Genetic diversity in non-native conifer woodlands under Continuous Cover Forestry management in England



Laura Guillardín Calvo

Linacre College & Department of Biology

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2023

In memory of Pablo Goikoetxea, great mentor and beloved friend

Table of Contents

Acknowledgements **6**
Abstract **8**
Chapter 1. General Introduction **11**
1.1 Literature review 12
1.1.1 Forestry in the UK and CCF historical perspective 12
1.1.2 Rapid Environmental Change 14
1.1.3 Importance of studying genetics in the context of CCF 14
1.1.4 Study species and study sites 15
1.1.5 DNA isolation and reducing plastic in the lab 20
1.1.6 Genotyping techniques 21
1.1.7 Detecting changes in genetic diversity 22
1.2 Rationale, objectives and outline of this thesis 24
1.3 References 26
Chapter 2. Comparing DNA isolation methods for forest trees: quality, plastic footprint, and time-efficiency **37**
2.1 Abstract 38
2.2 Background 39
2.3 Methods 42
2.3.1 Plant material 42
2.3.2 DNA isolation procedures and improvements 44
2.3.3 DNA recovery 45
2.3.4 DNA analyses 45
2.3.5 Plastic footprint and time needed for DNA isolations 46
2.4 Results 48
2.4.1 Improvement of DNA isolation methods and comparison of species 48
2.4.2 Analyses using the DNA recovered 51
2.4.3 DNA obtained in large populations 53
2.4.4 Plastic footprint and time needed for DNA isolations 54
2.5 Discussion 56
2.6 Conclusion 59
2.7 Acknowledgements 59
2.8 References 60
Chapter 3. GBS-Based SNP Genotyping and Genetic Diversity Assessment in *Thuja plicata* Adults and Juveniles in UK Continuous Cover Forestry **69**
3.1 Abstract 70
3.2 Background 71
3.3 Material and Methods 74
3.3.1 Study sites characterisation: structure and species diversity 74
3.3.2 Sampling and DNA isolation 77
3.3.3 Overview of genetic analyses 78
3.3.4 GBS library construction, sequencing and quality check 79
3.3.5 Clustering, assembly and variant calling using Ipyrad 80

3.3.6 Data post-filtering and SNP sets design....81
Sequence-based genotyping (SNP set1)....82
PCR-based genotyping (SNP set 2)....82
3.3.7 SNP set validation....83
3.3.8 Population genetics analysis....84
Genetic diversity and genetic variation....84
Kinship analysis within sites....85
3.4 Results....86
3.4.1 Study sites characterisation: structure and species diversity....86
3.4.2 DNA isolation, sequencing, quality check and SNP calling....89
3.4.3 Data post-filtering and SNP set design....92
3.4.4 SNP set validation....94
3.4.5 Population genetic analyses....97
Genetic diversity and genetic variation....97
Kinship analysis....99
3.5 Discussion....101
3.5.1 Sites biodiversity and level of transformation....101
3.5.2 Low number of variants and monomorphic SNPs....102
3.5.3 SNP genotyping assay development....103
3.5.4 Population genetic analysis using SNP set2....104
3.6 Conclusion....106
3.7 Acknowledgements....108
3.8 References....109
Chapter 4. High Heterozygosity but low levels of Effective Population Size in UK Continuous Cover Forestry planted stands of *Pseudotsuga menziesii*....121
4.1 Abstract....122
4.2 Background....123
4.3 Material and Methods....127
4.3.1 Study sites characterisation: structure and species diversity....127
4.3.2 Sampling and DNA isolation....130
4.3.3 Overview of genetic analyses....131
4.3.4 SNP assay design & Genotyping....132
4.3.5 Population genetics analysis....134
Genetic diversity and genetic variation....134
Kinship analysis....135
4.4 Results....136
4.4.1 Study sites characterisation: structure and species diversity....136
4.4.2 Genotyping assay development....139
4.4.3 Genetic diversity and genetic variation....139
4.4.4 Kinship analysis....141
4.5 Discussion....142
4.5.1 Sites biodiversity and levels of transformation....142
4.5.2 Population genetics analysis....143
4.6 Conclusion....146

4.7 Acknowledgements..... 146
4.8 References..... 148
Chapter 5. General Discussion.....159
5.1 An effective DNA isolation system that reduces carbon footprint.....159
5.2 Woodlands potential adaptation to environmental changes under CCF.....160
5.2.1 High levels of heterozygosity and monomorphic SNPs found in *T. plicata* stands.....161
5.2.2 Low levels of Ne in *P. menziesii* stands but high heterozygosity..... 162
5.2.3 Diversity in species, structure and genes..... 162
5.2.4. Assessment of the thesis objectives and hypothesis..... 164
5.3 Future research directions.....165
5.4 Impacts and potential applications of the research.....166
5.5 References..... 169

Acknowledgements

This thesis is for everyone who has supported me both personally and professionally since I decided to become a researcher. Thanks to the John Oldacre Foundation for sponsoring my PhD and to Mr Henry Hoare for contributing financially to the project. Thanks to the Oxford Biology Department and Linacre College for the immense support and help I received from every member of the staff throughout these 4 years.

The Climate Crisis we are witnessing is leading to such challenging times that contributing to expanding knowledge to overcome its impacts is key. This, together with my curiosity and willingness to understand forest species from the genetic point of view, provided the perfect context to begin the challenging and exciting journey of becoming a forest genetics researcher. This journey started with Unai López de Heredia, Alvaro Soto and. Pablo Goikoetxea†, they showed me what great science is and their dedication to their research inspired me and gave me the tools to learn and grow as a researcher. I am very grateful to you and cannot wait to celebrate. Pablo, I will always carry your love for Oaks with me.

After a year of working as a data analyst, John MacKay gave me the opportunity to continue with this adventure, and so I moved to the UK to start this PhD. Thank you, John, for sharing your insight and extensive knowledge in forestry while mentoring and guiding me through this thesis. Your comprehension and support helped me to continue even at the lowest moments. Working out together alternative plans and being able to see how to proceed even if it was not how we first planned, allowed me to move forward and keep progressing, which in the end was the end goal. I would also like to thank my second supervisor, Dr. Gary Kerr for sharing with me his knowledge and expertise on forestry systems. I am also grateful to all the past and current ForesTree Biology Group (former MacKay lab), Barley Rose, Ernest, Gabriele, Gavin, Hayley, Henry, Heather and Liam for their daily support, advice and suggestions, as well as for the fun and social events we had. Special thanks to Henry, for always having an answer to any of my (many) questions, to Barley Rose, for being a great friend, for listening when I needed and sharing very nice moments; and finally, to Hayley, for being a great support, role model and inspiration, in and outside academia, thanks for sharing with me your

love for science, trees and mountains. I would like to extend my sincere thanks to Ella, Emily, George and Will for assisting in some of the labwork and fieldwork, you all made hard work very enjoyable and this thesis could not have been undertaken without your enthusiasm and help. Thanks to Patty, Sarah and Caroline for the lab maintenance and to Tom for all the chats we shared in the department. Finally, I thank Stephen Harris, David Boshier and Dmitry Filatov for their valuable feedback that has improved this thesis.

On a more personal note, I would like to thank my family: Gracias por vuestro apoyo y amor incondicional, y por siempre creer en mí. Crecer rodeada de vuestra pasión por las montañas, árboles y naturaleza es lo que me ha convertido en lo que soy ahora. A Bergel y Kike, por las aventuras vividas y las que nos quedan juntos, gracias por estar siempre ahí. I would also like to thank the Oxford Basketball Team for being the safe place I needed, especially Rui, Rosa, Shivani, Luisa and Maike; playing basketball with you made me so happy. Thanks to Jenny, Phill and Matt, for all those quarantine birthdays and isolated moments we went through together, managing somehow, to make it enjoyable. Anna, thanks for being a great friend, your enthusiasm, tenacity, happiness and way of loving inspire me and I cannot wait to get into new exciting adventures with you. Finally, Alex, you kept me going when I did not want to, you gave me strength and made me always see the bright side of everything. You showed me that real happiness can be achieved and it is all about simply sharing moments together. We have the rest of our lives to live, explore, learn, enjoy and share together.

"*Your trees are here*

You are close but not near

Future forward

The only way is through here

I belong with you

Your mind made up

I belong with you

Your mind made up"

I belong with you - Alex Doherty

Abstract

Climate change is reshaping ecosystems at a pace that results in tree species not predicted to be able to adapt to new conditions fast enough. The variation in alleles is the main driver of adaptation in response to environmental change. Therefore, estimating the potential adaptability of tree species to Climate Change by monitoring their genetic diversity (GD) is crucial. Tree species appear in forests in varied ways, supporting valuable habitats while providing a full range of ecosystem services, including CO_2 capture, maintaining water quality, enhancing biodiversity and landscape, and producing timber. Due to the increasing concern for the future adaptability of forests, the Continuous Cover Forestry (CCF) system presents an opportunity to maintain and enhance forest services by prioritising the management of the ecosystem and encouraging natural regeneration rather than focusing on timber production; for that reason, this approach is being promoted by the UK Forestry Standards to replace other systems such as clear-felling. Therefore, even-aged plantations, usually composed of non-native conifer species are being transformed into CCF; however, there is a lack of studies on the transmission of GD to the naturally regenerated offspring.

This thesis aims to understand the GD within and between adults and juveniles under the CCF system. Our hypothesis suggests that in the long-term, the natural regeneration resulting from these plantations may show a low GD reducing the woodland's potential adaptability to rapid environmental changes. We analysed selected woodlands, focusing on two widely planted species, *Pseudotsuga menziesii* (Mirb.) Franco and *Thuja plicata* Donn ex D. Don. The project analysed over 1100 trees from seven woodlands, which required optimised and efficient techniques to process the samples and determine genotypes. We compared the DNA quality, time and plastic-efficiency of two DNA isolation methods and selected the optimal protocol for each species. We designed the genotyping assays using available SNP datasets for *P. menziesii* and performed a GBS to discover SNPs on *T. plicata*. We used a high-throughput PCR-based SNP scoring assay to genotype the sample sets. A detailed forestry survey across all sites allowed us to characterise and assess the level of transformation to CCF.

We found high levels of heterozygosity overall across species and stands, in both adults and juveniles; however, we found a low effective population size (N_e<500) in every stand for both species. In general, *T. plicata* stands had lower N_e compared to *P. menziesii*, in both adults and juveniles, suggesting that this species may have a lower adaptability to rapid environmental changes. In both species, some stands showed lower N_e in juveniles compared to the adults, which may indicate ineffective transmission of GD to the offspring; however, other stands showed the opposite trend, which may be due to pollen from neighbouring stands contributing to the reproductive process. Finally, in some cases, the stands at a late stage of transformation to CCF showed low levels of GD. This may indicate that logging and reducing the number of individuals in favour of species diversity may reduce the adaptive potential in the long-term. The variety of results across stands and species confirms the complexity of this subject and indicates that even though a woodland may appear close-to-nature visually, its genetic makeup may not be sufficiently diverse to adapt to changing climate conditions.

Chapter 1. General Introduction

Globally, conifers have largely been planted for timber production with limited age and species diversity managed in a clear-fell system; however, since the 1960s, a transition to a more sustainable and biodiverse approach has started in some areas including the UK (Helliwell & Wilson, 2012). Low-impact silvicultural systems such as the Continuous Cover Forestry (CCF) management are advised in the UK Forestry Standard (Forestry Commission, 2023), which encourages structural and species diversity while promoting natural regeneration to favour evolutionary adaptation. The CCF system can be developed from scratch in non-forested areas such as abandoned agricultural lands, after a clear-felling rotation or after disturbances, but in England, CCF is usually developed by transforming even-aged plantations of non-native conifers. Understanding how genetic diversity is transferred from the planted adults to their offspring in these woodlands is of interest in the context of CCF management. Specifically, a potential lack of genetic diversity in the planted adults may lead to an even narrower genetic diversity in the naturally regenerated trees (Finkeldey, 2002).

In order to support adaptation, it is crucial to maintain a high level of standing genetic diversity and to facilitate its maintenance from generation to generation. In most cases the level of genetic diversity in planted trees is unknown and the adaptation to existing conditions is untested (Konnert et al., 2011). In this thesis, we studied the genetic diversity levels in *Pseudotsuga menziesii* (Mirb.) Franco and *Thuja plicata* Donn ex D. Don. Studies in Europe have reported high levels of genetic diversity in planted *P. menziesii* stands (Eckhart et al., 2017; van Loo et al., 2019) but also found a reduction of diversity in natural regeneration (Neophytou et al., 2019). *T. plicata* in contrast, has not been studied in Europe or the UK, and its natural range is known to have low levels of genetic diversity and small effective population sizes (Shalev et al., 2022). This background information suggests very different scenarios between these two species thus we may expect very different characteristics between them.

This thesis delves into the genetic diversity in non-native coniferous woodlands, particularly focusing on differences between planted trees and their offspring under CCF management to inform sustainable management practices. To estimate the genetic diversity indicators in both adult and

juvenile trees from several woodlands, we optimised DNA isolation protocols and developed novel genotyping assays.

In this chapter, we describe the major effects of climate change on woodlands and review the potential for CCF to overcome some of these environmental impacts. Then we discuss the importance of understanding the genetic background of the planted trees and how logging and selective harvesting may affect genetic diversity in the offspring. Along the way, we will provide an introduction to the study sites and the species used in this work. Finally, we will discuss the DNA isolation methods and will review common genotyping techniques and methods to detect changes in genetic diversity.

1.1 Literature review

1.1.1 Forestry in the UK and CCF historical perspective

UK woodland area is estimated to be 3.25 million hectares, which represents 13.4% of the total land area. Over half (53%) of it is composed of broadleaves species (Forest research, 2023). Native tree species make up around 50% of the total forest cover. Some of the most common and significant ones include *Quercus robur* and *Quercus petraea*, *Fagus sylvatica*, *Pinus sylvestris*, *Betula pendula*, *Betula pubescens*, *Fraxinus excelsior*, *Sorbus aucuparia*, *Crataegus monogyna*, *Ilex aquifolium*, *Taxus baccata*. Some of the common and significant non-native species are *Picea sitchensi*s, *Pseudotsuga menziesi*i, *Thuja plicata*, *Picea abies*, *Larix decidua*, *Larix kaempferi*, *Pinus radiata*, *Chamaecyparis lawsoniana*, *Acer pseudoplatanus* and *Aesculus hippocastanum*. Across the UK, P. menziesii accounts for 3.7% and T. plicata < 1% of the land covered with conifer species (Forest research, 2023).

The history of UK forests is marked by changing needs and aims, which have shaped woodlands through the evolution of management strategies and forest policies. Due to Britain's dependence on timber for naval shipbuilding and other industries, periods of heavy felling and planting occurred from the 17th century but without specific management plans until the mid 19th century (Holmes, 1975). Around the middle 19th century, a debate started in Europe about the benefits of mixed forests with the increased interest in science along with a desire to both understand and

protect the environment (Helliwell & Wilson, 2012). It is around this period that arboretums were set up with both imported exotic and native tree species, e.g. Westonbirt National Arboretum, and when irregular management systems appeared (Holmes, 1975). Despite these developments, productive plantations gained in importance. The Forestry Commission was established in 1919 for the strategic industrialization of forests, which resulted in a rapid expansion of intensive conifer forestry based on clear felling systems mostly in the north and west of Britain (Raum & Potter, 2015). From 1968, a transition to a more sustainable approach began. In the second half of the 20th century, new forestry policies were developed in the UK to explicitly include other aims than timber harvest (Holmes, 1975). The 'Ancient Woodland Inventory' and 'Policy for Broadleaved Woodlands' (Forestry Commission, 1985) were the first laws with the specific objective of safeguarding the biodiversity of UK forests. In 1998, the Forestry Commission published the United Kingdom Forestry Standards (UKFS), the government's approach to sustainable forestry, now in the fourth revision (Forestry Commission, 2017). The UKFS is the present reference policy: it defines requirements and standards of sustainability and biodiversity and is mandatory for every woodland in the UK. It includes resilience objectives and multipurpose forests uses, encouraging the transformation of even-aged pure populations towards mixed irregular ones where a continuous cover is guaranteed, as proposed by the CCF approach. Already by the early 1990s, these objectives and the adoption of the CCF silvicultural system had gained in popularity as indicated by the foundation of the CCF Group (Helliwell & Wilson, 2012).

The CCF is an approach to forest management which uses silviculture based on ecological and environmental principles, maximising heterogeneity of species and sizes in the woodlands (Mason et al., 1999). "CCF aims to develop structurally, visually and biologically diverse forests, in which sustainable production of quality timber is achieved along with the provision of a wide range of ecosystem services" (Continuous Cover Foresty Group, 2024). Forests under Continuous Cover Management are described as woodlands in which canopy tree cover is sustained over time as opposed to forests managed based on systems that use clear felling. Maintaining a forest cover is expected to be more effective in curbing the impact of climate change than other traditional forestry practices (Helliwell & Wilson, 2012; Mason et al., 1999). The CCF approach is based upon the

following four key principles: i) ecosystem management rather than focus on trees management with an exclusive timber production aim; ii) using succession and disturbance as the basis of silvicultural prescriptions for stand management and renewal; iii) working within site limitations, and iv) development of a diverse (irregular) stand structure with a mixture of ages and species (Mason et al., 1999). Essentially, this approach provides the tools to manage an irregular forest of mixed species considering both the site potential and the multi-use purpose of woodlands (Kirby & Watkins, 2015).

1.1.2 Rapid Environmental Change

Nowdays, forests are being affected by a number of conditions, including the change of the land use and Climate Change, which is the main vector of rapid environmental changes (Rosenzweig et al., 2008). Forest tree species have long lifespans, which increases the risk of being outpaced by rapid changes and therefore, reduces their ability to adapt to new conditions (Taylor et al., 2015). To estimate the potential effects on forests, it is vital to understand both the phenotypic plasticity and adaptation ability of forest species to cope with observed and predicted climate conditions (Kremer et al., 2012, 2014; Peters, 1990). A change of temperatures are a component of climate change that is observed globally (Rosenzweig et al., 2008). Higher temperatures are expected to push species to higher altitudes and latitudes, causing a shift in their natural range, while lower temperatures may cause frosting damage, especially at seed early stages. Under the future expected conditions, pests and diseases can impact woodlands more seriously as they may affect non-adapted hosts due to the expansion of their distribution ranges and their life cycles changes due to variations in temperatures (Ramsfield et al., 2016; Volney & Fleming, 2000). Changes in precipitation regimes together with higher temperatures may lead to more severe drought events followed by an increased frequency and intensity of forest fires (Millar & Stephenson, 2015).

1.1.3 Importance of studying genetics in the context of CCF

These threats caused by a rapidly changing climate are expected in both natural and managed forests (Tew et al., 2023), thus implementing conservation genetics strategies (Aitken & Bemmels, 2016; Rellstab et al., 2015) while monitoring levels of genetic diversity (Andersson et al., 2022;

Heuertz et al., 2023; Hoban et al., 2021) is crucial to increase forests' resilience (Archambeau et al., 2023). Genetic variation within populations allows species to increase their adaptation capacity as their environment changes (Bitter et al., 2019), thus monitoring the levels of genetic diversity in forests using DNA-based techniques is key to ensuring the woodland's potential adaptability.

Even-aged plantations of single species, widely used for timber production, might hold a reduced genetic variation in comparison to mixed species forests, thus their potential to resist or adapt to several stresses is more limited (Straw et al., 2017). Plantations are also generally expected to be more severely affected by climate disruptions as they have a regular structure and trees of similar height, where the spread of pests, diseases and fires is easier (Klapwijk & Björkman, 2018; Straw et al., 2017). According to the UK forestry standards (Forestry Commission, 2023), even-aged plantations are advised to be transformed into more sustainable woodlands such as uneven-aged and mixed-species forests by using many different approaches; however, a popular strategy is to adopt a Continuous Cover Forestry (CCF) approach (Mason et al., 2003).

While the transformation to an irregular woodland structure is a well-defined silvicultural approach (Kerr et al., 2003; Mason & Kerr, 2000), there is a paucity of studies looking at the genetic context of its effects. Finkeldey (2002) discussed the genetic impact of forests being transformed into CCF due to a change in the population density as a result of selective harvesting. A decrease in the number of reproductive individuals may alter the breeding system which may lead to an increase in inbreeding and selfing, limiting the gene pool of future natural generations (Schaberg et al., 2018. The resulting stands may be unable to adapt to emerging climate impacts (Finkeldey, 2002), thus, if primary trees hold a reduced gene pool, the vigour of the future forest may be further compromised. Therefore, the assessment of genetic and evolutionary impacts of selective harvesting in CCF woodlands is of great interest (Frankham et al., 2002).

1.1.4 Study species and study sites

Currently, there are several sites in the UK in which CCF is well-developed; here, we selected five sites and a total of seven stands that may be at different stages of transformation (Table 1, Figure 1 and Figure 2). The Stourhead (Western) estate switched to CCF as the preferred management system

in the 1950s when the owners decided to improve the biodiversity by increasing the number of species while harvesting individual trees instead of using clear-felling practices. It is made up of different stands with a mixture of species and a combination of age classes throughout the site. In our project, we defined two stands within the Stourhead estate, Stourhead1 and Stourhead2, which had different species compositions and planted years (Table 1, Figure 2). The Longleat estate has been managed under CCF for the last 40 years on much of its woodland and the long-term aim is to maintain and expand CCF when suitable. Similarly, we defined two stands within the Longleat estate, Longleat1 and Longleat2 (Table 1, Figure 2). Forestry Research scientists are testing different methods in Mortimer (Figure 2) to transform pure even-aged stands towards irregular and mixed forests managed under the CCF approach. Ringwood Forest (Figure 2) is included in the East Dorset Blocks woodland and managed by Forestry England, the stand we selected is defined as mixed conifers and is managed under the Low Impact Silvicultural Systems (LISS), which includes the CCF management system (Mason et al., 1999). Finally, Bagley Wood (Figure 2) is a large ancient woodland owned by St John's College (Oxford) composed of native broadleaves and non-native conifers compartments. Differences in structure among the seven studied sites can also be observed in Figure 2.

Table 1. Stand names, year planted and species studied.

Stand	Year planted	Species studied	Location	Altitude (m)	Size (ha)	Ownership
Mortimer	1980	*P. menziesii*	Long: -2.74263 Lat: 52.362642	220	5.33	Forestry Research
Ringwood	1961	*P. menziesii*	Long: -1.62939 Lat: 50.874727	44	5.77	Forestry England
Stourhead2	1939, 1951	*P. menziesii*, *T. plicata*	Long: -2.35440 Lat: 51.10101	244	18.57	Hoare’s family
Longleat2	1925	*P. menziesii*	Long: -2.21294 Lat: 51.1765	151	6.87	Thynn family (Marquess of Bath)
Bagley Wood	1950	*T. plicata*	Long: -1.260311 Lat: 51.715921	93.1	3.73	St John's College (Oxford)
Longleat1	1958	*T. plicata*	Long: -2.179381 Lat: 51.78560	151	5.07	Thynn family (Marquess of Bath)
Stourhead1	1961	*T. plicata*	Long: -2.35114 Lat: 51.10856	244	5.00	Hoare’s family

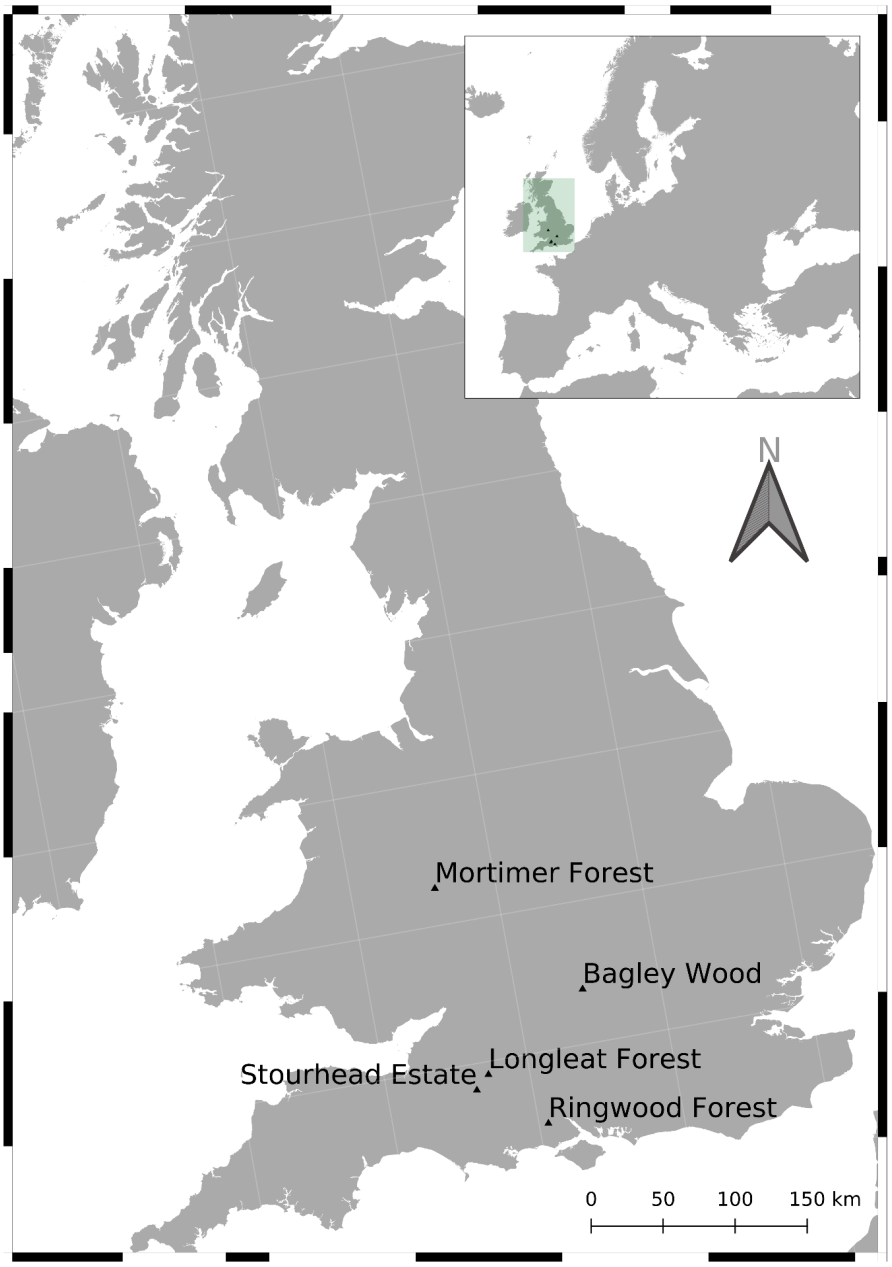


Figure 1. Location map of the five selected study sites.



Figure 2. Images of the seven study stands located on five different sites. Images taken by Laura Guillardin during the field work (Jun 2021 - October 2021).
A) Ringwood,
B) Stourhead1,
C) Stourhead2,
D) Bagley Wood,
E) Longleat1,
F) Mortimer,
G) Longleat2.

To investigate the genetic status of CCF woodlands in the UK, we selected *P. menziesii* and *T. plicata* as they both appear as the main species in several UK plantations that may be at different stages of transformation to CCF. *P. menziesii* and *T. plicata* are large evergreen coniferous trees, and both are native to western North America where they appear together in mixed natural forests and plantations (Figure 3). These species are monoecious, i.e. individuals have both the male and female reproductive separated organs, and anemochorous, i.e. their pollen and seed are dispersed by the wind. *P. menziesii* occurs in two different ecological varieties, *P. menziesii* var. *menziesii* or *viridis* (Schwer.) Franco (coastal variety), and *P. menziesii* var. *glauca* (Beissn.) Franco (interior variety, or Rocky-Mountain variety), which hybridise naturally (van Loo et al., 2015). While both species have a principally outcrossing mating system, inbreeding by mating genetically related individuals or self-reproduction may occur under different situations (Huang et al., 2022). *T. plicata* is known to have high selfing (Wang & Russel, 2006) which may imply an increase of inbreeding depression and lower heterozygosity.

