

Quantitative trait locus (QTL) analysis of ash dieback in *Fraxinus excelsior*

Dissertation

To achieve the degree of

Doctor rerum naturalium (Dr. rer. nat.)

Faculty of Mathematics, Informatics and Natural Sciences

Department of Biology

University of Hamburg

Submitted to the University of Hamburg by

Melina Krautwurst

Hamburg 2024

Thesis Committee:

Supervisor:

Prof. Dr. Julia Kehr

University of Hamburg

Co-supervisor:

PD Dr. Birgit Kersten

Thünen Institute of Forest Genetics

Other members:

PD Dr. Cornelia Heinze

University of Hamburg

Prof. Dr. Sigrun Reumann

University of Hamburg

Disputation: 22. November 2024

This dissertation is submitted as a cumulative thesis based on the work contained in two (2) accepted peer-reviewed articles and two (2) in preparation for submission. The articles are organized in a thematic order in the following four main chapters, with contributions to the quoted articles and chapters detailed accordingly:

Chapter 1

Melina Krautwurst*, Franziska Past*, Birgit Kersten, Ben Bubner & Niels A. Müller (2024). Identification of full-sibling families from natural single-tree ash progenies based on SSR markers and genome-wide SNPs. *Journal of Plant Diseases and Protection* (acceptance letter received, 02.07.2024)

- Coordination of results with project partners
- Mapping and variant calling of SNP markers
- Variant filtering
- Data visualization
- Analysing results
- Writing – original draft
- Writing – final draft

*equal contribution

Chapter 2

Melina Krautwurst, Birgit Kersten & Niels A. Müller. Construction and comparative analysis of high-density linkage maps in *Fraxinus excelsior* (in preparation).

- Mapping and variant calling of SNP markers
- Variant filtering
- Linkage map construction
- Analysing results
- Data visualization
- Writing – original draft

Chapter 3

Melina Krautwurst, Birgit Kersten, Franziska Past, Ben Bubner, & Niels A. Müller.
QTL analysis of ash dieback (in preparation).

- Analysing phenotypic traits
- QTL analysis
- Analysing results
- Data visualization
- Writing – original draft

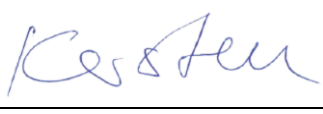
Chapter 4

Melina Krautwurst, Annika Eikhof, Sylke Winkler, Daniel Bross, Birgit Kersten & Niels A. Müller (2024). High-molecular-weight DNA extraction for broadleaved and conifer tree species. *Silvae Genetica* 73, 85–98. doi: 10.2478/sg-2024-0009

- Testing kits, ingredients and methods
- Optimizing the procedure for *Fraxinus excelsior*
- Coordination of results with project partners
- Oxford Nanopore Sequencing
- Data visualization
- Writing – original draft
- Writing – final draft

X 

Prof. Dr. Julia Kehr, Hamburg den 16.07.2024

X 

PD Dr. Birgit Kersten, Großhansdorf den 16.07.2024

Eidesstattliche Versicherung:

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Affidavit:

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated.

If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plagiarism that may have been generated by the gAI.

Hamburg, den

Unterschrift

This work was carried out from October 2020 to Juli 2024 at the Thünen Institute of Forest Genetics, Großhansdorf, Germany, in the Genome Research Group of Dr. Niels A. Müller. The FraxForFuture research network is funded by the German Federal Ministry of Food and Agriculture and the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety. The projects of the sub-networks are funded by the Waldklimafonds (funding code: 2219WK21A4) and the Thünen Institute. The project executing agency is the Fachagentur für Nachwachsende Rohstoffe e.V. (FNR).

‘Not all hope is gone’ – in memory of Eike Schöning

Index

Index	VII
List of figures	X
List of abbreviation	XI
Abstract	1
Kurzfassung	3
General introduction	5
Common ash and the ash dieback disease	5
Stem collar necrosis	9
Genetic variation in susceptibility to ash dieback	10
Quantitative Trait Locus (QTL) analysis for species with long generation times	11
Comparison of genetic with physical maps for the identification of structural variants	15
Long read sequencing for high-quality reference genomes of forest trees	16
Objectives and hypotheses	17
References	18
Identification of full-sibling families from natural single-tree ash progenies based on SSR markers and genome-wide SNPs	5
Abstract	5
Introduction	6
Material and Methods	7
Plant material	7
Sampling, DNA extraction and SSR genotyping	9
COLONY analysis	10
DNA extraction and Illumina low-coverage resequencing	11
Mapping and variant calling	12
Variant filtering	12
Data visualization	13
Results	13
Discussion	17
Acknowledgements	19
Competing interests statement	19
References	20
Supplementary	24
Construction and Comparative Analysis of High-Density Linkage Maps in <i>Fraxinus excelsior</i>	39
Abstract	40
Introduction	41
Material and Methods	42
Family identification, DNA extraction and Illumina low-coverage resequencing	42
	VII

Mapping and variant calling	43
Variant filtering	43
Segregation filtration	44
Linkage map construction	44
Comparison of genetic maps with the physical map	46
Results	46
Discussion	50
Conclusion	52
References	53
Supplementary	55
<u>QTL analysis of ash dieback</u>	67
Abstract	68
Introduction	68
Material and Methods	70
Family identification, variant calling, and linkage map construction	70
Phenotypic traits	70
QTL analysis	71
Structural and functional annotation of the QTL region	73
Results	74
Discussion	84
Conclusions	85
References	85
Supplements	88
<u>High-molecular-weight DNA extraction for broadleaved and conifer tree species</u>	95
<u>Abstract</u>	96
Introduction	96
Materials and methods	99
Biological samples	99
Reagents, equipment and supplies for HMW DNA extraction	99
Procedure for nuclei isolation and HMW DNA extraction	101
DNA extraction	107
Estimating DNA fragment size distribution through gel electrophoresis	108
Characterizing DNA fragment size distribution through precision analysis with the Agilent FEMTO Pulse System	109
Oxford Nanopore Sequencing Setup	109
PacBio high-fidelity sequencing setup	110
Results	110
HMW DNA concentration and quality	110
Fragment size distribution of the HMW DNA of <i>Fraxinus excelsior</i>	111
Fragment size distribution of the HMW DNA of <i>Taxus baccata</i>	112
Discussion	116
Limitations and critical steps in the procedure	116
Quality of plant material	116

Quantity of plant material	117
Grinding	117
Application of PVP	118
Centrifuge speed	118
DNA 'jellies'	119
Quality and quantity of DNA	119
Conclusion	120
References	121
Supplements	124
General Discussion	132
Genetic variation of plant disease resistance	132
Do we understand the genomic basis of ash dieback?	138
Conclusions and perspectives	146
References	150
Acknowledgements	156

List of figures

Figure 1: Historical distribution of ash dieback in Europe_____	7
Figure 2: Effects of ash dieback on <i>Fraxinus excelsior</i> and infection of <i>Hymenoscyphus fraxineus</i> _____	8
Figure 3: Distribution of genetic markers across individuals_____	12
Figure 4: Classical crossing in comparison to pseudo-backcross_____	4
Figure 5: Examples for the identification of chromosomal mutations using genetic and physical maps_____	15
Figure 6: Overview of future directions and enhancements of the existing research results_____	158

List of abbreviation

ADB	Ash dieback
BwB	Breeding-without-Breeding
BLASTP	Protein Basic Local Alignment Search Tool
CCS	Consensus sequence
CDS	Coding sequences
Chr	Chromosome
cM	CentiMorgan
DBH	Diameter at breast height
DNA	deoxyribonucleic acid
EUFROGEN	European Forest Genetic Resources Program
<i>F. excelsior</i>	<i>Fraxinus excelsior</i>
Full-sib	Full-siblings
GATK	Genome Analysis Toolkit
GBS	Genotyping-by-sequencing
GO	Gene ontology
GS	Genomic selection
<i>H. albidus</i>	<i>Hymenoscyphus albidus</i>
<i>H. fraxineus</i>	<i>Hymenoscyphus fraxineus</i>
HB	Homogenization Buffer Stock
HMW	High-molecular-weight
LG	Linkage group
MAMPs	Microbe-associated molecular pattern
MAS	Marker-assisted selection
NaCl	Natriumchlorid
NCBI	National Center for Biotechnology Information
NIB	Nuclei Isolation Buffer
ONT	Oxford Nanopore Technologies
PacBio	Pacific Biosciences
PADRE	Pathogen and abiotic stress response domain
PAMPs	Pathogen-associated molecular patterns
PCR	Polymerase chain reaction
pH	Potential of hydrogen
PRR	Pattern recognition receptors
PVP	Polyvinylpyrrolidone
QTL	Quantitative trait locus
RIL	Recombinant inbred line
SCN	Stem collar necrosis
SMRT	Single-molecule real-time
SNP	Single-nucleotide polymorphism
SSR	Simple sequence repeat
SV	Structural variants
TSB	Triton-Sucrose-Buffer
VCF	Variant Call Format

Abstract

Ash dieback, caused by *Hymenoscyphus fraxineus*, threatens common ash (*Fraxinus excelsior*) across Europe. It leads to a severe increase in the mortality rate and losses in natural stands. Multiple studies have shown that susceptibility is a heritable trait and that genotypes with low susceptibility exist. Resistance genes often underlie larger genomic patterns that must be identified to understand the susceptibility mechanism within the tree fully. This thesis aims to investigate the genomic basis of ash dieback susceptibility, attempting to support genetic and genomic conservation strategies for *F. excelsior*.

To achieve this, the identification and genotyping of full-sibling families within four progenies from potentially tolerant mother trees have been conducted. Using SSR markers and the COLONY program, full-sibling families were predicted and subjected to high-resolution genotyping with genome-wide SNP markers. This approach identified five large families suitable for genetic linkage mapping and quantitative trait locus (QTL) analyses.

Further, the focus was on constructing high-density genetic linkage maps from the full-sibling families using the SNP markers and the Lep-MAP3 program. These maps were compared with the chromosome-level reference genome assembly and revealed high collinearity and notable differences, such as the split in chromosome 2 and the merging of chromosomes 22 and 23. Moreover, individual structural variants were predicted. These findings underscore the utility of genetic mapping in enhancing reference genome assemblies.

To investigate the genomic basis of ash dieback and potentially connected genes, the genetic linkage maps were applied in a genome-wide QTL analysis on pseudo-backcross full-sibling families to identify loci associated with stem collar necrosis (SCN) as a key phenotype. One QTL for SCN was identified, explaining 15% of the phenotypic variance and highlighting the polygenic nature of susceptibility. Additionally, with the QTL locus range, a gene encoding protein PADRE domain-containing protein connected to fungus response was annotated.

To obtain a higher resolution of the genetic map, verify the predicted structural variants and establish a pan-genome of *F. excelsior* for future association studies, long-read sequencing is essential. High-molecular-weight DNA is the key to achieve long-read sequencing. A protocol for extracting high-molecular-weight DNA from *F. excelsior* was developed for this. Based on nuclei isolation and DNA extraction using the Nanobind® plant nuclei kit by Pacific Biosciences, this protocol yielded high-quality DNA suitable for Oxford Nanopore Technologies and PacBio sequencing. The study underscores the complexities and adaptations required for successful high-molecular-weight DNA extraction in tree species.

Kurzfassung

Das Eschentriebsterben, verursacht durch *Hymenoscyphus fraxineus*, bedroht die Gemeine Esche (*Fraxinus excelsior*) in ganz Europa. Dies führt zu einem starken Anstieg der Sterblichkeitsrate und Verlusten in natürlichen Beständen und städtischen Räumen. Zahlreiche Studien haben gezeigt, dass die Anfälligkeit eine erblich bedingte Eigenschaft ist und dass Genotypen mit geringer Anfälligkeit existieren. Resistenzgene liegen oft größeren genomischen Mustern zugrunde, die identifiziert werden müssen, um den Anfälligkeitsmechanismus im Baum vollständig zu verstehen. In dieser Dissertation wird die genomische Grundlage der Anfälligkeit für das Eschentriebsterben untersucht, um genetische und genomische Naturschutz- und Erhaltungsstrategien für *F. excelsior* zu unterstützen.

Zu diesem Zweck wurden die Identifizierung und Genotypisierung von Vollgeschwisterfamilien innerhalb von vier Nachkommenschaften potenziell toleranter Mutterbäume durchgeführt. Mit Hilfe von SSR-Markern und dem Programm COLONY wurden Vollgeschwisterfamilien vorhergesagt und einer hochauflösenden Genotypisierung mit SNP-Markern unterzogen. Die in diesem Ansatz identifizierten vier Familien, eignen sich für die genetische Kopplungskartierung und die Analyse quantitativer Merkmale (QTL).

Weiterhin lag der Fokus auf der Erstellung genetischer Kopplungskarten aus den Vollgeschwisterfamilien unter Verwendung der genomweiten SNP-Marker und des Programms Lep-MAP3. Diese Karten wurden mit der Chromosomen-Referenzgenomassemblierung verglichen und zeigten eine hohe Kollinearität sowie bemerkenswerte Unterschiede, wie die Teilung von Chromosom 2 und die Zusammenführung der Chromosomen 22 und 23. Zudem wurden individuelle strukturelle Varianten ermittelt. Diese Ergebnisse unterstreichen den Nutzen von genetischen Kartierung zur Verbesserung der Referenzgenomassemblierungen und tragen zum Verständnis der genomischen Basis der Anfälligkeit für das Eschentriebsterben bei.

Um die genomische Basis und potenziell gekoppelte Gene der Anfälligkeit von Eschentriebsterben, wurden genetischen Kopplungskarten in einer genomweiten QTL-Analyse an ‚pseudo-backcross‘ Vollgeschwisterfamilien angewandt. Hierdurch

können Loci, die im Zusammenhang mit dem Stammfußnekrosen (SCN) stehen, bestimmt werden. Ein QTL für SCN wurde identifiziert, der 15% der phänotypischen Varianz erklärt und die polygenetische Natur der Anfälligkeit hervorhebt. Zusätzlich wurde im Bereich des QTL-Locus ein Gen annotiert, das ein Protein mit einer PADRE-Domäne kodiert, das mit der Pilzabwehr in Verbindung steht.

Um eine höhere Auflösung der genetischen Karte und der strukturellen Varianten des Genoms zu erreichen und ein Pan-Genom für *F. excelsior* für zukünftige Studien zur Verfügung zu stellen, ist Long-Read-Sequenzierung unerlässlich. Hochmolekulare DNA ist der Schlüssel zu Long-Read-Sequenzierung. Ein Protokoll zur Extraktion hochmolekularer DNA aus *F. excelsior* wurde hierfür entwickelt. Das Protokoll basiert auf Zellkern Isolierung und der DNA-Extraktion mit dem Nanobind® Plant Nuclei Kit von Pacific Biosciences. Das Ergebnis ist DNA von hoher Qualität, die für die Sequenzierung mit Oxford Nanopore Technologies und PacBio geeignet ist. Die Studie unterstreicht die Komplexität und notwendige Anpassungen, die für eine erfolgreiche Extraktion hochmolekularer DNA bei unterschiedlichen Baumarten erforderlich sind.

General introduction

Forest ecosystems are of primary importance since they provide diverse ecosystem services essential for the economic, social and general well-being of human populations. In the Anthropocene, disturbances such as the rapid climate change and biological invasions seriously threaten different ecosystems. Ecosystem services provided by forests, such as climate regulation through carbon sequestration, habitat for wildlife or water cycle regulation are crucial for maintaining ecological balance in the future. New pests and pathogens cause destruction, fragmentation, habitat loss, and species decline in forests worldwide (Panzavolta et al., 2021; Guégan et al., 2023).

Various methods can be implemented in forest management to combat pests, with genomic toolsets offering promising approaches. Advances in statistical and computational tools, along with an increasing number of case studies, have enhanced these possibilities (Neophytou et al., 2022). This progress is fuelled by the availability of high-quality genomic resources and affordable genotyping methods (Feng and Du, 2022). Understanding the biological interactions and genetic adaptation of tree species, which have long generation times, in response to rapidly evolving pathogens is a real challenge. Genetic tools such as microsatellites, SNP datasets, genetic mapping and other methods are instrumental in addressing gaps in our knowledge for practical conservation (Holliday et al., 2017). By identifying the specific genetic polymorphisms associated with important complex traits, such as pathogen resistance, we can analyze their frequencies and understand their effects (Mackay, 2001; Swalarsk-Parry et al., 2022). While pathogen resistance is a trait with a complex genetic basis, different studies have demonstrated that it is possible to resolve this complexity and contribute valuable insights to our understanding (Lespinasse et al., 2000; Butler et al., 2016; Bartholomé et al., 2020; Fan et al., 2024).

Common ash and the ash dieback disease

Common ash, *Fraxinus excelsior* L. (Oleaceae), is one of ~50 species within the *Fraxinus* genus. It is the only native German ash species. *F. excelsior* colonised

northern Europe during the Cretaceous and spread southwards during periods of climate cooling (Marigo et al., 2000). The deciduous tree is known for the black buds on the shoots, which sit on the end of the stouts. Young trees are characterised by pale grey and smooth bark, which becomes thick and vertically cracked with age. Adult trees reach a height between 12-30 m; the age can go up to 250 years (Thomas, 2016). Flower production begins after 20-30 years. The species behaves as subdioecious; flowers can be fully female, fully male or hermaphroditic, though hermaphroditic flowers are either functionally female or male (Saumitou-Laprade et al., 2017). Seeds develop from July to September, and the samara (flattened wings) form clusters that hang in bunches. The seeds can be dispersed across distances of 100 m (Dobrowolska et al., 2011; Thomas, 2016). The ash tree holds significant ecological and economic value, playing a vital role in supporting biodiversity and providing high-quality wood for various industries. Ash is the host of 44 obligatorily associated species, 62 highly associated species, and partially associated species (Mitchell et al., 2016). Further, ash wood is highly valuable for furniture and construction due to its flexibility (Burggraaf, 1972).

F. excelsior has a diploid genome with 23 pairs of chromosomes ($2n = 46$). The genome of *F. excelsior* was first sequenced and assembled by Sollars et al. (2017). The genome sequencing utilised a combination of short-read sequencing technologies HiSeq 2000 (Illumina, San Diego, California, USA) along with optical mapping to improve the assembly's accuracy and contiguity. The second reference genome was published by Kazimierz Wielki University in Poland in 2021 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_019097785.1/). The Polish reference genome was assembled using a combination of advanced sequencing technologies. These include long-read sequencing by Pacific Bioscience (Sequal) and high-coverage short-read sequencing (Illumina NovaSeq) and Hi-C mapping. This hybrid approach ensures a more accurate and contiguous assembly. Further, a new reference genome from a less susceptible ash tree was published (10.07.2024 www.ashgenome.org; <http://oadb.tsl.ac.uk>). It is the third reference genome. One of the primary motivations for sequencing the *F. excelsior* genome was to understand the genetic basis of ongoing threats by introduced pests and pathogens,

that is the emerald ash borer in North America and the fungus causing ash dieback in Europe. These initial assemblies provided the basis for genomic studies.

In early 1990, the pathogenic fungus *Hymenoscyphus fraxineus* was first observed in Poland and spread from there across 23 European countries (Figure 1). The establishment of the diseases across Europe was recently confirmed by the National Forest Inventory (Klesse et al., 2021). Moreover, it seems that the range for *H. fraxineus* is further expanding, as it was now identified in north-western Spain as well (Stroheker et al., 2021). It has been reported in Germany for over 20 years (first observed in 2002, proven in 2006).

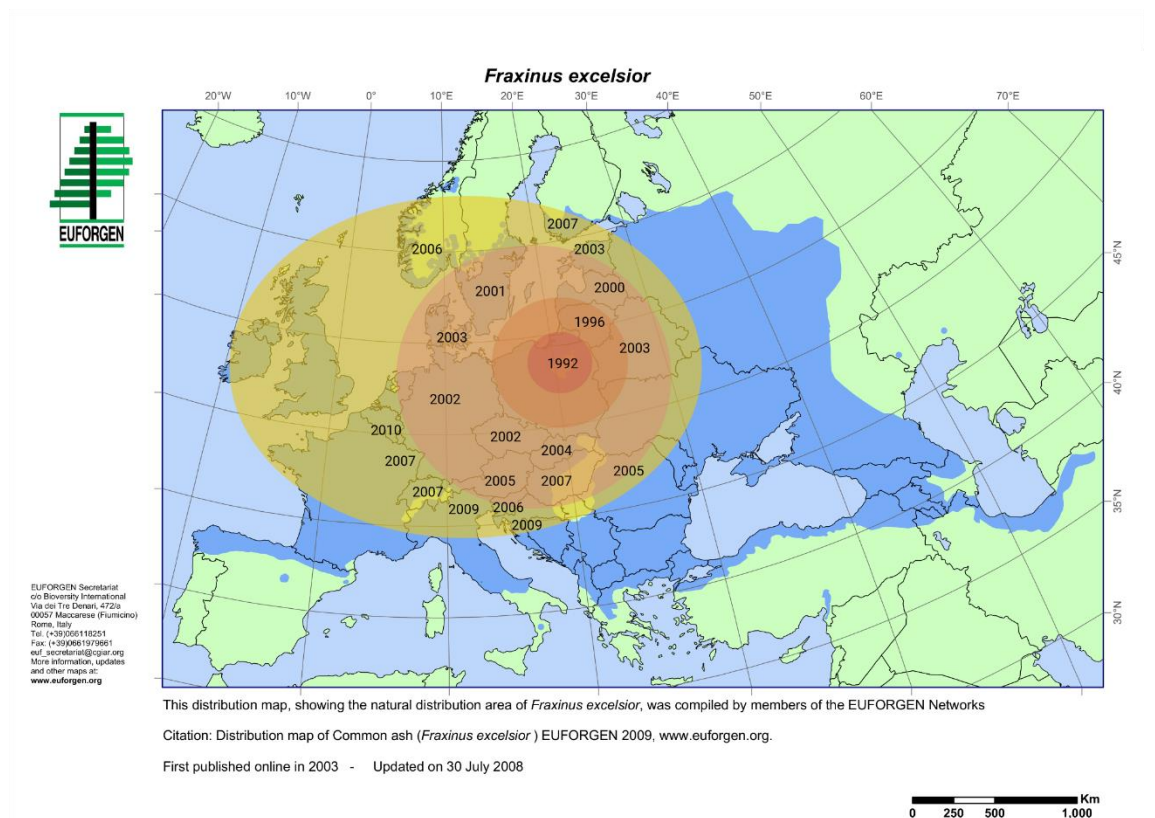


Figure 1: Historical distribution of ash dieback in Europe. The map shows the distribution of *Fraxinus excelsior* (blue) in Europe from the first observation of ash dieback in 1992. The number of years represents the first observation of symptoms of ash dieback in each country (Timmermann et al., 2011). The map is based on (EUFORGEN, 2009).¹ Created with Biorender.com.

The fungus is native to Japan, Korea, east Russia and China. It is a member of the order Helotiales (class *Leotymycetes*, *Ascomycota*) (Langer et al., 2022). On

¹ Copyright: © 2011 Timmermann et al. and This open-access article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

F. excelsior, which did not co-evolve with the fungus, it causes the ash dieback (ADB) disease. This disease leads to a dramatic decline in ash stands in forests and urban areas. The infected trees are impacted by symptoms such as stem collar necrosis or rot and secondary infestation with bark or wood rot fungi, as well as insects, leaf necrosis, premature leaf drop, shoot wilting and crown dieback (Fuchs et al., 2024) (Figure 2). The symptoms compromise growth, wood quality and ultimately often result in mortality. However, a small minority of ash trees remain fully healthy (Bakys et al., 2013; Coker et al., 2019; Davydenko et al., 2022). It is documented that the mortality rate was moderate between 1987 and 2000; from 2010–2020, the mortality rate increased strongly. The differences were mainly caused by the fact

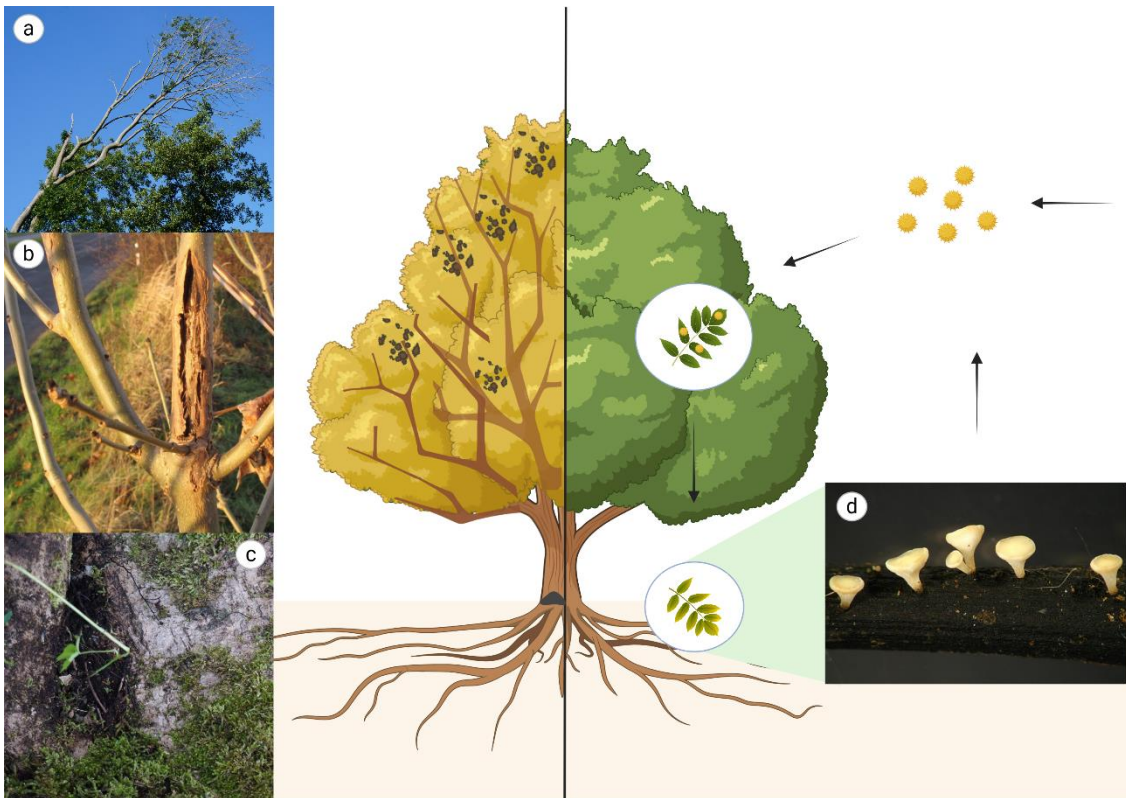


Figure 2: Effects of ash dieback on *Fraxinus excelsior* and infection of *Hymenoscyphus fraxineus*. The left side of the pictogram shows an ash tree with leaf yellowing, crown shedding and stem collar necrosis. Three examples of ash dieback symptoms are (a) crown dieback, (b) lesions caused by necrosis, and (c) stem collar necrosis. The right side of the pictogram shows a healthy ash tree. The ascospores of *Hymenoscyphus fraxineus* infect the leaves and grow over the surface, forming internal structures. (d) The fruiting bodies grow on the leaf litter. Created with BioRender.com. Picture credit: Fraziska Past (b) and Felix Rentschler (d).

that in some countries (e.g. Poland), *F. excelsior* was not assessed in inventories (George et al., 2022). Before ash dieback, common ash was known to have high regeneration capability and was dominant in mixed stands (Langer et al., 2022). In Sweden the status of *F. excelsior* was changed to ‘critically endangered’ already in

2010 (Cleary et al., 2017). George et al. (2022) suggest that *F. excelsior* is currently under extreme extinction risk, especially in northern Europe.

An infection of *F. excelsior* with *H. fraxineus* is primarily caused by ascospores, which are wind-dispersed over large distances. *H. fraxineus* exhibits a pleiomorphic mating system. It has sexual and asexual stages during its disease cycle (Fones et al., 2016). It can have three different phases: saprotrophic (feeding on dead tissue), biotrophic (feeding on live tissue), or necrotrophic (feeding on and killing live tissue) (Lo Presti et al., 2015). By late spring and summer, ascospores spread and enter the trees (Timmermann et al., 2011). Spores germinate into the fungal anamorph, with hyphae developing and entering the plant. The fungus enters through the leaf stomata, but it can also enter through the lenticels (microscopic pores in the bark) (Gross et al., 2012; Nemesio-Gorriz et al., 2019). Furthermore, it was reported that the fungus enters through the epidermis (Mansfield et al., 2019). It then extends into the space between the apoplast and finally enters living host cells. The hyphae further grow into the leaf petiole, spreading into the branches and stem, causing ash dieback (Gross et al., 2012; Mansfield et al., 2019; Carroll and Boa, 2024). In autumn and winter, the fungus enters the teleomorph phase (sexual), where the lifecycle continues in the tree litter. In the litter, the fruiting bodies evolve in the rachis of the leaf. The lifecycle restarts with the release of new ascospores (Schematic illustration in Figure 2). *H. fraxineus* has two mating types, MAT1-1 or MAT1-2. Reproduction thus happens between these two types (Gross et al., 2012).

Stem collar necrosis

This study focuses on Stem collar necrosis (SCN). SCN is the main cause of ash dieback and increases the mortality rates of ash (Häuser et al., 2024). Genotype-phenotype association studies mainly focus on crown symptoms, senescence or overall health status in genetic and genomic studies (Harper et al., 2016; Sollars et al., 2017; Doonan et al., 2023; Meger et al., 2024a; Meger et al., 2024b). In 2023-2024, data published about the biology and associated fungi of SCN increased (Lutz et al., 2023; Peters et al., 2023; Rozsypálek et al., 2023; Häuser et al., 2024; Peters et al., 2024). The knowledge about SCN gives a better understanding of the symptoms, how it acts with its environment, and how less susceptible trees interact. The

genomic basis of the susceptibility to necrosis is unknown and the focus of the QTL analysis in this study. SCN is defined as basal lesions with necrotic tissue on the outside and inside of the stem (Langer, 2017). It is mainly caused by fungi. The path of infection is not fully understood yet (Peters et al., 2023). Environmental conditions can promote infection; humid and strong moisture favours the condition (Marçais et al., 2016). Other fungi species, such as *Armillaria* species, exploit the infected tissue caused by *H. fraxineus*, leading to further tree instability (Chandelier et al., 2016; Lysenko et al., 2024). Severe necrosis is often observed later than severe crown damage on unhealthy trees (Madsen et al., 2021). SCN makes the ash increasingly vulnerable to other infections and pests, a progressive symptom that ultimately leads to the tree's death. Different scoring systems for SCN were developed for adult trees and saplings (Langer et al., 2022). A sapling has a height of up to 2 m and consists of the primary stem, the primary shoot, and side shoots. Since the primary stem has the most important role in the sapling's survival, it is especially important for the survey of SCNs.

Genetic variation in susceptibility to ash dieback

Phenotypic selection of trees with pathogen resistance in traditional breeding programs has shown promising results for other tree species and various pathogens (Alfaro et al., 2013; Sniezko et al., 2020a; Martín et al., 2023). However, ash dieback is a complex disease where genomic and genetic selection may produce more effective results than traditional breeding programs. The primary objective is to identify low-susceptibility *F. excelsior* genotypes with a broad geographical range across different populations. These genotypes can be utilized as breeding material for the reforestation of ash populations. The heritability value for ash dieback resistance is approximately 0.4 (McKinney et al., 2011; Pliura A et al., 2011; Lobo et al., 2014; Enderle et al., 2015; Lobo et al., 2015). Research has shown that low susceptibility to ash dieback is a multifactorial polygenic trait, with several genetic markers identified as being associated with increased tolerance (McKinney et al., 2014; Harper et al., 2016; Sollars et al., 2017; Stocks et al., 2019; Meger et al., 2024a; Meger et al., 2024b). These studies, which are based on non-related individuals, further indicate that the specific genetic population structure impacts the estimated effects of identified variants, precluding universal application outside of the studied

populations (Sollars et al., 2017; Meger et al., 2024a; Meger et al., 2024b; Plumb et al., 2024). Further research is therefore necessary to achieve a more comprehensive understanding of the genetic basis of ash dieback variation and the potential applications of genetic differences.

Quantitative Trait Locus (QTL) analysis for species with long generation times

A Quantitative Trait Locus (QTL) is a genetic locus (region of DNA in the genome) for which allelic variation is associated with the phenotype of a quantitative trait. Generally, quantitative traits are multifactorial and are influenced by several polymorphic genes and environmental conditions, so one or many QTLs can influence a trait or a phenotype. QTL analysis is a statistical method that links two types of information—phenotypic data (trait measurements) and genotypic data (molecular genetic markers)—in an attempt to explain the genetic basis of variation in complex traits (Falconer & Mackay, 1996; Kearsley, 1998; Lynch & Walsh, 1998). QTLs represent molecular differences among individuals in the genomic regions controlling a particular trait exhibiting quantitative inheritance. It can explain a portion of phenotypic variance and thus are particularly suited to analyse traits governed by a few large-effect loci with Mendelian inheritance patterns within controlled cross populations (Rani et al., 2023). The family-based mapping analysis relies on genetic recombination and segregation in the progenies of bi-parental crossings (Ashwath et al., 2023). Genetic linkage maps of segregating populations are necessary for QTL analyses. Genetic maps are constructed using genotype data from numerous related individuals (in this study, full-siblings) at selected markers to ascertain their relative genetic positions. Genotype data enables us to identify recombination events, which are directly related to genetic distances and the sequential order of markers on the chromosome (Cartwright et al., 2007). During meiosis, homologous chromosomes exchange genetic material through crossing over. The likelihood that two genetic markers will be separated by recombination depends on the distance between them: markers that are close together are less likely to be separated than markers that are far apart. The frequency of recombination between different markers is calculated based on how often the markers are inherited together in the offspring. Markers that frequently appear

together are considered to be closely linked. The genetic linkage is measured in centimorgan (cM), representing 1% of the recombination frequency. For example, in Figure 3 a, the blocks of the same colour (either black or white) indicate regions where no recombination has occurred, suggesting genetic linkage. Individual 1 has a long stretch of white circles from 0-50 cM, implying that the region's markers are genetically linked and have been inherited together without recombination. A linkage map is constructed by arranging the markers in an order based on their recombination frequencies.

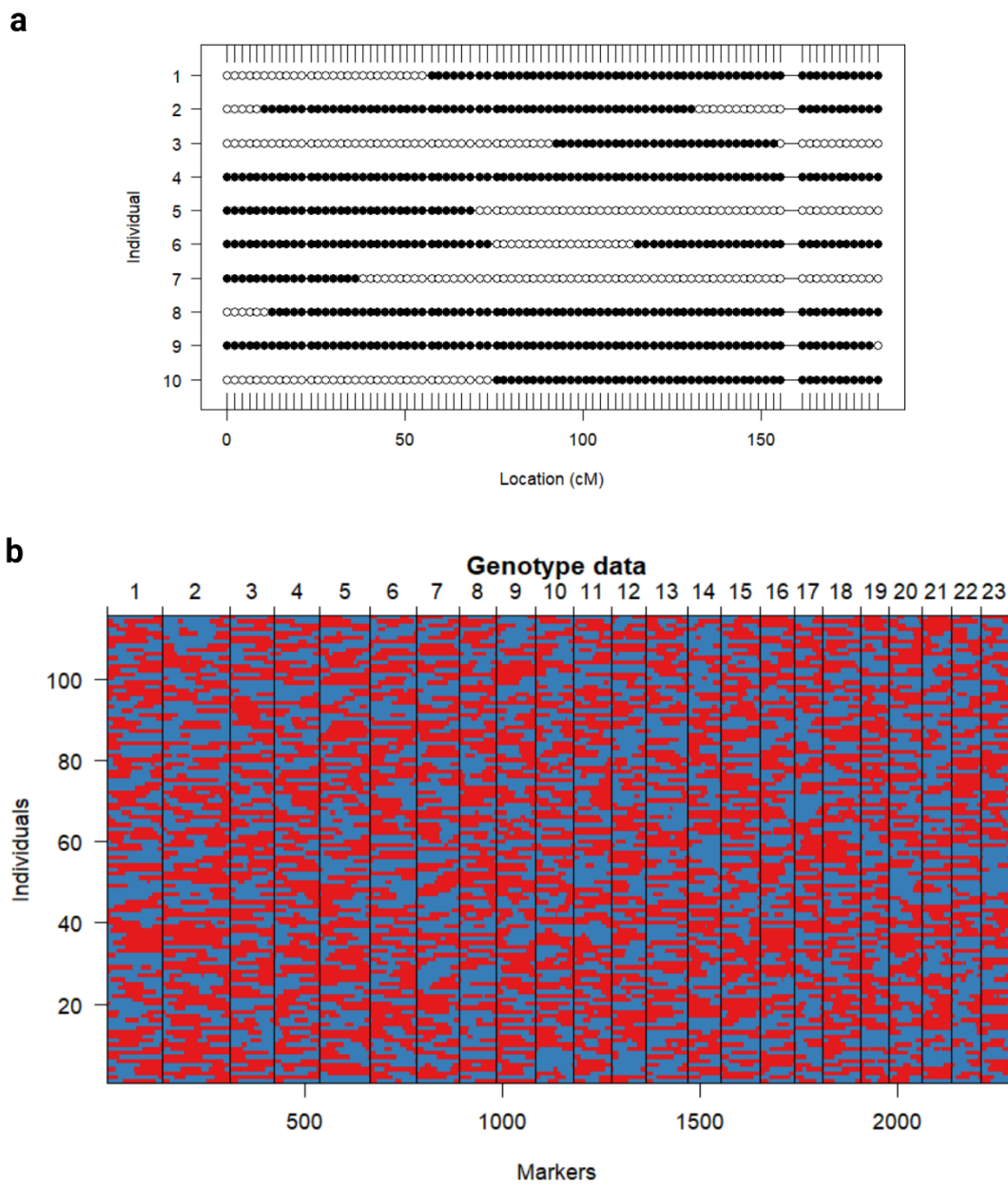


Figure 3: Distribution of genetic markers across individuals. The panels show allele segregation among individuals. Panel a displays genotype distribution on one linkage group for 10 individuals on a paternal map. White and black circles correspond to AA and AB genotypes, respectively. The circles correspond to the presence of different alleles at each marker location. The x-axis indicates the position on the chromosome in centimorgans (cM), while the y-axis lists

the individuals. Panel b shows a graphical representation of genotypes on 23 linkage groups (top) and markers (bottom) of 116 individuals. AA and AB are displayed in red and blue, respectively. Horizontal lines split linkage groups. The distribution and pattern of these colours reveal the genetic variation and recombination events across the genome.

In Figure 3 b, a recombination event can be identified in the full-sibling family, as indicated by the change in colour (red-to-blue). If an individual shows a red-to-blue switch, recombination occurs between those markers. Standard QTL analyses are typically conducted on F_2 , backcross populations, and recombinant inbred line (RIL) families. However, for species with long generation times, such as trees, their production could take decades, given that *F. excelsior*, for example, starts to flower after 20-30 years (Harper et al., 2016). In such cases, the 'Breeding-without-Breeding' (BwB) method can be utilised to perform QTL analysis. BwB enables studies with incomplete pedigrees (e.g., missing father information) or without controlled pollination (El-Kassaby et al., 2011). BwB could be equally named 'crossing-without-crossing' since one of the parents is not homozygous and no F_1 generation is formed. The parents pollinate naturally, and the progenies of dominant fathers are chosen for the study; no real crossing is conducted (Figure 4). The 'crossing' is described as a 'pseudo backcross', resulting in a 1:2:1 segregation ratio of progeny.

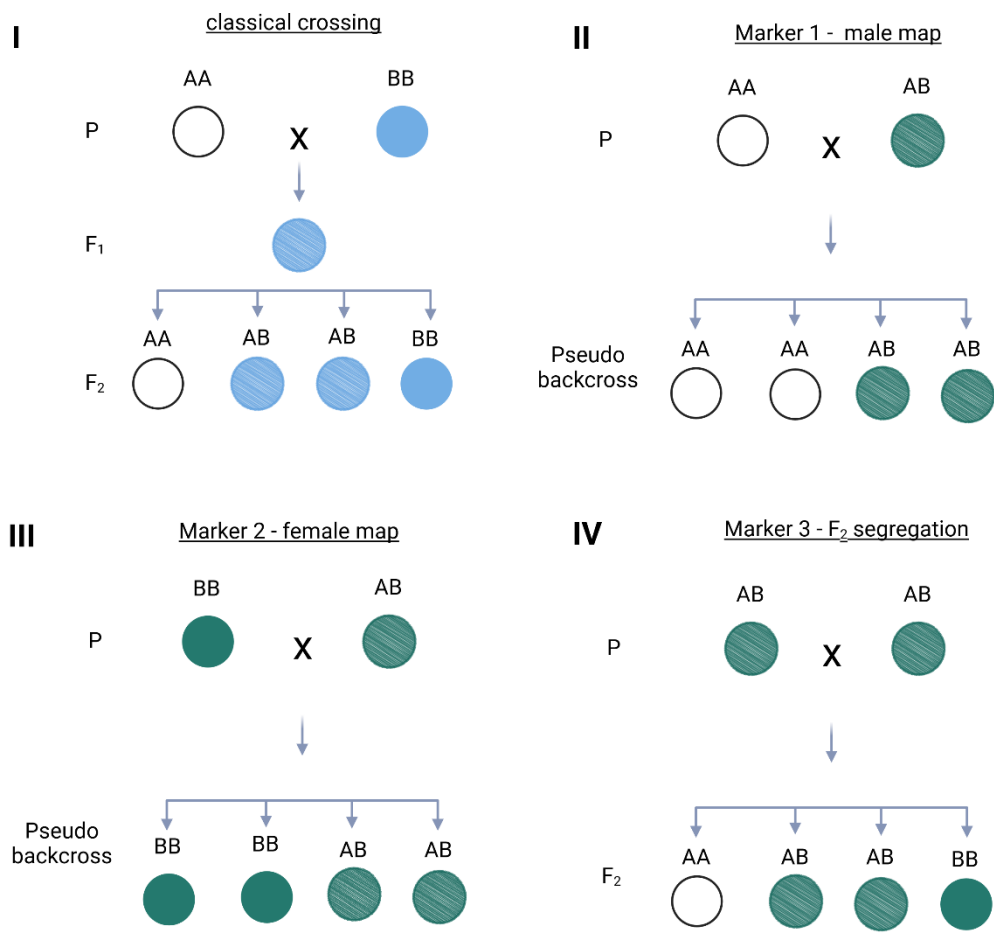


Figure 4: Classical crossing in comparison to pseudo-backcross. Panel I illustrates the classical crossing scheme where two parental lines (P) with genotypes AA and BB are crossed. The resulting F₁ generation consists of individuals with genotype AB. When these F₁ individuals are self-fertilized or crossed among themselves, the F₂ generation exhibits a 1:2:1 segregation ratio. It demonstrates Mendelian inheritance and segregation of alleles. Panels II, III, and IV show pseudo backcross. Panel II displays a homozygous mother and heterozygous father, the progenies exhibit a 1:1 segregation of genotypes AA and AB, demonstrating the inheritance pattern from the male parent. Panel III illustrates a pseudo backcross involving a marker associated with the female parent. The parental lines are BB and AB. The offspring display a 1:1 segregation of genotypes BB and AB, demonstrating the inheritance pattern from the female parent. Panel IV shows the segregation of a marker in the F₂ generation from an initial AB x AB cross. The resulting offspring exhibit a 1:2:1 segregation ratio of AA, AB, and BB genotypes. This panel reinforces the typical Mendelian inheritance and segregation patterns expected in an F₂ population. It shows if the pseudo backcross progenies are used to create an F₂ generation. Created with Biorender.com.

It combines genotypic or phenotypic pre-selection of superior individuals with the use of informative DNA markers for fingerprinting and pedigree reconstruction of offspring. This approach allows the assembly of naturally occurring full-sibling families resulting from mating among selected parents, thereby allowing selection of full-sibling families and QTL analysis (El-Kassaby et al., 2007). Controlled crosses are therefore not needed, strongly reducing the required time and resources.

Comparison of genetic with physical maps for the identification of structural variants

Genetic maps can also be used to assess large structural variants by comparing the genetic map of an individual of interest with the physical map of the reference genome (Choi et al., 2016; Tong et al., 2016; Ashwath et al., 2023). Identified structural variants can be validated using long-read sequencing.

Structural variants are genomic alterations that involve segments of DNA typically larger than 50 base pairs, for our study focuses on larger segments in megabase pair range (Alkan et al., 2011). These variations can significantly impact the structure and function of the genome, potentially leading to phenotypic diversity or disease. Different types of structural variants and chromosomal mutations can be detected. Structural variants include insertions (addition of nucleotides), deletions (loss of nucleotides), duplications (copies of a segment of DNA), inversions (reversal of a segment of DNA within the chromosome), translocations (rearrangement of segments between non-homologous chromosomes) and copy number variants (Feuk et al., 2006) (Figure 5). The changes can have pronounced phenotypic effects (Alonge et al., 2020; Huang and Rieseberg, 2020; Dallinger et al., 2024).

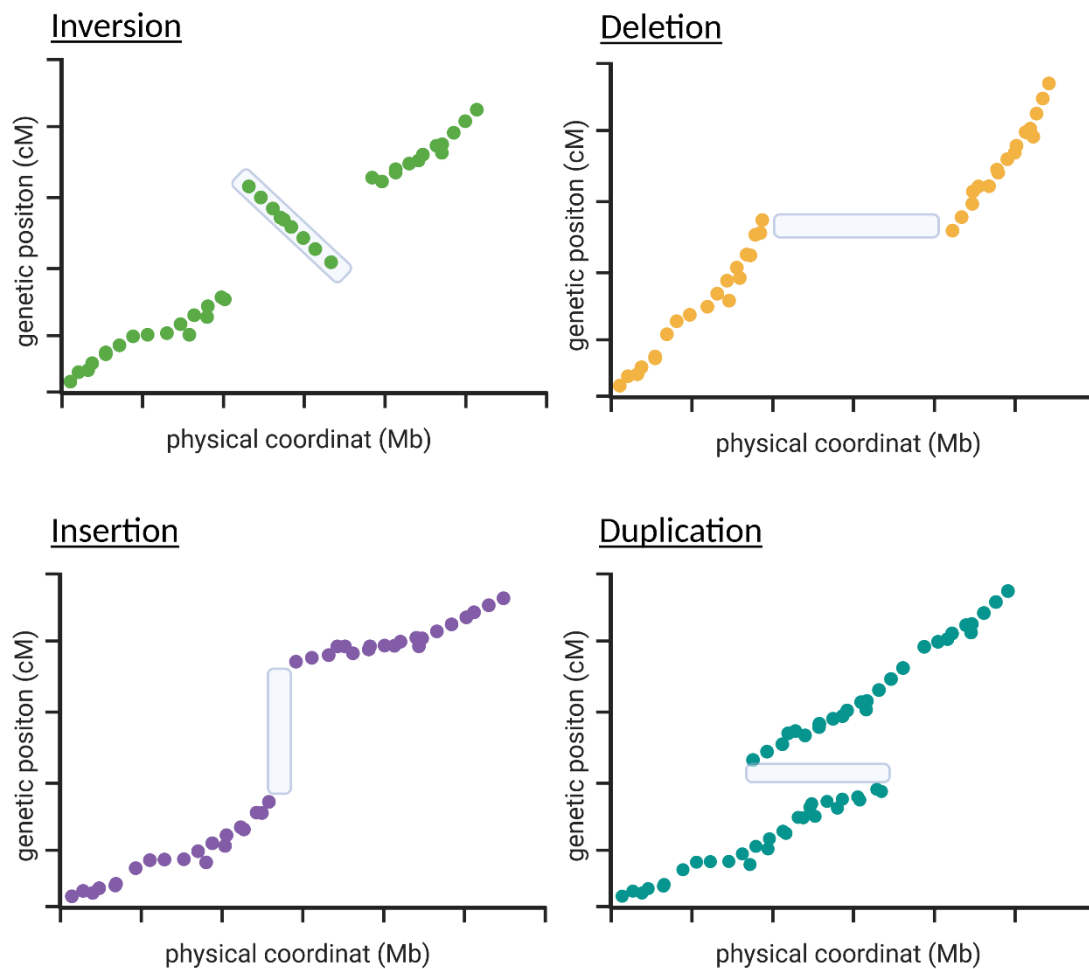


Figure 5: Examples for the identification of chromosomal mutations using genetic and physical maps. This figure illustrates four types of chromosomal mutations (inversion, deletion, insertion, and duplication) by comparing their genetic positions (cM) to their physical coordinates (Mb). Created with Biorender.com.

Long read sequencing for high-quality reference genomes of forest trees

Forest trees, which are mostly non-model species, often lack high-quality chromosome-level reference genomes (Plomion et al., 2016; Street, 2019; X. Yang et al., 2019). There are several challenges associated with sequencing and assembling tree and plant genomes in general, e.g., high levels of heterozygosity, large numbers of repeats and sometimes large genome sizes. High-quality, chromosome-level genomes have become more available with the development of long-read sequencing. Platforms such as the ones offered by Oxford-Nanopore Technologies (ONT) or Pacific Bioscience (PacBio) provide two different approaches for long-read sequencing. The output of the two technologies can be combined. In PacBio sequencing using single-molecule real-time (SMRT) sequencing, a single DNA polymerase enzyme synthesizes a complementary strand of DNA, and this process is observed in real-time. The DNA polymerase incorporates the fluorescently labelled nucleotides, and each incorporation event produces a unique fluorescent

signal that is captured by a detector. The read length ranges between 10,000 and 15,000 base pairs (Nishii et al., 2023). In comparison, ONT uses biological nanopores where the DNA strand is pulled through, and the nucleotides cause a change of ionic current within the pore. The ionic current change is characteristic of specific nucleotides. It is possible to generate long and ultra-long reads (100,000-2,000,000 base pairs), with the possibility for telomer to telomer assemblies (Dumschott et al., 2020). A chromosome-scale reference genome was recently published for *F. excelsior* (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_019097785.1/, 06.07.2021). The assembly includes 23 chromosomes and 415 scaffolds, with a total assembly size of 806.5 Mb. The DNA input for ONT or PacBio sequencing is high-molecular-weight (HMW) DNA. There are different extraction protocols and kits available, which are mostly for the general genre of 'plants' (Ashley Jones et al., 2021; Kang et al., 2023). Extracting HMW DNA from trees is more challenging, and general plant protocols may not work for tree material. Trees are rich in polyphenolic compounds, polysaccharides, tannins, and other secondary metabolites, which complicate extraction (Formato et al., 2022). Long-read sequencing can not only be used to create reference genomes but also to identify and investigate large structural variants, which are increasingly recognized as important drivers of trait variation (Nguyen et al., 2023).

Objectives and hypotheses

The important economic and ecological roles of *Fraxinus excelsior* in European forests underscore the necessity of investing in sustainable conservation methods and strategies to protect forests against pests and diseases. A comprehensive understanding of the genomic basis underlying susceptibility to ash dieback could enhance our understanding of pathogen responses, thereby contributing to effective conservation strategies. As the body of literature on this topic expands, so does its complexity. At the same time, new methods are becoming available, which might help to resolve this complexity. This growing tool kit enables the exploration of connections between genomic regions and specific traits, particularly regarding the susceptibility to ash dieback. Understanding the genomic basis underlying stem collar necrosis (SCN) could help unravel the mechanisms driving ash dieback susceptibility. This research aims to contribute valuable knowledge to the field of

forest genomics and enhance the understanding of genetic susceptibility to ash dieback, ultimately aiding in the conservation and management of ash populations. The first aim of this thesis was, to identify and select large full-sib families in single-tree progenies. We hypothesized that this will be possible due to one or few (unknown) male trees having a disproportionately large contribution to the generation of offspring, as has been reported for other tree species before. The second aim was to discover susceptibility loci through QTL analysis and determine the effect size of the identified loci. Specifically, the following hypotheses were tested: (i) dominant fathers in segregating ash populations can be accurately identified using SSR and SNP markers, (ii) the susceptibility of ash dieback can be partially explained with QTL analysis, (iii) the genomic regions connected with susceptibility have large effect sizes.

References

- Alfaro, R. I., King, J. N., and vanAkker, L. (2013). Delivering Sitka spruce with resistance against white pine weevil in British Columbia, Canada. *The Forestry Chronicle* 89, 235–245. doi: 10.5558/tfc2013-042
- Alkan, C., Coe, B. P., and Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nat Rev Genet* 12, 363–376. doi: 10.1038/nrg2958
- Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., et al. (2020). Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell* 182, 145–161.e23. doi: 10.1016/j.cell.2020.05.021
- Ashley Jones, Cynthia Torkel, David Stanley, Jamila Nasim, Justin Borevitz, and Benjamin Schwessinger (2021). High-molecular weight DNA extraction, clean-up and size selection for long-read sequencing.
- Ashwath, M. N., Lavale, S. A., Santhoshkumar, A. V., Mohapatra, S. R., Bhardwaj, A., Dash, U., et al. (2023). Genome-wide association studies: an intuitive solution for SNP identification and gene mapping in trees. *Funct Integr Genomics* 23, 297. doi: 10.1007/s10142-023-01224-8
- Bakys, R., Vasaitis, R., and Skovsgaard, J. P. (2013). Patterns and severity of crown dieback in young even-aged stands of european ash (*Fraxinus excelsior* L.) in relation to stand density, bud flushing phenotype, and season. *Plant Protect. Sci.* 49, 120–126. doi: 10.17221/70/2012-PPS
- Bartholomé, J., Brachi, B., Marçais, B., Mougou-Hamdane, A., Bodénès, C., Plomion, C., et al. (2020). The genetics of exapted resistance to two exotic pathogens in pedunculate oak. *New Phytol* 226, 1088–1103. doi: 10.1111/nph.16319
- Burggraaf, P. D. (1972). Some observations on the course of the vessels in the wood of *Fraxinus excelsior* L. *Acta Botanica Neerlandica* 21, 32–47. doi: 10.1111/j.1438-8677.1972.tb00744.x
- Butler, J. B., Freeman, J. S., Vaillancourt, R. E., Potts, B. M., Glen, M., Lee, D. J., et al. (2016). Evidence for different QTL underlying the immune and

- hypersensitive responses of *Eucalyptus globulus* to the rust pathogen *Puccinia psidii*. *Tree Genetics & Genomes* 12, 1–13. doi: 10.1007/s11295-016-0987-x
- Carroll, D., and Boa, E. (2024). Ash dieback: From Asia to Europe. *Plant Pathol* 73, 741–759. doi: 10.1111/ppa.13859
- Cartwright, D. A., Troglio, M., Velasco, R., and Gutin, A. (2007). Genetic mapping in the presence of genotyping errors. *Genetics* 176, 2521–2527. doi: 10.1534/genetics.106.063982
- Chandelier, A., Gerarts, F., San Martin, G., Herman, M., and Delahaye, L. (2016). Temporal evolution of collar lesions associated with ash dieback and the occurrence of *Armillaria* in Belgian forests. *For. Path.* 46, 289–297. doi: 10.1111/efp.12258
- Choi, K., Reinhard, C., Serra, H., Ziolkowski, P. A., Underwood, C. J., Zhao, X., et al. (2016). Recombination Rate Heterogeneity within *Arabidopsis* Disease Resistance Genes. *PLOS Genetics* 12, e1006179. doi: 10.1371/journal.pgen.1006179
- Cleary, M., Nguyen, D., Stener, L. G., Stenlid, J., and Skovsgaard, J. P. (2017). Ash and ash dieback in Sweden: A review of disease history, current status, pathogen and host dynamics, host tolerance and management options in forests and landscapes. Swedish University of Agricultural Sciences. ISBN: 978-91-576-8696-1.
- Coker, T. L. R., Rozsypálek, J., Edwards, A., Harwood, T. P., Butfoy, L., and Buggs, R. J. A. (2019). Estimating mortality rates of European ash (*Fraxinus excelsior*) under the ash dieback (*Hymenoscyphus fraxineus*) epidemic. *Plants, People, Planet* 1, 48–58. doi: 10.1002/ppp3.11
- Dallinger, H. G., Löschenberger, F., Azrak, N., Ametz, C., Michel, S., and Bürstmayr, H. (2024). Genome-wide association mapping for pre-harvest sprouting in European winter wheat detects novel resistance QTL, pleiotropic effects, and structural variation in multiple genomes. *Plant Genome* 17, e20301. doi: 10.1002/tpg2.20301
- Davydenko, K., Skrylnyk, Y., Borysenko, O., Menkis, A., Vysotska, N., Meshkova, V., et al. (2022). Invasion of Emerald Ash Borer *Agrilus planipennis* and Ash Dieback Pathogen *Hymenoscyphus fraxineus* in Ukraine—A Concerted Action. *Forests* 13, 789. doi: 10.3390/f13050789
- Dobrowolska, D., Hein, S., Oosterbaan, A., WAGNER, S., Clark, J., and Skovsgaard, J. P. (2011). A review of European ash (*Fraxinus excelsior* L.): implications for silviculture. *Forestry* 84, 133–148. doi: 10.1093/forestry/cproo1
- Doonan, J. M., Budde, K. B., Kosawang, C., Lobo, A., Verbylaite, R., Brealey, J. C., et al. (2023). Multiple, single trait GWAS and supervised machine learning reveal the genetic architecture of *Fraxinus excelsior* tolerance to ash dieback in Europe. *bioRxiv*, 2023.12.11.570802. doi: 10.1101/2023.12.11.570802
- Dumschott, K., Schmidt, M. H.-W., Chawla, H. S., Snowdon, R., and Usadel, B. (2020). Oxford Nanopore sequencing: new opportunities for plant genomics? *J Exp Bot* 71, 5313–5322. doi: 10.1093/jxb/eraa263
- El-Kassaby, Y. A., Cappa, E. P., Liewlaksaneeyanawin, C., Klápště, J., and Lstibůrek, M. (2011). Breeding without breeding: is a complete pedigree necessary for efficient breeding? *PLOS ONE* 6, e25737. doi: 10.1371/journal.pone.0025737
- El-Kassaby, Y. A., Lstibůrek, M., Liewlaksaneeyanawin, C., Slavov, G. T., and Howe G. T., eds (2007). Breeding without breeding: approach, example, and proof of concept. In *Proceedings of the IUFRO division 2 joint conference: low input breeding and conservation of forest genetic resources*. Antalya, Turkey.

