prof. Yulai Yanbaev in Russland).

Um Zugang zu in Asien beheimateten *Shorea*-Arten zu bekommen, wurden diverse botanische Gärten und asiatische Kollegen kontaktiert, was aber leider ohne Erfolg blieb. Deshalb wurde für die Gattung *Shorea* umfangreiches Holzmaterial (>500 Proben aus sieben asiatischen Herkünften) von den Kollegen des Thünen-Instituts für Holzforschung zur Verfügung gestellt. Es handelt sich hierbei um die Handelshölzer „White Meranti“ und „Red Meranti“.

Vom Kooperationspartner GD Holz wurden 31 Sperrholzproben und eine MDF-Probe geschickt. Diese Sperrholzproben beinhalten laut Deklaration alle der im Projekt bearbeiteten Gattungen mit Ausnahme von *Shorea*. Außerdem sind zusätzliche Gattungen enthalten, von denen einige ebenfalls inzwischen mit bearbeitet werden können, wie z.B. *Quercus*.

Die Probensammlung (bereits extrahierte DNA aus frischem Material) für das Projekt beläuft sich auf insgesamt mehr als 600 Individuen (Bäume und einige krautige Pflanzen). Der überwiegende Teil der Baumproben mit Art- und Herkunftsinformation des zugehörigen Baumes sowie anderen Metadaten wurden in der am Institut etablierten SDB („sample database“) erfasst und sind damit jederzeit recherchierbar. Die übrigen Baumproben werden derzeit noch erfasst.

2.2 Methoden und Vorgehensweise

Insgesamt sollten mit diesem Projekt folgende Teilziele verfolgt werden:

1. Validierung von DNA-Extraktions-Protokollen für Holzgemische (Validierung der DNA-Extraktion)
2. NGS und bioinformatische Datenanalysen zur Selektion potentiell differenzierender SNPs und InDels (NGS und Bioinformatik)
3. Validierung potentiell differenzierender SNPs und InDels in erweiterten Sets von Individuen zur Entwicklung komplexer „whole-plastome“ DNA-Barcodes (Markerentwicklung)
4. Umsetzung der DNA-Barcodes in genetische Markersets und Validierung an Holzverbundprodukten (Markersets)
5. Anwendung der Vielgattungs/Vielarten-Marker bei Holzprodukten in Zusammenarbeit mit dem Gesamtverband Deutscher Holzhandel (Praxistests)

Im Folgenden werden die Methodik und Vorgehensweise im Rahmen dieser Teilziele genauer beschrieben. Die erzielten Ergebnisse finden sich in Kapitel 3. Alle Publikationen, die im Projekt erstellt wurden, finden sich im Text in Fettdruck referenziert.

2.2.1 Methoden bei der Validierung der DNA-Extraktion (Teilziel 1)

Ein spezielles Extraktionsprotokoll für Holz wurde von unserem Institut in Kooperation mit der Universität Adelaide entwickelt und patentiert (Lowe et al. 2015). Dieses Protokoll wird routinemäßig für die DNA Extraktion aus Holz, z.B. im Rahmen der Bearbeitung der Prüfaufträge im Thünen-Kompetenzzentrum Holzherkünfte (<https://www.thuenen.de/de/thuenen-institut/verbundstrukturen/thuenen-kompetenzzentrum-holzherkuenfte>) verwendet, aber auch in Projekten für diverse Holzproben.

Im Rahmen von Arbeiten für eine geplante Doktorarbeit (Lasse Schindler) wurde dieses Protokoll optimiert. Neben Tests mit verschiedenen Temperaturen und Chemikalien wurden im Zuge dieser Doktorarbeit auch selbst Spanplatten aus definierten Holzgemischen hergestellt und die DNA extrahiert. Aus Spanplatten, die nur aus einer Holzart bestanden, konnte aus allen Platten DNA extrahiert werden, d.h. das vorliegende Extraktionsprotokoll funktioniert. In PCR-Tests mit ersten im Zuge der Doktorarbeit entwickelten cp-Primern wurde zusätzlich überprüft, ob alle im Gemisch enthaltenen Hölzer identifiziert werden können. Hierbei ergab sich, dass bestimmte Mischungen von Baumarten einen Einfluss auf das PCR-Ergebnis hatten. So gab es Mischungen, bei denen auch ein Anteil von 10% einer Holzart identifiziert werden konnte, aber auch solche, bei denen selbst 50% einer Holzart nicht wiederzufinden waren.

Mit diesem Phänomen haben wir uns im Rahmen dieses Teilziels genauer weiterbeschäftigt. Zur Untersuchung des Problem der Mischungen verschiedener Baumarten, insbesondere der Kombinationen von Laub- und Nadelhölzern wurden verschiedene Mischungsverhältnisse und Kombinationen von DNAs ausgewählter Laub- und Nadelbaumarten mit ausgewählten

DNA-Markern getestet, die in diesem Projekt entwickelt wurden (siehe 3.3), um optimale Primerkombinationen zu finden, bei denen dieses Phänomen so gering wie möglich auftaucht.

Diese Arbeiten erfolgten vor allem im Rahmen einer Bachelorarbeit (**Bünger, 2021**), wobei hierbei vier Nadel- und Laubbaumarten anhand der Marker der obersten taxonomischen Ebene (NL = Differenzierung Nadel/Laub) überprüft worden. Um die Mischungen zu erstellen, wurden folgende zwei Strategien ausprobiert:

- (i) Mischen nach Extraktion: Die DNA wurde einzeln aus den Individuen extrahiert, gemessen und auf die gleiche Konzentration verdünnt. Dann wurden die DNAs verschiedener Nadel-Laubkombinationen gemischt.
- (ii) Mischen vor Extraktion: Hierbei wurden in einem praxisnäheren Ansatz die Späne der einzelnen Holzarten vor der Extraktion in den Mischungen entsprechenden Gewichtsanteilen gemischt (10%/90%; 50%/50% und 90%/10%).

Diese experimentellen Analysen wurden durch bioinformatische Analysen im Rahmen der Bachelorarbeit ergänzt (Details und Ergebnisse in 3.1).

2.2.2 Methoden zu NGS und Bioinformatik (Teilziel 2)

2.2.2.1 NGS

Im Projekt wurden neue DNA-Sequenzdaten verschiedener Laub- und Nadelbaumarten mit folgenden unterschiedlichen Zielstellungen mittels NGS (Next Generation Sequencing) generiert:

1. Es wurden Sequenzdaten für das *de novo*-Assembly von cp- und mito-Genomen ausgewählter Arten benötigt. Diese Genome konnten dann als Referenzsequenzen für die Markerentwicklung in der jeweiligen Gattung bzw. Art verwendet werden. In diesem Zusammenhang wurde u.a. *Pinus cembra* mittels NGS sequenziert, um dann das komplette Chloroplastengenom zu generieren (siehe unten).
2. Für die Identifizierung von Taxon-spezifischen SNPs oder InDels in cp- und mito-Genomen (Grundlage für die Markerentwicklung) wurden ausserdem verschiedene Individuen einiger Laub- und Nadelbaumarten sequenziert bzw. mehrere Artvertreter einer Gattung. So wurden u.a. zahlreiche Individuen der Art *Picea abies* (über gesamtes Verbreitungsgebiet) sequenziert, um einen *Picea abies*-spezifischen cp-Marker zu entwickeln. Da innerhalb der Gattung *Pinus* neben cp-Markern auch mito-Marker für einige Arten entwickelt werden sollen, erfolgte eine Sequenzierung (*genome skimming*) von mehreren Individuen verschiedener Pinusarten.
3. Für einige Baumarten, wie z.B. innerhalb der Gattung *Quercus*, erfolgten Re-Sequenzierungen von Individuen, um zuvor entwickelte Artmarker (DBU-Projekt Weisseiche) nunmehr in diesem Projekt publikationsreif *in silico* zu validieren (**Schroeder und Kersten, 2022**).

Bei den o.g. Sequenzierung von cp- und mito-Genomen (1. und 2.) wurde das sogenannte „genome skimming“ angewandt (siehe Antrag), wobei die Sequenzierung von Gesamt-DNA in relativ geringer Sequenzabdeckung erfolgt (1-3X). Für die Re-Sequenzierungen erfolgte eine Sequenzierung in deutlich höherer Coverage.

Für die Extraktion von Gesamt-DNA wurde eine in unserem Labor etablierte Standardmethode (ATMAB) für Blattmaterial durchgeführt (**Bruegmann et al. 2022**). Da DNA für eine NGS besonders hohen Qualitätsansprüchen genügen muss, wurden die enthaltenen Waschschriffe (Dichlormethan, Ethanol) jeweils zweimal durchgeführt. Zusätzlich wurde nach dem RNase-Verdau nochmal gefällt und mit Ethanol aufgereinigt. Von den DNA-Proben wurden genomische Sequenzierbibliotheken generiert, welche dann mittels Illumina HiSeq 4000 sequenziert wurden (2x150 bp; GATC Biotech AG Konstanz/Genewiz/Azenta).

Daneben konnte die zunehmende Anzahl von inzwischen öffentlich verfügbaren kompletten cp- und mito-Sequenzen vieler verschiedener Baumarten (Tab. 1) sowie öffentliche Rohdaten aus NGS („short read archive“-SRA bei NCBI) für die bioinformatische Analysen im Rahmen der Markerentwicklung im Projekt genutzt werden.

Tab. 1: Übersicht der in NCBI verfügbaren kompletten Sequenzen (Stand 21.11.2022)

Art	Familie	cp (NCBI)	mito (NCBI)
<i>Pinus sylvestris</i>	Pinaceae	NC_035069	
<i>Pinus radiata</i>			
<i>Pinus taeda</i>		NC_021440	NC_039746
<i>Pinus contorta</i>		NC_011153	
<i>Pinus nigra</i>			
<i>Pinus massoniana</i>		NC_021439	
<i>Pinus sibirica</i>		NC_028552	
<i>Pinus cembra</i>		MN536531	
<i>Pinus pumila</i>		NC_041108	
<i>Pinus armandii</i>		NC_029847	
<i>Pinus nelsonii</i>		NC_011159	
<i>Picea abies</i>	Pinaceae	NC_021456	MN642623->26
<i>Picea sitchensis</i>		NC_011152	MK697696.1->MK697708.1
<i>Picea glauca</i>		KT634228	LKAM01000001->36
<i>Picea omorika</i>			
<i>Pseudotsuga menziesii</i>	Pinaceae	MH612867	
<i>Pseudotsuga sinensis</i>		NC_016064	
<i>Abies alba</i>	Pinaceae	NC_042410	ON378818-> ON378828
<i>Betula pendula</i>	Betulaceae	LT855378	LT855379
<i>Betula pubescens</i>		NC_039996	
<i>Betula nana</i>		NC_033978	
<i>Betula occidentalis</i>		NC_039993	
<i>Betula platyphylla</i>		NC_039994	
<i>Betula cordifolia</i>		NC_037473	
<i>Betula lenta</i>		NC_039992	
<i>Eucalyptus camaldulensis</i>	Myrtaceae	NC_022398	
<i>Eucalyptus grandis</i>		NC_014570	NC_040010
<i>Eucalyptus saligna</i>		NC_022397	
<i>Eucalyptus deglupta</i>		NC_022399	
<i>Eucalyptus urophylla</i>			
<i>Populus tremula</i>	Salicaceae	NC_027425	NC_028096
<i>Populus alba</i>		AP008956	NC_041085
<i>Populus nigra</i>		NC_037416	
<i>Populus davidiana</i>		NC_032717	NC_035157
<i>Fagus sylvatica</i>	Fagaceae	NC_041437	NC_050960.1
<i>Fagus grandifolia</i>			
<i>Fagus crenata</i>		NC_041252	
<i>Fagus engleriana</i>		NC_036929	
<i>Shorea laevis</i>	Dipterocarpaceae		
<i>Shorea assamica</i>			
<i>Shorea pachyphylla</i>		NC_040966	
<i>Shorea zeylanica</i>		NC_040965	

cp = Chloroplast, mito = Mitochondrium; Arten in fett: Arten, die laut Antrag in die Markerentwicklung einbezogen werden sollten; in **rot**: Komplette Referenzgenome, die im Rahmen dieses Projektes (sequenziert), assembliert, annotiert und in internationalen Zeitschriften publiziert wurden (siehe nächstes Kapitel und 3.2).

2.2.2.2 Bioinformatik

Zur *de novo* Assemblierung kompletter cp- und mito-Genome mittels CLC Genomics Workbench (CLC-GWB) wurden sowohl öffentliche NGS-Daten (vom SRA bei NCBI) als auch im Projekt generierte NGS-Daten (siehe 2.2.2.1) genutzt. Die Daten, Tools und Methoden, welche bei der Assemblierung und Annotierung der gewonnenen Sequenzen verwendet wurden, finden sich ausführlich in den entsprechenden im Projekt erstellten Publikationen beschrieben (**Schott et al. 2019; Mader et al. 2019; Mader, Schroeder et al. 2020; Kersten et al. 2022**).

Bei der Entwicklung von DNA-Markern lag der Fokus auf dem Chloroplasten- und mitochondrialen Genom. Als Grundlage für die Markerentwicklung wurden bioinformatisch DNA-Sequenzvarianten (insbesondere SNPs und InDels) identifiziert, die spezifisch für eine bestimmte taxonomische Kategorie sind, wobei der Fokus auf Gattungen lag. Es wurden aber auch spezifische SNPs und InDels für Subgattungen und Arten sowie übergeordnete taxonomische Kategorien, wie u.a. Familien und Ordnungen identifiziert.

Folgende bioinformatische Strategien und Methoden wurden bei der Identifizierung von Taxon-spezifischen SNPs und InDels angewandt (u.a. unter Verwendung der in Tab. 1 gelisteten umfangreichen Sequenzdaten):

1. Sogenannte „whole-plastome“-Alignments mit CLC-GWB. Dabei wurden komplette Sequenzen von cp-Genomen von mehreren Individuen der taxonomischen Kategorie des Interesses zusammen mit dem cp-Genom einer sogenannten „outgroup“ analysiert (z.B. wurden die Sequenzen der cp-Genome von möglichst vielen Arten der Gattung des Interesses zusammen mit dem cp-Genom eines Artvertreters einer anderen Gattung innerhalb derselben Familie aligniert). Diese Alignments dienen zunächst zur Erstellung von phylogenetischen Bäumen mit CLC-GWB, die eine erste Orientierung innerhalb des Taxons des Interesses ermöglichen (siehe Abb. 1 als Bsp. für die Familie der Dipterocarpaceae). Im Weiteren wurden dann Taxon-spezifische SNPs/inDels in diesen Alignments detektiert.
2. Alignments von großen Plastom-Fragmenten (F0frag, F1frag und F2frag) mit CLC-GWB im Falle einiger Nadelbäume. Da sich die cp-Genome vieler Nadelbaumgattungen strukturell voneinander unterscheiden (unterschiedliche Genstruktur), konnten in diesen Fällen keine kompletten Genome aligniert werden, sondern es wurden zuvor konservierte cp-Genomfragmente definiert und dann untereinander aligniert. Diese Alignments wurden insbesondere zur Identifizierung von InDels angewandt, die spezifisch für *Pinus* Subgattungen, Sektionen und einige Arten sind.
3. Genische Alignments von cp-Genen bzw. mito-Genen (Details in **Mader, Schroeder et al. 2020**). Dabei wurden für konservierte cp- bzw. mito-Gene die kompletten Gensequenzen vieler Laub- und Nadelbaumarten aligniert, d.h. pro Gen wurde ein umfangreiches Alignment mit Vertretern möglichst vieler Baumgattungen, Familien und Ordnungen erstellt (möglichst breite taxonomische Abdeckung in jedem genischen Alignment). Die genischen Alignments von cp-Genen enthalten inzwischen bis zu 255 Arten aus 69 Gattungen (51 Laubgehölze + krautige Pflanzen, 18 Nadelgehölze) und 25 Familien. Zusätzlich wurde Bambus mit in die Alignments aufgenommen.
4. Systematischer Vergleich der Längen intergenischer Linker in cp-Genomen von Laub- und Nadelbaumarten (möglichst breite taxonomische Abdeckung). Dazu wurde ein RubySkript

erstellt, womit eine Matrix mit intergenischen Linkerlängen in cp-Genomen verschiedener Baumarten generiert wurde. Diese Matrix wurde dann genutzt, um intergenische Linker mit Taxon-spezifischer Linkerlänge zu identifizieren.

5. Systematischer Vergleich der Längen intergenischer Linker in mito-Genomen. Es wurde ein PERL-Skript entwickelt, um systematisch die Längen intergenischer Linker in annotierten kompletten Sequenzen mitochondrialer Genome auszulesen und zwischen verschiedenen Taxa zu vergleichen.
6. Nutzung öffentlicher Daten für *in silico*-Validierung potentieller Taxon-spezifischer SNPs/InDels. Für einige Gattungen, wie z.B. *Shorea* und *Parashorea* wurden öffentlich verfügbare NGS-Daten aus Re-Sequenzierungen bzw. „genome skimming“-Ansätzen genutzt, um potentiell Taxon-spezifische SNPs, die mit den o.g. Strategien identifiziert wurden, weiter *in silico* zu validieren. Dafür wurden die entsprechenden NGS-Daten vom SRA (NCBI) heruntergeladen, mit CLC-GWB auf ein geeignetes Referenzgenom gemappt (Mappingansatz) und nachfolgend SNPs und kleine InDels in den Mappings detektiert. Im Falle einiger SNPs/InDels erfolgten weitere *in silico* Validierungen anhand von BlastN-Analysen unter Nutzung des NCBI-Blast-Tools.