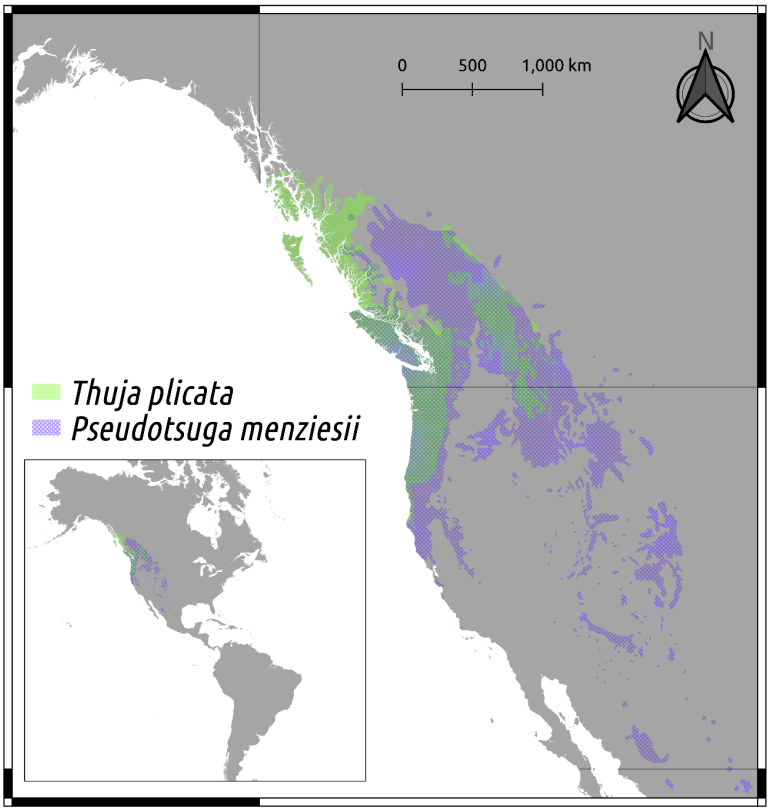


Figure 3. Natural ranges of *Thuja plicata* (in green) and *Pseudotsuga menziesii* (in purple).

1.1.5 DNA isolation and reducing plastic in the lab

Molecular experiments investigating populations across multiple sites may require the use of a large amount of single-use plastic consumables (Howes, 2019; Sawyer, 2019). An estimated 5.5 million tonnes of plastic is used in laboratories per year worldwide (Urbina et al., 2015). Therefore molecular studies, although essential to understanding crucial biological processes, have a large environmental impact (Di Paolo et al., 2022). One of the main steps and challenges when designing genomic experiments is the isolation of nucleic acids (DNA or RNA). Although several methods have been developed for recalcitrant tissue samples for isolating DNA from tree species (e. g. Emaus et al., 2022; Green & Sambrook, 2016; Vega-Vela & Chacón Sánchez, 2011), few studies have compared their plastic use.

Organisations and institutions are applying laboratory sustainability practises to reduce plastic use such as My Green Lab (https://www.mygreenlab.org/) and The Sustainable Laboratory Practices Working Group (SLPWG). The reuse of specific plastic items is increasing with the development of new cleaning technologies such as the on-bench Pipette Tip Cleaning Machine (TipNovus) (Grenova Solutions, Richmond, VA, USA). Other studies are probing the potential implementation of protocols that use less plastic without compromising experimental results (Alves et al., 2021; Torkamaneh et al., 2020).

An efficient experimental design by estimating the essential number of samples is required to reduce the potential waist of resources. However, larger sample sizes have recently become more common in genetics and genomics projects (Fumagalli, 2013), due to lower sequencing costs (Levy & Boone, 2019) and increased data analysis power (Hong & Park, 2012). This increases the time and plastic needed to complete a project, therefore, the optimisation and testing of available DNA isolation protocols efficiency is crucial to reduce the project footprint and promote sustainable practices.

1.1.6 Genotyping techniques

Genotyping techniques in forest species have experienced significant developments and have increased our ability to test the genetic variation of these habitats (Holliday et al., 2018). Specifically, coniferous species have very large genome sizes reaching up to 35 Gb in some *Pinus* species but whole genome sequencing is now feasible due to recent technological advances (Mackay et al., 2012; Neale et al., 2013, Woudstra et al., 2023). However, some of the most widely used techniques to study non-model tree species are restriction enzyme-based sequencing methods, which allow the collection of high amounts of data at a practical cost (Holliday et al., 2018; M. Müller et al., 2023). These high-throughput sequencing techniques include Genotyping by Sequencing (GBS) and Restriction-site associated DNA sequencing (RADseq), and require more complex bioinformatic analysis to deal with potentially high amounts of missing data, genotyping errors and to process the raw data to obtain the genotypes (Guillardín-Calvo et al., 2019; Mastretta-Yanes et al., 2015; Mora-Márquez et al., 2023). These methods sequence the fragments of the DNA digested by one or more enzymes and allow the discovery of thousands of single nucleotide polymorphisms (SNPs) (Elshire et al., 2011). SNPs are variations at single base pair locations, bi-allelic and codominant, they can appear within genes and in intergenic regions across the genome, therefore, they are widely used as genetic markers to estimate genetic variation (Zimmerman et al., 2020).

Alternatively to sequence-based genotyping, various high-throughput platforms allow genotyping using selected SNPs on a range of scales from 10s to 10,000s of SNPs. Illumina Infinium iSelect HD, (Illumina, San Diego, CA), Sequenom MassARRAY platform (Agena Bioscience, San Diego, CA), and Standard Biotools Dynamic Arrays (San Francisco, CA, USA), are PCR-based genotyping methods which allow the genotyping of large amounts of samples in a cost and time-efficient way. These techniques are simpler to analyse, have lower levels of missing data increasing reliability, and reduced costs compared to sequencing when analysing large sample sizes (Choi et al., 2020; Davey et al., 2011). Sequenom MassARRAY assays (Godbout et al., 2017; Telfer et al., 2013) and Standard Biotools assays (Karam et al., 2015; Pfenninger et al., 2021) have been

developed in several forest trees; however, assays for the project species have not been developed to date.

A set of highly variable microsatellites were developed by Slavov et al., (2004) to genotype *P. menziesii* individuals which are widely used in population genetic studies (e.g. Eckhart et al., 2017; Pakull et al., 2021; van Loo et al., 2015); additionaly, genotyping systems based on sequencing have been developed (Howe et al., 2020; Müller et al., 2012; Neale et al., 2017). These SNP resources have been applied to genomic studies to study the molecular basis of adaptation (George et al., 2021; Müller et al., 2015) and to detect candidate genes linked to preferred breeding traits (Eckert et al., 2009; Krutovsky & Neale, 2005). *T. plicata* on the other hand, has been deeply investigated to understand its reproductive process (O'Connell et al., 2001) and its high levels of inbreeding (O'Connell et al., 2004), using highly polymorphic microsatellites (O'Connell & Ritland, 2000). Recent studies have developed genomic resources and techniques in *T. plicata* (Shalev et al., 2022) to understand the levels of genetic diversity and the potential inbreeding depression across generations (Shalev et al., 2023).

1.1.7 Detecting changes in genetic diversity

Population genetics concerns the study of the genetic variability of populations and how it is affected by evolutionary forces (Brookfield, 1996; Meilan & Kirst, 2020). Genetic diversity and the changes in the genetic composition of populations are the basis of evolutionary adaptation (Allendorf et al., 2022). Genetic diversity is the study of genetic variants and can be measured at the molecular level by analysing allele frequencies in a population. Genetic variations in alleles and genotypes are the raw material for adaptive evolution and allow species to evolve in response to climatic changes (Neale & Kremer, 2011). Furthermore, a decrease in genetic diversity reduces the evolutionary potential due to a loss of reproductive fitness and an increase in inbreeding (Mills, 2007). The evolutionary forces that act on the allele frequencies and therefore modify the genetic diversity levels of a population are mutation, genetic drift, selection and migration or gene flow (Gillespie, 1998). Mutations are considered the main source of novel genetic variation while genetic drift contributes to their removal as the consequence of a finite number of individuals within a population. Genetic drift

leads to allele fixation and random loss of rare alleles, affecting smaller populations more (Meilan & Kirst, 2020). As mutation and genetic drift affect the entire genome, selection can affect multiple individual loci and occurs at any life stage. Selection can be positive if the individual that carries the allele has a higher reproductive success producing more offspring, or negative selection, which leads to fewer offspring from the individual that carries the selected allele (Meilan & Kirst, 2020). Furthermore, selection is natural when it is driven by environmental processes but can also be driven by human selective harvesting (Frankham et al., 2002) or through genetic resource management. Finally, the migration or exchange of alleles among populations also affects loci individually and changes the allele frequencies of the populations (Meilan & Kirst, 2020). In the absence of evolutionary forces acting on a population, the allele frequencies remain unchanged from generation to generation (Hardy, 1908) and the population is in agreement with the Hardy-Weinberg (HW) principle. Natural populations usually deviate from this perfect equilibrium and to study how the allele frequencies change due to evolutionary forces, Fisher (1956) and Wright (1950) developed statistical methods still in use to analyse population genetics results (Gillespie, 1998; Meilan & Kirst, 2020).

Parameters widely used to describe within-population genetic diversity are Heterozygosity (H), and effective population size (N_e), both useful for tracking population genetic variation and potential evolvability. Heterozygosity is calculated as the proportion of individuals carrying different allele variants at a given gene locus in a population (Hartl, 2020; Hartl & Clark, 1997). The observed heterozygosity (H_o) measured from a sample is the actual proportion of heterozygous genotypes often compared against the expected heterozygosity (H_e), which is calculated from the frequencies of different alleles assuming HW equilibrium. For example, significant discrepancies where H_o is larger than H_e may indicate a lack of heterozygotes potentially due to inbreeding in the population (Meilan & Kirst, 2020).

N_e allows us to predict the evolutionary change in a population that may result from genetic drift (Waples & Do, 2010). The actual number of individuals that form a population (N_c) may be different to the N_e, which is the number of ideal individuals calculated as contributing genetically to the offspring (Waples, 2022). Hence, N_e represents the required number of individuals in a population

that would be affected by genetic drift in the same way as the original population (Waples, 2022). Recently, the Convention on Biological Diversity (CBD, 2022; O'Brien et al., 2022) has recommended the threshold of $N_e > 500$ as a genetic diversity indicator to monitor genetic conservation.

Overall, both indicators (heterozygosity and N_e) are essential for understanding different aspects of genetic diversity and are often used together to gain a more complete view of a population's genetic makeup. The N_e may decrease more rapidly between generations than heterozygosity (H_e), therefore it may be a more sensitive indicator to inform on the potential genetic adaptability of populations faced with rapid climate change (Hoban et al., 2023). This approach is used in conservation genetics and forestry management to ensure the resilience and long-term adaptability of populations in the face of environmental changes and human impacts.

1.2 Rationale, objectives and outline of this thesis

Continuous Cover Forestry (CCF) woodlands in the UK have not been genetically studied. Additionally, one of the main CCF principles is to encourage natural regeneration. Therefore, if either the initial planted stand is not genetically diverse, or if its diversity is reduced by selective removal of potential reproductive individuals, the subsequent offspring may be genetically depauperate, resulting in forests that lack the ability to adapt to environmental changes. A population genetics evaluation comparing genetic diversity indicators across generations will inform on potential reductions in the gene pool in populations transforming to CCF.

Therefore this project hypothesises that in the long-term, the natural regeneration resulting from these plantations may show a low genetic diversity. This can result when: (i) the planted canopy trees hold a reduced gene pool, (ii) the gene pool is transferred poorly to the offspring, and (iii) the woodlands are logged and trees are selected to be removed decreasing the effective population size, which may result in a decrease of diversity over time. We will test this hypothesis by analysing selected woodlands, focusing on two widely planted species, *T. plicata* and *P. menziesii*. The main objectives are to:

Objective 1) To optimise DNA isolation methods for each species while reducing the time and plastic used in the project.

Chapter 2 compares the isolation of DNA using a novel automatic DNA isolation equipment (MicroGEM) and a commonly used DNA isolation kit (Plants mini kit, QIAGEN). It compares the two methods in terms of DNA quality, yield, time-efficiency and plastic footprint on both study species and an additional conifer tree.

Objective 2) To develop genotyping assays for the study species

Chapter 3 describes the SNP discovery step for *T. plicata* using the Genotyping by Sequencing (GBS) technique. Chapters 3 and 4 explain the filtering process and SNP assay design for each species and the SNP genotyping using a high-throughput microfluidic PCR amplification system.

Objective 3) To assess the level of CCF transformation of the study stands

Chapters 3 and 4 evaluate the species and structural diversity of each stand to assess the level of transformation into CCF by doing a forest inventory on the studied stands.

Objective 4) To estimate the genetic diversity within and between adult and juvenile generations

Chapters 3 and 4 assess the genetic diversity in canopy trees and compare it with the genetic diversity that appears in natural regeneration of both species.

1.3 References

Aitken, S. N., & Bemmels, J. B. (2016). Time to get moving: Assisted gene flow of forest trees. Evolutionary Applications, 9(1), 271–290. https://doi.org/10.1111/eva.12293

Allendorf, F. W., Funk, W. C., Aitken, S. N., Byrne, M., Luikart, G., Antunes, A., Allendorf, F. W., Funk, W. C., Aitken, S. N., Byrne, M., Luikart, G., & Antunes, A. (2022). Conservation and the Genomics of Populations (Third Edition, Third Edition). Oxford University Press.

Alves, J., Sargison, F. A., Stawarz, H., Fox, W. B., Huete, S. G., Hassan, A., McTeir, B., & Pickering, A. C. (2021). A case report: Insights into reducing plastic waste in a microbiology laboratory. Access Microbiology, 3(3). https://doi.org/10.1099/acmi.0.000173

Andersson, A., Karlsson, S., Ryman, N., & Laikre, L. (2022). Monitoring genetic diversity with new indicators applied to an alpine freshwater top predator. Molecular Ecology, 31(24), 6422–6439. https://doi.org/10.1111/mec.16710

Archambeau, J., Bianchi, S., Buiteveld, J., Callejas-Díaz, M., Cavers, S., Hallingbäck, H., Kastally, C., de Miguel, M., Mutke, S., Sánchez, L., Whittet, R., González-Martínez, S. C., & Bastien, C. (2023). Managing forest genetic resources for an uncertain future: Findings and perspectives from an international conference. Tree Genetics & Genomes, 19(3), 26. https://doi.org/10.1007/S11295-023-01603-Z

Bitter, M. C., Kapsenberg, L., Gattuso, J.-P., & Pfister, C. A. (2019). Standing genetic variation fuels rapid adaptation to ocean acidification. Nature Communications, 10(1), Article 1. https://doi.org/10.1038/s41467-019-13767-1

CBD. (2022). Decision adopted by the conference of the parties to the Convention on Biological Diversity. 15/15 Monitoring framework for the Kunming-Montreal Global Biodiversity Framework. In Convention of Biological Diversity. https://www.cbd.int/doc/decisions/cop-15/cop-15-dec-05-en.pdf

Choi, S. R., Oh, S. H., Dhandapani, V., Jang, C. S., Ahn, C. H., Rameneni, J. J., Kim, H., Jeon, I., & Lim, Y. P. (2020). Development of SNP markers for marker-assisted breeding in Chinese cabbage using Fluidigm genotyping assays. Horticulture Environment and Biotechnology,

61(2), 327–338. https://doi.org/10.1007/S13580-019-00211-Y

Continuous Cover Foresty Group. (2024, May). *Home.* Continuous Cover Foresty Group. https://www.ccfg.org.uk/

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics, 12(7), Article 7. https://doi.org/10.1038/nrg3012

Di Paolo, L., Abbate, S., Celani, E., Di Battista, D., & Candeloro, G. (2022). Carbon Footprint of Single-Use Plastic Items and Their Substitution. Sustainability 2022, Vol. 14, Page 16563, 14(24), 16563. https://doi.org/10.3390/SU142416563

Eckert, A. J., Bower, A. D., Wegrzyn, J. L., Pande, B., Jermstad, K. D., Krutovsky, K. V., St. Clair, J. B., & Neale, D. B. (2009). Association genetics of coastal Douglas fir (Pseudotsuga menziesii var. Menziesii, Pinaceae). I. Cold-hardiness related traits. Genetics, 182(4), 1289–1302. https://doi.org/10.1534/genetics.109.102350

Eckhart, T., Walcher, S., Hasenauer, H., & van Loo, M. (2017). Genetic diversity and adaptive traits of European versus American Douglas-fir seedlings. European Journal of Forest Research, 136(5–6), 811–825. https://doi.org/10.1007/s10342-017-1072-1

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE, 6(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Emaus, M. N., Cagliero, C., Gostel, M. R., Johnson, G., & Anderson, J. L. (2022). Simple and efficient isolation of plant genomic DNA using magnetic ionic liquids. Plant Methods, 18(1), 1–14. https://doi.org/10.1186/S13007-022-00860-8/FIGURES/10

Finkeldey, R. (2002). Reproduction in Continuous Cover Forests—The Geneticist's Perspective. In Continuous Cover Forestry (pp. 67–79). https://doi.org/10.1007/978-94-015-9886-6_6

Fisher, R. A. (1956). Statistical methods and scientific inference (pp. viii, 175). Hafner Publishing Co.

Forest Research. Forestry Statistics 2023. A Compendium of Statistics about Woodland, Forestry and Primary Wood Processing in the United Kingdom. Edinburgh, 2023.

Forestry Commission. (2023). The UK Forestry Standard.

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2002). Introduction to Conservation Genetics. Cambridge University Press. https://doi.org/10.1017/CBO9780511808999

Fumagalli, M. (2013). Assessing the Effect of Sequencing Depth and Sample Size in Population Genetics Inferences. PLOS ONE, 8(11), e79667. https://doi.org/10.1371/JOURNAL.PONE.0079667

George, J.-P., Schueler, S., Grabner, M., Karanitsch-Ackerl, S., Mayer, K., Stierschneider, M., Weissenbacher, L., & van Loo, M. (2021). Looking for the needle in a downsized haystack: Whole-exome sequencing unravels genomic signals of climatic adaptation in Douglas-fir (*Pseudotsuga menziesii*). Ecology and Evolution, 11(12), 8238–8253. https://doi.org/10.1002/ece3.7654

Gillespie, J. H. (1998). Population genetics: A concise guide. The Johns Hopkins University Press.

Godbout, J., Tremblay, L., Levasseur, C., Lavigne, P., Rainville, A., Mackay, J., Bousquet, J., & Isabel, N. (2017). Development of a traceability system based on a SNP array for large-scale production of high-value white spruce (*Picea glauca*). Frontiers in Plant Science, 8(July), 1–13. https://doi.org/10.3389/fpls.2017.01264

Green, M. R., & Sambrook, J. (2016). Precipitation of DNA with ethanol. Cold Spring Harbor Protocols, 2016(12), 1116–1120. https://doi.org/10.1101/pdb.prot093377

Guillardín-Calvo, L., Mora-Márquez, F., Soto, Á., & López de Heredia, U. (2019). RADdesigner: A workflow to select the optimal sequencing methodology in genotyping experiments on woody plant species. Tree Genetics and Genomes, 15(4). https://doi.org/10.1007/s11295-019-1372-3

Hardy, G. H. (1908). MENDELIAN PROPORTIONS IN A MIXED POPULATION. Science (New York, N.Y.), 28(706), 49–50. https://doi.org/10.1126/science.28.706.49

Hartl, D. L. (2020). A Primer of Population Genetics and Genomics. Oxford University PressOxford. https://doi.org/10.1093/oso/9780198862291.001.0001

Hartl, D. L., & Clark, A. G. (1997). Principles of Population Genetics. In Sinauer Associates. (Vol. 46, Issue 2). https://doi.org/10.2307/2531471

Helliwell, R., & Wilson, E. R. (2012). Continuous Cover Forestry in Britain. Quarterly Journal of Forestry, 106(3):214-224.

Heuertz, M., Carvalho, S. B., Galindo, J., Rinkevich, B., Robakowski, P., Aavik, T., Altinok, I., Barth, J. M. I., Cotrim, H., Goessen, R., González-Martínez, S. C., Grebenc, T., Hoban, S., Kopatz, A., McMahon, B. J., Porth, I., Raeymaekers, J. A. M., Träger, S., Valdecantos, A., Vella, A., Vernsesi, C., Garnier-Géré, P. (2023). The application gap: Genomics for biodiversity and ecosystem service management. Biological Conservation, 278, 109883. https://doi.org/10.1016/J.BIOCON.2022.109883

Hoban, S., Bruford, M. W., Chris Funk, W., Galbusera, P., Patrick Griffith, M., Grueber, C. E., Heuertz, M., Hunter, M. E., Hvilsom, C., Kalamujic Stroil, B., Kershaw, F., Khoury, C. K., Laikre, L., Lopes-fernandes, M., Macdonald, A. J., Mergeay, J., Meek, M., Mittan, C., Mukassabi, T. A., O'Brien, D., Ogden, R., Palma-Silva, C., Ramakrishnan, U., Segelbacher, G., Shaw, R. E., Sjogren-Gulve, P., Veličković, N., Vernesi, C. (2021). Global Commitments to Conserving and Monitoring Genetic Diversity Are Now Necessary and Feasible. Forum 964 BioScience, 71(9). https://doi.org/10.1093/biosci/biab054

Holliday, J. A., Hallerman, E. M., & Haak, D. C. (2018). Genotyping and Sequencing Technologies in Population Genetics and Genomics. 83–125. https://doi.org/10.1007/13836_2017_5

Holmes, G. D. (1975). History of Forestry and Forest Management. Philosophical Transactions of the Royal Society of London, 271(911), 69–80.

Hong, E. P., & Park, J. W. (2012). Sample Size and Statistical Power Calculation in Genetic Association Studies. Genomics & Informatics, 10(2), 117. https://doi.org/10.5808/GI.2012.10.2.117

Howe, G. T., Jayawickrama, K., Kolpak, S. E., Kling, J., Trappe, M., Hipkins, V., Ye, T., Guida, S., Cronn, R., Cushman, S. A., & McEvoy, S. (2020). An Axiom SNP genotyping array for Douglas-fir. BMC Genomics, 21(1), 9. https://doi.org/10.1186/s12864-019-6383-9

Howes, L. (2019). Can Laboratories Move Away from Single-Use Plastic? ACS Central Science, 5(12), 1904–1906. https://doi.org/10.1021/acscentsci.9b01249

Huang, R., Zeng, W., Deng, H., Hu, D., Wang, R., & Zheng, H. (2022). Inbreeding in Chinese Fir: Insight into the Rare Self-Fertilizing Event from a Genetic View. Genes, 13(11):2105. https://doi.org/10.3390/genes13112105

Brookfield, J. (1996). Population genetics. Current Biology, 6(4), 354-356

Karam, M.-J., Lefèvre, F., Dagher-Kharrat, M. B., Pinosio, S., & Vendramin, G. g. (2015). Genomic exploration and molecular marker development in a large and complex conifer genome using RADseq and mRNAseq. Molecular Ecology Resources, 15(3), 601–612. https://doi.org/10.1111/1755-0998.12329

Kerr, G., Boswell, R., & Mason, B. (2003). A sampling system to monitor the transformation from even-aged stands to continuous cover. Forestry, 76(4), 425–435. https://doi.org/10.1093/forestry/76.4.425

Kirby, K. J., & Watkins, C. (Eds.). (2015). *Europe's changing woods and forests: From wildwood to cultural landscapes*. CAB International.

Klapwijk, M. J., & Björkman, C. (2018). Mixed forests to mitigate risk of insect outbreaks. Scandinavian Journal of Forest Research, 33(8), 772–780. https://doi.org/10.1080/02827581.2018.1502805

Konnert M., Maurer W., Degen B., & Kätzel R (2011). Genetic monitoring in forests - early warning and controlling system for ecosystemic changes. iForest 4: 77-81. - doi: 10.3832/ifor0571-004

Kremer, A., Potts, B. M., & Delzon, S. (2014). Genetic divergence in forest trees: Understanding the consequences of climate change. Functional Ecology, 28(1), 22–36. https://doi.org/10.1111/1365-2435.12169

Kremer, A., Ronce, O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., Bridle, J. R., Gomulkiewicz, R., Klein, E. K., Ritland, K., Kuparinen, A., Gerber, S., & Schueler, S. (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecology Letters, 15(4), 378–392. https://doi.org/10.1111/j.1461-0248.2012.01746.x

Krutovsky, K. V., & Neale, D. B. (2005). Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in douglas fir. Genetics, 171(4), 2029–2041. https://doi.org/10.1534/genetics.105.044420

Levy, S. E., & Boone, B. E. (2019). Next-Generation Sequencing Strategies. Cold Spring Harbor Perspectives in Medicine, 9(7). https://doi.org/10.1101/CSHPERSPECT.A025791

Mackay, J., Dean, J. F. D., Plomion, C., Peterson, D. G., Cánovas, F. M., Pavy, N., Ingvarsson, P. K., Savolainen, O., Guevara, M. Á., Fluch, S., Vinceti, B., Abarca, D., Díaz-Sala, C., & Cervera, M. T. (2012). Towards decoding the conifer giga-genome. Plant Molecular Biology, 80(6), 555–569. https://doi.org/10.1007/s11103-012-9961-7

Mason, B., & Kerr, G. (2000). Transforming Even-aged Conifer Stands to Continuous Cover Management. 40, 1–8.

Mason, B., Kerr, G., Pommerening, A., Edwards, C., Hale, S. E., Ireland, D., & Moore, R. (2003). Continuous cover forestry in British conifer forests. Forest Research Annual Report and Accounts 2003-2004, January 2014, 38–53.