- Enderle, R., Nakou, A., Thomas, K., and Metzler, B. (2015). Susceptibility of autochthonous German *Fraxinus excelsior* clones to *Hymenoscyphus pseudoalbidus* is genetically determined. *Annals of Forest Science* 72, 183–193. doi: 10.1007/s13595-014-0413-1
- EUFORGEN (2009). Distribution map of common ash (*Fraxinus excelsior*). First published online in 2003; updated on 30 July 2008, https://www.euforgen.org/fileadmin/templates/euforgen.org/upload/Documents/Maps/PDF/Fraxinus_excelsior.pdf
- Fan, S., Georgi, L. L., Hebard, F. V., Zhebentyayeva, T., Yu, J., Sisco, P. H., et al. (2024). Mapping QTLs for blight resistance and morpho-phenological traits in inter-species hybrid families of chestnut (*Castanea* spp.). *Front Plant Sci* 15, 1365951. doi: 10.3389/fpls.2024.1365951
- Feng, L., and Du, F. K. (2022). Landscape Genomics in Tree Conservation Under a Changing Environment. *Front Plant Sci* 13, 822217. doi: 10.3389/fpls.2022.822217
- Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural variation in the human genome. *Nat Rev Genet* 7, 85–97. doi: 10.1038/nrg1767
- Fones, H. N., Mardon, C., and Gurr, S. J. (2016). A role for the asexual spores in infection of *Fraxinus excelsior* by the ash-dieback fungus *Hymenoscyphus fraxineus*. *Sci Rep* 6, 34638. doi: 10.1038/srep34638
- Formato, M., Scharenberg, F., Pacifico, S., and Zidorn, C. (2022). Seasonal variations in phenolic natural products in *Fagus sylvatica* (European beech) leaves. *Phytochemistry* 203, 113385. doi: 10.1016/j.phytochem.2022.113385
- Fuchs, S., Häuser, H., Peters, S., Knauf, L., Rentschler, F., Kahlenberg, G., et al. (2024). Ash dieback assessments on intensive monitoring plots in Germany: influence of stand, site and time on disease progression. *J Plant Dis Prot*, 1–18. doi: 10.1007/s41348-024-00889-y
- George, J.-P., Sanders, T. G. M., Timmermann, V., Potočić, N., and Lang, M. (2022). European-wide forest monitoring substantiate the necessity for a joint conservation strategy to rescue European ash species (*Fraxinus* spp.). *Sci Rep* 12, 4764. doi: 10.1038/s41598-022-08825-6
- Gross, A., Zaffarano, P. L., Duo, A., and Grünig, C. R. (2012). Reproductive mode and life cycle of the ash dieback pathogen *Hymenoscyphus pseudoalbidus*. *Fungal Genet Biol* 49, 977–986. doi: 10.1016/j.fgb.2012.08.008
- Harper, A. L., McKinney, L. V., Nielsen, L. R., Havlickova, L., Li, Y., Trick, M., et al. (2016). Molecular markers for tolerance of European ash (*Fraxinus excelsior*) to dieback disease identified using Associative Transcriptomics. *Sci Rep* 6, 19335. doi: 10.1038/srep19335
- Häuser, H. L., Hauswald, J., and Kätzel, R. (2024). Comparative studies on the identification and assessment of basal stem necrosis in ash trees (*Fraxinus excelsior*) against the backdrop of the European ash dieback.
- Holliday, J. A., Aitken, S. N., Cooke, J. E. K., Fady, B., González-Martínez, S. C., Heuertz, M., et al. (2017). Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. *Mol Ecol* 26, 706–717. doi: 10.1111/mec.13963
- Huang, K., and Rieseberg, L. H. (2020). Frequency, Origins, and Evolutionary Role of Chromosomal Inversions in Plants. *Front Plant Sci* 11, 296. doi: 10.3389/fpls.2020.00296
- Kang, M., Chanderbali, A., Lee, S., Soltis, D. E., Soltis, P. S., and Kim, S. (2023). High-molecular-weight DNA extraction for long-read sequencing of plant genomes: An optimization of standard methods. *Appl Plant Sci* 11, e11528. doi: 10.1002/aps3.11528

- Klesse, S., Abegg, M., Hopf, S. E., Gossner, M. M., Rigling, A., and Queloz, V. (2021). Spread and Severity of Ash Dieback in Switzerland – Tree Characteristics and Landscape Features Explain Varying Mortality Probability. *Front. For. Glob. Change* 4. doi: 10.3389/ffgc.2021.645920
- Langer, G. (2017). Collar rots in forests of Northwest Germany affected by ash dieback. *Baltic Forestry* 23, 44-19.
- Langer, G. J., Fuchs, S., Osewold, J., Peters, S., Schrewe, F., Ridley, M., et al. (2022). FraxForFuture—research on European ash dieback in Germany. *J Plant Dis Prot*, 1–11. doi: 10.1007/s41348-022-00670-z
- Lespinasse, D., Grivet, L., Troispoux, V., Rodier-Goud, M., Pinard, F., and Seguin, M. (2000). Identification of QTLs involved in the resistance to South American leaf blight (*Microcyclus ulei*) in the rubber tree. *Theor Appl Genet* 100, 975–984. doi: 10.1007/s001220051379
- Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., et al. (2015). Fungal effectors and plant susceptibility. *Annu Rev Plant Biol* 66, 513–545. doi: 10.1146/annurev-arplant-043014-114623
- Lobo, A., Hansen, J. K., McKinney, L. V., Nielsen, L. R., and Kjær, E. D. (2014). Genetic variation in dieback resistance: growth and survival of *Fraxinus excelsior* under the influence of *Hymenoscyphus pseudoalbidus*. *Scandinavian Journal of Forest Research* 29, 519–526. doi: 10.1080/02827581.2014.950603
- Lobo, A., McKinney, L. V., Hansen, J. K., Kjaer, E. D., and Nielsen, L. R. (2015). Genetic variation in dieback resistance in *Fraxinus excelsior* confirmed by progeny inoculation assay. *For. Path.* 45, 379–387. doi: 10.1111/efp.12179
- Lutz, T., Ridley, M., Hadel, B., Schulz, B., Enderle, R., Steinert, M., et al. (2023). Evaluation and identification of viruses for biocontrol of the ash dieback disease. *J Plant Dis Prot*, 1–11. doi: 10.1007/s41348-023-00804-x
- Lysenko, L., Griem, E., Wagener, P., and Langer, E. J. (2024). Fungi associated with fine roots of *Fraxinus excelsior* affected by ash dieback detected by next-generation sequencing. *J Plant Dis Prot*, 1–13. doi: 10.1007/s41348-024-00923-z
- Mackay, T. F. (2001). The genetic architecture of quantitative traits. *Annu Rev Genet* 35, 303–339. doi: 10.1146/annurev.genet.35.102401.090633
- Madsen, C. L., Kosawang, C., Thomsen, I. M., Hansen, L. N., Nielsen, L. R., and Kjær, E. D. (2021). Combined progress in symptoms caused by *Hymenoscyphus fraxineus* and *Armillaria* species, and corresponding mortality in young and old ash trees. *Forest Ecology and Management* 491, 119177. doi: 10.1016/j.foreco.2021.119177
- Mansfield, J., Brown, I., and Papp-Rupar, M. (2019). Life at the edge – the cytology and physiology of the biotroph to necrotroph transition in *Hymenoscyphus fraxineus* during lesion formation in ash. *Plant Pathol* 68, 908–920. doi: 10.1111/ppa.13014
- Marçais, B., Husson, C., Godart, L., and Caël, O. (2016). Influence of site and stand factors on *Hymenoscyphus fraxineus* -induced basal lesions. *Plant Pathol* 65, 1452–1461. doi: 10.1111/ppa.12542
- Marigo, G., Peltier, J.-P., Girel, J., and Pautou, G. (2000). Success in the demographic expansion of *Fraxinus excelsior* L. *Trees* 15, 1–13. doi: 10.1007/s004680000061
- Martín, J. A., Domínguez, J., Solla, A., Brasier, C. M., Webber, J. F., Santini, A., et al. (2023). Complexities underlying the breeding and deployment of Dutch elm disease resistant elms. *New For (Dordr)* 54, 661–696. doi: 10.1007/s11056-021-09865-y

- McKinney, L. V., Nielsen, L. R., Collinge, D. B., Thomsen, I. M., Hansen, J. K., and Kjaer, E. D. (2014). The ash dieback crisis: genetic variation in resistance can prove a long-term solution. *Plant Pathol* 63, 485–499. doi: 10.1111/ppa.12196
- McKinney, L. V., Nielsen, L. R., Hansen, J. K., and Kjær, E. D. (2011). Presence of natural genetic resistance in *Fraxinus excelsior* (Oleraceae) to *Chalara fraxinea* (Ascomycota): an emerging infectious disease. *Heredity* 106, 788–797. doi: 10.1038/hdy.2010.119
- Meger, J., Kozioł, C., Pałucka, M., Burczyk, J., and Chybicki, I. J. (2024a). Genetic resources of common ash (*Fraxinus excelsior* L.) in Poland. *BMC Plant Biol* 24, 186. doi: 10.1186/s12870-024-04886-z
- Meger, J., Ulaszewski, B., Pałucka, M., Kozioł, C., and Burczyk, J. (2024b). Genomic prediction of resistance to *Hymenoscyphus fraxineus* in common ash (*Fraxinus excelsior* L.) populations. *Evol Appl* 17, e13694. doi: 10.1111/eva.13694
- Mitchell, R. J., Hewison, R. L., Hester, A. J., Broome, A., and Kirby, K. J. (2016). Potential impacts of the loss of *Fraxinus excelsior* (Oleaceae) due to ash dieback on woodland vegetation in Great Britain. *New Journal of Botany* 6, 2–15. doi: 10.1080/20423489.2016.1171454
- Nemesio-Gorriz, M., McGuinness, B., Grant, J., Dowd, L., and Douglas, G. C. (2019). Lenticel infection in *Fraxinus excelsior* shoots in the context of ash dieback. *iForest* 12, 160–165. doi: 10.3832/ifor2897-012
- Neophytou, C., Heer, K., Milesi, P., Peter, M., Pyhäjärvi, T., Westergren, M., et al. (2022). Genomics and adaptation in forest ecosystems. *Tree Genetics & Genomes* 18, 1–7. doi: 10.1007/s11295-022-01542-1
- Nguyen, T. V., Vander Jagt, C. J., Wang, J., Daetwyler, H. D., Xiang, R., Goddard, M. E., et al. (2023). In it for the long run: perspectives on exploiting long-read sequencing in livestock for population scale studies of structural variants. *Genet Sel Evol* 55, 9. doi: 10.1186/s12711-023-00783-5
- Nishii, K., Möller, M., Foster, R. G., Forrest, L. L., Kelso, N., Barber, S., et al. (2023). A high quality, high molecular weight DNA extraction method for PacBio HiFi genome sequencing of recalcitrant plants. *Plant Methods* 19, 41. doi: 10.1186/s13007-023-01009-x
- Peters, S., Fuchs, S., Bien, S., Bußkamp, J., Langer, G. J., and Langer, E. J. (2023). Fungi associated with stem collar necroses of *Fraxinus excelsior* affected by ash dieback. *Mycol Progress* 22. doi: 10.1007/s11557-023-01897-2
- Peters, S., Gruschwitz, N., Bien, S., Fuchs, S., Bubner, B., Blunk, V., et al. (2024). The fungal predominance in stem collar necroses of *Fraxinus excelsior*: a study on *Hymenoscyphus fraxineus* multilocus genotypes. *J Plant Dis Prot*, 1–13. doi: 10.1007/s41348-024-00912-2
- Pliura A, Lygis V, Suchockas V, and Bartkevicius E. (2011). Performance of Twenty Four European *Fraxinus excelsior* Populations in three Lithuanian progeny trials with a special emphasis on resistance to *Chalara Fraxinea*. *Baltic Forestry* 17, 17–34.
- Plomion, C., Bastien, C., Bogeat-Triboulot, M.-B., Bouffier, L., Déjardin, A., Duplessis, S., et al. (2016). Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science* 73, 77–103. doi: 10.1007/s13595-015-0488-3
- Plumb, W. J., Kelly, L. J., Mullender, J., Powell, R. F., Nemesio-Gorriz, M., Carey, D., et al. (2024). Genetic barcodes for ash (*Fraxinus*) species and generation of new wide hybrids. *bioRxiv*, 2024.02.19.581010. doi: 10.1101/2024.02.19.581010
- Rani, K., Kumar, M., Razzaq, A., Ajay, B. C., Kona, P., Bera, S. K., et al. (2023). “Chapter 1 - Recent advances in molecular marker technology for QTL

- mapping in plants,” in *QTL mapping in crop improvement: Present progress and future perspectives*, eds. S. H. Wani, D. Wang, and G. Singh (Amsterdam: Academic Press), 1–15.
- Rozsypálek, J., Martinek, P., Palovčíková, D., and Jankovský, L. (2023). The protection of ash trees against ash dieback by tree injections. *Urban Forestry & Urban Greening* 88, 128071. doi: 10.1016/j.ufug.2023.128071
- Saumitou-Laprade, P., Vernet, P., Vekemans, X., Billiard, S., Gallina, S., Essalouh, L., et al. (2017). Elucidation of the genetic architecture of self-incompatibility in olive: Evolutionary consequences and perspectives for orchard management. *Evol Appl* 10, 867–880. doi: 10.1111/eva.12457
- Snieszko, R. A., Johnson, J. S., Reeser, P., Kegley, A., Hansen, E. M., Sutton, W., et al. (2020). Genetic resistance to *Phytophthora lateralis* in Port-Orford-cedar (*Chamaecyparis lawsoniana*) – Basic building blocks for a resistance program. *Plants, People, Planet* 2, 69–83. doi: 10.1002/ppp3.10081
- Sollars, E. S. A., Harper, A. L., Kelly, L. J., Sambles, C. M., Ramirez-Gonzalez, R. H., Swarbreck, D., et al. (2017). Genome sequence and genetic diversity of European ash trees. *Nature* 541, 212–216. doi: 10.1038/nature20786
- Stocks, J. J., Metheringham, C. L., Plumb, W. J., Lee, S. J., Kelly, L. J., Nichols, R. A., et al. (2019). Genomic basis of European ash tree resistance to ash dieback fungus. *Nat Ecol Evol* 3, 1686–1696. doi: 10.1038/s41559-019-1036-6
- Street, N. R. (2019). “Chapter One - Genomics of forest trees,” in *Advances in Botanical Research: Molecular Physiology and Biotechnology of Trees*, ed. F. M. Cánovas (Academic Press), 1–37.
- Stroheker, S., Queloz, V., and Nemesio-Gorriz, M. (2021). First report of *Hymenoscyphus fraxineus* causing ash dieback in Spain. *New Disease Reports* 44. doi: 10.1002/ndr2.12054
- Swalarsk-Parry, B. S., Steenkamp, E. T., van Wyk, S., Santana, Q. C., van der Nest, M. A., Hammerbacher, A., et al. (2022). Identification and Characterization of a QTL for Growth of *Fusarium circinatum* on Pine-Based Medium. *J Fungi (Basel)* 8. doi: 10.3390/jof811214
- Thomas, P. A. (2016). Biological Flora of the British Isles: *Fraxinus excelsior*. *Journal of Ecology* 104, 1158–1209. doi: 10.1111/1365-2745.12566
- Timmermann, V., Børja, I., Hietala, A. M., Kirisits, T., and Solheim, H. (2011). Ash dieback: pathogen spread and diurnal patterns of ascospore dispersal, with special emphasis on Norway*. *EPPO Bulletin* 41, 14–20. doi: 10.1111/j.1365-2338.2010.02429.x
- Tong, C., Li, H., Wang, Y., Li, X., Ou, J., Wang, D., et al. (2016). Construction of High-Density Linkage Maps of *Populus deltoides* × *P. simonii* Using Restriction-Site Associated DNA Sequencing. *PLOS ONE* 11, e0150692. doi: 10.1371/journal.pone.0150692
- Yang, X., Lee, W.-P., Ye, K., and Lee, C. (2019). One reference genome is not enough. *Genome Biol* 20, 104. doi: 10.1186/s13059-019-1717-0

Chapter 1

Identification of full-sibling families from natural single-tree ash progenies based on SSR markers and genome-wide SNPs

Melina Krautwurst^{1,*}, Franziska Past^{2,*}, Birgit Kersten¹, Ben Bubner² & Niels A. Müller¹

¹Thünen Institute of Forest Genetics, Großhansdorf, Germany

²Thünen Institute of Forest Genetics, Waldsieversdorf, Germany

* equal contribution

Correspondence to: ben.bubner@thuenen.de, niels.mueller@thuenen.de

Accepted by the 'Journal of Plant Diseases and Protection' (06.07.2024)



Abstract

Common ash, *Fraxinus excelsior*, is threatened by the invasive pathogen *Hymenoscyphus fraxineus*, which causes ash dieback. The pathogen is rapidly spreading throughout Europe with severe ecological and economic consequences. Multiple studies have presented evidence for the existence of a small fraction of genotypes with low susceptibility. Such genotypes can be targets for natural and artificial selection to conserve *F. excelsior* and associated ecosystems. To resolve the genetic architecture of variation in susceptibility it is necessary to analyze segregating populations. Here we employed about 1,000 individuals of each of four single-tree progenies from potentially tolerant mother trees to identify full-sibling (full-sib) families. To this end, we first genotyped all 4,000 individuals and the four mothers with eight SSR markers. We then used the program COLONY to predict full-sibs without knowledge of the paternal genotypes. For each single-tree progeny, COLONY predicted dozens of full-sib families, ranging from 3-166 individuals. In the next step, 910 individuals assigned to full-sib families with more than 28 individuals were subjected to high-resolution genotyping using over one million genome-wide SNPs which were identified with Illumina low-coverage resequencing. Using these SNP genotyping data in principal component analyses we were able to assign individuals to full-sib families with high confidence. Together the analyses revealed five large families with 73-212 individuals. These can be used to generate genetic linkage maps and to perform quantitative trait locus analyses for ash dieback susceptibility. The elucidation of the genetic basis of natural variation in ash may support breeding and conservation efforts and may contribute to more robust forest ecosystems.

Keywords: Common ash (*Fraxinus excelsior*), breeding-without-breeding, SSR markers, SNPs, whole-genome resequencing, full-sibling families, ash dieback.

Introduction

In the early 1990s the necrotrophic ascomycete fungus *Hymenoscyphus fraxineus*, causing the ash dieback (ADB) disease, was first observed in Europe (Coker et al., 2019; Evans, 2019). The fungus spread from its first introduction to Poland via wind-borne spores over most European ash populations (Kowalski, 2006). The fungus can be traced back to Eastern Asia where it is associated with native *Fraxinus* species (Husson et al., 2011; McKinney et al., 2012; Zhao et al., 2013; Landolt et al., 2016). In Europe, the host of the pathogenic fungus is common ash (*Fraxinus excelsior*). Infected trees suffer from crown dieback and necrotic lesions. In the end, infection often leads to the death of the trees causing severe losses to European woodlands (Bakys et al., 2013; Coker et al., 2019).

Notably, *Fraxinus excelsior* exhibits natural variation in ADB susceptibility and several studies have shown that part of this variation is heritable with estimated heritabilities of 0.25 – 0.57 (Pliura AL, 2007; McKinney et al., 2011; Pliura A et al., 2011; McKinney et al., 2012; Lobo et al., 2014; McKinney et al., 2014; Enderle et al., 2015; Lobo et al., 2015; Harper et al., 2016; Muñoz et al., 2016; Plumb et al., 2020). Several different phenotypes are presumed to be connected to ADB susceptibility. For example, the timing of bud burst and senescence may be important for variation in ADB although results of different studies are not always consistent (McKinney et al., 2011; Bakys et al., 2013; Stener, 2013; Pliura et al., 2016; Nielsen et al., 2017).

Genotype-phenotype associations could reveal the genetic basis of variation in ADB and highlight candidate genes. For genetic mapping or pedigree studies, the selection of parents and artificial mating needs to be conducted. In natural tree populations it can be challenging to perform artificial mating or to identify the father to the naturally occurring seedlings, especially in species with a complex mating system as in *F. excelsior*. Common ash is a wind-pollinated and wind-dispersed, polygamous subdioecious tree species. The method “breeding-without-breeding” (BwB) overcomes the need for artificial mating, by working with paternally unknown but maternally known material. Mothers can be selected based on their genotype or phenotype. Paired with DNA markers, it is possible to reconstruct pedigree structures with BwB and to use the identified full-sib families

for quantitative trait locus analyses or the assessment of various breeding values (El-Kassaby and Lstibůrek, 2009; Lstibůrek et al., 2011; Lstibůrek et al., 2015b). For pedigree prediction choosing a suitable downstream analysis for the sample set is important. Different molecular markers can be effective in assessing kinships, such as simple sequence repeat (SSR) and single-nucleotide polymorphism (SNP) markers (Amom et al., 2020; Jiang et al., 2020; Zeng et al., 2023b).

SSR and SNP markers can be powerful tools in combination or separately (; García et al., 2018; Capo-chichi et al., 2022; Zeng et al., 2023a). SNPs offer the opportunity to identify single base changes between individuals, are mostly biallelic, as well as the most abundant source of genetic polymorphism (Agarwal et al., 2008). With new sequencing technologies high numbers of SNPs can be reliably identified (Howe et al., 2020). SSRs are multi-allelic, highly polymorphic and currently the cheaper option for kinship assessment compared to SNPs, for which sequencing needs to be performed (Ramesh et al., 2020).

For kinship identification within high numbers of individuals, an application of both marker types could be of advantage, because SSR markers offer a low-cost possibility with sufficient performance and can be used for preselecting large full-sib-families before sequencing and SNP calling. The purpose of this study is to analyze the feasibility and advantages of the application of two consecutive methodological steps for kinship assessment in common ash: (i) low-cost genotyping with SSRs to predict potential full-sib families, followed by (ii) high-resolution genotyping using whole-genome sequencing.

Material and Methods

Plant material

The four mother trees are distributed across the state of Mecklenburg-Western Pomerania in the north-east of Germany (Table 1). These trees were selected using an assessment scheme developed in a previous project (ResEsche, FNR project number “FKZ 22019915”). This scheme should ensure the vitality and silvicultural quality of the selected trees. The vitality criteria are assessed on the basis of foliage,

shoot and trunk damage. The quality is recorded with parameters such as diameter at breast height (DBH), height and trunk shape. The selected mother trees all showed no or only a few dieback symptoms in the crown area (no more than 10% crown defoliation, no more than 15% replacement shoot proportion). The upright growth was of perfect or at least of normal quality (straight/upright growth, weak twisted growth, solid woodiness, etc.). Three of the four mother trees showed no signs of ADB, only Dar-18 had a slight stem necrosis. Around 3,000 seeds per tree were collected in 2018 as green seeds. All seeds were sown in a nursery bed within two weeks after harvesting. After germinating in spring 2019, 960 seedlings per spring progeny were planted in 6x4 QuickPot plates (24 pots, 16 cm deep, HerkuPlast Kubern GmbH, Ering, Germany). They were first cultivated in a greenhouse and then transferred under a shading net in the nursery. In September 2019 plants were re-potted in 4x3 QuickPot plates (12 pots, 18 cm deep, HerkuPlast Kubern GmbH, Ering, Germany) and stayed in the nursery until planting. From 15th to 17th of April 2021 the seedlings were planted in a semi-randomized block design at a trial site near Schulzendorf (Brandenburg, Germany; Table 1). The area around the trial site is characterised by agriculture and small forests. Infected trees of *F. excelsior* with ash dieback symptoms were observed adjacent to the trial site.

Table 1. Location information of mother trees and trial site.

Name	Forest District	Forest Area	Coordinates
Mother Trees			
Fri-8	Radelübbe	Friedensthal	53.6135667, 11.3055333
Kar-4	Jaegerhof	Karlsburg	54.0674667, 13.4514500
Eve-2	Grevesmühlen	Eversdorf	53.8784500, 11.2492833
Dar-18	Dargun	Dargun	53.8537899, 12.8458667
Trial Site			
Schulzendorf	Strausberg	Wriezen	52.686818, 14.125038

Sampling, DNA extraction and SSR genotyping

Sampling for SSR genotyping was conducted in the nursery in late summer 2019 (Fri-8) and early summer 2020 (other progenies). The QuickPots were arranged in a 96-well-plate-like format. This arrangement allowed sampling in 96-well-plates without the need for time-consuming individual labelling. The 96-well-plates intended for DNA extraction were filled with two ceramic beads (1,4 mm Omni Beads, Omni International, Kennesaw; United States) per well using the customized Brendan bead dispenser (<https://customlabinstitute.wordpress.com>). The plate was cooled during the sampling process with a plate fitting ice pack. The sample, a 2 x 3 mm piece of the youngest, fully developed leaflet, was taken with forceps, which were cleaned with Ethanol (70 %) between each sample. After sampling, the plates were stored at -80 °C. Because of the large number of samples, a “quick and dirty”-method for DNA extraction was chosen (Hu et al., 2014). The samples in the frozen sampling plates were homogenized (30 Hz; MM400, Retsch, Haan, Germany) for at least four minutes in a precooled container. Depending on the homogenization grade, additional homogenization was performed. After adding 200 µl buffer (50 mM Tris, pH 8; 300 mM NaCl, 0,1 g/ml saccharose), the plates were again homogenized for another minute. After centrifuging the plate (5,889 x g; 5 min), the upper phase, containing the DNA, was directly used for polymerase chain reaction (PCR) after diluting 1:10 on the same day. Plates with the “quick and dirty” DNA extract were stored at -20 °C and, after another round of centrifugation, could be used for another PCR.

The PCR was performed with the Multiplex PCR Kit (Type-It Microsatellite master mix; Qiagen, Hilden, Germany). The SSR primer mix consisted of eight primers (see Table 2 and Table S1). With this SSR-multiplex, a touchdown procedure was performed (Table S2). The first progeny (Fri-8) was analyzed with F24 which was later replaced by F12, because F12 was more variable and more reliable.

Table 2. SSR-primers used for genotyping of ash seedlings in a multiplex assay. Primer sequences are based on the reference genome BATG0.5

primer	locus name	primer sequence (5'-3')	motif	conc.	label (color)
F12	FEMSATL1 2C	F: TTTTGGAACCCTTGATTTTG R: CTATATACACCTACCTCTC	(GA) ₆ CA (GA) ₈	0.2 μM	DY-751 (blue)
F23	Contig184_ 179564_179 439	F: GCCATTGTTGGGTTTCACTC R: CCGGGCTTAGTATCCAACCTG	(AG) ₁₀	0.2 μM	DY-682 (green)
F25	Contig1992_ 131610_13 1822	F: CTCGGAGGTGGTTGATGAGT R: AGAGCTCCCAACGCTCAATA	(TA) ₇	0.2 μM	DY-682 (green)
F26	Contig918_ 100265_100 431	F: TGTTAGTGGTATGGTGGAGGC R: TTGCAGGCACAATTACATAAA G	(CT) ₁₃	0.2 μM	DY-751 (blue)
F27	Contig5418_ 11372_115 13	F: TTATTGCCGATGTCTGTGGA R: CGGGAAGTTTCACCTCAGAT	(TG) ₉	0.2 μM	Cy5 (black)
F28	Contig3670_ 10994_107 05	F: CACTTCAATGTCAGCATTCCA R: GAAAGTTCACCCAGTCAAAT	(CT) ₁₅	0.4 μM	DY-682 (green)
F29	Contig4738_ 19258_191 42	F: GCAAGGGAAGTAGCAACGAC R: TGTGACTCAAATAGGGCTCAG A	(AG) ₂₁	0.2 μM	DY-751 (blue)
F30	Contig344_ 81629_8137 5	F: GGAGCAGTACCATATGCGCT R: AGGAAGATCTAACGCTGCTTG	(AG) ₁₄	0.2 μM	Cy5 (black)
F24	Contig5748_ 109580_10 9480	F: ACATGCCTTCTCTTCGCC R: TTCTAGGGCCTGCAAACAAC	(TA) ₈	0.1 μM	DY-751 (blue)

F12 is adapted from marker FEMSATL12 (Lefort et al. 1999). All other primers are from Sollars et al. (2017).
conc. = final concentration.

The PCR products were analyzed with capillary electrophoresis (GenomeLab™ GeXP, Beckman Coulter) and the corresponding chemical kit (SCIEX, Framingham U.S.A). The peak scoring was done with the provided software (GenomeLab) to obtain a list of the alleles (Tables S3-S7).

COLONY analysis

In order to determine genetic sample relationships, the SSR-genotyping output was transferred to the COLONY Software ((Jones and Wang, 2010a)Version: 2.0.6.5/2.0.6.6). In addition to the input data, input parameters for the analysis of all

progenies had to be defined (see Supplement: Supplemental materials on COLONY parameters for family estimations with SSR markers). In contrast to other methods, the software can be used for monoecious and dioecious species and is not restricted to codominant markers without genotyping error. Also, instead of pairwise comparisons, it uses a full-pedigree likelihood approach, which takes into account the likelihood of the entire pedigree structure and allows the simultaneous inference of parentage and sibship (Jones and Wang, 2010a). This likelihood approach can lead to different outputs if the analysis is repeated. To ensure the reliability of the results, the software was run twice. The two runs were compared and the individuals that were observed in both runs were chosen. The chosen families are listed in Table 3 and named after the number of full-sib families identified by COLONY (e.g., FS 05).

DNA extraction and Illumina low-coverage resequencing

The first batch of 167 samples with young leaf material was collected in September 2020, and immediately placed on ice in 1.5 ml Eppendorf safe-lock tubes (Eppendorf, Wesslingen, Germany). In addition, leaves from the four mother tree clones (Table 1) were collected. All samples were stored at -80°C until DNA extraction. Frozen samples were homogenized using pestle and mortar in liquid nitrogen. All following steps were conducted following the DNA extraction protocol by Bruegmann et al., 2022. The second batch of 751 samples was collected in June 2021. For the 751 samples, we used the MagMAX™ Plant DNA Isolation Kit (Thermo Fisher Scientific, Germany) following purification using the KingFisher™ Apex (Thermo Fisher Scientific, Germany) with a 96 deep-well head. DNA sample QC and library preparation for sequencing were performed by Novogene (UK) Ltd. (Cambridge, UK) for both sample batches. In the first batch, 163 of 172 samples passed the quality control, including the four mother tree samples. In the second batch, 747 of 751 samples passed the quality control. Sequencing data (2 x 150 bp reads) were generated on the Novaseq 6000 platform for both batches. The first batch was sequenced to an average sequencing depth of 10.8X and the second batch to 11.3X according to the ash reference genome (Sollars et al., 2017). Both batches together comprise 906 samples plus the four mother trees.

Mapping and variant calling

For both batches, sequencing data were mapped against the common ash reference genome (Sollars et al., 2017) using bwa-mem (version bwa-0.7.17.tar.bz2 (Li and Durbin, 2009)). Grouping of the reads and duplicates were marked using Picard tool's (v2.26.2) (<http://broadinstitute.github.io/picard/>). Joint variant calling was performed for batch one with GATK (version 4.2.3.0), following the best practices for germline short variant discovery wherever possible (Poplin et al., 2017). For the second batch, the variant calling was performed by Novogene using Sentieon (Aldana and Freed, 2022). For generating gVCFs the 'HaplotypeCaller' from GATK was used. After combining the gVCFs with GATK's 'GenomicsDBImport', the 'GenotypeGVFs' tool was used for the joint genotyping. For batch one GATK version 4.2.3.0 and for batch two GATK version 4.0.5.1 were used (Kemp, 2003).

Variant filtering

The two sample sets were filtered separately, with the same filtering options. For hard filtering, we mostly followed the documentation on 'Hard-filtering germline short variants' on the GATK website. We filtered indels and SNPs separately. We removed variants based on strand bias (FisherStrand 'FS' > 60 & StrandOddsRatio 'SOR' > 3) and mapping quality (RMSMappingQuality 'MQ' < 40, MappingQualityRankSumTest 'MQRankSum' < -1). Based on the distribution of the variant confidence score QualByDepth 'QD' we chose a more stringent cutoff of QD > 10, to remove any low-confidence variants. Filtering was performed with bcftools v1.7 (Li, 2011). We then extracted the variant sequencing depth values 'DP' and minor allele frequency 'frq2' using vcftools v0.1.15 (Danecek et al., 2011). To visualize the DP and choose the parameters we used R (R Core Team, 2022). Minimal-mean DP was 5.2 and max-mean DP 10.3. Non-biallelic SNPs were excluded and SNPs with more than 10 % missing data were removed. Filtering resulted in a set of 14.42 million SNPs for the first batch and 11.78 million SNPs for the second batch. Before the resulting VCF (Variant Call Format) files could be merged, an intersect was calculated using the 'isec' function of bcftools (Danecek et al., 2021) to identify common SNPs in both VCF files. Then, the 'merge' function was used to create one multi-sample file. Further, the minor allele frequency was filtered

0.08-0.50 with PLINK v1.9 (Purcell et al., 2007). The merged file of both batches comprised 5.87 million SNPs for 910 individuals.

We used PLINK for principal component analyses (PCA) to determine family-clusters in the SNP datasets. Further, we used the R package ‘SNPRelate’ (Zheng et al., 2012) for the genome-wide identity-by-state analysis to create a dendrogram based on the SNP data. The dendrogram implements the formed clusters of the PCA and does not include the outliers. Except for Fri-8 in the dendrogram outliers included.

Data visualization

The results were further analyzed and visualized using the R packages ‘ggplot2’ (Wickham, 2016), ‘VCFR’ (Knaus and Grünwald, 2017), ‘MASS’ (Venables and Ripley, 2003) and ‘dendextend’ (Galili, 2015).

Results

A total of 960 samples per single-tree progeny of each of four mother trees, potentially tolerant against ash dieback, were collected and analyzed with eight SSR markers (Tables S3-S7). In the end, 3,476 of 3,840 (90.5 %) samples could be successfully genotyped. The program COLONY predicted between 151 and 179 different fathers per progeny (Figure 1 and Figure 2, Table S8) giving rise to single trees without full-sib-family membership up to full-sib families with 166 members. In total, 910 individuals, which were assigned to the nine largest predicted full-sib families, were selected for high-resolution genotyping using Illumina whole-genome resequencing (two sets). Only families with at least 28 individuals were chosen. With a sequencing depth of 10X we were able to identify a total of 5.87 million SNPs after read mapping to the *F. excelsior* reference genome (Sollars et al., 2017) and SNP filtering. Employing only largely independent SNPs ($r^2 < 0.2$) within each progeny, we were able to reliably assign full-sib families. The final file included six samples that were sequenced and analyzed in set 1 and set 2. With these ‘twins’ we were able to compare the two downstream analyses. The ‘twins’ are represented in the Fri-8 dendrogram (Figure 2), which shows the same results for both downstream analyses.

Table 3 summarizes all results of the full-sibling family identification using SSR and SNP markers. Full-sib families identified by COLONY are named with FS (for full-siblings) and the number given by COLONY. Full-sib families assigned by PCA analysis are named after the number of clusters formed in the PCA. The dendrogram is based on identity-by-state analysis, showing the genetic distance as the proportion of loci where the alleles are identical (Figure 1, 2 (e and f)). The vertical axis shows the proportion of the genetic distance between the individuals, with longer branches indicating greater genetic distance and shorter branches indicating closer genetic similarity. The height of the nodes corresponds with the genetic distance at which the identity-by-state algorithm decided to merge or split.

For the two single-tree progenies from Eve-2 and Dar-18, the family structure predicted by COLONY using the eight SSR markers was largely consistent with the SNP data analyses (Figure 1). Only a few outliers were detected. In Eve-2, family sizes were close to the predictions, with the largest full-sib family comprising 138 individuals (Table 3). This demonstrates the general feasibility of performing breeding-without-breeding in ash with the eight described SSR markers. For Dar-18, a single dominant pollen donor gave rise to a single large full-sib family, of which 116 individuals could be confirmed with the SNP markers. Similar to Eve-2, Dar-18 also demonstrates the feasibility of the SSR markers approach.

The other two single-tree progenies from Kar-4 and Fri-8 showed discrepancies between the SSR and SNP marker classifications (Figure 2). Fri-8 was predicted to be composed of three full-sib families using SSR markers. However, the SNP markers revealed it to be a single large family of 212 individuals, with additional outliers representing other pollen donors. The progenies of Fri-8 had the lowest genetic distance (0.20 individual dissimilarity) compared to the other dendrograms (0.25-0.30 individual dissimilarity). Further, the outliers of Fri-8 are placed at a higher level and show that they are genetically less similar. The lower individual dissimilarity of Fri-8 compared to the other progenies can be an indicator that the unknown fathers of the full-sib family are more closely related than the fathers of other full-sib families, especially compared with Dar-18, which shows the highest dissimilarity in all four dendrograms.

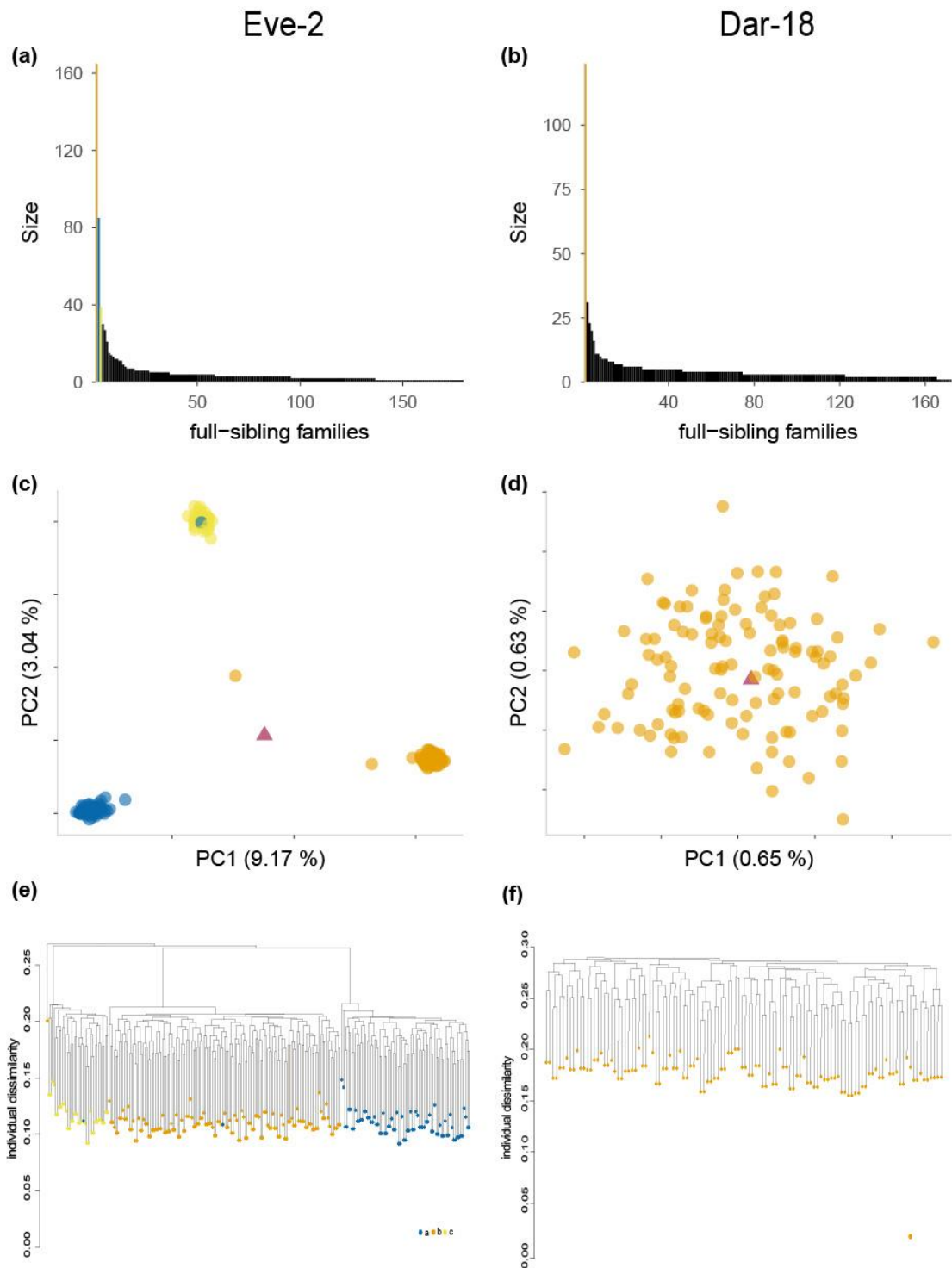


Figure 6: Comparison of SSR and SNP full-sibling identification from the mother trees Eve-2 and Dar-18. The barplots (a and b) show all progenies that were analyzed with the SSR markers and assigned to full-sibling families (ranging from 1 – 165 siblings). Each bar represents one predicted full-sibling family. Selected families with more than 30 individuals are indicated by colors, that is three families for Eve-2 (a) and one family for Dar-18 (b). The principal component analyses (c and d) show the SNP marker results. The colour scheme of the dots represents the results of the SSR markers. The red triangle represents the mother tree. Panel (e) and (f) show the results of genome-wide identity-by-state analyses using the SNP marker results in dendrograms. The y-axis represents the individual dissimilarity and the x-axis represents the individual samples being clustered. The dots represent the SSR results and the dendrogram clustering represents the SNP markers.

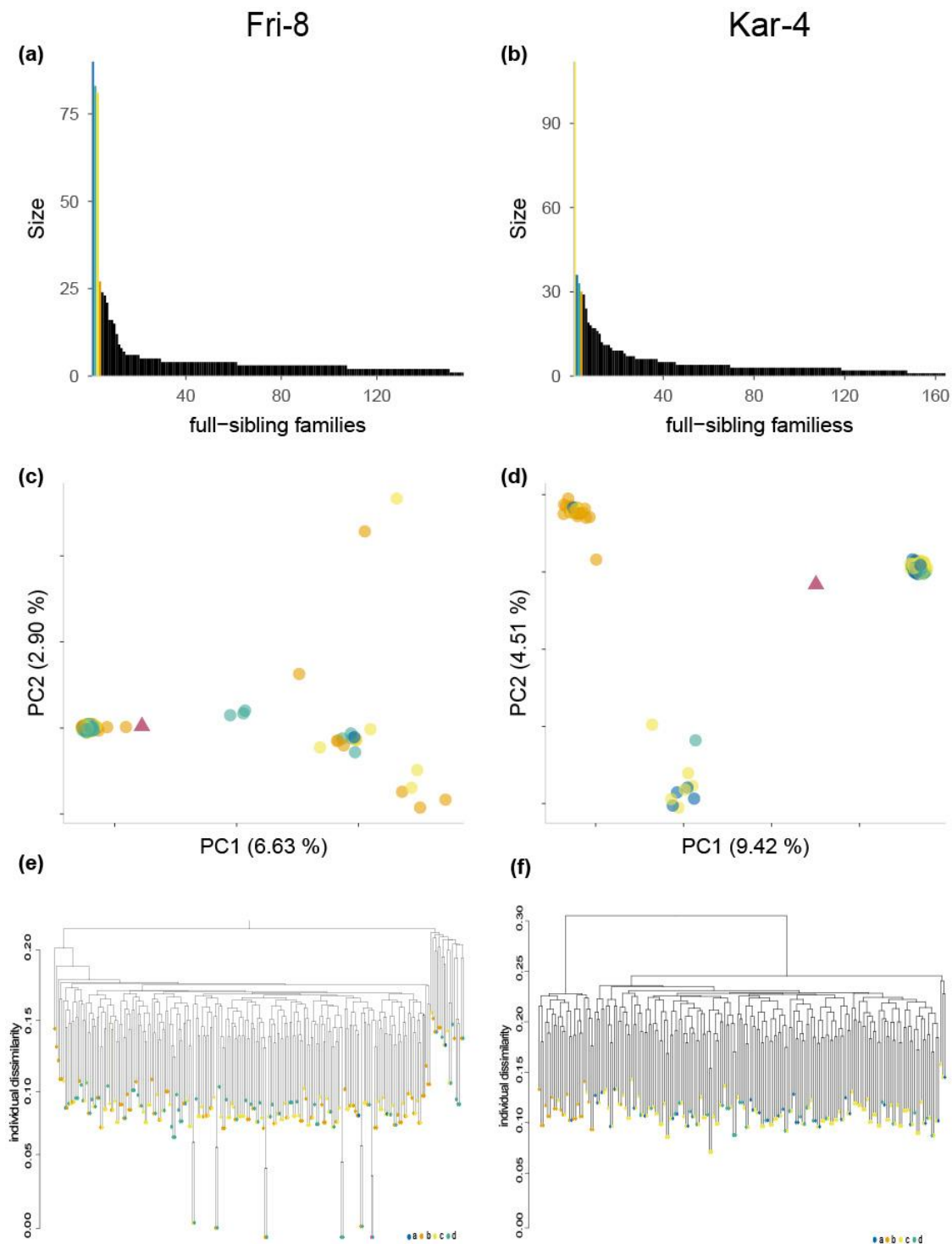


Figure 7: Comparison of SSR and SNP full-sibling identification from the mother trees Fri-8 and Kar-4. The barplots (a and b) show all progenies that were analyzed with the SSR markers and assigned to full-sibling families (ranging from 1 – 112 siblings). Each bar represents one predicted full-sibling family. Selected families with more than 30 individuals are indicated by colors, that is four families for Fri-8 (a) and four families for Kar-4 (b). The principal component analyses (c and d) show SNP marker results. The colour scheme of the dots represents the results of the SSR markers. The red triangle represents the mother tree. Panels (e) and (f) show the results of genome-wide identity-by-state analysis using the SNP marker in dendrograms. The y-axis represents the individual dissimilarity and the x-axis represents the individual samples being clustered. The dots represent the SSR results and the dendrogram clustering represents the SNP markers.

Table 3: Summary of all full-sibling families identified by SSR and SNP marker.

Progeny	Abbreviation	SSR-full-sibling family name	Size	SNP-full-sibling family name	Size
Friedrichsthal	Fri-8	FS 10	81	Cluster 1	212
		FS 17	90		
		FS 22	63		
Karlsburg	Kar-4	FS 29	113	Cluster 1	164
		FS 05	37	Cluster 2	25
		FS 21	30		
		FS 41	33		
Eversdorf	Eve-2	FS 20	166	Cluster 2	138
		FS 08	85	Cluster 1	73
		FS 30	39	Cluster 3	35
Dargun	Dar-18	FS 19	123	Cluster1	116
		FS 14	28		

FS = Full-sib families identified by COLONY

The description of Cluster indicates the groups of individuals clustering in the PCA (Figure 1 and 2)

In Kar-4, the SNP marker analysis showed two relatively large families (164 and 25 individuals) instead of four predicted by SSR markers. Full-sib families identified by COLONY for Kar-4, indicated by blue and yellow dots, formed one cluster in the SNP PCA (Figure 2), demonstrating these are one family instead of two. The predicted fathers of the four families identified by COLONY in both runs showed strong similarity (Table S9). The assigned offspring changed between the predicted fathers when comparing the runs. SNP analysis defined these families (predicted by COLONY) as one family.

To ensure the reliability of our full-sib determinations, we suggest considering a combined approach using SSRs for preselection and SNPs for validation of full-sibs, thus improving the robustness and accuracy of the family structure predictions. However, SSRs were used to preselect full-sibs, while SNPs were described as more reliable in full-sib identification.

Discussion

For genotype-phenotype association studies it is important to maximize the statistical power by employing large mapping populations with a high number of individuals. In BwB approaches, the reliable determination of full-sib families is critical. Although whole-genome resequencing and genome-wide SNP detection provided robust data for this purpose, sequencing costs are still be prohibitive with

thousands of samples, such as the 4,000 individuals in our study. The preselection of individuals prior to SNP genotyping was thus important, and SSR markers provided the required simplicity and relatively low costs.

Our results show that implementing a two-step genotyping process with SSR and whole-genome sequencing is an effective way to achieve the identification of full-sib families in ash. Similar methodologies have been employed in other forestry studies, highlighting the effectiveness of the BwB approach. For example, in a study on Douglas-fir, Slavov et al., 2005 used two paternity assignment methods to reconstruct full-sib families within 16 half-sibling families. The program CERVUS (version 2.0), based on likelihood-based assignment, identified parentage with high accuracy. This was complemented by the PFL (Pedigree-Free Likelihood) program, which assigns paternity by genotypic exclusion. Both methods demonstrated the feasibility and effectiveness of the BwB approach (El-Kassaby et al., 2007). Additionally, Lstibůrek et al., 2015a conducted a simulation study examining different sizes of three conceptual populations using the BwB approach. The study demonstrated that it is possible to obtain genetic information and select superior genotypes from commercial forest plantations without the need for controlled breeding. These findings were further supported by studies from Čepl et al., 2018 which also underlined the potential of the BwB strategy for genetic analysis and breeding in forestry.

SSRs can be used to do a preselection of individuals for resequencing based on full-sib prediction, thus avoiding sequencing of less relevant individuals, the genome-wide SNP data can be employed for full-sib validation and construction of genetic maps. Regarding the discrepancies, especially in Fri-8, it becomes clear how crucial the accurate scoring of the SSR markers is. Two SSR markers showed inconclusive peaks in some cases in this population. This further points to scoring errors and/or null alleles being responsible for the incongruence between SSR and SNP markers. It would be interesting to re-evaluate the SSR genotypes in the resequencing data to better understand the underlying cause. Additionally, strong similarity of predicted fathers may be a further indication of unreliable progeny reconstruction and should be carefully checked especially when relying on SSR markers alone. Despite some contradiction between SSR and SNP marker classification, the

preselection of individuals by SSR analysis allowed us to identify relatively large full-sib families for deeper analysis by SNP genotyping. Similar studies have shown comparable results (van Inghelandt et al., 2010). In a study by Zavinon et al. (2020) family identification in Beninese pigeon pea (*Cajanus cajan* (L.)) populations was conducted with 30 informative SSR loci and 794 genotyping-by-sequencing (GBS) derived SNPs. The results with both marker sets were similar, but the PCA based on SNP markers showed the more accurate results.

For genotype-phenotype association studies, crossing and back-crossing generations are essential. For tree species with long generation times, this can be challenging. With the BwB technique we were able to identify full-sib families within the F₁ generation of four different tolerant mother trees. The combination of SSR and SNP markers enabled the successful identification of families that are a valuable resource to perform quantitative trait locus analyses for susceptibility to *H. fraxineus* in future studies.

Acknowledgements

We thank our technical assistants Viktoria Blunk, Annika Eikhof, Katrin Groppe, Marlies Karaus, Heidrun Mattauch, and Regina Zimmermann for their help with laboratory and fieldwork. We thank members of the FraxForFuture project and the Thünen Institute of Forest Genetics for experimental advice and discussion on the project and data analysis. The FraxForFuture research network is funded by the German Federal Ministry of Food and Agriculture and the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety. The projects of the sub-networks are funded by the Waldklimafonds (funding code: 2219WK21A4) and the Thünen Institute. The project executing agency is the Fachagentur für Nachwachsende Rohstoffe e.V. (FNR).

Competing interests statement

The authors declare that they have no competing interests.

References

- Agarwal, M., Shrivastava, N., and Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep* 27, 617–631. doi: 10.1007/s00299-008-0507-z
- Aldana, R., and Freed, D. (2022). Data Processing and Germline Variant Calling with the Sentieon Pipeline. *Methods Mol Biol* 2493, 1–19. doi: 10.1007/978-1-0716-2293-3_1
- Amom, T., Tikendra, L., Apana, N., Goutam, M., Sonia, P., Koiyam, A. S., et al. (2020). Efficiency of RAPD, ISSR, iPBS, SCoT and phytochemical markers in the genetic relationship study of five native and economical important bamboos of North-East India. *Phytochemistry* 174, 112330. doi: 10.1016/j.phytochem.2020.112330
- Bakys, R., Vasaitis, R., and Skovsgaard, J. P. (2013). Patterns and severity of crown dieback in young even-aged stands of european ash (*Fraxinus excelsior* L.) in relation to stand density, bud flushing phenotype, and season. *Plant Protect. Sci.* 49, 120–126. doi: 10.17221/70/2012-PPS
- Bruegmann, T., Fladung, M., and 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
- Capo-chichi, L. J. A., Elakhdar, A., Kubo, T., Nyachiro, J., Juskiw, P., Capettini, F., et al. (2022). Genetic diversity and population structure assessment of Western Canadian barley cooperative trials. *Front Plant Sci* 13, 1006719. doi: 10.3389/fpls.2022.1006719
- Čepl, J., Stejskal, J., Lhotáková, Z., Holá, D., Korecký, J., Lstibůrek, M., et al. (2018). Heritable variation in needle spectral reflectance of Scots pine (*Pinus sylvestris* L.) peaks in red edge. *Remote Sensing of Environment* 219, 89–98. doi: 10.1016/j.rse.2018.10.001
- Coker, T. L. R., Rozsypálek, J., Edwards, A., Harwood, T. P., Butfoy, L., and Buggs, R. J. A. (2019). Estimating mortality rates of European ash (*Fraxinus excelsior*) under the ash dieback (*Hymenoscyphus fraxineus*) epidemic. *Plants, People, Planet* 1, 48–58. doi: 10.1002/ppp3.11
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10. doi: 10.1093/gigascience/giab008
- El-Kassaby, Y. A., and Lstibůrek, M. (2009). Breeding without breeding. *Genetics Research* 91, 111–120. doi: 10.1017/S001667230900007X
- El-Kassaby, Y. A., Lstibůrek, M., Liewlaksaneeyanawin, C., Slavov, G. T., and Howe G. T., eds (2007). Breeding without breeding: approach, example, and proof of concept. In *Proceedings of the IUFRO division 2 joint conference: low input breeding and conservation of forest genetic resources*. Antalya, Turkey.
- Enderle, R., Nakou, A., Thomas, K., and Metzler, B. (2015). Susceptibility of autochthonous German *Fraxinus excelsior* clones to *Hymenoscyphus pseudoalbidus* is genetically determined. *Annals of Forest Science* 72, 183–193. doi: 10.1007/s13595-014-0413-1

- Evans, M. R. (2019). Will natural resistance result in populations of ash trees remaining in British woodlands after a century of ash dieback disease? *R Soc Open Sci* 6, 190908. doi: 10.1098/rsos.190908
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720. doi: 10.1093/bioinformatics/btv428
- García, C., Guichoux, E., and Hampe, A. (2018). A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (*Juniperus phoenicea* ssp. *turbinata*). *Tree Genetics & Genomes* 14, 1–9. doi: 10.1007/s11295-018-1301-x
- Harper, A. L., McKinney, L. V., Nielsen, L. R., Havlickova, L., Li, Y., Trick, M., et al. (2016). Molecular markers for tolerance of European ash (*Fraxinus excelsior*) to dieback disease identified using Associative Transcriptomics. *Sci Rep* 6, 19335. doi: 10.1038/srep19335
- Howe, G. T., Jayawickrama, K., Kolpak, S. E., Kling, J., Trappe, M., Hipkins, V., et al. (2020). An Axiom SNP genotyping array for Douglas-fir. *BMC Genomics* 21, 9. doi: 10.1186/s12864-019-6383-9
- Hu, J.-Y., Zhou, Y., He, F., Dong, X., Liu, L.-Y., Coupland, G., et al. (2014). miR824-regulated AGAMOUS-LIKE16 contributes to flowering time repression in Arabidopsis. *Plant Cell* 26, 2024–2037. doi: 10.1105/tpc.114.124685
- Husson, C., Scala, B., Caël, O., Frey, P., Feau, N., Ioos, R., et al. (2011). *Chalara fraxinea* is an invasive pathogen in France. *Eur J Plant Pathol* 130, 311–324. doi: 10.1007/s10658-011-9755-9
- Jiang, K., Xie, H., Liu, T., Liu, C., and Huang, S. (2020). Genetic diversity and population structure in *Castanopsis fissa* revealed by analyses of sequence-related amplified polymorphism (SRAP) markers. *Tree Genetics & Genomes* 16, 1–10. doi: 10.1007/s11295-020-01442-2
- Jones, O. R., and Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol Ecol Resour* 10, 551–555. doi: 10.1111/j.1755-0998.2009.02787.x
- Kemp, F. (2003). Modern Applied Statistics with S. *J Royal Statistical Soc D* 52, 704–705. doi: 10.1046/j.1467-9884.2003.t01-19-00383_22.x
- Knaus, B. J., and Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* 17, 44–53. doi: 10.1111/1755-0998.12549
- Kowalski, T. (2006). *Chalara fraxinea* sp. nov. associated with dieback of ash (*Fraxinus excelsior*) in Poland. *For. Path.* 36, 264–270. doi: 10.1111/j.1439-0329.2006.00453.x
- Landolt, J., Gross, A., Holdenrieder, O., and Pautasso, M. (2016). Ash dieback due to *Hymenoscyphus fraxineus*: what can be learnt from evolutionary ecology? *Plant Pathol* 65, 1056–1070. doi: 10.1111/ppa.12539
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Lobo, A., Hansen, J. K., McKinney, L. V., Nielsen, L. R., and Kjær, E. D. (2014). Genetic variation in dieback resistance: growth and survival of *Fraxinus*

- excelsior* under the influence of *Hymenoscyphus pseudoalbidus*. Scandinavian Journal of Forest Research 29, 519–526. doi: 10.1080/02827581.2014.950603
- Lobo, A., McKinney, L. V., Hansen, J. K., Kjaer, E. D., and Nielsen, L. R. (2015). Genetic variation in dieback resistance in *Fraxinus excelsior* confirmed by progeny inoculation assay. For. Path. 45, 379–387. doi: 10.1111/efp.12179
- Lstibůrek, M., Hodge, G. R., and Lachout, P. (2015a). Uncovering genetic information from commercial forest plantations—making up for lost time using “Breeding without Breeding”. Tree Genetics & Genomes 11, 1–12. doi: 10.1007/s11295-015-0881-y
- Lstibůrek, M., Hodge, G. R., and Lachout, P. (2015b). Uncovering genetic information from commercial forest plantations—making up for lost time using “Breeding without Breeding”. Tree Genetics & Genomes 11, 1–12. doi: 10.1007/s11295-015-0881-y
- Lstibůrek, M., Ivanková, K., Kadlec, J., Kobliha, J., Klápště, J., and El-Kassaby, Y. A. (2011). Breeding without breeding: minimum fingerprinting effort with respect to the effective population size. Tree Genetics & Genomes 7, 1069–1078. doi: 10.1007/s11295-011-0395-1
- McKinney, L. V., Nielsen, L. R., Collinge, D. B., Thomsen, I. M., Hansen, J. K., and Kjaer, E. D. (2014). The ash dieback crisis: genetic variation in resistance can prove a long-term solution. Plant Pathol 63, 485–499. doi: 10.1111/ppa.12196
- McKinney, L. V., Nielsen, L. R., Hansen, J. K., and Kjær, E. D. (2011). Presence of natural genetic resistance in *Fraxinus excelsior* (Oleraceae) to *Chalara fraxinea* (Ascomycota): an emerging infectious disease. Heredity 106, 788–797. doi: 10.1038/hdy.2010.119
- McKinney, L. V., Thomsen, I. M., Kjaer, E. D., and Nielsen, L. R. (2012). Genetic resistance to *Hymenoscyphus pseudoalbidus* limits fungal growth and symptom occurrence in *Fraxinus excelsior*. For. Path. 42, 69–74. doi: 10.1111/j.1439-0329.2011.00725.x
- Muñoz, F., Marçais, B., Dufour, J., and Dowkiw, A. (2016). Rising Out of the Ashes: Additive genetic variation for crown and collar resistance to *Hymenoscyphus fraxineus* in *Fraxinus excelsior*. Phytopathology 106, 1535–1543. doi: 10.1094/PHYTO-11-15-0284-R
- Nielsen, L. R., McKinney, L. V., and Kjær, E. D. (2017). Host phenological stage potentially affects dieback severity after *Hymenoscyphus fraxineus* infection in *Fraxinus excelsior* seedlings. Baltic Forestry 23, 229–232.
- Pliura, A., Lygis, V., Marčiulyniene, D., Suchockas, V., and Bakys, R. (2016). Genetic variation of *Fraxinus excelsior* half-sib families in response to ash dieback disease following simulated spring frost and summer drought treatments. iForest 9, 12–22. doi: 10.3832/ifor1514-008
- Pliura A, Lygis V, Suchockas V, and Bartkevicius E. (2011). Performance of Twenty Four European *Fraxinus excelsior* Populations in three Lithuanian progeny trials with a special emphasis on resistance to *Chalara Fraxinea*. Baltic Forestry 17, 17–34.
- Pliura AL, B. V. I. (2007). Genetic variation in adaptive traits of progenies of Lithuanian and western European populations of *Fraxinus excelsior* L. Baltic Forestry 13, 28–38.
- Plumb, W. J., Coker, T. L. R., Stocks, J. J., Woodcock, P., Quine, C. P., Nemesio-Gorriz, M., et al. (2020). The viability of a breeding programme for ash in the British Isles in the face of ash dieback. Plants, People, Planet 2, 29–40. doi: 10.1002/ppp3.10060

- Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., van der Auwera, G. A., et al. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. doi: 10.1101/201178
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81, 559–575. doi: 10.1086/519795
- R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for*. Vienna, Austria.
- Ramesh, P., Mallikarjuna, G., Sameena, S., Kumar, A., Gurulakshmi, K., Reddy, B. V., et al. (2020). Advancements in molecular marker technologies and their applications in diversity studies. *J Biosci* 45, 1–15. doi: 10.1007/s12038-020-00089-4
- Slavov, G. T., Howe, G. T., and Adams, W. T. (2005). Pollen contamination and mating patterns in a Douglas-fir seed orchard as measured by simple sequence repeat markers. *Can. J. For. Res.* 35, 1592–1603. doi: 10.1139/x05-082
- Sollars, E. S. A., Harper, A. L., Kelly, L. J., Sambles, C. M., Ramirez-Gonzalez, R. H., Swarbreck, D., et al. (2017). Genome sequence and genetic diversity of European ash trees. *Nature* 541, 212–216. doi: 10.1038/nature20786
- Stener, L.-G. (2013). Clonal differences in susceptibility to the dieback of *Fraxinus excelsior* in southern Sweden. *Scandinavian Journal of Forest Research* 28, 205–216. doi: 10.1080/02827581.2012.735699
- van Inghelandt, D., Melchinger, A. E., Lebreton, C., and Stich, B. (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor Appl Genet* 120, 1289–1299. doi: 10.1007/s00122-009-1256-2
- Venables, W. N., and Ripley, B. D. (2003). *Modern applied statistics with S*. New York [u.a.]: Springer. ISBN: 9780387954578.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer International Publishing. ISBN: 9783319242774.
- Zavinon, F., Adoukonou-Sagbadja, H., Keilwagen, J., Lehnert, H., Ordon, F., and Perovic, D. (2020). Genetic diversity and population structure in Beninese pigeon pea [*Cajanus cajan* (L.) Huth] landraces collection revealed by SSR and genome wide SNP markers. *Genet Resour Crop Evol* 67, 191–208. doi: 10.1007/s10722-019-00864-9
- Zeng, W., Su, Y., Huang, R., Hu, D., Huang, S., and Zheng, H. (2023a). Insight into the complex genetic relationship of chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.) advanced parent trees based on SSR and SNP datasets. *Forests* 14, 347. doi: 10.3390/f14020347
- Zeng, W., Su, Y., Huang, R., Hu, D., Huang, S., and Zheng, H. (2023b). Insight into the complex genetic relationship of chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.) advanced parent trees based on SSR and SNP datasets. *Forests* 14, 347. doi: 10.3390/f14020347
- Zhao, Y.-J., Hosoya, T., Baral, H.-O., Hosaka, K., and Kakishima, M. (2013). *Hymenoscyphus pseudoalbidus*, the correct name for *Lambertella albida* reported from Japan. *Mycotaxon* 122, 25–41. doi: 10.5248/122.25
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606

Supplementary

Data availability

All genomic variants have been deposited in the European Variant Archive (EVA) under the accession number 'PRJEB64325'.