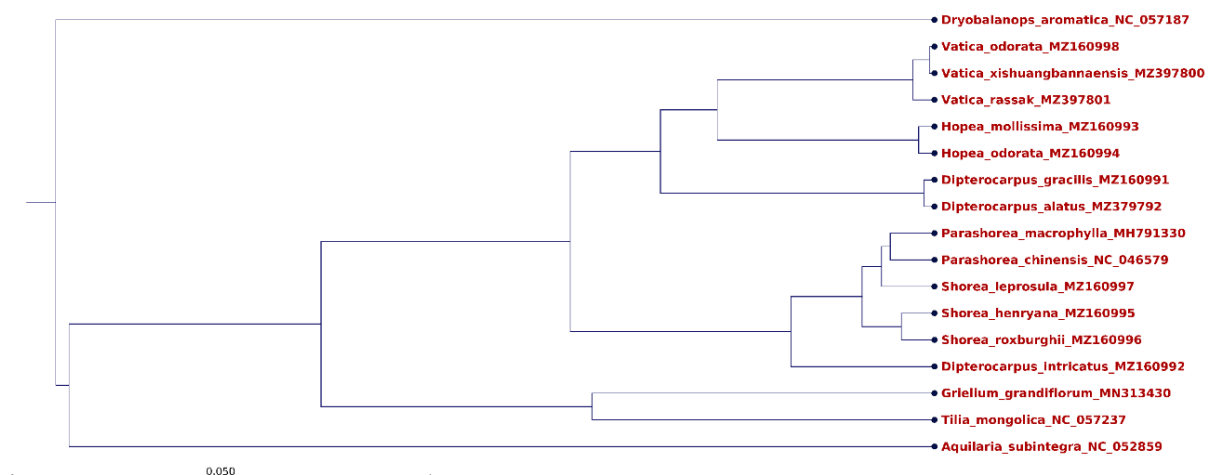


Abb. 1: Phylogenetischer Baum basierend auf einem „whole plastome“-Alignment verschiedener Arten der Familie der Dipterocarpaceae.

Es wurden öffentlich verfügbare Sequenzdaten von Genbank (NCBI) verwendet (siehe Accession-Nummern nach den Artnamen). Als „outgroups“ innerhalb der Ordnung der Malvales wurden folgende Arten verwendet: *Tilia mongolica* (Familie Malvaceae), *Griellium grandiflorum* (Neuradaceae) und *Aquilaria subintegra* (Thymelaeaceae). *Dryobalanops aromatica* ist eine Art innerhalb der Dipterocarpaceae. Vermutlich ist eine falsche Artzuordnung zu den Sequenzdaten die Ursache für die Ausgruppierung dieser Art im dargestellten Baum. Bei der Markerentwicklung für Dipterocarpaceae wurde *Dryobalanops* deshalb wie eine „outgroup“ behandelt. Der phylogenetische Baum zeigt auch, dass eine klare genetische Differenzierung zwischen den Gattungen *Shorea* und *Parashorea* vermutlich nicht möglich sein wird, da sich die Gattung *Shorea* nicht als monophyletische Gruppe im Baum darstellt.

Im Falle der Entwicklung Taxon-spezifischer cp-Marker wurde oft zunächst die o.g. Strategie 1 angewendet. SNPs bzw. InDels, die sich im „whole-plastome“-Alignment spezifisch für ein Taxon des Interesses darstellten und in einem konservierten Gen lagen, wurden dann mit

Strategie 2 weiteruntersucht, um herauszufinden ob diese SNPs/InDels auch über einen breiteren taxonomischen Bereich spezifisch waren (alternatives SNP/InDel-Allel tritt exklusiv nur in der nachzuweisenden taxonomischen Kategorie auf = „goldene SNPs/InDels“).

Um Taxon-spezifische SNPs in den genischen Alignments automatisiert zu identifizieren, wurde im Projekt die Software **SNPtax** entwickelt, deren Quellcode nunmehr öffentlich verfügbar ist (**Mader, Schroeder et al. 2020**). Mit dieser Software wird für jede einzelne Position im genischen Alignment bestimmt, für welches Taxon eine Base an dieser Position spezifisch ist, woraufhin für jedes Taxon des Interesses eine Liste mit spezifischen SNPs erstellt werden kann.

2.2.3 Methoden bei der Markerentwicklung (Teilziel 3)

Die Markerentwicklung erfolgte für ausgewählte Taxon-spezifische SNPs und InDels sowie intergenische cp/mito-Linker mit Taxon-spezifischen Längen (aus 2.2.2.2).

Zur Entwicklung sogenannter CAPS („cleaved amplified polymorphic site“)-Marker, auch PCR-RFLPs genannt, wurden aus Taxon-spezifischen SNPs solche SNPs herausgefiltert, die zur Änderung in der Erkennungssequenz einer Restriktionsschnittstelle führen. Für die gefilterten SNPs wurden dann geeignete Primersequenzen designt, die den SNP flankieren und PCR-Produkte in einem Längensbereich von etwa 100 bis 250 bp generieren. Die Primersequenzen wurden von der jeweils im Mapping bzw. Alignment verwendeten Referenzsequenz abgeleitet (siehe 2.2.2.2) und die Spezifität der Primer entsprechend der/den zu differenzierenden taxonomischen Ebene/Ebenen „justiert“. Grundsätzlich wurde versucht, Primer zu generieren, die auf einer möglichst hohen taxonomischen Ebene greifen. Wenn das nicht möglich war, so sollten die Primer zumindest für die zu differenzierende taxonomische Ebene (und möglichst die darüberliegende Ebene) eine hohe Spezifität aufweisen. Das Längenscreening der PCR-Produkte und Restriktionsfragmente erfolgte dann auf geeigneten Agarosegelen.

Für Taxon-spezifische InDels (Längenpolymorphismus) wurden wie für die SNPs geeignete Primersequenzen designt, die das InDel flankieren und PCR-Produkte in einem Längensbereich von 90 bis 250 bp generieren. Größere InDels, d.h. solche, bei denen der Längenunterschied zwischen den zu differenzierenden Taxa größer als 10 bp lag, konnten auf Agarosegelen sichtbar gemacht werden. Für kleinere InDels unter 10 bp reicht die Auflösung auf Agarosegelen nicht aus, daher erfolgte das Längenscreening mit Fluoreszenz-markierten Primern auf einem Kapillarsequenzierer („Genetic Analyzer“).

Für die Markerentwicklung basierend auf intergenischen Linkern mit Taxon-spezifischer Länge wurden Primer in die beiden Gene „gelegt“, welche den jeweiligen Linker flankieren. Das Längenscreening erfolgte dann, in Abhängigkeit von den zu detektierenden Längenunterschieden, entweder auf einem Gel oder mit dem „Genetic Analyzer“.

2.2.4 Methoden bei der Entwicklung von Multiplex-Markersets (Teilziel 4)

Die Entwicklung von Multiplex-Markersets erfolgte zunächst für das Screening der Fragmentlängen der verschiedenen enthaltenen Marker auf dem ABI. Dazu wurde für jeden Marker jeweils der *forward* oder der *reverse* Primer mit einem Fluoreszenzfarbstoff markiert,

um die Fragmente auf dem ABI- Sequenzierer sichtbar zu machen. Die Namen der Farbstoffe und die jeweilige Markierungsfarbe sind im Ergebnisteil aufgeführt.

Ein „Multiplex“ ist bei Markersets auf zwei Ebenen möglich: 1. Die Amplifikationen (PCRs) werden einzeln für jeden Primer durchgeführt und das Multiplexen erfolgt erst für die Analyse auf einem Genetic Analyzer oder 2. werden bereits Primer, die unterschiedliche Fragmente amplifizieren in der PCR gemeinsam verwendet. Die zweite Variante stellt sehr spezielle Anforderungen an die Primer, daher wurde hier zunächst immer nur die Variante 1 durchgeführt. Die zu amplifizierenden Fragmentlängen der einzelnen Primer werden auf Überschneidungen hin überprüft. Primer, die keine Überschneidungen in den Fragmentlängen produzieren, können dann dieselbe Fluoreszenzmarkierung erhalten. Sind Überschneidungen vorhanden, müssen unterschiedliche Markierungen verwendet werden, um die Primer im Multiplex anwenden zu können. Da neben Längenunterschieden (InDels) auch SNPs nach Restriktion mit in die Markersets aufgenommen wurden, musste zusätzlich darauf geachtet werden, dass immer das verschiedene Taxa differenzierende Fragment nach Restriktion noch die Fluoreszenzmarkierung enthält. Auf diesem Weg können bis zu etwa 11 Marker in einem Multiplex-Set im ABI 3730 Genetic Analyzer verwendet werden, da für dieses Gerät fünf verschiedene Fluoreszenzfarbstoffe zur Verfügung stehen. Häufig sind wir aber auf zwei kleinere Sets mit insgesamt mehr Markern ausgewichen, um zum Teil mehrere Marker für dieselbe Differenzierung verwenden zu können. So ist auch bei Ausfall eines Markers gewährleistet, dass ein Ergebnis erzielt werden kann. Längen und Markierungen sind beispielhaft im Ergebnisteil für ein Multiplex-Markerset gezeigt (Kap. 3.4).

Im März 2022 ist der ABI 3730 Genetic Analyzer ausgefallen. Eine kurzfristige Ersatzbeschaffung ist bei einem Gerät in der finanziellen Größenordnung nicht möglich. Es erfolgte im Mai eine Leihgabe eines Beckmann Kapillarsequenzierers von unseren Forstgenetik-Kollegen in Waldsiedersdorf. Der Beckmann arbeitet mit einem anderen Typ von Farbstoffen und es stehen nur vier Farbstoffe zur Verfügung. Daher mussten zunächst alle bereits für den ABI 3730 markierten Primer neu gelabelt und bestellt werden. Es folgte eine Anpassung und Optimierung der Markersets an die anderen Gegebenheiten des Beckmanns, d.h. die Markersets mussten umgestellt und zum Teil weiter verkleinert werden.

2.2.5 Methoden für Praxistests (Teilziel 5)

Am 28.04.2021 fand per Videokonferenz eine Besprechung mit Nils Olaf Petersen (GD Holz, Kooperationspartner im Projekt), Hilke Schröder und Birgit Kersten (beide Thünen-Institut) statt, um die Details des Praxistests und die Bereitstellung von Proben durch GD Holz zu besprechen.

Für den Praxistest erhielten wir von Mitgliedern des GD Holz diverse Sperrhölzer und eine MDF Platte unterschiedlichster Zusammensetzung. Für einige dieser Sperrhölzer wurden sowohl einzelne Marker zur Überprüfung der verschiedenen taxonomischen Ebenen als auch das Fagales-Markerset angewendet. Desweiteren waren in einigen Hölzern nur einzelne Baumarten deklariert, die derzeit noch mit den speziellen Gattungs- bzw. Artmarkern überprüft werden.

2.3 Projektablauf

Der ursprünglich im Antrag angegebene Zeitplan war mit Beginn 01.01.2018 datiert. Dieser wurde nach Bewilligung angepasst auf einen realen Projektbeginn zum 01.04.2018.

Im Antrag war das Prescreening als Überprüfung der Differenzierungskapazität zunächst einer noch großen Anzahl vorausgewählter SNP- und InDel-Marker geplant. Da in öffentlichen Datenbanken neben kompletten Chloroplastengenomen (in NCBI verfügbar) auch eine sehr große Menge an NGS Rohdaten (SRA bei NCBI) verfügbar waren, konnte die Strategie geändert werden. Die vorhandenen Rohdaten zahlreicher Individuen diverser Gattungen und Arten sowie die kompletten cp-Genominformationen dienten direkt bioinformatisch für eine erste Validierung der gefundenen differenzierenden SNPs und InDels. Ein Prescreening mittels MassARRAY-Technologie wurde daher nicht mehr durchgeführt.

Durch die Corona-Pandemie und die damit verbundenen Einschränkungen in der Labornutzungskapazität sowie durch die dem Projektträger jeweils zeitnah mitgeteilten Probleme bei Stellenbesetzungen war das Projekt zeitweise nicht ganz vollständig im ursprünglichen Zeitplan. Es kam zu Verzögerungen in wenigen Arbeitspaketen, wie in den jeweiligen Zwischenberichten ausführlich dargestellt. Diese Verzögerungen konnten aber im Rahmen der ersten kostenneutralen Projektverlängerung um ein Jahr weitestgehend ausgeglichen werden (Bewilligung vom 23.11.2020). Im Zuge einer weiteren kostenneutralen Verlängerung bis Ende September 2022 (Bewilligung vom 17.11.2021) konnte der Rückstand vollständig ausgeglichen werden und es wurden zusätzliche Marker (insbesondere zusätzliche Artmarker) entwickelt und NGS-Daten, insbesondere von *Pinus* zur weiteren Artmarkerentwicklung über das Projektende hinaus generiert.

Da im Verlängerungszeitraum leider der ABI-Sequenzierer, mit dem zuvor die Analyse der Markersets erfolgte, kaputt ging, wurde als Ersatzgerät vom Thünen-Institut ein Beckmann-Sequenzierer bereitgestellt, der aber zusätzliche Anpassungskosten beim Laborbedarf erforderte (Details im Umwidmungsantrag vom 21.06.2022). Ausserdem mussten bereits entwickelte Markersets an andere Fluoreszenzfarbstoffe (andere Gesamtzahl möglicher Fluoreszenzmarkierungen) angepasst werden. Insbesondere mussten alle bisher verwendeten Primer mit anderer Fluoreszenzmarkierung bestellt werden. Damit war es im letzten Verlängerungszeitraum zwar möglich die Umstellung bereits entwickelter Markersets vom ABI auf den Beckmann vorzunehmen, aber nicht – wie im Verlängerungsantrag zusätzlich geplant – noch weitere neue Markersets zu entwickeln. Es ist aber geplant, über das Projektende hinaus im Rahmen des Holzkompetenzzentrums weitere Markersets zu etablieren, basierend auf Markern, die bereits im Projekt entwickelt und erfolgreich validiert wurden.

3. Ergebnisse

3.1 Ergebnisse zur Validierung der DNA-Extraktion (Teilziel 1)

Das im Institut entwickelte Protokoll für die DNA-Extraktion aus Blatt- bzw. Nadelmaterial von Bäumen wurde weiter optimiert und publiziert (**Bruegmann et al. 2022**).

Ein spezielles Extraktionsprotokoll für Holz ist von unserem Institut in Kooperation mit der Universität Adelaide entwickelt und patentiert worden (Lowe et al. 2015). Dieses Protokoll wurde im Rahmen einer Doktorarbeit (Lasse Schindler) optimiert (Details siehe unter Methoden, 2.2.1). In PCR-Tests mit ersten im Zuge der Doktorarbeit entwickelten cp-Primern an Mischhölzern ergab sich, dass bestimmte Mischungen von Baumarten einen Einfluss auf das PCR-Ergebnis hatten. Die Untersuchungen führten zu dem Schluss, dass die meisten Laubhölzer die Nadelhölzer selbst bei den relativ gut funktionierenden cp-Markern in Mischungen überlagern. Für die Überprüfung, ob in Gemischen Nadelhölzer vorhanden sind, wurden neue Primer in den bereits ausgewählten Genen entwickelt, mit denen ausschließlich Nadelhölzer amplifiziert werden, nicht aber Laubhölzer. Eine Mischung der beiden Markertypen sollte das Problem der Überlagerung der Nadelhölzer durch Laubgehölze lösen. Mit diesem Phänomen haben wir uns im Rahmen dieses Teilziels genauer beschäftigt.

Zur Untersuchung des Problems der Mischungen verschiedener Baumarten, insbesondere der Kombinationen von Laub- und Nadelhölzern wurden verschiedene Mischungsverhältnisse und Kombinationen von DNAs ausgewählter Laub- und Nadelbaumarten mit ausgewählten DNA-Markern getestet, die in diesem Projekt entwickelt wurden (siehe 3.3).

Diese Arbeiten erfolgten vor allem im Rahmen einer Bachelorarbeit (**Bünger, 2021**), wobei hierbei die folgenden vier Nadel- und Laubbaumarten anhand der Marker der obersten taxonomischen Ebene (NL = Differenzierung Nadel/Laub) überprüft worden:

- NL_cox_1_F1a/R1a (mitochondrialer Marker)
- NL_cox_1_F2/R2a (mitochondrialer Marker)
- NL_petA_F1/R1b (Chloroplasten-Marker)
- NL_psbD_F1/R1 (Chloroplasten-Marker)

Die Mischungen wurden unter Anwendung von zwei verschiedenen Strategien hergestellt (Mischung nach Extraktion bzw. Mischen vor Extraktion; Details in 2.2.1). Diese Untersuchung hat gezeigt, dass der Erfolg der Differenzierung nicht nur allein von den verwendeten Primerpaaren abhängt, sondern auch von anderen Faktoren, wie dem Mischungsverhältnis oder den PCR-Bedingungen.

Insgesamt zeigt sich, dass es im Allgemeinen möglich ist, Kernholz-DNA von Laub- und Nadelbäumen durch die entwickelten Marker zu differenzieren (Abb. 2 für den Marker NL_petA_F1/R1b).