Mason, B., Kerr, G., & Simpson, J. (1999). What is Continuous Cover Forestry ? Forestry Commission Information Note 29. Forestry Commission, Edinburgh.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Molecular Ecology Resources, 15(1), 28–41. https://doi.org/10.1111/1755-0998.12291

Meilan, R., & Kirst, M. (2020). Forest-tree population genomics. In Forest genomics and biotechnology (pp. 87–108). CABI. https://doi.org/10.1079/9781780643502.0087

Millar, C. I., & Stephenson, N. L. (2015). Temperate forest health in an era of emerging megadisturbance. Science, 349(6250), 823–826. https://doi.org/10.1126/science.aaa9933

Mills, L. S. (2007). Conservation of Wildlife Populations: Demography, Genetics, and Management, 2nd Edition | Wiley. Wiley.Com. https://www.wiley.com/en-us/Conservation+of+Wildlife+Populations%3A+Demography%2C+Genetics%2C+and+Management%2C+2nd+Edition-p-9781118406670

Mora-Márquez, F., Nuño, J. C., Soto, Á., & Heredia, U. L. D. (2023). Missing genotype imputation in non-model species using Self-Organizing Maps [Preprint]. Preprints. https://doi.org/10.22541/au.168596316.62218099/v1

Müller, Freund, F., Wildhagen, H., & Schmid, K. J. (2015). Targeted re-sequencing of five Douglas-fir provenances reveals population structure and putative target genes of positive selection. Tree

Genetics and Genomes, 11(1). https://doi.org/10.1007/s11295-014-0816-z

Müller, M., Kües, U., Budde, K. B., & Gailing, O. (2023). Applying molecular and genetic methods to trees and their fungal communities. Applied Microbiology and Biotechnology, 107(9), 2783–2830. https://doi.org/10.1007/s00253-023-12480-w

Müller, Ensminger, I., & Schmid, K. J. (2012). A catalogue of putative unique transcripts from Douglas-fir (*Pseudotsuga menziesii*) based on 454 transcriptome sequencing of genetically diverse, drought stressed seedlings. BMC Genomics, 13(1). https://doi.org/10.1186/1471-2164-13-673

Neale, D. B., & Kremer, A. (2011). Forest tree genomics: Growing resources and applications. Nat Rev Genet 12, 111–122 (2011). https://doi.org/10.1038/nrg2931

Neale, D. B., Langley, C. H., Salzberg, S. L., & Wegrzyn, J. L. (2013). Open access to tree genomes: The path to a better forest. Genome Biology, 14(6), 120. https://doi.org/10.1186/gb-2013-14-6-120

Neale, D. B., McGuire, P. E., Wheeler, N. C., Stevens, K. A., Crepeau, M. W., Cardeno, C., Zimin, A. V., Puiu, D., Pertea, G. M., Sezen, U. U., Casola, C., Koralewski, T. E., Paul, R., Gonzalez-Ibeas, D., Zaman, S., Cronn, R., Yandell, M., Holt, C., Langley, C. H., … Wegrzyn, J. L. (2017). The Douglas-Fir genome sequence reveals specialization of the photosynthetic apparatus in Pinaceae. G3: Genes, Genomes, Genetics, 7(9), 3157–3167. https://doi.org/10.1534/g3.117.300078

Neophytou, C., van Loo, M., & Hasenauer, H. (2019). Genetic diversity in introduced Douglas-fir and its natural regeneration in Central Europe. Forestry: An International Journal of Forest Research, 1–10. https://doi.org/10.1093/forestry/cpz055

O'Brien, D., Laikre, L., Hoban, S., Bruford, M. W., Ekblom, R., Fischer, M. C., Hall, J., Hvilsom, C., Hollingsworth, P. M., Kershaw, F., Mittan, C. S., Mukassabi, T. A., Ogden, R., Segelbacher, G., Shaw, R. E., Vernesi, C., & MacDonald, A. J. (2022). Bringing together approaches to reporting on within species genetic diversity. Journal of Applied Ecology, 59(9), 2227–2233. https://doi.org/10.1111/1365-2664.14225

O'Connell, L. M., & Ritland, C. E. (2000). Characterization of microsatellite loci in western redcedar

(*Thuja plicata*). Molecular Ecology, 9(11), 1920–1922. https://doi.org/10.1046/j.1365-294x.2000.01055-2.x

O'Connell, L. M., Russell, J., & Ritland, K. (2004). Fine-scale estimation of outcrossing in western redcedar with microsatellite assay of bulked DNA. Heredity, 93(5), 443–449. https://doi.org/10.1038/sj.hdy.6800521

O'Connell, L. M., Viard, F., Russell, J., & Ritland, K. (2001). The mating system in natural populations of western redcedar (*Thuja plicata*). Canadian Journal of Botany, 79(6), 753–756. https://doi.org/10.1139/cjb-79-6-753

Pakull, B., Eusemann, P., Wojacki, J., Ahnert, D., & Liesebach, H. (2021). Genetic diversity of seeds from four German Douglas fir (*Pseudotsuga menziesii*) seed orchards. European Journal of Forest Research, 140(6), 1543–1557. https://doi.org/10.1007/s10342-021-01419-3

Peters, R. L. (1990). Effects of global warming on forests. Forest Ecology and Management, 35(1–2), 13–33. https://doi.org/10.1016/0378-1127(90)90229-5

Pfenninger, M., Reuss, F., Kiebler, A., Schönnenbeck, P., Caliendo, C., Gerber, S., Cocchiararo, B., Reuter, S., Blüthgen, N., Mody, K., Mishra, B., Bálint, M., Thines, M., & Feldmeyer, B. (2021). Genomic basis for drought resistance in European beech forests threatened by climate change. eLife, 10, e65532. https://doi.org/10.7554/eLife.65532

Ramsfield, T. D., Bentz, B. J., Faccoli, M., Jactel, H., & Brockerhoff, E. G. (2016). Forest health in a changing world: Effects of globalization and climate change on forest insect and pathogen impacts. Forestry, 89(3), 245–252. https://doi.org/10.1093/forestry/cpw018

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. Molecular Ecology, 24(17), 4348–4370. https://doi.org/10.1111/mec.13322

Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Menzel, A., Root, T. L., Estrella, N., Seguin, B., Tryjanowski, P., Liu, C., Rawlins, S., & Imeson, A. (2008). Attributing physical and biological impacts to anthropogenic climate change. Nature, 453(7193), 353–357. https://doi.org/10.1038/nature06937

Sawyer, A. (2019). The unsustainable lab. BioTechniques, 66(1), 5–7.

https://doi.org/10.2144/btn-2018-0185

Shalev, T. J., Gamal El-Dien, O., Yuen, M. M. S., Shengqiang, S., Jackman, S. D., Warren, R. L., Coombe, L., van der Merwe, L., Stewart, A., Boston, L. B., Plott, C., Jenkins, J., He, G., Yan, J., Yan, M., Guo, J., Breinholt, J. W., Neves, L. G., Grimwood, J., … Bohlmann, J. (2022). The western redcedar genome reveals low genetic diversity in a self-compatible conifer. Genome Research, 1952–1964. https://doi.org/10.1101/gr.276358.121

Shalev, T. J., Gamal El-Dien, O., Yuen, M. M. S., van der Merwe, L., Kirst, M., Yanchuk, A. D., Ritland, C., Russell, J. H., & Bohlmann, J. (2023). Genetic architecture of terpene chemistry and growth traits and the impact of inbreeding on these traits in western redcedar (*Thuja plicata*). Evolutionary Applications. https://doi.org/10.1111/EVA.13526

Slavov, G. T., Howe, G. T., Yakovlev, I., Edwards, K. J., Krutovskii, K. V., Tuskan, G. A., Carlson, J. E., Strauss, S. H., & Adams, W. T. (2004). Highly variable SSR markers in Douglas-fir: Mendelian inheritance and map locations. Theoretical and Applied Genetics. 108(5), 873–880. https://doi.org/10.1007/s00122-003-1490-y

Straw, N. A., Williams, D. T., Fielding, N. J., Jukes, M., Connolly, T., & Forster, J. (2017). The influence of forest management on the abundance and diversity of hoverflies in commercial plantations of Sitka spruce: The importance of sampling in the canopy. Forest Ecology and Management, 406, 95–111. https://doi.org/10.1016/j.foreco.2017.06.010

Taylor, S. A., Larson, E. L., & Harrison, R. G. (2015). Hybrid zones: Windows on climate change. Trends in Ecology and Evolution, 30(7), 398–406. https://doi.org/10.1016/j.tree.2015.04.010

Telfer, E., Graham, N., Stanbra, L., Manley, T., & Wilcox, P. (2013). Extraction of high purity genomic DNA from pine for use in a high-throughput Genotyping Platform. New Zealand Journal of Forestry Science, 43(1), 3. https://doi.org/10.1186/1179-5395-43-3

Tew, E. R., Ambrose-Oji, B., Beatty, M., Büntgen, U., Butterworth, H., Clover, G., Cook, D., Dauksta, D., Day, W., Deakin, J., Field, A., Gardiner, B., Harrop, P., Healey, J. R., Heaton, R., Hemery, G., Hill, L., Hughes, O., Khaira-Creswell, P. K., … Sutherland, W. J. (2023). A horizon scan of issues affecting UK forest management within 50 years. Forestry: An International Journal of Forest Research, cpad047. https://doi.org/10.1093/forestry/cpad047

Torkamaneh, D., Boyle, B., St-Cyr, J., Légaré, G., Pomerleau, S., & Belzile, F. (2020). NanoGBS: A Miniaturized Procedure for GBS Library Preparation. Frontiers in Genetics, 11. https://doi.org/10.3389/FGENE.2020.00067

Urbina, M. A., Watts, A. J. R., & Reardon, E. E. (2015). Labs should cut plastic waste too. Nature, 528(7583), 479–479. https://doi.org/10.1038/528479c

van Loo, M., Hintsteiner, W., Pötzelsberger, E., Schüler, S., & Hasenauer, H. (2015). Intervarietal and intravarietal genetic structure in Douglas-fir: Nuclear SSRs bring novel insights into past population demographic processes, phylogeography, and intervarietal hybridization. Ecology and Evolution, 5(9), 1802–1817. https://doi.org/10.1002/ece3.1435

van Loo, M., Lazic, D., Chakraborty, D., Hasenauer, H., & Schüler, S. (2019). North American Douglas-fir (*P. menziesii*) in Europe: Establishment and reproduction within new geographic space without consequences for its genetic diversity. Biological Invasions, 21(11), 3249–3267. https://doi.org/10.1007/s10530-019-02045-2

Vega-Vela, N. E., & Chacón Sánchez, M. I. (2011). Isolation of high-quality DNA in 16 aromatic and medicinal Colombian species using silica-based extraction columns. Agronomía Colombiana, 29(3), 349–357.

Volney, W. J. A., & Fleming, R. A. (2000). Climate change and impacts of boreal forest insects. Agriculture, Ecosystems and Environment, 82(1–3), 283–294. https://doi.org/10.1016/S0167-8809(00)00232-2

Wang, T., & Russel, J. H. (2006). Evaluation of selfing effects on western redcedar growth and yield in operational plantations using the tree and stand simulator (TASS). Forest Science, 52(3), 281–289. https://doi.org/10.5849/forsci.15-042

Waples, R. S. (2022). What Is Ne, Anyway? Journal of Heredity, 113(4), 371–379. https://doi.org/10.1093/jhered/esac023

Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. Evolutionary Applications, 3(3), 244–262. https://doi.org/10.1111/j.1752-4571.2009.00104.x

Woudstra, Y., Tumas, H., Ghelder, C. van, Hung, T. H., Ilska, J. J., Girardi, S., A'Hara, S., McLean, P.,

Cottrell, J., Bohlmann, J., Bousquet, J., Birol, I., Woolliams, J. A., & MacKay, J. J. (2023). *Conifers concentrate large numbers of NLR immune receptor genes on one chromosome* (p. 2023.10.20.563305). bioRxiv. https://doi.org/10.1101/2023.10.20.563305

Wright, S. (1950). The genetical structure of populations. Annals of Eugenics, 31(111), 323–354.

Zimmerman, S. J., Aldridge, C. L., & Oyler-Mccance, S. J. (2020). An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. BMC Genomics, 21(1), 1–16. https://doi.org/10.1186/s12864-020-06783-9

Chapter 2. Comparing DNA isolation methods for forest trees: quality, plastic footprint, and time-efficiency

Laura Guillardin[1], John J. MacKay[1]

1. Department of Biology, University of Oxford, Oxford OX1 3RB, United Kingdom

2.1 Abstract

Genetic and genomic studies are seeing an increase in sample sizes together with a wider range of species investigated in response to environmental change concerns. In turn, these changes may come with challenges including the time and difficulty of isolating nucleic acids (DNA or RNA), the sequencing cost and environmental impacts of the growing amount of plastic waste generated in the process. *Pseudotsuga menziesii* (Mirbel) Franco (PM), *Tsuga heterophylla* (Raf.) Sarg. (TH) and *Thuja plicata* Donn ex D.Don (TP) are conifer species found in diverse woodlands both as natives and naturalised exotics. Our study was carried out whilst investigating their genetics to understand their genetic diversity and potential for adaptation. In the present study, we compared two different DNA isolation methods, i.e., spin-column DNeasy plant mini kit (QIAGEN), and temperature-driven enzymatic cocktail Plant DNA Extraction (MicroGEM). The quantity of recovered DNA and the quality of DNA were assessed along with the plastic footprint and time needed for the three tree species. Both methods were optimised and proven to provide enough DNA for each studied species. The yield of DNA for each method depended on the species: QIAGEN showed higher yield in *P. menziesii* and *T. heterophylla*, while *T. plicata* recovered a similar amount of DNA for both methods. The DNA quality was investigated using DNA barcoding techniques by confirming species identity and species discrimination. No difference was detected in the PCR amplification of the two barcoding loci, (*rbcL* and *trnH-psbA*), and the recovered sequences between DNA isolation methods. Measurement of the plastic use and the processing time per sample indicated that MicroGEM had a 52.64% lower plastic footprint and was 51.8 % faster than QIAGEN. QIAGEN gave higher yields in two of the species although both methods showed similar quality results across all species. However, MicroGEM was clearly advantageous to decrease the plastic footprint and improve the time efficiency. Overall, MicroGEM recovers sufficient and reliable DNA to perform common downstream analyses such as PCR and sequencing. Our findings illustrate the benefits of research and efforts towards developing more sustainable methods and techniques to reduce the environmental footprint of molecular analyses.

2.2 Background

Genetic and genomic analyses often benefit from a large sample set to reliably detect associations and identify patterns relevant to key biological questions (Fumagalli, 2013) due to higher statistical power and accuracy (Hong & Park, 2012). However, larger sample sizes require more resources, such as time to collect and process the samples and the volume of consumables used in the laboratory (Alves et al., 2021). Analyses commonly used include Genome-wide association studies (GWAS) (Visscher et al., 2017), genomic selection (GS) and Genome environment association (GEA) to scan genomes for associations with complex traits of interest as in tree breeding programs (Müller et al., 2019) where the power to detect loci with small effects may be constrained by the sample size (Tan & Ingvarsson, 2022). The minimum sample size for each genetic analysis varies depending on the research question, the genetic markers used, the biological system and the available resources. Environmental change and its effects on natural habitats as well as in the distribution, population size and genetic diversity (Brown et al., 2020; Garden et al., 2015; Harley, 2011; Heuertz et al., 2023; Pauls et al., 2013) increases the number of species at risk and that need to be studied (Harley, 2011). Therefore, both the range and the size of studies involving molecular analyses are increasing (Alberto et al., 2013).

Molecular studies on non-model tree species may face challenges including obtaining high-quality DNA and RNA, accessing wild populations, the complexity and size of their genome, and the limited available genomic resources (Holliday et al., 2017). These factors increase the complexity and the resource intensity of genomic studies in trees compared to studies of many other organisms (Abril et al., 2011). The challenges of obtaining high-quality DNA and RNA from trees include the difficulty of breaking down the cell walls to release the nucleic acids (Rezadoost et al., 2016) and the inhibitory effects of secondary metabolites and polysaccharides on downstream applications such as polymerase chain reaction (PCR), sequencing and genotyping. Simple and robust protocols for isolation of high-quality nucleic acids from tree tissues that are suitable for downstream applications would help to overcome these challenges (Abdel-Latif, A., & Osman, G., 2017)

.

The main limitations and concerns when designing genomic experiments with large sample sizes include the time and difficulty to isolate nucleic acids (DNA or RNA), the sequencing cost and the environmental impact of plastic waste generated in the process. Sequencing platforms have reduced nucleic acid sequencing costs over the last two decades (Levy & Boone, 2019), making sequencing more accessible to researchers studying non-model species and feasible to study larger sample sizes. Some cost-effective pipelines, workflows and strategies have also been developed to reduce costs and process time by reusing consumables (Fu et al., 2017). Single-use plastic consumables are a significant source of waste in many scientific fields (Howes, 2019; Sawyer, 2019), with an estimated use of 5.5 million tonnes per year worldwide (Urbina et al., 2015); therefore, although they are essential to perform genomic analyses, they have a large environmental impact (Di Paolo et al., 2022). Improved laboratory sustainability is encouraged by organisations and labels such as My Green Lab (https://www.mygreenlab.org/) and The Sustainable Laboratory Practices Working Group (SLPWG) which promote green practices in the laboratory to reduce their plastic footprint. A different approach is to reuse specific plastic items, for example, Grenova Solutions (https://grenovasolutions.com/) developed an on-bench Pipette Tip Cleaning Machine (TipNovus) that is easy to integrate into laboratory routines. Another solution is to choose or develop protocols that use less plastic without compromising the experiment's outcome (Alves et al., 2021; Torkamaneh et al., 2020).

Several methods have been proposed to overcome the difficulties when isolating DNA from tree species. Broadly, there are five main types of DNA isolation systems including organic extraction methods which use organic solvents like phenol and chloroform (Doyle, 1990; Green & Sambrook, 2017), solid-phase extraction methods that use solid matrices, such as silica to bind and purify the DNA (Vega-Vela & Chacón Sánchez, 2011), precipitation methods which use salts and ethanol to precipitate the DNA (Green & Sambrook, 2016), enzymatic digestion methods that use individual or a combination of enzymes to digest the samples and release the DNA (Manen et al., 2005) and use of magnetic bead-based coated with an agent which binds with DNA and isolates it from the cellular suspension (Azimi et al., 2011), or magnetic ionic liquids (Emaus et al., 2022). These methods are

usually combined and modified depending on the experimenter's needs (Fu et al., 2017; Li et al., 2013).

In this study, we compared the performance of two different DNA isolation methods in three conifer forest tree species: *P. menziesii* (PM), *T. heterophylla* (TH) and *T. plicata* (TP). They are all large evergreen coniferous trees native to western North America. Coniferous species, including the study species, are widely found in Europe and North America in woodlands which provide a variety of ecosystem services including timber production, carbon sequestration, biodiversity and habitat conservation, water regulation, recreation, and tourism (Bastien, 2019; Mina et al., 2017). Additionally, conifer trees are commonly used as pioneer species in reforestation projects in poorer sites due to their resilience to extreme environmental conditions (Vadell et al., 2016). At the same time, conifer species are affected by the change in the climate (Alberto et al., 2013) and its subsequent increasing disturbances such as the increase of forest wildfires and their intensity, longer and higher number of drought periods and pests and disease outbreaks. Therefore, conifers are largely studied organisms to shed light on the genomic mechanisms of adaptation, the association to different environments, the levels of population genetic diversity of woodlands and the genetic resistance to pests & diseases, among others (Alberto et al., 2013).

We compared the recovered DNA quantity of the original and modified protocols: i) Plant DNA Extraction (MicroGEM International PLC, 2019) (MicroGEM) which uses a temperature-controlled enzymatic cocktail and ii) DNeasy plant mini kit (Qiagen USA, Valencia, CA) (QIAGEN), a silica column-based method widely used to isolate DNA from conifer trees. In this context, this study aimed to find an efficient DNA isolation method for forest trees by looking at the DNA quality and its potential downstream use, the plastic waste generated, the energy required to produce it and the total time needed to process the samples using both modified DNA isolation methods. To verified the potential downstream use of the DNA recovered we used PCR techniques with gene-specific primers followed by DNA sequencing.

2.3 Methods

2.3.1 Plant material

The study species are the conifers *T. plicata*, *P. menziesii* and *T. heterophylla*. They are all part of the Pinales order; *T. plicata* belongs to the Cupressaceae, whilst *P. menziesii* and *T. heterophylla* belong to the Pinaceae. Four samples per species (Figure 1) were retrieved from different planted stands in England using an arborist slingshot following a modified method based on Youngentob et al., (2016). The leaves collected were dried by storing them in silica gel beads after removing them from the tree and until processing for DNA isolation.

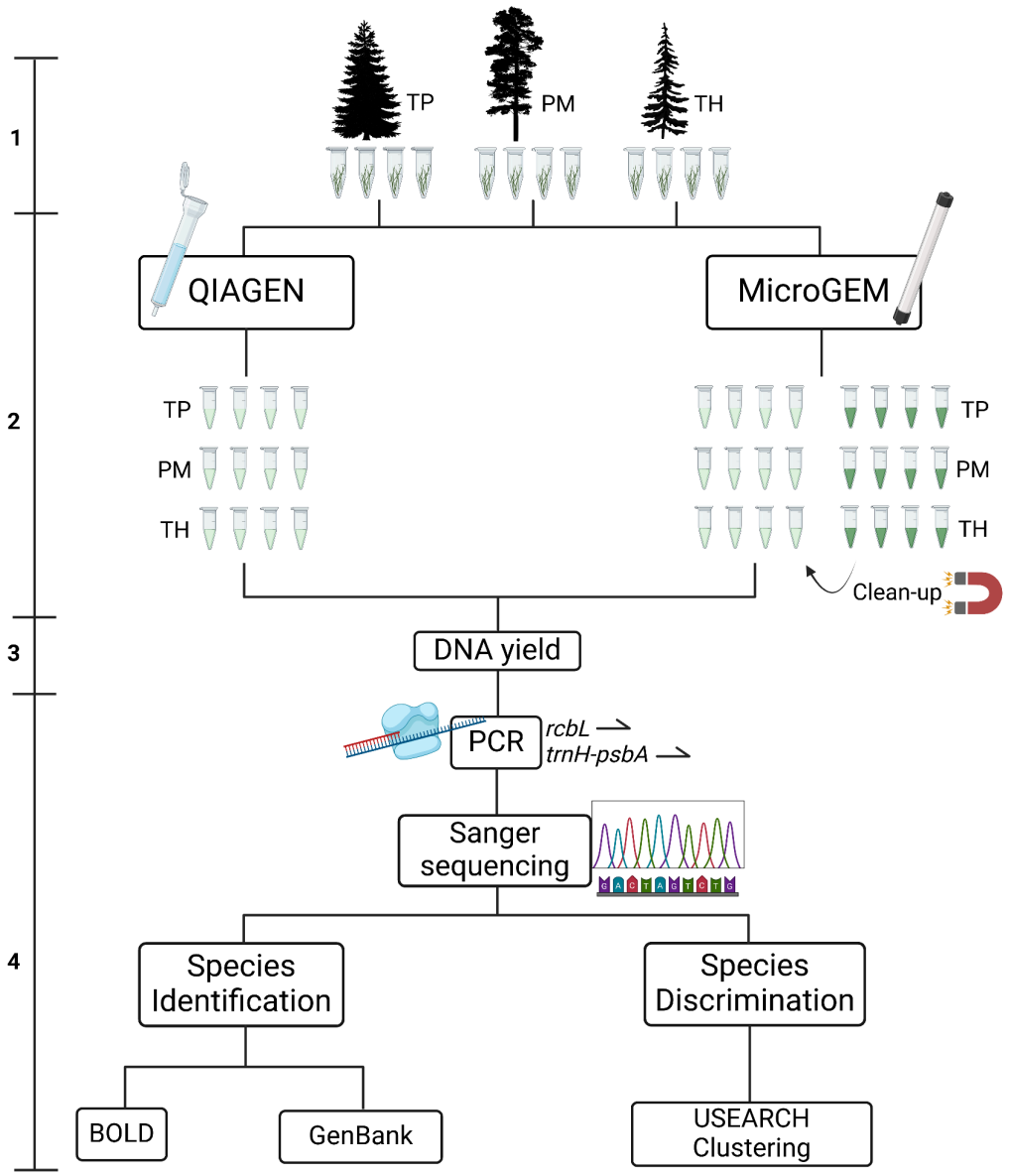


Figure 1. Diagram of the methodology followed. Illustration created with BioRender.com. The methodology and sequence of major steps implemented across this study, included: 1) plant material collection, 2) DNA isolation, 3) DNA yield quantification and 4) DNA barcoding to test the DNA potential use for downstream analysis, including species identification and species discrimination by performing PCR and followed by DNA sequencing. TP: *T. plicata*, PM: *P. menziesii* and TH: *T. heterophylla*.

2.3.2 DNA isolation procedures and improvements

We isolated the DNA using 20 mg of leaf tissue from the 12 samples (Figure 1) by using the manufacturer's instructions for the DNeasy plant mini kit (Qiagen USA, Valencia, CA) (QIAGEN), we also modified the procedure to enhance the quality and yield of the DNA recovered, by adding 20 µl of Protease K in the lysate step and increasing the incubation time to a minimum of an hour at 55°C (https://lauguillardin.github.io/2022-07-12-QIAGEN_DNAextraction_Pmenziesii/).

We processed the same 12 samples using the DNA isolation method Plant DNA Extraction (MicroGEM International PLC, 2019) (MicroGEM) (Figure 1) following the manufacturer's original instructions. First, 10 mg of leaf tissue is mixed with the lysis buffer and using a mechanical homogenizer to break the cells, second, the lysate is transferred to a PDQeX DNA-binding cartridge and third, the enzymatic cocktail is added to the cartridge, which is inserted into the PDQeX device which through the combination of changes in temperature and the physical design of the cartridge performs a chemical process that releases the DNA into the recovery tubes. We also modified this procedure to optimise the DNA yield as follows: 1) The tissue was disrupted alone and added with a micro scoop to the pre-sample mix which included the lysate buffer and the enzymatic cocktail in individual tubes, 2) the sample mix was pipetted into the cartridges, which were inserted into the PDQex device with a modified plant programme with an incubation time increased from 5 min to 15 min and extraction time from 5 min to 10 min (https://lauguillardin.github.io/2022-07-12-microGEM-DNAextraction_Thujaplicata/).

After recovering the DNA we performed a clean-up step by using the AMPpureXP beads (Deangelis et al., 1995), which use magnetic particles that bind the DNA to allow the clean-up. A concentration of 0.8x of AMPureXP beads (Beckman Coulter™, Brea, CA, USA) was used for the size selection needed to retrieve the isolated DNA fragments (0.8x ratio recommended by NewEngland Biolabs).

2.3.3 DNA recovery

The DNA concentration from all samples was quantified using Qubit® Fluorometer v.4.0 (Invitrogen™, Thermo Fisher Scientific, Cleveland, OH, USA). The DNA yield present in the final elution (μg) was calculated by multiplying the DNA concentration by the final volume eluted. The yield per unit mass (ng DNA/mg tissue) was estimated dividing the DNA yield (ng of DNA recovered) by the amount of mass used as starting material (mg).

The following analyses were performed using the samples isolated using the modified protocols. DNA quality of these samples was visualised by electrophoresis in 0.8% Agarose gels (TAE Buffer), which confirmed the DNA presence. We performed an ANOVA test (error type iii) using the R (R Core Team, 2018) package 'ape' (Paradis & Schliep, 2019) to analyse the differences in DNA yield recovered between methods and species followed by a post-hoc Tukey's test (Keselman & Rogan, 1977) for comparing all possible group pairings. To visualise the data, we used the R package ggplot2 (Wickham, 2009).

2.3.4 DNA analyses

To verify the potential use of the DNA recovered in downstream analyses we assessed the DNA amplifiability and its potential to be sequenced by using DNA barcoding techniques (Figure 1). The PCR amplification was performed with the intergenic spacer primers situated on the chloroplast DNA: the large subunit of the ribulose-bisphosphate carboxylase gene (*rbcL*) and the plastid *trnH-psbA*, and sequencing on each sample. We adjusted the DNA concentrations to 20ng/μl, when possible, for PCR amplification of *rbcL* and *trnH-psbA* barcoding loci (Table 1). We used Q5® High-Fidelity 2X Master Mix DNA Polymerase (NewEngland Biolabs®, Ipswich, MA, USA) and added 100 ng of template DNA and 0.5 μM of each primer into a 25 μl final reaction following the manufacturer's instructions. The thermocycler conditions were 98°C for 20 seconds, followed by 40 cycles starting at 98°C for 10 seconds, TA (*rbcL*: 54°C and *trnH-psbA*: 63°C, Table 1) for 30 seconds and 72°C for 30 seconds, with a final extension at 72°C for 2 minutes. The amplified regions were visualised by electrophoresis in a 1.8% Agarose gel with SYBR safe (Invitrogen™, Thermo Fisher

Scientific, Cleveland, OH, USA) as stain. We estimated visualy the proportion (%) of major and minor bands detected on the result gels.

Table 1. Locus and primer details for PCR. TA: Annealing temperature used.

Marker name	Primer's sequence	TA (°C)	Expected product size (bp)	Source
trnH-psbA	CGCGCATGGTGGATTCACAATCC GTTATGCATGAACGTAATGCTC	63	500	(Armenise et al., 2012)
rbcL	ATGTCACCACAAACAGAAAC TCGCATGTACCTGCAGTAGC	54	750	(Armenise et al., 2012)

PCR products of two samples per species for each of the extraction methods were analysed by Sanger sequencing (Source Biosciences). Pre-cleaning and post-quality checks were performed by the sequencing service facilitator. Retrieved sequences were visualised and the electropherograms manually checked by using SnapGene software (https://www.snapgene.com/). We exported the fasta files from the ab1 files which were trimmed with Trimmomatic (Bolger et al., 2014) and aligned with MUSCLE (Edgar, 2004). We estimated the proportion of samples retrieved with the expected sequence length from all the samples submitted.

We submitted the *rbcL* trimmed sequences to The Barcode of Life Data system v4 (BOLD) (Ratnasingham & Hebert, 2007) and the *trnH-psbA* to the nucleotide dataset from GenBank (Clark et al., 2016). BOLD enables a species-level identification of plant taxa by submitting queries of *MatK* and *rbcL* sequences to be searched against their reference library through their sequence threshold identification system. We verified the species discrimination efficacy of the *trnH-psbA* barcode by clustering the aligned sequences together by similarity using the software USEARCH v.11 (Edgar & Bateman, 2010). The UCLUST algorithm divides a set of sequences into clusters providing a high-throughput species-level discrimination. The minimum similarity level was set to 99%.

2.3.5 Plastic footprint and time needed for DNA isolations

The plastic used to isolate DNA with both modified methods was calculated for a single reaction by weighing every plastic item with a precision balance (Table 2) not including the packaging.

Table 2. Weight of plastic items, and number of items (N) needed per DNA isolation method (QIAGEN and MicroGEM) to process one sample.