Supplemental materials on COLONY parameters for family estimations with SSR marker

For the analytical process with the program COLONY, the following parameters were chosen: Mating system: male polygam and female polygam, with inbreeding, without clones, dioecious and diploid, length of run: medium, analysis method: full likelihood, likelihood precision: high, run specification: default and no sibship prior.

Supplementary Tables S3-S6 are available in the preprint version of the publication on <https://www.biorxiv.org/content/10.1101/2023.07.18.549475v1.supplementary-material> bioRxiv:



Table S1: PCR master mix

	one sample
Multiplex-Mix	6.25 μ l
ddH ₂ O	2.75 μ l
10x Primer-Mix SetF3	1.25 μ l
5x Q-Solution	1.25 μ l
total	11.5 μ l

Table S2: Polymerase chain reaction program

temperature	Time	Comments/notes
--------------------	-------------	-----------------------

95°C	15 min	
95°C	30 sec	16x (-0,5 °C/cycle)
60°C	90 sec	
72°C	60 sec	
95°C	30 sec	25x
52°C	90 sec	
72°C	60 sec	
60°C	30 min	
8°C	storage	Within PCR-machine

Table S3: Allele report of offspring genotypes of Eve-2 of the program COLONY

Table S4: Allele report of offspring genotypes of Dar-18 of the program COLONY.

Table S5: Allele report of offspring genotypes of Fri-8 of the program COLONY.

Table S6: Allele report of offspring genotypes of Kar -4 of the program COLONY.

Table S7: Mother alleles

Mothertree	F12		F23		F27		F26		F25		F30		F29		F28	
Fri-08	89	89	125	129	141	147	159	175	214	214	242	244	252	258	277	309
Kar-04	89	97	125	129	139	141	173	173	210	216	236	256	242	242	291	311
Eve-02	83	83	121	129	141	147	157	173	214	214	238	254	256	258	281	309
Dar-18	87	103	121	121	145	147	165	171	212	212	254	260	242	242	281	283
-	87	99	125	125	141	147	171	175	210	214	240	250	244	250	277	281

Table S8: Predicted full sib progenies COLONY. yellow marks: unique offsprings between colony runs per progeny.

Fri-8		F008		F008	
first run	second run	first run	second run	first run	second run
FS 10	FS 10	FS 17	FS 17	FS 22	FS 23
F008_I_12	F008_I_12	F008_I_25	F008_I_25	F008_I_34	F008_I_34
F008_I_47	F008_I_47	F008_I_40	F008_II_22	F008_I_41	F008_I_41
F008_I_75	F008_I_71	F008_I_66	F008_II_30	F008_I_42	F008_I_42
F008_II_19	F008_I_75	F008_I_71	F008_II_38	F008_II_04	F008_II_04
F008_II_36	F008_I_82	F008_I_82	F008_II_84	F008_II_10	F008_II_10
F008_II_47	F008_II_19	F008_I_88	F008_III_51	F008_II_15	F008_II_15
F008_II_78	F008_II_47	F008_II_22	F008_III_56	F008_II_26	F008_II_26
F008_II_79	F008_II_63	F008_II_30	F008_III_62	F008_II_48	F008_II_36
F008_II_81	F008_II_78	F008_II_38	F008_III_67	F008_II_49	F008_II_48
F008_II_93	F008_II_81	F008_II_61	F008_III_70	F008_II_53	F008_II_49
F008_III_08	F008_II_86	F008_II_63	F008_III_85	F008_II_60	F008_II_53
F008_III_22	F008_III_08	F008_II_65	F008_IV_08	F008_II_67	F008_II_60
F008_III_32	F008_III_22	F008_II_76	F008_IV_22	F008_II_75	F008_II_67
F008_III_39	F008_III_39	F008_II_84	F008_IV_34	F008_II_94	F008_II_75

F008_III_52	F008_III_60	F008_II_86	F008_IV_86	F008_III_02	F008_II_94
F008_III_56	F008_III_68	F008_III_10	F008_V_14	F008_III_28	F008_III_02
F008_III_62	F008_III_74	F008_III_23	F008_V_40	F008_IV_15	F008_III_28
F008_III_63	F008_III_78	F008_III_51	F008_V_81	F008_IV_44	F008_III_32
F008_III_68	F008_III_79	F008_III_57	F008_V_87	F008_IV_51	F008_III_52
F008_III_78	F008_III_84	F008_III_60	F008_V_90	F008_IV_62	F008_III_63
F008_III_79	F008_III_87	F008_III_67	F008_VI_85	F008_IV_69	F008_III_72
F008_III_84	F008_III_92	F008_III_70	F008_VII_03	F008_IV_89	F008_IV_09
F008_III_87	F008_IV_12	F008_III_72	F008_VII_51	F008_V_05	F008_IV_15
F008_IV_12	F008_IV_81	F008_III_74	F008_VII_86	F008_V_15	F008_IV_44
F008_IV_19	F008_IV_88	F008_III_85	F008_VII_89	F008_V_25	F008_IV_51
F008_IV_34	F008_IV_90	F008_III_92	F008_VIII_23	F008_V_38	F008_IV_62
F008_IV_86	F008_V_16	F008_IV_08	F008_VIII_40	F008_V_41	F008_IV_69
F008_V_16	F008_V_22	F008_IV_09	F008_IX_11	F008_V_47	F008_IV_79
F008_V_22	F008_V_39	F008_IV_11	F008_IX_38	F008_V_56	F008_IV_89
F008_V_33	F008_V_63	F008_IV_16	F008_IX_57	F008_V_62	F008_V_05
F008_V_39	F008_V_72	F008_IV_20	F008_IX_60	F008_V_71	F008_V_15
F008_V_40	F008_V_73	F008_IV_22	F008_IX_68	F008_V_79	F008_V_25
F008_V_55	F008_V_74	F008_IV_26	F008_IX_69	F008_VI_01	F008_V_38
F008_V_72	F008_VI_03	F008_IV_50	F008_IX_79	F008_VI_12	F008_V_41
F008_V_74	F008_VI_25	F008_IV_74	F008_IX_94	F008_VI_33	F008_V_47
F008_VI_03	F008_VI_50	F008_IV_79	F008_X_11	F008_VI_42	F008_V_55
F008_VI_08	F008_VI_82	F008_IV_81	F008_X_18	F008_VI_49	F008_V_56
F008_VI_25	F008_VII_12	F008_IV_88	F008_X_21	F008_VI_52	F008_V_62
F008_VI_50	F008_VII_15	F008_IV_90	F008_X_22	F008_VI_63	F008_V_71
F008_VI_82	F008_VII_19	F008_IV_91	F008_X_28	F008_VI_76	F008_V_79
F008_VII_03	F008_VII_37	F008_V_14	F008_X_42	F008_VII_22	F008_VI_01
F008_VII_11	F008_VII_49	F008_V_63	F008_X_50	F008_VII_30	F008_VI_12
F008_VII_12	F008_VII_74	F008_V_67	F008_X_51	F008_VII_35	F008_VI_33
F008_VII_15	F008_VII_76	F008_V_73	F008_X_64	F008_VII_47	F008_VI_42
F008_VII_19	F008_VIII_03	F008_V_81	F008_X_85	F008_VII_52	F008_VI_49
F008_VII_49	F008_VIII_65	F008_V_87	F008_X_86	F008_VII_54	F008_VI_52
F008_VII_74	F008_VIII_70	F008_V_90	F008_X_87	F008_VII_81	F008_VI_63
F008_VII_87	F008_VIII_87	F008_VI_85		F008_VII_83	F008_VI_76
F008_VII_89	F008_VIII_91	F008_VI_91		F008_VII_92	F008_VII_22
F008_VIII_13	F008_VIII_93	F008_VII_37		F008_VII_94	F008_VII_30
F008_VIII_65	F008_VIII_96	F008_VII_51		F008_VII_95	F008_VII_35
F008_VIII_87	F008_IX_09	F008_VII_55		F008_VII_96	F008_VII_47
F008_VIII_91	F008_IX_21	F008_VII_59		F008_VIII_10	F008_VII_52
F008_VIII_93	F008_IX_29	F008_VII_70		F008_VIII_11	F008_VII_54
F008_VIII_96	F008_IX_45	F008_VII_76		F008_VIII_16	F008_VII_81
F008_IX_09	F008_IX_47	F008_VII_86		F008_VIII_24	F008_VII_83
F008_IX_15	F008_IX_82	F008_VIII_03		F008_VIII_42	F008_VII_92
F008_IX_21	F008_IX_90	F008_VIII_23		F008_VIII_52	F008_VII_94
F008_IX_22	F008_X_09	F008_VIII_40		F008_VIII_56	F008_VII_95
F008_IX_37	F008_X_29	F008_VIII_70		F008_VIII_57	F008_VII_96
F008_IX_47	F008_X_32	F008_IX_11		F008_VIII_81	F008_VIII_10
F008_IX_50	F008_X_34	F008_IX_29		F008_VIII_82	F008_VIII_11
F008_IX_53	F008_X_43	F008_IX_34		F008_VIII_90	F008_VIII_16
F008_IX_71	F008_X_44	F008_IX_38			F008_VIII_24

F008_IX_72	F008_X_49	F008_IX_45	F008_VIII_42
F008_IX_76	F008_X_55	F008_IX_57	F008_VIII_52
F008_IX_79	F008_X_57	F008_IX_60	F008_VIII_56
F008_IX_82	F008_X_84	F008_IX_68	F008_VIII_57
F008_X_02		F008_IX_69	F008_VIII_81
F008_X_03		F008_IX_90	F008_VIII_82
F008_X_28		F008_IX_94	F008_VIII_90
F008_X_29		F008_X_04	F008_IX_15
F008_X_34		F008_X_09	F008_IX_22
F008_X_40		F008_X_11	F008_IX_37
F008_X_43		F008_X_18	F008_IX_50
F008_X_44		F008_X_21	F008_IX_53
F008_X_49		F008_X_22	F008_IX_71
F008_X_51		F008_II_01	F008_IX_72
F008_X_57		F008_X_37	F008_IX_76
F008_X_72		F008_X_39	F008_X_02
F008_X_91		F008_X_42	F008_X_03
		F008_X_50	F008_X_04
		F008_X_55	F008_X_37
		F008_X_60	F008_X_40
		F008_X_64	F008_X_72
		F008_X_82	F008_X_82
		F008_X_84	F008_X_91
		F008_X_85	
		F008_X_86	
		F008_X_87	

Kar-04		F009					
first run	second run	first run	second run	first run	second run	first run	second run
FS 05	FS 06	FS 21	FS 84	FS 29	FS 30	FS 41	FS 43
F009_I_05	F009_I_06	F009_I_24	F009_II_90	F009_I_35	F009_I_35	F009_I_56	F009_I_56
F009_I_15	F009_I_22	F009_II_90	F009_III_22	F009_I_48	F009_I_48	F009_I_58	F009_I_58
F009_I_22	F009_I_59	F009_III_22	F009_III_55	F009_I_50	F009_I_50	F009_I_64	F009_I_64
F009_I_59	F009_I_65	F009_III_55	F009_III_77	F009_I_61	F009_I_61	F009_I_75	F009_I_75
F009_I_65	F009_II_27	F009_III_77	F009_III_92	F009_I_74	F009_I_74	F009_I_83	F009_I_83
F009_II_27	F009_II_75	F009_III_92	F009_IV_36	F009_I_77	F009_I_77	F009_I_85	F009_I_85
F009_II_75	F009_III_31	F009_IV_36	F009_V_51	F009_I_79	F009_I_79	F009_II_38	F009_II_38
F009_III_26	F009_III_49	F009_V_51	F009_V_73	F009_I_87	F009_I_87	F009_II_65	F009_II_65
F009_III_31	F009_III_75	F009_V_73	F009_V_76	F009_I_88	F009_I_88	F009_II_73	F009_II_73
F009_III_49	F009_IV_33	F009_V_76	F009_V_96	F009_I_92	F009_I_92	F009_II_82	F009_II_82
F009_III_75	F009_IV_39	F009_V_96	F009_VI_17	F009_I_93	F009_I_93	F009_III_08	F009_III_08
F009_III_96	F009_V_27	F009_VI_17	F009_VI_78	F009_II_05	F009_II_05	F009_IV_04	F009_IV_04
F009_IV_24	F009_V_36	F009_VI_78	F009_VI_87	F009_II_21	F009_II_21	F009_IV_06	F009_IV_06
			F009_VII_2				
F009_IV_33	F009_V_75	F009_VI_87	3	F009_II_34	F009_II_34	F009_IV_07	F009_IV_07
		F009_VII_2	F009_VII_2				
F009_IV_39	F009_V_95	3	7	F009_II_74	F009_II_72	F009_IV_29	F009_IV_29
		F009_VII_2	F009_VII_3				
F009_V_27	F009_VI_22	7	3	F009_II_83	F009_II_74	F009_IV_40	F009_IV_40

F009_V_36	F009_VI_29	F009_VII_3	F009_VII_6	F009_II_85	F009_II_83	F009_V_20	F009_V_20
F009_V_46	F009_VIII_73	F009_VII_6	F009_VII_9	F009_II_88	F009_II_85	F009_V_45	F009_V_45
F009_V_75	F009_VIII_77	F009_VII_9	F009_VIII_08	F009_III_30	F009_II_88	F009_V_67	F009_V_67
F009_V_95	F009_IX_13	F009_VIII_08	F009_VIII_66	F009_III_48	F009_III_30	F009_VI_18	F009_VI_18
F009_VI_22	F009_IX_51	F009_VIII_66	F009_VIII_96	F009_III_52	F009_III_37	F009_VI_76	F009_VI_76
F009_VI_29	F009_IX_85	F009_VIII_96	F009_IX_10	F009_III_63	F009_III_48	F009_VII_4	F009_VII_4
F009_VIII_54	F009_IX_91	F009_IX_10	F009_IX_18	F009_III_68	F009_III_52	F009_VII_7	F009_VII_7
F009_VIII_73	F009_X_02	F009_IX_18	F009_IX_37	F009_III_72	F009_III_63	F009_VII_9	F009_VII_9
F009_VIII_77	F009_X_18	F009_IX_37	F009_IX_55	F009_III_78	F009_III_68	F009_VIII_82	F009_VIII_82
F009_IX_09	F009_X_62	F009_IX_55	F009_X_26	F009_IV_01	F009_III_72	F009_VIII_92	F009_VIII_92
F009_IX_13	F009_X_72	F009_X_26	F009_X_28	F009_IV_18	F009_III_78	F009_IX_11	F009_IX_11
F009_IX_51	F009_X_76	F009_X_28	F009_X_67	F009_IV_23	F009_IV_01	F009_IX_32	F009_IX_32
F009_IX_85		F009_X_67	F009_X_85	F009_IV_28	F009_IV_18	F009_IX_42	F009_IX_42
F009_IX_91		F009_X_85		F009_IV_56	F009_IV_23	F009_IX_92	F009_IX_92
F009_X_02				F009_IV_58	F009_IV_28	F009_X_41	F009_X_41
F009_X_18				F009_IV_63	F009_IV_56	F009_X_58	F009_X_58
F009_X_21				F009_IV_71	F009_IV_58	F009_X_92	F009_X_92
F009_X_59				F009_IV_73	F009_IV_63		
F009_X_62				F009_IV_74	F009_IV_71		
F009_X_72				F009_IV_88	F009_IV_73		
F009_X_76				F009_IV_95	F009_IV_74		
				F009_V_13	F009_IV_88		
				F009_V_16	F009_IV_95		
				F009_V_23	F009_V_13		
				F009_V_35	F009_V_16		
				F009_V_40	F009_V_23		
				F009_V_41	F009_V_35		
				F009_V_50	F009_V_40		
				F009_V_56	F009_V_41		
				F009_V_57	F009_V_50		
				F009_V_59	F009_V_56		
				F009_V_60	F009_V_57		
				F009_V_71	F009_V_59		
				F009_V_78	F009_V_60		
				F009_V_87	F009_V_71		
				F009_V_88	F009_V_78		
				F009_V_92	F009_V_87		
				F009_VI_02	F009_V_88		
				F009_VI_12	F009_V_92		
				F009_VI_26	F009_VI_02		
				F009_VI_41	F009_VI_12		
				F009_VI_45	F009_VI_26		
				F009_VI_57	F009_VI_41		
				F009_VI_63	F009_VI_45		
				F009_VI_67	F009_VI_57		
				F009_VI_73	F009_VI_63		

F009_VII_0	
5	F009_VI_67
F009_VII_0	
8	F009_VI_73
F009_VII_1	F009_VII_0
7	5
F009_VII_3	F009_VII_0
4	8
F009_VII_4	F009_VII_1
4	7
F009_VII_5	F009_VII_3
1	4
F009_VII_5	F009_VII_4
5	4
F009_VII_5	F009_VII_5
6	1
F009_VII_7	F009_VII_5
2	5
F009_VII_7	F009_VII_5
6	6
F009_VII_8	F009_VII_7
0	2
F009_VII_8	F009_VII_7
2	6
F009_VII_8	F009_VII_8
5	0
F009_VII_8	F009_VII_8
8	2
F009_VIII_	F009_VII_8
02	5
F009_VIII_	F009_VII_8
05	8
F009_VIII_	F009_VIII_
14	02
F009_VIII_	F009_VIII_
19	05
F009_VIII_	F009_VIII_
25	14
F009_VIII_	F009_VIII_
27	19
F009_VIII_	F009_VIII_
36	25
F009_VIII_	F009_VIII_
45	27
F009_VIII_	F009_VIII_
78	36
F009_VIII_	F009_VIII_
87	45
	F009_VIII_
F009_IX_02	78
	F009_VIII_
F009_IX_04	87
F009_IX_24	F009_IX_02
F009_IX_28	F009_IX_04
F009_IX_33	F009_IX_24
F009_IX_38	F009_IX_28
F009_IX_41	F009_IX_33
F009_IX_50	F009_IX_38
F009_IX_64	F009_IX_41
F009_IX_74	F009_IX_50
F009_IX_87	F009_IX_64
F009_X_03	F009_IX_74
F009_X_14	F009_IX_87
F009_X_17	F009_X_03

F009_X_20 F009_X_14
 F009_X_30 F009_X_17
 F009_X_39 F009_X_20
 F009_X_51 F009_X_30
 F009_X_52 F009_X_39
 F009_X_56 F009_X_51
 F009_X_57 F009_X_52
 F009_X_60 F009_X_56
 F009_X_63 F009_X_57
 F009_X_66 F009_X_60
 F009_X_68 F009_X_63
 F009_X_91 F009_X_66
 F009_X_93 F009_X_68
 F009_X_91
 F009_X_93

Eve-2	F010	first run	second run	first run	second run
first run	second run	first run	second run	first run	second run
FS 08	FS 08	FS 20	FS 20	FS 30	FS 32
F010_I_08	F010_I_08	F010_I_22	F010_I_22	F010_I_38	F010_I_38
F010_I_18	F010_I_18	F010_I_24	F010_I_24	F010_I_58	F010_I_58
F010_I_25	F010_I_25	F010_I_40	F010_I_40	F010_I_61	F010_I_61
F010_I_33	F010_I_33	F010_I_51	F010_I_51	F010_II_01	F010_II_01
F010_I_47	F010_I_47	F010_I_53	F010_I_53	F010_II_34	F010_II_34
F010_I_49	F010_I_49	F010_I_54	F010_I_54	F010_II_79	F010_II_79
F010_I_82	F010_I_82	F010_I_59	F010_I_59	F010_II_82	F010_II_82
F010_I_87	F010_I_87	F010_I_62	F010_I_62	F010_III_06	F010_III_06
F010_I_88	F010_I_88	F010_I_69	F010_I_69	F010_III_16	F010_III_16
F010_II_15	F010_II_15	F010_I_77	F010_I_77	F010_III_20	F010_III_20
F010_II_42	F010_II_42	F010_I_81	F010_I_81	F010_III_50	F010_III_50
F010_II_47	F010_II_44	F010_I_84	F010_I_84	F010_III_69	F010_III_69
F010_II_49	F010_II_47	F010_I_92	F010_I_92	F010_III_90	F010_III_90
F010_II_66	F010_II_49	F010_I_95	F010_I_95	F010_IV_30	F010_IV_30
F010_II_73	F010_II_66	F010_II_09	F010_II_09	F010_IV_41	F010_IV_41
F010_III_08	F010_II_73	F010_II_18	F010_II_18	F010_IV_88	F010_IV_88
F010_III_11	F010_III_08	F010_II_25	F010_II_25	F010_IV_91	F010_IV_91
F010_III_18	F010_III_11	F010_II_28	F010_II_28	F010_V_02	F010_V_02
F010_III_21	F010_III_18	F010_II_40	F010_II_40	F010_V_06	F010_V_06
F010_III_24	F010_III_21	F010_II_56	F010_II_56	F010_V_40	F010_V_40
F010_III_28	F010_III_24	F010_II_69	F010_II_69	F010_V_46	F010_V_46
F010_III_38	F010_III_28	F010_II_70	F010_II_70	F010_V_76	F010_V_76
F010_III_40	F010_III_38	F010_II_80	F010_II_80	F010_V_77	F010_V_77
F010_III_56	F010_III_40	F010_II_83	F010_II_83	F010_VI_32	F010_VI_32
F010_III_92	F010_III_56	F010_II_84	F010_II_84	F010_VI_61	F010_VI_61
F010_IV_27	F010_III_92	F010_II_85	F010_II_85	F010_VI_86	F010_VI_86
F010_IV_38	F010_IV_27	F010_III_03	F010_III_03	F010_VII_13	F010_VII_13
F010_IV_39	F010_IV_38	F010_III_07	F010_III_07	F010_VII_27	F010_VII_27
F010_IV_42	F010_IV_39	F010_III_12	F010_III_12	F010_VII_35	F010_VII_35
F010_IV_56	F010_IV_42	F010_III_19	F010_III_19	F010_VII_56	F010_VII_56

F010_IV_66	F010_IV_56	F010_III_22	F010_III_22	F010_VIII_49	F010_VIII_49
F010_IV_69	F010_IV_66	F010_III_25	F010_III_25	F010_IX_01	F010_IX_01
F010_IV_72	F010_IV_69	F010_III_30	F010_III_30	F010_IX_08	F010_IX_08
F010_V_27	F010_IV_72	F010_III_34	F010_III_34	F010_IX_36	F010_IX_36
F010_V_60	F010_V_27	F010_III_41	F010_III_41	F010_IX_89	F010_IX_89
F010_V_72	F010_V_60	F010_III_53	F010_III_53	F010_IX_96	F010_IX_96
F010_VI_06	F010_V_72	F010_III_57	F010_III_57	F010_X_09	F010_X_09
F010_VI_10	F010_VI_06	F010_III_67	F010_III_67	F010_X_23	F010_X_23
F010_VI_16	F010_VI_10	F010_III_72	F010_III_72	F010_X_26	F010_X_26
F010_VI_25	F010_VI_16	F010_III_74	F010_III_74		
F010_VI_29	F010_VI_25	F010_III_78	F010_III_78		
F010_VI_33	F010_VI_29	F010_III_85	F010_III_85		
F010_VI_45	F010_VI_33	F010_III_91	F010_III_91		
F010_VI_54	F010_VI_34	F010_III_96	F010_III_96		
F010_VI_58	F010_VI_45	F010_IV_08	F010_IV_08		
F010_VI_71	F010_VI_54	F010_IV_18	F010_IV_18		
F010_VI_85	F010_VI_58	F010_IV_29	F010_IV_29		
F010_VI_88	F010_VI_71	F010_IV_34	F010_IV_34		
F010_VI_96	F010_VI_85	F010_IV_40	F010_IV_40		
F010_VII_02	F010_VI_88	F010_IV_45	F010_IV_45		
F010_VII_20	F010_VI_96	F010_IV_47	F010_IV_47		
F010_VII_21	F010_VII_02	F010_IV_48	F010_IV_48		
F010_VII_37	F010_VII_20	F010_IV_50	F010_IV_50		
F010_VII_40	F010_VII_21	F010_IV_52	F010_IV_52		
F010_VII_41	F010_VII_37	F010_IV_57	F010_IV_57		
F010_VII_63	F010_VII_40	F010_IV_62	F010_IV_62		
F010_VII_65	F010_VII_41	F010_IV_64	F010_IV_64		
F010_VII_74	F010_VII_63	F010_IV_80	F010_IV_80		
F010_VII_89	F010_VII_65	F010_IV_87	F010_IV_87		
F010_VIII_03	F010_VII_74	F010_IV_89	F010_IV_89		
F010_VIII_22	F010_VII_89	F010_V_05	F010_V_05		
F010_VIII_30	F010_VIII_03	F010_V_07	F010_V_07		
F010_VIII_42	F010_VIII_22	F010_V_08	F010_V_08		
F010_VIII_44	F010_VIII_30	F010_V_12	F010_V_12		
F010_VIII_51	F010_VIII_42	F010_V_22	F010_V_22		
F010_VIII_53	F010_VIII_44	F010_V_39	F010_V_39		
F010_VIII_66	F010_VIII_51	F010_V_45	F010_V_45		
F010_VIII_79	F010_VIII_53	F010_V_51	F010_V_51		
F010_VIII_86	F010_VIII_66	F010_V_53	F010_V_53		
F010_VIII_87	F010_VIII_79	F010_V_58	F010_V_58		
F010_VIII_89	F010_VIII_86	F010_V_67	F010_V_67		
F010_IX_10	F010_VIII_87	F010_V_73	F010_V_73		
F010_IX_15	F010_VIII_89	F010_V_81	F010_V_81		
F010_IX_21	F010_IX_10	F010_V_83	F010_V_83		
F010_IX_32	F010_IX_15	F010_V_89	F010_V_89		
F010_IX_67	F010_IX_21	F010_V_91	F010_V_91		
F010_IX_71	F010_IX_32	F010_V_92	F010_V_92		
F010_X_17	F010_IX_67	F010_V_94	F010_V_94		
F010_X_20	F010_IX_71	F010_V_95	F010_V_95		

F010_X_21	F010_X_17	F010_V_96	F010_V_96
F010_X_25	F010_X_20	F010_VI_01	F010_VI_01
F010_X_30	F010_X_21	F010_VI_05	F010_VI_05
F010_X_80	F010_X_25	F010_VI_07	F010_VI_07
F010_X_85	F010_X_30	F010_VI_08	F010_VI_08
F010_X_88	F010_X_80	F010_VI_14	F010_VI_14
	F010_X_85	F010_VI_18	F010_VI_18
	F010_X_88	F010_VI_20	F010_VI_20
		F010_VI_21	F010_VI_21
		F010_VI_26	F010_VI_26
		F010_VI_27	F010_VI_27
		F010_VI_35	F010_VI_35
		F010_VI_39	F010_VI_39
		F010_VI_40	F010_VI_40
		F010_VI_51	F010_VI_51
		F010_VI_65	F010_VI_65
		F010_VI_68	F010_VI_68
		F010_VI_73	F010_VI_73
		F010_VII_14	F010_VII_14
		F010_VII_15	F010_VII_15
		F010_VII_16	F010_VII_16
		F010_VII_19	F010_VII_19
		F010_VII_32	F010_VII_32
		F010_VII_34	F010_VII_34
		F010_VII_36	F010_VII_36
		F010_VII_39	F010_VII_39
		F010_VII_43	F010_VII_43
		F010_VII_44	F010_VII_44
		F010_VII_53	F010_VII_53
		F010_VII_58	F010_VII_54
		F010_VII_59	F010_VII_58
		F010_VII_60	F010_VII_59
		F010_VII_82	F010_VII_60
		F010_VII_95	F010_VII_82
		F010_VIII_39	F010_VII_95
		F010_VIII_48	F010_VIII_39
		F010_VIII_55	F010_VIII_48
		F010_VIII_58	F010_VIII_55
		F010_VIII_59	F010_VIII_58
		F010_VIII_61	F010_VIII_59
		F010_VIII_72	F010_VIII_61
		F010_VIII_74	F010_VIII_72
		F010_VIII_76	F010_VIII_74
		F010_VIII_77	F010_VIII_76
		F010_VIII_82	F010_VIII_77
		F010_VIII_83	F010_VIII_82
		F010_IX_02	F010_VIII_83
		F010_IX_05	F010_IX_02
		F010_IX_18	F010_IX_05
		F010_IX_20	F010_IX_18

F010_IX_22	F010_IX_20
F010_IX_38	F010_IX_22
F010_IX_39	F010_IX_38
F010_IX_40	F010_IX_39
F010_IX_53	F010_IX_40
F010_IX_55	F010_IX_53
F010_IX_62	F010_IX_55
F010_IX_63	F010_IX_62
F010_IX_76	F010_IX_63
F010_IX_81	F010_IX_76
F010_IX_85	F010_IX_81
F010_IX_87	F010_IX_85
F010_IX_88	F010_IX_87
F010_IX_92	F010_IX_88
F010_IX_94	F010_IX_92
F010_X_02	F010_IX_94
F010_X_03	F010_X_02
F010_X_05	F010_X_05
F010_X_06	F010_X_06
F010_X_08	F010_X_08
F010_X_22	F010_X_22
F010_X_28	F010_X_28
F010_X_31	F010_X_31
F010_X_32	F010_X_32
F010_X_40	F010_X_40
F010_X_44	F010_X_44
F010_X_48	F010_X_48
F010_X_49	F010_X_49
F010_X_53	F010_X_53
F010_X_56	F010_X_56
F010_X_59	F010_X_59
F010_X_61	F010_X_61
F010_X_66	F010_X_66
F010_X_68	F010_X_68
F010_X_71	F010_X_71
F010_X_83	F010_X_83
F010_X_89	F010_X_89
	F010_X_91

Dar-18	F011	first run	second run	first run	second run
first run	second run	first run	second run	first run	second run
FS 13	FS 13	FS 14	FS 15	FS 19	FS 19
F011_I_16	F011_I_16	F011_I_17	F011_I_18	F011_I_22	F011_I_22
F011_I_65	F011_I_65	F011_III_21	F011_III_21	F011_I_23	F011_I_23
F011_II_81	F011_II_81	F011_III_30	F011_III_68	F011_I_26	F011_I_26
F011_III_28	F011_III_28	F011_III_68	F011_IV_06	F011_I_49	F011_I_49
F011_IV_05	F011_IV_05	F011_IV_06	F011_IV_09	F011_I_52	F011_I_52
F011_IV_67	F011_IV_67	F011_IV_09	F011_IV_17	F011_I_54	F011_I_54

F011_V_88	F011_V_88	F011_IV_17	F011_IV_46	F011_I_57	F011_I_57
F011_VI_07	F011_VI_07	F011_IV_46	F011_IV_72	F011_I_59	F011_I_59
F011_VI_86	F011_VI_86	F011_IV_72	F011_IV_76	F011_I_75	F011_I_75
F011_VI_93	F011_VI_93	F011_IV_76	F011_V_06	F011_I_77	F011_I_77
F011_VI_95	F011_VI_95	F011_V_06	F011_V_29	F011_II_23	F011_II_23
F011_VII_48	F011_VII_48	F011_V_29	F011_V_38	F011_II_28	F011_II_28
F011_VIII_67	F011_VIII_67	F011_V_30	F011_VI_78	F011_II_29	F011_II_29
F011_IX_03	F011_IX_03	F011_V_38	F011_VI_85	F011_II_43	F011_II_31
F011_IX_13	F011_IX_13	F011_V_49	F011_VII_65	F011_II_48	F011_II_43
F011_IX_23	F011_IX_23	F011_VI_78	F011_VII_66	F011_II_64	F011_II_48
F011_IX_33	F011_IX_33	F011_VI_85	F011_VIII_22	F011_II_87	F011_II_64
F011_IX_54	F011_IX_54	F011_VII_65	F011_VIII_33	F011_III_04	F011_II_87
F011_IX_88	F011_IX_88	F011_VII_66	F011_VIII_36	F011_III_07	F011_III_04
F011_X_20	F011_X_20	F011_VIII_22	F011_VIII_84	F011_III_08	F011_III_07
		F011_VIII_33	F011_IX_48	F011_III_18	F011_III_08
		F011_VIII_35	F011_IX_64	F011_III_23	F011_III_18
		F011_VIII_36	F011_X_86	F011_III_35	F011_III_23
		F011_VIII_55		F011_III_38	F011_III_35
		F011_VIII_84		F011_III_46	F011_III_38
		F011_IX_48		F011_III_53	F011_III_46
		F011_IX_64		F011_III_73	F011_III_53
		F011_X_86		F011_III_74	F011_III_73
				F011_III_78	F011_III_74
				F011_III_79	F011_III_78
				F011_IV_04	F011_III_79
				F011_IV_07	F011_IV_04
				F011_IV_12	F011_IV_07
				F011_IV_44	F011_IV_12
				F011_IV_51	F011_IV_44
				F011_IV_55	F011_IV_51
				F011_IV_75	F011_IV_55
				F011_IV_78	F011_IV_75
				F011_V_01	F011_IV_78
				F011_V_08	F011_V_01
				F011_V_17	F011_V_08
				F011_V_18	F011_V_17
				F011_V_19	F011_V_18
				F011_V_20	F011_V_19
				F011_V_24	F011_V_20
				F011_V_25	F011_V_24
				F011_V_51	F011_V_25
				F011_V_55	F011_V_51
				F011_V_59	F011_V_55
				F011_V_61	F011_V_59
				F011_V_70	F011_V_61
				F011_V_75	F011_V_70
				F011_V_81	F011_V_75
				F011_VI_15	F011_V_81
				F011_VI_39	F011_VI_15
				F011_VI_43	F011_VI_39

F011_VI_51	F011_VI_43
F011_VI_52	F011_VI_51
F011_VI_53	F011_VI_52
F011_VI_60	F011_VI_53
F011_VI_74	F011_VI_60
F011_VI_79	F011_VI_74
F011_VI_82	F011_VI_79
F011_VI_90	F011_VI_82
F011_VI_92	F011_VI_90
F011_VII_06	F011_VI_92
F011_VII_09	F011_VII_06
F011_VII_17	F011_VII_09
F011_VII_18	F011_VII_17
F011_VII_35	F011_VII_18
F011_VII_44	F011_VII_35
F011_VII_45	F011_VII_44
F011_VII_49	F011_VII_45
F011_VII_50	F011_VII_49
F011_VII_54	F011_VII_50
F011_VII_55	F011_VII_54
F011_VII_59	F011_VII_55
F011_VII_60	F011_VII_59
F011_VII_63	F011_VII_60
F011_VII_70	F011_VII_63
F011_VII_72	F011_VII_70
F011_VII_74	F011_VII_72
F011_VIII_24	F011_VII_74
F011_VIII_25	F011_VIII_24
F011_VIII_27	F011_VIII_25
F011_VIII_43	F011_VIII_27
F011_VIII_47	F011_VIII_43
F011_VIII_57	F011_VIII_47
F011_VIII_60	F011_VIII_57
F011_VIII_61	F011_VIII_60
F011_VIII_65	F011_VIII_61
F011_VIII_70	F011_VIII_65
F011_VIII_71	F011_VIII_70
F011_VIII_79	F011_VIII_71
F011_VIII_80	F011_VIII_79
F011_IX_05	F011_VIII_80
F011_IX_14	F011_IX_05
F011_IX_27	F011_IX_14
F011_IX_31	F011_IX_27
F011_IX_34	F011_IX_31
F011_IX_37	F011_IX_34
F011_IX_47	F011_IX_37
F011_IX_53	F011_IX_47
F011_IX_56	F011_IX_53
F011_IX_68	F011_IX_56

F011_IX_74	F011_IX_68
F011_IX_76	F011_IX_74
F011_IX_77	F011_IX_76
F011_IX_79	F011_IX_77
F011_IX_81	F011_IX_79
F011_IX_85	F011_IX_81
F011_IX_94	F011_IX_85
F011_X_02	F011_IX_94
F011_X_14	F011_X_02
F011_X_28	F011_X_14
F011_X_33	F011_X_28
F011_X_51	F011_X_33
F011_X_72	F011_X_51
F011_X_73	F011_X_72
F011_X_76	F011_X_73
F011_X_79	F011_X_76
F011_X_83	F011_X_79
F011_X_92	F011_X_83
	F011_X_92

Table S9: Predicted fathers COLONY. green marks: allels occurring more than twice within these fathers per progeny.

progeny Fri-8		F008				second run	
father	marker	allele 1	allele 2	father	marker	allele 1	allele 2
*10	F24	96	104	*10	F24	96	104
*10	F23	121	121	*10	F23	121	121
*10	F27	141	147	*10	F27	141	147
*10	F26	167	175	*10	F26	167	175
*10	F25	212	214	*10	F25	212	214
*10	F30	238	244	*10	F30	244	244
*10	F29	264	268	*10	F29	258	264
*10	F28	277	309	*10	F28	277	309
*17	F24	96	104	*17	F24	96	104
*17	F23	121	121	*17	F23	121	121
*17	F27	141	147	*17	F27	141	147
*17	F26	167	175	*17	F26	167	175
*17	F25	212	214	*17	F25	212	214
*17	F30	238	244	*17	F30	238	244
*17	F29	252	258	*17	F29	252	268
*17	F28	277	309	*17	F28	277	309
*22	F24	96	104	*23	F24	96	104
*22	F23	121	121	*23	F23	121	121
*22	F27	141	147	*23	F27	141	147
*22	F26	167	175	*23	F26	167	175
*22	F25	212	214	*23	F25	212	214
*22	F30	240	240	*23	F30	238	240

*22	F29	258	264	*23	F29	258	264
*22	F28	277	309	*23	F28	277	309

progeny Kar-4				progeny Eve-2			
F009				F010			
predicted fathers are identical in run one and run two				predicted fathers are identical in run one and run two			
fathers	marker	allel 1	allel 2	fathers	marker	allel 1	allel 2
*5 / *6	F12	101	103	*8 / *8	F12	87	101
*5 / *6	F23	125	127	*8 / *8	F23	121	125
*5 / *6	F27	141	147	*8 / *8	F27	141	147
*5 / *6	F26	167	175	*8 / *8	F26	167	175
*5 / *6	F25	212	214	*8 / *8	F25	208	214
*5 / *6	F30	238	256	*8 / *8	F30	250	256
*5 / *6	F29	250	272	*8 / *8	F29	242	242
*5 / *6	F28	281	287	*8 / *8	F28	281	295
*21 / *84	F12	87	87	*20 / *20*	F12	87	101
*21 / *84	F23	121	125	*20 / *20*	F23	125	127
*21 / *84	F27	145	147	*20 / *20*	F27	145	147
*21 / *84	F26	163	171	*20 / *20*	F26	171	173
*21 / *84	F25	210	218	*20 / *20*	F25	214	214
*21 / *84	F30	238	238	*20 / *20*	F30	242	268
*21 / *84	F29	250	262	*20 / *20*	F29	234	272
*21 / *84	F28	309	317	*20 / *20*	F28	281	285
*29 / *30	F12	101	103	*30 / *32	F12	89	91
*29 / *30	F23	121	127	*30 / *32	F23	121	127
*29 / *30	F27	141	147	*30 / *32	F27	141	147
*29 / *30	F26	167	175	*30 / *32	F26	157	171
*29 / *30	F25	212	214	*30 / *32	F25	212	214
*29 / *30	F30	236	238	*30 / *32	F30	238	268
*29 / *30	F29	250	272	*30 / *32	F29	242	258
*29 / *30	F28	281	287	*30 / *32	F28	281	307
*41 / *43	F12	101	103				
*41 / *43	F23	121	125				
*41 / *43	F27	139	147				
*41 / *43	F26	167	175				
*41 / *43	F25	212	214				
*41 / *43	F30	236	256				
*41 / *43	F29	250	272				
*41 / *43	F28	281	287				

progeny Dar-18

F011

predicted fathers are identical in run one and run two

fathers	marker	allele 1	allele 2
*13 / *13	F12	87	103
*13 / *13	F23	127	129
*13 / *13	F27	147	149
*13 / *13	F26	165	167
*13 / *13	F25	212	214
*13 / *13	F30	240	298
*13 / *13	F29	242	262
*13 / *13	F28	283	293
*14 / *15	F12	87	103
*14 / *15	F23	127	127
*14 / *15	F27	147	149
*14 / *15	F26	169	173
*14 / *15	F25	214	216
*14 / *15	F30	238	296
*14 / *15	F29	256	262
*14 / *15	F28	293	315
*19 / *19	F12	87	93
*19 / *19	F23	121	121
*19 / *19	F27	147	147
*19 / *19	F26	175	177
*19 / *19	F25	214	214
*19 / *19	F30	240	260
*19 / *19	F29	250	270
*19 / *19	F28	313	329

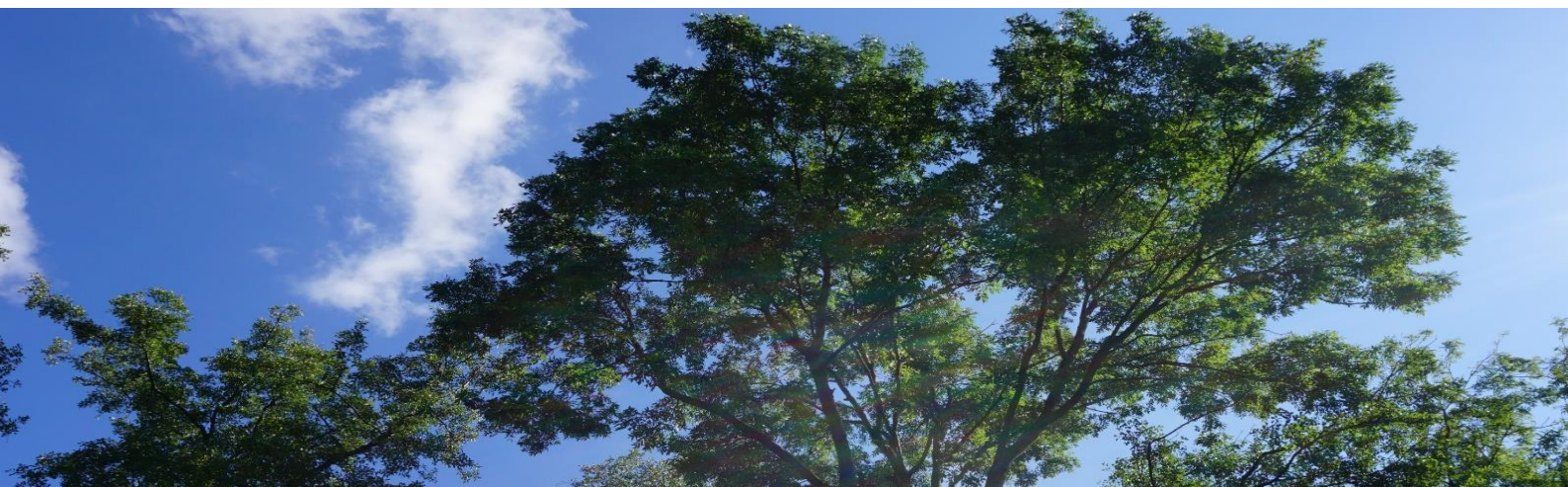
Chapter 2

Construction and Comparative Analysis of High-Density Linkage Maps in *Fraxinus excelsior*

Melina Krautwurst, Birgit Kersten & Niels A. Müller

Thünen Institute of Forest Genetics, Großhansdorf, Germany

In preparation



Abstract

Genetic maps are vital for understanding the genomic architecture. The present investigations aim to construct high-density genetic linkage maps from four full-sibling (full-sib) families of common ash (*Fraxinus excelsior* L.). Further, maternal and paternal differences were compared. We used single nucleotide polymorphism (SNP) markers and the Lep-MAP3 program to generate genetic linkage maps. The genetic linkage maps were compared with a chromosome-level reference genome assembly of *Fraxinus excelsior*. The comparison of physical positions of the SNPs against their genetic positions in the linkage map demonstrate a high level of collinearity between our individuals and the individual used for the generation of the reference genome, they also feature two important differences. A split in chromosome 2 was identified in all families and both sexes, thus assigning two parts of chromosome 2 of the reference genome to two different linkage groups. Additionally, chromosomes 22 and 23 were assigned to the same linkage group. These differences between linkage maps and the physical map of the reference genome highlight the possibility of using genetic mapping for improving the quality of reference genome assemblies in future studies. The new high-density genetic linkage maps can be applied in quantitative trait-loci studies, facilitating the identification of genetic components underpinning key biological traits to potentially support sustainable forest management.

Keywords: Common ash (*Fraxinus excelsior*), linkage mapping, genetic maps, Lep-MAP3

Introduction

Genetic maps facilitate the study of genome organisation, evolution, and the inheritance of traits. In trees, genetic mapping has significant implications for breeding, conservation, and understanding responses to environmental stresses (Sun et al., 2015; Tong et al., 2016; Di Wu et al., 2019). Genetic maps are built using genotypes of many related individuals at selected markers to determine the relative genetic locations of these markers. Genotype data allow us to infer where recombination has occurred, which is directly related to the genetic distance and the order of markers on the chromosome (Cartwright et al., 2007). It provides valuable genomic information and generates a better understanding of the genome and chromosome structure. Further, high-density genetic maps can be used for quantitative trait locus (QTL) analyses. The results may be useful for marker-assisted selection (Liu et al., 2015; Tong et al., 2016) and the identification of the biological mechanisms underlying specific traits, such as disease resistance (Herrero et al., 2020).

Common ash, *Fraxinus excelsior* L. (Oleaceae), is a temperate tree species distributed throughout Europe. It has excellent wood properties and high economic and ecological value (Meger et al., 2024a). The common ash behaves as a subdioecious system. This means there are two distinct reproductive groups: (i) hermaphrodites and females and (ii) males and male-biased hermaphrodites. Hermaphrodites and females are part of one self-incompatibility (SI) group and functionally act as females (because they only produce seeds but no viable pollen). Males and male-biased hermaphrodites on the other hand are part of another SI group and functionally act as males (because they produce only pollen but no fertile seeds) (Saumitou-Laprade et al., 2018).

With anthropogenic climate change and pests and pathogens spreading, pressure on forest ecosystem services is rising (May et al., 2024; Xu and Prescott, 2024). *F. excelsior*, for example, is severely threatened by the invasive pathogenic fungus *Hymenoscyphus fraxineus*, which causes ash dieback (Nielsen et al., 2017; Coker et al., 2019). The pathogen is rapidly spreading throughout Europe, with severe ecological and economic consequences.

Despite the importance of *F. excelsior*, there is limited knowledge regarding the genomic structure and recombination patterns. The recently published chromosome-scale reference genome (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_019097785.1/) includes 23 chromosomes and 415 scaffolds, with a total assembly size of 806.5 Mb. Furthermore, it comprises 58.90 % repetitive DNA and 41,355 high-confidence gene models (Meger et al., 2024b). This reference genome is a foundation for understanding the genomic structure of important traits in *F. excelsior*.

This study aims to construct high-density genetic linkage maps for four full-sibling families of *F. excelsior* using single nucleotide polymorphism (SNP) markers to compare them with the reference genome assembly. The four full-sibling families from Mecklenburg-Western Pomerania in the northeast of Germany, where the mother trees are potentially tolerant to ash dieback, have been identified in a previous study (Krautwurst et al., 2023)(Chapter 1). By comparing maternal and paternal linkage maps, we seek to uncover sex-specific recombination rates and identify potential chromosomal anomalies and large structural variations. These findings may contribute to future genetic research and practical applications in breeding and conservation. Understanding the genomic structure of diseases will help to preserve species in the face of ongoing ecological threats.

Material and Methods

Family identification, DNA extraction and Illumina low-coverage resequencing

Four potentially less susceptible mother trees across Mecklenburg-Western Pomerania in the northeast of Germany were chosen for the full-sibling (full-sib) family identification (Table 1). In 2018, around 3000 seeds per tree were collected from the mother trees. The family identification was conducted with eight SSR markers and the program Colony (Jones and Wang, 2010b) to predict full-sib families. The predicted families from Colony were resequenced by Novogene (UK) Ltd. (Cambridge, UK) on the Novaseq 6000 platform. Further details can be found in the study of Krautwurst et al. 2023 (Chapter 1). The sequencing included only the full-sibs and the mother trees. The father trees are unknown and not included.

Table 1: Summary of all full-sibling families identified by SSR and SNP marker and included in resequencing (Krautwurst et al. 2024).

Name	Size
Dar18	115
Eve2	134
Kar4	162
Fri8	148

Mapping and variant calling

The sequencing data of every family and all included individuals were mapped against the common ash reference genome (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_019097785.1/) using bwa-mem (version bwa-0.7.17.tar.bz (Li and Durbin, 2009)). Grouping of the reads and duplicates was marked using Picard tools (version (v) 2.26.2) (<http://broadinstitute.github.io/picard/>). Joint variant calling was performed with GATK v4.4.0.0, following the best practices for germline short variant discovery wherever possible (Poplin et al., 2017). The ‘HaplotypeCaller’ from GATK was used to generate variant files in genomic Variant Call Formate (gVCF). After combining the gVCFs with GATK’s ‘GenomicsDBImport’, the ‘GenotypeGVFs’ tool was used for the joint genotyping.

Variant filtering

Our selection process was rigorous and thorough. For hard filtering, we mostly followed the documentation on ‘Hard-filtering germline short variants’ on the GATK website. We filtered indels and SNPs separately and only proceeded with SNP data. We removed variants based on strand bias (FisherStrand ‘FS’ > 60 & StrandOddsRatio ‘SOR’ > 3) and mapping quality (RMSMappingQuality ‘MQ’ < 40, MappingQualityRankSumTest ‘MQRankSum’ < -1). Based on the distribution of the variant confidence score QualByDepth ‘QD’, we chose a more stringent cutoff of QD > 10 to remove any low-confidence variants. Filtering was performed with bcftools v1.17 (Li, 2011). We then extracted the variant sequencing depth values ‘DP’ and minor allele frequency ‘frq2’ using vcftools v0.1.15 (Danecek et al., 2011). To visualise the DP and choose the parameters, we used R (R Core Team, 2022) minimal-mean DP was 4, and max-mean DP was 12. Non-biallelic SNPs were

excluded and data from SNPs with more than 9 % missing genotype data were removed. Since GATK codes missing data as 0/0 in the section ‘DP’ in v4.4.0.0, we used bcftools ‘FORMAT/DP > 4’ for further elimination of missing data. Further, the minor allele frequency was filtered at 0.2 with vcftools, since we were only interested in variants with Mendelian segregation pattern (1:2:1 or 1:1). Before the resulting Variant Call Format (vcf) files could be merged, an intersect was calculated using the ‘isec’ function of bcftools (Danecek et al., 2021) to identify common SNPs in the vcf file. Then, the ‘merge’ function was used to create one multi-sample file (Table 2).

Table 2.: Variant counts (SNP counts) of the four full-sibling families after hard filtering.

Name	Variant count
Dar18	565,839
Eve2	691,999
Kar4	924,351
Fri8	979,820

Segregation filtration

Segregation filtering was conducted following the corresponding Mendelian segregation ratios in the R environment using ‘VCFR’ (Knaus and Grünwald, 2017). The paternal vcf was generated using the maternal and full-sib genotype information. The resulting maternal and paternal vcf files were further filtered using the R base for unique and duplicated variants. The ‘unique’ function ensured that only unique variant positions and their corresponding chromosomes were retained in the dataset.

Linkage map construction

We conducted linkage mapping with the Lep-MAP3 software package (Rastas, 2017). Before starting the mapping, we used GATK ‘CalculateGenotypePosteriors’ to calculate the posterior genotype probability for each sample genotype. Next we used the model ‘Parentcall2’ to call parental genotypes. Further, we used ‘Filtering2’ to remove non-informative markers and distorted markers, with a dataTolerance = 0.001. The next step is ‘SeparateChromosome2’, categorising markers into linkage groups (LGs). A LOD (Logarithm of Odds) score (scores likelihood that two genetic markers are in the same linkage group) limit of three was chosen. The assignment to the linkage groups was conducted by computing all pair-wise LOD scores

between markers and joining markers with LOD scores higher than the given lod limit. In the last step, markers were separated into their corresponding LGs and ordered using the function ‘OrderMarkers2’ (Table 3). LG numbers and numbers of markers for each family can be found in the Supplementary Tables 1-4.

Table 3.: Overview of the four family outcomes of linkage mapping using Lep-MAP3. A total number of linkage groups and mapped markers before the cut-off.

Family	Sex	No. of mapped markers	No. of linkage groups
Dar18	Female	7,470	31
Dar18	Male	8,291	45
Eve2	Female	35,129	28
Eve2	Male	7,184	36
Kar4	Female	53,022	27
Kar4	Male	24,686	31
Fri8	Female	45,426	29
Fri8	Male	13,378	45

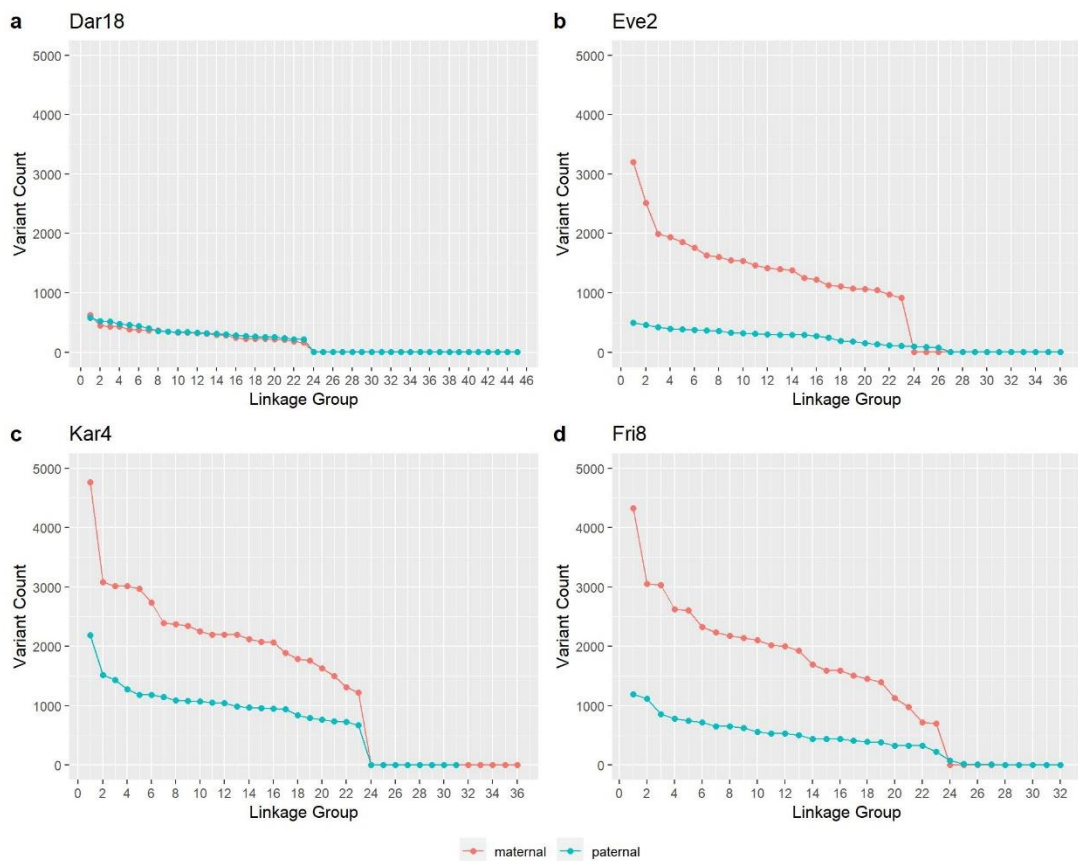


Figure 8: Graphs showing the linkage groups generated by Lep-MAP3 and the count of variants that have been sorted to the Linkage Group. The red line represents the maternal map, and the green represents the paternal map.