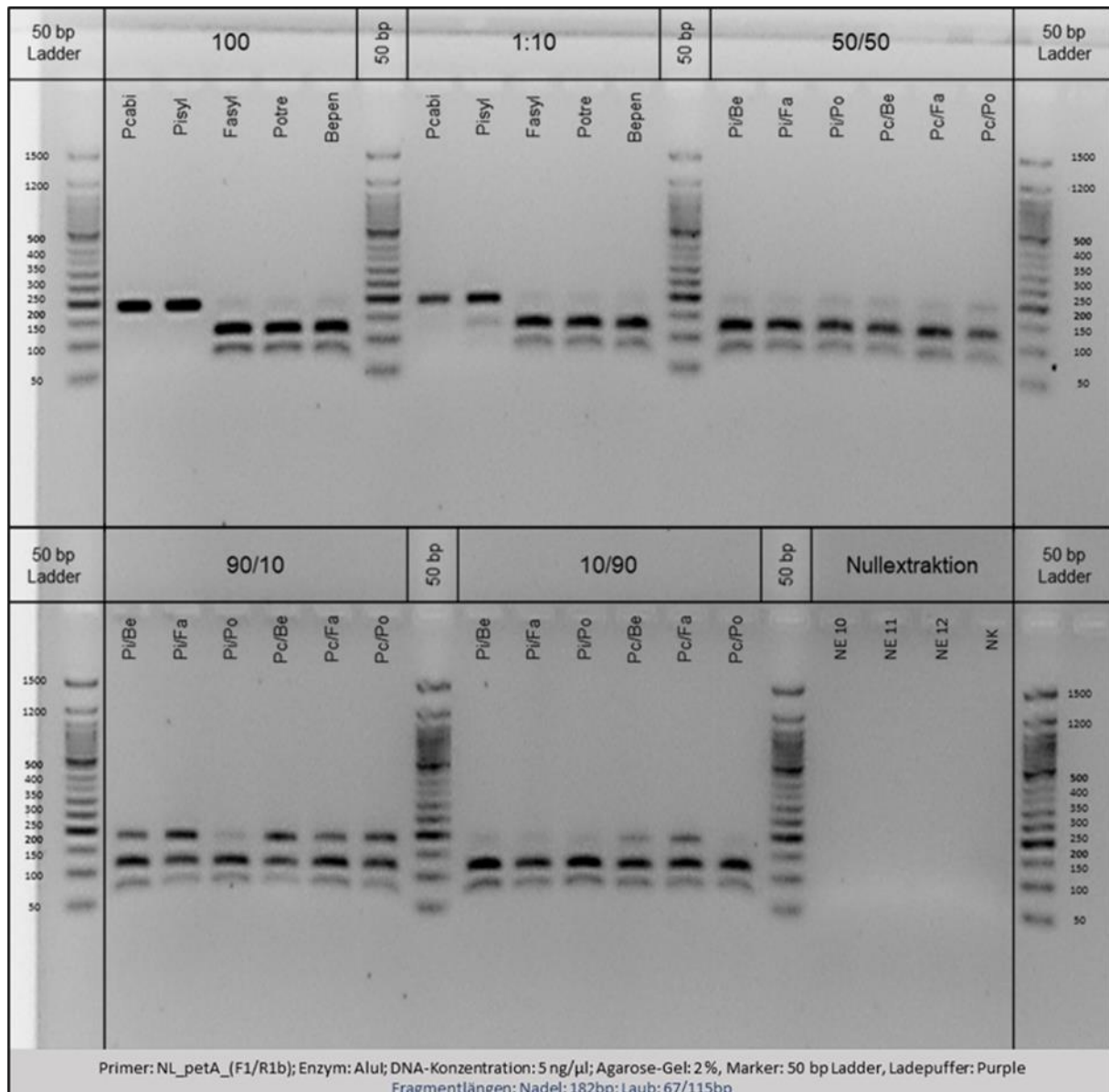


Abb. 2: Gelelektrophorese der Restriktion des Mischtestversuchs (Mischen vor Extraktion) nach PCR mit den Primern NL_petA_F1/R1b. Als Enzym wurde *AluI* verwendet. Pi (Pisyl): *Pinus sylvestris*; Pc (Pcabi): *Picea abies*; Fa (Fasyf): *Fagus sylvatica*; Po (Potre): *Populus tremula*; Be (Bepen): *Betula pendula*; NK: Nullkontrolle; NE: Nullextraktion; Fragmentlängen: Nadel: 182 bp; Laub: 67/115 bp.

Die in den Versuchen eingesetzten Primerpaare wiesen unterschiedliche Ergebnisse zur Identifikation von Nadel- und Laubbaum-DNA in verschiedenen Mischungsverhältnissen auf.

Die Versuche führen zu dem Ergebnis, dass in Holzgemischen die eingesetzten Chloroplasten-Marker geeigneter sind als die Mito-Marker NL_cox_1_F1/R1a und NL_cox_1_F2/R2a.

Diese Tendenz, dass Laubbaum-DNA im Gemisch besser in der PCR amplifiziert wird als Nadelbaum-DNA, wurde mittels bioinformatischer Analysen in der Bachelorarbeit genauer analysiert (Bünger, 2021). Dazu wurden NGS-Daten aus der Re-Sequenzierung von Gesamt-DNA aus Blättern bzw. Nadeln von 3 Laubbaum- bzw. 2 Nadelbaumarten verwendet. Die NGS-Datensätze wurden dazu auf cp- und Mitogenome der jeweils untersuchten Art mittels CLC-

GWB gemappt. Durch die Bestimmung der jeweiligen Mappingcoverage, war es möglich, die zelluläre Kopienzahl von Chloroplasten- bzw. mitochondrialem Genom in Relation zum Kerngenom abzuschätzen und dann zwischen Laub- und Nadelbäumen zu vergleichen. Tabelle 2 zeigt die Ergebnisse dieser Studie.

Tab. 2: Ergebnis der bioinformatischen Analyse von Nadel- und Laubgenomen. Aufgeführt sind die verschiedenen Baumarten, die Genomgrößen, die Coverage-Werte und das daraus berechnete Verhältnis zum Kerngenom. Tabelle aus **Bünger (2021)**

Baumart	Genom	Genomgröße (bp)	Coverage	Kopienanzahl im Verhältnis zum Kerngenom
<i>Fagus sylvatica</i>	Kern	542.302.265	7,93	2,00
	Mitochondrion	504.715	170,16	42,93
	Chloroplast	158.462	505,25	127,47
<i>Betula pendula</i>	Kern	391.815.048	32,23	2,00
	Mitochondrion	581.505	610,94	37,92
	Chloroplast	160.532	4343,72	269,58
<i>Populus tremula</i>	Kern	408.816.716	30,27	2,00
	Mitochondrion	783.442	650,18	42,96
	Chloroplast	156.067	7079,50	467,76
<i>Picea abies</i>	Kern	19.600.000.000	5,39	2,00
	Mitochondrion	4.900.000	92,22	34,22
	Chloroplast	124.000	10228,74	3795,15
<i>Pinus taeda</i>	Kern	22.000.000.000	3,82	2,00
	Mitochondrion	1.191.054	61,17	32,05
	Chloroplast	121.530	5391,37	2824,93

Die zellulären Kopienzahlen des mito-Genoms liegen in den untersuchten Laub- und Nadelproben in einem ähnlichen Bereich (Laub: 39-43; Nadel: 32-34; Tab. 2; **Bünger, 2021**). Da aber das nukleäre Genom der untersuchten Nadelbäume (Tab. 2) etwa 40x größer ist als das der Laubbäume (40-fach höheres Molekulargewicht), enthält die gleich DNA-Menge (Gewicht) eines Nadelbaums auch eine 40-fach geringere Kopienzahl des nukleären Genoms und damit auch des mito-Genoms im Vergleich zur gleichen DNA-Menge eines Laubbaumes. Damit ist zu erwarten, dass bei einem 50%:50%-Gemisch aus der DNA von Laub und Nadeln die PCR mit einem mito-Primerpaar auch drastisch weniger PCR-Produkt aus Nadeln im Vergleich zu Laub liefert.

Damit unterstützen die bioinformatischen Ergebnisse, die experimentell beobachtete Tendenz, dass Laubbaum-DNA im Gemisch besser in der PCR amplifiziert wird als Nadelbaum-DNA, insbesondere bei der Verwendung von mito-Markern.

Bei cp-Markern ist der Unterschied in der Amplifikation zwischen Laub- und Nadel-DNA nicht so drastisch, was durch die Ergebnisse der Bioinformatik unterstützt wird (Tab. 2). Auch wenn die Kopienzahl des nukleären Genoms bei gleicher DNA-Menge bei einem Nadelbaum etwa 40-fach geringer als beim Laubbaum ist, so wirkt sich das nicht im gleichen Verhältnis auf die jeweiligen Kopienzahlen des cp-Genoms aus, da die zelluläre Kopienzahl des cp-Genoms in Nadeln etwa 10-fach höher ist als in Laub (Tab. 2). Das erklärt, warum cp-Marker bei Laub-/Nadelgemischen besser funktionieren als mito-Marker.

Eine einheitliche Aussage diesbezüglich ist zum aktuellen Stand der Forschung jedoch noch nicht möglich, da hierfür noch mehr Sequenzdaten verschiedener Arten, Individuen und Gewebe (z.B. Holz) bioinformatisch analysiert werden müssten. Ein Vergleich der Kopienzahlen der verschiedenen Genome zwischen Nadel- und Laubbäumen fällt vermutlich je nach Gewebe etwas unterschiedlich aus.

Die in der Bachelorarbeit entwickelte Methode zur Abschätzung der zellulären Kopienzahl von cp- und mito-Genomen basierend auf NGS-Daten aus Re-Sequenzierungen wurde nachfolgend auch in folgender Publikation angewandt, die im Projekt erstellt wurde: **Kersten et al. 2022**.

3.2 Ergebnisse zu NGS und Bioinformatik (Teilziel 2)

Als Referenzen für die weitere Markerentwicklung (2.2.3) wurden im Projekt von folgenden Baumarten komplette cp- u./o. mito-Genome assembliert, annotiert und publiziert:

- von *Pinus cembra* das cp-Genom (**Schott et al. 2019**; Abb. 3)
- von *Fagus sylvatica* das cp-Genom (**Mader et al. 2019**)
- von *Fagus sylvatica* das mito-Genom (**Mader, Schroeder et al. 2020**; Abb. 4)
- von *Abies alba* das mito-Genom (**Kersten et al. 2022**)

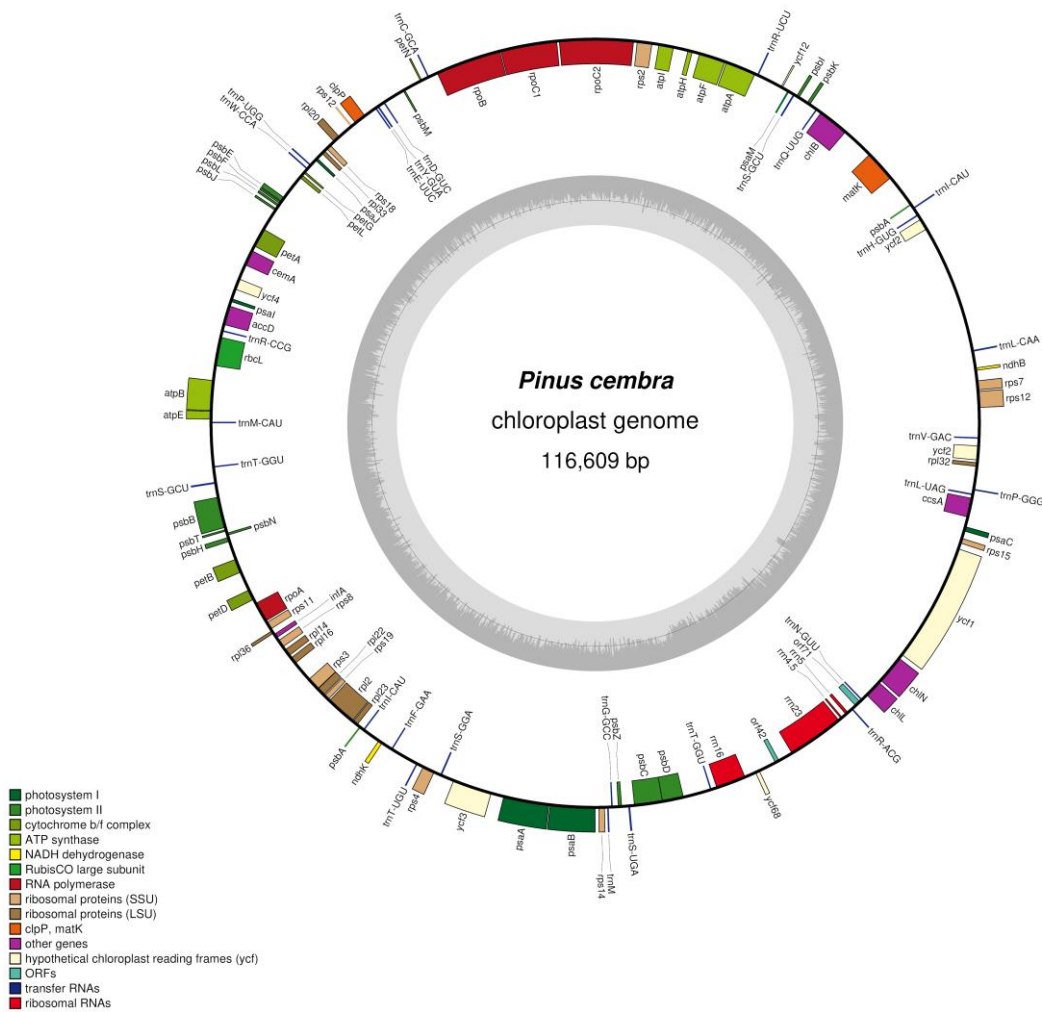


Abb. 3: Genkarte der assemblierten Chloroplasten-Genomsequenz von *Pinus cembra* L. Individuum PICEM_1; GenBank MN536531; siehe auch **Schott et al. 2019**

Die assemblierte und annotierte cpDNA-Sequenz von *Pinus cembra* (Subgattung Strobus; Abb. 3) konnte im weiteren Verlauf für die Markerentwicklung innerhalb der Gattung *Pinus* als wichtige Referenzsequenz verwendet werden.

NGS-Sequenzdaten (Rohdaten), die im Projekt GenMon für ein *Fagus sylvatica* Referenzindividuum (FASYL_29) generiert wurden, konnten im Projekt Holz-DNA-Barcoding erfolgreich genutzt werden, um die komplette cpDNA-Sequenz der Rotbuche zu assemblieren und zu annotieren (NC_041437.1; **Mader et al. 2019**). Parallel dazu wurden diese NGS-Rohdaten von Rotbuche (selbes Individuum) zur Assemblierung der kompletten mitochondrialen Genomsequenz im Rahmen dieses Projektes genutzt. Diese Sequenz wurde ebenfalls annotiert und publiziert (Abb. 4; GenBank Accession-Nummer MT446430; **Mader, Schroeder et al. 2020**). Die verwendeten Rohdaten wurden beim SRA eingereicht, um sie öffentlich zugänglich zu machen (**Mader and Kersten, 2020**).

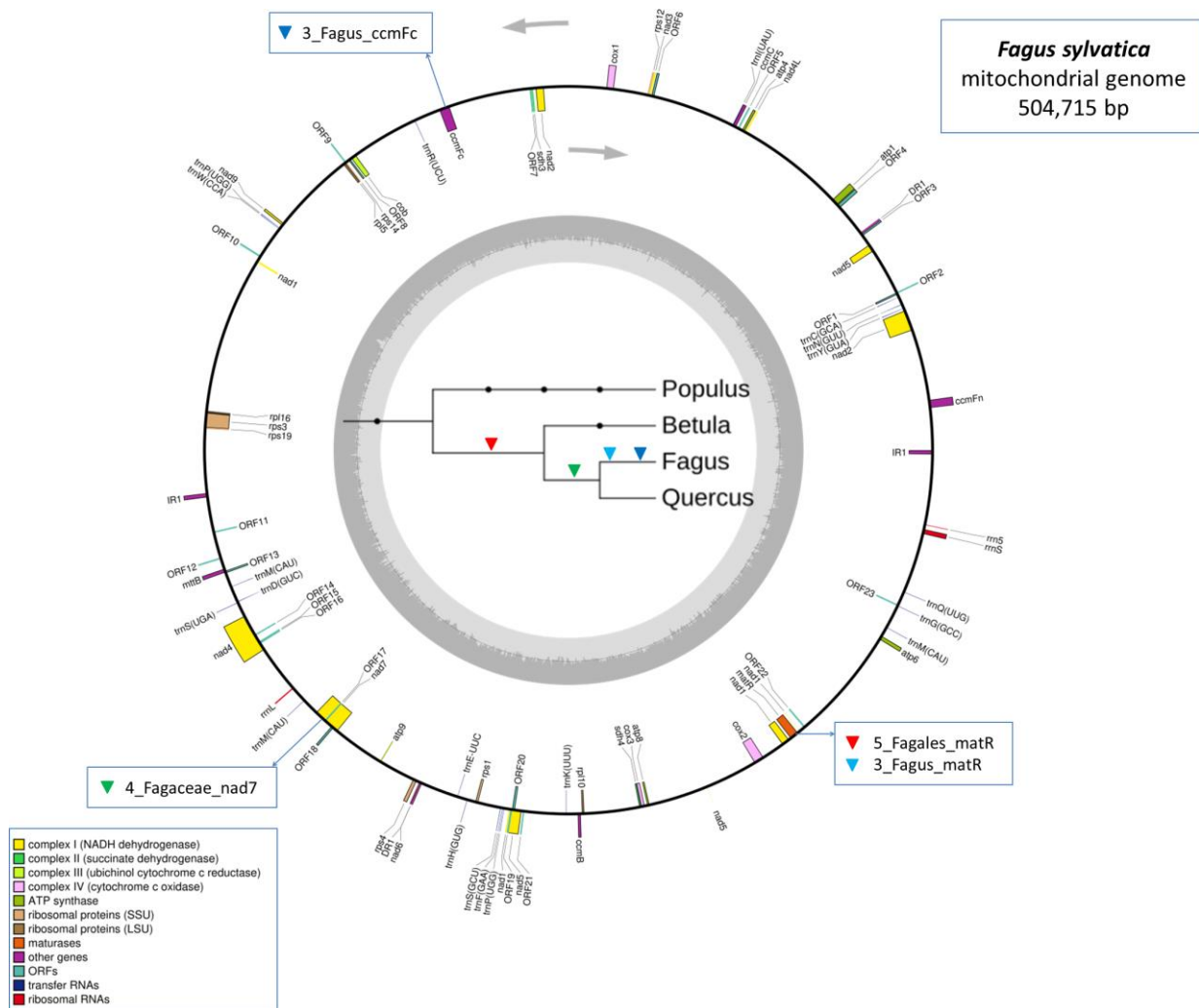


Abb. 4: Zirkuläre grafische Darstellung der assemblierten mitochondrialen Genomsequenz von *Fagus sylvatica* L. und abgeleitete genetische Marker.