Plastic item	Weight (g)	QIAGEN (N)	MicroGEM (N)
Cartridge	1.183	0.00	1.00
2ml tube	1.089	1.00	1.00
1.5mL tube	0.940	1.00	0.00
2mL no-Lid tube	0.932	3.00	0.00
Column	0.734	2.00	0.00
0.5mL tube	0.461	0.00	1.00
0.2mL tube	0.153	0.00	1.00
1000ul tip	1.033	5.13	0.17
200ul tip	0.315	2.04	0.08
20ul tip	0.219	0.04	0.00
10ul tip	0.275	1.00	0.00
cut tip	0.311	0.00	1.00
Total		15.21	5.25

The carbon dioxide emitted to produce 1 kg of polypropylene plastic commonly used to make laboratory supplies has been discussed by Harding et al., (2007) and 3.4 kg CO_2 is generally accepted. The total energy required to produce 1 kg of plastic from the extraction of raw materials to the final manufactured product is 85.9 MJ (Harding et al., 2007). Following these data, we calculated the kg of CO_2 emissions and total energy required to process one sample with each DNA isolation method based on the amount of plastic needed in each case.

The time required to isolate DNA by using both modified methods was timed for each step in the procedures when processing 24 samples and calculated for a single reaction. Incubation time and centrifuge steps time are fixed and remain the same for either 1 or 24 samples. We divided the hands-on time by 24 samples and added it to the measured fixed time.

2.4 Results

2.4.1 Improvement of DNA isolation methods and comparison of species

We isolated the DNA from four samples in each of the three study species by using MicroGEM and QIAGEN original and improved protocols (Figure 2). DNA samples isolated with QIAGEN showed higher levels of brightness than the ones isolated with MicroGEM in the agarose gel (Figure 2), which correlates with the overall higher levels of concentration found (Table 3). Samples 11 and 12 (PM3 and PM4) isolated with MicroGEM (Figure 2 panel B) appeared empty in the gel, which may be due to a missplace of the sample in the loading process.

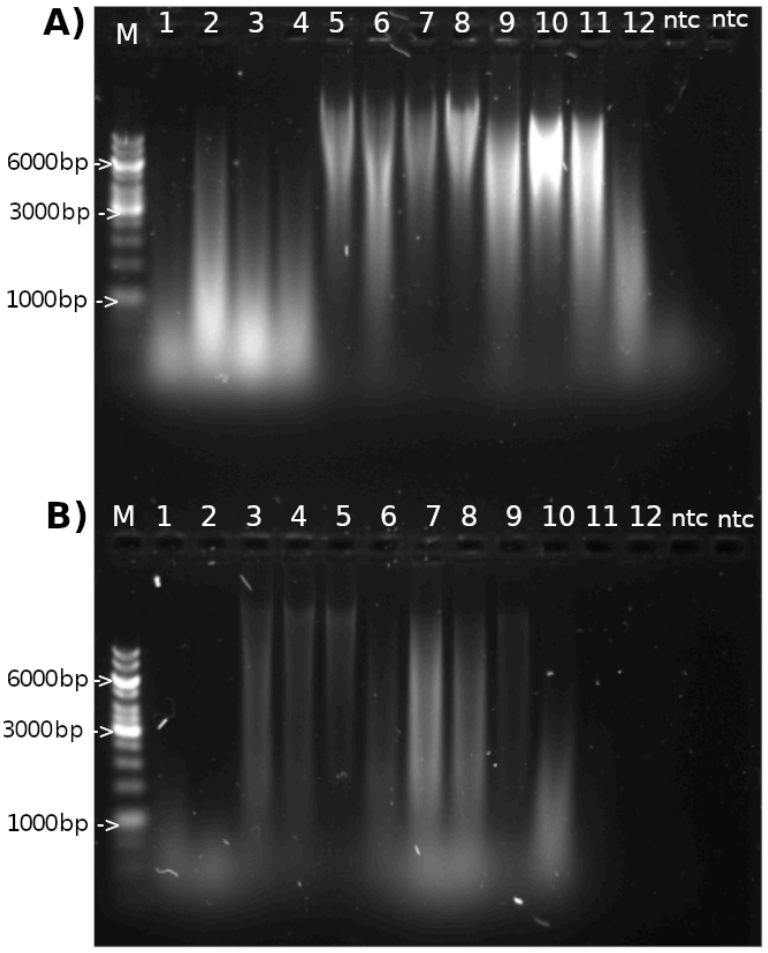


Figure 2. Agarose gel with the DNA isolated using A) QIAGEN and B)MicroGEM modified methods. M: 1Kb molecular-weight size marker, 1:TH1, 2:TH2, 3:TH3, 4:TH4, 5:TP1, 6:TP2, 7:TP3, 8:TP4, 9:PM1, 10:PM2, 11:PM3, 12:PM4, ntc: non-template control

The improvements to DNA isolation methods increased the DNA recovered both for the MicroGEM and QIAGEN methods in *T. plicata*, *P. menziesii* and *T. heterophylla* (Figure 3). The data suggest that both measures of DNA yield (final DNA yield eluted and yield per unit mass of starting material) were generally higher with QIAGEN. The recovered DNA yield per unit mass of starting material was higher when using MicroGEM in *T. plicata* (Figure 3, B and D and Table 3). However, statistical comparison of modified methods and species showed significant effects of species and the interaction of species and methods but not between methods (Table 4). A post hoc Tukey test among species within methods showed that *P. menziesii* yielded significantly better than the other species when using the original QIAGEN protocol, while *T. plicata* showed significantly better results when using the original MicroGEM method (Figure 3, A and C). The modified QIAGEN showed *P. menziesii* and *T. heterophylla* grouping together and yielding significantly higher than *T. plicata* results while with MicroGEM modified protocol *T. plicata* showed significantly higher yields compared to the other species (Figure 3, B and D).

Overall, the modified procedures yielded more than 1 µg across all the species starting from a minimum of 10 mg when using MicroGEM and 20 mg with QIAGEN of dried tissue (except for TH4 extracted with MicroGEM,), which is sufficient for many downstream analyses. Yields above 10 µg were obtained in *P. menziesii* and most of *T. heterophylla* but only for QIAGEN method (Table 3). The samples extracted with the modified protocols of both DNA isolation methods will be used for the subsequent analyses.

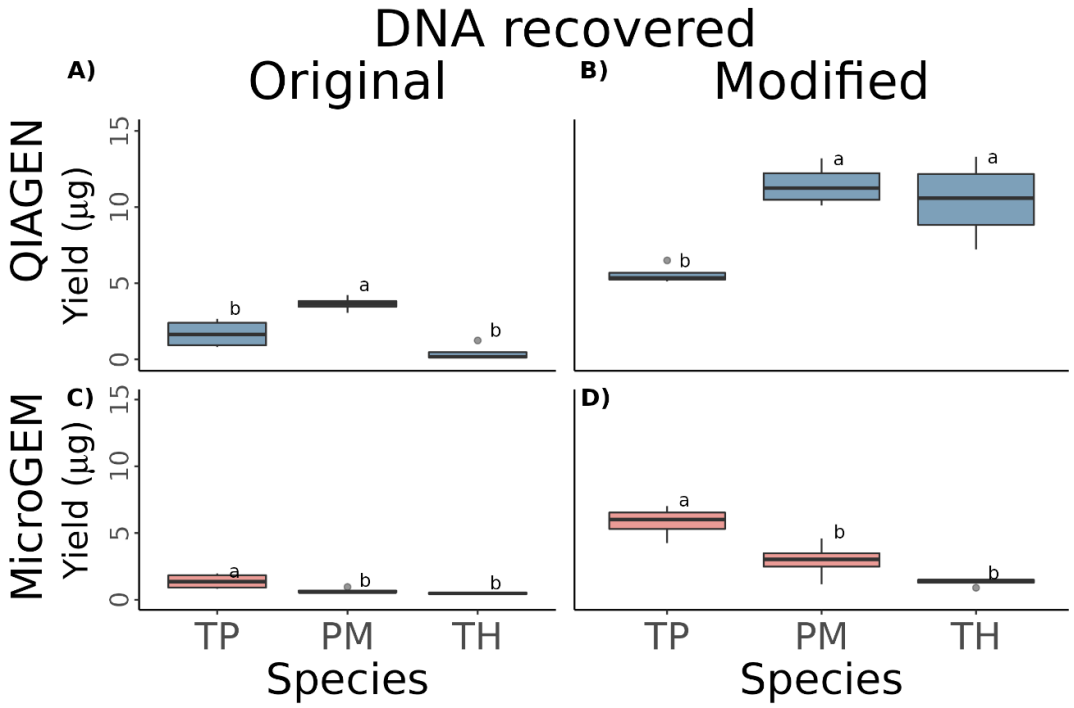


Figure 3. DNA yield recovered by A) using QIAGEN original protocol, B) QIAGEN modified protocol, C) MicroGEM original protocol and D) modified MicroGEM protocol. TP: *T. plicata*, PM: *P. menziesii* and TH: *T. heterophylla*. Letters represent the post hoc test (Tukey test) results showing the significant differences between species for each of the protocols tested.

Table 3. DNA concentration (Conc.), DNA yield in the final eluted volume and yield (ng) per unit mass (mg of starting tissue used: 10mg in MicroGEM and 20 mg in QIAGEN) across samples from each sample with both improved methods.

Sample ID	Conc. (ng/µl) MicroGEM	Yield in final volume (µg) MicroGEM	Yield per unit mass (ng DNA/mg tissue) MicroGEM	Conc.(ng/µl) QIAGEN	Yield in final volume (µg) QIAGEN	Yield per unit mass (ng DNA/mg tissue) QIAGEN
TH1	17.8	1.42	142	72.3	7.23	361.5
TH2	21.2	1.48	148	133.0	13.30	665
TH3	19.3	1.54	154	118.0	11.80	590
TH4	15.1	0.91	91	93.7	9.37	468.5
TP1	78.0	7.02	702	51.1	5.11	255.5
TP2	75.0	6.38	638	65.0	6.50	325
TP3	53.2	4.26	426	52.8	5.28	264
TP4	70.8	5.66	566	54.2	5.42	271
PM1	65.8	4.61	461	119.0	11.90	595
PM2	41.9	2.93	293	106.0	10.60	530
PM3	16.8	1.18	118	132.0	13.20	660
PM4	44.6	3.12	312	101.0	10.10	505

Table 4. Anova table of factors affecting DNA yield recovered including methods, species and their interaction. Sum Sq: Sum of squares, Df: Degrees of freedom, Pr: Probability. '***' corresponds to a P. value of 0.001.

Anova Table (Type III tests)	Sum Sq	Df	F value	Pr (>F)	Significance
(Intercept)	124434025.00	1	57.63	5.13E-07	***
Method	126253.13	1	0.06	8.12E-01	
Species	78713516.67	2	18.23	4.71E-05	***
Method:Species	109300794.75	2	25.31	5.88E-06	***
Residuals	38865031.50	18			

2.4.2 Analyses using the DNA recovered

The potential use of the DNA recovered from each method was analysed with standard DNA Barcoding techniques with commonly used gene-specific primers, the large subunit of the ribulose-bisphosphate carboxylase gene (*rbcL*) and the plastid *trnH-psbA* intergenic spacer. Major amplicons of the expected size (Table 1) were detected in all the samples tested for both barcodes *rbcL* and *trnH-psbA* (Table 5, Figure 4). Other minor bands appeared in the *rbcL* PCR products as displayed in the gel images while none were detected in the *trnH-psbA* gel (Figure 4, Table 5). There were no differences between DNA isolation methods in *trnH-psbA* PCR products, but samples extracted with MicroGEM show minor bands in 75% of the *rbcL* products compared to only 33.3% of QIAGEN samples (Figure 4, A and C).

Every *trnH-psbA* sequence retrieved was of the expected length amplicon size while 25% of the *rbcL* sequences were unreliable (Table 5). When looking at the differences between the DNA isolation methods, one QIAGEN sample and two MicroGEM samples could not be sequenced satisfactorily for *rbcL*.

Table 5. Summary of PCR and Sequencing results for each DNA isolation method per species and barcode marker. N: Number of samples. Proportion (%) was estimated by counting the major and minor bands visualy from the gel (PCR) and by counting the number of samples retrieved with the expected sequence length (Sequencing).

		PCR					Sequencing	
		rbcL		*trnH-psbA*			*rbcL*	*trnH-psbA*
DNA isolation Method	Species	N	Major amplicon detected	Minor bands detected	Major amplicon detected	Minor bands detected	N	Expected seq. length	Expected seq. length
QIAGEN	TP	4	100%	50%	100%	0%	2	100%	100%
	PM		100%	50%	100%	0%		50%	100%
	TH		100%	0%	100%	0%		100%	100%
MicroGEM	TP		100%	75%	100%	0%		50%	100%
	PM		100%	100%	100%	0%		50%	100%
	TH		100%	50%	100%	0%		100%	100%

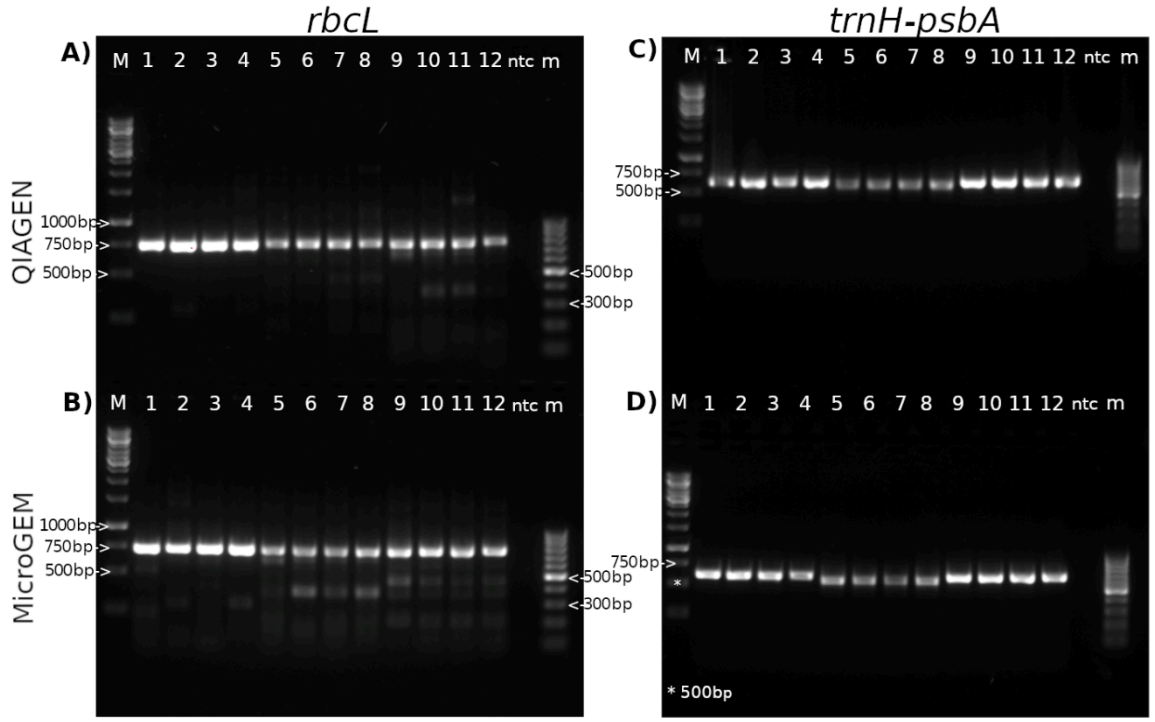


Figure 4. *rbcL* PCR product for A) QIAGEN B) MicroGEM samples and *trnH-psbA* for C) QIAGEN and D) MicroGEM samples. Sample names: M: 1Kb molecular-weight size marker, 1:TH1, 2:TH2, 3:TH3, 4:TH4, 5:TP1, 6:TP2, 7:TP3, 8:TP4, 9:PM1, 10:PM2, 11:PM3, 12:PM4, ntc: non-template control, m: 100bp molecular size marker.

We submitted the *rbcL* and *trnH-psbA* recovered sequences to BOLD and GENBANK, respectively to verify the species identities of the tested samples. All sequences returned the correct species identification, except for *rbcL* in samples whose sequence length was shorter than the expected size (Table 6). Two of the problematic sequences were from samples isolated with MicroGEM (*T. plicata* and *P. menziesii*) and one from QIAGEN (*P. menziesii*) (Table 6).

Table 6. Species identification using GENEBANK and BOLD datasets for both barcodes. N: Number of samples.

DNA isolation method	Species	N	BOLD *rbcL*	GENBANK *trnH-psbA*
QIAGEN	TP	2	2	2
	PM	2	1	2
	TH	2	2	2
MicroGEM	TP	2	1	2
	PM	2	1	2
	TH	2	2	2

The sequences obtained of the *trnH-psbA* amplicons were clustered using USEARCH v.11 to assess the species discrimination efficacy of both methods. A clustering was determined for the three species and the samples were classified in species-specific clusters with a level of similarity above 99% in every case, without any difference between DNA isolation methods.

2.4.3 DNA obtained in large populations

Populations with larger numbers of individuals have been processed in two of the species described above with both methods in a parallel project (Table 7). The QIAGEN method gave higher mean yields for *T. plicata* than *P. menziesii* with the latter also being more consistent as shown by lower levels of variation (CV) (Table 7). The yield recovered from *T. plicata* samples was lower and more variable when using MicroGEM than QIAGEN (Table 7).

Table 7. Summary table of the DNA yield in the final eluted volume and yield (ng) per unit mass (mg of starting tissue used: 10mg in MicroGEM and 20 mg in QIAGEN) recovered in Chapters 3 and 4 when using both DNA isolation methods. CV: Coefficient variation, N: Number of samples.

	Species	Yield in final volume (µg)	Yield per unit mass (ng DNA/mg tissue)	Yield CV	N
QIAGEN	TP	9.603	960.3	59.24%	79
	PM	7.766	388.3	40.37%	1146
MicroGEM	TP	3.312	331.2	66.09%	472

2.4.4 Plastic footprint and time needed for DNA isolations

The plastic consumed was determined for all steps used in each method, including the clean-up step using AMPureXP beads needed after the DNA isolation using MicroGEM (Figure 5). To extract one sample, 15.21g of plastic was used with QIAGEN and 5.25g with MicroGEM + AMPureXP beads from both tube and tip items (Figure 5). Overall, MicroGEM required 52.64% less plastic than QIAGEN to isolate DNA, per sample. The plastic footprint was determined for both CO_2 emissions and energy consumption to produce the required plastic to process a single sample. The data showed that the plastic footprint was at least 50% lower with MicroGEM (Table 8).

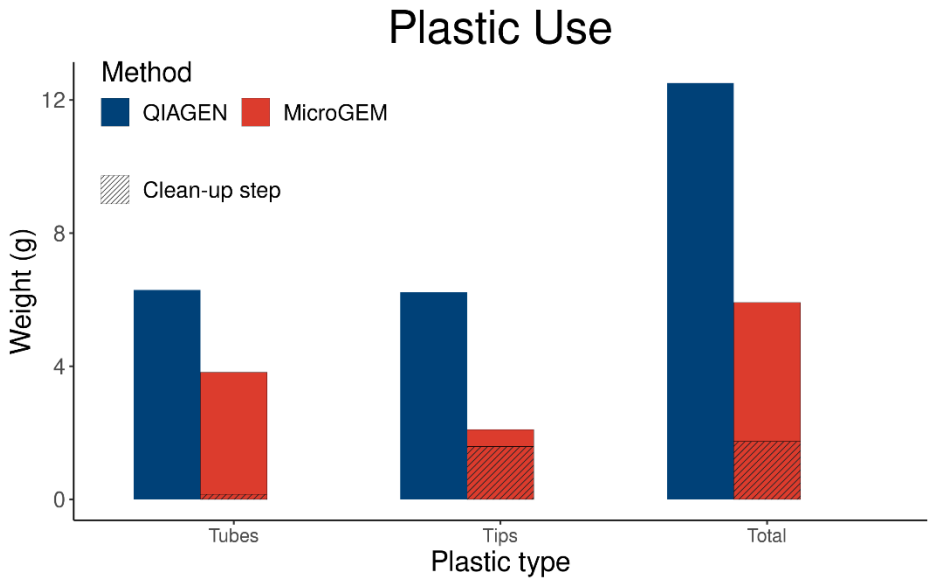


Figure 5. Total plastic required to process one sample by QIAGEN (blue) and MicroGEM (red). The clean-up step plastic use (MicroGEM only) hatched.

Table 8. Plastic footprint of the DNA isolation methods to process one sample.

	Carbon emissions (Kg CO_2)	Energy used (Mj)
QIAGEN	0.04255	1.07498
MicroGEM	0.02011	0.50809

We also measured the time required to isolate DNA with each method by determining the time needed to process 24 samples and calculated for a single sample (see methods for details). Overall, the time required to extract a single sample using MicroGEM was 51.8 % less than QIAGEN (Table 9). If we only include the hands-on time needed to purify the DNA of one sample by using both methods, then MicroGEM needed 34.6 % less time than QIAGEN.

Table 9. Time required to process one sample by using each of the DNA isolation methods.

	Total (min)	Total hands-on (min)
QIAGEN	119.4	5.4
MicroGEM	57.5	3.5

The plastic consumed to isolate 1146 *P. menziesii* and 79 *T. plicata* samples from the population project described above (see Table 7) by using the DNA isolation QIAGEN method was 15.3 kg, which represented 52.06 kg of CO_2 emitted, and 1310 Mj of energy required to produce the plastic used. If MicroGEM was used to isolate the DNA of this project rather than QIAGEN, 8.1 kg less plastic would have been utilised, 27.44 kg of CO_2 emission would have been avoided and 686 Mj of energy would have been saved. Based on an 8-hour working day, 25.9 days were needed to complete the 1225 isolations using QIAGEN while only 14.7 days would have been needed to isolate the same samples if using MicroGEM.

2.5 Discussion

The central question posed in this study was whether a rapid and plastic-efficient DNA isolation method will recover reliable DNA similarly to a commonly used kit. We compared the DNA yield, the potential use of the DNA recovered on downstream analyses, and efficiency on four samples of three different forest tree species and considered the outputs of a separate study in larger populations (Chapters 3 and 4).

We presented data on the use of two methods to isolate nucleic acids from forest trees including method optimisations to the basic protocols to isolate the DNA compared to the original methods. No significant difference in DNA yield recovered was found between the two methods when using the improved protocols, indicating that either method will deliver sufficient DNA yield for standard molecular analyses. However, significant differences were found among species, which suggested the need for specific optimisation of laboratory protocols when looking at various species. The significant difference found in the interaction between methods and species suggests that *P. menziesii* and *T. heterophylla* recovered more DNA when using QIAGEN, but *T. plicata* recovered similar amounts in both cases. Several studies looked at the differences in the amount of recovered DNA quality and time efficiency of different DNA isolation methods in plant species (Bashalkhanov & Rajora, 2008). Bashalkhanov & Rajora., (2008) tested several DNA extraction systems suitable for conifers obtaining a mean DNA yield of more than 10 µg when using the QIAGEN DNeasy kit, which aligns with our DNA yield results. They also concluded that the QIAGEN DNeasy kit is not a preferred method when dealing with large numbers of samples due to the long time that is required to complete the isolation. Our initial MicroGEM DNA concentration results were similar to previous research such as described by Ryan et al., (2020), which showed a very low recovered DNA concentration (less than 3 ng/µl) when using MicroGEM on different plant tissues. After optimisation, we substantially increased the concentration from less than 5 ng/µl to over 15 ng/µl and up to 78 ng/µl using MicroGEM. These results confirmed the utility of optimising protocols to obtain sufficient DNA to perform reliable downstream analyses. It is essential to acknowledge the relatively low number of samples in our dataset could limitate and impact our findings. However, similar sample

sizes were used in comparable studies evaluating a novel DNA isolation method (Moeller et al., 2014). They found significant trends in different species with datasets of four samples per species. Nonetheless, future tests with larger sample sizes and including more species may provide further validation and enhance the generalisation of our observations.

The potential use of the DNA for downstream analyses was analysed with standard DNA barcoding techniques. Abdel-Latif, A., & Osman, G. (2017) compared three DNA isolation methods on maize and assessed the amplification efficiency using PCR techniques. We used gene-specific primers to validate the species identification as done by Armenise et al. (2012). *rbcL* and *trnH-psbA* are commonly used to confirm the species identity and power of discrimination by PCR amplification and sequencing (Hollingsworth et al., 2009; Loera-Sánchez et al., 2020). For *trnH-psbA*, both DNA isolation methods delivered accurate size amplicons and no minor bands which match with Kress & Erickson, (2007) results, where *trnH-psbA* exhibited the highest PCR success. The sequences from *rbcL* and *trnH-psbA* markers were submitted to BOLD and GenBank, respectively, to confirm sample identities and the taxonomical assignment when using *trnH-psbA* was 100% successful, which relates to what Loera-Sánchez et al., (2020) found. Those sequences of unexpected length were the only ones incorrectly identified. In the species discrimination analysis with the *trnH-psbA*, all samples from the same species clustered together, without any differences between the DNA isolation method. Several studies have validated these two loci for efficient DNA barcoding in plants (Kress & Erickson, 2007; Loera-Sánchez et al., 2020), and specifically conifers (Armenise et al., 2012). In their study, Armenise et al., (2012) recovered the same expected fragment size and expected sequence length as our results when analysing *P. menziesii* using *rbcL* (710 bp) and *trnH-psbA* (565 bp) markers, showing consistency in the amplifications. Their findings suggest that combining these two markers may be preferable to perform DNA barcoding in conifers due to their PCR uniformity and sufficient sequence quality while showing enough variation to perform species identity analyses at the genus level. In contrast, our results suggest that using *trnH-psbA* alone retrieves sufficient evidence to identify and differentiate species from different genera which corresponds with Kress & Erickson, (2007) results.

To assess the plastic and time efficiency of each DNA isolation method, we reported the plastic footprint and the time needed to process one sample. The results of this study indicate that a

temperature-driven enzymatic cocktail isolation system reduces plastic footprint by 52.64% compared to a commonly used silica-based nucleic acid isolation method. This enzymatic method also reduces the average time needed to process a sample by 51.8%. When looking at the plastic footprint, we found that there is limited research on comparison of plastic consumption among different isolation nucleic acids methods. Marengo et al., (2019) developed a DNA isolation method for plant species where they reduced the sample processing time while maintaining the quality of the DNA recovered. The authors also discussed the reduction of plastic waste achieved although this was not quantified. The lack of previous research on this matter prevents a direct comparison with our DNA isolation plastic footprint results. Nonetheless, there is an increase of studies looking at how to reduce general plastic use in laboratories (Alves et al., 2021; Howes, 2019) and proposing protocols optimised to decrease the amount of plastic use (Torkamaneh et al., 2020). Alves et al., (2021) developed a 7-week study measuring the plastic used in a molecular laboratory after promoting very specific behavioural and protocol changes to reduce the use of plastic items. They were able to achieve a reduction of more than 10 kg of plastic per week on average, from an initial of 24 kg. The improved protocol for GBS library preparation proposed by Torkamaneh et al., (2020) reduces both the time and plastic needed by 75% and 89%, respectively, compared to standard methods. These findings, together with our results, confirm the potential to apply sustainable measures such as using alternative materials, reusing items when possible or developing protocols which use less plastic without compromising the experiment's outcome.

Our study comparing different methods shows the potential to significantly impact on the plastic consumption and efficiency of DNA isolation in coniferous forest tree species. Our findings suggest that MicroGEM is a highly suitable method as it provides sufficient DNA yield with good downstream analyses performance while producing the least amount of plastic waste and being the most time-efficient. Our study also shows that selecting the most suitable method will depend on the specific requirements of the project, the species studied and the resources available. We believe that our results will encourage researchers to select DNA isolation methods based on sustainable laboratory practices, although further research is needed to explore the performance of this method on a broader range of species and molecular analyses. Additional research to assess other DNA isolation

methods and their potential to reduce plastic use is also needed due to the importance of increasing the studies that evaluate the environmental impact of molecular laboratories.

2.6 Conclusion

In conclusion, the main result of this study is that this rapid and plastic-efficient method, Plant DNA Extraction, MicroGEM, recovers sufficient and reliable DNA to perform common downstream analyses such as PCR and sequencing, and performs as well as a commonly used spin-column kit, DNeasy plant mini kit, QIAGEN. Our study highlights the merit of efforts towards developing more sustainable and efficient laboratory practices in the field of molecular biology. A reduction of plastic waste in molecular laboratories deserves further research for the development of techniques and protocols that use less single-use plastic items.