The plots of the total number of markers ordered to the LG were used to choose a cut-off for defining the total number of reliable LGs per genetic map (Table 4), since some markers will fail to be integrated into larger LGs and will thus form additional

tiny LGs not corresponding to any chromosome (Figure 1, Table 3). The maximum length of each chromosome was extracted to examine recombination (in this case, provided by the LG length in centimorgan) and differences between males and females (Supplementary Tables S5-S8).

Table 4.: Number of sensible linkage groups for the female and male maps of each family (cut-off defined based on Figure 1).

Family	Sex	Linkage groups
Dar18	Female	23
Dar18	Male	23
Eve2	Female	23
Eve2	Male	26
Kar4	Female	23
Kar4	Male	23
Fri8	Female	27
Fri8	Male	23

Comparison of genetic maps with the physical map

To compare genetic and physical marker positions, the R packages ‘ggplot2’ (Wickham, 2016) and ‘VCFR’ (Knaus and Grünwald, 2017) were used.

Results

In this study, we constructed high-density genetic linkage maps for four ash tree full-sib families: Dar18, Eve2, Kar4, and Fri8. Using the program Lep-MAP3, we identified and categorised multiple linkage groups for each family (Table 4). This allowed us to compare genetic and physical marker positions. The plots of genetic positions of the SNP markers in the linkage maps against their physical positions (Figure 2) demonstrate a high level of collinearity between our genetic maps and the reference genome of *F. excelsior* (Supplementary Figures S1-S8).

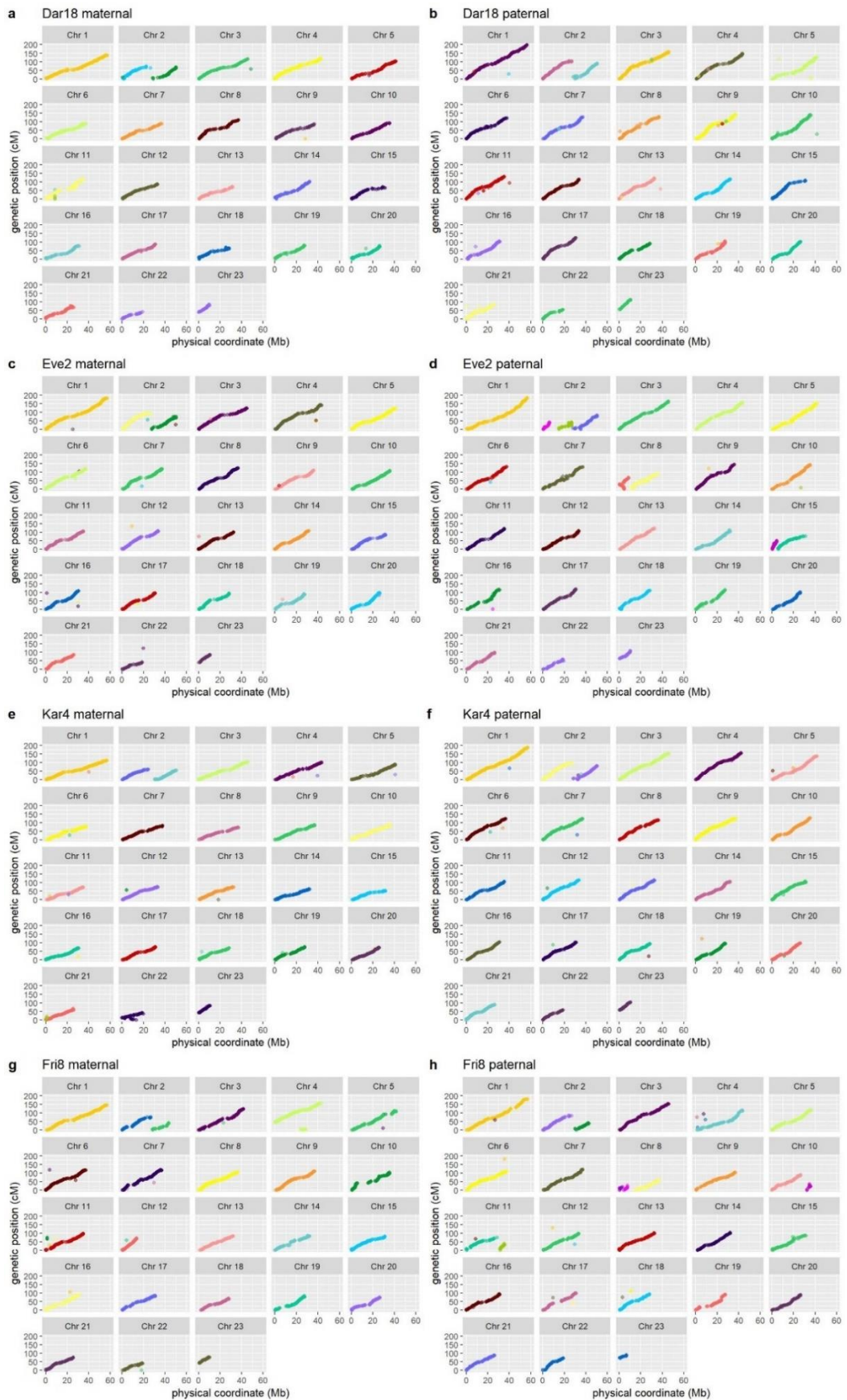


Figure 2: Visualization of accumulated genetic distance (cM) of our genetic maps on the y-axis and physical distance (Mb) of the ash reference genome on the x-axis. For each family the maternal and the paternal maps are shown. The colours of each marker represent the assignment to a linkage group. The number of linkage groups in each genetic map is given in Table 4.

Two discrepant features could be detected in all four families in the paternal and maternal maps. First, a split in chromosome (Chr) 2 was identified, thus assigning two parts of this chromosome to two different linkage groups (indicated by two different colours in Figure 2). This suggests that what is physically one chromosome in the reference genome assembly (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_019097785.1/) behaves as two separate linkage groups indicating that they represent two chromosomes. One exception is seen in the paternal map of the family Eve2 (Figure 2 d), where chromosome 2 is split into three different linkage groups. The second general discrepancy can be found for chromosomes 22 and 23. The linkage mapping results do not distinguish the two chromosomes and assign them to one linkage group (Figure 2), suggesting that they actually represent a single chromosome.

Some chromosomes in some families show evidence of structural variants compared to the reference genome, which may result from insertions causing the according chromosomes to split into several LGs, e.g. Eve2 paternal Chr 2, Chr 8 and Chr 15. But there is a possibility, especially of the size and arrangement, that these are only artifacts. The Eve2 paternal map for Chr 8 also shows evidence of an inversion. A variation on Chr 8 can also be found in the paternal map of Fri8. Further, larger horizontal gaps can be observed in Dar18 maternal Chr 11, Fri8 maternal Chr 10 and paternal Chr 17, which might indicate deletions. Small duplication can be observed in Fri8 maternal Chr 4 and Kar4 maternal Chr 22. However, some outliers can be detected in all families.

The comparison between females and males was similar in all families (Figure 3). However, the male linkage groups had higher recombination rates on average (example Dar18, Figure 4). The total genetic length (cM) between females and males also differs; the total length of the linkage groups in all male maps is higher than in females (Supplementary Tables S5-S8).

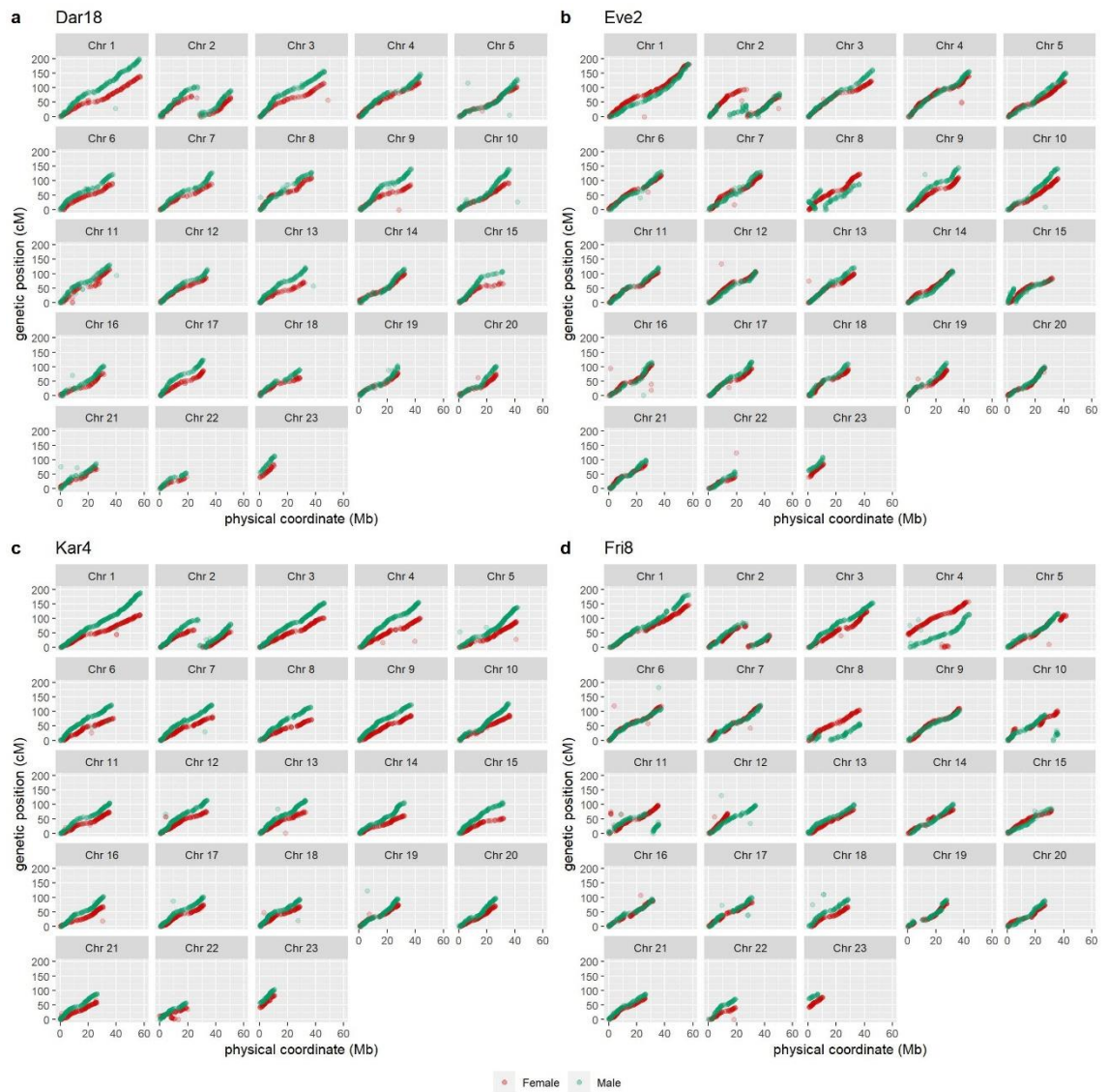


Figure 3: Comparison of female and male correlation of accumulated genetic distance (cM) to physical distance (Mb) in all four families. The female is coloured in red, and the male is coloured in green. The physical positions of the single nucleotide polymorphism markers in *Fraxinus excelsior* were plotted against their genetic positions in the linkage map.

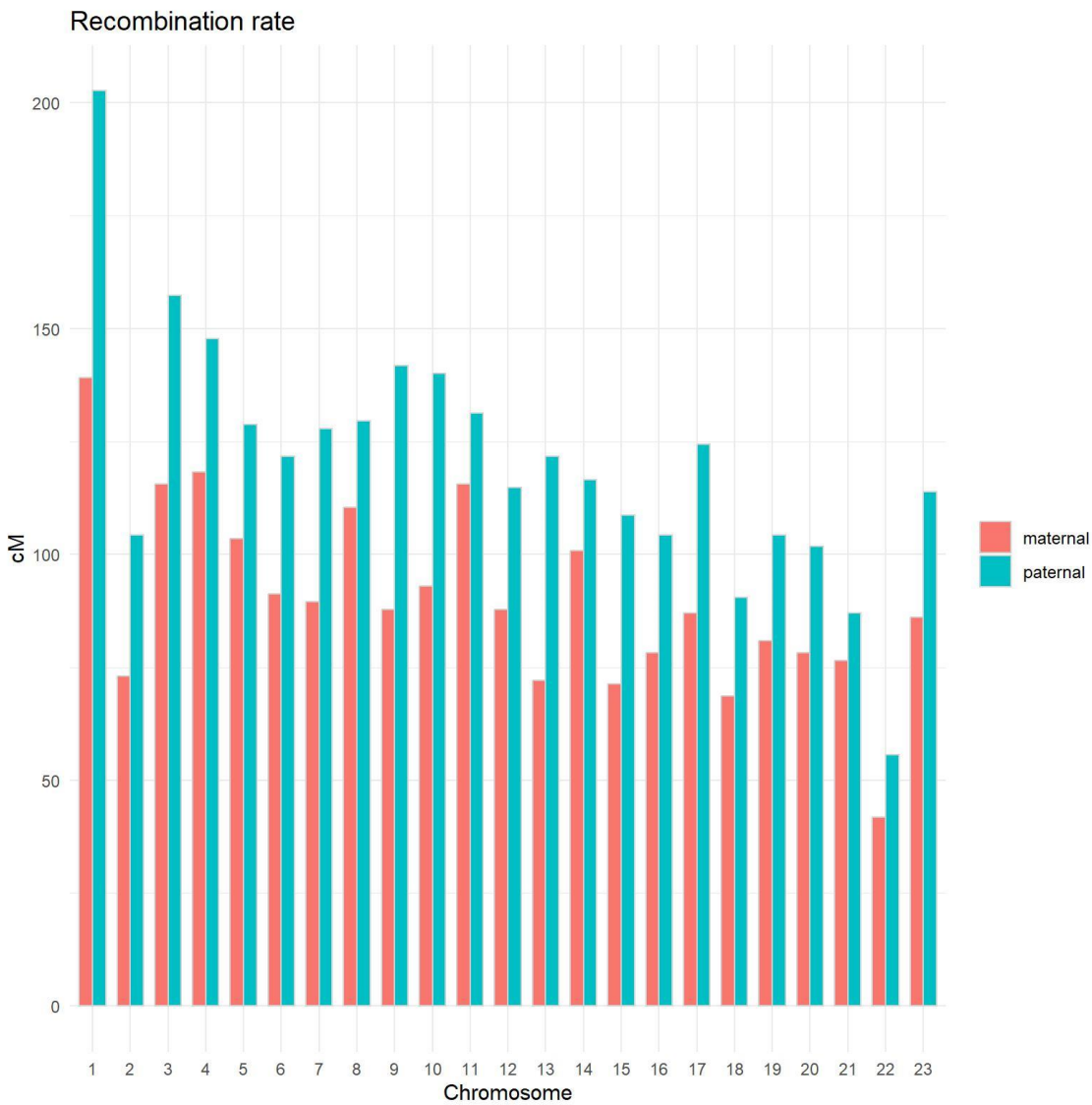


Figure 4: This figure illustrates the maximal recombination rate, measured in centimorgan (cM), for all chromosomes separated by sex (maternal and paternal) for Dar18.

Discussion

The main findings emphasise the high quality of the genetic maps, with 23 linkage groups corresponding to the 23 ash chromosomes. Also, the marker positions are aligned with the genome sequence represented by the chromosomes of a chromosome-level reference genome (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_019097785.1/). Interestingly, in all four families, chromosome two was split, appearing as two separate linkage groups, and chromosomes 22 and 23 should be one chromosome since they belong to the same linkage group. The genetic maps will be a valuable resource for QTL analyses.

The occurring structural variants seem to be family -and sex-specific (Figure 2). Most of them appear, except Dar18, on the paternal map. One reason could be the smaller number of markers used for the paternal maps than the maternal ones (Table 2). Fewer genetic markers can lead to a reduction in capturing the recombination events and can cause less accurate mapping, which means that it could be that the structural variants can be artefacts (Ott et al., 2011). The smaller number of genetic markers in the paternal file could result from different sequencing depths between the mother and full sibs. Since the mothers were sequenced with higher coverage than the siblings and for the unknown father, segregation was used to create the paternal file; this could have resulted in different marker numbers. For the validation of the structural variants, especially smaller ones, long-read sequencing emerges as a promising possibility, opening up new avenues for future research and applications (Pinosio et al., 2016; Garcia et al., 2017; Amarasinghe et al., 2020; Lian et al., 2024). An important pre-requisite for long-read sequencing in *F. excelsior* is the preparation of high-molecular-weight DNA, which can be done as presented in Chapter 2 (Krautwurst et al., 2024).

Studies show that the differences in the paternal and maternal map (Figure 4) could be due to differences in the degree of recombination. The degree and direction of recombination rates and landscapes underlying genetic architecture often differ between sexes (Brekke et al., 2023). Sexual dimorphism in the recombination landscape can often be observed biased towards the tips of chromosomes in male meiosis. In contrast, female recombination rates are likelier to be elevated towards the centromeres or more uniform across the chromosome. Fine-scale recombination patterns often vary between males and females (Rifkin et al., 2022). The molecular mechanisms underlying these differences are not fully understood. Some hypotheses propose that the chromatin landscape could be a factor, with sexually antagonistic selection on coding genes and their regulatory elements also playing a role. Additionally, meiotic drive in females, selection during the haploid phase of the life cycle, and inherent mechanistic constraints may all contribute to these sex-specific differences (Sardell and Kirkpatrick, 2020; Rifkin et al., 2022).

Errors in the genetic map (e.g., Dar18 Chr 9 paternal, three variants belong to different LGs even when it should be the same as the main LG; Figure 2) can be caused by genotyping errors (Akbarpour et al., 2021), which may be due to homozygous genotypes mistakenly classified as heterozygous, missing values, low quantity or quality of DNA, or operator errors (Bonin et al., 2004; Whitlock et al., 2008). These errors can lead to incorrect detection of inversions among individuals, which can lead to differences between linkage order and physical order (Abecasis et al., 2001; Cartwright et al., 2007; Gomez-Raya et al., 2022). A solution could be to perform independent genotyping, at least for some individuals (Wang et al., 2024). However, errors (e.g., Fri8 paternal Chr 8, 10 and 11, small sections with different LG groupings; Figure 2) could also arise during linkage group assignment in the step ‘SeparateChromosomes2’ using Lep-MAP3. Variants may be mapped to the wrong LG, and linked variants are assigned to this LG in the next step. However, the horizontal gaps located in Dar18 maternal Chr 11, Fri8 maternal Chr 10 and paternal Chr 17 may indicate a low SNP coverage in the particular genomic region. Which could indicate insertions, but as discussed above it is a possibility that these are only artefacts (Herrero et al., 2020).

Conclusion

The present research declares high-quality genetic maps of four full-sib families of *F. excelsior*. All maps contained 23 LGs, consistent with the number of chromosomes in *F. excelsior* ($2n = 46$). Comparative analysis with the reference genome revealed two differences: the unity of chromosome two, which should be two chromosomes, and the split of chromosomes 22 and 23, which should be one chromosome. These findings may support future improvements of the reference genome. The comparison of female and male genetic maps showed a bias in length, with the male maps being longer than the female maps. The high-quality genetic maps generated in this study are valuable resources for QTL analyses, which might contribute to conserving ash as an important forest tree species in Europe.

References

- Abecasis, G. R., Cherny, S. S., and Cardon, L. R. (2001). The impact of genotyping error on family-based analysis of quantitative traits. *Eur J Hum Genet* 9, 130–134. doi: 10.1038/sj.ejhg.5200594
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., and Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 21, 30. doi: 10.1186/s13059-020-1935-5
- Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., and Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Mol Ecol* 13, 3261–3273. doi: 10.1111/j.1365-294X.2004.02346.x
- Brekke, C., Johnston, S. E., Knutsen, T. M., and Berg, P. (2023). Genetic architecture of individual meiotic crossover rate and distribution in a large Atlantic Salmon (*Salmo salar*) breeding population. *bioRxiv*.
- Cartwright, D. A., Troggio, M., Velasco, R., and Gutin, A. (2007). Genetic mapping in the presence of genotyping errors. *Genetics* 176, 2521–2527. doi: 10.1534/genetics.106.063982
- Coker, T. L. R., Rozsypálek, J., Edwards, A., Harwood, T. P., Butfoy, L., and Buggs, R. J. A. (2019). Estimating mortality rates of European ash (*Fraxinus excelsior*) under the ash dieback (*Hymenoscyphus fraxineus*) epidemic. *Plants, People, Planet* 1, 48–58. doi: 10.1002/ppp3.11
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10. doi: 10.1093/gigascience/giab008
- Di Wu, Koch, J., Coggeshall, M., and Carlson, J. (2019). The first genetic linkage map for *Fraxinus pennsylvanica* and syntenic relationships with four related species. *Plant Mol Biol* 99, 251–264. doi: 10.1007/s11103-018-0815-9
- Garcia, S., Williams, S., Xu, A. W., Herschleb, J., Marks, P., Stafford, D., et al. (2017). Linked-Read sequencing resolves complex structural variants. *bioRxiv*.
- Gomez-Raya, L., Gómez Izquierdo, E., La Mercado Peña, E. de, Garcia-Ruiz, F., and Rauw, W. M. (2022). First-degree relationships and genotyping errors deciphered by a high-density SNP array in a Duroc × Iberian pig cross. *BMC Genom Data* 23, 14. doi: 10.1186/s12863-022-01025-1
- Herrero, J., Santika, B., Herrán, A., Erika, P., Sarimana, U., Wendra, F., et al. (2020). Construction of a high density linkage map in Oil Palm using SPET markers. *Sci Rep* 10, 9998. doi: 10.1038/s41598-020-67118-y
- Jones, O. R., and Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol Ecol Resour* 10, 551–555. doi: 10.1111/j.1755-0998.2009.02787.x
- Knaus, B. J., and Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* 17, 44–53. doi: 10.1111/1755-0998.12549
- Krautwurst, M., Past, F., Kersten, B., Bubner, B., and Müller, N. A. (2023). Identification of full-sibling families from natural single-tree ash progenies based on SSR markers and genome-wide SNPs. doi: 10.1101/2023.07.18.549475

- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Lian, Q., Huettel, B., Walkemeier, B., Mayjonade, B., Lopez-Roques, C., Gil, L., et al. (2024). A pan-genome of 69 *Arabidopsis thaliana* accessions reveals a conserved genome structure throughout the global species range. *Nat Genet* 56, 982–991. doi: 10.1038/s41588-024-01715-9
- Liu, H., Niu, Y., Gonzalez-Portilla, P. J., Zhou, H., Wang, L., Zuo, T., et al. (2015). An ultra-high-density map as a community resource for discerning the genetic basis of quantitative traits in maize. *BMC Genomics* 16, 1078. doi: 10.1186/s12864-015-2242-5
- May, M., Hirsch, S., and Abramson, M. (2024). Transformation of Plantation Forestry Productivity for Climate Change Mitigation and Adaptation. *Cold Spring Harb Perspect Biol* 16. doi: 10.1101/cshperspect.a041670
- Meger, J., Kozioł, C., Pałucka, M., Burczyk, J., and Chybicki, I. J. (2024a). Genetic resources of common ash (*Fraxinus excelsior* L.) in Poland. *BMC Plant Biol* 24, 186. doi: 10.1186/s12870-024-04886-z
- Meger, J., Ulaszewski, B., Pałucka, M., Kozioł, C., and Burczyk, J. (2024b). Genomic prediction of resistance to *Hymenoscyphus fraxineus* in common ash (*Fraxinus excelsior* L.) populations. *Evol Appl* 17, e13694. doi: 10.1111/eva.13694
- Nielsen, L. R., McKinney, L. V., and Kjær, E. D. (2017). Host phenological stage potentially affects dieback severity after *Hymenoscyphus fraxineus* infection in *Fraxinus excelsior* seedlings. *Baltic Forestry* 23, 229–232.
- Ott, J., Kamatani, Y., and Lathrop, M. (2011). Family-based designs for genome-wide association studies. *Nat Rev Genet* 12, 465–474. doi: 10.1038/nrg2989
- Pinosio, S., Giacomello, S., Faivre-Rampant, P., Taylor, G., Jorge, V., Le Paslier, M. C., et al. (2016). Characterization of the Poplar Pan-Genome by Genome-Wide Identification of Structural Variation. *Mol Biol Evol* 33, 2706–2719. doi: 10.1093/molbev/msw161
- Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., van der Auwera, G. A., et al. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. doi: 10.1101/201178
- R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for*. Vienna, Austria.
- Rastas, P. (2017). Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* 33, 3726–3732. doi: 10.1093/bioinformatics/btx494
- Rifkin, J. L., Hnatovska, S., Yuan, M., Sacchi, B. M., Choudhury, B. I., Gong, Y., et al. (2022). Recombination landscape dimorphism and sex chromosome evolution in the dioecious plant *Rumex hastatulus*. *Philos Trans R Soc Lond B Biol Sci* 377, 20210226. doi: 10.1098/rstb.2021.0226
- Sardell, J. M., and Kirkpatrick, M. (2020). Sex Differences in the Recombination Landscape. *Am Nat* 195, 361–379. doi: 10.1086/704943
- Saumitou-Laprade, P., Vernet, P., Dowkiw, A., Bertrand, S., Billiard, S., Albert, B., et al. (2018). Polygamy or subdioecy? The impact of diallelic self-incompatibility on the sexual system in *Fraxinus excelsior* (Oleaceae). *Proc Biol Sci* 285. doi: 10.1098/rspb.2018.0004

- Sun, L., Zhu, X., Zhang, Q., and Wu, R. (2015). A unifying experimental design for dissecting tree genomes. *Trends in Plant Science* 20, 473–476. doi: 10.1016/j.tplants.2015.05.012
- Tong, C., Li, H., Wang, Y., Li, X., Ou, J., Wang, D., et al. (2016). Construction of High-Density Linkage Maps of *Populus deltoides* × *P. simonii* Using Restriction-Site Associated DNA Sequencing. *PLOS ONE* 11, e0150692. doi: 10.1371/journal.pone.0150692
- Wang, X., Wang, J., Xia, X., Xu, X., Li, L., Cao, S., et al. (2024). Effect of genotyping errors on linkage map construction based on repeated chip analysis of two recombinant inbred line populations in wheat (*Triticum aestivum* L.). *BMC Plant Biol* 24, 306. doi: 10.1186/s12870-024-05005-8
- Whitlock, R., Hipperson, H., Mannarelli, M., Butlin, R. K., and Burke, T. (2008). An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Mol Ecol Resour* 8, 725–735. doi: 10.1111/j.1755-0998.2007.02073.x
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer International Publishing. ISBN: 9783319242774.
- Xu, W., and Prescott, C. E. (2024). Can assisted migration mitigate climate-change impacts on forests? *Forest Ecology and Management* 556, 121738. doi: 10.1016/j.foreco.2024.121738

Supplementary

Supplementary Table S1.: Linkage groups and variant count of Eve2 left paternal and right maternal. Linkage group contains the variants that could not be sorted to any linkage group.

Variant count	Linkage group	Variant count	Linkage group
494	1	3200	1
460	2	2520	2
422	3	1993	3
400	4	1940	4
391	5	1857	5
380	6	1762	6
365	7	1632	7
357	8	1601	8
334	9	1547	9
326	10	1537	10
317	11	1462	11
307	12	1421	12
297	13	1401	13
295	14	1385	14
293	15	1251	15
274	16	1227	16
248	17	1130	17
188	18	1108	18
180	19	1074	19
154	20	1063	20
137	21	1051	21

113	22	971	22
109	23	915	23
102	24	71	0
87	25	24	24
76	26	2	25
57	0	2	26
3	27	2	27
2	28	2	28
2	29		
2	30		
2	31		
2	32		
2	33		
2	34		
2	35		
2	36		

Supplementary Table S2.: Linkage groups and variant count of Dar18 left paternal and right maternal. Linkage group contains the variants that could not be sorted to any linkage group.

Variant count	Linkage group	Variant count	Linkage group
586	1	632	1
524	2	451	2
515	3	431	3
484	4	430	4
461	5	391	5
442	6	378	6
406	7	370	7
360	8	364	8
350	9	350	9
344	10	332	10
340	11	328	11
332	12	326	12
319	13	309	13
317	14	292	14
301	15	289	15
282	16	248	16
276	17	232	17
266	18	228	18
258	19	228	19
258	20	222	20
235	21	212	21
219	22	185	22
219	23	162	23
152	0	61	0
3	24	4	24

2	25	3	25
2	26	2	26
2	27	2	27
2	28	2	28
2	29	2	29
2	30	2	30
2	31	2	31
2	32		
2	33		
2	34		
2	35		
2	36		
2	37		
2	38		
2	39		
2	40		
2	41		
2	42		
2	43		
2	44		
2	45		

Supplementary Table S3.: Linkage groups and variant count of Kar4 left paternal and right maternal. Linkage group contains the variants that could not be sorted to any linkage group.

Variant count	Linkage group	Variant count	Linkage group
2192	1	4767	1
1515	2	3085	2
1437	3	3020	3
1276	4	3015	4
1187	5	2965	5
1180	6	2738	6
1147	7	2397	7
1091	8	2374	8
1081	9	2348	9
1070	10	2257	10
1050	11	2201	11
1045	12	2195	12
989	13	2194	13
968	14	2127	14
959	15	2078	15
948	16	2070	16
942	17	1888	17
839	18	1790	18
793	19	1761	19

763	20	1628	20
740	21	1505	21
733	22	1314	22
671	23	1222	23
54	0	53	0
2	24	4	24
2	25	3	26
2	26	3	25
2	27	2	36
2	28	2	35
2	29	2	34
2	30	2	33
2	31	2	32
		2	31
		2	30
		2	29
		2	28
		2	27

Supplementary Table S4.: Linkage groups and variant count of Fri8 left paternal and right maternal. Linkage group contains the variants that could not be sorted to any linkage group.

Variant count	Linkage group	Variant count	Linkage group
1195	1	4330	1
1115	2	3057	2
860	3	3030	3
782	4	2628	4
743	5	2605	5
718	6	2331	6
658	7	2236	7
655	8	2183	8
631	9	2142	9
565	10	2104	10
535	11	2025	11
534	12	2007	12
503	13	1926	13
445	14	1694	14
442	15	1593	15
437	16	1591	16
411	17	1514	17
395	18	1450	18
383	19	1400	19
333	20	1130	20
325	21	979	21
325	22	723	22
226	23	705	23

81	24	31	0
22	0	2	24
20	25	2	25
14	26	2	26
11	27	2	27
6	28	2	28
2	29	2	29
2	30		
2	31		
2	32		
2	44		
2	45		

Supplementary Table S5.: Recombination rate of Eve2 for each chromosome and sex. The max value represents the maximal length of the chromosomes in centimorgan (cM).

Chromosome	Sex	Max (cM)
1	Female	181.343
	Male	182.836
2	Female	94.03
	Male	80.597
3	Female	123.134
	Male	163.433
4	Female	142.537
	Male	156.716
5	Female	121.642
	Male	152.239
6	Female	120.149
	Male	131.343
7	Female	119.403
	Male	129.851
8	Female	123.881
	Male	89.552
9	Female	111.194
	Male	145.522
10	Female	108.955
	Male	142.537
11	Female	105.97
	Male	121.642
12	Female	135.075
	Male	108.955
13	Female	100.746
	Male	122.388
14	Female	108.209
	Male	111.94
15	Female	87.313
	Male	78.358
16	Female	108.209
	Male	115.672
17	Female	94.776
	Male	117.91
18	Female	94.03
	Male	110.448
19	Female	91.791

	Male	113.433
20	Female	97.761
	Male	100.746
21	Female	86.567
	Male	99.254
22	Female	123.881
	Male	59.701
23	Female	86.567
	Male	109.701

Supplementary Table S6.: Recombination rate of Kar4 for each chromosome and sex. The max value represents the maximal length of the chromosomes in centimorgan (cM).

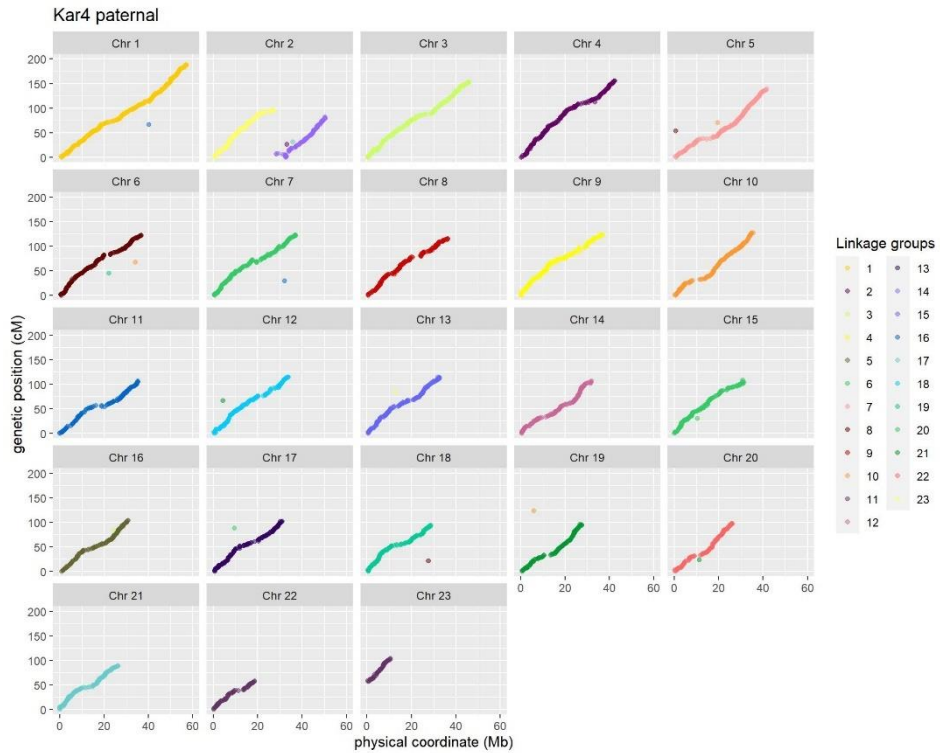
Chromosome	Sex	Max (cM)
1	Female	112.963
	Male	188.272
2	Female	60.494
	Male	96.296
3	Female	102.469
	Male	153.704
4	Female	100.617
	Male	155.556
5	Female	88.889
	Male	138.272
6	Female	76.543
	Male	122.222
7	Female	82.716
	Male	122.84
8	Female	72.222
	Male	114.815
9	Female	86.42
	Male	123.457
10	Female	88.272
	Male	127.16
11	Female	74.691
	Male	106.173
12	Female	75.926
	Male	114.815
13	Female	75.926
	Male	114.815
14	Female	62.963
	Male	106.79
15	Female	53.086
	Male	108.642
16	Female	68.519
	Male	103.704
17	Female	75.309
	Male	101.852
18	Female	69.753
	Male	93.827
19	Female	75.926
	Male	123.457
20	Female	72.222
	Male	97.531
21	Female	66.049
	Male	88.889
22	Female	41.975
	Male	57.407
23	Female	83.333
	Male	103.704

Supplementary Table S7.: Recombination rate of Fri8 for each chromosome and sex. The max value represents the maximal length of the chromosomes in centimorgan (cM).

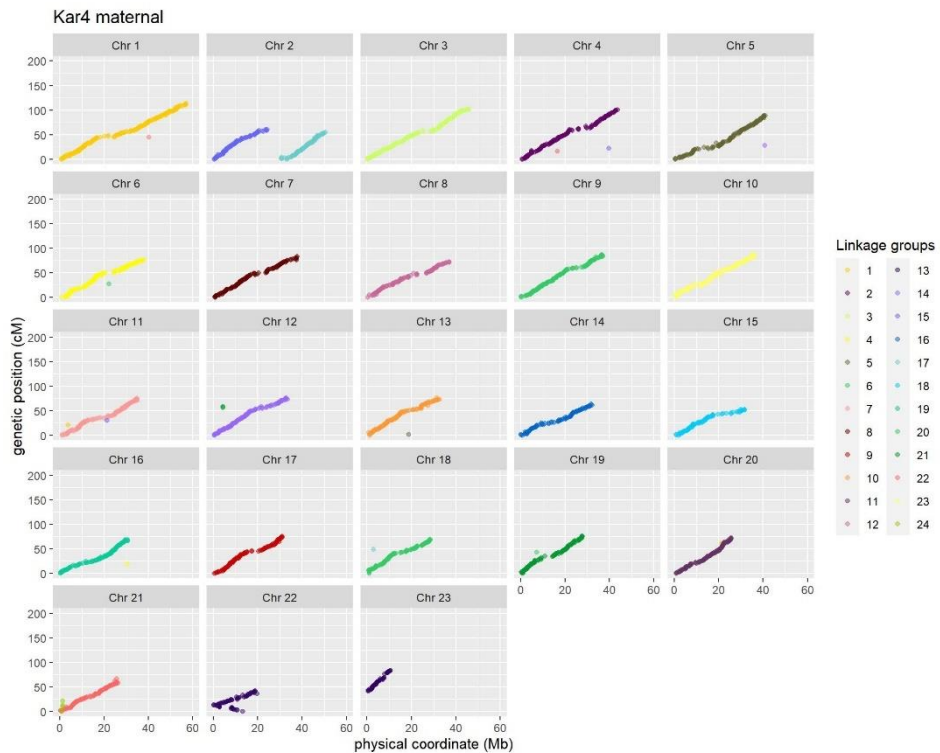
Chromosome	Sex	Max (cM)
1	Female	145.946
	Male	181.081
2	Female	75.676
	Male	85.135
3	Female	124.324
	Male	154.054
4	Female	157.432
	Male	114.865
5	Female	112.162
	Male	117.568
6	Female	119.595
	Male	181.757
7	Female	118.919
	Male	121.622
8	Female	105.405
	Male	58.784
9	Female	110.811
	Male	103.378
10	Female	103.378
	Male	88.514
11	Female	97.973
	Male	73.649
12	Female	69.595
	Male	130.405
13	Female	83.108
	Male	100.676
14	Female	84.459
	Male	102.703
15	Female	81.081
	Male	87.162
16	Female	107.432
	Male	95.946
17	Female	85.811
	Male	101.351
18	Female	68.919
	Male	110.811
19	Female	83.108
	Male	91.892
20	Female	75.676
	Male	89.865
21	Female	75
	Male	87.838
22	Female	41.892
	Male	70.946
23	Female	77.703
	Male	88.514

Table S8.: Recombination rate of Dar18 for each chromosome and sex. The max value represents the maximal length of the chromosomes in centimorgan (cM).

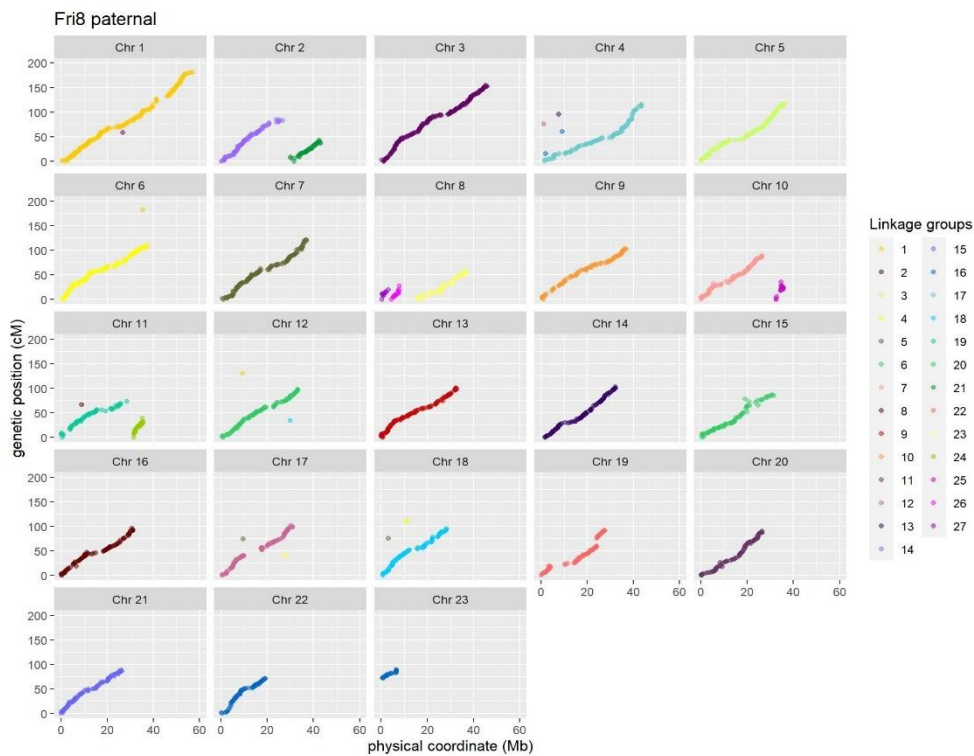
Chromosome	Sex	Max (cM)
1	Female	139.13
	Male	202.609
2	Female	73.043
	Male	104.348
3	Female	115.652
	Male	157.391
4	Female	118.261
	Male	147.826
5	Female	103.478
	Male	128.696
6	Female	91.304
	Male	121.739
7	Female	89.565
	Male	127.826
8	Female	110.435
	Male	129.565
9	Female	87.826
	Male	141.739
10	Female	93.043
	Male	140
11	Female	115.652
	Male	131.304
12	Female	87.826
	Male	114.783
13	Female	72.174
	Male	121.739
14	Female	100.87
	Male	116.522
15	Female	71.304
	Male	108.696
16	Female	78.261
	Male	104.348
17	Female	86.957
	Male	124.348
18	Female	68.696
	Male	90.435
19	Female	80.87
	Male	104.348
20	Female	78.261
	Male	101.739
21	Female	76.522
	Male	86.957
22	Female	41.739
	Male	55.652
23	Female	86.087
	Male	113.913



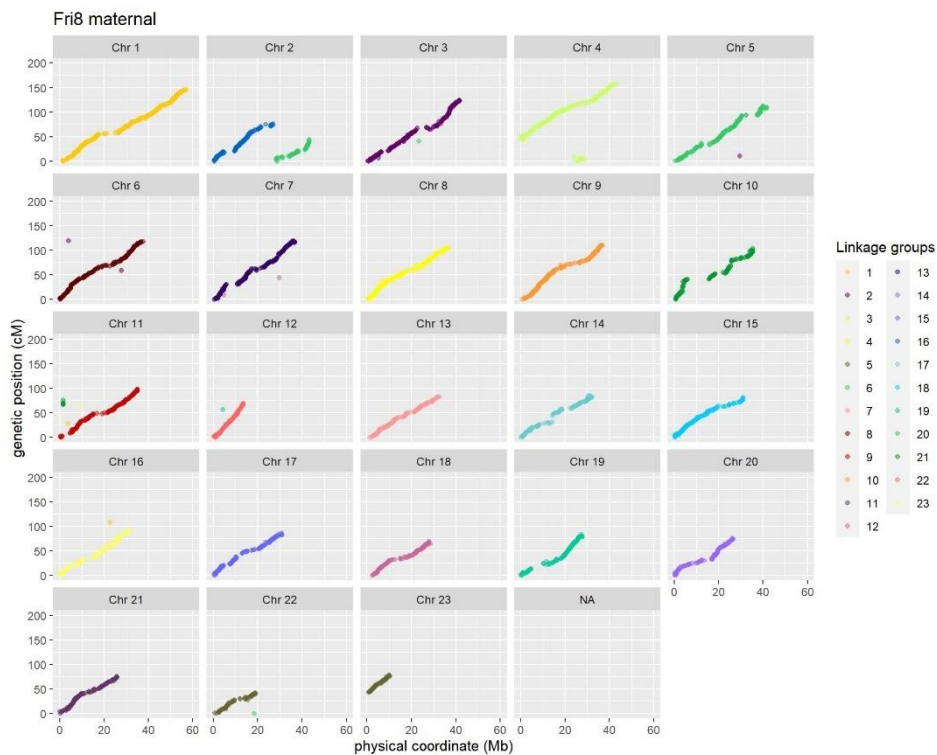
Supplementary Figure S1: Correlation of accumulated genetic distance (cM) and physical distance (Mb). The physical positions of the single nucleotide polymorphism (SNP) markers in *Fraxinus excelsior* were plotted against their genetic positions in the linkage map. The colours of each chromosome represent the assignment to a linkage group Kar4 paternal.



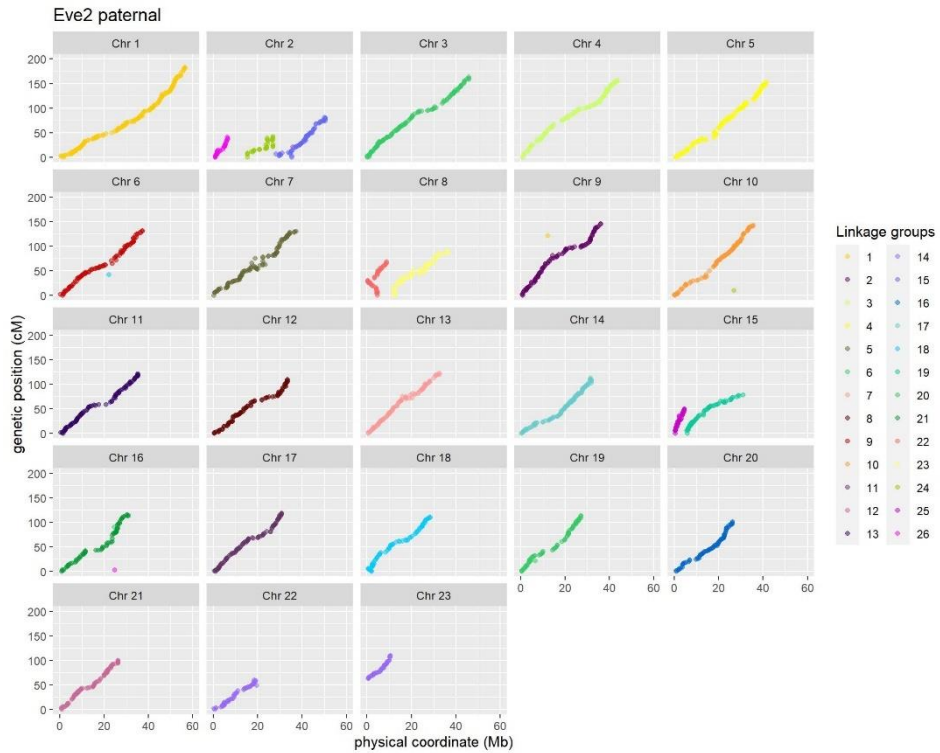
Supplementary Figure S2: Correlation of accumulated genetic distance (cM) and physical distance (Mb). The physical positions of the single nucleotide polymorphism (SNP) markers in *Fraxinus excelsior* were plotted against their genetic positions in the linkage map. The colours of each chromosome represent the assignment to a linkage group Kar4 maternal.



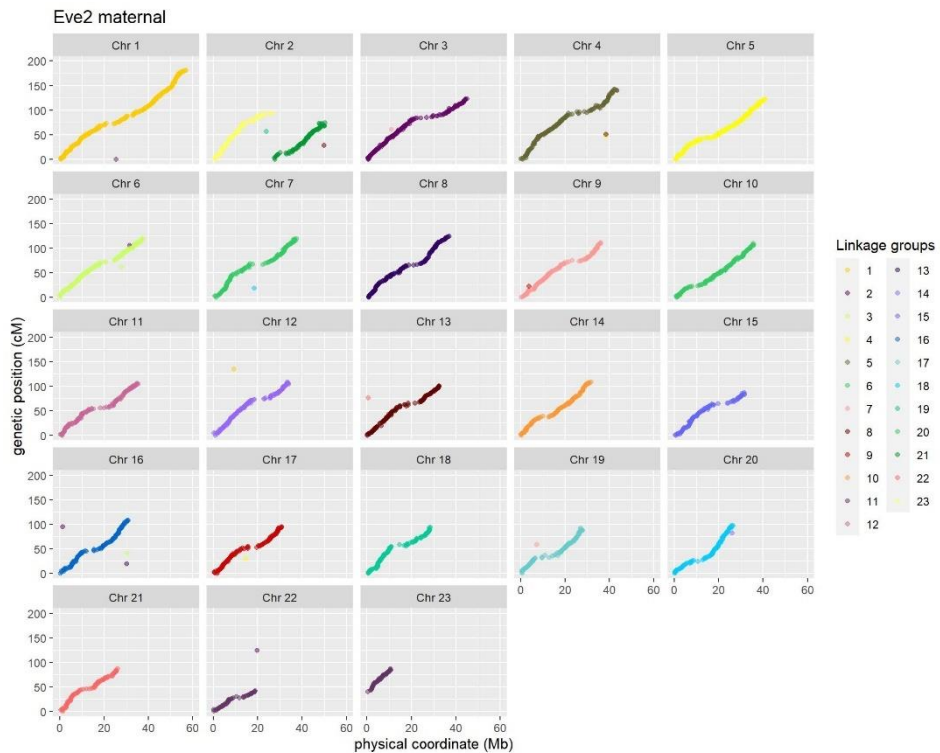
Supplementary Figure S3: Correlation of accumulated genetic distance (cM) and physical distance (Mb). The physical positions of the single nucleotide polymorphism (SNP) markers in *Fraxinus excelsior* were plotted against their genetic positions in the linkage map. The colours of each chromosome represent the assignment to a linkage group Fri8 paternal.



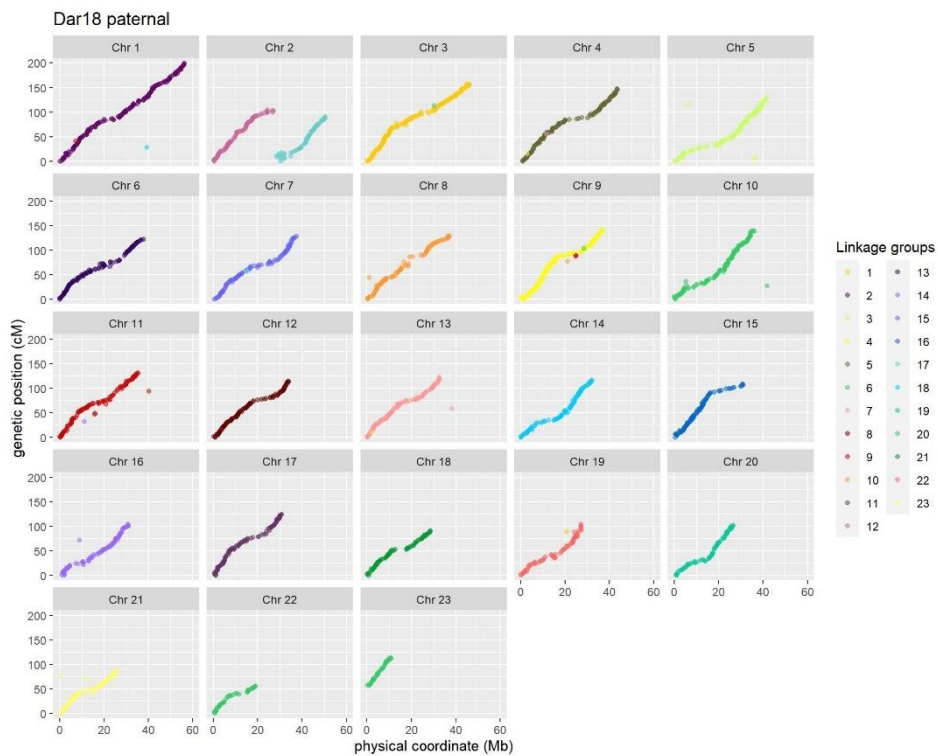
Supplementary Figure S4: Correlation of accumulated genetic distance (cM) and physical distance (Mb). The physical positions of the single nucleotide polymorphism (SNP) markers in *Fraxinus excelsior* were plotted against their genetic positions in the linkage map. The colours of each chromosome represent the assignment to a linkage group Fri8 maternal.



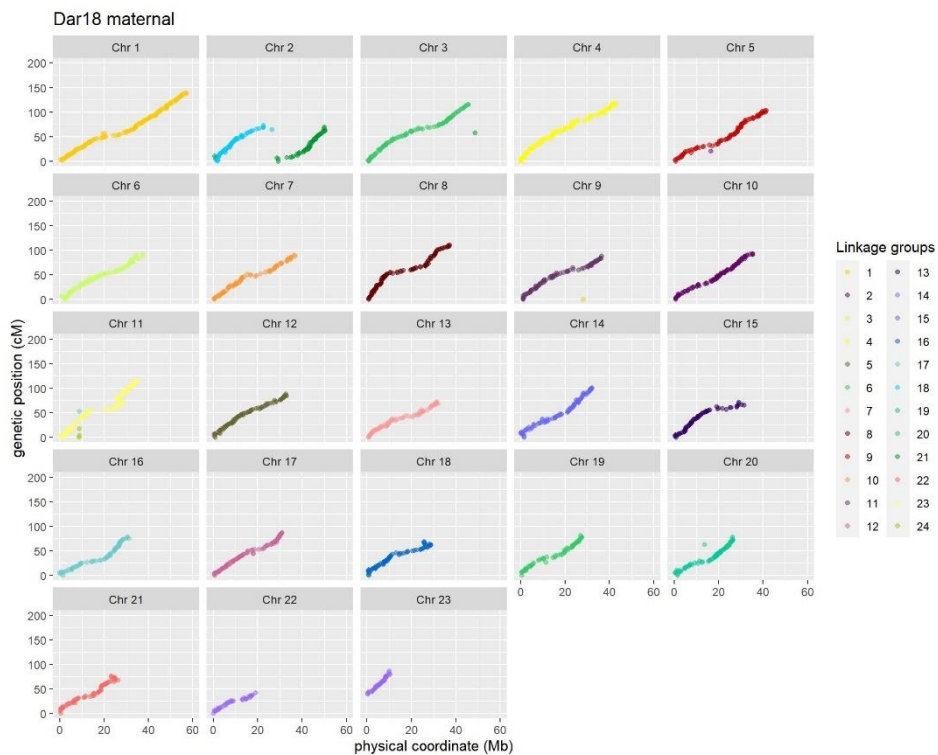
Supplementary Figure S5: Correlation of accumulated genetic distance (cM) and physical distance (Mb). The physical positions of the single nucleotide polymorphism (SNP) markers in *Fraxinus excelsior* were plotted against their genetic positions in the linkage map. The colours of each chromosome represent the assignment to a linkage group Eve2 paternal.



Supplementary Figure S6: Correlation of accumulated genetic distance (cM) and physical distance (Mb). The physical positions of the single nucleotide polymorphism (SNP) markers in *Fraxinus excelsior* were plotted against their genetic positions in the linkage map. The colours of each chromosome represent the assignment to a linkage group Eve2 maternal.



Supplementary Figure S7: Correlation of accumulated genetic distance (cM) and physical distance (Mb). The physical positions of the single nucleotide polymorphism (SNP) markers in *Fraxinus excelsior* were plotted against their genetic positions in the linkage map. The colours of each chromosome represent the assignment to a linkage group Dar18 paternal.



Supplementary Figure S8: Correlation of accumulated genetic distance (cM) and physical distance (Mb). The physical positions of the single nucleotide polymorphism (SNP) markers in *Fraxinus excelsior* were plotted against their genetic positions in the linkage map. The colours of each chromosome represent the assignment to a linkage group Dar18 maternal.

Chapter 3

QTL analysis of ash dieback

Melina Krautwurst¹, Birgit Kersten¹, Franziska Past², Ben Bubner² & Niels A. Müller¹

¹Thünen Institute of Forest Genetics, Großhansdorf, Germany

² Thünen Institute of Forest Genetics, Waldsieversdorf, Germany

In preparation



Abstract

Ash dieback, caused by *Hymenoscyphus fraxineus*, is threatening common ash (*Fraxinus excelsior*) across Europe. Evidence suggests that susceptibility to ash dieback is a heritable polygenic trait. We conducted a genome-wide quantitative trait locus (QTL) analysis on pseudo-backcross full-sibling families to identify chromosomal regions associated with ash dieback susceptibility in *Fraxinus excelsior*. The analysis was performed on four full-sibling families over three years, focusing on stem collar necrosis (SCN), a symptom of ash dieback. We identified one QTL for SCN, which explained 15% of the phenotypic variance, indicating a strong contribution of this locus to susceptibility to the disease.

Keywords: quantitative trait locus (QTL) analysis, ash dieback, stem collar necrosis

Introduction

Ash dieback (ADB) is a disease affecting common ash (*Fraxinus excelsior* L.), which is caused by the invasive necrotrophic ascomycete fungus *Hymenoscyphus fraxineus* (Kowalski, 2006). The fungus, originally native to East Asia, was first observed in Europe in north-eastern Poland in 1992 and has since spread throughout the continent (Husson et al., 2011; McKinney et al., 2012; Zhao et al., 2013). Contrary to the ADB epidemic it causes in Europe, the East Asian host of *H. fraxineus* is not affected by the pathogen.

H. fraxineus is heterothallic and reproduces sexually on ash petioles in leaf litter annually, with 90% of isolated fungi being infectious (Shamsi et al., 2024). The fungus disperses via wind-borne ascospores, and the mycelium spreads from infected leaves into the woody parts of the tree, causing necrosis and crown dieback. Stem collar necrosis (SCN) is defined as a basal lesion with necrotic tissue on the outside; the shape depends on the progress of the infection and/or the associated fungi (Peters et al., 2023). Multiple studies confirm that SCN is caused by *H. fraxineus* (Chandelier et al., 2016; Langer, 2017; Peters et al., 2023; Peters et al., 2024). The

fungus triggers the SCN, which may subsequently involve other fungus species, explaining why *H. fraxineus* is not always identified (Peters et al., 2023). Trees of all ages are susceptible, with younger trees exhibiting higher mortality rates of (~ 82%) compared to older trees (~70%) (Coker et al., 2019; Madsen et al., 2021; Laubray et al., 2024).

To secure the future of *F. excelsior* in commercial forestry settings but also for ecosystem conservation, breeding for lower disease susceptibility could be an important contribution. Understanding the genetic architecture underlying ADB susceptibility is a key component towards this goal. Genetics impact an individual's susceptibility to *H. fraxineus* (Stocks et al., 2017). Quantitative trait locus (QTL) mapping can identify specific regions of the genome associated with disease susceptibility, as quantitative phenotypic differences arise from the segregation of alleles at multiple QTL (Mackay, 2001). High-density genetic maps and SNP markers enhance the precision of QTL identification (Jamann et al., 2015; Torello Marinoni et al., 2018). The presence of a gene within a QTL can indicate its role in controlling trait expression (Thumma et al., 2010).

Due to the time and resource constraints associated with generating specific crosses, we utilised the breeding-without-breeding (BwB) method (El-Kassaby et al., 2007) to create four full-sibling families for QTL analysis. Pseudo-testcross breeding is most effective when a small number of genes control the target trait (Fan et al., 2024). While single-family QTL mapping studies have significant power to detect marker-trait associations, they are limited to loci segregating in the two parents. Therefore, using families with different genetic backgrounds is essential to broaden the scope of detectable loci and capture diverse sources of resistance genes (Lespinasse et al., 2000).

To identify the chromosomal regions affecting ash dieback susceptibility, we conducted a genome-wide QTL analysis of SCN on ash saplings from four different families. A QTL in one family was detected, which explained 15% of the phenotypic variance and is associated with one gene whose expression may be connected to fungus susceptibility.