Die dargestellte annotierte Genomsequenz des Individuums FASYL_29 ist bei GenBank verfügbar (Accession-Nummer MT446430). Die zirkuläre Darstellung entspricht nicht der physikalischen Struktur des Genoms *in vivo*, wo es vermutlich in unterschiedlichen Konformationen existieren kann. Vorhergesagte Gene sind als farbige Kästchen auf dem äußeren Kreis dargestellt. Der innere graue Kreis ist eine Darstellung des GC-Gehaltes. Die abgeleiteten genetischen Marker für verschiedene Taxa finden sich in den Textboxen am äußeren Kreis. Im phylogenetischen Baum im inneren Kreis ist gezeigt, für welche taxonomische Ebene der jeweilige Marker spezifisch ist. Die Abbildung wurde unverändert aus **Mader, Schroeder et al. (2020)** kopiert (© 2020 by the authors; Licensee MDPI, Basel, Switzerland; distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license; <http://creativecommons.org/licenses/by/4.0/>).

Diese Genomsequenz war die erste komplette mitochondriale Genomsequenz innerhalb der Familie der Fagaceae. Das mito-Genom von *Fagus sylvatica* bildete zusammen mit weiteren Daten eine hervorragende Grundlage für die Identifizierung verschiedener Taxon-spezifischer SNPs und InDels innerhalb der Ordnung der Fagales im Projekt (Abb. 4; 3.3).

Ausserdem wurde unter Verwendung bereits publizierter kurzer Illumina-Reads (Mosca *et al.*, 2020) sowie langer PacBio Reads, die vom internationalen AForGen-Netzwerk (<https://aforgen.wsl.ch/en/>) von einem *Abies alba* Individuum bereitgestellt wurden, die komplette mitochondriale Genomsequenz assembliert und annotiert (**Kersten et al. 2022**;

Tabelle 1). Diese Sequenz bildet eine wichtige Grundlage für die zukünftige Markerentwicklung in der Gattung *Abies*.

Die bioinformatische Auswertung der in Tab. 1 aufgeführten umfangreichen Sequenzdaten zusammen mit im Projekt neu generierten Referenzsequenzen (siehe oben) mittels Alignment- und Mappingansätzen und nachfolgender SNP- und InDel-Detektion (siehe Methoden) führte zur Identifizierung von einer Vielzahl von Taxon-spezifischen SNPs und InDels in cp- und mito-Genomen, die die Grundlage für die Markerentwicklung auf allen in Abb. 5 (unter 3.3) dargestellten taxonomischen Ebenen waren.

Die automatisierte Erstellung einer Matrix intergenischer Linker in cp-Genomen (2.2.2.2) erlaubte die Identifizierung einiger Linker mit Taxon-spezifischer Länge, wie innerhalb der Gattung *Pinus* und der Familie der Dipterocarpaceae. Im Falle der Dipterocarpaceae erfolgte eine Vorauswahl von Linkern, die möglicherweise spezifisch für Dipterocarpaceae bzw. untergeordnete taxonomische Kategorien sind. Die Länge dieser intergenischen Linker wurden dann für die verschiedenen Arten im Dipterocarpaceae „whole plastome“-Alignment (Abb. 1) bestimmt und verglichen. Für die weitere Markerentwicklung (nächstes Kapitel) wurde der Linker *ndhK_ndhJ* ausgewählt, da er möglicherweise eine spezifische Länge von 115 bp in Dipterocarpaceae hat. Außerdem zeigt auch *Tilia* theoretisch eine Gattung-spezifische Länge von 135 bp für diesen Linker. Der Linker *petB_psbH* kann möglicherweise die Gattungsgruppe *Shorea/Parashorea/Hopea* (Linkerlänge: 110 bp) von allen anderen Arten inklusive anderer Dipterocarpaceae-Arten unterscheiden. *Tilia* zeigt auch für diesen Linker wieder eine spezifische Länge (128 bp). Der Linker *petB_psbH* wurde deshalb auch für die weitere Markerentwicklung ausgewählt.

Da für die Gattung *Pinus*, neben Gattungsmarkern auch Marker für untergeordnete taxonomische Ebenen geplant waren, wie die Subgattungen und einige *Pinus*-Arten, wurden insbesondere hier die intergenischen Bereiche des cp-Genoms in die bioinformatischen Analysen einbezogen, da diese eine höhere Variabilität als die genischen Bereiche aufweisen. Da die cp-Genome verschiedener Nadelbaumgattungen unterschiedliche Genstruktur zeigen, erfolgten dazu keine „whole plastome“-Alignments, sondern separate Alignments für große Plastom-Fragmente (F0frag, F1frag und F2frag; Kapitel 2.2.2.2). In diesen Alignments konnten einige InDels identifiziert werden, die spezifisch für *Pinus* Subgattungen und eine Sektion sind (siehe nächstes Kapitel).

Auch für andere Baumarten konnten potentiell spezifische bzw. differenzierende SNPs/InDels bioinformatisch identifiziert werden, die die Grundlage für die Entwicklung genetischer Artmarker waren (Anhang 1). Die bereits generierten NGS-Daten von *Picea abies* (verschiedene Individuen) sollen über das Projektende hinaus zur Identifizierung Art-spezifischer Sequenzvarianten als Grundlage für die Entwicklung eines Artmarkers genutzt werden.

Im Falle vieler Nadelbaum-Gattungen kann die Kombination von Art-spezifischen cp- und mito-Markern genutzt werden, um Hybride zu identifizieren. Das ist möglich, da in den meisten dieser Gattungen das mito-Genom maternal und das cp-Genom paternal vererbt wird (siehe Antrag). Im Falle der Gattung *Taxus* wurden Art-spezifische InDels im mito-Genom sowie Art-spezifische SNPs im cp-Genom gefunden, die nunmehr die genetische Identifizierung von

Hybriden aus *T. cuspidata* und *T. baccata* mit abgeleiteten Markern ermöglichen (*T. x. media*; siehe nächstes Kapitel).

Um auch einige Hybriden innerhalb von *Pinus* genetisch bestimmen zu können, werden über das Projektende hinaus die bereits vorliegenden NGS-Daten verschiedener *Pinus*-Arten mit ähnlichen Strategien ausgewertet.

Ein Teil der identifizierten Taxon-spezifischen bzw. differenzierenden SNPs und InDels wurde bereits publiziert (**Mader, Schroeder et al. 2020; Bruegmann et al. 2022**). Es ist geplant, alle weiteren Taxon-spezifischen Marker, die erfolgreich zur Markerentwicklung verwendet wurden, zusammen mit den entsprechenden Markern und Markersets in einer großen Publikation zusammenzufassen und damit einem breiten internationalen Publikum zugänglich zu machen.

3.3 Ergebnisse zur Markerentwicklung (Teilziel 3)

Bei der Markerentwicklung lag der Fokus auf Markern, die spezifisch für eine definierte taxonomische Ebene sind („goldene Marker“). Dazu wurden Taxon-spezifische SNPs und InDels über die bioinformatische Analyse ausgewählt wurden (siehe Methoden).

Es erfolgte die Entwicklung zahlreicher PCR-basierter Marker für diverse taxonomische Ebenen: CAPS-Marker (PCR-RFLP) für SNPs (wenn diese zur Änderung in der Erkennungssequenz einer Restriktionsschnittstelle führen) sowie einfache PCR-Marker für InDels (Längenscreening). Zu diesem Zweck wurden zunächst spezifische Primer designt, die den SNP bzw. das InDel flankieren und PCR-Produkte in einem Längenbereich von etwa 100 bis 250 bp generieren. Die Primersequenzen wurden von der jeweils im Mapping bzw. Alignment verwendeten Referenzsequenz abgeleitet und die Spezifität der Primer wurde entsprechend der/den zu differenzierenden taxonomischen Ebene/Ebenen „justiert“.

Grundsätzlich wurde immer versucht, Primer zu generieren, die auf einer möglichst hohen taxonomischen Ebene greifen. Die Primer-Spezifität auf verschiedenen taxonomischen Ebenen wurde in den genischen Sequenzalignments (Kapitel 3.2.) bzw. in BlastN-Analysen der Primer gegen nichtredundante Nukleotidsequenzen bei NCBI geprüft.

In Abbildung 5 ist ein Überblick gegeben, für welche der taxonomischen Ebenen erfolgreich Marker entwickelt werden konnten (Markerzusammenstellung in Anhang 1). Einige dieser Marker sind in den folgenden Kapiteln für die einzelnen taxonomischen Ebenen im Detail beschrieben.

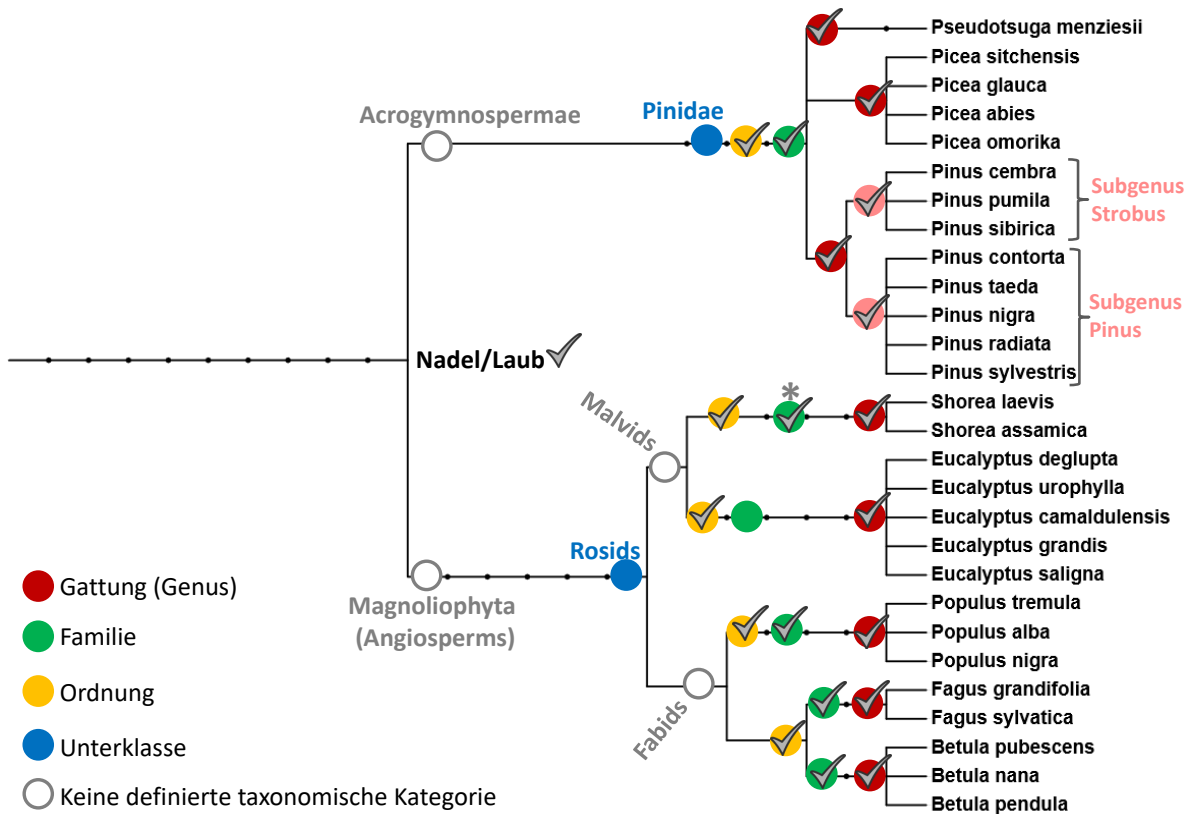


Abb. 5: Stand der Markerentwicklung für taxonomische Ebenen des Interesses

Phylogenetischer Baum für alle Baumarten, die in die Entwicklung von Gattungsmarkern einbezogen werden sollten. Für alle im Antrag geplanten Gattungen sowie die Subgattungen von *Pinus* wurde jeweils mindestens 1 Marker entwickelt (rote/rosa Kreise mit grauen Häkchen). Zusätzlich wurden für höhere taxonomische Ebenen (andere Kreise mit grauen Häkchen) Marker entwickelt, die die Bestimmung der in Holzmischproben enthaltenden Gattungen unterstützen sollen. Außerdem wurden Marker für die Familie der Dipterocarpaceae entwickelt (Häkchen mit Stern darüber), die noch weiter validiert werden. Darüber hinaus wurden Marker für Poaceae entwickelt, die nicht im phylogenetischen Baum dargestellt sind. Eine Zusammenfassung aller im Projekt entwickelten Marker finden sich im Anhang 1.

3.3.1 Nadel-Laub Differenzierung

Die höchste taxonomische Ebene, die den Ausgangspunkt bildet, ist die Unterscheidung von Nadel- und Laubgehölzen. Die Herausforderung bei einem Alignment (Sortierung der Sequenzen) auf dieser hohen taxonomischen Ebene war es, Bereiche für das Primerdesign zu finden, die bei allen Gattungen möglichst konserviert sind.

Für diese Unterscheidung haben wir zunächst die Sequenzen der Untereinheit I des mitochondrialen Gens Cytochromoxidase (coxI) diverser Laub- und Nadelbäume verglichen und nach SNPs gesucht. Das Ziel war hier, Variationen zu finden, die für eine große Anzahl an Laub- bzw. Nadelgewächsen eine 100%ige Differenzierung bieten. Daher wurde außer den Gattungen von Interesse (Tab. 1) eine große Anzahl zusätzlicher Gattungen mit in die Alignments einbezogen (Kapitel 3.2). Aus einer bioinformatischen Vorauswahl von 26 SNPs blieben zwei, auf deren Grundlage 2 CAPS-Marker entwickelt wurden. Mit diesen Markern ist eine Trennung von Nadel- und Laubgehölzen zu 100% der genannten Gattungen möglich. Diese beiden mitochondrialen Marker sind im Kapitel 3.1 bereits aufgeführt.

Als weitere Bereiche konnten in vier (aus 27 vorausgewählten) konservierten Chloroplastengenomen (*petN*, *psbB*, *psbC*, *psbD*) jeweils ein bzw. zwei SNPs gefunden werden, die den oben genannten Kriterien entsprachen und mit Hilfe der daraus entwickelten CAPS-Marker eine 100%ige Unterscheidung ermöglichen. Nach Validierungen mit diversen Individuen verblieben zwei der beiden konservierten Chloroplastengenome *petN* und *psbD*, in denen jeweils ein SNP gefunden wurde, der den oben genannten Kriterien entsprach und mit Hilfe der daraus entwickelten CAPS-Marker eine 100%ige Unterscheidung ermöglichte. Beide SNPs enthalten Erkennungssequenzen für das Restriktionsenzym *AluI*. Im *petA*-Gen werden die Laubgehölze geschnitten, im *psbD*-Gen die Nadelgehölze (*NL_petA/AluI* und *NL_psbD/AluI*). Die Auswahl mehrerer SNPs, die demselben Ziel dienen, ermöglicht die unter 3.1 beschriebene Problematik zu klären.

Für die Nadel-Laub Differenzierung wurden drei zusätzliche mitochondriale Marker, *NL_ccmFn*, *ccmFc_intron* und *NL_matR_intron*, entwickelt, die alle auf theoretischen Längenpolymorphismen basieren. Die weitere praktische Testung zeigte, dass die Marker *ccmFc_intron* und *NL_matR_intron* zwar die vorhergesagten Längenunterschiede zeigten, jedoch nur eingeschränkt einsetzbar sind.

Der Marker *ccmFc_intron* zeigt zwar spezifische Längen für die Nadelbaumgattungen, hat aber den Nachteil, dass nur einige Nadelbaumgattungen amplifiziert werden (*Pinus*, *Abies* und *Larix*). Die praktische Testung ergab, dass u.a. in folgenden Nadelbaum-Gattungen keine PCR-Produkte amplifiziert werden: *Picea*, *Sequoia*, *Thuja*, *Calocedrus*. Damit ist dieser Marker leider nicht für die universelle Anwendung bei der Nadel-Laub Differenzierung geeignet.

Der Marker *NL_matR_intron* wurde zwar erfolgreich mit vielen Einzelproben von Laub- bzw. Nadelhölzern validiert, zeigte sich aber leider problematisch bei der Testung von gemischten Laub-/Nadelhölzern (Kapitel 3.1). Deshalb wurden auch für diesen Marker spezielle Nadelprimer entwickelt, die sich aber in der experimentellen Testung ebenfalls nicht bewährt haben.

Generell sollte erwähnt werden, dass derzeit das Design von Primern für die Amplifikation von Bereichen des mitochondrialen Genoms in Nadelbäumen noch sehr schwierig ist, da bisher nur von sehr wenigen Nadelbaumarten die mitochondrialen Genome sequenziert wurden (Tab. 1). Das erklärt, warum mit Praxistests nicht immer die theoretischen Ergebnisse aus der Bioinformatik bestätigt werden können.