2.7 Acknowledgements

The authors acknowledge the forestry estates where the samples came from: Stourhead (Western), Longleat (Wiltshire) and Bagley Woods. Additionally, we thank Caterina Branca for technical advice on MicroGem methodology. We also thank Barley Rose Collier Harris for sampling support and Tin Hang (Henry) Hung for DNA Barcoding technical advice. LG received financial support from the Oxford-John Oldacre Graduate Scholarship and from Mr Henry Hoare. The project was funded in part by the OxLEP fund for the Oxford Plant Sciences Centre for Innovation.

2.8 References

Abdel-Latif, A., & Osman, G. (2017). Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. Plant Methods, 13, 1. https://doi.org/10.1186/s13007-016-0152-4

Abril, N., Gion, J. M., Kerner, R., Müller-Starck, G., Cerrillo, R. M. N., Plomion, C., Renaut, J., Valledor, L., & Jorrin-Novo, J. V. (2011). Proteomics research on forest trees, the most recalcitrant and orphan plant species. Phytochemistry, 72(10), 1219–1242. https://doi.org/10.1016/J.PHYTOCHEM.2011.01.005

Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., Lefèvre, F., Lenormand, T., Yeaman, S., Whetten, R., & Savolainen, O. (2013). Potential for evolutionary responses to climate change – evidence from tree populations. Global Change Biology, 19(6), 1645–1661. https://doi.org/10.1111/GCB.12181

Alves, J., Sargison, F. A., Stawarz, H., Fox, W. B., Huete, S. G., Hassan, A., McTeir, B., & Pickering, A. C. (2021). A case report: Insights into reducing plastic waste in a microbiology laboratory. Access Microbiology, 3(3). https://doi.org/10.1099/acmi.0.000173

Armenise, L., Simeone, M. C., Piredda, R., & Schirone, B. (2012). Validation of DNA barcoding as an efficient tool for taxon identification and detection of species diversity in Italian conifers. European Journal of Forest Research, 131(5), 1337–1353. https://doi.org/10.1007/s10342-012-0602-0

Azimi, S. M., Nixon, G., Ahern, J., & Balachandran, W. (2011). A magnetic bead-based DNA extraction and purification microfluidic device. Microfluidics and Nanofluidics, 11(2), 157–165. https://doi.org/10.1007/s10404-011-0782-9

Bashalkhanov, S., & Rajora, O. P. (2008). Protocol: A high-throughput DNA extraction system suitable for conifers. Plant Methods 2008 4:1, 4(1), 1–6. https://doi.org/10.1186/1746-4811-4-20

Bastien, J.-C. (2019). Douglas-fir biomass production and carbon sequestration. 4.2. Douglas-fir – an option for Europe. In European Forest Institute, What Science can tell us (Vol. 9).

https://hal.inrae.fr/hal-02791000

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brown, S. C., Wigley, T. M. L., Otto-Bliesner, B. L., Rahbek, C., & Fordham, D. A. (2020). Persistent Quaternary climate refugia are hospices for biodiversity in the Anthropocene. Nature Climate Change 2020 10:3, 10(3), 244–248. https://doi.org/10.1038/s41558-019-0682-7

Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2016). GenBank. Nucleic Acids Research, 44(D1), D67–D72. https://doi.org/10.1093/nar/gkv1276

Deangelis, M. M., Wang, D. G., & Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Research, 23(22), 4742. https://doi.org/10.1093/NAR/23.22.4742

Di Paolo, L., Abbate, S., Celani, E., Di Battista, D., & Candeloro, G. (2022). Carbon Footprint of Single-Use Plastic Items and Their Substitution. Sustainability 2022, Vol. 14, Page 16563, 14(24), 16563. https://doi.org/10.3390/SU142416563

Doyle, J. J. (1990). Isolation of plant DNA from fresh tissue. Focus, 12(1).

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Edgar, R. C., & Bateman, A. (2010). Search and clustering orders of magnitude faster than BLAST. BIOINFORMATICS APPLICATIONS NOTE, 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Emaus, M. N., Cagliero, C., Gostel, M. R., Johnson, G., & Anderson, J. L. (2022). Simple and efficient isolation of plant genomic DNA using magnetic ionic liquids. Plant Methods, 18(1), 1–14. https://doi.org/10.1186/S13007-022-00860-8/FIGURES/10

Fu, Z. yu, Song, J. cheng, & Jameson, P. E. (2017). A rapid and cost effective protocol for plant genomic DNA isolation using regenerated silica columns in combination with CTAB extraction. Journal of Integrative Agriculture, 16(8), 1682–1688. https://doi.org/10.1016/S2095-3119(16)61534-4

Fumagalli, M. (2013). Assessing the Effect of Sequencing Depth and Sample Size in Population Genetics Inferences. PLOS ONE, 8(11), e79667. https://doi.org/10.1371/JOURNAL.PONE.0079667

Garden, J. G., O'Donnell, T., & Catterall, C. P. (2015). Changing habitat areas and static reserves: Challenges to species protection under climate change. Landscape Ecology, 30(10), 1959–1973. https://doi.org/10.1007/S10980-015-0223-3/FIGURES/6

Green, M. R., & Sambrook, J. (2016). Precipitation of DNA with ethanol. Cold Spring Harbor Protocols, 2016(12), 1116–1120. https://doi.org/10.1101/pdb.prot093377

Green, M. R., & Sambrook, J. (2017). Isolation of High-Molecular-Weight DNA Using Organic Solvents. Cold Spring Harbor Protocols, 2017(4), pdb.prot093450. https://doi.org/10.1101/pdb.prot093450

Harding, K. G., Dennis, J. S., von Blottnitz, H., & Harrison, S. T. L. (2007). Environmental analysis of plastic production processes: Comparing petroleum-based polypropylene and polyethylene with biologically-based poly-β-hydroxybutyric acid using life cycle analysis. Journal of Biotechnology, 130(1), 57–66. https://doi.org/10.1016/J.JBIOTEC.2007.02.012

Harley, C. D. G. (2011). Climate change, keystone predation, and biodiversity loss. Science, 334(6059), 1124–1127. https://doi.org/10.1126/SCIENCE.1210199/SUPPL_FILE/1210199.HARLEY.SOM.PDF

Heuertz, M., Carvalho, S. B., Galindo, J., Rinkevich, B., Robakowski, P., Aavik, T., Altinok, I., Barth, J. M. I., Cotrim, H., Goessen, R., González-Martínez, S. C., Grebenc, T., Hoban, S., Kopatz, A., McMahon, B. J., Porth, I., Raeymaekers, J. A. M., Träger, S., Valdecantos, A., Vella, A., Vernsesi, C., Garnier-Géré, P. (2023). The application gap: Genomics for biodiversity and ecosystem service management. Biological Conservation, 278, 109883. https://doi.org/10.1016/J.BIOCON.2022.109883

Holliday, J. A., Aitken, S. N., Cooke, J. E. K., Fady, B., González-Martínez, S. C., Heuertz, M., Jaramillo-Correa, J. P., Lexer, C., Staton, M., Whetten, R. W., & Plomion, C. (2017). Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. Molecular Ecology, December 2017.

https://doi.org/10.1111/mec.13963

Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M. W., Cowan, R. S., Erickson, D. L., Fazekas, A. J., Graham, S. W., James, K. E., Kim, K. J., John Kress, W., Schneider, H., van AlphenStahl, J., Barrett, S. C. H., van den Berg, C., Bogarin, D., … Little, D. P. (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America, 106(31), 12794–12797. https://doi.org/10.1073/PNAS.0905845106/SUPPL_FILE/0905845106SI.PDF

Hong, E. P., & Park, J. W. (2012). Sample Size and Statistical Power Calculation in Genetic Association Studies. Genomics & Informatics, 10(2), 117. https://doi.org/10.5808/GI.2012.10.2.117

Howes, L. (2019). Can Laboratories Move Away from Single-Use Plastic? ACS Central Science, 5(12), 1904–1906. https://doi.org/10.1021/acscentsci.9b01249

Keselman, H. J., & Rogan, J. C. (1977). The Tukey multiple comparison test: 1953–1976. Psychological Bulletin, 84(5), 1050–1056. https://doi.org/10.1037/0033-2909.84.5.1050

Kress, W. J., & Erickson, D. L. (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. PLOS ONE, 2(6), e508. https://doi.org/10.1371/JOURNAL.PONE.0000508

Levy, S. E., & Boone, B. E. (2019). Next-Generation Sequencing Strategies. Cold Spring Harbor Perspectives in Medicine, 9(7). https://doi.org/10.1101/CSHPERSPECT.A025791

Li, J., Wang, S., Yu, J., Wang, L., & Zhou, S. (2013). A Modified CTAB Protocol for Plant DNA Extraction. Chinese Bulletin of Botany, 48(1), 72. https://doi.org/10.3724/SP.J.1259.2013.00072

Loera-Sánchez, M., Studer, B., & Kölliker, R. (2020). DNA barcode trnH-psbA is a promising candidate for efficient identification of forage legumes and grasses. BMC Research Notes, 13(1), 1–6. https://doi.org/10.1186/S13104-020-4897-5/TABLES/3

Manen, J.-F., Sinitsyna, O., Aeschbach, L., Markov, A. V., & Sinitsyn, A. (2005). A fully automatable enzymatic method for DNA extraction from plant tissues. BMC Plant Biology, 5(1), 23. https://doi.org/10.1186/1471-2229-5-23

Marengo, A., Cagliero, C., Sgorbini, B., Anderson, J. L., Emaus, M. N., Bicchi, C., Bertea, C. M., & Rubiolo, P. (2019). Development of an innovative and sustainable one-step method for rapid plant DNA isolation for targeted PCR using magnetic ionic liquids. Plant Methods, 15(1), 1–11. https://doi.org/10.1186/S13007-019-0408-X/FIGURES/7

MicroGEM International PLC. (2019). Microgem: Nucleic Acid Extraction: Extremely Fast, Extremely Simple, Extremely Effective. Genetic Engineering & Biotechnology News, 39(S2), S15–S15. https://doi.org/10.1089/gen.39.S2.05

Mina, M., Bugmann, H., Cordonnier, T., Irauschek, F., Klopcic, M., Pardos, M., & Cailleret, M. (2017). Future ecosystem services from European mountain forests under climate change. Journal of Applied Ecology, 54, 389–401. https://doi.org/10.1111/1365-2664.12772

Moeller, J. R., Moehn, N. R., Waller, D. M., & Givnish, T. J. (2014). Paramagnetic cellulose DNA isolation improves DNA yield and quality among diverse plant taxa1. Applications in Plant Sciences, 2(10), apps.1400048. https://doi.org/10.3732/apps.1400048

Müller, B. S. F., de Almeida Filho, J. E., Lima, B. M., Garcia, C. C., Missiaggia, A., Aguiar, A. M., Takahashi, E., Kirst, M., Gezan, S. A., Silva-Junior, O. B., Neves, L. G., & Grattapaglia, D. (2019). Independent and Joint-GWAS for growth traits in Eucalyptus by assembling genome-wide data for 3373 individuals across four breeding populations. New Phytologist, 221(2), 818–833. https://doi.org/10.1111/NPH.15449

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pauls, S. U., Nowak, C., Bálint, M., & Pfenninger, M. (2013). The impact of global climate change on genetic diversity within populations and species. Molecular Ecology, 22(4), 925–946. https://doi.org/10.1111/MEC.12152

R Core Team. (2018). R: A language and environment for statistical computing (4.3.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ratnasingham, S., & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes, 7(3), 355–364.

https://doi.org/10.1111/J.1471-8286.2007.01678.X

Rezadoost, M. H., Kordrostami, M., & Kumleh, H. H. (2016). An efficient protocol for isolation of inhibitor-free nucleic acids even from recalcitrant plants. 3 Biotech, 6(1), 1–7. https://doi.org/10.1007/S13205-016-0375-0

Ryan, A. L., O'Hern, C. P., & Elkins, K. M. (2020). Evaluation of Two New Methods for DNA Extraction of "Legal High" Plant Species. Journal of Forensic Sciences, 65(5), 1704–1708. https://doi.org/10.1111/1556-4029.14478

Sawyer, A. (2019). The unsustainable lab. BioTechniques, 66(1), 5–7. https://doi.org/10.2144/btn-2018-0185

Tan, B., & Ingvarsson, P. K. (2022). Integrating genome-wide association mapping of additive and dominance genetic effects to improve genomic prediction accuracy in Eucalyptus. The Plant Genome, 15(2), e20208. https://doi.org/10.1002/TPG2.20208

Torkamaneh, D., Boyle, B., St-Cyr, J., Légaré, G., Pomerleau, S., & Belzile, F. (2020). NanoGBS: A Miniaturized Procedure for GBS Library Preparation. Frontiers in Genetics, 11. https://doi.org/10.3389/FGENE.2020.00067

Urbina, M. A., Watts, A. J. R., & Reardon, E. E. (2015). Labs should cut plastic waste too. Nature, 528(7583), 479–479. https://doi.org/10.1038/528479c

Vadell, E., de-Miguel, S., & Pemán, J. (2016). Large-scale reforestation and afforestation policy in Spain: A historical review of its underlying ecological, socioeconomic and political dynamics. Land Use Policy, 55, 37–48. https://doi.org/10.1016/j.landusepol.2016.03.017

Vega-Vela, N. E., & Chacón Sánchez, M. I. (2011). Isolation of high-quality DNA in 16 aromatic and medicinal Colombian species using silica-based extraction columns. Agronomía Colombiana, 29(3), 349–357.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. The American Journal of Human Genetics, 101(1), 5–22. https://doi.org/10.1016/J.AJHG.2017.06.005

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer New York. https://doi.org/10.1007/978-0-387-98141-3

Youngentob, K. N., Zdenek, C., & van Gorsel, E. (2016). A simple and effective method to collect leaves and seeds from tall trees. Methods in Ecology and Evolution, 7(9), 1119–1123. https://doi.org/10.1111/2041-210X.12554

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Comparing DNA isolation methods for forest trees: quality, plastic footprint, and time-efficiency
Publication Status	☑Published □ Accepted for Publication □Submitted for Publication □Unpublished and unsubmitted work written in a manuscript style
Publication Details	Guillardín, L., MacKay, J.J. Comparing DNA isolation methods for forest trees: quality, plastic footprint, and time-efficiency. Plant Methods 19, 111 (2023). https://doi.org/10.1186/s13007-023-01086-y

Student Confirmation

Student Name:	Laura Guillardin Calvo	
Contribution to the Paper	Designed the study, conceived and planned the experiments, performed sampling, laboratory, bioinformatic and statistical analyses and prepared the figures. Drafted and revised the manuscript.	
Signature 	Date	15/11/2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: John MacKay, Wood Professor of Forest Sciences		
Supervisor comments I agree with the above comments.		
Signature 	Date	05/01/2024

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 3. GBS-Based SNP Genotyping and Genetic Diversity Assessment in *Thuja plicata* Adults and Juveniles in UK Continuous Cover Forestry

Laura Guillardin[1], Ella Glover[1], Gary Kerr[2], John J. MacKay[1]

1. Department of Biology, University of Oxford, Oxford OX1 3RB, United Kingdom

2. Forest Research, Alice Holt Lodge, Farnham, Surrey, GU10 4LH, United Kingdom

3.1 Abstract

Thuja plicata Donn ex D.Don is a conifer tree of great importance for its cultural, ecological and wood quality features in its natural range in western North America and has been widely used in the UK for timber production. They are included in even-aged plantations that are now being converted into more diverse woodlands, which will use natural regeneration and aim to establish a Continuous Cover Forestry (CCF) management. The planted canopy trees may not show a large genetic diversity resulting in a poor gene pool being transmitted to the offspring. Thus, the ability of these forests to face ongoing environmental shifts and disease threats, among other stressors, may be compromised. To investigate this matter, we will estimate the genetic diversity transmission from the canopy trees towards naturally regenerated individuals. Therefore, to estimate the genetic diversity, we genotyped a total of 163 adults and 176 naturally regenerated individuals and assessed the level of transformation into CCF of four UK woodlands. We used the Genotyping-by-Sequencing (GBS) method in 80 individuals selected from the different sites and found 504 SNPs after conservative filtering from which we selected 154 SNPs and validated them using a PCR-based genotyping method. A final number of 67 SNPs were included in our reduced SNP set and was used to genotype the whole sample size. A large number of variant sites were private monomorphic SNPs (40.3%) from which 70% were found in Longleat1 adults, which may indicate a high number of potential fixed alleles for this plantation revealing significant genetic differences among woodlands, also shown in the F_{st}-pairwise results. Overall, high levels of heterozygosity were found across sites ($H_o = 0.38$), with the exception of Longleat1 ($H_o = 0.26$). No major differences were found in the heterozygosity levels comparison between adults and juveniles within woodlands, only significant in Longleat1 and Stourhead1 where adults exhibited higher values than the juveniles. However, N_e was very low for both adults and juveniles across all sites ($N_e < 100$), which may reveal the inexistence of sufficient genetic diversity in *T. plicata* UK plantations to face future disturbances. Our findings and genotyping methods provide insights into the ability of CCF-managed woodlands, including instances of low genetic diversity in *T. plicata*, to face environmental shifts and disease threats.

3.2 Background

Thuja plicata Donn ex D.Don (TP) is a large evergreen coniferous tree native to western North America, which may be found in UK woodlands under Continuous Cover Forest (CCF) management. CCF is an approach to managing forests and is defined as a variety of sustainable practices adapted to local conditions in which the aim is to increase and safeguard biodiversity while preserving timber harvest. It is believed to be a better approach for curbing the impact of climate change than other traditional practices based on clear-felling (Helliwell & Wilson, 2012; Mason et al., 1999). CCF also matches aspirations to extend the range of ecosystem services that are provided to the community (Kirby & Watkins, 2015). The approach is based upon the following four key principles: i) ecosystem management rather than focus on trees management with timber aim; ii) using succession and disturbance as the basis of silvicultural prescriptions for stand management and renewal; iii) working within site limitations, and iv) development of a diverse (irregular) stand structure with a mixture of species (Mason et al., 1999). The forest structure is defined by the spatial distribution, species diversity and variation in tree dimensions (Pommerening, 2002). A higher number of species and diversity in the structure indicates the development from pure even-aged plantation into Continuous cover forest woodlands (Gärtner & Reif, 2004). Pommerening & Murphy (2004) developed the original structure diversity diagram which we modified and presented in Figure 1. This diagram shows four different woodland's structure types, from an even-aged plantation to a late stage of CCF, with multi and uneven-aged woodlands in between. Although not every woodland may go through every stage, this diagram together with the estimation of species diversity indexes allows us to broadly categorise the study sites stage of transformation towards CCF. Changes in thinning practices to open gaps are one of the main drivers to achieve the transformation into CCF allowing the appearance of naturally regenerated trees (Kerr et al., 2017; Macdonald et al., 2010),

resulting in lower tree-specific canopy density while ensuring sufficient canopy coverage.

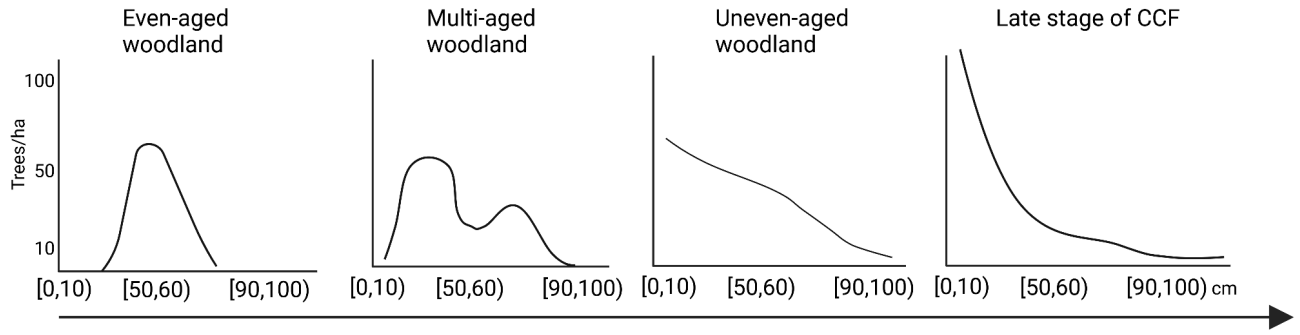


Figure 1. Structure diversity diagram transition from a plantation towards a CCF woodland. Modified from Pommerening & Murphy (2004). The x-axis represent the diametric classes in cm and the y-axis is the number of trees/ha.

This species is of great importance in Northern Western America as it has been used and managed by First People Nations for centuries for cultural and utility items (Hebda & Mathewes, 1984). Due to its valuable characteristics such as wood durability and resistance to rot (Stirling et al., 2017), the species also plays an important role in the forestry sector for building purposes and has been widely planted in Europe for this very reason (Aldhous et al., 1974).

T. plicata is a monoecious species, i.e. individuals have both the male and female reproductive separated organs, and anemochorous, i.e. their pollen and seed are dispersed by the wind. Some *T. plicata* populations present low levels of standing genetic diversity, which could potentially impact the species' long-term adaptability. While it has principally an outcrossing mating system, inbreeding by mating genetically related individuals or self-reproduction may occur under different situations, moreover, *T. plicata* shows relatively high selfing although studies have shown low levels of inbreeding depression, specifically in growth and chemistry traits (Shalev et al., 2023; Wang & Russel, 2006); nevertheless, low genetic diversity levels have been found in its natural range when estimating heterozygosity (H_e and H_o), nucleotide diversity (π) and effective population size (N_e) (Shalev et al., 2022). Variability in genetic diversity levels was reported along a latitudinal gradient within its natural distribution, showing higher levels in the south coastal region (O'Connell et al., 2008). It has been reported that during the last glacial period, *T. plicata* was only found in this southern region which act as a refugium. Afterwards, it spread towards the north and to the east to its

current natural distribution (O'Connell et al., 2008). Additionally, predictions were made of its spread from coastal areas towards the interior due to changing climatic conditions (Hamann & Wang, 2006). However, a big impact on tree species has been predicted by climate change which could largely affect *T. plicata* prevalence and potential for adaptation (Alberto et al., 2013) both in their natural range and in woodlands outside the original distribution.

There are relatively few genomic resources to support investigations of genetic diversity in *T. plicata*, although a transcriptome (Shalev et al., 2018) and a reference genome were recently released (Shalev et al., 2022); however, several methods are available to fill this gap efficiently. Restriction site-associated DNA sequencing (RADseq) and Genotyping-by-Sequencing (GBS) have emerged as powerful and cost-effective tools for high-throughput genotyping for non-model organisms (Holliday et al., 2018). GBS originally was used to describe a specific method using only one restriction enzyme (Elshire et al., 2011), while double digestion RADseq (ddRADseq) describes the use of both, a common and a rare cut-site enzymes to allow a more accurate fragment selection size (Parchman et al., 2018). In this chapter, we used two enzymes to digest the *T. plicata* DNA and we refer to this approach as GBS, which is nowadays used to describe any methods that use two restriction enzyme cut sites for sequencing and genotyping (Andrews et al., 2016). GBS allows the discovery of thousands of single nucleotide polymorphisms (SNP) to apply in population genetics studies although deep data curation and use of complex bioinformatic pipelines are needed (Vaux et al., 2023). Estimating genotyping errors (Mastretta-Yanes et al., 2015) and missing data handling by imputation (Mora-Márquez et al., 2023) are key steps to process GBS data to obtain reliable results. Alternatively, genotyping methods that use a reduced number of markers allow the use of PCR-based genotyping methods that are usually simpler to analyse, have lower levels of missing data increase reliability, and reduce costs compared to sequencing when analysing large sample sizes. Various automated high-throughput platforms, such as Illumina Infinium iSelect HD, (Illumina, San Diego, CA) the Sequenom MassARRAY platform (Agena Bioscience, San Diego, CA), and Standard Biotools Dynamic Arrays (San Francisco, CA, USA), allows the genotyping of large amounts of samples in a cost and time-efficient way (Choi et al., 2020; Davey et al., 2011).

T. plicata gained importance in the UK forestry sector at the end of the eighteenth century (Jarvis, 1973) but studies investigating UK plantations' genetic diversity levels are scarce. Additionally, records of the provenance of the material used to establish these woodlands are usually unknown which may increase the uncertainty of the capacity of adaptation of these populations due to the potential discordance of habitats and environments between source and established plantation.

This work is based on the hypothesis that low standing genetic diversity may be present in *T. plicata* among woodlands being converted to a CCF regime, together with human selection and logging of individual trees may result in naturally-regenerated juveniles with an even narrower genetic base or a low potential to adapt to changing conditions. In order to test this hypothesis, we used planted *T. plicata* trees from four woodlands and conducted an SNP discovery step by using GBS and subsequently selected a subset of informative SNP markers for larger scale genotyping. The overarching goal of this study is to assess the differences in genetic diversity between planted trees and their offspring considering four different *T. plicata* plantations in the UK. This main goal can be divided into four distinct objectives: 1) To characterise the woodlands' level of transformation into CCF; 2) To perform SNP discovery in *T. plicata*; 3) To design and validate a reduced SNP genotyping set; 4) To assess the genetic diversity differences between adults and juveniles on four UK-planted woodlands managed under the CCF approach.

3.3 Material and Methods

3.3.1 Study sites characterisation: structure and species diversity

We sampled four CCF sites located in England (UK) that included *T. plicata* as part of the canopy and are identified as Bagley Wood, Longleat1, Stourhead1, and Stourhead2. Each site had at least 100 adult trees as recommended by Wojacki et al., (2019) to minimise possible low genetic diversity due to being a small population, as well as having enough viable natural regeneration.
We quantified the total surface area of each site using QGIS software (QGIS, 2023) and defined the stand distribution as patchy if the site includes separated areas, or as continuous. We assessed the level of isolation of each site by determining the existence of *T. plicata* woodlands within a 2 Km radius

from the middle point of each site, based on the National Forest Inventory Woodland public dataset (Forestry Commission, 2019) and private land cover shapefiles (supplied by landowners).

We performed a forest survey between September 2020 and June 2021 to estimate the level of transformation from a plantation towards an old-growth CCF woodland (Figure 2) following the recommended methodology described in the Forest Inventory and monitoring guidelines (Northwest Natural Resource Group, 2014). We used QGIS (QGIS, 2023) to randomly select the centre of the plots on each site and exported it into a GPX file format to upload to a GPS device. The number of plots varied in each site due to its size and the density of adult trees, ranging from 10 in Bagley Wood to 25 in Stourhead2. The total number of adult trees measured per site varied, ranging from 64 in Stourhead1 to 104 in Bagley Wood. Overall, the average number of adult trees within one plot ranges from 2.68 in Stourhead2 to 10.4 in Bagley Wood. These data can be found in the Supplementary online repository within the Inventory directory. (https://github.com/LauGuillardin/Thesis_Supplementary_Information/tree/main/Inventory). To exemplify the study design, we created a diagram that showed a simulated site with 16 plots (Figure 3).

The size of the tree plots ranged from 6m to 8m radius, depending on the site, while every natural regeneration plot was 3m across (Figure 2). We measured diameter at breast height (dbh) using a dbh tape, and height, using the Bitterlich Relascope method (Bitterlich, 1984) on every tree inside the tree plot over 10cm dbh. We measured the dimensions of regenerated seedlings and saplings within the regeneration plot using a digital calliper and a metre stick and took a census of each species (Figure 2). Additionally, we identified the species of shrubs, its average height and evaluated the proportion of each species ground coverage within the regeneration plot. We estimated the canopy density of each plot by using a spherical crown densiometer (Figure 2) and determined the *T. plicata* average height and tree density.