Material and Methods

Family identification, variant calling, and linkage map construction

The program COLONY (Jones and Wang, 2010a) was used based on SSR markers to identify full-sibling (full-sib) families within segregating populations. Out of 4,000 individuals, about 1000 per mother tree, four large full-sib families (n=115-162) were identified. They were further confirmed with high-resolution genotyping of more than one million genome-wide SNPs identified with Illumina low-coverage resequencing. For further details, see Chapter 1 (Krautwurst et al., 2023). Linkage map construction for paternal and maternal data was conducted with Lep-MAP3 (Rastas, 2017) software package, details are given in Chapter 2.

Phenotypic traits

The necrosis assessment on all full-sib families was conducted every year in March for three years (2022-2024). In 2022, the saplings were two years old and for 1,5 years at the trial side. The trial site was Schulzendorf, Mecklendorf- Mecklenburg Western Pomerania Germany (Coordinates 52.686818, 14.125038), and ash trees with ash dieback symptoms are growing next to the trial side thereby generating infection pressure for our experimental plants. The assessment is based on the SCN scoring key for saplings of the FraxforFuture project (Langer et al., 2022). The SCN scoring focuses on the saplings' primary stem, the area below the uppermost lateral shoot (Figure 1). It is scored with a 0 for non-infected individuals (no SCN can be detected) and 1 for infected trees (SCN is present).

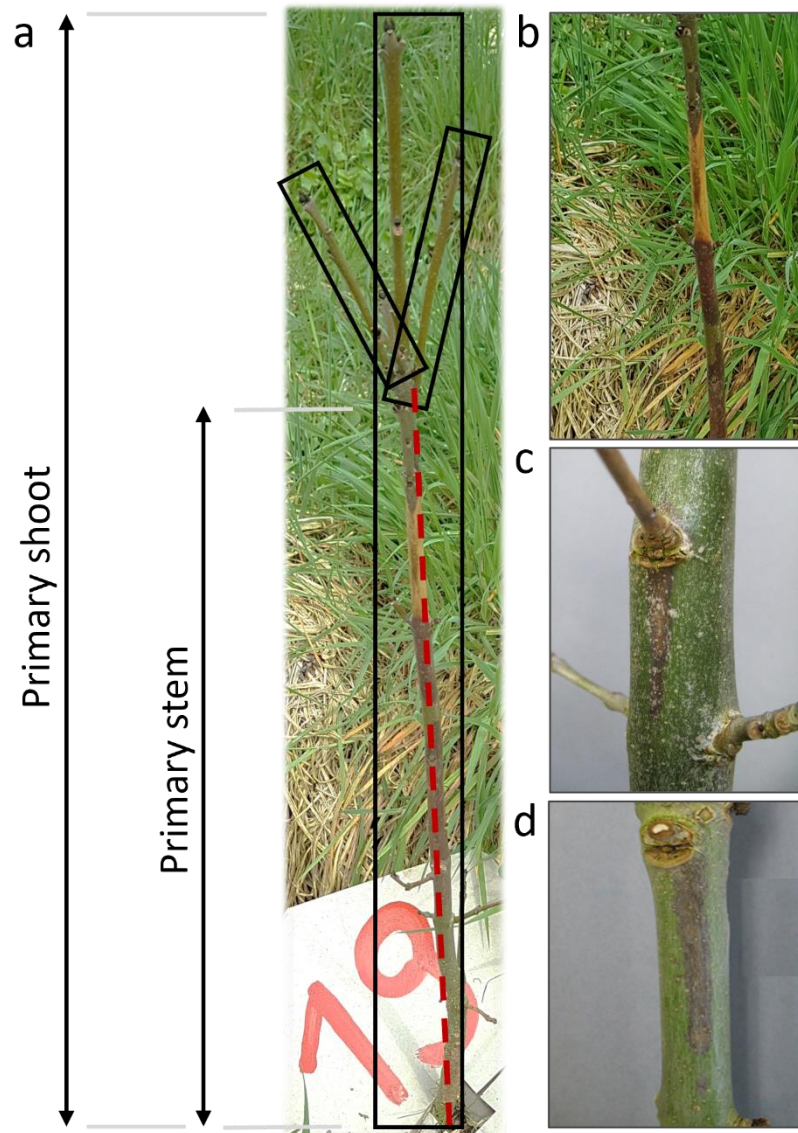


Figure 1: Schematic overview of stem collar necrosis scoring key. A common ash sapling illustrates the primary shoot and primary stem in panel a. Panels b-d demonstrate different scales of necrosis severity with examples. Panel b shows a necrosis that is fully covering the stem. Panel c necrosis double the length of the shoot diameter. Panel d shows a necrosis three times the length of the shoot diameter. Panel b-c had been scored with a 1 (infection, SVN appearing). Photo credits: Franziska Past and Dirk Wegner.

QTL analysis

For the QTL detection, we used the R-package *r/qtl* (Broman et al., 2003). The data for each year and family, and for paternal and maternal data were analysed separately; a total of twelve QTL analyses were conducted. The physical maps (with respect to the chromosome-level reference genome (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_019097785.1/)) were used based on the markers successfully placed into the genetic linkage maps created with *Lep-MAP3* (Chapter 2). Since the data is categorised as 'pseudo-backcross', *r/qtl*

identified the data as 'back cross' in the function 'read.cross'. The function 'calc.genoprob' was used to calculate conditional genotype probabilities using the hidden Markov model technology. The settings for every physical map were repeated (step = 1, off.end=0.0, error.prob=1.0e-4, stepwidth = "fixed"). The genome scan with a single QTL model, function 'scanone', was performed. The 'binary' model was chosen since the phenotype is either 0 or 1. Additionally, the method 'hk' is based on the Haley-Knott regression (Haley and Knott, 1992). The 'scanone' function was used to test two more methods: 'em' (maximum likelihood performed with EM algorithm (Dempster et al., 1977) and 'mr' (marker regression (Soller et al., 1976)), which is a regression of the phenotypes on the multipoint QTL genotype probabilities. Individuals with missing phenotypes were dismissed (Figure 2). The missing phenotypes resulted mainly from saplings whose primary stem was not clearly identified. The inconclusive growth forms resulted mostly from a late frost in 2021 and 2022, deceased shots, and deformed regrowth. Only a small number of saplings were identified as dead.

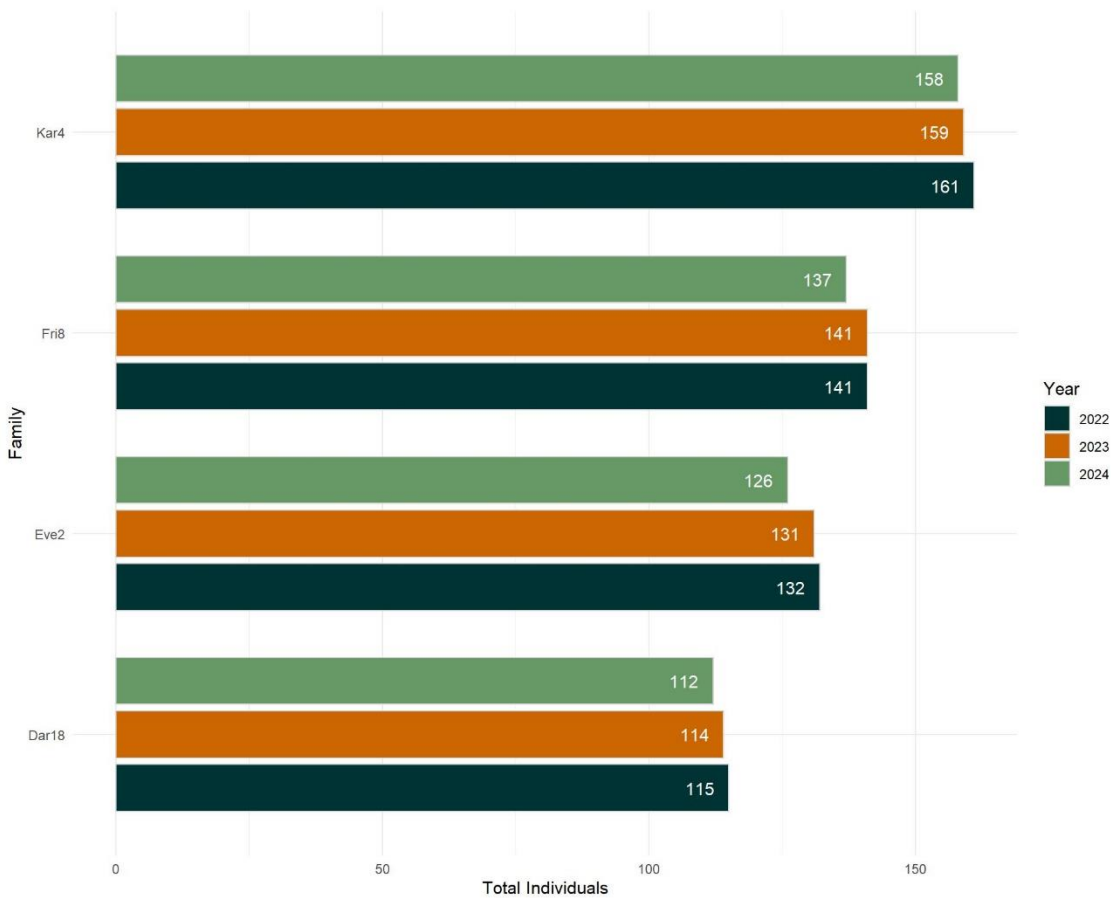


Figure 2: The number of individuals in each family every year (2022, 2023, 2024) for QTL analysis. Individuals with missing phenotypes have been dismissed.

A 'n.perm' of 1000 was chosen to calculate a permutation test. The function performs a permutation test with 1000 replicates. The effect size of significant QTL was calculated with the functions 'makeqtl' and 'fitqtl'. For 'fitqtl', the method 'hk' and the 'binary' model were chosen. The visualisation of the QTL plots was conducted with the package 'ggplot2' and 'ggpubr' (Wickham, 2016).

Structural and functional annotation of the QTL region

A partial nucleotide sequence of *F. excelsior* chromosome 16 (24,000,000-25,000,000 bp) including the QTL region (24,521,848-24,960,370 bp) was extracted from the total sequence of chromosome 16 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_019097785.1/) using CLC Genomics Workbench (CLC GWB; v.23.05) and subjected to structural annotation by AUGUSTUS (Stanke et al., 2004, <https://bioinf.uni-greifswald.de/augustus/>; "organism": *Arabidopsis thaliana*; "report genes on": both strands; "alternative transcripts": few). The resulting gff-file and a fasta-file, including the protein sequences related to the annotated transcript models, were downloaded from the AUGUSTUS web server.

For functional annotation, the predicted *F. excelsior* protein sequences were analysed versus *A. thaliana* protein sequences using the "BLAST" tool of CLC GWB (BLASTP). To create the BLAST database, coding sequences (CDS) from *A. thaliana* (file "Araport11_cds_20240409_representative_gene_model") were downloaded from TAIR (<https://www.arabidopsis.org/>), translated to protein sequences using the tool "translate to protein" of CLC GWB and used as an input for the tool "create BLAST database" of CLC GWB. The BLAST hits were filtered by e-value keeping only hits with e-values below e-30.

Gene Ontology (GO) terms were assigned to *A. thaliana* gene identifiers based on functional annotations in STRING v.12.0 (<https://string-db.org/>, Szklarczyk et al., 2019). Using a nonredundant list of *A. thaliana* gene identifiers as input, functional annotations were downloaded after analysis in STRING and screened for GO terms related to "fungus" (GO terms that include "fungus" in their description).

Functional interaction networks of single *A. thaliana* genes were created using STRING v.12.0 using the gene identifier as an input (default parameters). Networks, protein annotations and functional annotations were exported from STRING. Functional annotation files were screened for GO terms related to fungus.

Results

The prevalence of necrosis varied across the four families and three years (Figure 3). The distribution shows an increase in infected individuals over the three years in all families. The only discrepancy in Dar18 (Figure 3 a) is probably a result of incorrect phenotyping. However, this did not affect the overall identification of significant QTLs. This was tested by running the QTL analysis with different ‘test’ datasets.

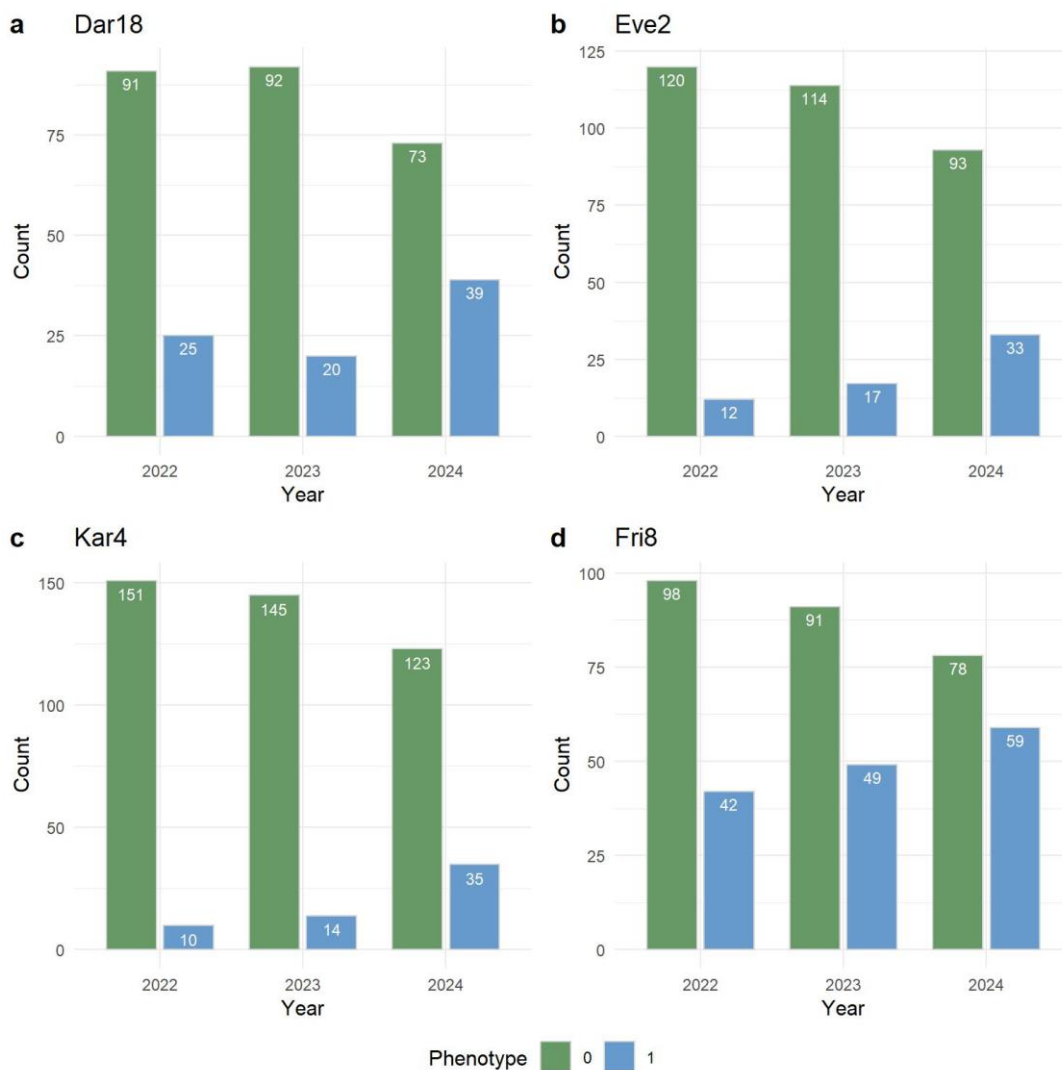


Figure 3: An overview of the distribution of necrosis over the years in all four families. 0 = no infection on the main stem (no necrosis), and 1 = infection on the main stem (necrosis).

The QTL analyses for ash dieback only revealed a single significant ($p < 0.05$) locus in one family (Dar18) and one year (2022; Figure 4-7 and Supplementary Figure S1-S4). Three different methods ('hr', 'em' and 'mr') have been evaluated using the 'scanone' function in r/qtl. They all showed virtually identical results. The QTL analysis of other families also showed peaks (Figure 4-7) but only marginally significant and without consistency across years. Kar4's paternal peaked at chromosome 16 in 2022 (Figure 7 a). Fri8 had a nearly significant peak at chromosome 20 on the paternal and maternal map in 2024, but not in the other years (Figure 5 c). Nevertheless, this might indicate a locus involved in tolerance to ash dieback, which can only be identified once the disease develops a certain level of infection. It will be exciting to continue performing QTL analyses with the phenotypes of the next years.

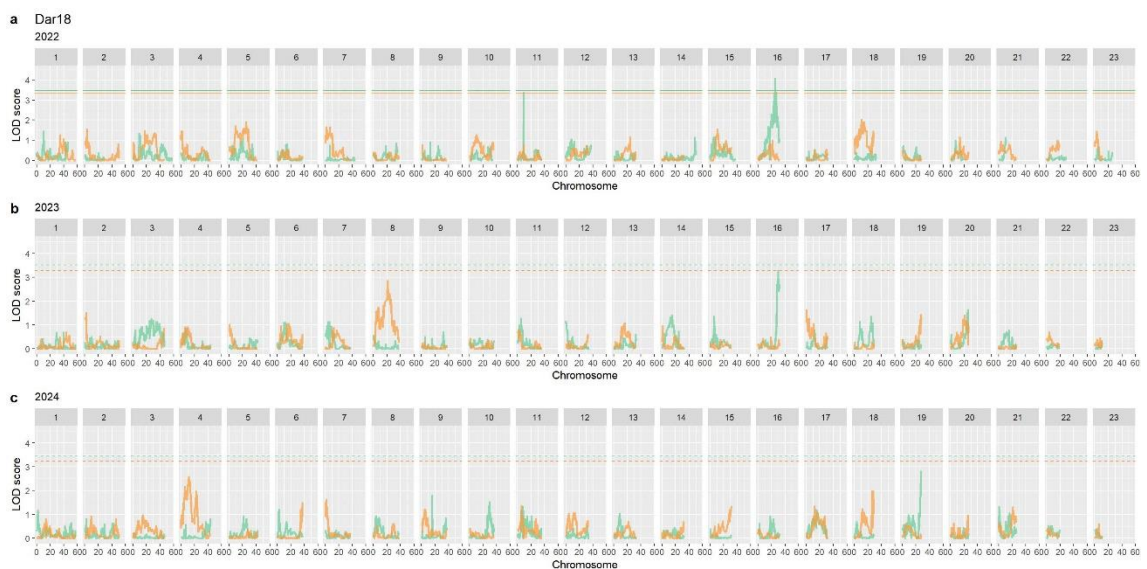


Figure 4: Quantitative Trait Loci (QTL) mapping results for Dar18 across three years (2022-2024; based on the physical map and LOD threshold 5%). Each panel represents a separate chromosome, numbered 1 to 23, with the x-axis indicating the position on the chromosome (in mega base pairs) and the y-axis showing the LOD (logarithm of the odds) score. Panel a: year 2022, Panel b: year 2023, Panel c: year 2024. The LOD score is plotted for each marker along the chromosomes in each panel, with significant peaks indicating potential QTLs. The horizontal dashed lines represent the LOD threshold for significance. Green lines indicate paternal significance and orange lines represent maternal significance.

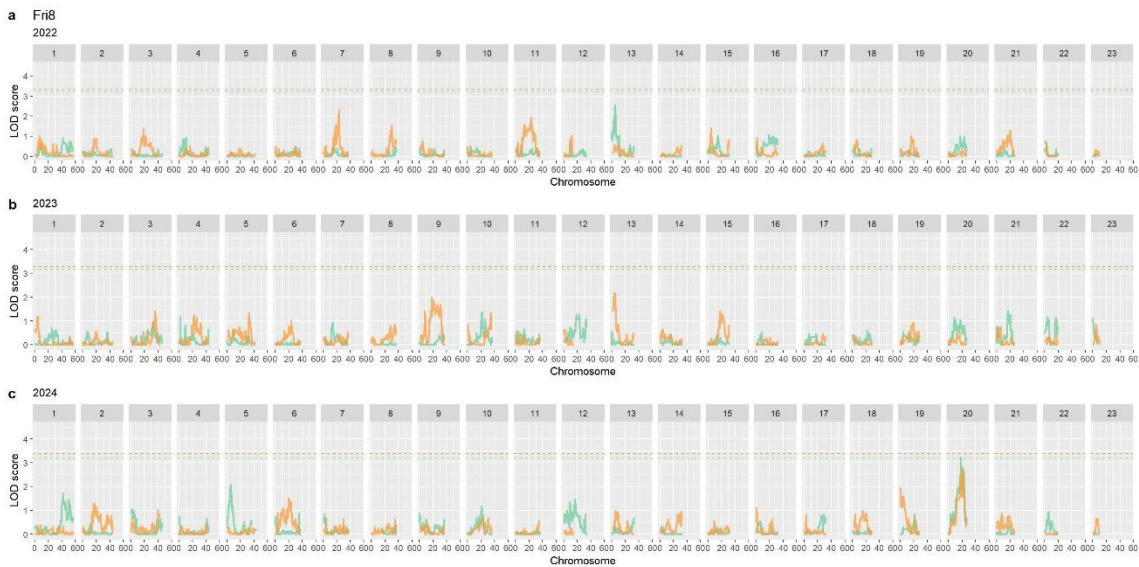


Figure 5: Quantitative Trait Loci (QTL) mapping results for Fri8 across three years (2022-2024; based on the physical map and LOD threshold 5%). Each panel represents a separate chromosome, numbered 1 to 23, with the x-axis indicating the position on the chromosome (in mega base pairs) and the y-axis showing the LOD (logarithm of the odds) score. Panel a: year 2022, Panel b: year 2023, Panel c: year 2024. The LOD score is plotted for each marker along the chromosomes in each panel, with significant peaks indicating potential QTLs. The horizontal dashed lines represent the LOD threshold for significance. Green lines indicate paternal significance and orange lines represent maternal significance.

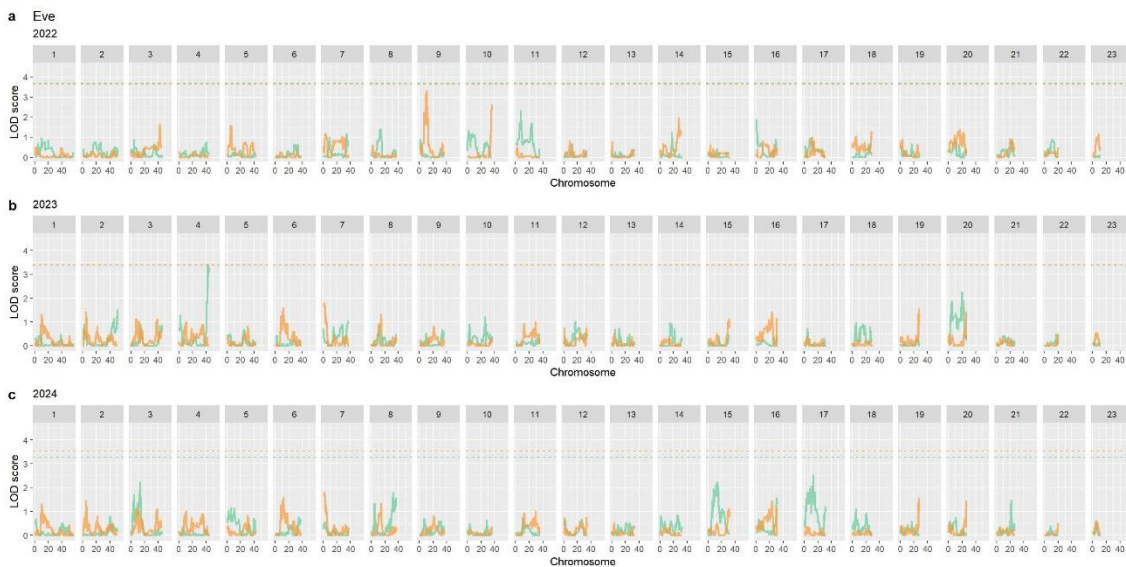


Figure 6: Quantitative Trait Loci (QTL) mapping results for Eve2 across three years (2022-2024; based on the physical map and LOD threshold 5%). Each panel represents a separate chromosome, numbered 1 to 23, with the x-axis indicating the position on the chromosome (in mega base pairs) and the y-axis showing the LOD (logarithm of the odds) score. Panel a: year 2022, Panel b: year 2023, Panel c: year 2024. The LOD score is plotted for each marker along the chromosomes in each panel, with significant peaks indicating potential QTLs. The horizontal dashed lines represent the LOD threshold for significance. Green lines indicate paternal significance and orange lines represent maternal significance.

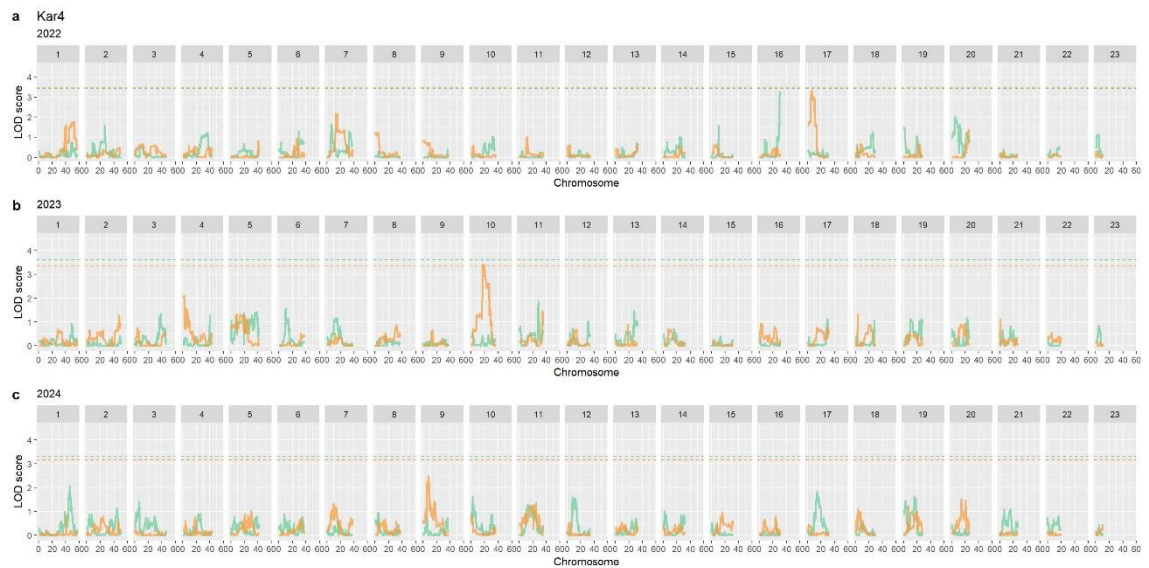


Figure 7: Quantitative Trait Loci (QTL) mapping results for Kar4 across three years (2022-2024; based on the physical map and LOD threshold 5%). Each panel represents a separate chromosome, numbered 1 to 23, with the x-axis indicating the position on the chromosome (in mega base pairs) and the y-axis showing the LOD (logarithm of the odds) score. Panel a: year 2022, Panel b: year 2023, Panel c: year 2024. The LOD score is plotted for each marker along the chromosomes in each panel, with significant peaks indicating potential QTLs. The horizontal dashed lines represent the LOD threshold for significance. Green lines indicate paternal significance and orange lines represent maternal significance.

The only significant QTL on the paternal map of family Dar18 identified in 2022 (Figure 4 a) explains 15% of the phenotypic variance (Chi-square test, $p = 1.51e^{-05}$). The low p -value indicates that the association between the QTL and the binary phenotype is highly significant and unlikely to be due to a random chance. It also strongly supports the existence of the QTL effect on the phenotype. This locus remained in 2023, with a lower LOD score not passing the 5% significance threshold (Figure 4 b).

The significant QTL is located on chromosome 16 between 24,520-24,960 kb. The structural annotation of the broader genomic region from 24-25 Mbp including the QTL revealed 140 putative genes with 155 transcript models (Figure 8).

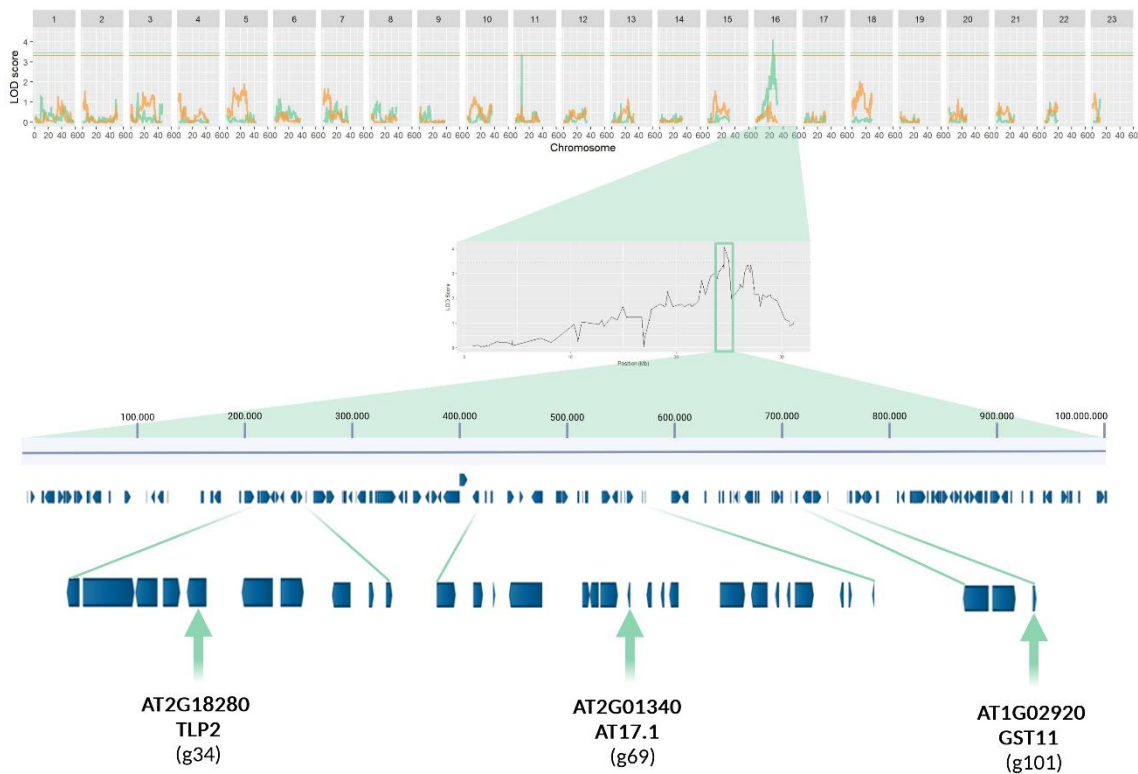


Figure 8: This figure illustrates the identification and localization of the quantitative trait locus (QTL) for stem collar necrosis (SCN) susceptibility in the family Dar18 on chromosome 16. Top panel shows LOD (Logarithm of the Odds) score plots across all 23 chromosomes. Each plot represents the results of the QTL analysis for SCN susceptibility, with the x-axis indicating chromosomal positions and the y-axis showing the LOD scores. Peaks in the LOD score plots represent potential QTLs. Middle panel shows a zoomed-in view of chromosome 16, where a significant QTL for SCN susceptibility was identified. The x-axis displays the chromosomal position in mega base pairs (Mbp), and the y-axis shows the LOD score. The highlighted region indicates the QTL peak with the highest LOD score. Bottom panel shows a detailed view of the identified QTL region on chromosome 16. This section illustrates the annotated genes (140) within the QTL region. Each blue rectangle represents a gene, with the size corresponding to the gene length. Created with Biorender.

For functional annotation, we mapped the *F. excelsior* genes to *A. thaliana* genes based on BLASTP-proven protein sequence similarity of related transcript models, thus assigning 69 *F. excelsior* gene identifiers to 51 different *A. thaliana* gene identifiers (Supplementary Tables S1; protein sequence identities of related mapped genes: 34.1% - 89.5%). These 51 non-redundant *A. thaliana* gene identifiers (Supplementary Table S2) were analysed by STRING (Szklarczyk et al., 2019) to identify genes potentially related to plant's defense responses to fungus. Three *A. thaliana* genes were identified to be assigned to related GO terms (Table 1). These *A. thaliana* genes map to the following *F. excelsior* genes annotated in the broader QTL region: g34, g69, g101 (Table 1; Supplementary Table S1; Figure 8). Table 2 presents annotation details and BLASTP results related to the three *A. thaliana* genes with “fungus”-related GO-terms.

Table 1: Putative *Fraxinus excelsior* genes in the chromosome 16-region at 24-25 Mbp (including the QTL region) with potential function in plant's defense responses to fungus (function assigned based on GO annotation of most similar *Arabidopsis thaliana* genes (Supplementary Table S1; annotation details in Table 2) according to the STRING database (<https://string-db.org/>; Szklarczyk et al. (2019)).

<i>F. excelsior</i> gene ID	Chromo-somal position (start)	Chromo-somal Position (end)	Str.	<i>A. thaliana</i> gene ID	<i>A. thaliana</i> other name	GO term(s) related to response to fungus
g34	24,234,442	24,237,838	-	AT2G18280	TLP2	GO:0009620
g69	24,503,512	24,504,360	-	AT2G01340	AT17.1	GO:0009620
g101	24,740,492	24,741,641	+	AT1G02920	GST11	GO:0009620 GO:0050832

Gene ID, gene identifier; GO:0009620, "response to fungus"; GO:0050832, "defense response to fungus". Both GO terms are assigned to the main GO domain "biological process". Str., DNA-strand.

Table 2: Blast results of selected *Fraxinus excelsior* genes (Table 1) versus *Arabidopsis thaliana* (BLASTP of related protein sequences) and functional annotation of the most similar *Arabidopsis thaliana* genes according to STRING (<https://string-db.org/>; Szklarczyk et al. (2019)).

<i>F. excelsior</i> gene ID	<i>A. thaliana</i> gene ID of best BLASTP hit	<i>A. thaliana</i> other name	<i>A. thaliana</i> gene description	E-value	Identity (%)	Query coverage (%)
g34	AT2G18280	TLP2	TUBBY LIKE PROTEIN 2	4e ⁻¹³¹	52.0	100
g69	AT2G01340	AT17.1	Encodes a protein whose expression is responsive to nematode infection; PADRE protein up-regulated after infection by <i>Sclerotinia sclerotiorum</i>	2e ⁻⁴⁰	47.8	94
g101	AT1G02920	GST11	GLUTATHIONE S-TRANSFERASE 11	3e ⁻⁷⁷	55.5	97

Related E-value: Indicating the statistical significance of the match. Lower values signify higher significance. Identity: The percentage of amino acids that are identical between the query sequence and the BLASTP hit. Query coverage: The percentage of the query sequence that is covered by the alignment with the BLASTP hit.

According to the annotation in STRING, the three *A. thaliana* proteins with "fungus"-related GO terms were identified as TLP2, AT17.1 and GST11 (Table 2). TLP is a 'tubby like protein 2' and, according to TAIR (<https://www.arabidopsis.org/>) a 'member of TLP family', the Thaumatin-like protein (TLP) family, which are part of a large pathogenesis-related (PR) gene family, involved in a broad range of defense and developmental processes in plants, fungi and animals (Brandazza et al., 2004). AT17.1 is described as a 'a protein whose expression is responsive to nematode infection; PADRE protein up-regulated after infection by *Sclerotinia sclerotiorum*' (Table 2). Interestingly, the fungus *Sclerotinia sclerotiorum* is assigned to the same order as *H. fraxineus* (order Helotiales; according to NCBI taxonomy; <https://www.ncbi.nlm.nih.gov/taxonomy>). AT17.1 includes - according to TAIR - the plant-specific PADRE domain (PFAM DUF4228; IPR025322), which is the pathogen and abiotic stress response, cadmium tolerance, disordered region-containing (PADRE) domain, that typically occurs in small single-domain proteins with a bipartite architecture. This domain is associated with plant defense against diverse stress stimuli and has a role in disease resistance to fungi (Didelon et al., 2020). In the STRING database (<https://string-db.org>), At17.1 is described as 'plastid movement impaired protein At17.1'. GST11 is a putative glutathione S-transferase (GST). Plant GSTs are ubiquitous and multifunctional enzymes highly inducible by a wide range of stress conditions, including biotic stress. Numerous transcriptome-wide investigations demonstrated that distinct groups of GSTs are induced in the early phase of bacterial, fungal and viral infections (Gullner et al., 2018).

The protein sequences of these *F. excelsior* genes (Supplementary Table S3) were further analysed by BLASTP versus *Fraxinus pennsylvanica* protein sequences at NCBI (<https://blast.ncbi.nlm.nih.gov/>). In BLASTP, the protein sequences related to the three genes provided hits with low e-values to genes described as 'unnamed protein product' (Table 3). Furthermore, the PADRE domain was also found in the BLASTP hit g69 (At17.1) against *Fraxinus pennsylvanica* protein sequences at NCBI (accession CAI9777728).

Table 3: BLASTP analyses were performed for the protein sequences of three *Fraxinus excelsior* genes (g34, g69, and g101; sequences in Table S3) against the protein database of *Fraxinus pennsylvanica* (green ash) at NCBI.

<i>F. excelsior</i> gene ID	<i>F. pennsylvanica</i> gene ID of best BlastP hit	<i>F. pennsylvanica</i> gene description	Related e-value	Identity (%)	Query coverage (%)
g34	CAI9777705	unnamed protein product	0	82.6	100
g69	CAI9777728	unnamed protein product	1e-94	93.0	100
g101	CAI9781113	unnamed protein product	1e-148	93.9	100

Gene ID, gene identifier. Related E-value: Indicating the statistical significance of the match. Lower values signify higher significance. Identity Percentage: The percentage of amino acids that are identical between the query sequence and the BLASTP hit. Query Coverage: The percentage of the query sequence that is covered by the alignment with the BLASTP hit.

The STRING network analysis for the protein At17.1 in *A. thaliana* revealed potential interactions and functional associations (Figure 9). The analysis identified multiple proteins interacting with At17.1. Notably, a significant subset of these interacting proteins, highlighted by red-coloured nodes, shares the GO term 'response to fungus' and/or 'defense response to fungus' (Table 4). Interestingly, all interaction partners in the STRING network (Figure 9) as well as At17.1 (bait in the network) are annotated as "PADRE proteins" according to TAIR (Table 4). These results further support a potential role for At17.1 in fungal response pathways, indicating that it may participate in mechanisms activated during fungal infections.

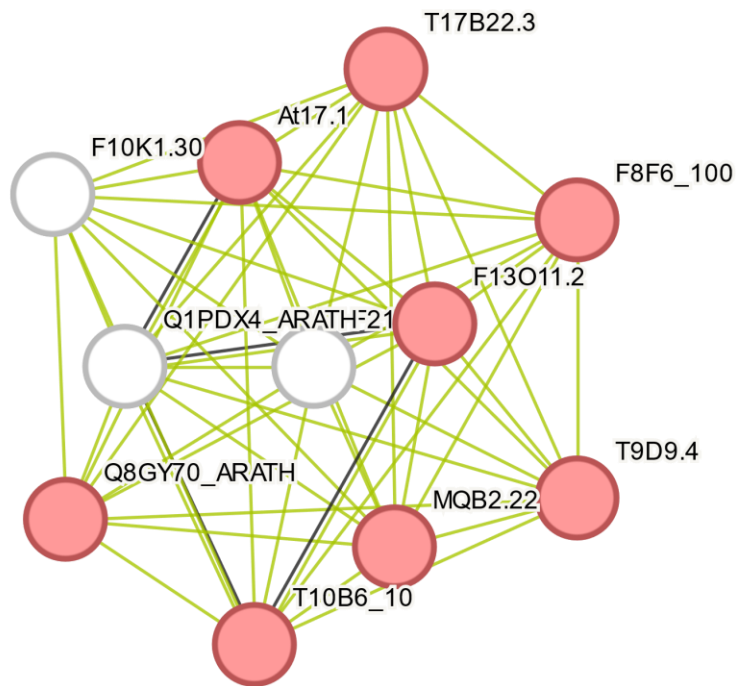


Figure 9: The STRING network analysis of the protein At17.1 (AT2G01340). Each node in the network represents a protein, and each edge (line) represents an interaction between the proteins. The red nodes indicate proteins with the following assigned Gene Ontology (GO) terms: 'response to fungus' (GO:0009620) and/or 'defense response to fungus' (GO:0050832). The edges represent distinct types of known and predicted functional interactions. Green lines: These lines indicate interactions inferred from text mining derived from the co-occurrence of protein names in literature. Black lines: These lines represent co-expression data.

Table 4: The table presents protein descriptions all functional interaction partners of At17.1 identified by STRING network analysis (Figure 9).

Node (STRING network)	A. thaliana gene ID	Description (STRING)	Description (TAIR)
At17.1	AT2G01340	Plastid movement impaired protein	PADRE protein up-regulated after infection by <i>S. sclerotiorum</i> *
F10K1.30	AT1G06980	6,7-dimethyl-8-ribityllumazine synthase	PADRE protein
F13O11.2	AT1G64700	DUF4228 domain protein	PADRE protein up-regulated after infection by <i>S. sclerotiorum</i>
F21F14.90	AT3G61920	UvrABC system protein C	PADRE protein
F8F6_100	AT5G03890	Emb CAB85509.1 (hypothetical protein)	PADRE protein up-regulated after infection by <i>S. sclerotiorum</i>
MQB2.22	AT5G62900	Basic-leucine zipper transcription factor K.	PADRE protein down-regulated after infection by <i>S. sclerotiorum</i>
Q1PDX4_ARATH	AT5G12340	DUF4228 domain protein	PADRE protein up-regulated after infection by <i>S. sclerotiorum</i>
Q8GY70_ARATH	AT3G10120	Hypothetical protein	PADRE protein down-regulated after infection by <i>S. sclerotiorum</i>
T10B6_10	AT5G17350	Uncharacterized protein At5g17350.	PADRE protein up-regulated after infection by <i>S. sclerotiorum</i> .
T17B22.3	AT3G03280	T17B22.3 protein. Hypothetical protein	PADRE protein up-regulated after infection by <i>S. sclerotiorum</i> .
T9D9.4	AT2G30230	6,7-dimethyl-8-ribityllumazine synthase.	6,7-dimethyl-8-ribityllumazine synthase (with DUF4228)

* Additional description in Table 2.

Furthermore, STRING network analyses were performed for the proteins TLP2 and GST11 in *A. thaliana* (Supplementary Figures S5-S6). Unlike At17.1, the GO terms of the potential interactors of TLP2 and GST11 did not show any connection to the fungal response. This distinction underscores the specific involvement of At17.1 in the response to fungal pathogens and highlights the differing roles of these proteins in *A. thaliana*.

Discussion

A QTL with a LOD score of 4.066 for stem collar necrosis (SCN) in the family Dar18 was detected. This LOD score exceeds the 5% significance threshold of 3.47 based on 1000 permutations, indicating statistical significance. However, it should be noted that a LOD score of 4 does not indicate a strong association.

Despite the statistical limitations, the biological relevance of this QTL should not be dismissed. The QTL is associated with three genes similar to fungus response-related *A. thaliana* genes (Tables 1-2), including AT17.1, a PADRE protein known to be involved in fungal resistance (Didelon et al., 2020), thus supporting its potential biological relevance. This finding implies that this gene may provide insights into the mechanisms of susceptibility to *Hymenoscyphus fraxineus*.

The fact that the peaks in all families do not repeat in other years indicates that the associated loci are not broadly involved in controlling susceptibility but may rather be genotype-specific and depend on specific environmental conditions differing from year to year. It has been shown that ash dieback takes many years to develop on trial sites fully (Timmermann et al., 2017; Carroll and Boa, 2024). Thus, we can only identify especially susceptible genotypes so far and have not yet differentiated tolerant from susceptible ones. However, it is necessary to identify QTL responsible for ash dieback tolerance. It will be exciting to follow the plants' phenotypes over the next years and perform repeated QTL analyses.

The identified protein At17.1 (AT2G01340), analysed through the STRING network, shows connections to other PADRE proteins (Figure 9, Table 4), most of which are linked to fungal response based on GO annotation. The STRING connections in this network are mainly established through text mining, indicating the prevalence of co-occurrence of protein names in the literature (Doncheva et al., 2019). Especially the co-expression interaction with the 'DUF44228 domain protein' (ID: Q1PDX4_ARATH, AT5G12340; Figure 9, Table 4), a putative PADRE protein up-regulated after infection by *S. sclerotiorum*, is noteworthy (Figure 9, Table 4). More reliable results may emerge from extended functional interaction studies in the future.

Despite the LOD threshold and STRING network limitations, this QTL explains a substantial proportion of the phenotype (15% of the total phenotypic variance). An effect size of this magnitude suggests that the identified QTL may have a strong influence on the trait, making it an interesting target for further genetic and functional studies.

Conclusions

This study highlights the challenges of detecting robust QTLs for stem collar necrosis (SCN) representative of ash dieback susceptibility, especially considering the limitations imposed by stringent LOD thresholds and the limited number of years for ash dieback progression and phenotyping. Extending the monitoring period will be crucial for better understanding the genetic architecture underlying ash dieback susceptibility. Future studies should additionally focus on validation experiments to confirm the identified QTL and elucidate their role in the response of *F. excelsior* to fungal infection. Despite the limitations, the detected QTL explains a considerable proportion of phenotypic variance, suggesting a relatively strong influence on SCN. This finding underscores the importance of continued research to unravel the genetic complexity of susceptibility traits in ash trees.

References

- Brandazza, A., Angeli, S., Tegoni, M., Cambillau, C., and Pelosi, P. (2004). Plant stress proteins of the thaumatin-like family discovered in animals. *FEBS Lett* 572, 3–7. doi: 10.1016/j.febslet.2004.07.003
- Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112
- Carroll, D., and Boa, E. (2024). Ash dieback: From Asia to Europe. *Plant Pathol* 73, 741–759. doi: 10.1111/ppa.13859
- Chandelier, A., Gerarts, F., San Martin, G., Herman, M., and Delahaye, L. (2016). Temporal evolution of collar lesions associated with ash dieback and the occurrence of *Armillaria* in Belgian forests. *For. Path.* 46, 289–297. doi: 10.1111/efp.12258
- Coker, T. L. R., Rozsypálek, J., Edwards, A., Harwood, T. P., Butfoy, L., and Buggs, R. J. A. (2019). Estimating mortality rates of European ash (*Fraxinus excelsior*) under the ash dieback (*Hymenoscyphus fraxineus*) epidemic. *Plants, People, Planet* 1, 48–58. doi: 10.1002/ppp3.11

- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 39, 1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x
- Didelon, M., Khafif, M., Godiard, L., Barbacci, A., and Raffaele, S. (2020). Patterns of Sequence and Expression Diversification Associate Members of the PADRE Gene Family With Response to Fungal Pathogens. *Front. Genet.* 11, 491. doi: 10.3389/fgene.2020.00491
- Doncheva, N. T., Morris, J. H., Gorodkin, J., and Jensen, L. J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res* 18, 623–632. doi: 10.1021/acs.jproteome.8b00702
- Fan, S., Georgi, L. L., Hebard, F. V., Zhebentyayeva, T., Yu, J., Sisco, P. H., et al. (2024). Mapping QTLs for blight resistance and morpho-phenological traits in inter-species hybrid families of chestnut (*Castanea* spp.). *Front Plant Sci* 15, 1365951. doi: 10.3389/fpls.2024.1365951
- Gullner, G., Komives, T., Király, L., and Schröder, P. (2018). Glutathione S-Transferase Enzymes in Plant-Pathogen Interactions. *Front Plant Sci* 9, 1836. doi: 10.3389/fpls.2018.01836
- Haley, C. S., and Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69, 315–324. doi: 10.1038/hdy.1992.131
- Husson, C., Scala, B., Caël, O., Frey, P., Feau, N., Ioos, R., et al. (2011). *Chalara fraxinea* is an invasive pathogen in France. *Eur J Plant Pathol* 130, 311–324. doi: 10.1007/s10658-011-9755-9
- Jamann, T. M., Balint-Kurti, P. J., and Holland, J. B. (2015). QTL mapping using high-throughput sequencing. *Methods Mol Biol* 1284, 257–285. doi: 10.1007/978-1-4939-2444-8_13
- Jones, O. R., and Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol Ecol Resour* 10, 551–555. doi: 10.1111/j.1755-0998.2009.02787.x
- Kowalski, T. (2006). *Chalara fraxinea* sp. nov. associated with dieback of ash (*Fraxinus excelsior*) in Poland. *For. Path.* 36, 264–270. doi: 10.1111/j.1439-0329.2006.00453.x
- Krautwurst, M., Past, F., Kersten, B., Bubner, B., and Müller, N. A. (2023). Identification of full-sibling families from natural single-tree ash progenies based on SSR markers and genome-wide SNPs. doi: 10.1101/2023.07.18.549475
- Langer, G. (2017). Collar rots in forests of Northwest Germany affected by ash dieback. *Baltic Forestry* 23, 44–19.
- Langer, G. J., Fuchs, S., Osewold, J., Peters, S., Schrewe, F., Ridley, M., et al. (2022). FraxForFuture—research on European ash dieback in Germany. *J Plant Dis Prot*, 1–11. doi: 10.1007/s41348-022-00670-z
- Laubray, S., Buée, M., and Marçais, B. (2024). *Hymenoscyphus fraxineus* persistence in the ash litter. *Plant Pathol.* doi: 10.1111/ppa.13948
- Lespinasse, D., Grivet, L., Troispoux, V., Rodier-Goud, M., Pinard, F., and Seguin, M. (2000). Identification of QTLs involved in the resistance to South American leaf blight (*Microcyclus ulei*) in the rubber tree. *Theor Appl Genet* 100, 975–984. doi: 10.1007/s001220051379
- Mackay, T. F. (2001). The genetic architecture of quantitative traits. *Annu Rev Genet* 35, 303–339. doi: 10.1146/annurev.genet.35.102401.090633
- Madsen, C. L., Kosawang, C., Thomsen, I. M., Hansen, L. N., Nielsen, L. R., and Kjær, E. D. (2021). Combined progress in symptoms caused by *Hymenoscyphus*

- fraxineus* and *Armillaria* species, and corresponding mortality in young and old ash trees. *Forest Ecology and Management* 491, 119177. doi: 10.1016/j.foreco.2021.119177
- McKinney, L. V., Thomsen, I. M., Kjaer, E. D., and Nielsen, L. R. (2012). Genetic resistance to *Hymenoscyphus pseudoalbidus* limits fungal growth and symptom occurrence in *Fraxinus excelsior*. *For. Path.* 42, 69–74. doi: 10.1111/j.1439-0329.2011.00725.x
- Peters, S., Fuchs, S., Bien, S., Bußkamp, J., Langer, G. J., and Langer, E. J. (2023). Fungi associated with stem collar necroses of *Fraxinus excelsior* affected by ash dieback. *Mycol Progress* 22. doi: 10.1007/s11557-023-01897-2
- Peters, S., Gruschwitz, N., Bien, S., Fuchs, S., Bubner, B., Blunk, V., et al. (2024). The fungal predominance in stem collar necroses of *Fraxinus excelsior*: a study on *Hymenoscyphus fraxineus* multilocus genotypes. *J Plant Dis Prot*, 1–13. doi: 10.1007/s41348-024-00912-2
- Rastas, P. (2017). Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* 33, 3726–3732. doi: 10.1093/bioinformatics/btx494
- Shamsi, W., Mittelstrass, J., Ulrich, S., Kondo, H., Rigling, D., and Prospero, S. (2024). Possible Biological Control of Ash Dieback Using the Mycoparasite *Hymenoscyphus Fraxineus* Mitovirus 2. *Phytopathology* 114, 1020–1027. doi: 10.1094/PHYTO-09-23-0346-KC
- Soller, M., Brody, T., and Genizi, A. (1976). On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theor Appl Genet* 47, 35–39. doi: 10.1007/BF00277402
- Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 32, W309–12. doi: 10.1093/nar/gkh379
- Stocks, J. J., Buggs, R. J. A., and Lee, S. J. (2017). A first assessment of *Fraxinus excelsior* (common ash) susceptibility to *Hymenoscyphus fraxineus* (ash dieback) throughout the British Isles. *Sci Rep* 7, 16546. doi: 10.1038/s41598-017-16706-6
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47, D607–D613. doi: 10.1093/nar/gky1131
- Thumma, B. R., Southerton, S. G., Bell, J. C., Owen, J. V., Henery, M. L., and Moran, G. F. (2010). Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens*. *Tree Genetics & Genomes* 6, 305–317. doi: 10.1007/s11295-009-0250-9
- Timmermann, V., Nagy, N. E., Hietala, A. M., Børja, I., and Solheim, H. (2017). Progression of Ash Dieback in Norway Related to Tree Age, Disease History and Regional Aspects. 1392–1355.
- Torello Marinoni, D., Valentini, N., Portis, E., Acquadro, A., Beltramo, C., Mehlenbacher, S. A., et al. (2018). High density SNP mapping and QTL analysis for time of leaf budburst in *Corylus avellana* L. *PLOS ONE* 13, e0195408. doi: 10.1371/journal.pone.0195408
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer International Publishing. ISBN: 9783319242774.

Zhao, Y.-J., Hosoya, T., Baral, H.-O., Hosaka, K., and Kakishima, M. (2013). *Hymenoscyphus pseudoalbidus*, the correct name for *Lambertella albid* reported from Japan. *Mycotaxon* 122, 25–41. doi: 10.5248/122.25

Supplements

Figures

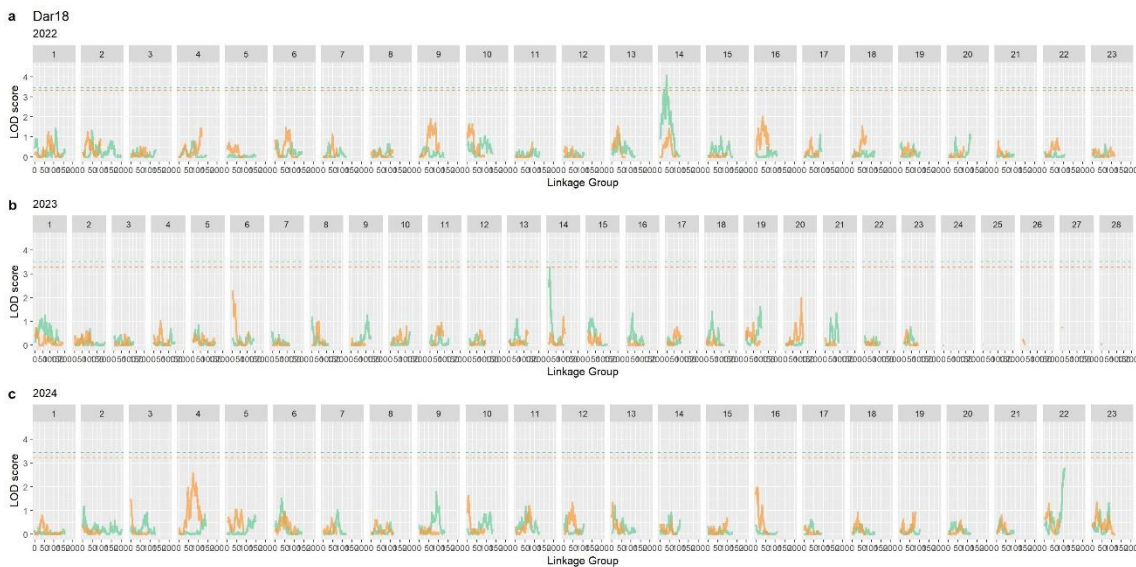


Figure S1: Quantitative Trait Loci (QTL) mapping results for Dar18 across three years (2022-2024).). Based on the genetic map and LOD threshold 5%. Each panel represents a separate chromosome, numbered 1 to 23, with the x-axis indicating the position on the chromosome (cM) and the y-axis showing the LOD (logarithm of the odds) score. Panel a: year 2022, Panel b: year 2023, Panel c: year 2024. The LOD score is plotted for each marker along the chromosomes in each panel, with significant peaks indicating potential QTLs. The horizontal dashed lines represent the LOD threshold for significance. Green lines indicate paternal significance and orange lines represent maternal significance.

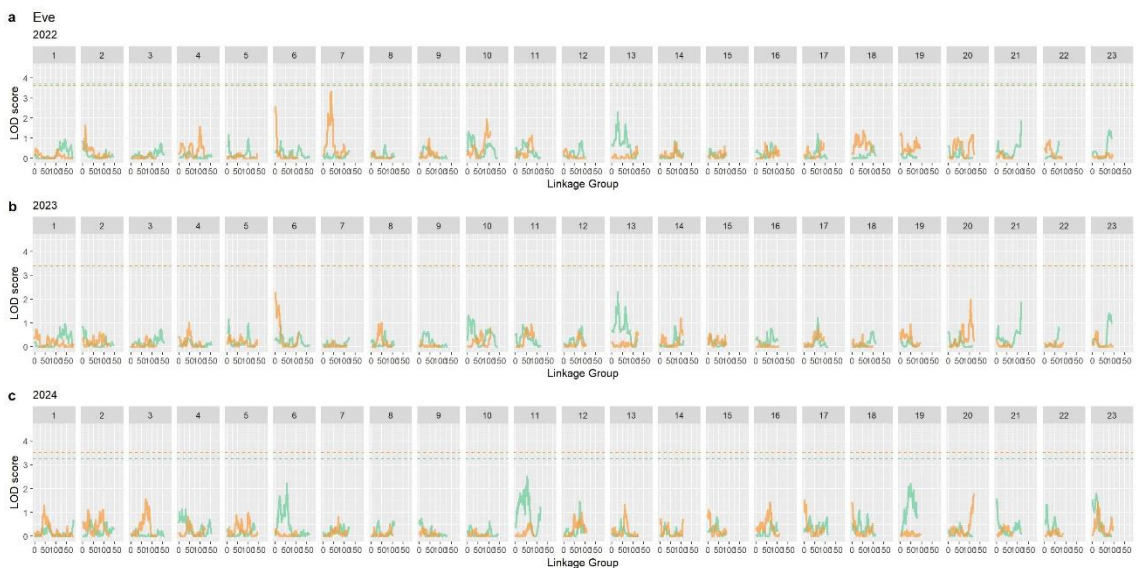


Figure S2: Quantitative Trait Loci (QTL) mapping results for Eve2 across three years (2022-2024).). Based on the genetic map and LOD threshold 5%. Each panel represents a separate chromosome, numbered 1 to 23, with the x-axis indicating the position on the chromosome (cM) and the y-axis showing the LOD (logarithm of the odds) score. Panel a: year 2022, Panel b: year 2023, Panel c: year 2024. The LOD score is plotted for each marker along the chromosomes in each panel, with significant peaks indicating potential QTLs. The horizontal dashed lines represent the LOD threshold for significance. Green lines indicate paternal significance and orange lines represent maternal significance.