3.3.2 Differenzierung von Ordnungen

Da die Ordnung **Coniferales** die einzige Ordnung innerhalb der Nadelgehölze ist, sind die unter 3.3.1 beschriebenen Nadel-Laub-Marker automatisch auch diagnostisch für die Ordnung Coniferales.

Die unter 3.2 beschriebene Generierung und Annotierung der kompletten mitochondrialen Sequenz von *Fagus sylvatica* bildete eine hervorragende Grundlage für die Entwicklung mitochondrialer Marker innerhalb der Ordnung der **Fagales**. Dazu erfolgten verschiedene Alignments mitochondrialer Gene. So konnte auf der taxonomischen Ebene der Ordnung für die Fagales ein SNP im mitochondrialen Gen *matR* identifiziert werden, der diese Ordnung

von allen anderen differenziert. Die entwickelten Primer greifen bei allen verwendeten Laub- und Nadelgewächsen, und alle Fagales werden mit dem Restriktionsenzym *NciI* geschnitten (Marker = 5_FA_matR/*NciI*; Mader, Schroeder et al. 2020; Abb. 4). Ebenfalls erfolgreich validiert und angewendet wurde ein weiterer mitochondrialer Marker, 5_FA_ccmFc_F1/R1. In dem Gen *ccmFc* ist ebenfalls ein Fagales-spezifischer SNP enthalten. Das Fagales-spezifische Allel wird mit dem Restriktionsenzym *MluCI* geschnitten, sodaß nach PCR und Restriktion zwei Fragmente entstehen, während alle anderen - nicht Fagales - ungeschnitten verbleiben.

Im mitochondrialen *matR*-Gen wurde basierend auf den bioinformatischen Analysen ebenso ein SNP detektiert, der spezifisch für die Ordnung **Malpighiales** ist. Auf dieser Grundlage konnte ein CAPS-Marker entwickelt werden (5_MP_matR/*HaellI*). In den Genotypisierungen mit diesem Marker wurden erwartungsgemäß die PCR-Produkte der untersuchten Malpighiales-Vertreter mit dem Restriktionsenzym *HaellI* geschnitten, wogegen Laubbaumarten anderer Ordnungen das ungeschnittene PCR-Produkt zeigten (Abb. 6).



Abb. 6: Experimentelle Validierung des Malpighiales-Markers 5_MP_matR/*HaellI*.

Die Proben verschiedener Malpighiales-Arten sind sandfarben unterlegt. Die PCR-Produkte aller Malpighiales-Proben liegen nach Verdau geschnitten vor und zeigen ein Fragment der Größe 53 bp. Die weiteren Fragmente von 46 und 13 bp, die beim Verdau entstehen, sind auf dem Gel nicht erkennbar. Alle anderen Laubhölzer zeigen ein Fragment der Größe 112 bp (ungeschnittenes PCR-Fragment). Nadelhölzer werden gar nicht amplifiziert (nicht gezeigt).

Für die Ordnungen **Malvales** und **Brassicales** wurde ein spezifisches InDel (6 bp) im mitochondrialen Gen *ccmFc* identifiziert, wobei Malvales und Brassicales im Vergleich zu anderen Ordnungen die Insertion zeigen. Basierend auf diesem Längenpolymorphismus wurde ein PCR-Marker entwickelt (5_MA_BR_ccmFc). Das Längenscreening zur Genotypisierung einer größeren Anzahl von Individuen verschiedener Arten mit diesem Marker erfolgte noch auf dem ABI-Sequenzierer (Tab. 3). Alle Vertreter der Ordnung Malvales zeigten ein Fragment der Größe 143 bp auf dem ABI im Vergleich zu Vertretern von 8 anderen Ordnungen, die ein Fragment von 136 bp Länge zeigten bzw. nicht amplifiziert wurden.

Tab. 3: Screening des Malvales-Markers 5_MA_BR_ccmFc auf dem ABI

Ordnung	Art/Gattung	ABI-Längen	Ordnung	Art	ABI-Längen
Malvales	<i>Hibiscus rosa</i>	143	Fagales	<i>Quercus robur</i>	136
Malvales	<i>Hibiscus syriacus</i>	143	Fagales	<i>Quercus robur</i>	136
Malvales	<i>Alcea rosea</i>	143	Fagales	<i>Castanea mollissima</i>	136
Malvales	<i>Tilia Hybride 1</i>	143	Fagales	<i>Fagus sylvatica</i>	136
Malvales	<i>Tilia Hybride 2</i>	143	Fagales	<i>Fagus orientalis</i>	136
Malvales	<i>Tilia Hybride 3</i>	143	Fagales	<i>Betula pubescens</i>	136
Malvales	<i>Tilia petiolaris</i>	143	Fagales	<i>Betula luminifera</i>	136
Malvales	<i>Tilia cordata</i>	143	Fagales	<i>Carpinus betulus</i>	136
Malvales	<i>Tilia platyphyllos</i>	143	Fagales	<i>Corylus colchica</i>	136
Myrtales	<i>Epilobium sp.</i>	136	Fagales	<i>Alnus glutinosa</i>	136
Myrtales	<i>Eucalyptus grandis</i>	136	Fagales	<i>Juglans nigra</i>	136
Myrtales	<i>Eucalyptus urophylla</i>	136	Laminales	<i>Fraxinus americana</i>	136
Malpighiales	<i>Populus tremula</i>	**	Sapindales	<i>Acer pseudoplatanus</i>	136
Fabales	<i>Robinia pseudoacacia</i>	136	Sapindales	<i>Aesculus turbinata</i>	136
			Ginkgoales	<i>Ginkgo biloba</i>	**

Nur Vertreter der Ordnung der Malvales (farblich hinterlegt) und Brassicales zeigen eine Länge von 143 bp beim Längenscreening der PCR-Produkte auf dem ABI. **: kein PCR-Produkt.

Zur Differenzierung von Baumarten der Ordnung **Myrtales** von den Arten anderer Ordnungen konnte - basierend auf einem spezifischen SNP im mitochondrialen *matR*-Gen - ein CAPS-Marker entwickelt und validiert werden (5_MY_matR/BsaWI). Alle Arten, die der Ordnung Myrtales angehören werden mit dem restriktionsenzym *BsaWI* in zwei Fragmente geschnitten (Abb. 7).

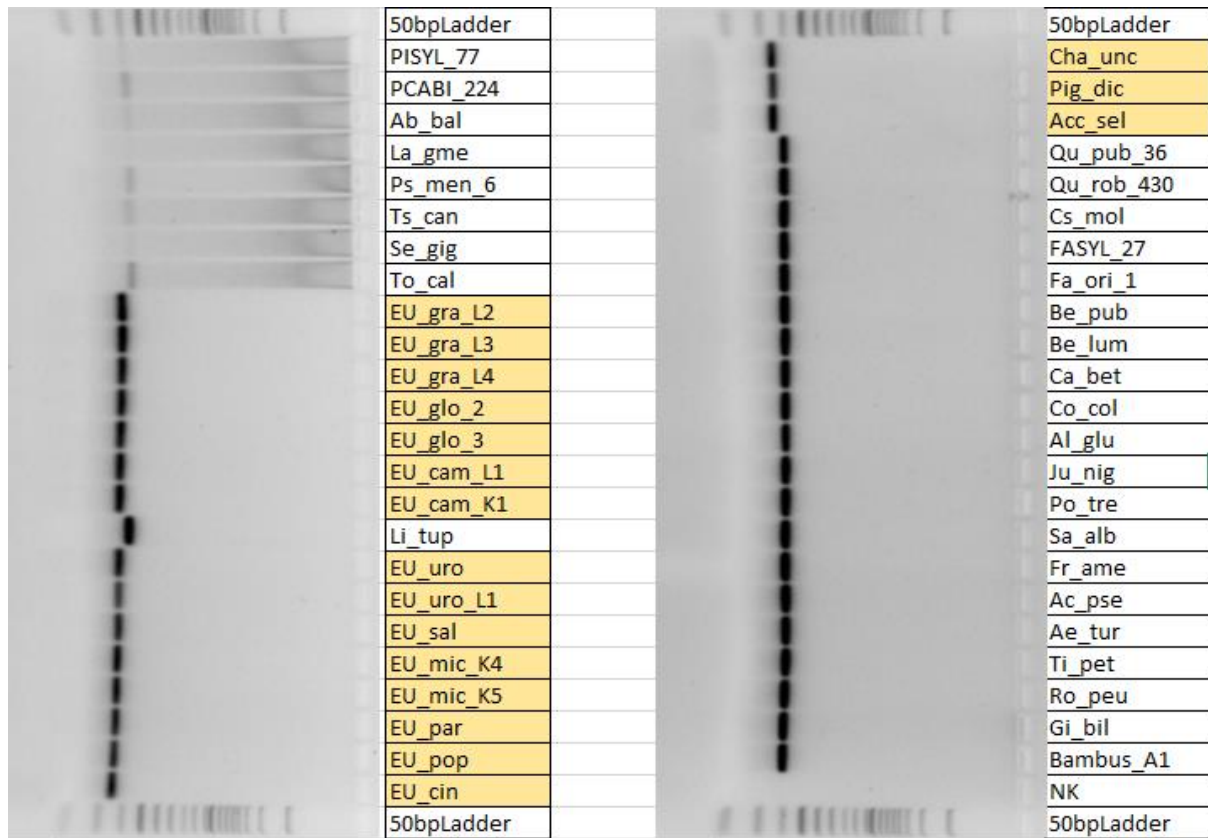


Abb. 7: Darstellung eines Ordnungsmarkers für Myrtales (5_MY_matR/BsaWI)

Die Proben verschiedener Myrtales-Arten sind sandfarben unterlegt. Alle Myrtales-Proben sind geschnitten und zeigen ein Fragment der Größe 121 bp (das abgeschnittene kleinere Stück von 13 bp ist auf dem Gel nicht erkennbar). Alle anderen Laubbölzer zeigen ein Fragment der Größe 134 bp (ungeschnittenes PCR-Fragment). Die Nadelhölzer (links oben) werden nur schwach oder gar nicht amplifiziert.

3.3.3 Differenzierung von Familien

Für die Unterscheidung der **Pinaceae** von allen anderen Familien wurde im mitochondrialen Gen *mttB* ein Längenunterschied von 4 bp identifiziert (4_PI_mttB/Del, die Pinaceae sind 4 bp kürzer). Diese Längenunterschiede lassen sich sehr einfach über PCR und Längenscreening auf einem Kapillarsequenzierer analysieren.

Für die **Cupressaceae** wurde ein mitochondrialer Marker gefunden, der ursprünglich für die Differenzierung von Subgattungen innerhalb der Gattung *Pinus* gedacht war. Bei diesem Marker (4_CU_nad6) zeigen die Cupressaceae ein anderes Schnittmuster als alle anderen getesteten Familien und Gattungen. Dieser Marker ist erfolgreich mit zahlreichen Individuen validiert worden.

Für die Identifizierung der Familie der **Taxaceae** konnten im Chloroplastengen *matK* im Abstand von knapp 1000 Basenpaaren zwei SNPs identifiziert werden, an denen jeweils mit dem Restriktionsenzym *TfiI* die Vertreter der Taxaceae in zwei Fragmente geschnitten werden. Eine Validierung mit 28 Individuen aus der Familie der Taxaceae bestätigte diese SNPs als Marker.

Für die **Fagaceae** konnten zwei Marker aus dem mitochondrialen Genom identifiziert werden, die ebenso wie die Marker für die Ordnung Fagales aus der bioinformatischen Analyse des

Mitochondrien-Genoms von *Fagus sylvatica* hervorgegangen sind (Kapitel 3.1). Im Gen *ccmb* werden mit dem Restriktionsenzym *MspI* alle Fagaceae geschnitten (4_FA_ccmb/*MspI*) und im Gen *nad7* schneidet das Restriktionsenzym *Sfcl* ebenfalls alle Fagaceae (4_FA_nad7/*Sfcl*; Abb. 4; Mader, Schroeder et al. 2020). Die **Salicaceae** können über einen Größenunterschied von 6 bp ebenfalls im mitochondrialen Gen *ccmb* unterschieden werden (4_SA_ccmb/*Del*, die Salicaceae sind 6 bp kürzer).

Für die Familie **Betulaceae** (Ordnung Fagales) wurden in einem „whole plastome“-Alignment verschiedener Fagales-Arten drei SNPs identifiziert, die potentiell Betulaceae-Arten innerhalb der Fagales differenzieren. Einer dieser SNPs liegt im *rpoB*-Gen und konnte für die Entwicklung eines CAPS-Markers genutzt werden (4_BE_rpoB/*Scal*), da der SNP zu einer Änderung einer Restriktionsschnittstelle für *Scal* führt. Da dieser Marker kein „goldener“ Marker ist, kann er nur verwendet werden, wenn zuvor bestätigt wurde, dass die Probe nur Fagales-Arten enthält. Als ein weiterer Familien-spezifischer Marker wurde ein InDel im cp-Gen *ndhB* identifiziert, wobei die Betulaceae eine 6-bp Insertion zeigen. Im Längenscreening mit dem daraus abgeleiteten PCR-Marker 4_BE_ndhB_InDel zeigten alle untersuchten Betulaceae-Individuen ein PCR-Produkt in erwarteter Länge von 126 bp, mit Ausnahmen von 2 Individuen der Gattung *Juglans*, die eine PCR-Produkt-Länge von 120 bp zeigten (wie auch die getesteten Nicht-Betulaceae-Arten). Bei den Abweichlern scheint es sich jedoch nur um seltene Vertreter der Gattung *Juglans* zu handeln. Eine BlastN-Analyse der *ndhB*-Sequenz von *Betula pendula* bei NCBI ergab, dass die Mehrzahl der 41 *Juglans*-Sequenzen mit Blast-Hits die Betulaceae-spezifische Insertion zeigten. Nur 2 *Juglans*-Sequenzen zeigten sich hier als Abweichler.

Ein SNP, der in der bioinformatischen Analyse als potentiell Fagales-spezifisch selektiert wurde, führte zur Entwicklung des mitochondrialen Markers 5_FA_ccmFc/*MluCI*. Die Validierung dieses Markers mittels PCR-RFLP ergab, dass dieser wegen weniger Abweichler zwar nicht für Fagales geeignet ist, jedoch bei Vertretern der **Juglandaceae** und **Betulaceae** innerhalb der Fagales ein spezifisches Restriktionsmuster ergab. Weitere Validierungen bestätigten, dass dieser Marker spezifisch für Juglandaceae und Betulaceae (gemeinsam) ist.

Für **Dipterocarpaceae** wurde - basierend auf potentiellen Familien-spezifischen SNPs im cp-Gen *atpA* ein CAPS-Marker entwickelt, dessen experimentelle Testung demnächst geplant ist: 4_DI_atpA_495/*Tsel*. Außerdem wurde der potentielle Dipterocarpaceae-spezifischer Marker *ndhK_ndhJ* entwickelt, der auf einem Längenpolymorphismus im intergenischen Linker *ndhK_ndhJ* basiert (3.2) und dessen experimentelle Testung auch noch geplant ist.

3.3.4 Identifizierung von Gattungsgruppen und Gattungen/Arten

Bei den Nadelhölzern wurden zunächst für die Gattungen ***Pinus***, ***Picea*** und ***Abies*** verschiedene Gene des Chloroplastengenoms bioinformatisch auf SNPs und InDels hin durchsucht. Im cp-Gen *rpoC2* wurde eine große Anzahl potentieller SNPs und InDels gefunden, von denen für fünf Primer generiert wurden. Aus diesen konnten für die Gattung ***Pinus*** zwei Marker (3_PI_rpoC2/*AcI* und 3_PI_rpoC2/*MseI*) und für die Gattung ***Picea*** ein Marker (3_PC_rpoC2/*MseI*) etabliert werden. Bei allen drei Markern handelt es sich um SNPs, die Restriktionsstellen für Enzyme aufweisen und daher als einfache CAPS-Marker genutzt

werden können. Diese Marker führen nur innerhalb der Pinaceae (3_PI_rpoC2/*AccI*) bzw. nur bei allen Nadelhölzern (3_PI_rpoC2/*MseI* und 3_PC_rpoC2/*MseI*) zu Amplifikationsprodukten (Abb. 8). Daher wurden keine weiteren Laubgehölze für die Validierung verwendet. Zusätzlich wurden zwei CAPS-Marker für die Gattung **Abies** entwickelt. Beide befinden sich im Chloroplasten-Genom, einer im *cemA*-Gen (3_AB_cemA/*Bst*UI), der andere im *matK*-Gen (3_AB_matK/*Hinf*I). Beide Marker konnten mit einer größeren Anzahl an *Abies*-Individuen validiert werden und stehen ebenfalls als Gattungsmarker zur Verfügung.