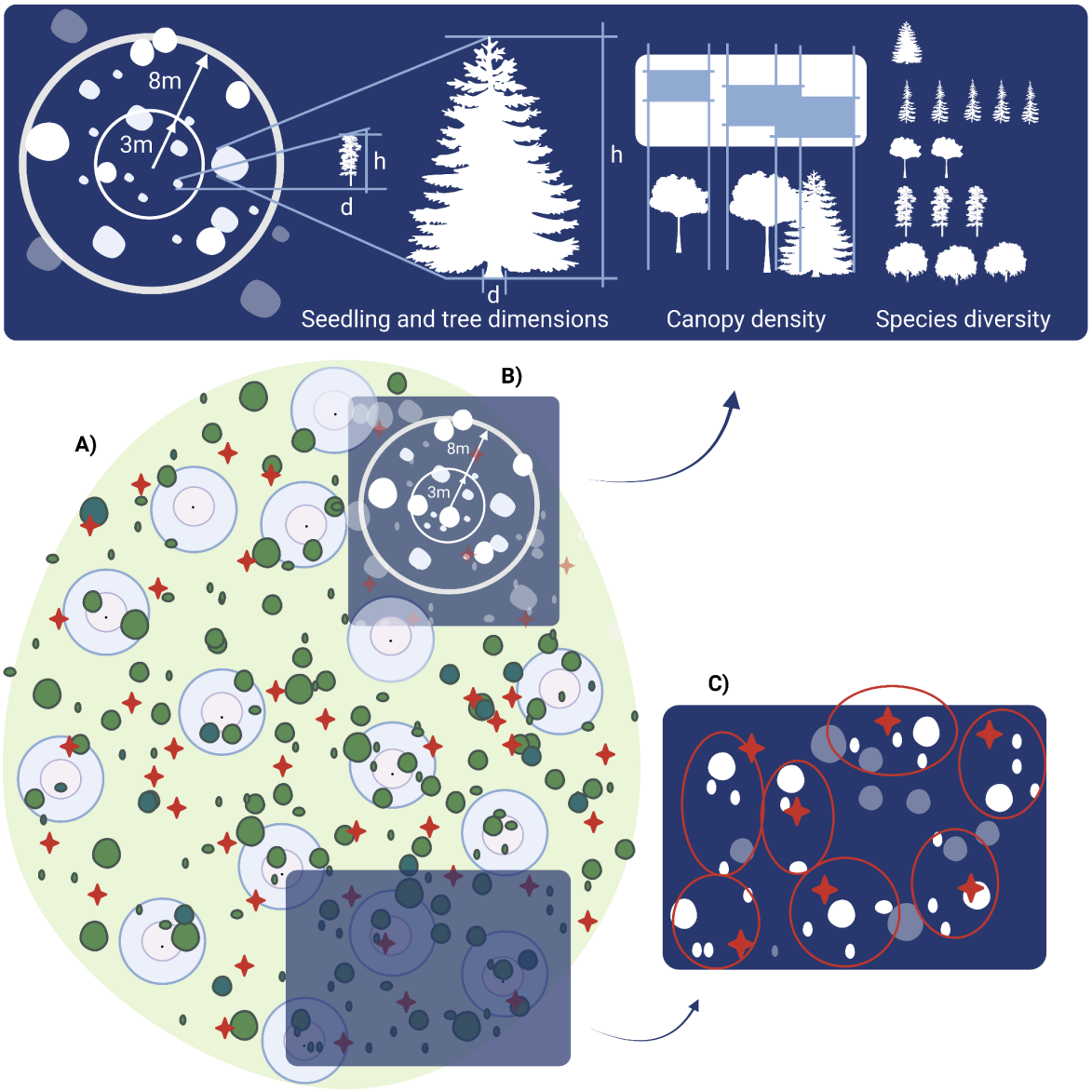


Figure 2. Diagram of field work carried out for both, the forest survey and the sampling of tissue for DNA extraction. A) Represents a simplified site where trees are the green shapes of different sizes, the larger are adult trees and the smaller are seedlings or saplings. The circular concentric circles represent the tree and natural regeneration plots for the forest survey, and the red stars are the GPS random points for the tissue sampling. B) Forest inventory plots description and parameters measured. Seedling and tree dimensions: d = diameter at breast height (cm), h= height (m), We included the context of the location of the plots within a site. C) Example of the selection of seven adult trees (white) and the three seedlings or saplings (white) assigned per adult for tissue sampling.

To measure the canopy, offspring snd shrub species diversity, we estimated the following indexes: 1) the Species richness (S) or actual number of different species, 2) The Shannon index, *H'*, (Shannon, 1948), which considers both species richness and evenness, 3) the Species evenness (E) (Moore, 2013), which describes the commonness or rarity of a species within a stand, and 4) the Effective number of species (Moore, 2013), considered the number of equally-common species required to give a particular value of a Shannon index (H').

The Shannon entropy index was calculated by counting the number of different species and their frequency of appearances (Figure 2), and then applying the following formula (1) where S = species richness and pi = relative abundance of species i.

$$H' = -\sum_{i=1}^{S} p_i \cdot log\, p_i \qquad (1)$$

The Species evenness, which ranges from 0 to 1, and the Effective number of species were calculated using the formulas (2) and (3), respectively:

$$E = \frac{H'}{ln(S)} \qquad (2)$$

$$Effective\ number\ of\ species = EXP(H') \qquad (3)$$

The raw inventory data for each study site, the estimated canopy density and the data needed to calculate the species diversity indexes can be found in the Supplementary online repository (https://github.com/LauGuillardin/Thesis_Supplementary_Information/tree/main/Inventory).

3.3.2 Sampling and DNA isolation

We sampled leaf material from 34 to 48 adults and 41 to 46 juveniles across sites (Table 8) for genetic analysis (Figure 3). These sampling design is independent to the forest survey described in section *3.3.1 Study sites characterisation: structure and species diversity,* therefore, the trees measured for dbh and height are not the ones we selected to sample leaves from. To spread the sampled trees fairly throughout the site area, we calculated random points and exported them into a GPX file format to upload to a GPS device, resulting in a uniform distribution across the site. Figure

3, panel C describes the protocol followed to select the adult trees by choosing the closest tree to the GPS point (red star) and the three seedlings or saplings that were closer to the selected adult tree. The material recovered was stored in labelled paper envelopes which were kept in plastic zip bags together with silica gel beads to dry the tissue. A subsample of 10 adults and 10 juveniles per site, 80 samples in total, was used to perform the GBS step.

The DNA of the 80 used for SNP discovery was isolated using the spin-column DNeasy plant mini kit (Qiagen USA, Valencia, CA) (Qiagen), as in Chapter 2 (see section *DNA isolation procedures and improvements*). For the other 259 samples, we used the temperature-driven enzymatic cocktail Plant DNA Extraction (MicroGEM International PLC, 2019), as described in Chapter 2. This protocol simplifies the DNA isolation and provides sufficient DNA yield and quality to perform the genotyping while reducing by half the time and plastic used compared to using the Qiagen method (Chapter 2). After recovering the DNA we performed a clean-up step by using the AMPpureXP beads, which use magnetic particles that bind the DNA. A concentration of 0.8X of AMPureXP beads (Beckman Coulter™, Brea, CA, USA) was used to retrieve the cleaned isolated DNA (recommended by NewEngland Biolabs).

The DNA concentration was determined by using Qubit 4 Fluorometer (Invitrogen, Thermo Fisher Scientific, Cleveland, OH, USA). We used NanoDrop One Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) to estimate the purity based on absorbance ratios A260/A280 and A260/A230. A minimum DNA concentration of 20 ng/ul and absorbance ratios of 1.8 were required on each sample.

3.3.3 Overview of genetic analyses

Two complementary genotyping approaches (Figure 3) were used to estimate the genetic diversity of the study populations and the differences between adults and offspring: 1) Sequence-based genotyping using GBS data and 2) PCR-based assay for SNP genotyping using Standard Biotools microfluidics SNP genotyping platform (Standard Biotools, San Francisco, CA, USA). The sequence-based method used a SNP discovery step using Genotyping by sequencing techniques. The variants were called and then used to design the SNP set to perform the PCR-based

SNP genotyping. We compared the results from both approaches to validate the use of a small SNP set which provides a high-throughput, time-efficient, highly reliable and shareable protocol when the objective is to assess the genetic diversity of large sample size experiments.

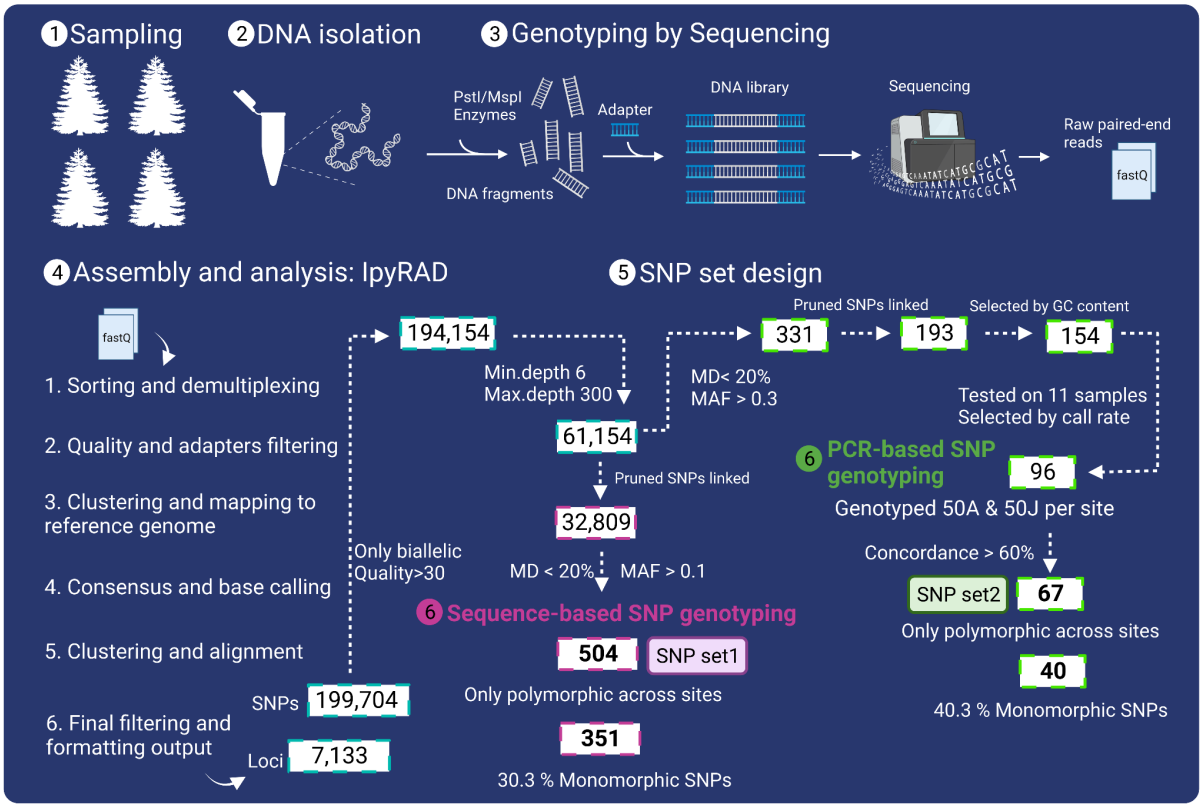


Figure 3. Genetic analysis workflow. A detailed description of each step performed from the sampling (1) and DNA isolation (2) , the GBS (3), the assembly and variant calling using IpyRad (4) to the design of both SNP sets (5) to use in the genotyping step. MD = missing data, MAF = minor allele frequency.

3.3.4 GBS library construction, sequencing and quality check

Genotyping by Sequencing (GBS) is a technique where the DNA is fragmented by the use of one or more digestion enzymes and then sequenced usually with the use of short-read sequencing techniques aiming to uncover polymorphic sites across the genome. To select the combination of enzymes producing a suitable cut rate to be used in the library preparation we analysed the *T. plicata* genome sequence (Shalev et al., 2022) with the software ddRADseqtools (Mora-Márquez et al., 2017) which simulates the digestion of a reference genome with pre-selected enzymes. We compared the

number and length of fragments recovered from each pair of available enzymes tested and selected the optimal combination.

The DNA was normalised to 20 ng/ul and used to prepare the library (Genomic Analysis Platform of Laval University, Canada) and further sequencing (Genome Quebec, Canada). The library was constructed with PstI/MspI enzyme combination (NEB, Ipswich, MA, USA) and the Illumina TruSeq HT Library Preparation Kit (Illumina, Inc. San Diego, CA, USA) (Figure 3), retaining on the fragments above 300bp from a fragment selection step. Sequencing was performed on an Illumina NovaSeq 6000 PE150 platform (Illumina Inc. San Diego, CA, USA), and 150 bp paired-end raw files were obtained. The fastq files for each pair-end were checked for quality using the FastQC software (Andrews, 2010); the initial demultiplexing step used SABRE (Barry et al., 1989) based on the barcode per sample supplied by the Illumina sequencer provider, not allowing any mismatch or Ns, and the Illumina adapters were trimmed using Trimmomatic (Bolger et al., 2014).

3.3.5 Clustering, assembly and variant calling using Ipyrad

We used the Ipyrad pipeline (Eaton & Overcast, 2020) to cluster, assemble and call the bases from the trimmed fastq files. Ipyrad is divided into six steps (Figure 3), allowing the repetition of steps separately once the results have been reviewed. The parameters set are shown in Table 1. After the first Ipyrad run including all the samples, we performed exploratory depth, missing data and number of loci per sample tests. We excluded the samples that showed an average depth below 6 reads or more than 50% of missing data. We reran Ipyrad from the 3rd step without the excluded samples. All the runs were performed in the University of Oxford Advanced Research Computing (Oxford, UK) facility and used a Capability cluster node with 48 CPUs and 384GB RAM.

Table 1. Parameters name and value used in the Ipyrad pipeline

Parameter name	value
Assembly method	reference
Reference sequence	pairddrad
Restriction overhang	TGCAG,CG
Max low quality bases	5
Phred Qscore offset	43
Min depth statistical	6
Min depth majrule	6
Max depth	10000
Cluster threshold	0.90
Max barcode mismatch	0
Filter adapters	2
Filter min trim len	35
Max alleles consens	2
Max Ns consens	0.05
Max Hs consens	0.05
Min samples locus	4
Max SNPs locus	0.2
Max Indels locus	8
Max shared Hs locus	0.5
Trim reads	0, 0, 0, 0
Trim loci	0, 0, 0, 0

3.3.6 Data post-filtering and SNP sets design

To assess the potential for designing a practical and reliable subset of informative SNP markers set for future research and conservation efforts in *T. plicata*, we followed the two-step filtering process described in Figure 3.

First, we performed a pre-filtering step from the initial GBS dataset in which we removed multi-allelic SNPs (--max-alleles 2) using PLINK (Purcell et al., 2007) and indels (--remove-indels) from the original dataset followed by a read-depth pre-filtering allowing a read-depth value per loci between 6 and 300 (--min-meanDP --max-meanDP) and a minimum quality of 30 across the SNPs (--min quality 30) using VCFtools software (Danecek et al., 2011). Second, we performed a post-filtering procedure to identify a set of SNPs distributed across the genome, which were polymorphic and had low missing data levels. As shown in Figure 3, we followed stronger filtering criteria to design the reduced PCR-based SNP genotyping set compared to the sequence-based SNP genotyping approach to reach the desired number of 96 SNPs to test in the genotyping platform.

Sequence-based genotyping (SNP set1)

The GBS SNP genotyping used standard filtering criteria to reduce the missing data while ensuring that the markers are distributed across the genome and have sufficient polymorphism within and among sites.

For the sequenced-based genotyping SNP set (SNP set1), we started by pruning SNPs under strong linkage disequilibrium. This was performed using PLINK (Purcell et al., 2007), calculating the pairwise correlation in a 50-SNP window and discarding SNPs with a correlation higher than 0.1, in increments of 10 SNPs at a time (-indep pairwise 50 0.1 10) (Figure 3).

Next, we uploaded the resulting VCF file into RStudio (R Core Team, 2018; RStudio Team, 2020) using the read.vcfR function from the vcfR package (Knaus & Grünwald, 2017) to convert it into a genlight object for processing in the dartR (Mijangos et al., 2022) package environment using the vcfR2genlight tool. DartR facilitates the import and analysis of SNP data and includes the use of 3rd party population genetics software. Using the gl.filter.callrate function, we kept only the individuals with less than 50% of missing data and the SNPs that showed less than 20% of missing data. To ensure that the genetic variants are sufficiently common and informative, we used a cut-off of minor allele frequency (MAF) <0.1, using gl.filter.maf function across the whole dataset. We quantified the monomorphic SNPs on each site using gl.filter.monomorphs on each subsample.

PCR-based genotyping (SNP set 2)

For the PCR-based SNP genotyping platform (SNP set2) we aimed to create a small robust SNP set and thus followed a similar approach but some stronger filters. After the read-depth pre-filtering, we used VCFtools (Danecek et al., 2011) to filter out those SNPs with missing data above 20 % (--max-missing) and a MAF < 0.3 (--maf). Then, we pruned strongly linked SNPs using PLINK (Purcell et al., 2007) with the same criteria as for the SNP set 1. We selected 100 bases on each side of the variant to retrieve the SNP sequences by using *T. plicata* reference genome (v3) (Shalev et al., 2022) by processing the data in the Python package PySAM (Gilman et al., 2019). These sequences were sent to Standard Biotools (San Francisco, CA, USA) to design the assays and perform a quality assay check. The assays were filtered by GC content, discarding those with

GC>65%. The resulting SNP-type assays were ordered and tested on 11 samples. This test was performed using FlexSix IFC plates on the Standard Biotools microfluidics SNP genotyping platform using the Juno system to load and combine the microfluidics, the Biomark (all from Standard Biotools, San Francisco, CA, USA) to perform the PCR, and the FluidiGM SNP Genotyping software to digitalised and analyse the image outputs (Standard Biotools, San Francisco, CA, USA). We used Fast Probe Master Mix (Biotium Inc. Fremont, CA) and followed the recommended preparation protocol (Standard Biotools, 2015b). We selected the best 96 assays based on the level of polymorphisms and call rate to genotype the total sample of 339 individuals. We used eight 192.24 IFC plates following the manufacturer protocol (Standard Biotools, 2015a). The results were visualised using the FluidiGM SNP genotyping software and exported to an SNP table file.

The last step to design the SNP set2 was to remove those SNPs with <60% consistency between the two approaches. Finally, we quantified the number of exclusive monomorphic SNPs among sites.

3.3.7 SNP set validation

The use of fewer SNPs is driven by sample size and aims to strike a balance between genotyping costs and statistical power. The primary goal of comparing the results from the two SNP sets was to assess whether the smaller SNP set (SNP set2) provides comparable information and yields meaningful insights that are consistent with the larger SNP set (SNP set1). The SNP set2 markers were validated to ensure their accuracy and reliability for subsequent genetic diversity analyses. We performed the same analyses on both SNP sets (1 and 2) to evaluate the SNP set2 performance.

From both sets, we excluded the private-site monomorphic SNPs to avoid bias in genetic structure due to site-exclusive monomorphic SNPs (Nazareno et al., 2017; Schmidt et al., 2021). First, we assessed the ability of both SNP sets to identify population structure using the Principal Component Analysis (PCA) and Discriminant Analysis of Principal Components methods (DAPC) (Jombart et al., 2010). To compare the same number of samples from each dataset, we selected randomly 20 adult individuals per site from the SNP set2 data using the gl.subsample.ind function

from dartT (Mijangos et al., 2022). We used the dudi.pca function to implement the PCA analysis from ade4 package (v1.7-22; Dray & Dufour, 2007) and the dapc function from adegenet package (v2.1.10; Jombart, 2008). The data were procesed with ggplot2 (Wickham, 2009) to visualise the results.

To compare the level of genetic differentiation among sites, we used Arlequin (Excoffier & Lischer, 2010) to estimate the sites' F_{st} pairwise genetic distances, and we performed hierarchical Analyses of Molecular Variance (AMOVA) (Meirmans, 2012) to evaluate the amount of population genetic variation within and among sites.

3.3.8 Population genetics analysis

Genetic diversity and genetic variation

We estimated the genetic diversity within each site and compared adults and juveniles with data from SNP set2. We compared results when including both polymorphic and monomorphic SNPs and only polymorphic SNPs. We estimated observed heterozygosity (H_e) and expected heterozygosity (H_e) using the gl.report.heterozygosity tool in DartR package (Mijangos et al., 2022) and we adjusted the values for the unequal number of species per site. Significant differences among expected heterozygosity between adults and juveniles within sites were tested using the tool gl.test.heterozygosity. We used the tool divBasic from the package diveRsity (Keenan et al., 2013) to calculate the inbreeding coefficient (F_{is}) executing 1000 bootstraps to estimate the 95% confidence intervals. We compared the contemporary effective population size (N_e) of adults and juveniles using only monomorphic SNPs (40 SNPs in Longleat1 and 67 SNPs in the other sites) using the linkage disequilibrium method implemented in NeEstimator V2.1 software (Do et al., 2014) and used the Jackknife method to estimate the confidence intervals (Jones et al., 2016).

We evaluated the amount of genetic variation among sites and within sites defining two categories, adults and juveniles. We excluded the private monomorphic SNPs of the analysis to avoid bias in the genetic structure. We performed a hierarchical Analysis of Molecular Variance (AMOVA) using Arlequin (Excoffier & Lischer, 2010) with P-value = 0.05.

Kinship analysis within sites

To assign offspring to putative parents we included both monomorphic and polymorphic SNPs and we evaluated the parent-offspring rate per site using the software CERVUS (Kalinowski et al., 2007). We also removed SNPs not in HWE in each site to focus on relationships that are less influenced by biological processes like inbreeding or selection, which can lead to deviations from HWE; therefore, the final number of SNPs used per site is shown in Table 9. We used the GPS waypoints of each individual to determine the average, minimum and maximum parent-offspring assigned distance.

3.4 Results

3.4.1 Study sites characterisation: structure and species diversity

Four study sites were characterised in terms of area, isolation status, stand distribution, canopy cover, tree density, *T. plicata* specific tree density, and average height (Table 2). Stourhead2 is the largest site and the other three sites share similar sizes. The total tree density ranged from 206.9 to 373 trees/ha; Stourhead2 had the largest difference between the overall tree density (226.4 trees/ha) and the *T. plicata* density (21.2 trees/ha) (Table 2).

Only one site (Longleat1) was isolated from other *T. plicata* woodlands by at least 2 Km. Both Bagley Wood and Stourhead1 showed a patchy distribution as trees were from smaller compartments within the sites that were planted in the same year (information provided by owner). On average, the tallest trees were found in Stourhead2, which also had the lowest canopy cover.

Table 2. Site characterisation results, including overall site estimations and specific estimations for *Thuja plicata* individuals.

		Overall estimations					*Thuja plicata*		
Site	Year planted	Site area (ha)	Isolated (>2Km)	Stand distribution	Canopy cover (%)	Tree density (trees/ha)	Tree density (trees/ha)	Avg. Height (m)	Max. Height (m)
Bagley Wood	1950	3.73	No	Patchy	87.2	373.0	278.5	22.3	28
Longleat1	1958	5.07	Yes	Continuous	88.6	206.9	156.5	24.7	28
Stourhead1	1961	5.00	No	Patchy	80.8	283.9	237.3	27.7	36
Stourhead2	1951	18.57	No	Continuous	73.5	226.4	21.2	29.3	34

A list of the different adult, offspring and shrub species found in each site is shown in Table 3. Every site had more tree species in its offspring than in the canopy strata, which may indicate a colonisation from neighboring sites or that the adult trees of these species were outside every measured plot. Only Stourhead2 had broadleaves in their canopy (*Quercus petraea* and *Fraxinus excelsior*). Due to the variety of species' abundance, the effective number of species is smaller than the actual number of species (S) in every strata, for instance, the site with the highest S (Stourhead2 canopy, S = 8) differs from the effective number of species estimated (4.5) by 44% (Table 3 and Table

4). One site (Stourhead2) had a high canopy species diversity (*H'* = 1.5), while the other three sites were all below 0.6 (Table 4). The offspring species diversity ranged from 0.9 in Bagley Wood to 1.8 in Stourhead2.

Considering the distribution of tree diameter classes, only Stouhead2 exhibited a reverse-j profile, which is found in woodlands in uneven-aged woodlands or woodlands in transition towards CCF, while only one peak was found in the other sites (Figure 1). The most developed site in terms of structural diversity was Stourhead2 (Figure 4), additionally, it was the site that showed highest levels of species diversity (Table 4). The species evenness (E) showed high values in Bagley Wood shrubs and Longleat canopy and shrubs (S = 0.9, S = 0.8 and S= 1, respectively). Stourhead2 canopy and offspring had S = 0.72, indicating a close to even number of individuals per species; however, every other site showed S < 0.7. A minimum of S was found in the canopy of Stourhead1 (S = 0.5) (Table 4).

Table 3. List of species found in each stand separated by strata. The code represents the scientific name of the species found as follows: Ps: *Picea sitchensis*, Th: *Tsuga heterophylla*, Tp: *Thuja plicata*, Ia: *Ilex aquifolium*, Fs: *Fagus sylvatica*, Paq: *Pteridium aquilinum*, Asp: *Agrostis* spp., Rf: *Rubus fructicosa*, Pa: *Picea abies*, Tb: *Taxus baccata*, Qr: *Quercus robur*, Ag: *Abies grandis*, Pm: *Pseudotsuga menziesii*, Ld: *Larix decidua,* Cs: *Castanea sativa*, Cv: *Calluna vulgaris*, Qp: *Quercus petraea*, Fe: *Fraxinus excelsior*, Ap: *Acer pseudoplatanus*, Ca: *Corylus avellana*, Bp: *Betula pubescens*, Ud: *Urtica dioica*, Jsp: *Juncus* spp., Csp: *Carex* spp., Ea: *Equisetum arvense*.

	Bagley Wood			Longleat1		
	Canopy	Offspring	Shrubs	Canopy	Offspring	Shrubs
	Ps Th Tp	Ia Tp Th Fs Ps	Paq Asp Rf	Tp Pa	Ia Tb Qr Ag Tp Pa Pm	Paq Rf
Species richness (S)	3	5	3	2	7	2
	Stourhead1			Stourhead2		
	Canopy	Offspring	Shrubs	Canopy	Offspring	Shrubs
	Tp Ld Pm	Ia Th Tp Fs Pm Ps Cs Qr	Rf Paq Cv	Th Ps Pm Ld Tp Qp Pa Fe	Fs Ia Th Pm Ps Tp Ap Ag Ld Qr Ca Bp	Paq Rf Ud Jsp Csp Ea
Species richness (S)	3	8	3	8	12	6

Table 4. Species diversity indexes estimated for each strata and study site.

	Bagley Wood			Longleat1		
	Canopy	Offspring	Shrubs	Canopy	Offspring	Shrubs
Shannon entropy index (H')	0.619	0.934	1.040	0.555	1.355	0.693
Species evenness (E)	0.563	0.580	0.946	0.801	0.696	1.000
Effective number of species	1.857	2.544	2.828	1.742	3.876	2.000

	Stourhead1			Stourhead2		
	Canopy	Offspring	Shrubs	Canopy	Offspring	Shrubs
Shannon entropy index (H')	0.546	1.331	0.790	1.512	1.807	1.032
Species evenness (E)	0.497	0.640	0.719	0.727	0.727	0.576
Effective number of species	1.727	3.783	2.203	4.534	6.091	2.807

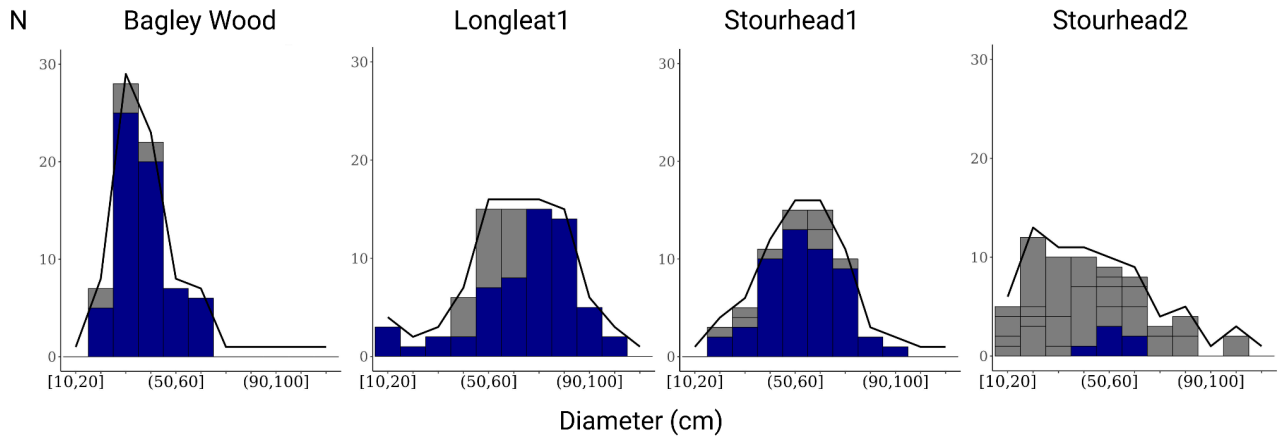


Figure 4. Structural diversity per site shown as the number of trees (from the canopy strata) per diametrical class in cm. *T. plicata* individuals coloured in blue, other species are shown in grey.