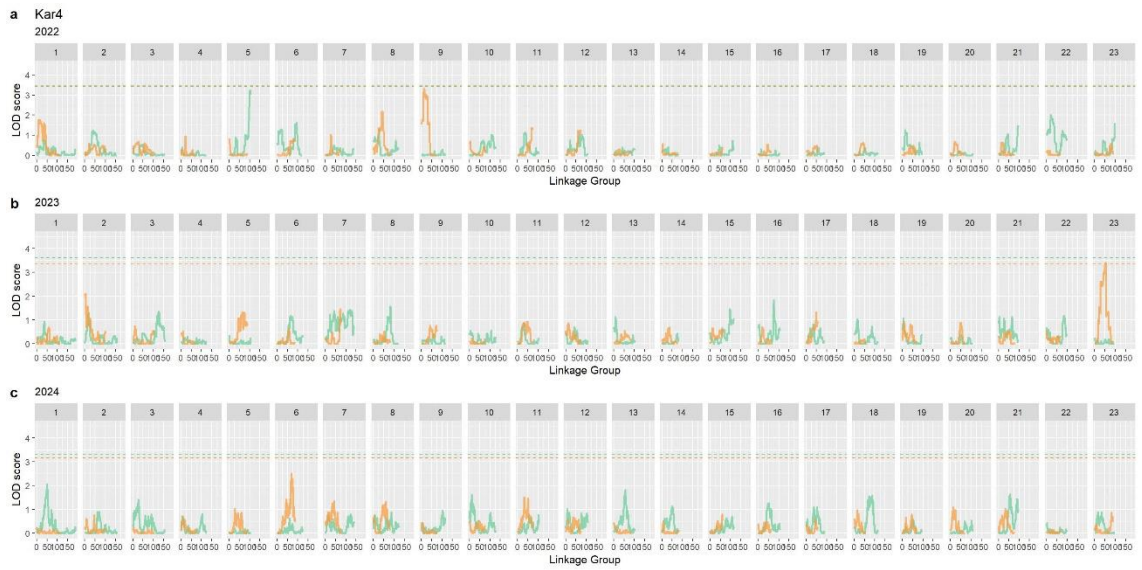


Figure S3: Quantitative Trait Loci (QTL) mapping results for Kar4 across three years (2022-2024).). Based on the genetic map and LOD threshold 5%. Each panel represents a separate chromosome, numbered 1 to 23, with the x-axis indicating the position on the chromosome (cM) and the y-axis showing the LOD (logarithm of the odds) score. Panel a: year 2022, Panel b: year 2023, Panel c: year 2024. The LOD score is plotted for each marker along the chromosomes in each panel, with significant peaks indicating potential QTLs. The horizontal dashed lines represent the LOD threshold for significance. Green lines indicate paternal significance and orange lines represent maternal significance.

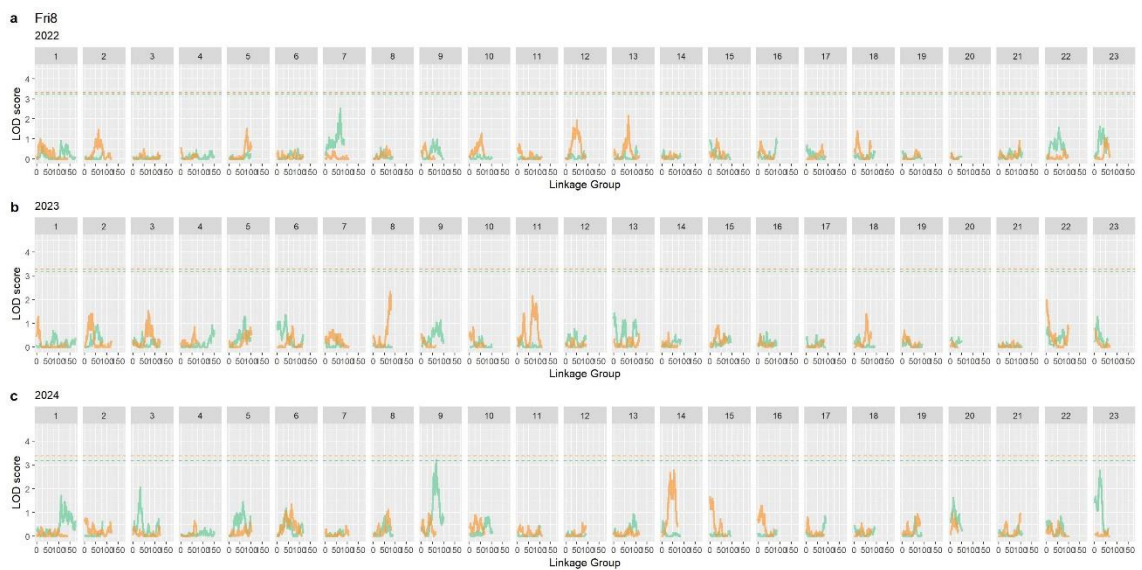


Figure S4: Quantitative Trait Loci (QTL) mapping results for Fru8 across three years (2022-2024).). Based on the genetic map and LOD threshold 5%. Each panel represents a separate chromosome, numbered 1 to 23, with the x-axis indicating the position on the chromosome (cM) and the y-axis showing the LOD (logarithm of the odds) score. Panel a: year 2022, Panel b: year 2023, Panel c: year 2024. The LOD score is plotted for each marker along the chromosomes in each panel, with significant peaks indicating potential QTLs. The horizontal dashed lines represent the LOD threshold for significance. Green lines indicate paternal significance and orange lines represent maternal significance.

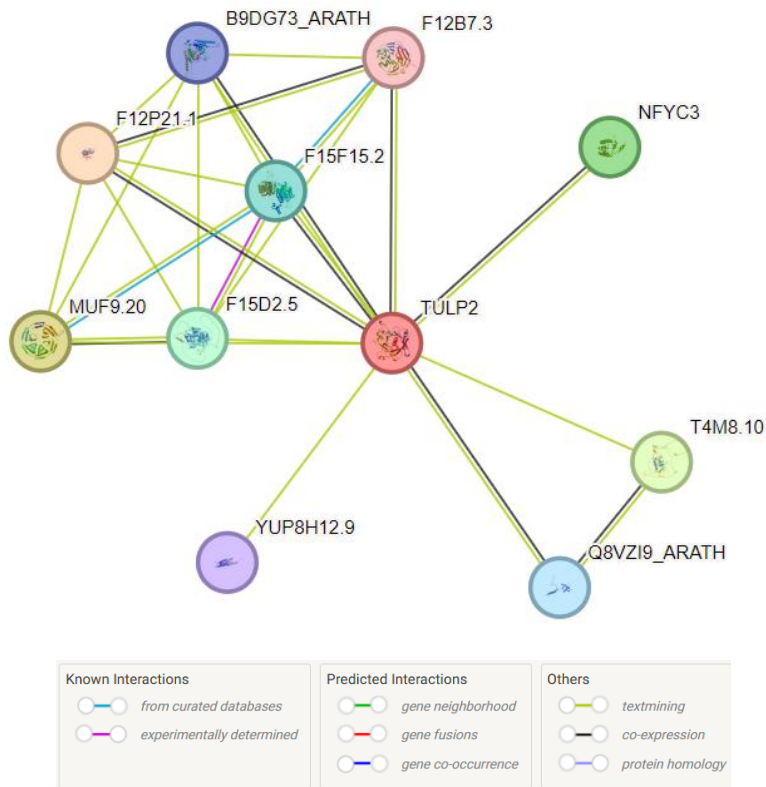


Figure S5: Functional interaction network of *A. thaliana* TLP2 (AT2G18280; TULP2 in the network) created with STRING.

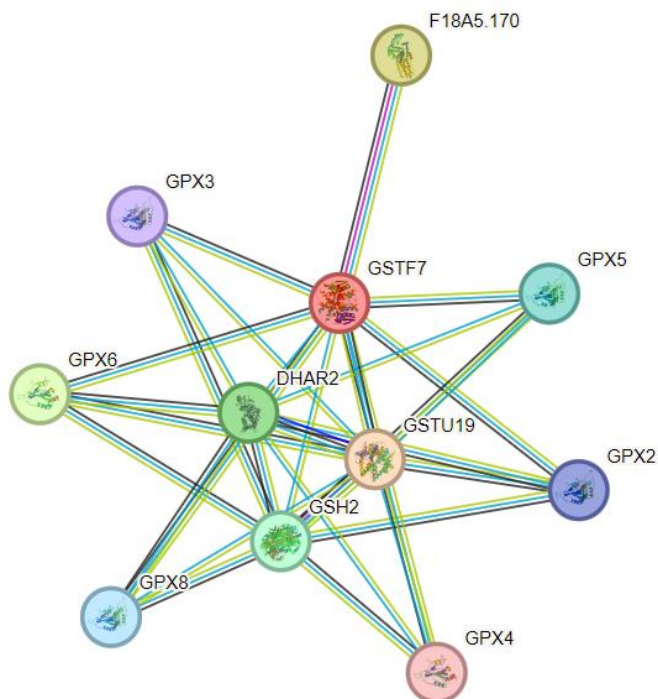


Figure S6: Functional interaction network of *A. thaliana* GST11 (AT1G02920, GSTF7 in the network) created with STRING.

Tables

Table S1: Selected results from BlastP analysis of protein sequences predicted in the QTL-including region of *Fraxinus excelsior* chromosome 16 (24,000,000-25,000,000 bp) versus *Arabidopsis thaliana* protein sequences (only those 69 out of 140 putative *Fraxinus excelsior* genes are presented which provided hits with e-values below e-30).

Query	Number of HSPs	Greatest identity %	Lowest E-value	Accession (E-value)
unnamed-1:g38.t1	52	58,62069	0	AT3G51120.1 Symbols: no_symbol_ available no_full_name_ available
unnamed-1:g38.t2	51	58,62069	0	AT3G51120.1 Symbols: no_symbol_ available no_full_name_ available
unnamed-1:g46.t1	1083	50	0	AT3G49740.1 Symbols: no_symbol_ available no_full_name_ available
unnamed-1:g54.t1	12	82,39203	0	AT2G22780.1 Symbols: PMDH1 peroxisomal_NAD-malate_dehydrogenase_1
unnamed-1:g65.t1	10	79,55882	0	AT5G10540.1 Symbols: TOP2 thimet_metalloendopeptidase_2
unnamed-1:g81.t1	7	80,84507	0	AT5G65620.2 Symbols: OOP, TOP1 thimet_metalloendopeptidase_1, organellar_oligopeptidase
unnamed-1:g90.t1	229	82,46073	0	AT4G36690.1 Symbols: ATU2AF65A
unnamed-1:g99.t1	42	72,88889	0	AT1G21480.1 Symbols: no_symbol_ available no_full_name_ available
unnamed-1:g99.t2	57	72,30392	0	AT1G21480.1 Symbols: no_symbol_ available no_full_name_ available
unnamed-1:g112.t1	28	54,43038	0	AT1G78280.1 Symbols: no_symbol_ available no_full_name_ available
unnamed-1:g112.t2	27	54,42177	0	AT1G78280.1 Symbols: no_symbol_ available no_full_name_ available
unnamed-1:g131.t1	10	84,04908	0	AT3G16150.1 Symbols: ASPGB1 asparaginase_B1
unnamed-1:g138.t1	132	60,2459	0	AT1G47056.1 Symbols: VFB1 VIER_F-box_proteine_1
unnamed-1:g55.t1	707	72,23975	6E-177	AT3G49660.1 Symbols: AtWDR5a, WDR5a human_WDR5 (WD40_repeat)_homolog_a
unnamed-1:g107.t1	166	44,54545	6E-176	AT5G49665.1 Symbols: WAV3
unnamed-1:g51.t1	40	75,34247	2,6E-170	AT4G37800.1 Symbols: XTH7 xyloglucan_endotransglucosylase/hydrolase_7
unnamed-1:g126.t1	310	74,13793	7,2E-162	AT4G24480.1 Symbols: no_symbol_ available no_full_name_ available
unnamed-1:g83.t1	23	71,60883	2,6E-160	AT1G67340.1 Symbols: no_symbol_ available no_full_name_ available
unnamed-1:g63.t1	30	55,55556	7,1E-138	AT4G37880.1 Symbols: no_symbol_ available no_full_name_ available
unnamed-1:g106.t1	3	62,23404	9,8E-134	AT4G38630.1 Symbols: MBP1, RPN10, ATMCB1, MCB1 MULTIUBIQUITIN_CHAIN_BINDING_PROTEIN_1, regulatory_particle_non-ATPase_10, MULTIUBIQUITIN_CHAIN-BINDING_PROTEIN_1
unnamed-1:g56.t2	17	51,93237	2E-132	AT5G65685.5 Symbols: SS5 STARCH_SYNTHASE_5
unnamed-1:g100.t1	67	64,11765	2,1E-132	AT1G77110.1 Symbols: PIN6 PIN-FORMED_6
unnamed-1:g34.t1	28	51,02041	3,2E-130	AT2G18280.1 Symbols: AtTLP2, TLP2 tubby_like_protein_2

unnamed-1:g4.t1	10	62,06897	2,8E-127	AT3G02570.1 Symbols: MEE31, PMI1 MATERNAL_EFFECT_EMBRYO_ARREST_31, PHOSPHOMANNOSE_ISOMERASE_1
unnamed-1:g62.t1	6	80,28169	1,2E-124	AT3G49640.1 Symbols: no_symbol_available no_full_name_available
unnamed-1:g29.t1	43	65,86207	8,7E-124	AT2G18360.1 Symbols: no_symbol_available no_full_name_available
unnamed-1:g88.t1	28	37,80718	3,6E-119	AT4G38180.1 Symbols: FRS5 FAR1-related_sequence_5
unnamed-1:g88.t2	29	37,80718	4,1E-119	AT4G38180.1 Symbols: FRS5 FAR1-related_sequence_5
unnamed-1:g73.t1	98	71,8254	1,7E-115	AT3G50670.1 Symbols: U1SNRNP, U1-70K U1_small_nuclear_ribonucleoprotein-70K
unnamed-1:g3.t1	7	78,01047	7E-113	AT3G02560.1 Symbols: no_symbol_available no_full_name_available
unnamed-1:g32.t2	35	76,92308	5,7E-100	AT3G54280.2 Symbols: RGD3, CHA16, BTAF1, ATBTAF1, CHR16 ROOT_GROWTH_DEFECTIVE_3
unnamed-1:g32.t1	32	76,92308	4,59E-97	AT3G54280.2 Symbols: RGD3, CHA16, BTAF1, ATBTAF1, CHR16 ROOT_GROWTH_DEFECTIVE_3
unnamed-1:g137.t1	41	53,89831	8,77E-94	AT4G34160.1 Symbols: CYCD3, CYCD3;1 CYCLIN_D3;1
unnamed-1:g31.t1	21	57,62712	4,33E-91	AT2G18360.1 Symbols: no_symbol_available no_full_name_available
unnamed-1:g133.t1	5	42,22222	2,99E-80	AT4G23160.2 Symbols: CRK8 cysteine-rich_RLK_(RECEPTOR-like_protein_kinase)_8
unnamed-1:g101.t1	51	54,80769	8,3E-78	AT1G02920.1 Symbols: ATGSTF8, GSTF7, ATGSTF7, ATGST11, GST11 GLUTATHIONE_S-TRANSFERASE_11, glutathione_S-transferase_7, ARABIDOPSIS_Glutathione_S-transferase_11
unnamed-1:g50.t2	22	53,1746	1,21E-73	AT4G10200.1 Symbols: no_symbol_available no_full_name_available
unnamed-1:g50.t1	80	53,1746	6,29E-73	AT4G10200.1 Symbols: no_symbol_available no_full_name_available
unnamed-1:g48.t1	161	85,49618	2,34E-71	AT5G65790.1 Symbols: ATMYB68, MYB68 MYB_DOMAIN_PROTEIN_68, myb_domain_protein_68
unnamed-1:g128.t1	186	86,48649	4,37E-70	AT4G37260.1 Symbols: ATMYB73, MYB73 myb_domain_protein_73
unnamed-1:g36.t1	15	46,15385	5,12E-62	AT1G06740.1 Symbols: MUG3 MUGSTANG_3
unnamed-1:g76.t1	10	61,40351	1,71E-61	AT4G36750.1 Symbols: no_symbol_available no_full_name_available
unnamed-1:g89.t1	10	61,40351	2,34E-61	AT4G36750.1 Symbols: no_symbol_available no_full_name_available
unnamed-1:g56.t1	9	52,17391	4,05E-58	AT5G65685.5 Symbols: SS5 STARCK_H_SYNTHASE_5
unnamed-1:g42.t1	5	45,69733	5,43E-57	AT5G66230.2 Symbols: no_symbol_available no_full_name_available
unnamed-1:g40.t1	6	89,53488	1,08E-52	AT2G16460.1 Symbols: no_symbol_available no_full_name_available
unnamed-1:g103.t1	26	60	4,83E-52	AT1G19260.1 Symbols: LOH3
unnamed-1:g70.t1	75	64,28571	2,56E-51	AT1G71050.1 Symbols: AtHMP15, HIPP20 HEAVY_METAL_ASSOCIATE

unnamed-1:g127.t1	8	43,65079	2,03E-48	D_PROTEIN_15, heavy metal associate d_isoprenylated_plant_protein_20_ AT4G23160.2_ Symbols: CRK8_ cyste ine-rich_RLK_(RECEPTOR- like_protein_kinase)_8_
unnamed-1:g105.t1	24	37,9562	2,51E-45	AT3G05850.1_ Symbols: MUG7_ MU STANG_7_
unnamed-1:g33.t1	26	71,11111	4,77E-45	AT4G36610.1_ Symbols: no_symbol_av ailable_ no_full_name_available_
unnamed-1:g75.t1	7	86,48649	1,64E-43	AT4G36750.1_ Symbols: no_symbol_av ailable_ no_full_name_available_
unnamed-1:g86.t1	7	86,48649	4,37E-43	AT4G36750.1_ Symbols: no_symbol_av ailable_ no_full_name_available_
unnamed-1:g35.t1	17	77,77778	1,67E-42	AT4G36520.1_ Symbols: AUXILIN- LIKE4_ AUXILIN-LIKE4_
unnamed-1:g39.t1	9	75,5102	1,41E-41	AT2G16460.1_ Symbols: no_symbol_av ailable_ no_full_name_available_
unnamed-1:g44.t1	12	47,84946	2,62E-41	AT5G66240.2_ Symbols: S2Lb, ULCS1 _ Ubiquitin_Ligase_Complex_Subunit_1, _SWD2-LIKE-b_
unnamed-1:g62.t2	8	74,4186	3,02E-41	AT3G49640.1_ Symbols: no_symbol_av ailable_ no_full_name_available_
unnamed-1:g95.t1	21	34,09091	2,03E-39	AT3G05850.1_ Symbols: MUG7_ MU STANG_7_
unnamed-1:g57.t1	17	43,75	2,09E-39	AT5G28780.1_ Symbols: no_symbol_av ailable_ no_full_name_available_
unnamed-1:g110.t1	38	82,05128	2,22E-39	AT3G54220.1_ Symbols: SGR1, SCR_ _SCARECROW, SHOOT_GRAVITROP ISM_1_
unnamed-1:g64.t1	18	64,58333	1,02E-38	AT5G65660.1_ Symbols: no_symbol_av ailable_ no_full_name_available_
unnamed-1:g28.t1	32	75,75758	2,42E-37	AT3G50870.1_ Symbols: GATA18, M NP, AtGATA18, HAN_ GATA_TRAN SCRIPTION_FACTOR_18, MONOPOL E, HANABA_TANARU_
unnamed-1:g71.t1	5	71,42857	7,26E-36	AT4G35980.1_ Symbols: no_symbol_av ailable_ no_full_name_available_
unnamed-1:g22.t1	37	79,48718	7,76E-36	AT5G48600.2_ Symbols: ATSMC4, S MC3, SMC4, ATCAP- C, ATSMC3_ ARABIDOPSIS_THALI ANA_STRUCTURAL_MAINTENANCE _OF_CHROMOSOME_4, structural_mai ntenance_of_chromosome_3, ARABIDO PSIS_THALIANA_CHROMOSOME_AS SOCIATED_PROTEIN-C_
unnamed-1:g124.t1	5	57	9,53E-36	AT4G23160.2_ Symbols: CRK8_ cyste ine-rich_RLK_(RECEPTOR- like_protein_kinase)_8_
unnamed-1:g22.t2	23	79,48718	1,72E-34	AT5G48600.2_ Symbols: ATSMC4, S MC3, SMC4, ATCAP- C, ATSMC3_ ARABIDOPSIS_THALI ANA_STRUCTURAL_MAINTENANCE _OF_CHROMOSOME_4, structural_mai ntenance_of_chromosome_3, ARABIDO PSIS_THALIANA_CHROMOSOME_AS SOCIATED_PROTEIN-C_
unnamed-1:g113.t1	4	51,63043	3,19E-34	AT5G65520.1_ Symbols: no_symbol_av ailable_ no_full_name_available_
unnamed-1:g37.t1	6	50,33557	3,41E-34	AT4G36520.1_ Symbols: AUXILIN- LIKE4_ AUXILIN-LIKE4_

Tabelle S2: Nonredundant list of 51 *Arabidopsis thaliana* gene identifiers from Table S1 used as input for GO-annotation by STRING

AT1G02920	AT1G06740	AT1G21480	AT1G19260	AT1G47056
AT1G67340	AT1G71050	AT1G77110	AT1G78280	AT2G01340
AT2G16460	AT2G18280	AT2G18360	AT2G22780	AT3G02560
AT3G02570	AT3G05850	AT3G16150	AT3G49640	AT3G49660
AT3G49740	AT3G50670	AT3G50870	AT3G51120	AT3G54220
AT3G54280	AT4G10200	AT4G23160	AT4G24480	AT4G34160
AT4G35980	AT4G36520	AT4G36610	AT4G36690	AT4G36750
AT4G37260	AT4G37800	AT4G37880	AT4G38180	AT4G38630
AT5G10540	AT5G28780	AT5G48600	AT5G49665	AT5G65520
AT5G65620	AT5G65660	AT5G65685	AT5G65790	AT5G66230
AT5G66240				

Tabelle S3: Predicted protein sequences of three *Fraxinus excelsior* genes annotated at/in the QTL locus which are assigned to Gene Ontology (GO) terms “response to fungus” (GO:0009620) and/or “defense response to fungus” (GO:0050832) according to STRING analysis of the related *Arabidopsis thaliana* gene identifier (Table 1).

Gene	Protein Sequence
g34	MSFKGIICDLKEIKDGRRSTSKREGIEGKHWLNKNRSHIAPDVAPSKPIQQGHWANLP QELLVDIIRRVEESETSWPARTVVVFCASVCKSWRDITKEIVKTPEECGRLTFPISLKQP GPRDAPIQCFIKRDISTSTYCLSNFLGTFKFTTYDCQPLNDAAVQHNSRSSRRFHTSQVS PRLHAYNNSVATISYELNVFRPRGPRRMNCFMNSIPVSSIQEGGTAPTPISEFPQCSDEKF SPPSVSEGKDSVINFSSRSLSSSSFSRPLSGEPLVLKNKAPRWHEQLQCWCLNFKGRVT VTSVKNFQLVAAVDPSLNISTAEQEKVILQFGKIGKDIFTMDYHYPLTAFQAFSICLSSF DSKPACE
g69	MGNNIGGRNKAKIMKINGEVFKVRLPAITQDVLKHFPGHVLEPEAVKKYGIRAEPL PEEELKAKKIYFLVELPKLREEKTPRRRSRAVHMDAKEWIALRRRSISEPLGAGSGSVR VKMRLPKAEIEKLIIESKDEKEMAEKIIDLCLLNSSKKIRE
g101	MAIKVHGHPLSTSTCRVLATLAAKELDYEFILVDLATDQQKSQAFLSLNPFQVPPFE DGDCLKLFESRAITQYLAHTYADKGTPLITRDPKKMAIISQWTEIEAHCNPPASKLAYE LVIKPISETREESVVSQLEARLAEVLDIYELHLAQSKYLGGSFTLADLHHLAVINV LMGTCLKALFDARPHVSAWCADILARPAWKKIVAATKY

Chapter 4

High-molecular-weight DNA extraction for broadleaved and conifer tree species

Melina Krautwurst¹, Annika Eikhof¹, Sylke Winkler², Daniel Bross¹, Birgit Kersten¹, Niels A. Müller¹

¹ Thünen Institute of Forest Genetics, 22927 Grosshansdorf, Germany

² Max-Planck Institute of Molecular Cell Biology and Genetics, 01307, Dresden, Germany

Correspondence to: niels.mueller@thuenen.de, birgit.kersten@thuenen.de

Published in *Silvae Genetica* (2024)

Doi: 10.2478/sg-2024-0009



High-molecular-weight DNA extraction for broadleaved and conifer tree species

Melina Krautwurst¹, Annika Eikhof¹, Sylke Winkler², Daniel Bross¹, Birgit Kersten^{1*}, Niels A. Müller^{1*}

Abstract

Plant genome sequencing based on long reads has increasingly been applied also to tree species in recent years. A crucial step in these genome projects is the successful extraction of high-molecular-weight DNA in high quality and sufficient quantity, which is imperative for long-read sequencing. The extraction of high-molecular-weight DNA in trees has been limited and difficult to conduct. To achieve successful extraction, modification of the protocol for tree species is necessary. Here, we present a protocol for high-molecular-weight DNA extraction from the broadleaved tree species *Fraxinus excelsior* and from the conifer species *Taxus baccata*. The protocol is based on nuclei isolation and is divided into two main steps, i.e. nuclei separation based on the nuclei isolation buffer and DNA extraction using the Nanobind® plant nuclei kit by Pacific Biosciences (PacBio). The protocol can be applied to different tree species to obtain high-molecular-weight DNA in high quality, which can be used for Oxford Nanopore Technologies (ONT) and PacBio sequencing. ONT sequencing using four high-molecular-weight DNA preparations from *Fraxinus excelsior* resulted in different read length N50 values (12.91-24.71 kb) and total base output (5.81-23.17 Gb), emphasizing the complex nature of the sequencing pipeline of high-molecular-weight DNA. PacBio HiFi sequencing of *Taxus baccata* produced circular consensus sequencing reads with an average length of 11.9 kb and an N50 of 13.59 kb. Altogether, this study presents a protocol to obtain high-molecular-weight DNA from different challenging tree species and discusses several important points that can be considered when adapting the protocol to additional species.

Introduction

The past 20 years have seen increasingly rapid advances in the genomics field, especially through the impressive development of high-throughput DNA sequencing technologies, ranging from short-read (e.g., Illumina) to long-read sequencing technologies. Advanced long-read sequencing technologies include single-molecule real-time sequencing developed by Pacific Biosciences (PacBio) and (ultra-)long-read nanopore sequencing developed by Oxford Nanopore Technologies (ONT). Long-read sequencing is fundamental to modern genome

studies (Pucker et al., 2022). The according technologies require high-molecular-weight (HMW) DNA of excellent quality and usually relatively high quantity. Each long-read sequencing platform has different requirements for DNA extraction. While some studies advocate avoiding guanidine lysis buffer, phenol, and chloroform purification for PacBio high-fidelity (HiFi) sequencing (PacBio, 2022b; Nishii et al., 2023), others have reported successful genome assemblies using DNA extracted with these methods (Grohme et al., 2018; Jebb et al., 2020). The rationale behind these divergent practices may stem from considerations such as potential toxicity associated with specific extraction components. For ONT, the purity of the DNA can influence pore activity and sequence quantity (Fleck et al., 2022; La Cerda et al., 2023; Oxford Nanopore Technologies, 2023).

Understanding the nuances of DNA extraction protocols is crucial for optimising sequencing outcomes and elucidating their impact on downstream analyses. Over the years, multiple protocols have been published for extracting HMW DNA from plants, using different approaches mostly optimized for only one sequencing platform (Li et al., 2020; Aggerwal et al., 2022; Schenk et al., 2023). The approaches to extract HMW DNA range from nuclei isolation followed by DNA extraction to direct precipitation of DNA with Cetyltrimethylammonium bromide (CTAB)-based methods or a combination of both. Establishing one protocol that fits all plant species may be difficult because of the diversity of cellular components in different species and tissues. Trees, for example, are rich in polyphenolic compounds, polysaccharides, tannins, and other secondary metabolites (Couch and Fritz, 1990; Formato et al., 2022). The combination and concentration of these compounds vary significantly among tree species (Witzell and Martín, 2008; Huang et al., 2023). These compounds affect DNA in different ways during extraction (Healey et al., 2014). Oxidized and co-precipitated DNA and/or nucleases and physical shearing can result from insufficient removal of these compounds. To regulate the negative impact of such compounds individually, the adaptation of the following steps/components in the HMW DNA extraction protocol is essential: the grinding process, washing buffer, centrifugation force and lysis. This indicates the need for establishing protocols that are suitable for recalcitrant tree species and applicable for ONT and PacBio HiFi sequencing.

The protocol developed here – using broadleaved and conifer tree species - seeks to obtain HMW DNA (after nuclei isolation) for multiple purposes and is feasible to be applied to different tree species. This protocol is based on HMW DNA preparation trials using two tree species with different genome sizes and secondary metabolite compositions: common ash (*Fraxinus excelsior* L.), and European yew (*Taxus baccata* L.). The prepared HMW DNA of the deciduous broadleaved tree species *F. excelsior* was sequenced with ONT, and that of the conifer species *T. baccata* was sequenced with PacBio HiFi. The methodological approach was separated into two main steps: (1) cell wall disruption and nuclei release. Removing debris, nuclei separation with an osmotic nuclei isolation buffer (NIB) and nuclei isolation; (2) lysis of the nuclear membrane and extraction of DNA using the Nanobind® plant nuclei kit by PacBio (Figure 1). For the presented protocol, the average cost per sample is about 40 €, and compared to other protocols of HMW DNA extraction and sequencing preparation (Zerpa-Catanho et al., 2021), it is relatively time-efficient (1-2 days).

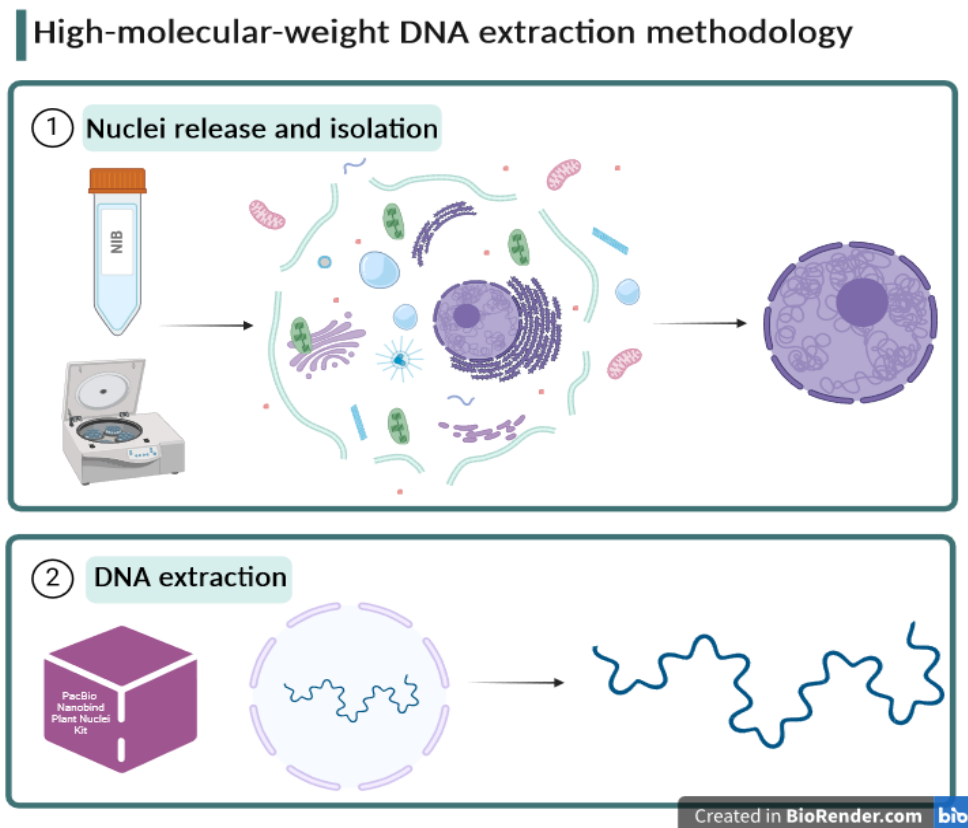


Figure 1: Schematic overview of the high-molecular-weight DNA extraction. (1) Cell wall disruption breaking of cell organelles, and nuclei separation with osmotic nuclei isolation buffer (NIB) and centrifugation. (2) Lysis of nuclear membrane and extraction of DNA using Nanobind® plant nuclei kit (PacBio).

Materials and methods

Biological samples

As starting material, 1-2 grams of plant tissue, stored at -80 °C or fresh, were used. The amount of plant tissue required depends on the tree species and the sample. Two grams of *F. excelsior* leaf material resulted in the highest HMW DNA quantity. For *T. baccata*, 1-2 grams of needles were used as starting material.

Reagents, equipment and supplies for HMW DNA extraction

Table 1-4 list the reagents, equipment, and supplies essential for conducting the HMW extraction.

Table 3: List of reagents.

Reagent	Source
Trizma® hydrochloride solution	Sigma-Aldrich
Potassium chloride (KCL)	Sigma-Aldrich
Ethylenediaminetetraacetic acid (EDTA) 0.5 M pH 8.0	AppliChem
Spermidine trihydrochloride	Sigma-Aldrich
Spermine tetrahydrochloride	Sigma-Aldrich
Sucrose	Duchefa Biochemie
Triton X-100	Sigma-Aldrich
Polyvinylpyrrolidone (MW ~360 kDa, PVP360)	Sigma-Aldrich
Polyvinylpyrrolidone 1% (MW ~ 40 kDa, PVP40)	Sigma-Aldrich
Tris	Roth
Sorbitol	Sigma-Aldrich
β-mercaptoethanol, 14 M	Merck
Nuclease-free water	Invitrogen
Water for HPLC (filtered 0.1 μM)	ChemSolute
Ethanol (96-100%)	Roth
Isopropanol (100 %)	Roth
Qubit 1x dsDNA BR Assay kit	Thermo Fisher
Nanobind® plant nuclei kit	PacBio
Short Read Eliminator (SRE <25 kb) Kit	PacBio

Table 4: Equipment and other supplies.

Reagent	Source
Ice	X
Liquid nitrogen (LN2)	X
DNA LoBind tubes, 1.5 ml	Eppendorf
Wide bore 200 – 1000 μl pipette tips	Axygen
Micro-centrifuge, Fresco 21	Thermo Fisher
Mini-centrifuge	Sprout
Centrifuge, Mega Star 600	VWR
Steriflip 20 μm pore size, 50 mL process volume	Millipore

50 mL conical tubes	Sarstedt
Tube rotator, shaker, platform rocker or tumbling rocker	Biometra
Vortex shaker	neoLab
Thermo-Shaker	BioSan
Hula mixer or multi rotator	biolab
Magnetic tube rack, DynaMag®-2	Invitrogen
Nanodrop 1000 spectrophotometer	Peqlab
Fluorescent DNA quantification, Qubit™3 fluorometer	Invitrogen
Small nylon or synthetic paintbrush	X

Table 5: Equipment for using a pestle and mortar.

Reagent	Source
Pestle and mortar	X
Ice	X
Glass beakers, of different sizes	X
Magnetic stir bar	X
Stainless steel laboratory spoon	Thermo Fisher

Table 6: Equipment for using TissueRuptor.

Reagent	Source
TissueRuptor II	Qiagen
TissueRuptor II disposable Probes	Qiagen

Buffer preparation for HMW DNA extraction

For HMW DNA extraction, the five buffers described below must be prepared in advance: 10x HB (Homogenization Buffer Stock), 1x HB (Homogenization Buffer), TSB (Triton-Sucrose-Buffer), Nuclei Isolation Buffer (NIB), and Sorbitol wash buffer. 10x HB, 1x HB and TSB were used as stock solutions and stored at 4 °C in a glass stock bottle. The NIB should not be older than one day. Owing to the aggregation of PVP360, a longer stirring time should be planned. Sorbitol wash buffer should be made fresh before every extraction. The amount of sorbitol wash buffer required depended on the efficiency of the washing steps. Some tree species seem to require fewer washing steps than others.

10x HB (Homogenization Buffer Stock), 500 ml

Reagent	Amount	Final concentration
Trizma	6.06 g	0.1 M
KCl	29.8 g	0.8 M
EDTA (0,5 M)	100 ml	0.1 M
Spermidine	1.28 g	17 mM
Spermine	1.74 g	17 mM

Add the reagents to a clean beaker and stir until dissolved; fill up to 500 ml with pure water in a glass stock bottle and set to pH 9 with 10 M NaOH (storage up to one year at 4 °C).

1x HB (Homogenization Buffer), 1 l

Reagent	Amount
Buffer 10x HB	100 ml
Sucrose	171.2 g

Add the reagents to a clean beaker, stir until dissolved, and fill up to 1 l with pure water in a glass stock bottle (stored for up to three months at 4 °C).

TSB-Buffer (Triton-Sucrose Buffer), 100 ml

Reagent	Amount
Triton X-100	20 ml
Buffer 10x HB	10 ml
Sucrose	17.1 g

Add the reagents to a clean beaker, stir until dissolved, and fill to 100 ml with pure water in a glass stock bottle (stored for up to one year at 4 °C).

NIB-Buffer (Nuclei Isolation Buffer)

Reagent	Amount (50ml)	Amount (100ml)
Buffer 1x HB	48.75 ml	97.5 ml
Buffer TSB	1.25 ml	2.5 ml
PVP360	0.5 g	1 g

Add the reagents to a clean beaker and stir until dissolved (storage at 4 °C).

Sorbitol wash buffer

Reagent	Stock	10 ml	20 ml
100 mM Tris	1 M	1 ml	2 ml
0.35 M Sorbitol	-	0.64 g	1.28 g
5 mM EDTA	0.5 M	100 µl	200 µl
1% PVP40	-	0.1 g	0.2 g
1% β-mercaptoethanol (just before use)	-	100 µl	200 µl

Add the reagents to a clean beaker and stir until dissolved.

! CAUTION: β-mercaptoethanol is toxic, only use under fume hood.

Procedure for nuclei isolation and HMW DNA extraction

The protocol for nuclei isolation is based on a related protocol used for nuclei isolation from *Abies alba* needles for HMW DNA preparation and PacBio sequencing (Kersten et al., 2022), which was further modified and adapted to the tree species analyzed in our study. The protocol for HMW DNA extraction represents an adaptation of the following protocol: 'Extracting HMW DNA from plant nuclei using Nanobind® kits' (PacBio, 2022a), where only a few steps have been adapted. The protocol, including nuclei isolation and DNA extraction, takes 2-3 days. Isolation and extraction can be conducted on the same day or two separate

days. If the protocol is performed in two days, the extracted nuclei pellet can be stored at -80 °C. After the extraction, the DNA must be incubated overnight at room temperature (12 h or longer). A detailed visualisation of the individual steps of the nuclei isolation and DNA extraction is shown in Figure 2.

High-molecular-weight gDNA extraction

Step by Step guide

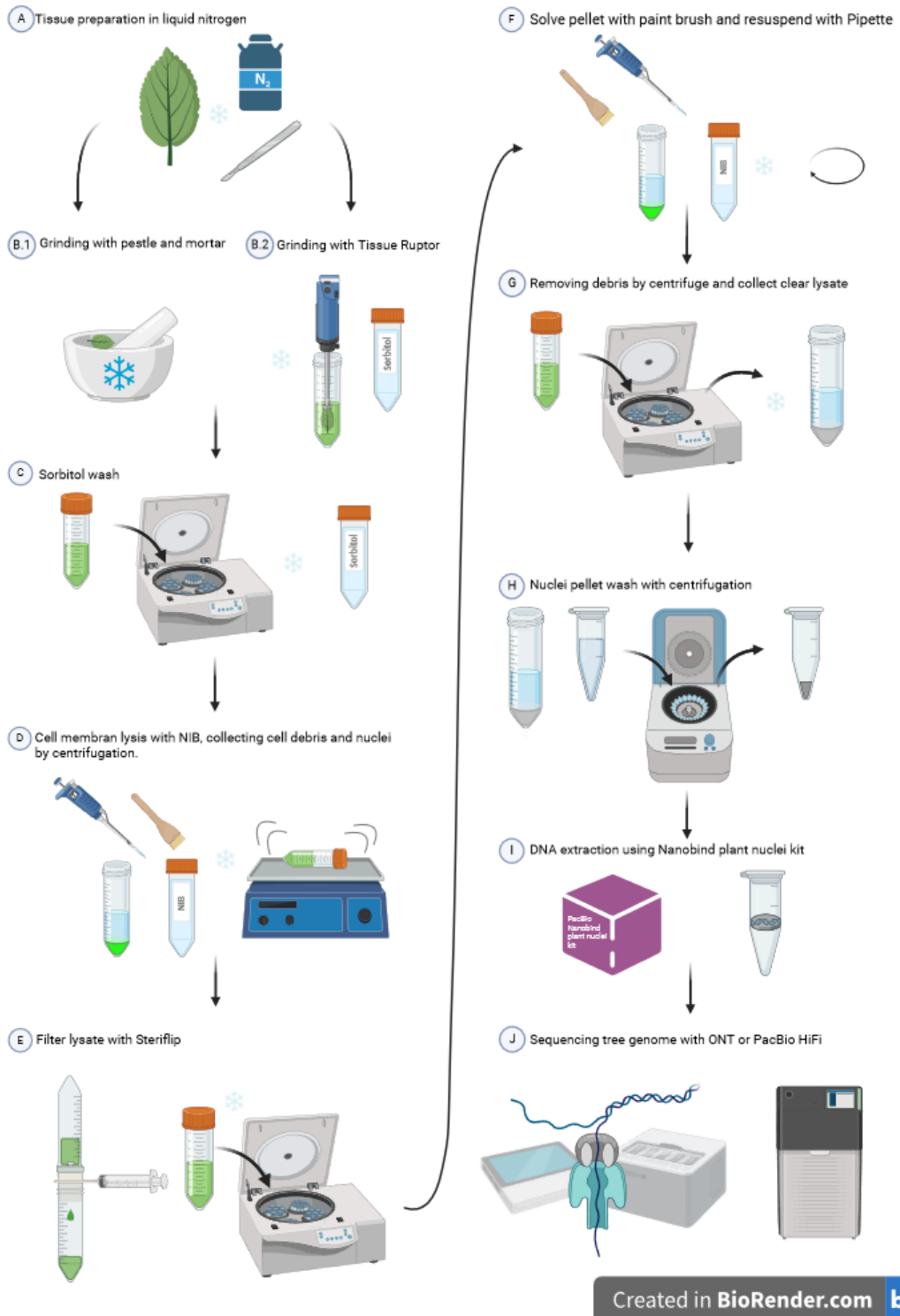





Figure 2: Illustration of the steps involved in the high-molecular-weight DNA extraction based on nuclei isolation and extraction.




Nuclei isolation

 Total duration 3 – 4 h

1. Place three tubes of 50 ml on ice. Chill the centrifuge for conical tubes to 4 °C.
2. Start with preparing the NIB and sorbitol wash buffer.
 -  Owing to the aggregation of the NIB, it is best to prepare it one day in advance.
3. Place NIB and sorbitol wash buffer without β -mercaptoethanol in precooled 50 ml tubes on ice.

Nuclei Isolation with pestle and mortar

 **30 - 40 min**


- a. Precool pestle, mortar, and laboratory spoon on ice.
- b. If fresh leaves or needles are collected, cut material with a sterile scalpel into pieces smaller than $\sim 0.5 \text{ cm}^2$ and shock freeze in liquid nitrogen.
 -  The cutting process should be done quickly.
- c. Before transferring the plant material into the mortar, add liquid nitrogen into the mortar to additionally lower the temperature of the equipment. Dip the pestle and laboratory spoon into liquid nitrogen that was added to the mortar. Everything needs to be as cold as possible.
 -  Thawing the plant material during the grinding process can strongly influence the extraction results. If the plant material starts sticking to the equipment it is not cold enough.
- d. When all liquid nitrogen is evaporated in the mortar, start adding the plant material.
- e. Grind with pestle for at least 20-30 min, and add liquid nitrogen in between to cool plant material and equipment.
 -  Add the liquid nitrogen slowly, adding it too fast can lead to overflow of the plant powder. To ensure that the mortar is kept cold, it can be placed into a box of ice during the grinding process.
- f. After 20 min the material should be a fine powder. Transfer the powder to a glass beaker and add a mix stir bar. Add 10 ml sorbitol wash buffer (without β -

mercaptoethanol). Transfer the glass beaker to a container or petri dish with ice.

Mix on magnetic stirrer for 10 min.

- g. Transfer the sample to a precooled 50 ml conical tube.


Isolation with TissueRuptor

 **10 min**

- a. Keep the frozen plant material cold as long as possible—transfer plant material to an empty precooled 50 ml conical tube. If plant material is larger than $\sim 0.5 \text{ cm}^2$ mince it with a mini pestle or scalpel until material is smaller than $\sim 0.5 \text{ cm}^2$.
- b. Add 10-20 ml precooled sorbitol wash buffer (without β -mercaptoethanol) to plant material. Keep the tube cold the whole time.


 The quantity of sorbitol buffer depends on the amount of plant material.

- c. Pull over TissueRuptor Disposable Probes on TissueRuptor device. Mix 2-4 times for 30 sec on max speed. Move through the tube. The plant material should be homogeneous and no bigger plant material pieces should be visible.

 Mix no longer than 30 sec, buffer can foam intensely. The foam should not reach the end of the probes of TissueRuptor; there can be a risk of damaging the device. The foam is not a concern for nuclei isolation.

4. Incubate for 5 min on ice.


5. Centrifuge for 15 min at 3,000-10,000 x g (depending on genome size) at 4 °C.

 3,000 x g are recommended for tree genome size $>1 \text{ Gb}$. For tree genome sizes $< 1 \text{ Gb}$ or unknown genome sizes 7,000 x g are recommended. For *F. excelsior* the centrifugation speed was lowered to 5,000 x g from the recommended speed. Too fast centrifugation can result in damaging the isolated nuclei.


Species	Centrifuge spin speed	Haploid genome size
<i>Fraxinus excelsior</i>	5,000 x g	0.867 Gb
<i>Taxus baccata</i>	3,000 x g	ca. 10.5 Gb

6. Repeat until the supernatant is clear or only a hint of green remains.

7. If repeat is necessary: discard supernatant, add 1 ml precooled Sorbitol and resuspend the pellet using a small paintbrush. Add again 10 ml sorbitol with β -mercaptoethanol and mix with a P1000 pipette until the solution is homogenous.


 Depending on the pellet's consistency, suspension with a paintbrush is optional. Light shaking can be performed instead.

8. If the supernatant is clear, discard the supernatant and add 1 ml NIB (add β -mercaptoethanol to NIB, use 125 μ l β -mercaptoethanol per 50 ml NIB), resuspend the pellet using a small paint brush and mix with a P1000 pipette until the solution is homogenous. Fill up to 15 ml with NIB.

 CAUTION: The procedure should be further performed under a fume hood to avoid inhalation of β -mercaptoethanol.


9. Place the tube on ice on the tube rotator, shaker, or platform rocker for 15 min.
Horizontal shaker on 150 rpm or a platform rocker at max speed.

10. Use a Steriflip to filter the lysate. Approximately 10-15 ml of lysate should be left after filtering.

 Depending on the consistency of the lysate, the filter of the Steriflip can get clogged. Additional buffer or/and using a second Steriflip can be an option.


11. Centrifuge for 10 min; spin speed depends on the genome size of the tree species (step 6).

12. Discard supernatant and, add 1 ml precooled NIB and, resuspend the pellet using a small paint brush and mix with a P1000 pipette until the solution is homogenous.
Fill up to 10 ml with NIB.

 The pellet mostly consists of nuclei at this point, mixing with the pipette should be done gently and not more than 5-7 times.

13. Centrifuge at 4 °C for 10 min; again, spin speed depends on the genome size of the tree species (step 6).

14. Repeat steps until the supernatant is clear, 4– 7 times.

 Nuclei can be damaged by repeating the centrifugation step too often. Examples of nuclei pellet shades can be found in Supplementary Figures 1-4.

Species	NIB wash step repeats
<i>Fraxinus excelsior</i>	5-7
<i>Taxus baccata</i>	4


15. Discard supernatant; add 1 ml precooled 1x HB to the pellet.
16. Solve the pellet with a paintbrush and resuspend with P1000 pipette.
17. Transmit 1 ml nuclei suspension into a 1.5 ml DNA LoBind microcentrifuge tube.
18. Centrifuge at 7,000 x g for 5 min at room temperature. To avoid overheating, it is best to use a cooling centrifuge and set it to 18-22 °C.
19. Discard supernatant. Continue with Nanobind® plant nuclei kit DNA (PacBio® 102-302-000) extraction protocol or freeze nuclei pellet in liquid nitrogen and store at -80 °C.

DNA extraction


Only those steps were listed that have been adapted from the original protocol 'Extracting HMW DNA from plant nuclei using Nanobind® kits' by PacBio (PacBio, 2022a). This protocol consists of sixteen steps. Researchers interested in the detailed protocol are directed to the original source.

Total duration 3 - 4 h

1. Consistency of nuclei pellet can differ between tree species. *F. excelsior* nuclei pellet can be viscous. The nuclei pellet of *T. baccata* is highly viscous and of jelly-like consistency. The nuclei lysate can get stuck in the pipette tips (wide bore tips) due to viscosity. Mixing by vortexing and light stirring should be conducted for the efficiency of lysis.


 For an efficient lysis, the amount of Proteinase K, RNase A and PL1 can be increased. For *T. baccata*, it can be necessary to double the amounts of reagents.
3. Mixing step is important, additionally mixing with the vortexer is necessary.

4. Incubation on a ThermoMixer can be expanded to 2 h or more. After 15 min 5x 1 sec pulse mix on vortexer, invert gently every 30 min. If the lysate is less viscous, the rpm can be lowered (approximately 450 rpm).

 Mixing with a vortexer after a 15 min period can be critical for the DNA quality.

Species	Incubation time
<i>Fraxinus excelsior</i>	2 h
<i>Taxus baccata</i>	1 – 1.5 h

8. Adaptation of HulaMixer settings, DNA can be sticky and the Nanobind disk should not get stuck at the bottom or top of the tube.

 Observe mixing on the HulaMixer closely, to adapt if the Nanobind disk is stuck.

Step	Setting	Time (s)
Rotation	9 rpm	OFF
Tilting	30°	12
Vibration	2°	1

11. Repeat the washing step 3-5 times until the DNA is clear.

Quantity and purity (A260 /A280 and A260 /A230 ratios) measurement.

If required, follow the Short Read Elimination Kit protocol by PacBio (cat.no. SS-100-101-01, since March 2023: SKU 102-208-300). If HMW DNA concentration is sufficient, continue with ONT library preparation or PacBio protocol. The SRE aims to reduce the number of short DNA fragments. Short genomic DNA fragments have a higher binding preference than long fragments to the flow cell pores or SMRT cell ZMW in single molecule sequencing platforms such as ONT and PacBio. This higher preference for short fragments can shift the mean fragment length or N50 towards shorter reads.

Estimating DNA fragment size distribution through gel electrophoresis

For a rough estimation of the DNA fragment size, agarose gel electrophoresis with Marker Lambda Mix 19 (Fermentas #SM0231) can be performed, bearing in mind the inherent limitations of this method for precise sizing. The Marker Lambda Mix 19 was prepared using 2 µl Lambda Mix, 2 µl 6x loading dye (Fermentas #R0611) and 8 µl H₂O. The mix was denatured at 65° C for 5 min and stored on ice

afterwards. The samples were prepared using 2 µl HMW DNA, 2 µl Orange G and 5 µl H₂O. Agarose gel was prepared at a concentration of 0.3 % in 0.5x TBE and ran at 100 V in 0.5x TBE for 5h.

Characterizing DNA fragment size distribution through precision analysis with the Agilent FEMTO Pulse System

To examine the DNA's size and integrity, we used the FEMTO Pulse device (Agilent Technologies), which applies pulse field gel electrophoresis. HMW DNA was separated using the Genomic DNA 165 kb Kit (Agilent Technologies, FP-1002-0275) according to the manufacturer's instructions (<https://www.agilent.com/cs/library/usermanuals/public/quick-guide-fp-1002-gdna-165-kb-kit-SD-AT000141.pdf>).

In brief, each DNA sample was equilibrated to room temperature for at least 30 min and then diluted with 0.25x TE buffer to a final concentration of 125 – 250 pg/µl by performing serial dilution steps. Between each step, a 30-minute interval incubation at room temperature was applied. 2 µl of final diluted samples and the HMW ladder were taken into a FEMTO pulse run for 1.5 hours. These pulse field conditions allow the separation of HMW DNA fragments up to 165 kb. The visualization and analysis of the FEMTO Pulse runs were carried out with the Prosize software tool (Agilent Technologies).

Oxford Nanopore Sequencing Setup

The input DNA mass can influence the outcome as well, which means choosing the right sequencing kits and amount of DNA is crucial. For ONT the input mass for the flow cells varies, depending on the ligation sequencing kits. When using the ligation sequencing kit SQK-LSK110 the input mass is 150-300 fmol (ca. 1.5-3 µg) DNA for the flow cell R10.4.1 and 100-200 fmol (or 1.5 - 3 µg) DNA for the flow cell R9.4.1 (which will be soon discontinued). Ligation sequencing kit SQK-LSK114 is only compatible with the R10.4.1 flow cell; the HMW DNA input mass for this combination is 100-200 fmol (ca. 1 µg) DNA. The purity should be measured using Nanodrop (A260/A280 of 1.8 and A260/A230 of 2.0–2.2). All DNA input for sequencing and purity are recommended by the manufacturer.

PacBio high-fidelity sequencing setup

PacBio HiFi libraries have been prepared making use of the most recent protocol (Preparing whole genome and metagenome libraries using SMRTbell® prep kit 3.0). gDNA has been sheared with the MegaRuptor™ device (Diagenode) to fragment sizes of 20 kb. The final PacBio libraries have either been size selected for fragments larger than 8 kb with the Blue Pippin™ device or alternatively by AMPure beads (Beckman) as described in the Pacbio library preparation guide. Input into the library preparation protocol was 3 ug of sheared DNA. PacBio requires a minimum of 2-3 µg of DNA per 1 Gb of genome length. Pacbio HiFi libraries were loaded on the SEQUEL2 device on SMRT cells with 8 million zero mode waveguides (ZMW) and run with the SEQUEL II sequencing kit 2.0 for 30 hours. For circular consensus sequence (CCS) calling, either the standard PacBio SMRTlink pipeline for CCS calling could be used or, alternatively, the deep consensus tool (Baid et al., 2023).

Results

HMW DNA concentration and quality

For a reliable estimation of HMW DNA concentration, it is necessary to measure DNA yield three times for every sample from the top, middle and bottom of the sample. Yield can differ up to 10-20 ng/µl between the fractions (Table 5). The main reason for the fluctuation is the inhomogeneity of the HMW DNA in the tube. The DNA yield measured by Nanodrop and Qubit should ideally be within a 1:1 to 2:1 ratio (Nanodrop/Qubit). DNA yield was measured before and after applying the short read eliminator kit (PacBio).

Table 5: High-molecular-weight DNA extraction quality and quantity values after application of the protocol to the two tree species *Fraxinus excelsior* and *Taxus baccata*. For each measurement 1 μ l DNA was used.

Before Short Read Eliminator				
	Nanodrop (ng/μl)	A260/A280	A260/A230	Qubit (ng/μl)
<i>Fraxinus excelsior</i>	120-900	1.60-1.90	1.50-1.80	300-500
<i>Taxus baccata</i>	220-800	1.70-1.80	1.50-1.90	200-500
After Short Read Eliminator				
	Nanodrop (ng/μl)	A260/280	A260/A230	Qubit (ng/μl)
<i>Fraxinus excelsior</i>	25-175	1.30-1.90	1.40-2.00	40-420
<i>Taxus baccata</i>				

Between 6-12 different HMW DNA preparations were considered for each of the tree species. Detailed values of each sample of *Fraxinus excelsior* are presented in supplement table-S3. For detailed values of *Taxus baccata* see table 7. HMW DNA extraction from *Fraxinus excelsior* is indicated for sequencing with Oxford Nanopore. The *Taxus baccata* HMW DNA extraction (without short-read elimination) is indicated for PacBio HiFi sequencing.

Fragment size distribution of the HMW DNA of *Fraxinus excelsior*

The DNA of *F. excelsior* extracted with our protocol was tested on the ONT MinION sequencing platform. Four samples were sequenced on four flow cells, as outlined in Table 6, aiming to provide a comprehensive understanding of their respective characteristics. For all four *F. excelsior* samples, the short read eliminator was employed before library preparation, and the sequencing was performed with an R9.4.1 flowcell (MinION) and the SQK-LSK110 ligation sequencing kit from ONT (Input DNA for the flow cell in Table 6). Notably, samples I and II underwent identical protocol steps including grinding with the TissueRuptor (Table S.1). Nevertheless, these samples exhibited different read N50 values, with 12.91 kb and 17.88 kb, respectively. The total output for sample I was 23.17 Gb (26.1X genome coverage), while sample II yielded 16.45 Gb (18.5X genome coverage), highlighting the considerable genomic content captured in both cases. Sample IV, representing a different genotype, followed the same isolation and extraction steps as samples I and II. Despite a good N50 value of 24.71 kb, it generated the lowest output with of 5.81 Gb (6.5X genome coverage; Figure 3). Sample III, while exhibiting the highest N50 value of 38.19 kb, also had low output (7.54 Gb, 8.5X coverage) (Table 6).

Table 6: Comparative analysis of high-molecular-weight DNA sequencing results in *F. excelsior* genotypes using Oxford Nanopore Technology. The DNA concentration measurement with Nanodrop and Qubit is the mean out of three measurements (top, middle, bottom). All DNA samples were stored at -80 °C, samples I, II and III have been stored for 5 months and sample IV for 12 months. The input DNA volume for ONT was adjusted according to Qubit measurement to achieve an input DNA amount of about 100-200 fmol DNA.

Sample	Genotype	Tissue weight (g)	Nanodrop (ng/μl)	A260/280	A260/A230	Qubit (ng/μl)	Input (μl)	N50 (kb)	Bases (Gb)	Genome coverage
I	A	2	36.3	1.92	2.12	42.4	23.5	12.91	23.17	26.1X
II	A	2	56.9	1.76	1.88	69.2	14.45/ 14.5/18.0	17.88	16.45	18.5X
III	A	1	43.0	1.26	0.56	41.0	24.3	38.19	7.54	8.5X
IV	B	2	59.5	1.61	1.40	83.8	14.23/28.6	24.71	5.81	6.5X

A flow cell for sequencing was used 1-3 times for one sample depending on DNA quality and quantity (row input (μl)).