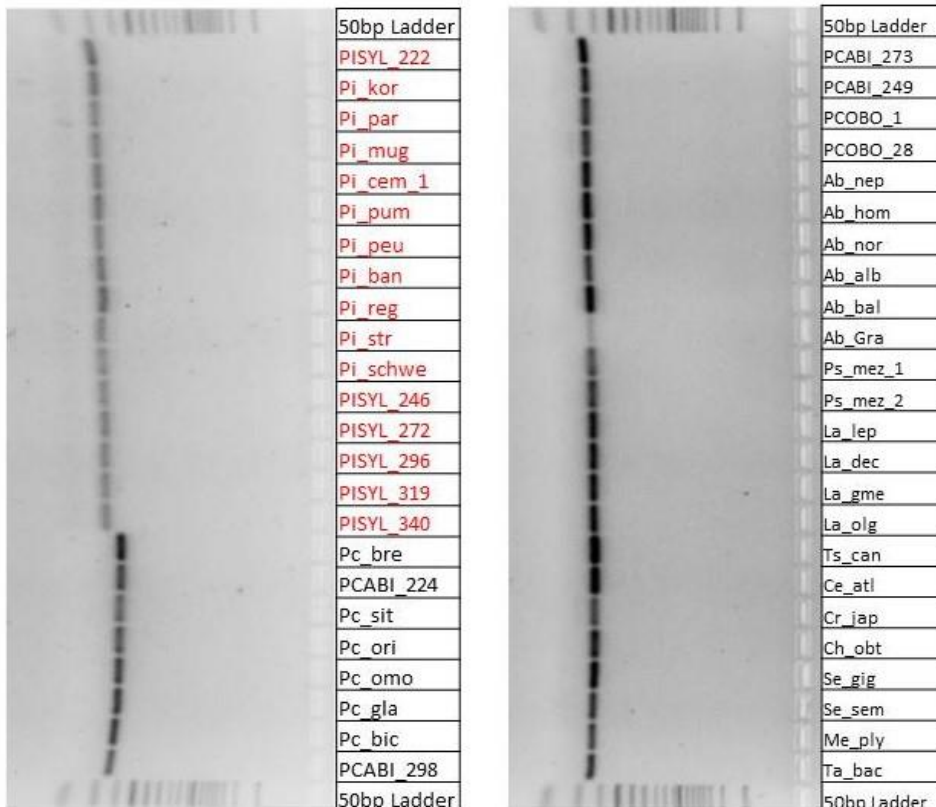


Abb. 8: Darstellung eines Gattungsmarkers für *Pinus* (3_PI_rpoC2/*MseI*).

Die Proben verschiedener *Pinus*-Arten sind rot markiert. Alle *Pinus*-Proben sind geschnitten und zeigen ein Fragment der Größe 80 bp (das abgeschnittene kleinere Stück von 36 bp ist auf dem Gel nicht erkennbar). Alle anderen Nadelhölzer zeigen ein Fragment der Größe 116 bp (ungeschnittenes PCR-Fragment).

Des Weiteren konnten für zwei weitere Gattungen der Nadelholzgewächse ebenfalls im *rpoC2*-Gen ein SNP bzw. Längenpolymorphismus identifiziert werden. Für die Identifizierung der Gattung **Larix** dient der Marker 2_PI_rpoC2_F1/R1. Wird an diesem SNP mit dem Restriktionsenzym *Bfal* geschnitten, ergeben sich für diese Gattung andere, spezifische, Fragmente als für die anderen Nadelgehölze. Für die Bestimmung der Gattung **Pseudotsuga** handelt es sich um den Marker 2_PI_F2frag_1/Ins. Diese Gattung bildet in diesem spezifischen Abschnitt des Gens ein 24 bis 36 bp längeres Amplifikationsprodukt als die anderen zur Validierung verwendeten Gattungen.

Zwei weitere Polymorphismen für die Gattung **Pseudotsuga** wurden ebenfalls bei der Suche nach Art-spezifischen Markern innerhalb der Gattung *Pinus* gefunden. Bei beiden handelt es

sich um Längenpolymorphismen: Marker 1_PI_F2frag_1 LP im intergenischen spacer psbJ_petA (ursprünglich nur vorgesehen für *Pinus sylvestris*) und 1_PI_F0frag2 LP aus dem petD-Gen (ursprünglich nur vorgesehen für *Pinus pinea*). Dieser zweite Marker zeigt zusätzlich auch eine andere Länge für die Gattung ***Tsuga***, die somit ebenfalls mit diesem Marker identifiziert werden kann. Ein weiterer Längenpolymorphismus (2_PI_F1frag_1 LP), mit dem ***Tsuga*** differenziert werden kann, ist einer, der ursprünglich für die Sektion *Pinus* entwickelt wurde, dort aber nicht standhielt. Auch ein Marker für die Gattung ***Cedrus*** ist bei der Suche nach artspezifischen Markern innerhalb der Gattung *Pinus* entdeckt worden (1_PI_F0frag_3 LP) (Anhang 1).

Für die Identifizierung von Vertretern der Gattung ***Taxus*** eignet sich der Chloroplasten-Marker 3_TA_atpB_F1/R1. Nur die untersuchten Vertreter dieser Gattung zeigten im Längenscreening (Sequenzierer) PCR-Produkte von 157 bp. Alle anderen Nadelbaum-Vertreter zeigten andere Längen. In Laubbäumen wird das Fragment nicht amplifiziert.

Für die Gattung ***Betula*** wurden mittels Bioinformatik fünf SNPs in fünf Genen im cp-Genom identifiziert sowie ein InDel. Ein goldener SNP im matK-Gen konnte für die Entwicklung eines CAPS-Markers genutzt werden (Marker 3_BE_matK/Mealll). Dieser Marker wurde erfolgreich validiert. 53 *Betula*-Individuen, die mindestens 16 verschiedenen *Betula*-Arten angehören, zeigten das *Betula*-spezifische Verdauemuster im Gegensatz zu 19 Bäumen von 19 verschiedenen anderen Fagales-Arten, die mit diesem Marker getestet wurden. Da dieser Marker nur in Fagales zu einem Amplifikationsprodukt führt, wurden auch keine Vertreter anderer Ordnungen getestet. Das identifizierte InDel wurde zur Entwicklung eines einfachen PCR-Markers genutzt (Marker 3_BE_rpoC2_Ins) und erlaubt mittels Screening der Länge des jeweiligen PCR-Produktes die Differenzierung von Birken innerhalb der Familie Betulaceae (PCR-Produkt-Längen: 111 bp bei Birken, 99 bp bei anderen Betulaceae).

Für die Gattung ***Populus*** (Pappeln) wurde ein CAPS-Marker basierend auf einem goldenen SNP im cp-Gen rpl33 entwickelt und erfolgreich validiert (Marker 3_PO_rpl33/Stul). Abbildung 9 zeigt die Validierung des Markers im rpl33-Gen mit Proben verschiedener Pappelarten im Vergleich zu Laubbaumarten anderer Gattungen und Nadelbaumarten. Nur in Pappeln erfolgt der Verdau des PCR-Produktes (187 bp) in 2 kleinere Fragmente (123 bp und 64 bp). Dieser Marker ist hervorragend geeignet, um Pappeln selbst aus komplexeren Mischproben zu identifizieren, da es ein goldener Marker ist. Außerdem erlaubt der Marker die zusätzliche Aussage, ob außer Pappeln eventuell andere Laubbaumarten in der Probe vertreten sind. DNA von Nadelbäumen wird mit diesem Marker nicht amplifiziert (z.B. *Pinus sylvestris* in Abb.9, rechts).

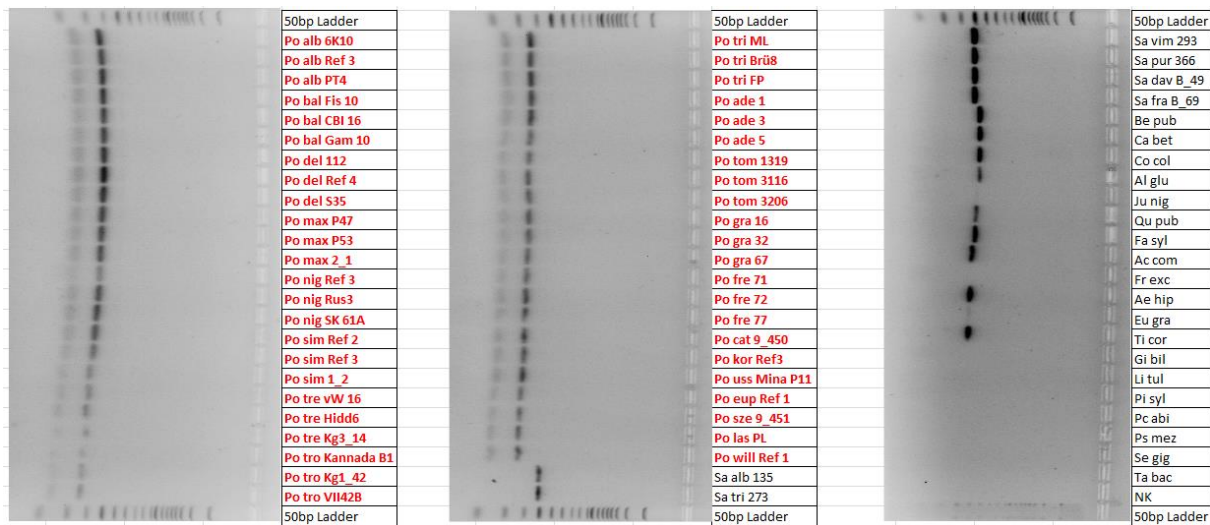


Abb. 9: Validierung eines Gattungsmarkers für *Populus* (3_PO_rpl33/Stul).

Proben verschiedener Pappelindividuen sind in roter Schrift. Nadelbäume sind auf dem rechten Gel unten (ab Pi syl). Alle restlichen Proben sind von verschiedenen Laubbaumarten.

Für die Gattung *Fagus* wurden zwei mitochondriale CAPS-Marker entwickelt (Abb. 4; **Mader, Schroeder et al. 2020**), jeweils basierend auf einem „goldenen“ SNP im matR-Gen (3_Fagus_matR/BstXI) bzw. im ccmFc-Gen (3_Fagus_ccmFc/BsmBI). Nach PCR und Verdau mit dem jeweiligen Restriktionsenzym liefern nur Individuen der Gattung *Fagus* zwei Fragmente (alle anderen zeigen nur das größere ungeschnittene PCR-Fragment). Der ccmFc-Marker ist von besonderem Interesse, da die zugehörigen Primer in Nadel- und Laubbäumen Amplicons liefern, womit eine zusätzliche Aussage darüber möglich ist, ob außer *Fagus* noch Vertreter anderer Gattungen in einer Mischprobe vorhanden sind. Es wurde bereits ausführlich über die bioinformatische Analyse des mitochondrialen Genoms von *Fagus sylvatica* und die Ableitung von Markern auf verschiedenen taxonomischen Ebenen berichtet (3.2). Analysen von Chloroplastengen führten zu einem Marker, der spezifisch für *Fagus sylvatica* ist. Es handelt sich um einen Marker im Chloroplastengen rpoA (1_FA_rpoA/HpyCH4III), bei dem nur *F. sylvatica* mit dem Restriktionsenzym geschnitten wird, während alle anderen untersuchten Arten und Gattungen ungeschnitten bleiben. Mit diesem Marker kann *F. sylvatica* von anderen *Fagus*-Arten (*F. crenata*, *F. engleriana*, *F. grandifolia*, *F. orientalis*) differenziert werden.

Während der Validierung anderer Laubbaumgattungen wurde auch versucht, für die Gattung *Quercus* (Eichen) einen eigenen Gattungsmarker zu entwickeln. Diese Gattung ist jedoch weiteren Gattungen in dieser eng verwandten Gruppe genetisch so ähnlich, dass kein reiner *Quercus*-Marker entwickelt werden konnte. Jedoch ist es gelungen, einen Marker für die Gruppe *Quercus/Castanea/Lithocarpus/Castanopsis* zu entwickeln. Da die drei zuletzt genannten Gattungen für Holzprodukte eher nicht relevant sein werden (evtl. wird manchmal noch Kastanienholz, *Castanea* verwendet), kann dieser Marker zur Identifizierung von Eichenholz in Verbundprodukten verwendet werden. Es wurden zwei Chloroplasten CAPS-Marker entwickelt, die jeweils für die o.a. Gruppe auf einem goldenen SNP basieren. Der eine SNP befindet sich im atpB-Gen (3_QU_atpB/Hpy188I), der andere im atpE-Gen (3_QU_atpE/TaqI) (Anhang 1). In beiden Fällen werden die Individuen der *Quercus*-Gruppe

mit den entsprechenden Restriktionsenzymen in zwei Fragmente geschnitten während alle anderen Taxa ungeschnitten bleiben.

Für die Gattung *Eucalyptus* wurden - basierend auf genischen Alignments plastidärer Gene - verschiedene SNPs identifiziert, die potentiell spezifisch für die Gattung sind. Einer dieser SNPs im *ndhI*-Gen wurde in einen CAPS-Marker umgesetzt und erfolgreich validiert (3_EU_ndhI_1 /*Ddel*). Mit diesem Marker liefern nur Bäume der Gattung *Eucalyptus* zwei Fragmente nach dem Verdau mit *Ddel*, wogegen alle anderen Gattungen das größere ungeschnittene PCR-Fragment auf dem Gelbild zeigen.

Für die Gattung *Fraxinus* sind bei den Auswertungen der ersten Multiplextests zwei Marker aufgefallen, bei denen diese Gattung andere Größen zeigt als alle anderen verwendeten Gattungen. Bei den Markern handelt es sich um den für die Nadel-Laub-Differenzierung entwickelten Marker NL_psbD/*AluI* und einen Marker, mit dem die Familie Fagaceae identifiziert wird, 3_FA_ccmFc/*BsmBI*. Bei beiden handelt es sich um SNPs, die eine Restriktionsstelle für ein Enzym enthalten. *Fraxinus* ergibt jeweils andere Fragmentmuster als die übrigen Taxa. Diese Marker sind sowohl mit frischem Material validiert als auch bereits bei der Validierung des Multiplex-Sets mit Holzproben analysiert worden.

Die beiden Gattungen *Shorea* und *Parashorea* sollen hier zusammengefasst als eine Gruppe betrachtet werden (siehe auch Legende zu Abb. 1). Für diese Gruppe konnten bisher drei spezifische „goldene“ SNP-Marker identifiziert werden, basierend auf SNPs in Chloroplastengen (3_Shorea_rpl36, 3_Shorea_accD sowie 3_Shorea_ycf1). Da uns bisher keinerlei biologisches Material dieser Gruppe für weitere Markervalidierungen zur Verfügung stand, wurden die identifizierten SNP-Marker durch Vergleich mit 34 öffentlich zugänglichen unassemblierten *Shorea* und *Parashorea* NGS-Datensätzen verifiziert (SRA PRJNA419625; Heckenhauer et al. 2019). Ein Probenscreening mit diesen drei SNP-Markern kann nur mit PCR und Sanger-Sequenzierung erfolgen, da die drei SNPs nicht zu einer Veränderung einer Restriktionsschnittstelle führen. Erste Validierungen mit Holzmaterial sind durchgeführt worden und werden weiter fortgesetzt.

Außerdem ergaben bioinformatische Analysen (Kapitel 3.2.) eine spezifische Länge von 151 bp für den intergenischen cp-Linker *psbH_petB* in der Gattungsgruppe *Shorea/Parashorea/Hopea*. Entsprechende Primer für ein Längenscreening mittels PCR wurden design, und die experimentelle Testung mit Holzmaterial ist demnächst geplant.

3.3.5 Identifizierung von Subgattungen, Sektionen und Arten innerhalb der Gattung *Pinus*

Für die Gattung *Pinus* sollen auch für die Identifizierung verschiedener Arten Marker bereitgestellt werden. Die Gattung *Pinus* umfasst weltweit über 110 Arten, von denen wir bereits im Antrag einige ökonomisch wichtige erwähnt haben, für die eine Artidentifizierung notwendig und sinnvoll ist. Da sich die *Pinus*-Arten genetisch nur wenig unterscheiden, war es sinnvoll, zunächst Marker für taxonomische Ebenen zwischen Gattungs- und Art-Ebene, wie die Subgattung (Subgenus) und die Sektion zu entwickeln und innerhalb dieser Subgattungen und Sektionen dann in einem nächsten Schritt die Arten zu differenzieren.

Für die Differenzierung der taxonomisch niedrigeren Ebenen der Subgattung und Sektion hat sich das cp-Gen *rpoC2* als sehr nützlich erwiesen. Die beiden **Subgattungen *Pinus*** und

Strobis sind voneinander durch einen Längenunterschied von 9 bp zu differenzieren (Pinus = 236 bp, Strobis = 227) und von allen anderen Pinaceae durch einem Unterschied von 3 bzw. 6 bp (alle anderen = 230 bp) (2_PI_rpoC2/Ins). Für die übrigen Nadelhölzer und die Laubholzgewächse ergeben diese Primer kein Amplifikationsprodukt. Zusätzlich kann die **Subgattung Strobis** mit Hilfe des Gens psbF ebenfalls durch einen Längenunterschied von 9 bp (Strobis = 136 bp, alle anderen = 127 bp) von allen anderen Arten innerhalb der Gattung *Pinus* unterschieden werden (2_PI_F2frag_4/Ins) sowie mit einem weiteren Längenpolymorphismus in einem der Chloroplasten-Fragmente (2_PI_F1frag_2 LP) (Abb. 10, Anhang 1).

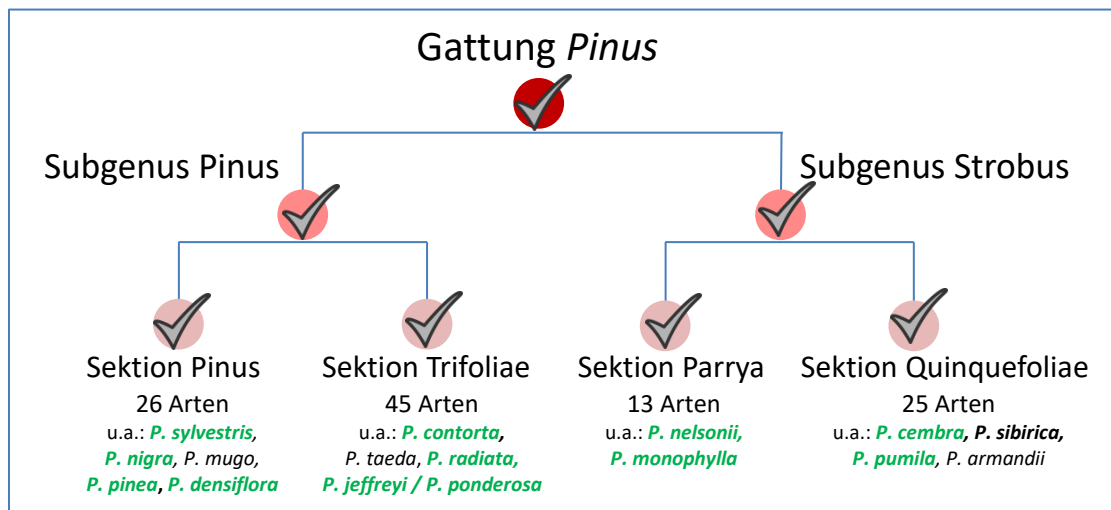


Abb. 10: Taxonomie und Markerentwicklung innerhalb der Gattung *Pinus*.