3.4.2 DNA isolation, sequencing, quality check and SNP calling

Simulations were performed on ddRADseqTools (Mora-Márquez et al., 2017) and the reference genome (Shalev et al., 2022), using different enzymes in order to predict which ones would yield a suitable number of DNA fragments to ensure high depth across samples (Table 5). We selected the restriction enzymes PstI/MspI to construct the library (see methods for library construction and sequencing details). The enzymes ApeKI, BfaI and NsiI/MspI were excluded due to the large number of potential fragments estimated (> 300,000 fragments), while SbfI/MspI was estimated at <10,000 fragments of 125-300bp size, which can compromise a full coverage of the *T. plicata* genome (Table 5). Total DNA was isolated from 80 individuals, including two duplicates for the GBS analysis by use of the DNeasy plant mini kit, the average DNA concentration was 100.2 ng/ul and the quality ratios A260/A230 and A260/A280 were 1.82 and 1.83, respectively.

Table 5. Number of fragments recovered by simulating the genome digestion using the potential enzymes.

Enzyme	Fragment size 126-300 bp
ApeKI	565,079
BfaI	7,962,110
PstI_MspI	112,191
NsiI_MspI	319,908
SbfI_MspI	8,236

The FastQC analysis of the resulting raw reads gave high-quality scores with an average Phred quality score above 30. We trimmed the 3' end of the sequences which showed low-quality results. Overrepresented sequences corresponding to adapter sequences were identified and removed in the data preprocessing step.

After the first run of the assembler and base calling software Ipyrad (Eaton & Overcast, 2020), we performed an exploratory analysis and detected that three samples showed a depth coverage under 3.5 and two other samples had >70% of missing data. We discarded the data from those five samples and reran the pipeline using the data of the other 75 samples. The total number of raw reads recovered was 1,273,904,871 and the number of raw reads recovered per individual are shown in Table 6, followed by the Ipyrad filtering process results. The final number of variants called were 199,704 SNPs in 7,133 loci. The raw NGS read data was deposited to NCBI's Short Read Archive (SRA). The BioProject ID is PRJNA1081148 and the SRA accession numbers range from SAMN40149401 to SAMN40149480.

Table 6. Final summary of statistics of the Ipyrad assembly run. Reads raw: number of raw reads demultiplexed for each sample, Reads passed filter: number of reads passed filter based on quality scores, Refseq mapped reads : number of reads mapped to the reference genome, Refseq unmapped reads : number of reads that did not mapped to the reference genome, Clusters total: Total clusters found, Clusters hidepth: number of clusters that pass the mindepth thresholds, Hetero est: heterozygosity estimation, Error est: Sequencing error rate, Reads consens: number of final reads after error filtering and Loci in assembly: number of loci recovered per sample.

SampleID	Reads raw	Reads passed filter	Refseq mapped reads	Refseq unmapped reads	Clusters total	Clusters hidepth	Hetero est	Error est	Reads consens	Loci in assembly
B1PA16	18456767	18159596	17515113	644483	104294	55355	0.007844	0.000187	24855	21154
B1PA17	19537438	19223586	18637723	585863	74214	41425	0.00232	0.00019	21986	19379
B1PA26	28719642	28244069	26843172	1400897	156332	80851	0.002458	0.000154	28959	23162
B1PA3	25120931	24719238	23853196	866042	115337	58625	0.002632	0.000155	25615	21778
B1PA33	4271424	4192103	3906045	286058	61659	29361	0.00222	0.000489	15902	14543
B1PA35	6716150	6598027	6332132	265895	58435	28958	0.002604	0.000423	17651	15856
B1PA37	17692561	17370021	16777371	592650	86969	46385	0.002109	0.000252	21590	19018
B1PA45	24266675	23840745	23196252	644493	80547	45430	0.002645	0.000168	24998	21195
B1PA52	23574560	23159694	21962088	1197606	174182	85330	0.002387	0.000204	27087	22236
B1PR106	17990112	17695293	17054392	640901	92342	48204	0.002945	0.000166	22155	19065
B1PR109	19286705	19009397	18457941	551456	69569	41041	0.002565	0.000189	23261	20228
B1PR135	7647035	7491046	7270352	220694	37373	22359	0.002909	0.000489	17403	15562
B1PR142	5877271	5768749	5609080	159669	34349	21602	0.002651	0.000452	18196	16137
B1PR35	31891988	31345061	30209781	1135280	146590	73539	0.002682	0.00017	28682	22940
B1PR52	26661277	26222864	25315304	907560	98487	52388	0.002458	0.000154	26621	21773
B1PR59	17072105	16828577	16411793	416784	65088	34944	0.007935	0.000235	24319	20150
B1PR78	18568664	18303819	17077403	1226416	162838	77873	0.001835	0.000205	21128	17909
B1PR92	6575160	6460195	6232016	228179	37887	22656	0.002692	0.000438	17297	15542
B1PR96	14025729	13803804	13449098	354706	44235	29046	0.002229	0.000249	21222	18665
L1PA1	2412254	2373280	2056446	316834	67128	28803	0.001986	0.000467	14211	13052
L1PA1b	23906361	23535822	20495243	3040579	265382	128505	0.001765	0.000187	25119	19700
L1PA21	30201716	29637923	27767279	1870644	158627	78742	0.007844	0.000187	26399	21439
L1PA22	39012765	38410828	30384598	8026230	501110	250448	0.001662	0.000233	31668	20970
L1PA39	17677918	17369674	15448261	1921413	237883	113294	0.006437	0.000475	23385	18931
L1PA41	19134922	18802077	16217219	2584858	204633	94774	0.001909	0.000407	23064	19137
L1PA42	28246955	27788324	20862661	6925663	393548	192749	0.001983	0.000303	29769	21323
L1PA7	12670635	12438997	11490295	948702	121892	59835	0.006312	0.000625	21586	18531
L1PA8	6209772	6078319	4503369	1574950	153593	62983	0.002035	0.000475	16536	14597
L1PR109	17335906	17050214	15568784	1481430	122685	61607	0.006796	0.000239	22071	18885
L1PR121	2656813	2604449	2507573	96876	24775	16587	0.002279	0.000534	14088	12868
L1PR19	20536022	20182869	17490884	2691985	254434	117766	0.002166	0.000298	26231	20887
L1PR35	30849686	30251526	29205931	1045595	90113	50158	0.002429	0.000181	26741	21787
L1PR42	10345995	10168553	9674771	493782	61262	33094	0.002268	0.000563	18788	16758
L1PR49	4662460	4561428	4047368	514060	66770	31223	0.002195	0.000626	16240	14301
L1PR56	41737756	40933375	25899321	15034054	703355	343379	0.002073	0.000286	37026	20280
L1PR87	14384768	14155605	13504166	651439	69035	37976	0.002046	0.000357	20043	17659
L1PR88	5803441	5713306	5134337	578969	89960	41148	0.002706	0.000597	17857	16106
S1PA2	14730633	14480400	12768272	1712128	123015	63660	0.002724	0.000477	23585	19845
S1PA24	15875952	15620984	14920390	700594	81087	44884	0.001958	0.000237	21540	18532
S1PA26	15749993	15490544	15117856	372688	49424	31152	0.002387	0.000307	23171	19724
S1PA31	15230305	14969584	14599620	369964	49725	30977	0.003032	0.000377	22745	19627
S1PA36	21879133	21538217	21034287	503930	66716	39344	0.002725	0.00031	26445	21798
S1PA37	18909594	18615027	18207444	407583	51199	31747	0.007844	0.000187	24174	20175

SampleID	Reads raw	Reads passed filter	Refseq mapped reads	Refseq unmapped reads	Clusters total	Clusters hidepth	Hetero est	Error est	Reads consens	Loci in assembly
S1PA39	9267306	9101483	8807791	293692	47059	26326	0.002502	0.000475	18886	16705
S1PA44	2326566	2269205	2171755	97450	25530	17368	0.002129	0.000501	15333	13714
S1PA46	2268813	2229595	2133191	96404	25106	16849	0.002478	0.000517	14121	12997
S1PA50	14990288	14757917	14026688	731229	79870	44086	0.00202	0.000392	19750	17218
S1PR101	19852831	19556414	18580276	976138	120189	60282	0.00331	0.00027	25108	21128
S1PR11	6636533	6525054	6352926	172128	38017	25533	0.0032	0.000771	19800	17585
S1PR135	5032566	4949922	3978431	971491	106795	46611	0.002405	0.000397	16174	14529
S1PR167	16106380	15819117	15473566	345551	46363	30349	0.003114	0.00028	23515	20108
S1PR172	23853209	23452636	22913318	539318	55948	36680	0.002544	0.000197	28269	23245
S1PR32	29744043	29290876	26259610	3031266	211572	103585	0.002422	0.000216	27346	21997
S1PR56	33989129	33437259	31945966	1491293	119509	65475	0.002682	0.00017	29109	23559
S1PR58	17511511	17192210	16641665	550545	63253	37268	0.009565	0.000256	23574	19951
S1PR82	6770413	6662271	6461296	200975	41827	26698	0.002584	0.000523	19201	17172
S4PA10	11461965	11287605	11001190	286415	49520	29274	0.00247	0.000529	18808	16509
S4PA13	12825847	12608168	12014037	594131	82998	43300	0.002376	0.00039	20037	17617
S4PA15	13147178	12967756	12486692	481064	82658	40308	0.002236	0.000343	19650	17192
S4PA19	24376469	23966060	23223028	743032	74549	41222	0.002199	0.000199	22305	19308
S4PA20	18205817	17919781	17124914	794867	106849	56087	0.002294	0.0003	21138	18210
S4PA21	9418500	9290060	8994494	295566	55394	28405	0.00187	0.000406	17511	15689
S4PA25	14498193	14287536	13795052	492484	63040	35967	0.006906	0.000362	21844	18958
S4PA26	16420977	16177287	15614581	562706	67147	37456	0.001406	0.000283	21738	18709
S4PA3	8182460	8036503	7609384	427119	61057	34424	0.002889	0.000943	18606	16362
S4PA35	7502708	7359091	6796538	562553	66266	32027	0.006312	0.000625	17433	15293
S4PR35	2620308	2556327	2456159	100168	27314	18067	0.002655	0.000572	15621	13694
S4PR42	6163660	6037215	5875635	161580	32959	21088	0.002805	0.000596	17983	15864
S4PR47	9331681	9167501	8272593	894908	89289	43577	0.002919	0.000757	18760	16294
S4PR77	20247411	19894843	19348648	546195	69211	38244	0.002078	0.00033	21702	18612
S4PR81	17159707	16837509	16321420	516089	75501	39781	0.001983	0.000303	22697	19299
S4PR89	23326493	22974434	22061128	913306	111321	55644	0.00105	0.000171	21896	18030
S4PR90	24711091	24296552	23811475	485077	52411	32906	0.007844	0.000187	24094	19943
S4PR96	15996841	15721257	15322535	398722	57627	32872	0.002522	0.000339	21798	18367
S4PR96b	19355832	19048090	18602150	445940	63978	36894	0.002591	0.000246	23532	19281

3.4.3 Data post-filtering and SNP set design

We recovered a total of 61,154 SNPs within 4,134 loci, which showed a read-depth average across loci of 100.5 (range of 6-300), after filtering to remove indels, multi-allelic SNPs, low low-quality SNPs (see methods for details).

The sequenced-based SNP set was reduced to a set of 504 SNPs (SNP set1) by filtering for missing data and distribution of the SNPs across the genome and level of polymorphism. Starting from the initial set of 61,154 SNPs, strongly linked SNPs were pruned, then samples with >50% of missing data and SNPs with >20% of missing data were removed, resulting in a dataset of 66 samples and 9,343 SNPs in 1,983 loci. Finally, we kept only the SNPs with MAF >0.1 across the four sites, reducing the total to 504 SNPs found in 443 loci (Table 7). The average missing data and MAF across

the SNP set1 is shown in Table 5. We detected 153 private-site monomorphic SNPs (30.3%) (Table 7) and explored their distribution by site (Figure 5) showing an unbalanced distribution, with above 76% of SNPs found only in Longleat1.

We developed a PCR-based genotyping assay and obtained data from 67 SNPs (SNP set2) across the full sample set, with pre-filtering and post-processing selection steps. From the 504 SNPs (SNP set1), we kept the SNPs with <20% missing data and those that had MAF>0.3 and pruned the SNPs strongly linked, resulting in 193 SNPs in 187 loci. We used 100 base pairs on each side of the variant, obtained from the reference genome (Shalev et al., 2022), in order to design the assay primers on the D3™ assay design website (Standard Biotools, San Francisco, CA, USA). This excluded 12 SNPs due to the high levels of GC content found (>65%) and 27 SNPs due to high levels of repetitive regions. The remaining 154 pairs of primers were tested using the Standard Biotools microfluidics SNP genotyping platform (Standard Biotools, San Francisco, CA, USA). We selected the best 96 assays based on call rate and polymorphism criteria to genotype all samples, including the samples we used for SNP discovery. The 96 primer pairs were used to genotype our 360 samples; four assays did not recover any data and were discarded. We performed a consistency test between the duplicated samples genotyped by the SNPs in common between SNP set1 and SNP set2 and discarded 25 SNPs that showed a concordance <60%, resulting in a final number of 67 SNPs (SNP set2). The SNP set2 mean missing data and MAF are reported in Table 3, and we found 27 SNPs (40.3%) that were private-site monomorphic (Table 7), 25 of them exclusively found in adults.

Table 7. Description of the two SNP sets used. N: Number.

	N SNPs	N private-site Monomorphic SNPs %	N samples genotyped	Mean missing data %	Mean minor allele frequency
SNP set1	504	30.4	75	8.53	0.280
SNP set2	67	40.3	360	4.23	0.397

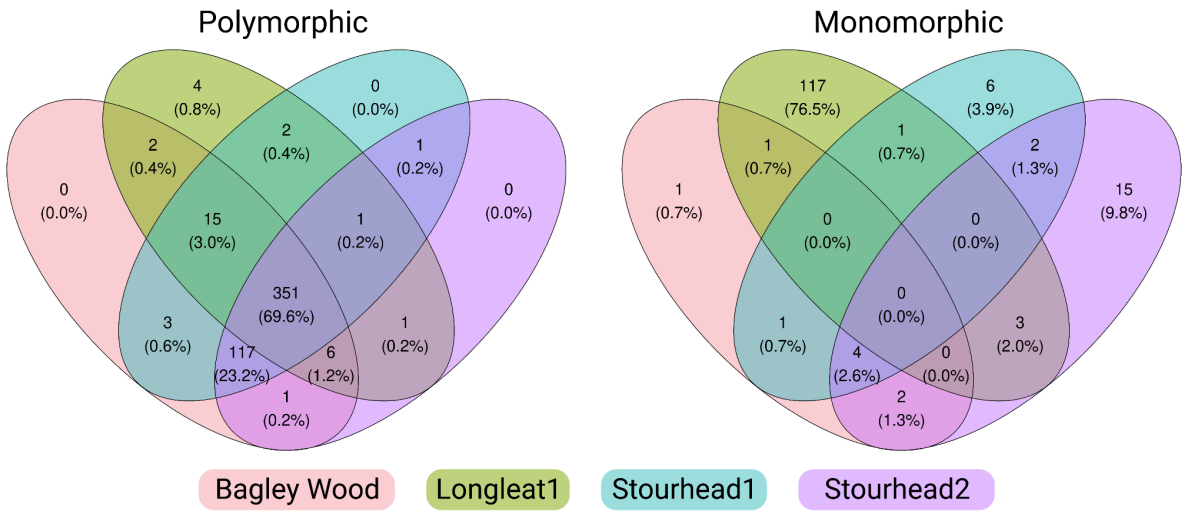


Figure 5. Number (and proportion) of polymorphic and monomorphic SNPs found in SNP set1 across sites.

3.4.4 SNP set validation

We assessed the ability of both SNP sets 1 and 2 to capture the genetic structure among sites using PCA and DAPC methods. We found a similar pattern of population structure (Figure 6). The PCA axis 1 (10.3% and 11.7%) of both tests showed a genetic differentiation of Longleat1 and Stourhead2 against the other two sites (Figure 6 panel A and B). A similar pattern was found in the DAPC analysis where axis 1 of both SNP sets (85.5% and 82.1%) explains Longleat1 and Stourhead2 differentiation while the second axis (13.2% and 13.3%) seems to shown a partial differentiation of Stourhead2 from the two other remaining sites.

The analyses of molecular variance (AMOVA) for the four sites using both SNP sets are reported in Table 8. This analysis assesses the partitioning of genetic variation among and within the study sites. Significant p-values (<0.05) indicate the presence of genetic differentiation. For SNP set1, the analysis revealed that 6.98% of the total genetic variation is attributed to differences among the four sites while SNP set2 showed 11.34%, both with a significant p-value (<0.001). The remaining genetic variation is found within the sites (93.02% and 88.66%) also with significant p-value (<0.001) (Table 8).

The pairwise F_{st} comparisons between sites (Table 9), show that Longleat1 exhibited the highest genetic differentiation with a pairwise F_{st} >0.1 when compared to every other site when using both SNP sets. The lowest pairwise F_{st} was found between Bagley Wood and Stourhead1, F_{st} = 0.01211 when using SNP set1 and F_{st} = 0.00234 when using SNP set2. These F_{st} values were further supported by significant p-values. We found that every site using either SNP set is genetically different from each other except for Bagley Wood and Stourhead1.

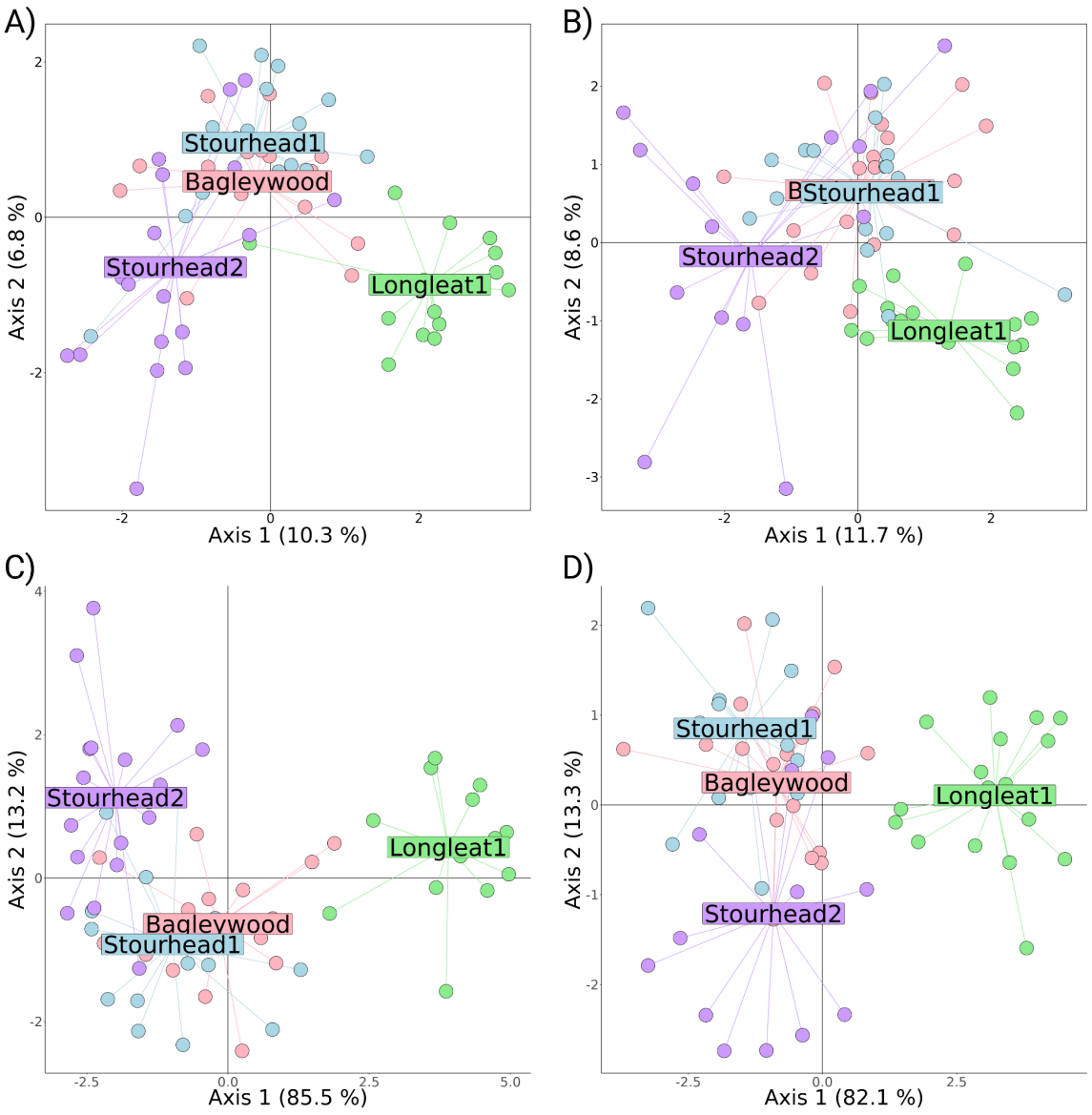


Figure 6. Analysis of principal components (PCA) results using A) SNP set1 and B) SNP set2 and discriminant analysis of principal component (DAPC) results using C) SNP set1 and D) SNP set2 across the four study sites.

Table 8. Comparison of results from SNP set 1 vs SNP set2 components of molecular variation (AMOVA) for the four study sites. Significant P-value <0.05. d.f = degrees of freedom.

	Source of variation	d.f	Sum of Squares	Percentage of variation	Pvalue
SNP set1	Among sites	3	274.886	6.98	<0.01
	within sites	716	4535.281	93.02	<0.01
	Total	719	4810.167		
SNP set2	Among sites	3	366.065	11.34	<0.01
	within sites	128	3000.049	88.66	<0.01
	Total	131	3366.114		

Table 9. Results of F_{ST} pairwise site comparisons for a) SNP set1 and b) SNPset 2. Above the diagonal, the population pairwise F_{ST}s, below the diagonal the calculated F_{ST}s P-values. Significant P-value <0.05. B1: Bagley Wood, L1: Longleat1, S1: Stourhead1 and S2: Stourhead4.

a)

	B1	L1	S1	S2
B1	-	0.1908	0.0121	0.0250
L1	<0.001	-	0.2137	0.2288
S1	0.14414	<0.001	-	0.0311
S2	0.00901	<0.001	<0.001	-

b)

	B1	L1	S1	S2
B1	-	0.1057	0.0023	0.0186
L1	<0.001	-	0.1220	0.1403
S1	0.0991	<0.001	-	0.0180
S2	<0.001	<0.001	<0.001	-

The SNP set 1 and SNP set2 lists can be found in the Supplementary online repository (https://github.com/LauGuillardin/Thesis_Supplementary_Information/tree/main/Genomic_data/TP).

3.4.5 Population genetic analyses

Genetic diversity and genetic variation

The genetic diversity indicators (Table 10) compare the results from using both the monomorphic and polymorphic SNPs to only polymorphic SNPs. Overall, we found high heterozygosity levels across and within populations. For the analysis of both marker groups, observed heterozygosity (H_o) ranged from 0.255 to 0.410 across sites, with the highest observed in Stourhead1 juveniles and the lowest in Longleat1 juveniles. With the polymorphic SNPs alone, H_o generally increased across sites, with the highest H_o in Longleat1 adults (0.437) and the lowest in Stourhead2 juveniles (0.338). The largest change when excluding the monomorphic SNPs happened in Longleat1 adults and juveniles (H_o from 0.261 to 0.437 and 0.255 to 0.386, respectively). We found significant differences between adults and juveniles expected heterozygosity (H_e) in Longleat1 (0.259 and 0.291) and Stourhead2 (0.439 and 0.456) when both monomorphic and polymorphic SNPs were used. When we excluded the monomorphic markers, the only significant differences were for H_e in Stourhead2 adults and juveniles (0.441 and 0.462).

The inbreeding coefficient (F_{is}) varied less when comparing the results from both types of markers, which ranged from -0.007 to 0.262 against only polymorphic SNPs, which ranged from –0.007 to 0.268 (Table 10). Longleat1 had the lowest inbreeding values and Stourhead2 the highest, although no significant differences were found between adults and juveniles within sites. The levels of N_e ranged from very low ($N_e < 10$) in Stourhead2 (adults & juveniles) and Longleat1 (adults) to low ($10< N_e < 100$) in all the other sites (Table 11). Juveniles in Stourhead1 and Longleat1 exhibited substantially higher N_e compared to adults.

Table 10. Genetic diversity results. Comparisons were tested between adults (A) and juveniles (J) within each site and is noted as "*" when the differences were significant (p-value <0.05). N_s = number of samples collected, C.I = confidence intervals, S.D = standard deviation, H_o = observed heterozygosity, H_e = expected heterozygosity, F_{is} = Inbreeding index.

	Bagley Wood		Longleat1		Stourhead1		Stourhead2	
	A	J	A	J	A	J	A	J
N_s	48	44	44	45	37	46	34	41
Including monomorphic								
H_o	0.398	0.393	0.261	0.255	0.367	0.410	0.381	0.337
H_o S.D	0.084	0.086	0.249	0.192	0.096	0.099	0.097	0.088
H_e	0.460	0.459	0.259	0.291*	0.448	0.453	0.439	0.456*
H_e S.D	0.041	0.041	0.243	0.210	0.056	0.058	0.070	0.060
F_{is}	0.134	0.144	-0.007	0.125	0.179	0.094	0.132	0.262
lower C.I	0.086	0.078	-0.087	0.049	0.095	0.048	0.073	0.162
upper C.I	0.182	0.217	0.076	0.194	0.263	0.142	0.191	0.350
Only polymorphic								
H_o	0.407	0.408	0.437	0.386	0.372	0.419	0.377	0.338
H_o S.D	0.081	0.089	0.162	0.132	0.106	0.092	0.095	0.076
H_e	0.463	0.461	0.434	0.440	0.451	0.455	0.441	0.462*
H_e S.D	0.041	0.039	0.149	0.132	0.057	0.055	0.070	0.049
F_{is}	0.122	0.115	-0.007	0.124	0.174	0.081	0.144	0.268
lower C.I	0.071	0.044	-0.088	0.049	0.083	0.026	0.060	0.178
upper C.I	0.171	0.188	0.074	0.201	0.260	0.134	0.227	0.360

Table 11. Effective population size (N_e) estimated in adults and juveniles across sites. min and max correspond to the 95% confidence intervals calculated using the Jackknife method (Jones et al., 2016).

	Adults			Juveniles		
	N_e	min	max	N_e	min	max
Bagley Wood	55.1	41	79.4	48.7	36.6	68.6
Longleat1	8.2	5.3	11.6	20.8	14.5	30.8
Stourhead1	20.3	16.4	25.6	97.8	61.8	200.9
Stourhead2	9.7	7.9	11.9	7.4	5.9	9.0

The analysis of molecular variation (AMOVA) results (Table 12) demonstrated significant genetic differentiation among the four sites, accounting for 6.93% of the total genetic variance. A small proportion of genetic variation is attributed to differences among adults and juveniles within sites (0.10%), although statistically insignificant. The majority of the genetic variation (92.98%) was found within groups (adults and juveniles).

Table 12. Analysis of molecular variance (AMOVA) results table among the four studies and between adults and juveniles within each site. d.f = degree of freedom. Significant P- value < 0.05.

Source of variation	d.f	Sum of Squares	Variance components	Percentage of variation (%)	P-value
Among sites	3	274.886	0.4717	6.93	0.0156
Among groups within sites	4	27.714	0.0067	0.10	0.4839
Within groups	712	4507.567	6.3309	92.98	<0.01
Total	719	4810.167	6.8092		

Kinship analysis

The number of SNPs used to perform this analysis after HWE filtering in each site is shown in Table 13, only four markers were in departure of HW in Stourhead2. The ability to assign regenerated individuals to their parents ranged from 11% in Longleat1 to 37% in Bagley Wood. The average distance between assigned parents and offspring across sites is 192.28 m. Stourhead1 is a patchy stand (Table 1) and showed the lowest average distance (65.8 m) and the lowest maximum distance (310.4 m) across sites while showing the second highest number of parents-offspring assigned (35%) (Table 13). Stourhead2 tree density was only 21.2 trees/ha, and its surface area was the largest (18.57 ha) (Table 1); it also showed 22% of assigned parents and the minimum assignment distance was the highest across sites (15.9 m). Bagley Wood is a patchy stand (Table 1) and showed the largest average and maximum distance of assigned parents-offspring while the highest number of parents-offspring assigned (Table 13). Longleat1 had the lowest number of parents-offspring assigned (11%) and the parent assigned to the highest number of offspring (Table 13).