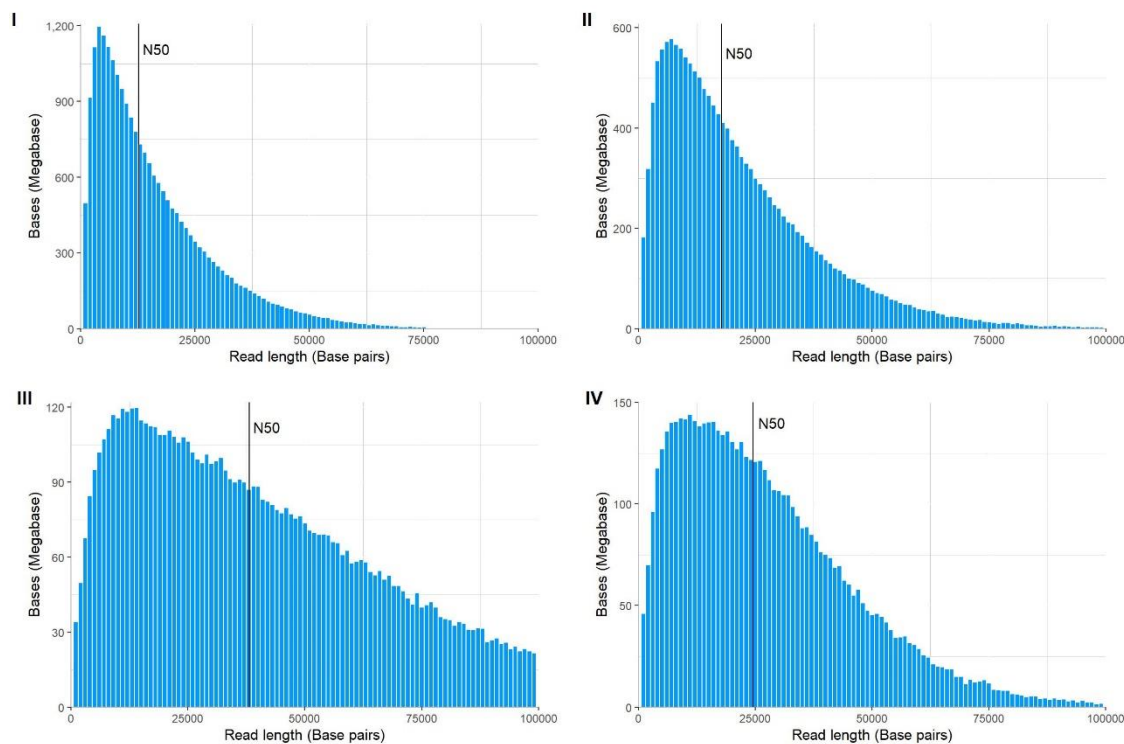


Figure 3: Sequencing results of *Fraxinus excelsior* DNA using the Oxford Nanopore Technologies MinION platform. The bar plots illustrate the read length distribution for the four samples, vertical lines highlight the N50 values. The Y-axis represents the cumulative number of bases in Megabase pairs (Mb) and the X-axis the read length range in base pairs (bp).

Fragment size distribution of the HMW DNA of *Taxus baccata*

Using the presented protocol, HMW DNA was extracted from 12 needle samples (1-1.6 grams material) of *T. baccata* for PacBio HiFi sequencing (Tables 7 and 8). The fragment size distribution of the DNA was assessed by the FEMTO Pulse system (Agilent) (Figure 4, Figures S.5-S.15). The average DNA fragment size, as measured by the Agilent FEMTO Pulse System and/or gel electrophoresis, was >30 kb.

Table 7: Details of 12 high-molecular-weight DNA extractions from needles of two *Taxus baccata* individuals.

ID	Genotype	Starting weight (g)	Nanodrop (ng/μl)	A260/A280	A260/A230	Qubit (ng/μl)	total amount (μg)	GQN (50kbp)
1	A	1.0	521.50	1.6 9	1.29	356	17.80	2.7
2	A	1.0	160.50	1.8 6	1.69	124	6.20	6.2
3	B	1.0	375.90	1.8 1	1.82	288	14.40	3.7
4	A	1.0	518.24	1.7 5	1.49	339	16.95	5.8
5	A	1.0	379.43	1.7 9	1.34	194	9.70	5.6
6	A	1.3	524.15	1.7 7	1.51	405	20.25	6.4
7	A	1.3	197.46	1.8 0	1.62	155	7.75	4.7
8	B	1.2	438.42	1.7 6	1.56	316	15.80	5.9
9	B	1.2	529.68	1.7 7	1.68	376	18.80	3.7
10	B	1.4	387.51	1.8 0	1.69	230	11.50	6.1
11	B	1.5	368.18	1.7 9	1.63	253	12.65	6.1
12	B	1.6	691.32	1.8 2	1.82	422	21.10	3.9

NanoDrop measurements were performed on sample parts taken from the middle of the tube. m denotes the total amount of extracted DNA based on Qubit concentration measurements. GQN (genome quality number) estimates the proportion of DNA that is contained in fragments longer than 50,000 bp based on FEMTO Pulse results on a scale from 0-10. Before GQN measurements, samples were frozen at -20°C for 1-3 months. HMW DNA extraction exact steps for each sample are detailed in Table S2.

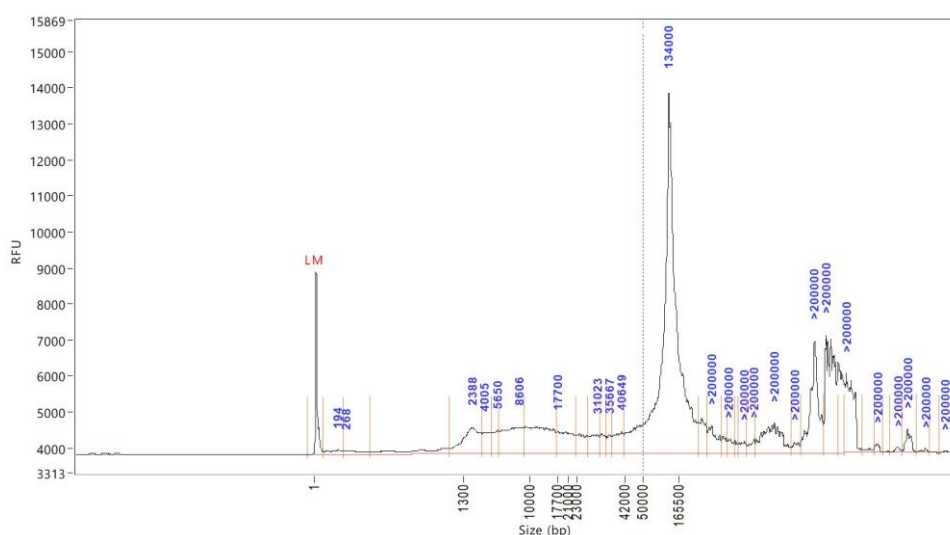


Figure 4: FEMTO Pulse DNA sizing QC analysis of *Taxus baccata* sample 2 (see Table 7). RFU = relative fluorescence unit.

Successful PacBio HiFi sequencing depends on long and intact DNA without contaminants such as secondary metabolites that might interfere with sequencing polymerase and/or the SMRT cell loading by sticking to the native DNA molecules. To test for these requirements, not only the size and integrity of the DNA is analyzed by pulse field gel electrophoresis, but also the purity of the DNA is determined spectrophotometrically. The A260/A280 and A260/A230 ratios should be in the range of 2. Another indication of high-quality DNA is the comparison of concentration measurements by Qubit and Nanodrop (Table 7). This ratio is preferably between 1 and 3.

For PacBio sequencing in general, and especially considering the large genome size of *T. baccata* (circa 10.5 Gb), it is recommended to add a size-selection step to the PacBio HiFi library preparation, e.g., with a BluePippin device (SAGE). The size selection removes short fragments of a given size range. Especially for large genome assemblies, it is highly recommended to sequence long and continuous fragments with high accuracy. These highly precise and continuous reads increase the chance to span repetitive regions in large genomes, which will provide better contiguity across the genome. In addition, a size selection by preparative pulse-field gel electrophoresis adds another purification step to the HiFi library preparation protocol. Although the DNA is enriched for longer fragments, this step comes with the caveat of DNA input. The HiFi library preparation includes a mechanical shearing step to achieve a defined library size. In addition, some shearing cannot be avoided when pipetting long gDNA, and some gDNA nicking occurs. Depending on the size and integrity of the input gDNA, about 20-50% of the HiFi library gets lost during the size selection by preparative gel electrophoresis. However, this loss of the HiFi library comes with the big advantage of a HiFi library that contains only fragments larger than, e.g. 10 or 12 kb, depending on the size exclusion. Figure 5 shows the result of a representative PacBio HiFi run based on a HiFi library generated from *T. baccata* HMW DNA that was prepared according to our protocol. The average length of the generated CCS reads of this SMRT cell is 11.9 kb, with the majority of CCS reads larger than 8 kb (Figure 5). The median QV score of these CCS reads is Q37. The total yield of CCS reads of this SMRT cell was 30.6 Gb (Figure S.16).

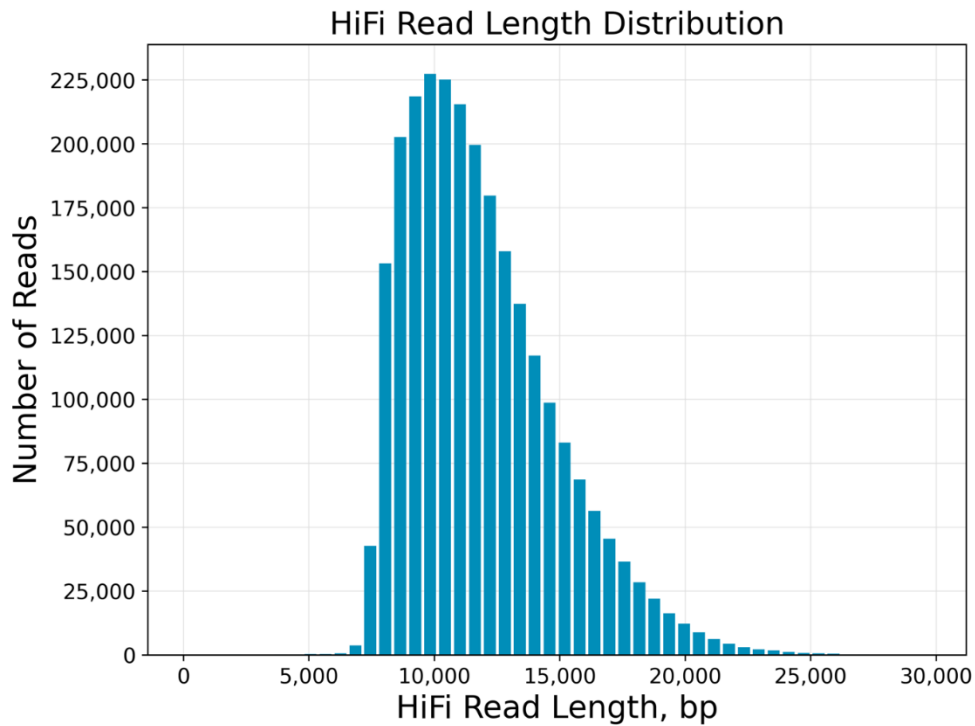


Figure 5: Sequencing results of *Taxus baccata* DNA using PacBio HiFi sequencing (Run ID 21, Table 8). The histogram shows the number of circular consensus sequencing (CCS) reads (Y-axis) of a given length range (X-axis). The average length of the CCS reads of this SMRT cell is 11.9 kb with the majority of CCS reads larger than 8 kb. HiFi reads provide base-level resolution with 99.9% single-molecule read accuracy. A description of the HiFi read generation is found here: <https://www.pacb.com/technology/hifi-sequencing/>. CCS reads have been called with the PacBio SMRT link pipeline (SMRTlink version 11.0.0.146107).

Table 8 shows further PacBio HiFi sequencing results based on deep consensus (Baid et al. 2023) calling of 24 runs with libraries based on samples 2 and 12 (Table 7), with an average N50 HiFi read length of 13.59 kb and a mean yield of HiFi reads of 32.05 Gbps. These findings underscore the importance of size selection in PacBio HiFi library preparation, particularly for larger and more repetitive genomes, facilitating efficient contig generation and the need for high-quality, high-molecular-weight DNA to reach good-quality results.

Table 8: HiFi yield, the mean HiFi read length and the N50 of HiFi reads of *Taxus baccata*. The deep consensus tool was used to calculate the HiFi reads. The first part of the Library ID refers to the sample ID of the input material (see Table 7).

Run ID	Library ID	Yield HiFi reads (Gbp)	Mean length of HiFi reads (kb)	N50 of HiFi reads (kb)
1	12-1	34.55	11.01	13.39
2	12-1	39.08	13.30	14.86
3	12-1	39.06	13.20	14.75
4	12-1	39.32	13.04	14.69
5	12-1	37.50	12.76	14.45
6	12-1	35.42	11.59	13.91
7	12-1	34.81	12.01	14.13
8	12-1	32.93	10.73	12.93
9	12-2	32.58	11.33	13.59
10	12-2	34.07	10.85	13.58
11	2-1	21.72	10.02	13.02
12	2-1	27.73	10.94	13.76
13	2-1	19.54	11.35	14.41
14	2-1	27.61	11.72	14.49
15	2-2	27.61	12.08	12.33
16	2-2	29.71	12.90	13.17
17	2-2	33.40	13.44	13.80
18	2-2	33.57	13.82	14.25
19	2-2	30.28	14.01	14.49
20	2-2	27.75	12.23	13.17
21	2-3	34.51	11.89	12.13
22	2-3	32.40	11.86	12.11
23	2-3	32.92	12.07	12.34
24	2-3	31.27	12.07	12.35

Discussion

Limitations and critical steps in the procedure

The presented results are based on five to twelve successful HMW DNA preparations of the two tree species, *F. excelsior* and *T. baccata*. Our proposed methodology can likely be adapted to other tree species. Notably, various improvements contribute to its overall effectiveness.

Quality of plant material

The sample material quality is essential. The timing of plant material harvest stands as a crucial determinant, varying significantly among species. It arises due to diverse factors such as life cycle variations, needle or leaf characteristics, and other species-specific considerations (Formato et al., 2022; Gaytán et al., 2022; Lin et al., 2024). For *F. excelsior*, young tree leaves a month after flushing did not work well for HMW DNA extraction with our protocol. The material was difficult to wash with NIB.

Presumably, secondary metabolites, whose composition depends on the age of the leaf, could be a cause (Jouve et al., 2007) The harvest between June and July produced acceptable results for *F. excelsior*. *T. baccata* was more flexible with the harvest time. In general, if the trees are cultivated in a greenhouse, it is important to keep them as healthy as possible and away from plants with any disease.

Quantity of plant material

The quantity of plant material used affects the extraction results. We used different amounts of input material for the two species. Increasing the amount of *T. baccata* needles resulted in a drastically higher DNA yield. The resulting nuclei pellet was difficult to handle during lysis, particularly in resuspension, filtration, and mixing steps. This trend was observed consistently. For other plant species and protocols, input material can be as high as 5-10 grams (Li et al., 2020; Zerpa-Catanho et al., 2021; Kang et al., 2023). Depending on the plant used, the amount needs to be adapted. While increasing the amount of material may lead to higher DNA yield in some cases, as observed in *T. baccata*, it does not necessarily guarantee the same outcome for all species or sample types. Zerpa-Cetanho et al. 2021 mentioned that the white colour of the nuclei pellet can be an indicator of the number of nuclei. If the nuclei pellet is not white, this may indicate that not enough nuclei are available.

Grinding

The grinding is one of the most delicate steps of the protocol. The grinding process can cause abrasion of the DNA, and bruised and thawed leaves produce polyphenols that corrupt the results, and DNA fragment length can be influenced (Couch and Fritz, 1990; Kang et al., 2023). The grinding tool used has the biggest impact. We observed an increase in DNA yield after using the TissueRuptor, with an increase between 100-600 ng/ μ l compared to the results using mortar and pestle; however, the TissueRuptor may negatively affect fragment length. The grinding process with mortar and pestle involves several steps, and the quality and quantity of DNA output can suffer. The temperature, the time of grinding and the right magnitude of sample material influence the grinding success. Other options were tested, since a TissueRuptor is not always available and expensive. A beater mill or similar devices that are often used for grinding plant material for DNA extraction seem to be an

option. Unfortunately, both species showed similarly deficient results. Mostly, the amount of plant material and the division into separate 1.5 ml tubes and the cooling process seem to be an issue. In the context of Oxford Nanopore sequencing, it has been observed that the grinding process significantly impacts read length distribution and N50 values. Notably, the use of mortar and pestle grinding has demonstrated optimal results in this regard (La Cerda et al., 2023). However, in other protocols based on CTAB and other tree species (e.g., *Populus tremula*), the use of cyclone mills or similar devices worked well for HMW DNA extraction and long-read sequencing (Inglis et al., 2018; Müller et al., 2020). Some authors recommend keeping the time of grinding to a minimum of 2 min and using mechanical separation devices for best results (Jones et al., 2023; Kang et al., 2023; Schenk et al., 2023).

Application of PVP

PVP can reduce the number of polyphenolic compounds that contaminate the extracted DNA and reduce browning of DNA (Loomis and Battaile, 1966; Rogers and Bendich, 1985). PVP causes foaming of the NIB and sorbitol wash buffers, which affects handling during the grinding process and washing steps. We used PVP360 for the NIB and PVP40 for sorbitol wash buffer. Other protocols use PVP10, which resulted in higher DNA yield (Carey et al., 2023; Schenk et al., 2023). Further, for the CTAB extraction method no PVP is included in the NIB but only used in the CTAB solution. In most CTAB protocols, purifying steps with chloroform are added, which removes the carryover of contaminants (Zerpa-Catanho et al., 2021; Russo et al., 2022). Additionally, there is the option to use PVPP (Polyvinylpolypyrrolidon) as a substitute (Porebski et al., 1997; Schenk et al., 2023).

Centrifuge speed

The genome size determines the optimal centrifuge force. Out of the two tested species, *T. baccata* had the largest haploid genome size (10.5 Gb). Here, the best extraction results were achieved using 3,000 x g. For *F. excelsior* (genome size of 0.867 Gb), the centrifugation speed was lowered to 5,000 x g from the recommended speed (7,000 x g for genome size <1 Gb). The quantity of DNA increased after that adaption.

DNA 'jellies'

Another aspect observed with both species is the formation of DNA 'jellies' after completing the nuclei extraction. The DNA 'jellies' are supposed to disperse overnight. Trials to leave the DNA resuspended over 24 h to 72 h did not make a difference; the 'jellies' dispersion was incomplete. The 'jellies' have been a problem for DNA yield measurement during the cleaning steps and when proceeding with the Oxford Nanopore ligation sequencing kit. One reason for the 'jellies' can be a DNA-phenol complex (Japelaghi et al., 2011).

Quality and quantity of DNA

Other protocols recommend the increase of proteinase K during nuclei lysis if the A260/280 is too low (Zerpa-Catanho et al., 2021). Lysis period varies between different protocols; a longer lysis can increase the quality and quantity of extracted HMW DNA. Kang et al. 2023 tested multiple factors which can influence the DNA results (CTAB method used). One was the pipetting during the process. They recommend minimizing the pipetting and using wide-bore tips during the process. In another study, they tested different incubation times and temperatures using the CTAB method on herbarium leaf tissue. They concluded that lower or higher temperatures and shorter incubation times resulted in DNA being less fragmented and less contaminated (Carey et al., 2023). Different approaches are proposed for purification, such as DNA precipitation or liquid phase extraction. They suggest using solid phase extraction (high salt gel electrophoresis trap), which separates the DNA with electrophoretic mobility (Kalendar et al., 2021).

In a comparison of the output of the four *F. excelsior* samples that have been sequenced with ONT, sample III had the most controversial output (Table 6). The N50 is high with a value of 38.19 kb, but a low output with 7.54 Gb and 8.5X coverage. Interestingly, sample III was the only one for which mortar and pestle were used for grinding, indicating that the use of the TissueRuptor might compromise fragment length but increase data output. The utilization of the different grinding method may have also adversely affected the purity of the DNA in sample III, leading to failure to meet the quality criteria for ONT sequencing. This may be another reason

for the lower output of that sequencing run. Overall, the flow cell quality, e.g. such as the number of active pores over the sequencing run, can affect the outcome.

Conclusion

Protocols for extracting high-molecular-weight DNA from various species are becoming more applicable. Yet, a noticeable gap persists for perennial species such as trees, which often have a large repertoire of secondary metabolites. Excellent quality HMW DNA is essential for long-read sequencing and subsequent *de novo* genome assemblies, which have been increasingly applied in recent years. This study addressed key aspects of HMW DNA extraction for broadleaved and conifer tree species. While the study could benefit from a higher number of repetitions, our protocol's implications are nevertheless relevant. It opens avenues for extracting high-quality HMW DNA from tree species for Oxford Nanopore and PacBio HiFi sequencing. In addition to its application in long-read sequencing for genome assembly, our nuclei extraction protocol has demonstrated versatility in other genomic techniques. For instance, nuclei extracted using our protocol were successfully utilized for chromatin conformation capture using the ARIMA Hi-C protocol, allowing for comprehensive exploration of the three-dimensional chromatin structure in *T. baccata*. Overall, this study underscores the importance of the quality and quantity of the plant material and choosing the right grinding technique and centrifuge speed. Our protocol for extracting high-quality genomic material from two tree species has the potential to be adapted to various other tree species in future studies.

Acknowledgements:

We thank our technical assistant Katrin Groppe for her help with laboratory work. We thank members of the FraxForFuture project and the Thünen Institute of Forest Genetics for experimental advice and discussion on the project and data analysis. The FraxForFuture research network is funded by the German Federal Ministry of Food and Agriculture and the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety. The projects of the sub-networks are funded by the Waldklimafonds and the Thünen Institute (funding code:

2219WK21A4). The project executing agency is the Fachagentur für Nachwachsende Rohstoffe e.V. (FNR).

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 497528752 – in the scope of the TaxGen project (funding of DFG to B.K.; DFG grant no. KE 916/10-1). NGS data production and data analysis for *T. baccata* were carried out at the DRESDEN-concept Genome Center, supported by the DFG Research Infrastructure Program (Project 407482635) and part of the Next Generation Sequencing Competence Network NGS-CN (Project 423957469).

References

- Aggerwal G, Edhigalla P, Walia P (2022) A comprehensive review of high-quality plant DNA isolation. *The Pharma Innovation Journal* 11(6):171–176
- Carey SJ, Becklund LE, Fabre PP, Schenk JJ (2023) Optimizing the lysis step in CTAB DNA extractions of silica-dried and herbarium leaf tissues. Wiley-Blackwell. *Appl Plant Sci* 11(3):e11522. <https://doi.org/10.1002/aps3.11522>
- Couch JA, Fritz PJ (1990) Isolation of DNA from plants high in polyphenolics. Springer; Kluwer Academic Publishers. *Plant Mol Biol Rep* 8(1):8–12. <https://doi.org/10.1007/BF02668875>
- Fleck S, Tomlin C, Da Coelho FS, Richter M, Danielsen E, Backenstose N, Krabbenhoft T, Lindqvist C, Albert V (2022) High quality long-read genomes produced from single MinION flow cells clarify polyploid and demographic histories of critically endangered ash species (*Fraxinus*: *Oleaceae*). <https://doi.org/10.21203/rs.3.rs-2350866/v1>
- Formato M, Scharenberg F, Pacifico S, Zidorn C (2022) Seasonal variations in phenolic natural products in *Fagus sylvatica* (European beech) leaves. *Phytochemistry* 203):113385. <https://doi.org/10.1016/j.phytochem.2022.113385>
- Gaytán Á, Moreira X, Castagneyrol B, van Halder I, Frenne P de, Meeussen C, Timmermans BGH, Hoopen JPJG ten, Rasmussen PU, Bos N, Jaatinen R, Pulkkinen P, Söderlund S, Covelo F, Gotthard K, Tack AJM (2022) The co-existence of multiple oak leaf flushes contributes to the large within-tree variation in chemistry, insect attack and pathogen infection. *New Phytol* 235(4):1615–1628. <https://doi.org/10.1111/nph.18209>
- Grohme MA, Schloissnig S, Rozanski A, Pippel M, Young GR, Winkler S, Brandl H, Henry I, Dahl A, Powell S, Hiller M, Myers E, Rink JC (2018) The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. Nature Publishing Group. *Nature* 554(7690):56–61. <https://doi.org/10.1038/nature25473>
- Healey A, Furtado A, Cooper T, Henry RJ (2014) Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods*. *Plant Methods* 10):21. <https://doi.org/10.1186/1746-4811-10-21>

- Huang J, Hartmann H, Ogaya R, Schöning I, Reichelt M, Gershenzon J, Peñuelas J (2023) Hormone and carbohydrate regulation of defense secondary metabolites in a Mediterranean forest during drought. *Environmental and Experimental Botany* 209):105298. <https://doi.org/10.1016/j.envexpbot.2023.105298>
- Inglis PW, Pappas MdCR, Resende LV, Grattapaglia D (2018) Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One* 13(10):e0206085. <https://doi.org/10.1371/journal.pone.0206085>
- Japelaghi RH, Haddad R, Garoosi G-A (2011) Rapid and efficient isolation of high quality nucleic acids from plant tissues rich in polyphenols and polysaccharides. Springer; Humana Press Inc. *Mol Biotechnol* 49(2):129–137. <https://doi.org/10.1007/s12033-011-9384-8>
- Jebb D, Huang Z, Pippel M, Hughes GM, Lavrichenko K, Devanna P, Winkler S, Jermiin LS, Skirmuntt EC, Katzourakis A, Burkitt-Gray L, Ray DA, Sullivan KAM, Roscito JG, Kirilenko BM, Dávalos LM, Corthals AP, Power ML, Jones G, Ransome RD, Dechmann DKN, Locatelli AG, Puechmaille SJ, Fedrigo O, Jarvis ED, Hiller M, Vernes SC, Myers EW, Teeling EC (2020) Six reference-quality genomes reveal evolution of bat adaptations. Nature Publishing Group. *Nature* 583(7817):578–584. <https://doi.org/10.1038/s41586-020-2486-3>
- Jones MM, Nagalingum NS, Handley VM (2023) Testing protocols to optimize DNA extraction from tough leaf tissue: A case study in *Encephalartos*. *Appl Plant Sci*. *Appl Plant Sci* 11(3):e11525. <https://doi.org/10.1002/aps3.11525>
- Jouve L, Jacques D, Douglas GC, Hoffmann L, Hausman J-F (2007) Biochemical characterization of early and late bud flushing in common ash (*Fraxinus excelsior* L.). *Plant Science* 172(5):962–969. <https://doi.org/10.1016/j.plantsci.2007.02.008>
- Kalendar R, Boronnikova S, Seppänen M (2021) Isolation and Purification of DNA from Complicated Biological Samples. Humana, New York, NY. *Methods Mol Biol* 2222):57–67. https://doi.org/10.1007/978-1-0716-0997-2_3
- Kang M, Chanderbali A, Lee S, Soltis DE, Soltis PS, Kim S (2023) High-molecular-weight DNA extraction for long-read sequencing of plant genomes: An optimization of standard methods. *Appl Plant Sci* 11(3):e11528. <https://doi.org/10.1002/aps3.11528>
- Kersten B, Rellstab C, Schroeder H, Brodbeck S, Fladung M, Krutovsky KV, Gugerli F (2022) The mitochondrial genome sequence of *Abies alba* Mill. reveals a high structural and combinatorial variation. *BMC Genomics* 23(1):776. <https://doi.org/10.1186/s12864-022-08993-9>
- La Cerda GY de, Landis JB, Eifler E, Hernandez AI, Li F-W, Zhang J, Tribble CM, Karimi N, Chan P, Givnish T, Strickler SR, Specht CD (2023) Balancing read length and sequencing depth: Optimizing Nanopore long-read sequencing for monocots with an emphasis on the Liliales. *Appl Plant Sci*. *Appl Plant Sci* 11(3):e11524. <https://doi.org/10.1002/aps3.11524>
- Li Z, Parris S, Saski CA (2020) A simple plant high-molecular-weight DNA extraction method suitable for single-molecule technologies. *BioMed Central. Plant Methods* 16(1):38. <https://doi.org/10.1186/s13007-020-00579-4>
- Lin S, Wang H, Dai J, Ge Q (2024) Spring wood phenology responds more strongly to chilling temperatures than bud phenology in European conifers. *Tree Physiol* 44(1). <https://doi.org/10.1093/treephys/tpad146>
- Loomis WD, Battaile J (1966) Plant phenolic compounds and the isolation of plant enzymes. *Phytochemistry* 5(3):423–438. [https://doi.org/10.1016/S0031-9422\(00\)82157-3](https://doi.org/10.1016/S0031-9422(00)82157-3)

- Müller NA, Kersten B, Leite Montalvão AP, Mähler N, Bernhardsson C, Bräutigam K, Carracedo Lorenzo Z, Hoenicka H, Kumar V, Mader M, Pakull B, Robinson KM, Sabatti M, Vettori C, Ingvarsson PK, Cronk Q, Street NR, Fladung M (2020) A single gene underlies the dynamic evolution of poplar sex determination. Nature Publishing Group. *Nat. Plants* 6(6):630–637. <https://doi.org/10.1038/s41477-020-0672-9>
- Nishii K, Möller M, Foster RG, Forrest LL, Kelso N, Barber S, Howard C, Hart ML (2023) A high quality, high molecular weight DNA extraction method for PacBio HiFi genome sequencing of recalcitrant plants. *BioMed Central. Plant Methods* 19(1):41. <https://doi.org/10.1186/s13007-023-01009-x>
- Oxford Nanopore Technologies (2023) Input DNA/RNA QC. https://community.nanoporetech.com/docs/prepare/library_prep_protocols/input-dna-rna-qc/v/idi_s1006_v1_revb_18apr2016/assessing-input-dna
- PacBio (2022a) Extracting HMW DNA from plant nuclei using Nanobind kits: Procedure & checklist. <https://www.pacb.com/wp-content/uploads/Procedure-checklist-Extracting-HMW-DNA-from-plant-nuclei-using-Nanobind-kits.pdf>
- PacBio (2022b) Technical note Preparing DNA for PacBio HiFi sequencing— extraction and quality control. <https://www.pacb.com/wp-content/uploads/Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf>
- Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Springer; Springer-Verlag. *Plant Mol Biol Rep* 15(1):8–15. <https://doi.org/10.1007/BF02772108>
- Pucker B, Irisarri I, Vries J de, Xu B (2022) Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. Cambridge University Press. *Quantitative Plant Biology* 3):e5. <https://doi.org/10.1017/qpb.2021.18>
- Rogers SO, Bendich AJ (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Springer; Martinus Nijhoff, The Hague/Kluwer Academic Publishers. *Plant Mol Biol* 5(2):69–76. <https://doi.org/10.1007/BF00020088>
- Russo A, Mayjonade B, Frei D, Potente G, Kellenberger RT, Frachon L, Copetti D, Studer B, Frey JE, Grossniklaus U, Schlüter PM (2022) Low-Input High-Molecular-Weight DNA Extraction for Long-Read Sequencing From Plants of Diverse Families. *Front Plant Sci* 13):883897. <https://doi.org/10.3389/fpls.2022.883897>
- Schenk JJ, Becklund LE, Carey SJ, Fabre PP (2023) What is the “modified” CTAB protocol? Characterizing modifications to the CTAB DNA extraction protocol. *Appl Plant Sci*). <https://doi.org/10.1002/aps3.11517>
- Witzell J, Martín JA (2008) Phenolic metabolites in the resistance of northern forest trees to pathogens — past experiences and future prospects. *Can. J. For. Res.* 38(11):2711–2727. <https://doi.org/10.1139/X08-112>
- Zerpa-Catanho D, Zhang X, Song J, Hernandez AG, Ming R (2021) Ultra-long DNA molecule isolation from plant nuclei for ultra-long read genome sequencing. *STAR Protocols* 2(1):100343. <https://doi.org/10.1016/j.xpro.2021.100343>

Supplements

Figures



Figure S.1: *Taxus baccata* nuclei pellet in 50 ml tube in NIB wash buffer after last washing step with NIB.



Figure S.2.: *Taxus baccata* Nuclei pellet in 1.5 ml Eppendorf tube before nuclei extraction with Nanobind plant nuclei kit.



Figure S.3.: *Fraxinus excelsior* Nuclei pellet in 50 ml tube in NIB wash buffer after last washing step with NIB.



Figure S.4.: *Fraxinus excelsior* Nuclei pellet in 1.5 ml Eppendorf tube before nuclei extraction with Nanobind plant nuclei kit.

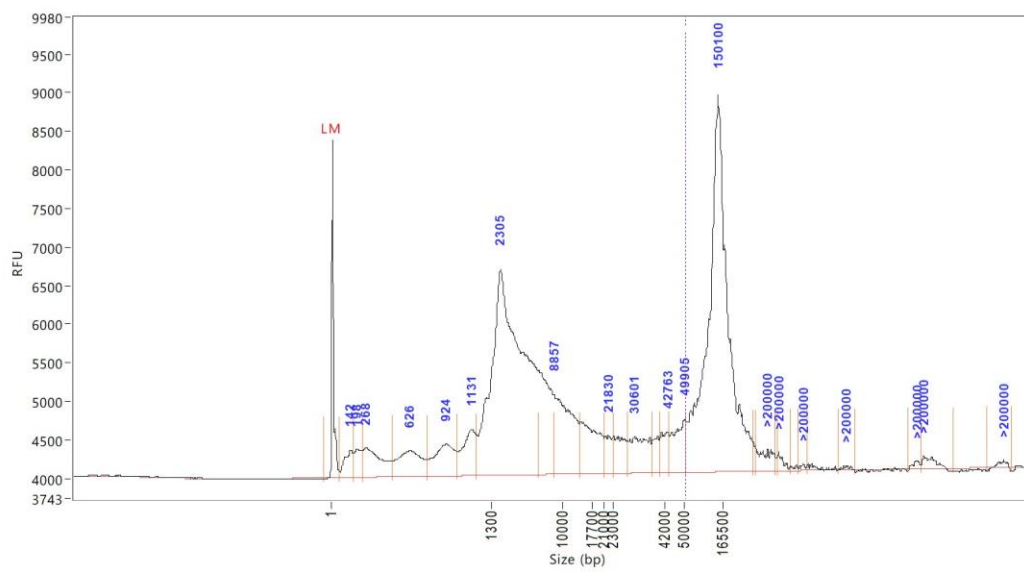


Figure S.5.: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 1.

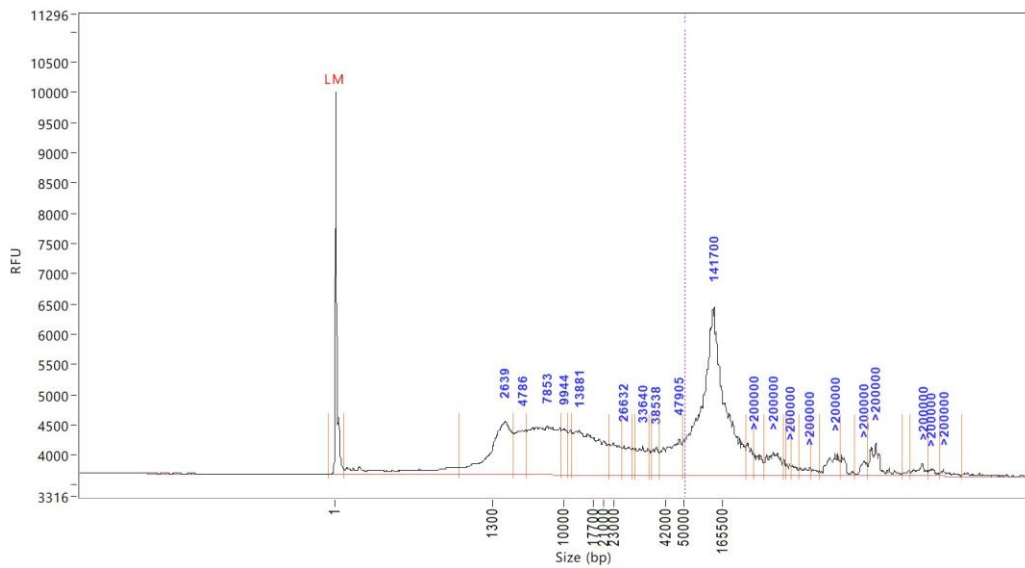


Figure S.6: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 3.

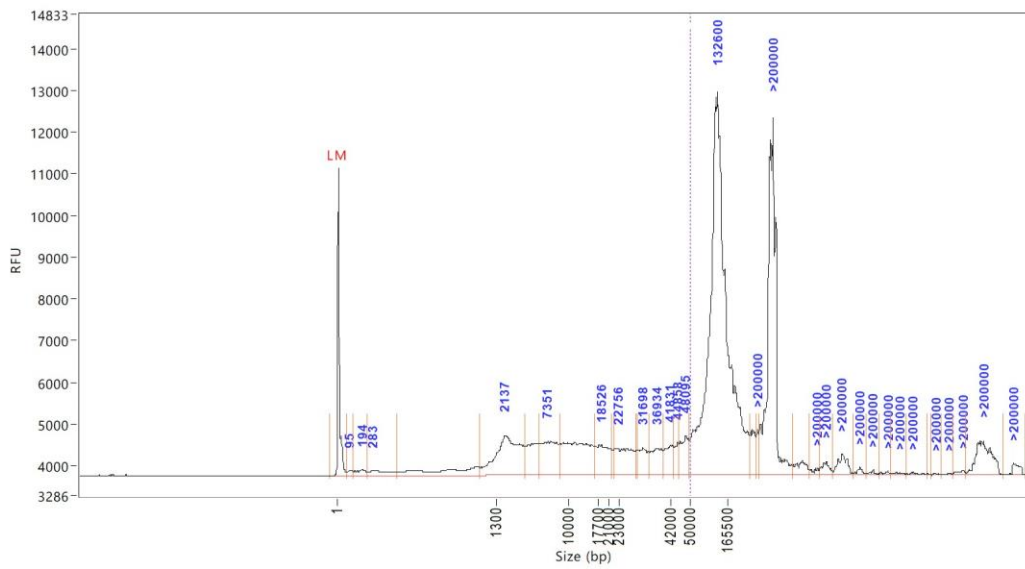


Figure S.7: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 4.

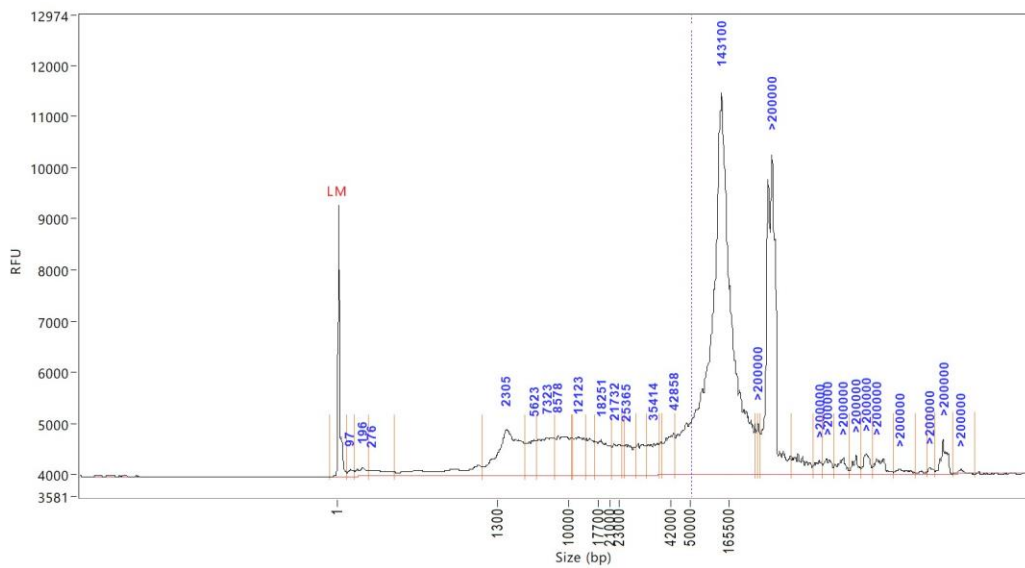


Figure S.8: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 5.

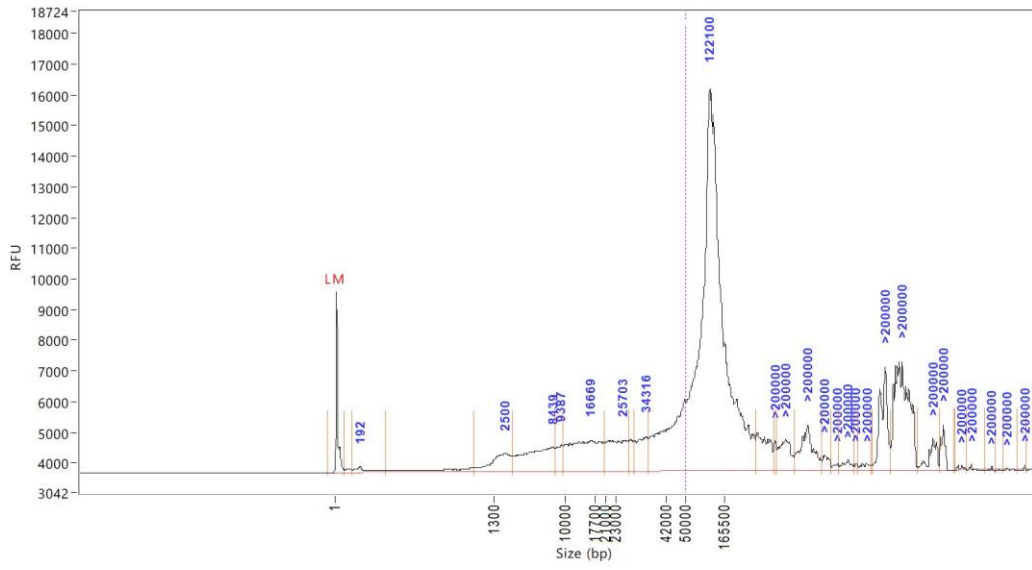


Figure S.9.: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 6.

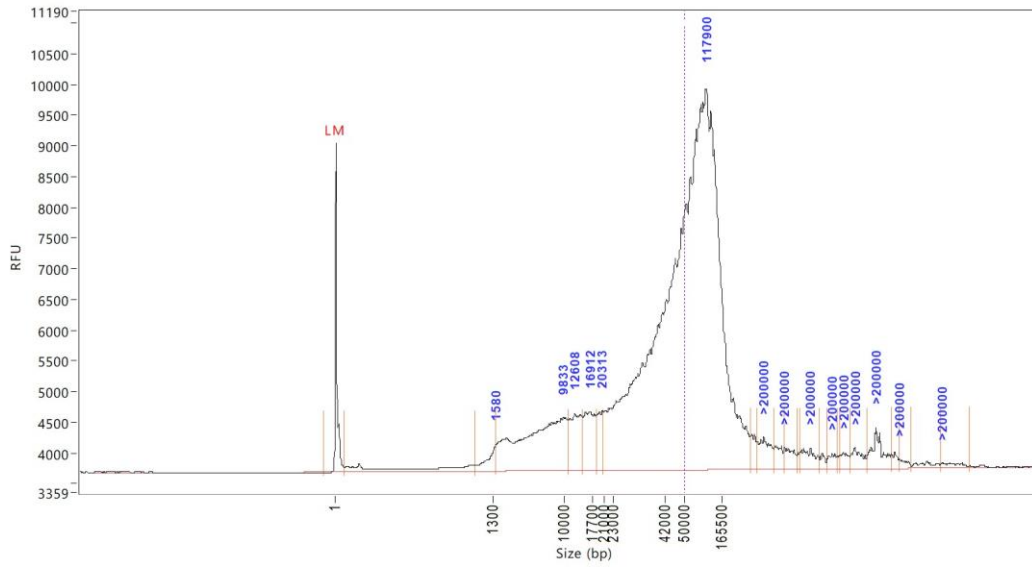


Figure S.10.: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 7.

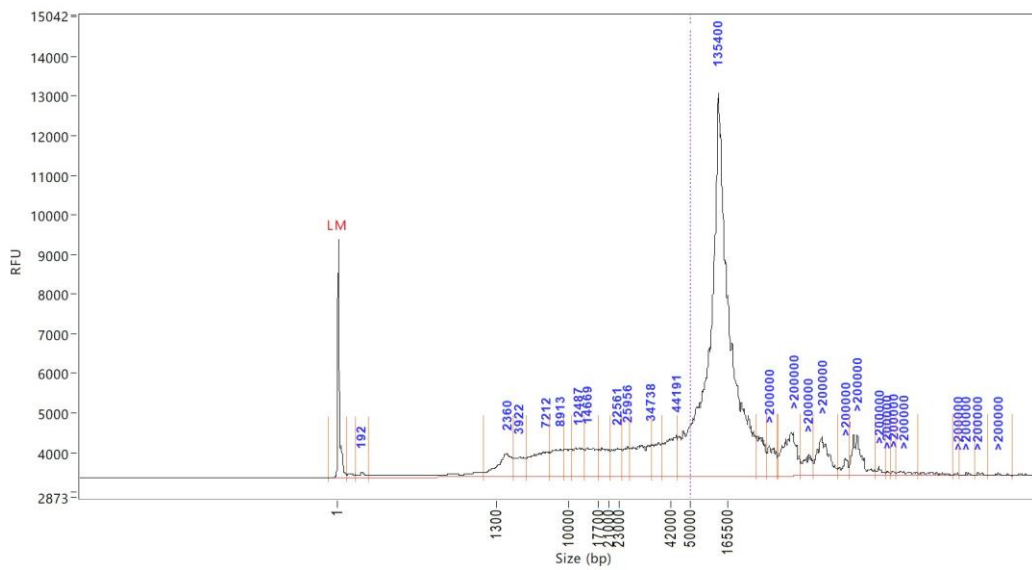


Figure S.11.: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 8.

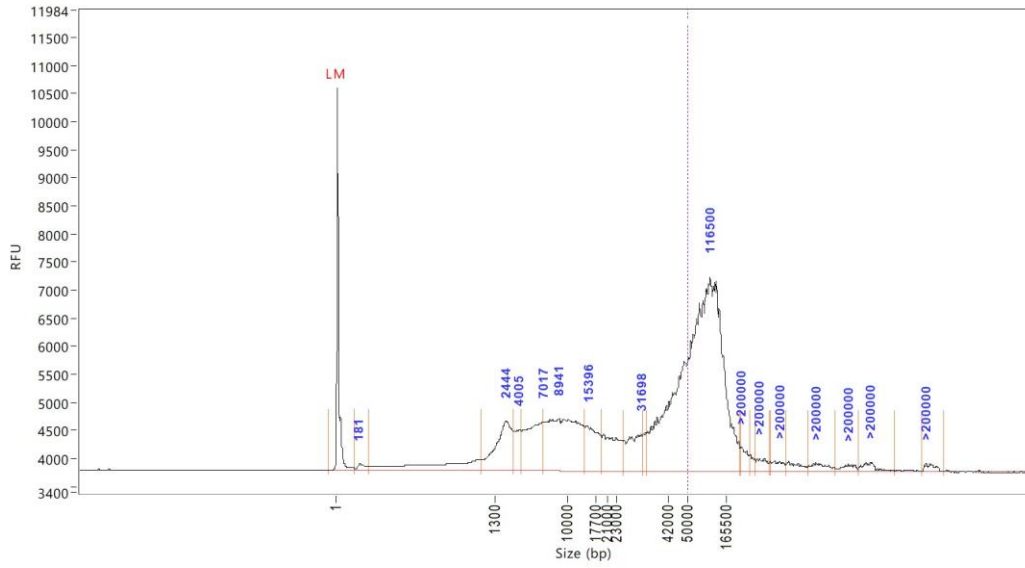


Figure S.12.: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 9.

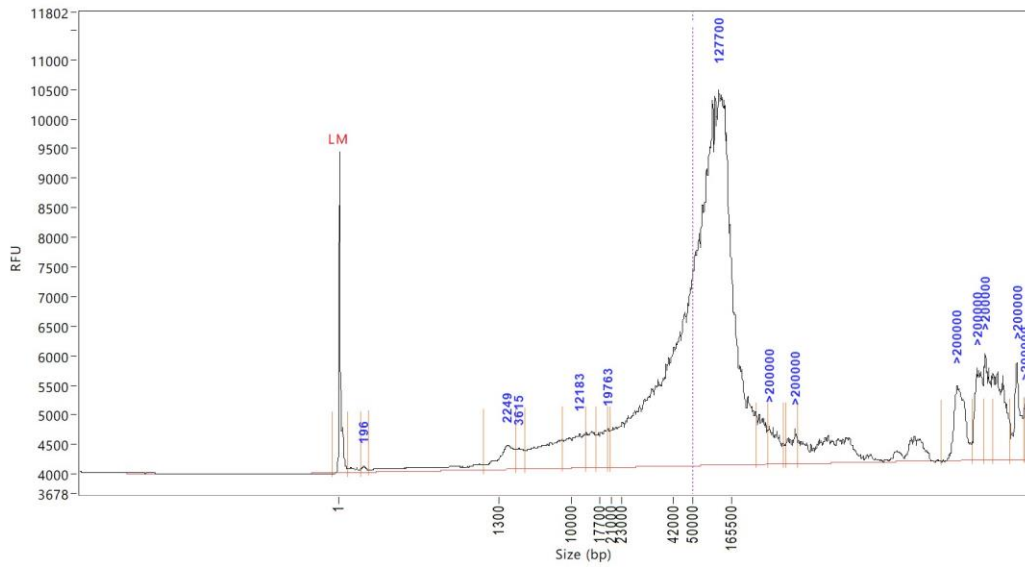


Figure S.13.: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 10.

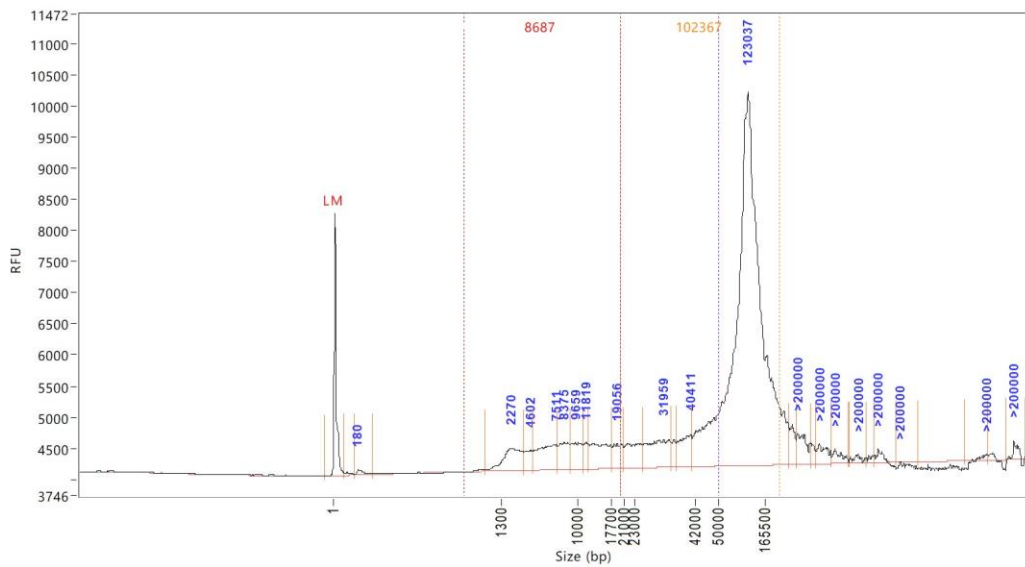


Figure S.14.: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 11.

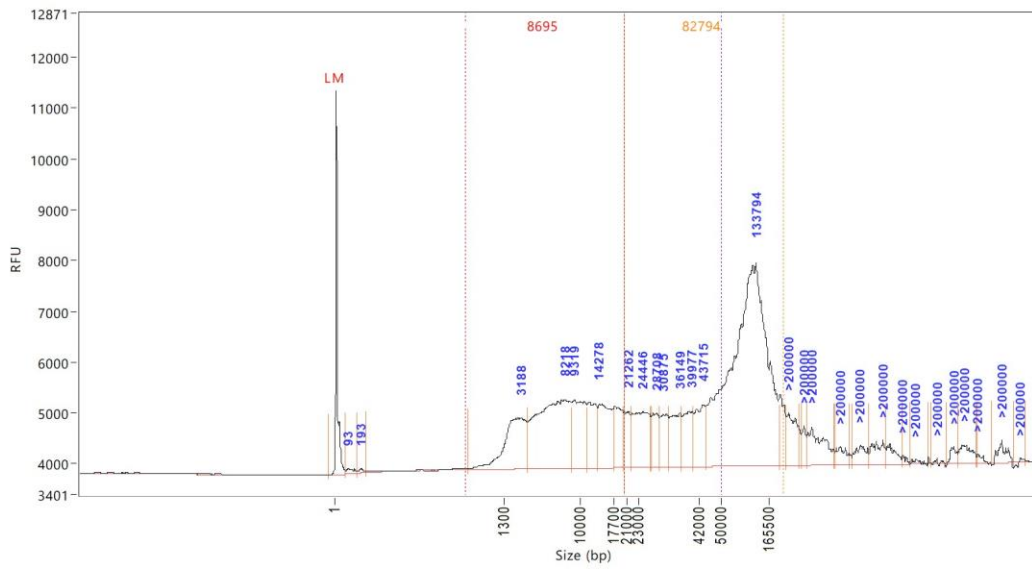


Figure S.15.: Femto Pulse DNA sizing QC analysis of *Taxus baccata* L. sample 12.

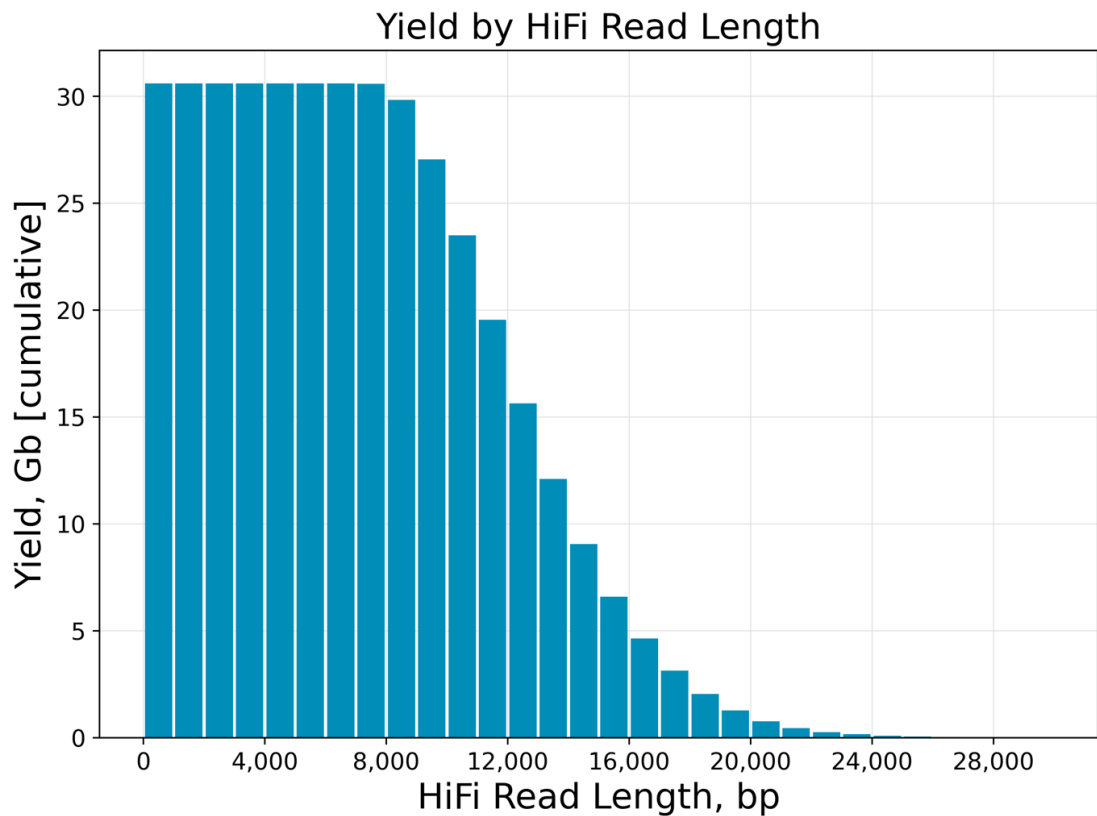


Figure S.16.: Sequencing results of *Taxus baccata* DNA using PacBio HiFi sequencing (Run ID 21, Table 8). The total yield of circular consensus sequencing (CCS) reads of this SMRT cell was 30.6 Gb. These CCS reads have been called with the PacBio SMRT link pipeline (SMRTlink version 11.0.0.146107)

Tables

Table S.1: High-molecular-weight DNA extraction protocol steps of Faxinus excelsior.

sample	Grinding	Washing steps (NIB), Isolation	Sorbitol wash	Washing steps (NIB), Extraction	Lysation	Centrifugation force
I	TissueRuptor	4	1	4	2 h	5000 x g
II	TissueRuptor	4	1	4	2 h	5000 x g
III	Mortal + Pestel	4	1	4	2 h	7000 x g
VI	TissueRuptor	4	1	4	2 h	5000 x g

Table S.2: High-molecular-weight DNA extraction protocol steps of Taxus baccata.

sample	Grinding	Washing steps (NIB), Isolation	Sorbitol wash	Washing steps (NIB), Extraction	Lysation	Centrifugation force
1	Mortal + Pestel	1	2	4	1 h	3000 x g
2	Mortal + Pestel	1	2	4	1.15 h	3000 x g
3	Mortal + Pestel	1	2	4	1.15 h	3000 x g
4	Mortal + Pestel	1	2	4	1.30 h	3000 x g
5	Mortal + Pestel	1	2	4	1.30 h	3000 x g
6	Mortal + Pestel	1	2	4	1 h	3000 x g
7	Mortal + Pestel	1	2	4	1 h	3000 x g
8	Mortal + Pestel	1	2	4	1.15 h	3000 x g
9	Mortal + Pestel	1	2	4	1.15 h	3000 x g
10	Mortal + Pestel	1	2	4	1 h	3000 x g
11	Mortal + Pestel	1	2	4	1 h	3000 x g
12	Mortal + Pestel	1	2	4	1.20 h	3000 x g

Table S.3: High-molecular-weight DNA extraction quality and quantity values after applying the protocol to Fraxinus excelsior. For each measurement, 1 µl DNA was used. The green-marked rows are the samples that were used for ONT Sequencing.

	Before Short Read Eliminator				After Short Read Eliminator			
	Nanodrop p (ng/µl)	A260/A280	A260/A230	Qubit (ng/µl)	Nanodrop p (ng/µl)	A260/280	A260/A230	Qubit (ng/µl)
1	127.7	1.61	1.31	61.4	33.6	1.26	0.55	41.0
2	359.9	1.76	1.59	149	162.3	1.43	0.99	87.0
3	912.3	1.58	1.13	314.0	174.1	1.63	1.45	131.0
4	199.00	1.67	1.39	122	59.5	1.61	1.40	83.8

5	143	1.78	1.68	250	56.9	1.76	1.88	69.2
6	217.4	1.81	1.64	330	36.3	1.92	2.12	424

General Discussion

The overall aim of this PhD project was to unravel the genomic basis of variation in ash dieback in common ash (*Fraxinus excelsior*). To achieve this goal, I pursued four main research topics. i) **The identification of full-sibling families.** Using SSR and SNP markers, I identified full-sibling families in four single-tree progenies of potential ash dieback-tolerant mother trees (verification of hypothesis i; General introduction). ii) **The generation of high-density linkage maps.** High-density linkage maps were created for the identified full-sibling families. These maps revealed structural genomic differences between the families and sexes, providing insights into their genetic makeup and recombination rates. iii) **Quantitative Trait Locus (QTL) Analysis.** QTL analyses for ash dieback susceptibility identified one significant locus in one of the four full-sibling families that is associated with stem collar necrosis, explaining 15% of the phenotypic variance (verification of hypothesis ii and for one genomic region of hypothesis iii; General introduction). This locus might be linked to a protein involved in fungal response, highlighting a potential genetic basis for susceptibility to ash dieback. iv) **High-molecular-weight DNA extraction.** I developed and optimized a protocol for extracting high-molecular-weight DNA from ash, overcoming the challenges posed by its complex biochemical compositions. This advancement enabled long-read sequencing on the Oxford Nanopore Technologies platform and paves the way for future association studies based on pangenomes.