Dargestellt sind Subgattungen (Subgenera) und Sektionen der Gattung *Pinus* mit den für das Projekt als relevant ausgewählten Arten (fettgedruckt; in grün dargestellt sind Arten, für die Marker entwickelt wurden). Alle taxonomischen Ebenen, für die bereits mindestens 1 Marker entwickelt wurde, sind mit einem grauen Häkchen markiert.

Auch für die Identifizierung der **Sektion Pinus** kann das Gen rpoC2 verwendet werden. Ein Amplifikationsprodukt wird auch in diesem Fall nur für die Familie Pinaceae erzeugt. Für diese Identifizierung wird ein SNP verwendet, für den eine Ansatzstelle eines Restriktionsenzym vorhanden ist. Alle Pinaceae werden in diesem Fall geschnitten, nur die Sektion Pinus nicht (2_PI_rpoC2/*Bfal*). Auch die weiteren **Sektionen Trifoliae** (2_PI_F1frag_1 LP) in der Subgattung Pinus und **Quinquefoliae** sowie **Parrya** (2_PI_F2frag_3 LP) in der Subgattung Strobis können anhand von Längenpolymorphismen identifiziert werden (Abb. 10, Anhang 1).

Für drei der Subsektionen innerhalb der Subgattung Pinus wurden Längenpolymorphismen in den Chloroplastenfragmenten gefunden: 1_PI_F2frag_1 LP differenziert die **Subsektion Pinus** von allen anderen und anhand des Fragments 1_PI_F0frag_2 LP unterscheiden sich die **Subsektionen Australes** und **Trifoliae** von den übrigen Subsektionen (Abb. 10, Anhang 1).

Auch auf der Artebene konnten gute Erfolge bei der Markerentwicklung erzielt werden. So sind für die Arten *Pinus sylvestris* und *P. densiflora* (beide 1_PI_F2frag_1 LP) sowie *P. pinea* (1_PI_F0frag_2 LP; 2_PI_F1frag_2 LP) je einer bzw. zwei Marker für die Differenzierung

dieser Arten von allen anderen untersuchten *Pinus*-Arten entwickelt worden. Bei allen vier Markern handelt es sich um Längenpolymorphismen, die auf dem Sequenzierer sehr gut darstellbar sind. Ein weiterer Längenpolymorphismus erlaubt die fast 100%ige Identifizierung von ***Pinus cembra*** (1_PI_F0frag_3 LP; Abb. 4). Bei der Validierung zeigten einige wenige Exemplare von ***P. sibirica*** dieselbe Länge wie *P. cembra* (Anhang 1).

Ein CAPS-Marker wurde für die Identifizierung von ***Pinus pumila*** (1_PI_psaA/PspGI) entwickelt, bei dem *P. pumila* zweimal geschnitten wird, also drei Fragmente zeigt, während alle anderen Nadelgehölze nur einmal geschnitten werden und demnach 2 Fragmente zeigen (Abb. 10, Anhang 1).

Für die Identifizierung von Arten innerhalb der Gattung *Pinus* geriet die Strategie, „goldene“ Marker auch über höhere taxonomische Ebenen hinaus zu finden, an ihre Grenzen. Daher wurde inzwischen zusätzlich nur innerhalb der Gattung *Pinus* in Chloroplastengen nach Art-spezifischen Unterschieden gesucht. Hierfür wurden umfangreiche Sequenzen aus NCBI für diverse *Pinus*-Arten verwendet. Im *rbcL*-Gen konnten auf diesem Weg zusätzliche Art-spezifische SNPs für ***P. pumila*** und ***P. cembra/P. sibirica*** gefunden werden, sowie weitere SNPs, die Art-spezifisch für ***P. nigra***, ***P. contorta*** und ***P. radiata*** sowie gemeinsam für die beiden Arten ***P. jeffreyi/P. ponderosa*** sind. Außerdem sind in diesem Gen weitere SNPs für die Unterscheidung von Arten innerhalb der Sektion *Strobus* enthalten wie ***P. monophylla*** und ***P. nelsonii***. Im *matK*-Gen konnten zwei zusätzliche SNPs für die Identifizierung von ***P. nigra*** und ***P. contorta*** gefunden werden (Anhang 1). Für die Artidentifizierung innerhalb der Gattung *Pinus* wird ein Abschnitt des *rbcL*-Gens sequenziert und nicht auf PCR-RFLPs zugegriffen.

3.4. Ergebnisse zu Multiplex-Markerset (Teilziel 4)

Basierend auf den bisher entwickelten Markern sollten mehrere Multiplex-Markerset entwickelt werden, die es ermöglichen, die im Projekt relevanten Gattungen in Holzmischprodukten völlig unbekannter Zusammensetzung zu identifizieren. Das erste Multiplex-Markerset ist für die Ordnung Fagales und alle darin befindlichen niedrigeren taxonomischen Gruppen entwickelt und auch an Holzproben validiert worden. Dies wurde ursprünglich für den ABI entwickelt, ist aber auch bereits für die Anwendung auf dem Beckmann umgestellt worden, d.h. es sind neue Primer mit den entsprechenden anderen Farben bestellt und getestet worden. Eine neue Zusammenstellung der Marker war nicht notwendig.

In diesem Set sind insgesamt 11 von 13 verfügbaren Markern enthalten, mit deren Hilfe die Ordnung Fagales, die Familien Fagaceae, Betulaceae und Juglandaceae sowie die Gattungen *Fagus*, *Betula* und *Quercus/Castanea* und außerdem die Art *Fagus sylvatica* identifiziert werden können (Anhang 1: Spalte „Fagales“ und daneben stehende taxonomische Gruppen). Die Marker sind nach einigen Tests in zwei Multiplex-Sets aufgeteilt worden, die alternativ angewendet werden können, so dass mit jedem der zwei Sets alle o.a. angegebenen Taxa identifiziert werden können (Tab. 4). Für jeden Marker ist jeweils der *forward* oder der *reverse* Primer mit einem Fluoreszenzfarbstoff markiert worden, um die Fragmente auf den Sequenzierern sichtbar zu machen (Tab. 4, Abb. 11). Die Namen der Farbstoffe und die jeweilige Markierungsfarbe sind in Tabelle 4 aufgeführt.

Tab. 4: Zusammenstellung der zwei alternativen Fagales-Multiplex-Sets mit den Farbstoffen für Anwendung auf dem ABI (oben) bzw. den Beckmann (unten).

Set 1 (ABI)		Set 2 (ABI)	
Taxa	Marker	Taxa	Marker
Fagales	5_FA_ccmFc_FAM	Fagales	5_FA_matR_NED
Fagaceae	4_FA_ccmB_NED	Fagaceae	4_FA_nad7_PET
<i>Fagus</i>	3_FA_ccmFc_VIC	<i>Fagus</i>	3_FA_matR_PET
<i>Quercus/Castanea</i>	3_QU_atpB_VIC	<i>Quercus/Castanea</i>	3_QU_atpE_FAM
Betulaceae	4_BE_ndhB_FAM	Betulaceae	4_BE_ndhB_FAM
<i>Betula</i>	3_BE_rpoC2_PET	<i>Betula</i>	3_BE_rpoC2_PET
<i>Fagus sylvatica</i>	1_FA_rpoA_FAM	<i>Fagus sylvatica</i>	1_FA_rpoA_FAM
Set 1 (Beckmann)		Set 2 (Beckmann)	
Fagales	5_FA_ccmFc_Dy682	Fagales	5_FA_matR_Cy5
Fagaceae	4_FA_ccmB_Cy7	Fagaceae	4_FA_nad7_Dy682
<i>Fagus</i>	3_FA_ccmFc_Dy682	<i>Fagus</i>	3_FA_matR_Cy7
<i>Quercus/Castanea</i>	3_QU_atpB_Cy5	<i>Quercus/Castanea</i>	3_QU_atpE_Dy682
Betulaceae	4_BE_ndhB_Cy5	Betulaceae	4_BE_ndhB_Cy5
<i>Betula</i>	3_BE_rpoC2_Cy7	<i>Betula</i>	3_BE_rpoC2_Cy5
<i>Fagus sylvatica</i>	1_FA_rpoA_Cy5	<i>Fagus sylvatica</i>	1_FA_rpoA_Cy5

In Abbildung 11 ist ein Ausschnitt der im Fagales-Multiplex-Set verwendeten Marker, angewendet auf dem ABI Sequenzierer, dargestellt. Beispielhaft sind zwei Vertreter aus der Ordnung Fagales (*Quercus* und *Fagus*) und ein Vertreter aus einer anderen Ordnung (*Fraxinus*) abgebildet.

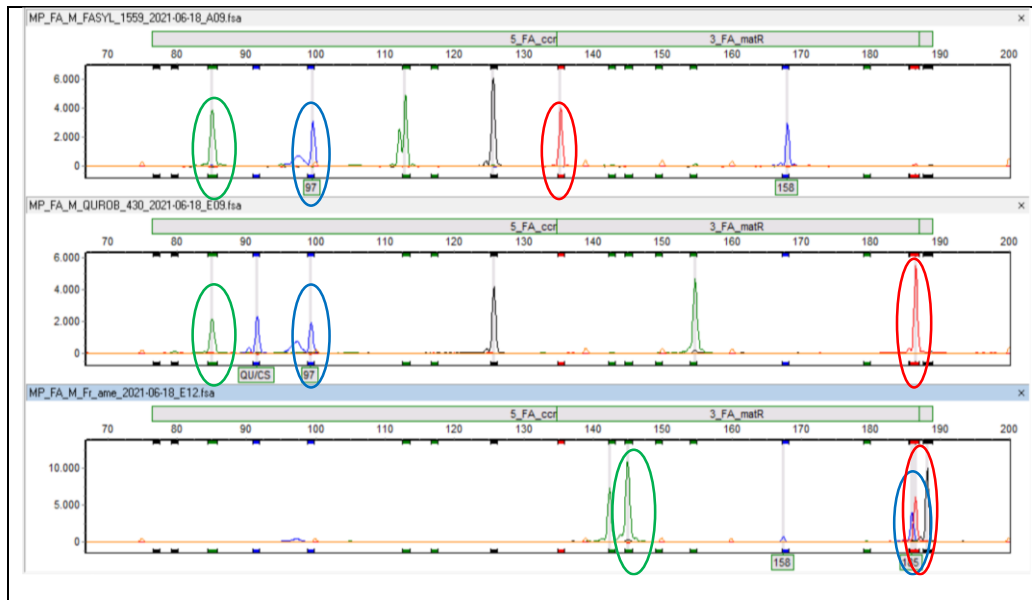


Abb. 11: Ausschnitt aus dem Multiplex-Markerset zur Identifizierung von Vertretern der Ordnung Fagales.

Das Markerset besteht aus den „goldenen“ Markern für die jeweils genannte taxonomische Kategorie (Tab. 4) und erlaubt deren eindeutige Identifizierung in Holzmischprodukten. Allein die drei farblich eingekreisten Marker erlauben anhand der gewonnenen Fragmentmuster eine klare Differenzierung zwischen *Fagus sylvatica* („FASYL“; oben), *Quercus robur* (QUROB; mittig) und *Fraxinus americana* („Fr_ame“; unten), einem Vertreter einer anderen Ordnung (Ordnung Laminales).

3.5. Ergebnisse zu den Praxistests (Teilziel 5)

Insgesamt erhielten wir 30 Sperrhölzer und eine MDF Probe von GD Holz-Mitgliedern. Siebzehn dieser Proben, bei denen Laubbaumarten deklariert waren, wurden mit dem Fagales Multiplex-Markerset getestet. Für sechs dieser Proben, bei denen nur jeweils eine Gattung/Art deklariert war, konnten als Bestandteile Birke bzw. Pappel bestätigt werden und es waren auch keine weiteren Arten enthalten (Tab. 5). Bei drei weiteren Proben wurden die deklarierten Arten (Birke bzw. Buche) zwar bestätigt, aber es wurden auch andere Gattungen gefunden. Bei zwei weiteren Proben wurden einige der deklarierten Arten nicht gefunden und drei Proben sind für die detaillierte Gattungsbestimmung noch in Bearbeitung. Nur für drei Proben konnten nicht ausreichend Marker amplifiziert werden, um eine Aussage treffen zu können (Tab. 5)

Tab. 5: Deklaration und Ergebnis der Markertests für 17 Sperrhölzer der GD-Holz Mitglieder.

FG-Nr.	deklariert	bestätigt	Zusätzliche Gattungen
FG_314_2	<i>Betula</i>	<i>Betula</i>	Nein
FG_314_3	<i>Populus</i>	<i>Populus</i>	Nein
FG_314_4	<i>Populus</i>	<i>Populus</i>	Nein
FG_314_6	<i>Betula</i>	<i>Betula</i>	Nein
FG_314_7	<i>Fagus</i>	<i>Fagus</i>	Ja
FG_315_3	<i>Populus</i> spp, <i>Eucalyptus globulus</i>	Zu wenig Marker erfolgreich	
FG_315_4	<i>Populus</i> spp, <i>Eucalyptus globulus</i>	Salicaceae	<i>Eucalyptus</i> muss noch überprüft werden
FG_315_5	<i>Betula pendula</i> , <i>Alnus glutinosa</i>	Betulaceae (<i>Alnus</i>), aber nicht <i>Betula</i>	Nein
FG_315_7	<i>Betula pendula</i> , <i>Populus</i> spp, <i>Eucalyptus globulus</i>	<i>Betula</i> und andere (<i>Eucalyptus</i>)	Test auf <i>Populus</i> steht noch aus
FG_315_8	<i>Populus</i> spp, <i>Eucalyptus globulus</i>	Salicaceae, keine anderen (kein <i>Eucalyptus</i>)	Test auf <i>Populus</i> steht noch aus
FG_315_9	<i>Quercus</i>	Zu wenig Marker erfolgreich	
FG_315_10	<i>Betula pendula</i> , <i>Betula pubescens</i>	<i>Betula</i>	Ja
FG_315_11	<i>Betula pendula</i> , <i>Betula pubescens</i>	<i>Betula</i>	Ja, andere Betulaceae
FG_315_12	<i>Fagus</i>	Zu wenig Marker erfolgreich	
FG_315_14	Diverse Arten	Nicht <i>Betula</i> , <i>Populus</i> , <i>Quercus</i> (obwohl u.a. deklariert)	
FG_317_1	<i>Betula pendula</i>	<i>Betula</i>	Nein
FG_317_2	<i>Betula pendula</i>	<i>Betula</i>	Nein

Die Deklaration für elf Proben beliefen sich ausschließlich auf Nadelhölzer, zum Teil mit Artangaben innerhalb der Gattungen *Pinus* bzw. *Picea*, aber bei drei Proben war nur die Gattung deklariert. Diese Überprüfungen sowohl auf die Gattungen als auch auf die Arten innerhalb der Gattung *Pinus* sind derzeit noch in der Bearbeitung. Für die verbleibenden

zwei Sperrholzproben waren tropische Hölzer deklariert, die mit unseren Markern nicht überprüfbar sind. MDF ist generell ein sehr stark verarbeitetes Material. Daher konnten wir für diese Probe keine ausreichende DNA Qualität für Überprüfungen gewinnen.