Genetic variation of plant disease resistance

The mechanisms of plant's resistance to pathogen infection and the underlying genetic basis are complex and multifaceted fields that are crucial to understand in forest improvement and conservation efforts (Naidoo et al., 2019; Wilson et al., 2023).

Disease resistance is the reduction of pathogen growth on or in the plant (and hence a reduction of disease), while the term disease tolerance describes plants that exhibit little disease damage despite substantial pathogen levels. Plant disease resistance protects plants from pathogens in two ways: (i) by pre-formed structures and chemicals, and (ii) by infection-induced responses of the immune system. Pre-

formed structures and chemicals may comprise plant cuticle/surface, plant cell walls, antimicrobial chemicals and peptides, enzyme inhibitors and detoxifying enzymes breaking down pathogen-derived toxins. Inducible post-infection plant defenses depend on specific receptors.

The plant immune system carries two interconnected tiers of receptors, one most frequently sensing molecules outside the cell and the other most frequently sensing molecules inside the cell. Both systems sense the intruder and respond by activating antimicrobial defenses in the infected cell and neighboring cells. In some cases, defense-activating signals spread to the rest of the plant. The two systems detect different types of pathogen molecules and classes of plant receptor proteins (Jones and Dangl, 2006; Dodds and Rathjen, 2010; Striganavičiūtė et al., 2021). The first tier is primarily governed by pattern recognition receptors (PRRs) that are activated by recognition of evolutionarily conserved pathogen or microbial-associated molecular patterns (PAMPs or MAMPs). Activation of PRRs leads to intracellular signaling, transcriptional reprogramming, and a complex output response that limits colonization. The system is known as PAMP-triggered immunity or as pattern-triggered immunity (PTI) (Jones and Dangl, 2006; Couto and Zipfel, 2016; Li et al., 2016).

The second tier, primarily governed by R gene products, is often termed effector-triggered immunity (ETI). ETI is typically activated by the presence of specific pathogen "effectors" and then triggers strong antimicrobial responses which are reliant on resistance (R) genes and are activated by specific pathogen strains. Major R genes often confer high levels of resistance to specific pathogens. Plant ETI often causes an apoptotic hypersensitive response.

In addition to PTI and ETI, plant defenses can be activated by the sensing of damage-associated compounds (DAMP), such as portions of the plant cell wall released during pathogenic infection (Molina et al., 2021). Plant immune system activity is regulated in part by signaling hormones such as salicylic acid (SA), jasmonic acid (JA) and ethylene (Moore et al., 2011; Bürger and Chory, 2019). In most plants, SA and JA trigger defenses against fungal biotrophic and necrotrophic

pathogens, respectively, in an antagonistic manner (Bari and Jones, 2009; Robert-Seilaniantz et al., 2011).

Plant sRNA pathways are understood to be important components of PTI and ETI (Jin, 2008; Padmanabhan et al., 2009). Plants can transport viral RNAs, mRNAs, miRNAs and siRNAs systemically through the phloem (Kehr and Buhtz, 2008). Advances in genome-wide studies revealed a massive adaptation of host miRNA expression patterns after infection by different fungal pathogens (e.g., Lu et al. (2007)).

Disease outcome is determined by the three-way interaction of the pathogen, the plant individual (specific genotype), and the environmental conditions (an interaction known as the disease triangle) (Scholthof, 2007). Individual differences in plant disease resistance are often incremental or quantitative rather than qualitative. The term quantitative resistance (QR) refers to plant disease resistance that is controlled by multiple genes and multiple molecular mechanisms that each have small or minor effects on the overall resistance trait (Cowger and Brown, 2019). QR is important in plant breeding because the resulting resistance is often more durable (effective for more years), and more likely to be effective against most or all strains of a particular pathogen. QR is typically effective against one pathogen species or a group of closely related species, rather than being broadly effective against multiple pathogens (Cowger and Brown, 2019).

There are several examples of pathogen outbreaks in trees and their consequences, e.g., Dutch elm disease caused by a sac fungus (Ascomycota) (Nielsen and Kjær, 2010), sudden oak death caused by an oomycete (*Phytophthora ramorum*) (Grünwald et al., 2019), or chestnut blight caused by *Cryphonectria parasitica* (Griffin, 2000). Several factors play a role in the outcome of pathogen infection, including (1) the severity of pathogen pressure, (2) the likelihood of initial establishment or repellence, (3) the success of subsequent pathogen development, and (4) the level of tolerance to tissue invasion (Budde et al., 2016). If the pathogen coevolves with the tree, the defense mechanisms can propagate an 'arms race' between the tree and the pathogen (Meaux and Mitchell-Olds, 2003). Often, the pathogen was first living mutualistic or co-existing without any harm to the tree,

but unfortunately, especially endophytes can switch between a mutualistic and a pathogenic lifestyle (Delaye et al., 2013). For many pathogenic pests, the pathogen is an invasive species (e.g., *H. fraxineus*), and long-term coevolution is not given, only coadaptation (Langer et al., 2022).

Trees are biologically complex species with long-generation times. Genetic adaptation is an arms race with quickly changing environmental factors and adaptation of the tree and involvement of pathogens (Neophytou et al., 2022). Trees have local systemic responses induced by an infection of the pathogen. The responses can be controlled by complex genetics (quantitative resistance). As described above, or single genes (qualitative resistance) (Kovalchuk et al., 2013). The underlying resistance genes interact with the response to induce effector-triggered immunity (Naidoo et al., 2019).

To approach pathogens and disease response, it is necessary to apply multiple techniques and methodological approaches. To mitigate threats to trees, a combination of methods is needed: Bioinformatics, sets of genes and complete genomes (with different sequencing approaches), mRNA, proteins, genome mapping, metabolites, epigenomic modification, evolutionary mechanisms, and community- and metagenomics (Plomion et al., 2016). It is desired to investigate resistant or less susceptible genotypes in forestry to reduce losses and increase the frequency of advantageous alleles for traits in the target populations (Grattapaglia et al., 2018). This study contributed to this effort by the identification, repeated phenotyping and sequencing of *F. excelsior* mother trees potentially tolerant against ash dieback - and related full-siblings (Chapter 1).

Genome-wide approaches for understanding the genomic architecture of plant disease resistance

In recent years, advances in genomic technologies, such as cost-efficient short-read sequencing as well as long-read sequencing, have markedly accelerated the discovery of markers or loci associated with pathogen resistance (Muchero et al., 2018). These technologies allowed the generation of plant reference genomes as the basis for detailed analysis of individual genetic variation and the identification of

genomic loci, genes or specific genetic markers linked to resistance traits. Integrating these markers into genomic prediction models (Chen et al., 2023) and subsequently into forest management holds the potential to develop pathogen-resistant varieties more efficiently, ensuring sustainable forestry practices and the preservation of valuable tree species.

During the last decades, work mainly focused on the understanding the genomic architecture of qualitative disease resistance mediated by a few genes conferring an almost complete resistance, while QR remains poorly understood despite the fact that it represents the predominant and more durable form of resistance in natural populations (Chauveau and Roby, 2024). One reason is that the selective pressure on any single gene is reduced by OR, thereby slowing down the pathogen's ability to adapt (Snieszko and Liu, 2023).

QTL mapping is an genome-wide approach to identify QTLs, which represent regions of the genome that contribute to resistance in a quantitative or qualitative manner, involving multiple or a view genes or a single gene (Torello Marinoni et al., 2018; Gebhardt, 2023). Identifying and characterizing these genetic loci is essential for understanding the molecular mechanisms underlying disease resistance and developing strategies to enhance resistance in plant populations. In the history of QTL used in forest trees, many QTLs were reported with supposedly 'major effect' sizes. Unfortunately, a lot were proved to be overestimated in effect sizes and underestimated in numbers (Grattapaglia et al., 2018). With time, the ability to use larger sample sizes and whole genome sequencing led to improvements of the outcome of QTL mapping. The limitations of QTL detection to be restricted to single families or single genetic crosses were overcome by genome-wide association studies (GWAS). However, , the downside of GWAS in forest trees is that only a few polymorphisms with very modest effects have been detected, and these largely lack independent validation, which is crucial for the scientific credibility of GWAS results.

In this thesis, we conducted a QTL analysis for ash dieback susceptibility to gain a deeper understanding of its genetic basis within a confined set of related individuals (Chapter 3) representing four families (Chapter 1). The aim was not to identify

genetic markers but to elucidate the number and effect sizes (i.e., the genetic architecture) of natural gene variants involved in variation in susceptibility within a family. Complex genetic architectures in global populations, as the ones often used for GWAS, may not be representative for what is happening on a small scale in a specific forest with a limited number of unrelated parental trees. Our analysis identified many SNPs and one significant QTL at chromosome 16 explaining a proportion of the phenotypic variation of stem necrosis (Chapter 3).

However, our results are insufficient to effectively estimate individual breeding values for a breeding program. QTL, compared to GWAS, gives insight into a less complex context. While GWAS performs on the whole species, QTL reflects a smaller portion of information of one family. These findings underscore the need for ongoing research on the genomic architecture of ash dieback and integration of genomic patterns to develop robust strategies for managing ash dieback and other forest tree diseases.

Although QTL mapping and GWAS provide resistance trait-associated SNPs, which could be applied in marker assisted selection (MAS), genomic selection (GS) may be an interesting alternative for the selection of resistant trees in future breeding programs. Degen and Müller (2023a) performed a comparative simulation study between MAS and GS, using the software ‘SNPscan breeder’ (Degen and Müller, 2023b). GS using gBLUP performed best in nearly all simulated scenarios. MAS based on GWAS results outperformed GS only when allelic effects were estimated in large populations (approximately 10,000 individuals) of unrelated individuals. The primary reason for this failure of MAS was the underestimated complexity of the genetic architecture for most traits, which involved numerous causal SNPs, most having small effects. In most studies, the number of causal gene markers was significantly underestimated, while their effect sizes were overestimated.

Consequently, even though research focused on the genetic dissection of quantitative traits, these efforts have not yet translated into practical applications in operational tree breeding (Grattapaglia et al., 2009; Grattapaglia, 2017). Research should focus on understanding the genomic basis of the pathogen resistance and susceptibility before developing and using genetic markers. The gap of knowledge

about the genomic basis should be closed before trying to use markers for MAS in practical application, especially in case of ash dieback disease. This is underlined by the study of Menkis et al. (2020), where molecular markers were tested against traditional monitoring of healthy-looking trees. The results revealed, that the markers only showed a moderate capacity for identification of ash dieback susceptibility, and morphological monitoring was more effective. The importance of detecting genome-wide patterns of pathogen susceptibility is crucial, and the importance should be communicated to the greater public.

Do we understand the genomic basis of ash dieback?

Most studies on the genomic bases of ash dieback susceptibility have used GWAS (Stocks et al., 2019; Doonan et al., 2023; Meger et al., 2024b). They employed several different phenotypes, such as crown damage, autumn leaf yellowing, and overall health status. So far, no GWAS using stem collar necrosis which was used as phenotype in this study (Chapter 3) or comparable phenotypes has been done.

Sollars et al. (2017) provided a comprehensive analysis of the genomic landscape and genetic diversity of *F. excelsior*. The study identified candidate genes that may be involved in resistance to ash dieback. The reference genome developed by Stocks et al. (2019) was also used in the study. The study used association mapping to link genetic markers with phenotypic traits related to ADB resistance. This method helped to identify significant loci contributing to the resistance phenotype. Key loci identified include regions on several chromosomes significantly associated with reduced disease severity. Population structure analysis indicated distinct genetic clusters, which are crucial for understanding the spread and impact of resistance traits. This involved gene expression analysis and other molecular biology techniques to verify the genes' involvement in the resistance response.

In 2021, the Polish (chromosome-level) reference genome was published. The same workgroup published a study that utilized genomic prediction models to estimate the susceptibility to ADB. The study reported high prediction accuracies for susceptibility traits, demonstrating the effectiveness of genomic prediction in identifying resistant individuals within ash populations. Accuracy varied depending on the model used and the genetic architecture of the trait. Unfortunately, the

model is not continuously stable. Only within the studied population did the GWAS model show sufficient results, thus it is highly inconvenient for the practical use. Single nucleotide polymorphism (SNP) arrays were employed to genotype ash populations. The study highlighted the importance of dense marker coverage for accurate genomic predictions. Key genetic markers associated with resistance were identified (Meger et al., 2024b).

While QTL mapping identifies specific loci associated with resistance, genomic prediction can integrate this information into a more comprehensive model, mainly relying on pairwise relationship between all individuals, to predict resistance across larger populations. The candidate genes identified in the Sollars and Stocks studies can be cross-referenced with the loci identified in our QTL analysis (Chapter 3) in future studies. This comparison might help pinpoint specific genes within the significant QTL region responsible for SCN resistance. All three studies have shown that the SNPs identified mostly apply to the specific trial site and are not transferable. Since this is the case in multiple studies with different trial sites, generating general markers to identify susceptibility does not work. Understanding the greater genomic basis of susceptibility of ash dieback should be prioritized.

For ash trees, it is predicted that despite the intensive investigations into useful conservation strategies ash will probably recover slowly through natural selection. The typically large effective population sizes and high-standing genetic variation in tree species will play a crucial role, this allows a fast shift in adaptive allele frequencies, when exposed to high selection pressure (Petit and Hampe, 2006; Barrett and Schluter, 2008). With the high selection pressure from *H. fraxineus*, ash trees may shift their adaptive allele frequencies (Barrett and Schluter, 2008), or this shift may already occur. Strong selection pressure is key to this theory, as it can favour individuals with low susceptibility. When the mortality rate is high, natural regeneration and selection for resistant offspring become possible, promoting the survival of individuals with advantageous traits (Budde et al., 2016). Adaptation from standing genetic variation will likely lead to faster evolution, with many small effect alleles rising in frequency, compared to adaptation to environmental change or pathogens by new mutations (Barrett and Schluter, 2008). Standing variation has

the advantage of many possibly beneficial genetic changes (McGregor et al., 2007). In *F. excelsior*, it is estimated that the frequency of genotypes with high resistance is expected to be relatively low, probably in the range of 1-5% (McKinney et al., 2014). The theory that ash is potentially already undergoing rapid evolution for higher levels of susceptibility can be underscored by multiple studies (McKinney et al., 2011; Lobo et al., 2014; McKinney et al., 2014; Pliūra et al., 2014; Lobo et al., 2015; Muñoz et al., 2016; Stocks et al., 2019; Meger et al., 2024b). It is reported that genetic variation of resistance is implemented in the Baltic areas, which have been the ash dieback hotspots at the beginning of the outbreak, and the mortality rate is decreasing (Pliura A et al., 2011). Besides the actual tree, mutualistic endophytes, bacteria and other fungi also undergo selection (Arnold et al., 2003). Since the life cycle of these organisms is much shorter than that of trees, they have an equal time span as *H. fraxineus*. Adaptation from standing genetic variants can be adapted much quicker. This can be additionally beneficial to the susceptibility of ash dieback. Fascinatingly, *H. fraxineus* has a closely related fungus, *Hymenoscyphus albidus*, native to Europe. It is a harmless decomposer for ash leaves. The comparison of the genomes of the two fungi revealed differences in secondary metabolite and transposable element repertoires. These differences may contribute to the pathogenicity of *H. fraxineus*. Specifically, *H. fraxineus* has an expanded repertoire of specialized metabolites and transposable elements associated with virulence and adaptation to different hosts (Elfstrand et al., 2021). These genomic differences may underlie its ability to cause dieback in ash, unlike its non-pathogenic relative *H. albidus*. Genetic basis is not restricted to the host; understanding the genomic basis of virulence from the pathogen is also essential.

Advancements in ash genomics

The construction of high-density genetic linkage maps in this study (Chapter 2) represents an important advancement in the genomics of *Fraxinus excelsior*. The high level of collinearity between our maps and the reference genome confirms the reliability of our approach. At the same time, the identified discrepancies underscore the importance of continuous refinement of reference genomes. These linkage maps enhance our understanding of the genomic structure of ash and

provide a robust framework for future genetic studies, including identifying QTLs associated with ash dieback susceptibility or other important traits.

The bias in the number of variants between paternal and maternal data could cause a higher count of mutations in the maternal genetic map. A dense marker coverage is essential for precise genetic mapping. Genetic maps gain on markers per cM (saturation) with more markers used (Colasuonno et al., 2021). Physical mapping, that is, mapping to a reference genome, has even higher requirements for marker saturation (Plomion et al., 2016). Physical maps require tens to hundreds of thousands of markers to achieve a high resolution on the whole genome to ensure only minimal gaps. High-density markers ensure that every region of the genome is accounted for and that the map accurately reflects the physical layout of the genome. Genetic maps can be constructed with hundreds to a few thousand markers. The distances are measured in centimorgans (cM), which are not directly proportional to physical distances. Genetic maps can tolerate larger gaps because they are more about the overall order of markers rather than representation of each genomic region. The first RFLP map of maize had approximately 400 markers (Helentjaris, 1987); nowadays, genetic maps consist of many more markers. For example, in *Hevea brasiliensis* (rubber tree), an ultrahigh-density genetic map of 200,000 SNP markers was constructed (Wu et al., 2022). Several thousand markers are currently used; 2,796 markers were used in tropical maize to create a physical map with the program BioNano IrysView (Pendleton et al., 2015; N. Yang et al., 2019). The recombination rate is an important metric which explains the frequency at which recombination occurs between two loci on a chromosome. Understanding recombination patterns can help predict how traits might be inherited, as it allows linked loci to be considered. Genetic maps reveal areas with high and low recombination rates. Physical maps can help pinpoint the exact locations of these hotspots and cold spots, providing insights into genome architecture and recombination mechanisms. In *Arabidopsis*, it is already shown that clusters of recombination rates influence disease-resistance genes (Choi et al., 2016). So far, we have investigated the differences in maternal and paternal recombination rates (Chapter 3, Figure 2). Further, it will be interesting to investigate the recombination across each chromosome and compare the results between maternal and paternal

data. An assessment could be the R package 'MareyMap' (Siberchicot et al., 2017). It relies on comparing genetic and physical maps of a chromosome to estimate local recombination rates. The output will complete the data information about paternal and maternal differences. When physical and genetic maps are compared, the concordance between genetic and physical distances is assessed. Regions with high genetic map distances should correspond to large physical distances. Discrepancies between the maps can indicate the presence of structural variants (Chapter 2) or regions of suppressed recombination.

Whole-genome alignments of complete genome assemblies are the key to investigating genome rearrangements that impact genetic diversity, phenotypes, recombination, and local adaptation. This can be achieved using long-read Oxford Nanopore Technologies (ONT) and PacBio high-fidelity (HiFi) technologies, which facilitate the identification of large and complex structural variants (SVs). In the thesis, we achieved the long-read sequence with ONT of *F. excelsior* (Chapter 4). Two of the four mother trees have already been successfully sequenced by ONT using the protocol for HMW DNA extraction developed in the thesis (Chapter 4, Krautwurst et al., 2024). Earlier, McKinney et al. (2014) observed that the genes that have an additive effect on ash dieback susceptibility are more likely from quantitative resistance (multiple genes) and not one single gene. This led to the decision that we want to create a pan-genome of the potential ash dieback-tolerant mother trees as a basis for future association studies. Pan-genomes, which compile multiple genomes representing the diversity within a species, offer more comprehensive insights into genetic variation than a single reference genome. It will further help better understand the natural phenotypic diversity of *F. excelsior*. Pan-genomes offer the opportunity to get an insight into large and complex SVs. This allows a more complete view of the diversity in the *F. excelsior* genome (Lian et al., 2024). Our specific case with the four mother trees will only represent the ash in northern Germany (Mecklenburg Pomerania). Since the word pan, a Greek word that means 'whole', the ideal outlook for common ash would be to use individuals from multiple areas, inside countries' borders, or, even better, from multiple countries (Schreiber et al., 2024). For ash, differences in susceptibility may be considered for the selection of individuals. The group of Laura Kelly and Richard

Buggs at Kew Botanical Gardens is currently working on a pan-genome for ash trees across the UK and Europe (<https://ashgenome.org/pan-genome/>).

Pangenomes are being generated for an increasing number of model and non-model species. For the important model plant *Arabidopsis thaliana*, 69 individuals were used to create a pan-genome from diverse geographical locations. The pan-genome differentiates core genes, present in all individuals, from dispensable genes, that occur only in some individuals. This differentiation helps to identify essential genetic elements and the variation contributing to phenotypic diversity. By integrating pan-genome data with phenotypic information, it is possible to link specific genetic variants to observable traits. The study uncovered numerous structural variants, such as insertions, deletions, and duplications, contributing to genetic and phenotypic variation within *A. thaliana*. Integrating the pan-genome data with phenotypic information made it possible to link specific genetic variants to functional traits, providing insights into how genetic diversity drives phenotypic diversity (Lian et al., 2024). So far, creating pan genomes is still challenging, from tissue sampling, isolation of high-molecular-weight DNA, to data management (Schreiber et al., 2024). In the context of forest trees, obtaining HMW DNA is particularly challenging due to their recalcitrance caused by secondary metabolites, which appear important for pest and disease resistance in perennial species (Kalendar et al., 2021; Jones et al., 2023). However, for *F. excelsior*, we can now extract high quality and quantity of HMW DNA (Chapter 4). Furthermore, with the existing genetic maps (Chapter 2), we will be able to identify more SVs consistent in all four families and unique within the families.

Integrating genomic advances in the practical management of ash dieback disease

Transferring the research data into practical applications is often not simple. One practical application in the phenological mismatch between the pathogen and its new host may result in disease escape of host genotypes in the extreme ends of the natural variation in growth rhythm (Budde et al., 2016). Avoidance of the pathogen has already been reported for the pathosystem *Ulmus* and ‘Dutch Elm Disease’. Early flushing can allow disease avoidance due to a phenological/physiological mismatch

with the occurrence of the pathogen vector (Ghelardini and Santini, 2009). Similar strategies have been reported for ash as well. *F. excelsior* shows evidence of an association between budburst, senescence and susceptibility (Bakys et al., 2013; Pliura et al., 2016; Andrić and Kajba, 2017; Doonan et al., 2023). This connection might be used as a tool for tree breeding.

The introduction of genome editing opens the door to new possibilities in forest trees, especially the clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR-associated protein 9 (Cas9) method. This highly efficient and effective technique has been used to implement targetable changes at specific places in a forest tree's genome (Bruegmann et al., 2024). The first edited tree was *Populus x tomentosa* (Di Fan et al., 2015). For successful editing in trees, three steps are involved: (1) establishment of an in vitro culture system that can regenerate plant shoots, (2) establishment of a protocol for the transfer of the Cas/gRNA into living cells, i.e., by classical genetic transformation, and (3) establishment of a protocol for genome editing. Hebda et al. (2021) reported progress in genetic engineering techniques for ash, showing successful genetic transformation of ash using *Agrobacterium tumefaciens*. This transformation involved callus tissue derived from embryos of *F. excelsior* and was demonstrated using the β -glucuronidase (GUS) reporter system. The study confirmed the formation of stable transgenic callus lines through RT-PCR experiments, ensuring the stable expression of the transgene.

The three genes potentially involved in fungal defense (TLP2, AT17.1, GST11), which we identified as closely associated with the QTL identified in our study (Chapter 3), could be interesting future targets for CRISPR/Cas9 editing in *F. excelsior* to potentially confer enhanced tolerance. AT17.1 potentially encoding a PADRE domain-including protein is of special interest in that respect as it seems to be functionally connected to other proteins involved in fungal responses (described in Chapter 3). Further, a tubby-like protein similar to TLP2 was identified. Tubby-like proteins have been reported to have an important role in multiple physiological and developmental processes and various environmental stress responses (Bano et al., 2022). Still, the functional mechanisms are not fully clear, especially in plants. Understanding the function and regulation of PADRE genes (Didelon et al., 2020)

could inform the design of CRISPR/Cas9 strategies to enhance disease resistance in *F. excelsior* by targeting or modifying these genes.

The functional validation methods used by Stocks et al. (2019) could be applied to our candidate genes to confirm their role in SCN resistance. Gene expression analysis, knock-out experiments, and other molecular biology techniques could help to validate the function of these genes. Other studies identified genes associated with ash dieback susceptibility as well (Harper et al., 2016; Downie, 2017; Sollars et al., 2017; Chaudhary et al., 2020). For example Downie (2017) attempted to identify genetic markers for low susceptibility using associative transcriptomics in ash. Eight of these encoded MADS-box proteins typical of transcription factors. Moreover, 2 genes encoding cinnamoyl-coenzyme A (CoA) reductase 2-like were identified. These genetic markers were identified using trees that had been exposed to the disease for up to 15 years. The phenotypes used for the study varied and were not directly associated with stem collar necrosis. This could be the reason why our findings of the PADRE-domain including proteins or Tubby-like proteins could not be aligned with prior studies. Despite the lack of corroborating results, finding the PADRE domain that is specifically connected with fungal responses (Didelon et al., 2020) warrants further study.

In the case of ash dieback, it has become increasingly evident that a structured approach to resources and strategies is necessary to manage the disease, especially when it spreads across multiple countries. The genomic and genetic basis, genetic architecture, and reference genome analysis must be integral to conservation strategies. Essential facilities for such programs include greenhouses, seed handling and cold storage capacity, inoculation infrastructure, field testing sites, database capabilities, and seed orchard development areas. Although there are existing programs for gene conservation, they often lack long-term funding and evolve slowly due to limited governmental and international commitment (Koskela et al., 2013; Fugeray-Scarbel et al., 2023).

The challenge is to apply scientific knowledge of conservation genetics to ensure recommendations are practical and implementable across different countries. The current attempts concentrate on seedlings/seedbanks, breeding programs, MAS

and GS (Sniezko and Koch, 2017). The methods have shown sufficient research outcomes in the history of pathogen and disease resistance programs (Thavamanikumar et al., 2013; Sniezko et al., 2020b; Pike et al., 2021). One general attempt is creating seed banks with less susceptible trees. Unfortunately, it comes with some downsides. Even when the screening of the seed donors is done with a high level of quality, the seedlings will have varying levels of susceptibility. Seedling survival from the orchard will be higher than in natural stands, but aspects such as endophytes, bacteria and other fungi play a role in susceptibility. Ash dieback is a complex disease, and there still are gaps in the knowledge of the signs of actual resistance genetically and morphologically. By choosing only seedlings based on phenotypical aspects, the survival rate of the seedlings can be lowered. Since resistance can be rare, hundreds to thousands of candidate trees (or their progenies) might need to be screened to provide enough resistant trees for the diverse genetic base required for reforestation or seedbanks. MAS showed reasonable results in breeding programs in other plant species; unfortunately, it is debatable and could not fulfil expectancy in forestry (Grattapaglia and Kirst, 2008; Kiszonas and Morris, 2018; Nadeem et al., 2018).

Conclusions and perspectives

In this study, we have successfully leveraged a combination of genetic, genomic and molecular biological methods to identify and analyze four full-sibling families from four potentially ash dieback-tolerant mother trees. Our research progressed through several key phases: (1) the identification of full-sibling families using SSR and SNP markers, (2) the generation of high-density linkage maps using the full-sibling families, revealing structural genomic differences among them, (3) a quantitative trait locus (QTL) analysis for ash dieback susceptibility, identifying one significant QTL associated with necrosis and explaining 15% of the phenotypic variance in one family, and (4) method development and optimization for the extraction of high-molecular-weight DNA from various tree species, overcoming challenges posed by recalcitrant tree species. A key gene linked to this QTL region was implicated in the fungal response. Our findings underscore the importance of understanding patterns in genomic pathogen susceptibility and the importance of

structural genomic variations, which can be more accurately detected using advanced sequencing technologies.

Looking forward, our study paves the way for several future directions. An overview of possible methods and future study opportunities is presented in Figure 3. Studies with firm attention (inner cycle) and studies that can be accumulated from the future results (outer cycle). The ambition is to conduct more Oxford nanopore sequencing with all four mothers, using the high-molecular-weight DNA protocol of Chapter 4. The produced sequencing results will provide several *de novo* assemblies. We want to use the assemblies for a higher resolution and improve the informational content of the genomic regions possibly involved in variation in ash dieback identified in the QTL analyses. Since there is a noticeable gap between maternal and paternal variants in the linkage groups, some anomalies could be caused by genotyping errors. We consider repeating the genotyping of the mother trees and full siblings of the four families established in this study (Chapter 1). A possible cause of the genotyping error or shifts in the data could be that the mother tree raw data from Illumina sequencing have a higher sequencing depth than the other samples.

Adapting the downstream pipeline and redoing the genotyping could solve some of these obstacles. The data is based on a reference genome from an ash tree from Poland (Kazimierz Wielki University). Before that, a highly fragmented reference genome with almost 90,000 contigs reported by Sollars et al. (2017) was widely used. The research group of Richard Buggs (Kew Royal Botanic Gardens), which led the generation of the first reference genome, published another reference genome in July 2024 (www.ashgenome.org; <http://oadb.tsl.ac.uk>). Since our genetic maps shown in Chapter 2, indicate that chromosomes 2, 22 and 23 are not completely accurate in the Polish reference, this could be an opportunity to use another reference genome to compare existing results and further improve them.

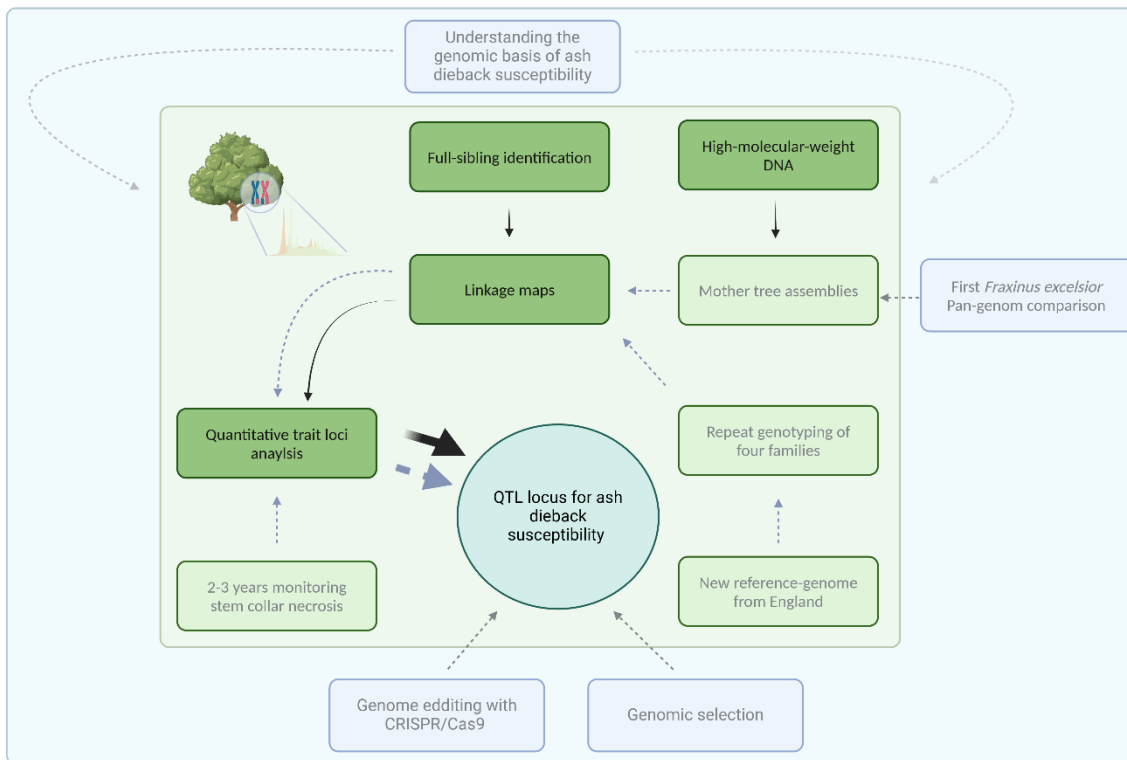


Figure 6: Overview of future directions and enhancements of the existing research results. Current results are featured in a darker green rectangle with black font. Future planned experiments and advances in lighter green rectangles with grey font and green background. The connection between each step is connected with arrows; the dotted grey arrow points out the future addition to each step. The blue background is future research that can be conducted with the ongoing research. Considering that a new reference genome of ash will be published by the group of Laura Kelly and Richard Buggs at Kew Botanical Gardens soon, it is an option to use this reference genome for genotyping.

Considering that the saplings were young, and the necrosis monitoring had only been conducted for three years, additional years of monitoring will be crucial. This may overcome environmental and other fluctuations and thus yield more robust phenotypic data. For *F. excelsior*, monitoring QTL phenotypes should be more extensive due to their complexity of traits and long development phase of the pathogen within the ash. At least 5-10 years of phenotypic data should be assessed to represent different growth stages and environmental conditions and to allow full development of the fungus within the trial site.

The combination of improved linkage mapping and longer monitoring of the trait will improve understanding of the genomic basis of ash dieback susceptibility. We have already shown that detecting a QTL for ash dieback susceptibility in one full-sib family is possible. An effect size of this magnitude suggests that the identified QTL has a major influence on the trait (Chapter 3). Our findings contribute to the greater outline of genomic ash dieback susceptibility and could be implemented for

studies of genomic selection (Chen et al., 2023). By integrating advanced sequencing technologies and continuing interdisciplinary collaborations, we can further unravel the genomic basis of disease resistance and contribute to the sustainable management and preservation of ash populations. These efforts may not only contribute to enhancing the resilience of ash trees but also provide a framework for addressing similar challenges in other tree species, thereby supporting broader forest conservation initiatives.

References

- Andrić, I., and Kajba, D. (2017). The impact of environmental drivers on narrow-leaved ash (*Fraxinus angustifolia* Vahl) budburst dates. *Šumar. list (Online)* 141, 13. doi: 10.31298/sl.141.1-2.1
- Arnold, A. E., Mejía, L. C., Kyllö, D., Rojas, E. I., Maynard, Z., Robbins, N., et al. (2003). Fungal endophytes limit pathogen damage in a tropical tree. *Proc Natl Acad Sci USA* 100, 15649–15654. doi: 10.1073/pnas.2533483100
- Bakys, R., Vasaitis, R., and Skovsgaard, J. P. (2013). Patterns and severity of crown dieback in young even-aged stands of european ash (*Fraxinus excelsior* L.) in relation to stand density, bud flushing phenotype, and season. *Plant Protect. Sci.* 49, 120–126. doi: 10.17221/70/2012-PPS
- Bano, N., Aalam, S., and Bag, S. K. (2022). Tubby-like proteins (TLPs) transcription factor in different regulatory mechanism in plants: a review. *Plant Mol Biol* 110, 455–468. doi: 10.1007/s11103-022-01301-9
- Bari, R., and Jones, J. D. G. (2009). Role of plant hormones in plant defence responses. *Plant Mol Biol* 69, 473–488. doi: 10.1007/s11103-008-9435-0
- Barrett, R. D. H., and Schluter, D. (2008). Adaptation from standing genetic variation. *Trends Ecol Evol* 23, 38–44. doi: 10.1016/j.tree.2007.09.008
- Bruegmann, T., Fendel, A., Zahn, V., and Fladung, M. (2024). “Genome Editing in Forest Trees,” in *A roadmap for plant genome editing*, eds. A. Riccio, D. Eriksson, D. Miladinović, J. Sweet, K. van Laere, and E. Woźniak-Gientka (Cham: Springer), 347–372.
- Budde, K. B., Nielsen, L. R., Ravn, H. P., and Kjær, E. D. (2016). The Natural Evolutionary Potential of Tree Populations to Cope with Newly Introduced Pests and Pathogens—Lessons Learned From Forest Health Catastrophes in Recent Decades. *Curr Forestry Rep* 2, 18–29. doi: 10.1007/s40725-016-0029-9
- Bürger, M., and Chory, J. (2019). Stressed Out About Hormones: How Plants Orchestrate Immunity. *Cell Host Microbe* 26, 163–172. doi: 10.1016/j.chom.2019.07.006
- Chaudhary, R., Rönneburg, T., Stein Åslund, M., Lundén, K., Durling, M. B., Ihrmark, K., et al. (2020). Marker-Trait Associations for Tolerance to Ash Dieback in Common Ash (*Fraxinus excelsior* L.). *Forests* 11, 1083. doi: 10.3390/f11101083
- Chauveau, C., and Roby, D. (2024). Molecular complexity of quantitative immunity in plants: from QTL mapping to functional and systems biology. *Comptes Rendus Biologies* 347, 35–44. doi: 10.5802/crbio.153
- Choi, K., Reinhard, C., Serra, H., Ziolkowski, P. A., Underwood, C. J., Zhao, X., et al. (2016). Recombination Rate Heterogeneity within Arabidopsis Disease Resistance Genes. *PLOS Genetics* 12, e1006179. doi: 10.1371/journal.pgen.1006179
- Colasuonno, P., Marcotuli, I., Gadaleta, A., and Soriano, J. M. (2021). From Genetic Maps to QTL Cloning: An Overview for Durum Wheat. *Plants* 10, 315. doi: 10.3390/plants10020315
- Couto, D., and Zipfel, C. (2016). Regulation of pattern recognition receptor signalling in plants. *Nat Rev Immunol* 16, 537–552. doi: 10.1038/nri.2016.77
- Cowger, C., and Brown, J. K. M. (2019). Durability of Quantitative Resistance in Crops: Greater Than We Know? *Annu Rev Phytopathol* 57, 253–277. doi: 10.1146/annurev-phyto-082718-100016

- Degen, B., and Müller, N. A. (2023a). A simulation study comparing advanced marker-assisted selection with genomic selection in tree breeding programs. *G3 (Bethesda)* 13. doi: 10.1093/g3journal/jkad164
- Degen, B., and Müller, N. A. (2023b). SNPscan breeder – a computer program to test genomic tools in breeding programs. *Silvae Genetica* 72, 126–131. doi: 10.2478/sg-2023-0013
- Delaye, L., García-Guzmán, G., and Heil, M. (2013). Endophytes versus biotrophic and necrotrophic pathogens—are fungal lifestyles evolutionarily stable traits? *Fungal Diversity* 60, 125–135. doi: 10.1007/s13225-013-0240-y
- Di Fan, Liu, T., Li, C., Jiao, B., Li, S., Hou, Y., et al. (2015). Efficient CRISPR/Cas9-mediated Targeted Mutagenesis in *Populus* in the First Generation. *Sci Rep* 5, 12217. doi: 10.1038/srep12217
- Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet* 11, 539–548. doi: 10.1038/nrg2812
- Doonan, J. M., Budde, K. B., Kosawang, C., Lobo, A., Verbylaite, R., Brealey, J. C., et al. (2023). Multiple, single trait GWAS and supervised machine learning reveal the genetic architecture of *Fraxinus excelsior* tolerance to ash dieback in Europe. *bioRxiv*, 2023.12.11.570802. doi: 10.1101/2023.12.11.570802
- Downie, J. A. (2017). Ash dieback epidemic in Europe: How can molecular technologies help? *PLoS Pathog* 13, e1006381. doi: 10.1371/journal.ppat.1006381
- Elfstrand, M., Chen, J., Cleary, M., Halecker, S., Ihrmark, K., Karlsson, M., et al. (2021). Comparative analyses of the *Hymenoscyphus fraxineus* and *Hymenoscyphus albidus* genomes reveals potentially adaptive differences in secondary metabolite and transposable element repertoires. *BMC Genomics* 22, 503. doi: 10.1186/s12864-021-07837-2
- Fugeray-Scarbel, A., Irz, X., and Lemarié, S. (2023). Innovation in forest tree genetics: A comparative economic analysis in the European context. *Forest Policy and Economics* 155, 103030. doi: 10.1016/j.forpol.2023.103030
- Gebhardt, C. (2023). A physical map of traits of agronomic importance based on potato and tomato genome sequences. *Front Genet* 14, 1197206. doi: 10.3389/fgene.2023.1197206
- Ghelardini, L., and Santini, A. (2009). Avoidance by early flushing: a new perspective on Dutch elm disease research. *iForest* 2, 143–153. doi: 10.3832/ifor0508-002
- Grattapaglia, D. (2017). “Status and Perspectives of Genomic Selection in Forest Tree Breeding,” in *Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement*, eds. M. Roorkiwal, M. E. Sorrells, and R. K. Varshney (Cham: Springer International Publishing; Imprint: Springer), 199–249.
- Grattapaglia, D., and Kirst, M. (2008). Eucalyptus applied genomics: from gene sequences to breeding tools. *New Phytol* 179, 911–929. doi: 10.1111/j.1469-8137.2008.02503.x
- Grattapaglia, D., Plomion, C., Kirst, M., and Sederoff, R. R. (2009). Genomics of growth traits in forest trees. *Curr Opin Plant Biol* 12, 148–156. doi: 10.1016/j.pbi.2008.12.008
- Grattapaglia, D., Silva-Junior, O. B., Resende, R. T., Cappa, E. P., Müller, B. S. F., Tan, B., et al. (2018). Quantitative Genetics and Genomics Converge to Accelerate Forest Tree Breeding. *Front Plant Sci* 9, 1693. doi: 10.3389/fpls.2018.01693

- Griffin, G. J. (2000). Blight Control and Restoration of the American Chestnut. *Journal of Forestry* 98, 22–27. doi: 10.1093/jof/98.2.22
- Grünwald, N. J., LeBoldus, J. M., and Hamelin, R. C. (2019). Ecology and Evolution of the Sudden Oak Death Pathogen *Phytophthora ramorum*. *Annu Rev Phytopathol* 57, 301–321. doi: 10.1146/annurev-phyto-082718-100117
- Harper, A. L., McKinney, L. V., Nielsen, L. R., Havlickova, L., Li, Y., Trick, M., et al. (2016). Molecular markers for tolerance of European ash (*Fraxinus excelsior*) to dieback disease identified using Associative Transcriptomics. *Sci Rep* 6, 19335. doi: 10.1038/srep19335
- Hebda, A., Liszka, A., Zgłobicki, P., Nawrot-Chorabik, K., and Lyczakowski, J. J. (2021). Transformation of European Ash (*Fraxinus excelsior* L.) Callus as a Starting Point for Understanding the Molecular Basis of Ash Dieback. *Plants* 10, 2524. doi: 10.3390/plants10112524
- Helentjaris, T. (1987). A genetic linkage map for maize based on RFLPs. *Trends in Genetics* 3, 217–221. doi: 10.1016/0168-9525(87)90239-3
- Jin, H. (2008). Endogenous small RNAs and antibacterial immunity in plants. *FEBS Lett* 582, 2679–2684. doi: 10.1016/j.febslet.2008.06.053
- Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286
- Jones, M. M., Nagalingum, N. S., and Handley, V. M. (2023). Testing protocols to optimize DNA extraction from tough leaf tissue: A case study in *Encephalartos*. *Appl Plant Sci* 11, e11525. doi: 10.1002/aps3.11525
- Kalendar, R., Boronnikova, S., and Seppänen, M. (2021). Isolation and Purification of DNA from Complicated Biological Samples. *Methods Mol Biol* 2222, 57–67. doi: 10.1007/978-1-0716-0997-2_3
- Kehr, J., and Buhtz, A. (2008). Long distance transport and movement of RNA through the phloem. *J Exp Bot* 59, 85–92. doi: 10.1093/jxb/erm176
- Kiszonas, A. M., and Morris, C. F. (2018). Wheat breeding for quality: A historical review. *Cereal Chem* 95, 17–34. doi: 10.1094/CCHEM-05-17-0103-FI
- Koskela, J., Lefèvre, F., Schueler, S., Kraigher, H., Olrik, D. C., Hubert, J., et al. (2013). Translating conservation genetics into management: Pan-European minimum requirements for dynamic conservation units of forest tree genetic diversity. *Biological Conservation* 157, 39–49. doi: 10.1016/j.biocon.2012.07.023
- Kovalchuk, A., Keriö, S., Oghenekaro, A. O., Jaber, E., Raffaello, T., and Asiegbu, F. O. (2013). Antimicrobial defenses and resistance in forest trees: challenges and perspectives in a genomic era. *Annu Rev Phytopathol* 51, 221–244. doi: 10.1146/annurev-phyto-082712-102307
- Krautwurst, M., Eikhof, A., Winkler, S., Bross, D., Kersten, B., and Müller, N. A. (2024). High-molecular-weight DNA extraction for broadleaved and conifer tree species. *Silvae Genetica* 73, 85–98. doi: 10.2478/sg-2024-0009
- Langer, G. J., Fuchs, S., Osewold, J., Peters, S., Schrewe, F., Ridley, M., et al. (2022). FraxForFuture—research on European ash dieback in Germany. *J Plant Dis Prot*, 1–11. doi: 10.1007/s41348-022-00670-z
- Li, B., Meng, X., Shan, L., and He, P. (2016). Transcriptional Regulation of Pattern-Triggered Immunity in Plants. *Cell Host Microbe* 19, 641–650. doi: 10.1016/j.chom.2016.04.011
- Lian, Q., Huettel, B., Walkemeier, B., Mayjonade, B., Lopez-Roques, C., Gil, L., et al. (2024). A pan-genome of 69 *Arabidopsis thaliana* accessions reveals a conserved genome structure throughout the global species range. *Nat Genet* 56, 982–991. doi: 10.1038/s41588-024-01715-9

- Lobo, A., Hansen, J. K., McKinney, L. V., Nielsen, L. R., and Kjær, E. D. (2014). Genetic variation in dieback resistance: growth and survival of *Fraxinus excelsior* under the influence of *Hymenoscyphus pseudoalbidus*. *Scandinavian Journal of Forest Research* 29, 519–526. doi: 10.1080/02827581.2014.950603
- Lobo, A., McKinney, L. V., Hansen, J. K., Kjaer, E. D., and Nielsen, L. R. (2015). Genetic variation in dieback resistance in *Fraxinus excelsior* confirmed by progeny inoculation assay. *For. Path.* 45, 379–387. doi: 10.1111/efp.12179
- McKinney, L. V., Nielsen, L. R., Collinge, D. B., Thomsen, I. M., Hansen, J. K., and Kjaer, E. D. (2014). The ash dieback crisis: genetic variation in resistance can prove a long-term solution. *Plant Pathol* 63, 485–499. doi: 10.1111/ppa.12196
- McKinney, L. V., Nielsen, L. R., Hansen, J. K., and Kjær, E. D. (2011). Presence of natural genetic resistance in *Fraxinus excelsior* (Oleraceae) to *Chalara fraxinea* (Ascomycota): an emerging infectious disease. *Heredity* 106, 788–797. doi: 10.1038/hdy.2010.119
- Meaux, J. de, and Mitchell-Olds, T. (2003). Evolution of plant resistance at the molecular level: ecological context of species interactions. *Heredity* 91, 345–352. doi: 10.1038/sj.hdy.6800342
- Meger, J., Ulaszewski, B., Pałucka, M., Kozioł, C., and Burczyk, J. (2024). Genomic prediction of resistance to *Hymenoscyphus fraxineus* in common ash (*Fraxinus excelsior* L.) populations. *Evol Appl* 17, e13694. doi: 10.1111/eva.13694
- Menkis, A., Bakys, R., Stein Åslund, M., Davydenko, K., Elfstrand, M., Stenlid, J., et al. (2020). Identifying *Fraxinus excelsior* tolerant to ash dieback: Visual field monitoring versus a molecular marker. *For. Path.* 50. doi: 10.1111/efp.12572
- Molina, A., Miedes, E., Bacete, L., Rodríguez, T., Mélida, H., Denancé, N., et al. (2021). Arabidopsis cell wall composition determines disease resistance specificity and fitness. *Proc Natl Acad Sci USA* 118. doi: 10.1073/pnas.2010243118
- Moore, J. W., Loake, G. J., and Spoel, S. H. (2011). Transcription dynamics in plant immunity. *Plant Cell* 23, 2809–2820. doi: 10.1105/tpc.111.087346
- Muchero, W., Sondreli, K. L., Chen, J.-G., Urbanowicz, B. R., Zhang, J., Singan, V., et al. (2018). Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant-pathogen interactions in a tree. *Proc Natl Acad Sci USA* 115, 11573–11578. doi: 10.1073/pnas.1804428115
- Muñoz, F., Marçais, B., Dufour, J., and Dowkiw, A. (2016). Rising Out of the Ashes: Additive genetic variation for crown and collar resistance to *Hymenoscyphus fraxineus* in *Fraxinus excelsior*. *Phytopathology* 106, 1535–1543. doi: 10.1094/PHYTO-11-15-0284-R
- Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., et al. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment* 32, 261–285. doi: 10.1080/13102818.2017.1400401
- Naidoo, S., Slippers, B., Plett, J. M., Coles, D., and Oates, C. N. (2019). The Road to Resistance in Forest Trees. *Front Plant Sci* 10, 273. doi: 10.3389/fpls.2019.00273
- Neophytou, C., Heer, K., Milesi, P., Peter, M., Pyhäjärvi, T., Westergren, M., et al. (2022). Genomics and adaptation in forest ecosystems. *Tree Genetics & Genomes* 18, 1–7. doi: 10.1007/s11295-022-01542-1
- Nielsen, L. R., and Kjær, E. D. (2010). Gene flow and mating patterns in individuals of wych elm (*Ulmus glabra*) in forest and open land after the

- influence of Dutch elm disease. *Conserv Genet* 11, 257–268. doi: 10.1007/s10592-009-0028-5
- Padmanabhan, C., Zhang, X., and Jin, H. (2009). Host small RNAs are big contributors to plant innate immunity. *Curr Opin Plant Biol* 12, 465–472. doi: 10.1016/j.pbi.2009.06.005
- Pendleton, M., Sebra, R., Pang, A. W. C., Ummat, A., Franzen, O., Rausch, T., et al. (2015). Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat Methods* 12, 780–786. doi: 10.1038/nmeth.3454
- Petit, R. J., and Hampe, A. (2006). Some Evolutionary Consequences of Being a Tree. *Annu. Rev. Ecol. Evol. Syst.* 37, 187–214. doi: 10.1146/annurev.ecolsys.37.091305.110215
- Pike, C. C., Koch, J., and Nelson, C. D. (2021). Breeding for Resistance to Tree Pests: Successes, Challenges, and a Guide to the Future. *Journal of Forestry* 119, 96–105. doi: 10.1093/jofore/fvaa049
- Pliura, A., Lygis, V., Marčiulyniene, D., Suchockas, V., and Bakys, R. (2016). Genetic variation of *Fraxinus excelsior* half-sib families in response to ash dieback disease following simulated spring frost and summer drought treatments. *iForest* 9, 12–22. doi: 10.3832/ifor1514-008
- Pliūra, A., Marčiulynienė, D., Bakys, R., and Suchockas, V. (2014). Dynamics of genetic resistance to *Hymenoscyphus pseudoalbidus* in juvenile *Fraxinus excelsior* clones. *Baltic Forestry* 20, 10–27.
- Pliura A, Lygis V, Suchockas V, and Bartkevicius E. (2011). Performance of Twenty Four European *Fraxinus excelsior* Populations in three Lithuanian progeny trials with a special emphasis on resistance to *Chalara Fraxinea*. *Baltic Forestry* 17, 17–34.
- Plomion, C., Bastien, C., Bogeat-Triboulot, M.-B., Bouffier, L., Déjardin, A., Duplessis, S., et al. (2016). Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science* 73, 77–103. doi: 10.1007/s13595-015-0488-3
- Robert-Seilantantz, A., Grant, M., and Jones, J. D. G. (2011). Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. *Annu Rev Phytopathol* 49, 317–343. doi: 10.1146/annurev-phyto-073009-114447
- Scholthof, K.-B. G. (2007). The disease triangle: pathogens, the environment and society. *Nat Rev Microbiol* 5, 152–156. doi: 10.1038/nrmicro1596
- Schreiber, M., Jayakodi, M., Stein, N., and Mascher, M. (2024). Plant pangenomes for crop improvement, biodiversity and evolution. *Nat Rev Genet.* doi: 10.1038/s41576-024-00691-4
- Siberchicot, A., Bessy, A., Guéguen, L., and Marais, G. A. B. (2017). MareyMap Online: A User-Friendly Web Application and Database Service for Estimating Recombination Rates Using Physical and Genetic Maps.
- Snieszko, R. A., Johnson, J. S., and Savin, D. P. (2020). Assessing the durability, stability, and usability of genetic resistance to a non-native fungal pathogen in two pine species. *Plants, People, Planet* 2, 57–68. doi: 10.1002/ppp3.49
- Snieszko, R. A., and Koch, J. (2017). Breeding trees resistant to insects and diseases: putting theory into application. *Biol Invasions* 19, 3377–3400. doi: 10.1007/s10530-017-1482-5
- Snieszko, R. A., and Liu, J.-J. (2023). Prospects for developing durable resistance in populations of forest trees. *New For (Dordr)* 54, 751–767. doi: 10.1007/s11056-021-09898-3

- Sollars, E. S. A., Harper, A. L., Kelly, L. J., Sambles, C. M., Ramirez-Gonzalez, R. H., Swarbreck, D., et al. (2017). Genome sequence and genetic diversity of European ash trees. *Nature* 541, 212–216. doi: 10.1038/nature20786
- Stocks, J. J., Metheringham, C. L., Plumb, W. J., Lee, S. J., Kelly, L. J., Nichols, R. A., et al. (2019). Genomic basis of European ash tree resistance to ash dieback fungus. *Nat Ecol Evol* 3, 1686–1696. doi: 10.1038/s41559-019-1036-6
- Striganavičiūtė, G., Žiauka, J., Sirgedaitė-Šėžienė, V., and Vaitiekūnaitė, D. (2021). Priming of Resistance-Related Phenolics: A Study of Plant-Associated Bacteria and *Hymenoscyphus fraxineus*. *Microorganisms* 9, 2504. doi: 10.3390/microorganisms9122504
- Thavamanikumar, S., Southerton, S. G., Bossinger, G., and Thumma, B. R. (2013). Dissection of complex traits in forest trees — opportunities for marker-assisted selection. *Tree Genetics & Genomes* 9, 627–639. doi: 10.1007/s11295-013-0594-z
- Torello Marinoni, D., Valentini, N., Portis, E., Acquadro, A., Beltramo, C., Mehlenbacher, S. A., et al. (2018). High density SNP mapping and QTL analysis for time of leaf budburst in *Corylus avellana* L. *PLOS ONE* 13, e0195408. doi: 10.1371/journal.pone.0195408
- Wilson, S. K., Pretorius, T., and Naidoo, S. (2023). Mechanisms of systemic resistance to pathogen infection in plants and their potential application in forestry. *BMC Plant Biol* 23, 404. doi: 10.1186/s12870-023-04391-9
- Wu, W., Zhang, X., Deng, Z., An, Z., Huang, H., Li, W., et al. (2022). Ultrahigh-density genetic map construction and identification of quantitative trait loci for growth in rubber tree (*Hevea brasiliensis*). *Industrial Crops and Products* 178, 114560. doi: 10.1016/j.indcrop.2022.114560
- Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., et al. (2019). Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat Genet* 51, 1052–1059. doi: 10.1038/s41588-019-0427-6

Acknowledgements

Before I started working at the Thünen Institut, I was devastated about Science. I thought there were no places where I could work in a safe environment where I could learn and preserve my passion. That changed when I had my first phone call with Niels. A few months later, I started my PhD journey at Thünen in Großhansdorf with Birgit and Niels as my supervisors. My first office in the library with a wonderful view was just the start of nearly four years that I will have in good memory.

First of all, I want to thank my supervisor, Niels. You were the person who always made me think I was smart and competent, even if I could not remember a single name or number without my notes. You never got tired of explaining or giving me a hint in the right direction. You are among the smartest people I know, and your mind functions differently, in a good way. With all the obstacles, you always supported me and had my back to the last. I'm very grateful for the time I was allowed to follow your journey. Next, I want to thank Birgit. Without you, I would probably not have finished my PhD; your realism and talent for planning were always on point. You always made time for me, even if there were more important tasks. You are an incredible scientist, especially as a woman, you inspire me to follow this path further. I want to thank our project leader, Ben, for giving me a chance to be a part of the FraxforFuture project and for always being helpful and believing in my competence. There would be not FraxGen without Franz. Thank you for sharing your passion for tea with me and all the travels together in this project. We had a lot of fun together, even if there was nothing to laugh about. I promise you, one day we will see the patriarchy fall.

Annika, you were among the first people I met in Thünen on my first day. Which was a very friendly welcome. Since then, I have greatly enjoyed our work on high-molecular-weight DNA. We always had fun, even if nothing worked, flowcells were dying, and DNA turned brown or disappeared. Ultimately, we won the hunt and know we have mastered the extraction of high-molecular-weight DNA in beech and ash. Besides the work, I enjoyed our conversations about everyday life and Thünen shenanigans. The third partner in ruling over the HMW extraction is Katrin (and

Ayla). I'm so glad I can call you my colleague and friend. I will miss our daily dog walks (Captain as well). Talking about friends and family, ups and downs in our lives. Our vibe was always uplifting, even on more difficult days.

Khira, I wish we had more time together in our office. The days we spent together were more fun, and we always had something to laugh about. I wish you and your family all the best. I'm looking forward to meeting your son.

Next, I want to thank Desanka. I think our time overlap was meant to be. I have not met a person in a while with whom I had so much in common. Soon, we will not only dream together of having a PhD, but we will also proudly wear our hats! I'm looking forward to when we have free minds and a proper vacation, where we will only read, hike, look for wonderful trees, and watch some trash. You are my PhD buddy. Remember, it is okay to have a tree as a friend.

I want to thank my two best friends, Laura and Benny. Even if we are far apart, you are always with me. Doing one by one, our contribution to saving nature from humankind and climate crises.

Finally, I want to thank my family. Mama and Papa, you always believed in me, even when I was not. You supported me with everything you had. I always have a place where I can go, filled with love and support, even when my compass needle is spinning around. You are my north star; I can always follow. My sister and her small family always have a sofa to sleep on for me and entrust me with little Liam. Without you three, home is not home. I'm looking forward to family vacations, where we can laugh until we are out of breath and watch Liam grow up.

I think everyone should have a Samewise Gamgee in their life. Someone with courageous, wears his heart on their sleeve, is not afraid to show true feelings, and always sees the light when everything seems dark. I'm very glad I found this person in you, Daniel, who would carry me up a mountain to destroy a ring (or finish my PhD). I would always do the same for you (maybe stealing the Falcon or Imperial I-class Star Destroyer for that). It is dangerous to go alone, so I take you and Captain. You are my person.

Ich versichere, dass dieses gebundene Exemplar der Dissertation und das in elektronischer Form eingereichte Dissertationsexemplar (über den Docata-Upload) und das bei der Fakultät (zuständiges Studienbüro bzw. Promotionsbüro Physik) zur Archivierung eingereichte gedruckte gebundene Exemplar der Dissertationsschrift identisch sind.

Ort, Datum
Unterschrift

Vorname und Nachname,

I, the undersigned, declare that this bound copy of the dissertation and the dissertation submitted in electronic form (via the Docata upload) and the printed bound copy of the dissertation submitted to the faculty (responsible Academic Office or the Doctoral Office Physics) for archiving are identical.

Place, Date

First name and surname, signatur