4. Öffentlichkeitsarbeit / Veröffentlichungen / Vorträge

Öffentlichkeitsarbeit:

- Fernsehbericht zum Thema „Holz“ am 28. November 2018, WDR, Planet Wissen; <https://www.planet-wissen.de/sendungen/sendung-holz-108.html>; mit Beteiligung des Thünen-Instituts für Forstgenetik, Holzherkunftsidentifizierung
- Besuch der Firma „Formafantasma“ am 19.09.2019 für Filmaufnahmen/Interviews über das Holzkompetenzzentrum für eine Ausstellung in London (04. März bis 17. Mai 2020).
- Stand in der BMEL-Halle zum Kompetenzzentrum Holzherkünfte auf der Internationalen Grünen Woche, 17.01. bis 26.01.2020, in Berlin.
- Die Ausstellung der Firma „Formafantasma“ in London (04. März bis 17. Mai 2020), für die am 19.09.2019 Filmaufnahmen und Interviews vom / über das Holzkompetenzzentrum gemacht wurden, ist Corona-bedingt leider abgesagt worden.
- Am 17. Juli 2021 wurde der nachfolgende Artikel online veröffentlicht, der über die Arbeit des Thünen-Kompetenzzentrums für Holzherkünfte berichtet: „Germany’s timber detectives take on illegal loggers“; <https://atlasofthefuture.org/project/centre-of-competence-on-the-origin-of-timber/>
- Am 10. März 2022 wurde der Artikel "The 'timber detectives' on the front lines of illegal wood trade - Germany's Center of Competence on the Origin of Timber tracks down which EU wood imports come from illegal sources in the world's third-biggest criminal sector" in National Geographic veröffentlicht: <https://www.nationalgeographic.com/environment/article/the-timber-detectives-on-the-front-lines-of-illegal-wood-trade>

Veröffentlichungen:

Akhmetzyanov L, Copini P, Sass-Klaassen U, **Schroeder H**, de Groot GA, Laros I, Daly A (2020) DNA of centuries-old timber can reveal its origin. Scientific Reports 10:20316. doi: 10.1038/s41598-020-77387-2

Braga WB, Deklerck V, Espinoza E, Groening M, Koch G, Monteiro Pastore TC, Ramanantoandro T, **Schröder H**, Watkinson C, Wiedenhoef AC, van Brusselen J, Bolanos J, Schmitz N (2020) Scientific methods for taxonomic and origin identification of timber [online]. GTTN (Global Timber Tracking Network), 6 p, zu finden in <<https://www.researchgate.net/publication/342003654>> [zitiert am 11.08.2021], [DOI:10.13140/RG.2.2.28416.46087](https://doi.org/10.13140/RG.2.2.28416.46087)

Bruegmann T, Fladung M, **Schroeder H** (2022). Flexible DNA isolation procedure for different tree species as a convenient lab routine. *Silvae Genetica* 71: 20-30

- Bünger K** (2021) Validierung und Anwendung molekularer Marker zur Holzarten-Identifizierung in Holzverbundprodukten. Fakultät für Maschinenbau und Bioverfahrenstechnik der Hochschule Hannover. Bachelorarbeit von Kathrin Bünger (angefertigt beim Thünen-Institut für Forstgenetik im Rahmen von Holz-DNA-Barcoding; 23.02.2021)
- Degen B**, Blanc-Jolivet C, Bakhtina S, Yanbaev R, Yanbaev Y, Mader M, Nürnberg S, **Schroeder H** (2021) Applying targeted genotyping by sequencing with a new set of nuclear and plastid SNP and Indel loci for *Quercus robur* and *Quercus petraea*. Conservation Genetics Resources 13: 345-347. doi: 10.1007/s12686-021-01207-6
- Kersten B**, Rellstab C, **Schroeder H**, Fladung M, Krutovsky K, Gugerli F (2022) The mitochondrial genome sequence of *Abies alba* Mill. reveals a high structural and combinatorial variation. BMC Genomics 23:776. doi: 10.1186/s12864-022-08993-9
- Kersten B**, Rellstab C, Gugerli F (2022) *Abies alba* isolate AA_WSL01 mitochondrion, scaffold sequences: NCBI, accession numbers ON378818 → ON378828
- Kersten B**, **Schott T**, Mader M (2020) *Fagus sylvatica* isolate FASYL_29_1 mitochondrion, complete genome: NCBI, accession MT446430 (version MT446430.1), DNA sequence, 504715 bp; <https://www.ncbi.nlm.nih.gov/nuccore/MT446430>
- Mader M, **Schroeder H**, **Schott T**, **Schöning-Stierand K**, Leite Montalvão AP, Liesebach H, Liesebach M, Fussi B, **Kersten B** (2020) Mitochondrial genome of *Fagus sylvatica* L. as a source for taxonomic marker development in the Fagales. Plants 9(10), 1274. doi: 10.3390/plants9101274 (This article belongs to the Special Issue Genetics, Genomics and Biotechnology of Plant Cytoplasmic Organelles).
- Mader M, **Kersten B** (2020) Mitochondrial genome of *Fagus sylvatica* L. as a source for taxonomic marker development in the Fagales: NCBI SRA, accession PRJNA648273, 4 SRA Experiments, 3 BioSamples; DNA-Seq, mRNA-Seq, 111 Gbases; <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA648273/>
- Mader M, Liesebach H, Liesebach M and **B Kersten** (2019) The complete chloroplast genome of *Fagus sylvatica* L. (Fagaceae). Mitochondrial DNA Part B 4(1): 1818-1819. doi: 10.1080/23802359.2019.1612712
- Schmitz N, Beeckman H, Cabezas JA, Cervera MT, Espinoza E, Fernandez-Golfín J, Gasson P, Hermanson JC, Arteaga MJ, Koch G, Lens F, Martínez-Jarquín S, Paredes-Villanueva K, Pastore TCM, Ramanantoandro T, Schraml R, **Schröder H**, Sebbenn AM, Tysklind N, Watkinson C, et al (2019) The Timber Tracking Tool Infogram: overview of wood identification methods' capacity. GTTN, 5 p. doi: 10.13140/RG.2.2.27920.25603
- Schoening-Stierand K**, **Schroeder H**, **Degen B**, **Kersten B** (2019) Identification of tree species in wood composite products by DNA barcoding. In: Adamowicz SJ, Ekrem T, Stur E, Bakken T, Martin MD (eds). Scientific abstracts from the 8th International Barcode of Life Conference; Trondheim, Norway; 17 to 20 June 2019. Genome 62 (6): 431-432. doi: 10.1139/gen-2019-0083
- Schott T**, **Schroeder H**, **Schöning-Stierand K**, **Kersten B** (2019) The complete chloroplast genome sequence of *Pinus cembra* L. (Pinaceae). Mitochondrial DNA Part B 4 (2): 4202-4203. doi: 10.1080/23802359.2019.1693297
- Schroeder H**, **Kersten B** (2022) A small set of nuclear markers for reliable differentiation of the two closely related oak species *Quercus robur* and *Q. petraea*. Manuscript in preparation

Präsentationen:

In der folgenden Tabelle 6 sind sowohl allgemeinere Vorträge zur Holzherkunftsidentifizierung mit Bezug zum Projekt als auch spezielle Vorträge und Poster, die direkt mit dem Projekt korreliert sind und in denen Projektergebnisse präsentiert wurden, aufgeführt.

Tab. 6: Vorträge/Poster von Wissenschaftlern zu dem Thema und zu den Arbeiten des Projekts. Präsentierender Autor ist jeweils fettgedruckt

Datum	Veranstaltung	Ort	Autoren	Thema
15.03.2018	Außenhandelstag des GD Holz	Bremen, Deutschland	Hilke Schröder , Birgit Kersten, Bernd Degen	Entwicklung genetischer Marker zur Gattungs- und Arterkennung in Holzverbundprodukten mittels Next Generation DNA-Barcoding (Vortrag)
24.04. bis 26.04.2018	Deutsche Baumpflegetage	Augsburg, Deutschland	Hilke Schröder , Bernd Degen	Einsatz molekularer Marker zur Art- und Herkunftsbestimmung von Bäumen und Holz (Poster)
29.07.2018	Jahrestreffen der Deutschen Dendrologischen Gesellschaft	Ahrensburg, Deutschland	Bernd Degen	Genetische Identifizierung von Baumart und Baumherkunft (Vortrag)
08.10. bis 09.10.2018	Holzhandel und Walderhaltung – fünf Jahre Erfahrungen mit der EU-Holzhandelsverordnung und dem Thünen-Kompetenzzentrum Holzherkünfte	Hamburg, Deutschland	Hilke Schröder , Céline Blanc-Jolivet, Bernd Degen	Anwendung genetischer Methoden zur Bestimmung von Holzart und -herkunft (Vortrag)
21.01. bis 22.01.2019	„A Bed of Roses – Royal or Revival“ Symposium	London, Großbritannien	Hilke Schröder	The Bed of Roses – DNA analysis (Vortrag)
23.05.2019	Besuchergruppe “Rantzau-Haus” Ahrensburg	Großhansdorf, Deutschland	Hilke Schröder , Bernd Degen	Anwendung genetischer Methoden zur Bestimmung von Holzart und –herkunft (Vortrag)
17.06. bis 20.06.2019	Barcode of Life conference	Trondheim, Norwegen	Hilke Schröder , Katrin Schöning-Stierand Bernd Degen, Birgit Kersten	Identification of tree species in wood composite products (“Lightning talk” und Poster)
06.02.2020	Besuch Oberstufen-Schülergruppe aus Hamburg-Niendorf	Großhansdorf, Deutschland	Hilke Schröder , Bernd Degen	Anwendung genetischer Methoden zur Bestimmung von Holzart und –herkunft (Vortrag)
13.09. bis 16.09.2021	Forstwissenschaftliche Tagung 2021	digital	Hilke Schröder , Birgit Kersten, Bernd Degen	Genetische Marker zur Gattungs- und Arterkennung in Holzverbundprodukten (Vortrag)
03.12.2021	Workshop des „Thünen-Kompetenzzentrum Holzherkünfte“	digital	Hilke Schröder , Céline Blanc-Jolivet, Bernd Degen	Entwicklung genetischer Methoden zur Bestimmung von Holzart und -herkunft (Vortrag)
09.03.2022	Thünen Themenfeldgespräch „Buche“	digital	Hilke Schröder	Thünen-Kompetenzzentrum Holzherkünfte – Prüfaufträge bei der Buche (Vortrag)

06.04.2022	Besuch Oberstufen-Schülergruppe aus Hamburg-Nienstedten	Großhansdorf, Deutschland	Hilke Schröder	Anwendung genetischer Methoden zur Bestimmung von Holzart und –herkunft (Vortrag)
19.04. bis 21.04.2022	“From Forests to Heritage” Conference	Amsterdam, Niederlande	Hilke Schröder	Genetic analysis of „The Bed of Roses” (Vortrag)

Aufgrund der Corona-Pandemie fanden ab Anfang 2020 kaum wissenschaftliche Tagungen statt, auf denen die Ergebnisse hätten präsentiert werden können. Außerdem konnten auch keine Besuchergruppen empfangen werden.

5. Fazit / Ausblick

Die im Projekt erfolgreich entwickelten und validierten Marker und Multiplex-Markersets erlauben eine effiziente genetische Erkennung vieler Gattungen und Arten von Hölzern, die in Holzverbundprodukten am häufigsten eingesetzt werden und fließen direkt in die praktische Anwendung im Rahmen des Thünen-Kompetenzzentrum Holzherkünfte ein. Wichtige Beiträge für die Praxis sind dabei auch die optimierten Protokolle zur DNA-Extraktion und die gewonnenen Ergebnisse zur Eignung verschiedener Markertypen für Mischhölzer aus Nadel und Laub. Die mittels NGS und bioinformatischer Auswertung erzeugten neuen Chloroplasten und mitochondrialen Referenzgenome, z.B. von *Fagus sylvatica*, bieten eine gute Grundlage für die Entwicklung weiterer genetischer Marker, deren Validierung durch die im Projekt gesammelten Referenzproben unterstützt wird. Außerdem können diese Referenzgenome in zukünftigen Studien für andere wissenschaftliche Fragestellungen verwendet werden. Das im Projekt generierte Wissen zu Taxon-spezifischen DNA-Varianten könnte auch in potentielle zukünftige Ansätze zum Metabarcoding einfließen.

6. Weitere Literaturangaben

Degen B, Höltken AM (2011) DNA-Methoden zur Kontrolle der Holzart und Holzherkunft. Holz-Zentralblatt 19:461.

Heckenhauer J, Paun O, Chase MW, Ashton PS, Kamariah AS, Samuel R (2019) Molecular phylogenomics of the tribe Shoreeae (Dipterocarpaceae) using whole plastid genomes. Ann Bot 123(5):857-865.

Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, et al. (2009) A DNA barcode for land plants. Proc Natl Acad Sci U S A 106: 12794–12797.

Kress WJ, Erickson DL (2008) DNA barcodes: genes, genomics, and bioinformatics. Proc Natl Acad Sci U S A, 105 (8), 2761-2762

Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V (2008) DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci U S A 105 (8), 2923-2928

Lowe AJ, Jardine DI, Cross HB, Degen B, Schindler L and AM Hoeltken (2015) A method of extracting plant nucleic acids from lignified plant tissue. Patent WO/2015/070279 filed 2014-11-14 and issued 2015-05-21.

- Mader M, Schroeder H, Schott T, Schöning-Stierand K, Leite Montalvão AP, Liesebach H, Liesebach M, Fussi B and B Kersten (2020) Mitochondrial genome of *Fagus sylvatica* L. as a source for taxonomic marker development in the Fagales. *Plants* 9(10):1274
- Mosca E, Cruz F, Gómez-Garrido J, Bianco L, Rellstab C, Brodbeck S, Csilléry K, Fady B, Fladung M, Fussi B, Gömöry D, González-Martínez SC, Grivet D, Gut M, Hansen OK, Heer K, Kaya Z, Krutovsky KV, Kersten B, Liepelt S, Opgenoorth L, Sperisen C, Ullrich KK, Vendramin GG, Westergren M, Ziegenhagen B, Alioto T, Gugerli F, Heinze B, Höhn M, Troggio M and DB Neale (2019) A reference genome sequence for the European silver fir (*Abies alba* Mill.): a community-generated genomic resource. *G3: Genes, Genomes, Genetics* 9(7):2039-2049
- Taberlet P, Coissac E, Pompanon F et al. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* 35: e14
- Xiufang S and K Canby (2010) China: Overview of forest governance, market and trade. *Forest Trends for FLEGT Asia regional program*, European Forest Institute (EFI)

Anhang

Anhang 1: Übersicht der verfügbaren Marker auf verschiedenen taxonomischen Ebenen

Ordnung	Familie	Gattung/Art
Coniferales NL_cox1_F1/R1 SNP NL_cox1_F2_R2 SNP NL_petA_F1/R1 SNIP NL_psbD_F1/R1 SNP	Pinaceae 4_PI_mttb_F1/R1 LP	<i>Pinus</i> 3_PI_rpoC2_F1/R1 SNP; 3_PI_rpoC2_F3/R3_SNP
		<i>Picea</i> 3_PC_rpoC2_F1/R1 SNP; 1_PI_F0frag_3 LP
		<i>Abies</i> 3_AB_cemA_F1/R1 SNP; 3_AB_matK_F1/R1 SNP
		<i>Larix</i> 2_PI_rpoC2_F1/R1 SNP
		<i>Pseudotsuga</i> 1_PI_F2frag_1 LP; 2_PI_F1frag_1 LP; 1_PI_F0frag_2 LP
		<i>Cedrus</i> 1_PI_F0frag_3 LP
	Cupressaceae 4_CU_nad6_F1/R1 SNP	<i>Cryptomeria</i>
		<i>Chamaecypari</i>
		<i>Sequoiadendron</i>
		<i>Sequoia</i>
		<i>Metasequoia</i>
	Taxaceae 4_TA_matK_F1/R1 SNP 4_TA_matK_F2/R2 SNP	<i>Taxus</i> 3_TA_atpB_F1/R1 SNP <i>T. baccata</i> 1_TA_Indel1 LP; 1_TA_Indel2 LP 1_TA_nad4L SNP
		<i>Torreya</i>
	Fagales 5_FA_matR_F1/R1 SNP 5_FA_ccmFc_F1/R1 SNP	Fagaceae 4_FA_ccmb_F1/R1 SNP 4_FA_nad7_F1/R1 SNP
<i>Fagus</i> 3_FA_matR_F1/R1 SNP 3_FA_ccmFc_F1/R1 SNP <i>F. sylvatica</i> 1_FA_rpoA_F1/R1 SNP		
<i>Betula</i> 3_Betula_matK SNP 3_Betula_rpoC2_Ins		
Betulaceae 4BE_rpoB_F1/R1 SNP 4BE_ndhB_F1/R1 LP		<i>Alnus</i>
		<i>Carpinus</i>
		<i>Corylus</i>
Juglandaceae 5_FA_ccmFc_F1/R1 SN		<i>Juglans</i>
		<i>Carya</i>
Nothofagaceae		<i>Nothofagus</i>

Malpighiales 5_MP_matR F1/R1 SNP	Salicaceae 4_SA_ccmb_F1/R1 LP 4_BA_petB_1 SNP 4_FA_nad7_F1/R1 SNP	Populus 3_PO_rpl33_F1/R1 SNP
		<i>Salix</i>
Sapindales	Sapindaceae	<i>Acer</i>
		<i>Aesculus</i>
Malvales 5_MA_BR_ccmFc_F1/R1 LP	Malvaceae	<i>Tilia</i>
	Dipterocarpaceae	Shorea
Lamiales	Oleaceae	Fraxinus NL_psbD_F1/R1 SNP 3_FA_ccmFc_F1/R1 SNP
Fabales	Fabaceae	Robinia
Myrtales 5_MY_matR_F1/R1 SNP	Myrtaceae	Eucalyptus 3_EU_ndhl_F1/R1 SNP

Sub-Gattung	Sektion	Sub-Sektion	Art
Pinus 2_PI_rpoC2_LP	Pinus 2_PI_rpoC2_SNP	Pinus 1_PI_F2frag_1 LP	<i>P. sylvestris</i> 1_PI_F2frag_1 LP
			<i>P. densiflora</i> 1_PI_F2frag_1 LP
		Pinaster	<i>P. pinea</i> 2_PI_F1frag_2 LP 1_PI_F0frag_2 LP
	Trifoliae 2_PI_F1frag_1 LP	Australes 1_PI_F0frag_2 LP	<i>P. radiata</i> in rbcL (innerhalb Gattung Pinus)
			Contortae 1_PI_F0frag_2 LP
		Ponderosae	<i>P. jeffreyi/P. ponderosa</i> in rbcL (innerhalb Gattung Pinus)
Strobis 2_PI_F2frag_4 LP 2_PI_F1frag_2 LP	Parrya 2_PI_F2frag_3 LP	Balfourianae	<i>P. monophylla</i> <i>P. nelsonii</i> in rbcL (nur innerhalb Gattung Pinus)
	Quinquefoliae 2_PI_F2frag_3 LP	Strobis	<i>P. pumila</i> 1_PI_psaA_F1/R1 SNP in rbcL (innerhalb Gattung Pinus)
			<i>P. cembra/(P. sibirica)</i> 1_PI_F0frag_3 LP in rbcL (innerhalb Gattung Pinus)

In dunkelgrün und fett sind die bereits validierten Marker angegeben.