Table 13. Kinship analysis results, including the number of SNPs used per site (N SNPs), the percentage of parent-offspring related found (Average %), the mean, maximum and minimum distance these relationships were found and the maximum number of offspring related to one parent found (Max offspring from one parent). N: number, Max: maximum, Min: minimum.

	N SNPs	Assigned (%)	Average distance (m)	Max distance (m)	Min distance (m)	Max offspring from one parent
Bagley Wood	67	37	282.8	1456.0	3.0	2
Longleat1	67	11	156.4	403.7	3.0	4
Stourhead1	67	35	65.8	310.4	3.0	3
Stourhead2	63	22	264.1	734.4	15.9	2

3.5 Discussion

In this study we assessed the genetic diversity differences between *T. plicata* planted trees and their offspring on four UK woodlands managed under the CCF approach. We identified 504 informative SNP markers by using a GBS approach and designed a reduced SNP set (67 SNPs) for genotyping on a PCR-based platform. Our study has revealed significant genetic variation differences and substantial variability in the number of private monomorphic SNPs among the study sites. Overall, we found high heterozygosity levels (H_e and H_o) across and within sites and only one site (Stourhead2) showed H_e significantly higher in the juvenile trees. However, we found very low N_e across and within sites, ranging in the adults from 8.2 in Longleat1 to 55 in Bagley Wood and in the juveniles from 7.4 in Stourhead2 to 97.8 in Stourhead1, which may indicate low levels of genetic diversity, and may lead to a lack of potential adaptation in these *T. plicata* stands.

3.5.1 Sites biodiversity and level of transformation

The four studied planted woodlands composed of *T. plicata* showed variable stand characteristics such as tree density, canopy coverage, structure and species diversity. An increase in species composition is expected in transformed stands (Bravo-Oviedo et al., 2014; Gärtner & Reif, 2004; Kerr et al., 2010, 2017). We estimated the canopy species diversity using the Shannon index parameter and recovered values ranging from 0.54 in Bagley Wood to 1.5 in Stourhead2, showing higher values compared to other UK stands in transition into CCF studied by Pommerening (2002), which ranged from 0.0997 to 0.6165.

The diameter class distribution changes from even-aged plantation to late stage of CCF as described in Pommerening & Murphy, (2004) and Vitkova & Dhubháin, (2013). Bagley Wood, Longleat1 and Stourhead1 showed a diametric class structure usually found in pure plantations similar to what was found in pure stands composed of > 90% of *Pinus sylvestris* and *Picea abies* by Zaytsev et al., (2019). Alternatively, Stourhead2 had a reverse-j structure typically found in unmanaged or near-natural woodlands (De Quesada & Kuuluvainen, 2020; Kerr, 2002) and late-stage CCF forests (Figure 1) (Kerr et al., 2010, 2017). A low overall tree density was found on the study sites ranging

from 206.9 trees/ha in Longleat1 to 373.0 trees/ha in Bagley Wood compared to the five sites in transformation examined by Kerr et al., (2017) which ranged from 557 to 716 trees/ha. These sites are composed of different tree species including *T. plicata* and have been managed using the Bradford-Hutt system, which is used to transform even-aged plantations into CCF focusing on managing small areas actively using plating, thinning and felling practices.

Overall, Stourhead2 exhibited the highest biodiversity values and the most diverse structure showing a heterogeneous mix of species and sizes. Bagley Wood, Longleat1 and Stourhead1 appeared to share a similar level of transformation into CCF and similar canopy species diversity values although Bagley Wood had the lowest variability of seedling diversity, which may compromise the future canopy diversity on this stand. Further monitoring may contribute to understanding the effects of low levels of species and structure diversity in transforming pure even-aged plantations into CCF woodlands to evaluate the future successful transition of these sites.

3.5.2 Low number of variants and monomorphic SNPs

Genotyping by sequencing allows for the discovery of thousands of variant sites in non-model species in a cost and time-efficient way but requires a multi-step bioinformatic process including data cleaning due to potential genotyping errors (Mastretta-Yanes et al., 2015) and high levels of missing data (López de Heredia et al., 2020; Vaux et al., 2023). Depending on the project aims, different filtering may be applied so standard data handling criteria are lacking, although a typical workflow includes missing data, minor allele frequency, minimum and maximum depth, hardy-weinberg and linkage disequilibrium filtering (Vaux et al., 2023). A way to manage missing data is to impute the missing alleles predicting the lacking nucleotide bases using statistical methods (López de Heredia et al., 2020; Mora-Márquez et al., 2023). Our study aim is to retain high-quality SNPs to design a reduced marker set thus a strong filtering was applied (see methods section). We found 199,704 variants before filtering, which is relatively low compared to the 1.3M SNPs found in Chinese fir (*Cunninghamia lanceolata*) also a member of the Cupressaceae family (Zheng et al., 2019) or the 349,542 SNPs found in the foxtail pine (*Pinus balfouriana*) (Friedline et al., 2015). After filtering by depth, MAF, missing data and pruning the loci linked, we retained 504 high-quality SNPs which

exhibited a high proportion of private monomorphic SNPs (30.3%) (Table 7), the majority being found in Longleat1 (Figure 5). The MAF>0.2 filter usually removes rare, private and monomorphic SNPs across the population studied. In our case, 70% (Figure 5) of the exclusive-site monomorphic SNPs are found in Longleat1, indicating that this site has a high level of fixed alleles compared to the others. The majority were found in the adults revealing low genetic variance and potential adapted alleles suggesting subsequent differences in the seed source provenance used when establishing the plantations.

3.5.3 SNP genotyping assay development

The validation of the SNPs discovered using GBS and filtered to perform population genetic analysis followed a similar approach to several genomic studies (Choi et al., 2020; Kishor et al., 2020; Nguyen et al., 2020). The 67 resulting SNPs (SNP set2) were validated by testing and comparing the genetic variation and genetic differentiation among sites found using the larger SNP set (SNP set1). The DAPC, PCA and the AMOVA results using both SNP sets, with only the polymorphic SNPs, showed the same genetic variation, and genetic clustering (Table 8 and Figure 6) supported by the F_{st}-pairwise results (Table 8). F_{st} pairwise comparison across sites indicated that only Stourhead1 and Bagley Wood do not show significant genetic differences while Longleat1 had the highest F_{st}-pairwise values ($F_{st} > 0.19$) across all the other sites. This approach is similar to Kishor et al., (2020), who developed an SNP set to investigate melon cultivars and validated their approach by comparing the genetic differentiation of the cultivars captured to a larger data set of 5,644 SNPs to a smaller SNP set (184 SNPs). The robust SNP markers reported in our study may facilitate further genetic studies on *T. plicata* using a repeatable, cost and time-efficient approach to investigate woodlands' genetic diversity and genetic variation, moreover when large sample sizes are required and GBS or other sequencing techniques may result in expensive costs and challenges of managing large amounts of missing data.

3.5.4 Population genetic analysis using SNP set2

The genetic variation analysis among sites revealed significant differences although no consistent genetic variation was found between adults and juveniles within each site and the maximum variation is found among individuals within sites.

In *T. plicata*, the mating system combines outcrossing with high levels of self-fertilization (O'Connell et al., 2001) while avoiding high levels of inbreeding depression (Wang & Russel, 2006). *T. plicata* from the natural range exhibited a H_o = 0.219 estimated from 112 trees by Shalev et al., (2022), including high levels of S.D = 0.110. Very subtle signs of genetic structure were found across the natural range by dividing the 112 trees geographically into northern, central and southern subpopulations and testing using STRUCTURE (Pritchard et al., 2000); however, no structure was observed when using the discriminant analysis of principal components (DAPC) (Jombart et al., 2010). This, together with non-significant results of F_{st}-pairwise across subpopulations may indicate large levels of gene flow occurring across the natural range despite the low levels of heterozygosity found.

In our study, the heterozygosity levels (H_o and H_e) were estimated including and excluding private monomorphic SNPs following the recommendations of Schmidt et al. (2021). The heterozygosity (both H_o and H_e) appeared lower across the sites when monomorphic SNPs were included, most remarkably in Longleat1 (Table 10). This same pattern was found in Schmidt et al., (2021) when they compared the autosomal heterozygosity (including monomorphic SNPs) to the SNP heterozygosity (only polymorphic SNPs) in two species of insects using ddRADseq and DarTseq data. They indicated that reporting both is advised as different biases can affect the genetic diversity estimations when using only polymorphic SNPs, such as the amount of missing data and a small number of samples.

Overall, genetic diversity levels across sites were relatively high in juveniles, ranging from H_o = 0.337 in Stourhead2 to H_o = 0.410 in Stourhead1, which contrasts the low levels of H_o = 0.219 found in the species natural range (Shalev et al., 2022). Only Longleat1 showed similar levels with H_o = 0.259 and H_o = 0.291 in adults and juveniles, respectively. The high S.D found in Longleat1

indicated that the He and Ho results are largely spread out across loci and may suggest variability in coverage (Høy Hansen et al., 2022). Low levels of He in Longleat1 was explained by the large number of monomorphic SNPs found, which could suggest the potential fixation of alleles due to adaptation to specific conditions or high levels of inbreeding in the source populations. This may explain the difference in the genetic diversity levels we found in Bagley Wood, Stourhead1 and Stourhead2 compared to Longleat1, although this would contradict the low levels of inbreeding detected (Table 10). Generally, the inbreeding levels remained unchanged after excluding monomorphic SNPs from the analysis and no differences between adults and juveniles were found. However, in three of the sites, the F_{is} in juveniles was higher than in adults, which may indicate a trend that could be of concern for the future. F_{is} ranged from -0.007 in Longleat1 adults to 0.262 in Stourhead2 juveniles, indicating lower inbreeding in these planted woodlands than in the natural range (F_{is} = 0.331), estimated by (Shalev et al., 2022).

The H_e was significantly higher in juveniles from natural regeneration than in planted adults in two of the sites, i.e. Stourhead2 and Longleat1; while the other sites had similar H_e in juveniles and adults. This may indicate that the frequency of the heterozygous on each site has not been reduced and in some cases has increased, suggesting a potential efficient transfer of alleles to the naturally regenerated trees.

The N_e is a genetic diversity estimator that is gaining importance due to its recent inclusion as a genetic diversity indicator in the the Convention on Biological Diversity (CBD) Global Biodiversity Framework (Frankham, 2022; Heuertz et al., 2023; Hoban et al., 2020, 2023). None of the sites had an $N_e > 500$, which is the current target established to ensure sufficient maintenance of genetic variation in the long-term (Hoban et al., 2020). Only Stourhead1 juveniles (N_e = 97.8) is close to N_e = 100 which is considered the minimum N_e to avoid short-term inbreeding depression (Frankham et al., 2014). We estimated the N_e using only polymorphic SNPs, resulting in less than the 180 minimum SNPs recommended, thus, these results must be interpreted with caution; however, we ensured no linkage among the study SNPs, which is one of the main assumptions of the method (Luikart et al., 2010; R. K. Waples et al., 2016; R. S. Waples & Do, 2010). Although the N_e estimated in the species'

natural range by Shalev et al. (2022) is higher than our results (N_e = 270.3), it also revealed low levels of genetic diversity across the natural distribution which follows the low He levels found.

On average across all the study woodlands, the proportion of assigned parents-offspring relationships found across sites was 26.25% at 95% confidence revealing a high proportion of parents being either from individuals located within the sites but not sampled or from outside the sites. Slightly lower results were found in Duminil et al. (2016), who investigated the parental assignment on a high selfing tree species (*Baillonella toxisperma*) and found only 10% of the offspring could be assigned to a sampled parent. The lowest proportion of assigned parent-offspring was found in Longleat1 (11%), which is the only site that is isolated by > 2 Km from other *T. plicata* stands (Table 2). It also had the highest number of offspring related to one single parent, indicating the potential for a relatively high level of half-sibship from few parents. Bagley Wood gave the highest proportion of assignments (37%) and a low number of offspring assigned to one single parent, suggesting a lower level of half-sibships. Bagley Wood is the smallest site in terms of area but had the highest density of *T. plicata* trees (Table 2), which could explain the high number of assigned trees. Stourhead1 had 35% of assigned parents, indicating that the patchy distribution of these sites (Table 2) does not prevent gene flow among the patches. This is also demonstrated by looking at the distance between individuals from separated patches being assigned, which ranged from 310.4 m in Stourhead1 to 1,456.0 m in Bagley Wood (Table 13). The high rate of selfing found in *T. plicata* (O'Connell et al., 2001) may also have affected the proportion of assigned parent-offspring found in our study. Further research looking at levels of selfing on these sites may shed light on the number of half-sibships and related individuals revealing how this may affect the genetic diversity of future generations.

3.6 Conclusion

In this study we developed and validated a reduced SNP set (SNP set2) to evaluate genetic variation and genetic diversity in *T. plicata* populations. The high-quality SNP markers identified in our research could be useful in subsequent genetic investigations of *T. plicata* populations, offering a reliable, cost-effective, and time-efficient method for exploring genetic diversity and variation,

especially when dealing with large sample sizes. We found a large number of monomorphic SNPs in the adults from Longleat1, which may indicate the existence of fixed alleles compared to the other study sites implying a lower genetic variation within this site as well as a potentially distinct genetic background. When the private monomorphic SNPs were excluded, the genetic differentiation was low between Bagley Wood and Stourhead1, suggesting they may have been established with a similar source of plant or seed material. Juvenile individuals had similar levels of heterozygosity compared to adults across all sites, revealing that the natural regeneration may hold similar genetic diversity and gene pool size than the parents. This observation suggests there is a potential for genetic adaptation to current and future changes in the climate. However, although the heterozygosity levels are high within and across three of the study sites (Bagley Wood, Stourhead1 and Stourhead2), only Stourhead1 juveniles had $N_e \approx 100$. The low N_e suggests that the genetic makeup of *T. plicata* UK plantations may be rather narrow and potentially need an injection of diversity to ensure their genetic viability in the long-term. Although Stourhead2 was at a more advanced stage of transformation toward CCF and had the highest levels of both species and structural diversity, its N_e was very low for both adults and juveniles. In contrast, Bagley Wood is at a very early stage of transformation but its higher heterozygosity and N_e levels may indicate it is the most suitable site for transformation to CCF based on natural regeneration. These findings suggest that woodlands we visually perceive as diverse and naturalised may not hold enough genetic diversity to face current and future disturbances caused by the variation in the climate.

We presented a detailed study of four *T. plicata* populations, both for the level of CCF transformation and genetic diversity, producing useful insights for genetic resource management. Further analysis may be useful to investigate the bigger picture of the species as a non-native tree. Analyses including individuals from its natural range may allow us to determine the origin of seed sources used in the UK and thus to understand the origin of the fixed and potentially adapted alleles as observed in Longleat1 adults. Together with the investigation of the genetic diversity within provenance regions, this understanding may improve our knowledge of regional variations in N_e across the natural range and the implications in genetic diversity and natural levels of inbreeding.

3.7 Acknowledgements

The authors acknowledge the forestry estates where the samples came from: Stourhead (Western), Longleat (Wiltshire) and Bagley Wood. We also thank Barley Rose Collier Harris for helping with sampling. We want to thank Will Hoare, George MacKay and Emily Farley for their assistance in both laboratory and field work. LG received financial support from the Oxford-John Oldacre Graduate Scholarship and from Mr Henry Hoare. The project was funded in part by the OxLEP fund for the Oxford Plant Sciences Centre for Innovation.

3.8 References

Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., Lefèvre, F., Lenormand, T., Yeaman, S., Whetten, R., & Savolainen, O. (2013). Potential for evolutionary responses to climate change – evidence from tree populations. Global Change Biology, 19(6), 1645–1661. https://doi.org/10.1111/GCB.12181

Aldhous, J. R., & Low, A. J. (1974). The potential of Western hemlock, Western red cedar, Grand fir and Noblefir in Britain. H.M. Stat. Office.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, 17(2), Article 2. https://doi.org/10.1038/nrg.2015.28

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Barry, J. T., Francis, B. J., & Davies, R. B. (1989). SABRE: Software for the Analysis of Binary Recurrent Events. In A. Decarli, B. J. Francis, R. Gilchrist, & G. U. H. Seeber (Eds.), Statistical Modelling. Lecture Notes in Statistics, vol 57. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-3680-1_7

Bitterlich, W. (1984). The relascope idea: Relative measurements in forestry. Farnham Royal: Commonwealth Agricultural Bureaux.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bravo-Oviedo, A., Pretzsch, H., Ammer, C., Andenmatten, E., Barbati, A., Barreiro, S., Brang, P., Bravo, F., Coll, L., Corona, P., Den Ouden, J., Ducey, M. J., Forrester, D. I., Giergiczny, M., Jacobsen, J. B., Lesinski, J., Löf, M., Mason, B., Matovic, B., … Zlatanov, T. (2014). European mixed forests: Definition and research perspectives. Forest Systems, 23(3), 518–533. https://doi.org/10.5424/fs/2014233-06256

Choi, S. R., Oh, S. H., Dhandapani, V., Jang, C. S., Ahn, C. H., Rameneni, J. J., Kim, H., Jeon, I., &

Lim, Y. P. (2020). Development of SNP markers for marker-assisted breeding in Chinese cabbage using Fluidigm genotyping assays. Horticulture Environment and Biotechnology, 61(2), 327–338. https://doi.org/10.1007/S13580-019-00211-Y

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics, 12(7), Article 7. https://doi.org/10.1038/nrg3012

De Quesada, G., & Kuuluvainen, T. (2020). Tree diametric-species diversity is affected by human impact in old Scots pine dominated forest in boreal Fennoscandia. https://doi.org/10.1186/s40663-020-0219-6

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources, 14(1), 209–214. https://doi.org/10.1111/1755-0998.12157

Dray, S., & Dufour, A. B. (2007). Theade4Package: Implementing the duality diagram for ecologists. Journal of Statistical Software, 22(4). https://doi.org/10.18637/jss.v022.i04

Duminil, J., Mendene Abessolo, D. T., Ndiade Bourobou, D., Doucet, J.-L., Loo, J., & Hardy, O. J. (2016). High selfing rate, limited pollen dispersal and inbreeding depression in the emblematic African rain forest tree Baillonella toxisperma – Management implications. Forest Ecology and Management, 379, 20–29. https://doi.org/10.1016/j.foreco.2016.08.003

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics, 36(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE, 6(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, 10(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Forestry Commission. (2019). National Forest Estate Subcompartments England. https://data-forestry.opendata.arcgis.com/datasets/3993555ec8124b1e91b55a4a8b84567c_0/explore?location=52.881100%2C-1.770299%2C8.24

Frankham, R. (2022). Evaluation of proposed genetic goals and targets for the Convention on Biological Diversity. Conservation Genetics, 23(5), 865–870. https://doi.org/10.1007/s10592-022-01459-1

Frankham, R., Bradshaw, C. J. A., & Brook, B. W. (2014). Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation, 170, 56–63. https://doi.org/10.1016/j.biocon.2013.12.036

Friedline, C. J., Lind, B. M., Hobson, E. M., Harwood, D. E., Mix, A. D., Maloney, P. E., & Eckert, A. J. (2015). The genetic architecture of local adaptation I: The genomic landscape of foxtail pine (*Pinus balfouriana* Grev. & Balf.) as revealed from a high-density linkage map. Tree Genetics & Genomes, 11(3), 49. https://doi.org/10.1007/s11295-015-0866-x

Gärtner, S., & Reif, A. (2004). The impact of forest transformation on stand structure and ground vegetation in the southern Black Forest, Germany. Plant and Soil, 264(1–2), 35–51. https://doi.org/10.1023/B:PLSO.0000047751.25915.77

Gilman, P., Janzou, S., Guittet, D., Freeman, J., DiOrio, N., Blair, N., Boyd, M., Neises, T., & Wagner, M. (2019). PySAM (Python Wrapper for System Advisor Model 'SAM') [SWR-19-57] (4.2.0) [Computer software]. National Renewable Energy Laboratory (NREL), Golden, CO (United States). https://doi.org/10.11578/DC.20190903.1

Hamann, A., & Wang, T. (2006). Potential effects of climate change on ecosystem and tree species distribution in British Columbia. Ecology, 87(11), 2773–2786. https://doi.org/10.1890/0012-9658(2006)87[2773:peocco]2.0.co;2

Hebda, R. J., & Mathewes, R. W. (1984). Holocene History of Cedar and Native Indian Cultures of the North American Pacific Coast. Science, 225(4663), 711–713.

https://doi.org/10.1126/science.225.4663.711

Helliwell, R., & Wilson, E. R. (2012). Continuous Cover Forestry in Britain. Quarterly Journal of Forestry.

Heuertz, M., Carvalho, S. B., Galindo, J., Rinkevich, B., Robakowski, P., Aavik, T., Altinok, I., Barth, J. M. I., Cotrim, H., Goessen, R., González-Martínez, S. C., Grebenc, T., Hoban, S., Kopatz, A., McMahon, B. J., Porth, I., Raeymaekers, J. A. M., Träger, S., Valdecantos, A., Vella, A., Vernsesi, C., Garnier-Géré, P. (2023). The application gap: Genomics for biodiversity and ecosystem service management. Biological Conservation, 278, 109883. https://doi.org/10.1016/J.BIOCON.2022.109883

Hoban, S., Bruford, M., D'Urban Jackson, J., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., Paz-Vinas, I., Sjögren-Gulve, P., Segelbacher, G., Vernesi, C., Aitken, S. N., Bertola, L. D., Bloomer, P., Breed, M., Rodríguez-Correa, H., Funk, W. C., Grueber, C. E., Hunter, M. E., Jaffe, R., Liggins , L., Mergeay , J., Moharrek , F., O'Brien , D., Ogden , R., Palma-Silva , C., Pierson , J., Ramakrishnan , U., Simo-Droissart , M., Tani , N., Waits , L. & Laikre, L. (2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biological Conservation, 248, 108654. https://doi.org/10.1016/j.biocon.2020.108654

Hoban, S., Bruford, M. W., da Silva, J. M., Funk, W. C., Frankham, R., Gill, M. J., Grueber, C. E., Heuertz, M., Hunter, M. E., Kershaw, F., Lacy, R. C., Lees, C., Lopes-Fernandes, M., MacDonald, A. J., Mastretta-Yanes, A., McGowan, P. J. K., Meek, M. H., Mergeay, J., Millette, K. L., Mittan-Moreau, C.S., Navarro, L.M., O'Brien, D., Ogden, R., Segelbacher, G., Paz-Vinas, I., Vernesi, C. & Laikre, L. (2023). Genetic diversity goals and targets have improved, but remain insufficient for clear implementation of the post-2020 global biodiversity framework. Conservation Genetics. https://doi.org/10.1007/s10592-022-01492-0

Holliday, J. A., Hallerman, E. M., & Haak, D. C. (2018). Genotyping and Sequencing Technologies in Population Genetics and Genomics. 83–125. https://doi.org/10.1007/13836_2017_5

Høy Hansen, M., Steensboe Lang, C., Abildgaard, N., & Nyvold, C. G. (2022). Comparative evaluation of the heterozygous variant standard deviation as a quality measure for

next-generation sequencing. Journal of Biomedical Informatics, 135, 104234. https://doi.org/10.1016/j.jbi.2022.104234

Jarvis, P. J. (1973). North American Plants and Horticultural Innovation in England, 1550-1700. Geographical Review, 63(4), 477–499. https://doi.org/10.2307/213917

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics, 11(1), 1–15. https://doi.org/10.1186/1471-2156-11-94/FIGURES/9

Jones, A. T., Ovenden, J. R., & Wang, Y. G. (2016). Improved confidence intervals for the linkage disequilibrium method for estimating effective population size. Heredity, 117(4), 217–223. https://doi.org/10.1038/hdy.2016.19

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Molecular Ecology, 16(5), 1099–1106. https://doi.org/10.1111/J.1365-294X.2007.03089.X

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution, 4(8), 782–788. https://doi.org/10.1111/2041-210X.12067

Kerr, G. (2002). The potential for sustainable management of semi-natural woodlands in southern England using uneven-aged silviculture. Forestry, 75(3), 227–243. https://doi.org/10.1093/forestry/75.3.227

Kerr, G., Morgan, G., Blyth, J., & Stokes, V. (2010). Transformation from even-aged plantations to an irregular forest: The world's longest running trial area at Glentress, Scotland. Forestry, 83(3), 329–344. https://doi.org/10.1093/forestry/cpq015

Kerr, G., Snellgrove, M., Hale, S., & Stokes, V. (2017). The Bradford-Hutt system for transforming young even-aged stands to continuous cover management. Forestry, 90(4), 581–593. https://doi.org/10.1093/forestry/cpx009

Kirby, K. J., & Watkins, C. (Eds.). (2015). Europe's changing woods and forests: From wildwood to cultural landscapes. CAB International.

Kishor, D. S., Song, W. H., Noh, Y., Lee, G. P., Park, Y., Jung, J. K., Shim, E. J., & Chung, S. M. (2020). Development of SNP markers and validation assays in commercial Korean melon cultivars, using Genotyping-by-sequencing and Fluidigm analyses. Scientia Horticulturae, 263, 109113. https://doi.org/10.1016/J.SCIENTA.2019.109113

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: A package to manipulate and visualize variant call format data in R. Molecular Ecology Resources, 17(1), 44–53. https://doi.org/10.1111/1755-0998.12549

López de Heredia, U., Mora-Márquez, F., Goicoechea, P. G., Guillardín-Calvo, L., Simeone, M. C., & Soto, A. (2020). ddRAD Sequencing-Based Identification of Genomic Boundaries and Permeability in Quercus ilex and Q. suber Hybrids. Frontiers in Plant Science, 11, 1330. https://doi.org/10.3389/FPLS.2020.564414/BIBTEX

Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., & Allendorf, F. W. (2010). Estimation of census and effective population sizes: The increasing usefulness of DNA-based approaches. Conservation Genetics, 11(2), 355–373. https://doi.org/10.1007/s10592-010-0050-7

Macdonald, E., Gardiner, B., & Mason, W. (2010). The effects of transformation of even-aged stands to continuous cover forestry on conifer log quality and wood properties in the UK. Forestry: An International Journal of Forest Research, 83(1), 1–16. https://doi.org/10.1093/forestry/cpp023

Mason, B., Kerr, G., & Simpson, J. (1999). What is Continuous Cover Forestry ? Forestry Commission, 1–8.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Molecular Ecology Resources, 15(1), 28–41. https://doi.org/10.1111/1755-0998.12291

Meirmans, P. G. (2012). AMOVA-Based Clustering of Population Genetic Data. Journal of Heredity, 103(5), 744–750. https://doi.org/10.1093/jhered/ess047

Mijangos, J. L., Gruber, B., Berry, O., Pacioni, C., & Georges, A. (2022). dartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. Methods in Ecology and

Evolution, 13(10), 2150–2158. https://doi.org/10.1111/2041-210X.13918

Moore, J. C. (2013). Diversity, Taxonomic versus Functional. In S. A. Levin (Ed.), Encyclopedia of Biodiversity (Second Edition) pp. 648–656. Academic Press. https://doi.org/10.1016/B978-0-12-384719-5.00036-8

Mora-Márquez, F., García-Olivares, V., Emerson, B. C., & López de Heredia, U. (2017). ddradseqtools: A software package for in silico simulation and testing of double-digest RADseq experiments. Molecular Ecology Resources, 17(2), 230–246. https://doi.org/10.1111/1755-0998.12550

Mora-Márquez, F., Nuño, J. C., Soto, Á., & Heredia, U. L. D. (2023). Missing genotype imputation in non-model species using Self-Organizing Maps [Preprint]. Preprints. https://doi.org/10.22541/au.168596316.62218099/v1

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. Molecular Ecology Resources, 17(6), 1136–1147. https://doi.org/10.1111/1755-0998.12654

Neophytou, C., van Loo, M., & Hasenauer, H. (2019). Genetic diversity in introduced Douglas-fir and its natural regeneration in Central Europe. Forestry: An International Journal of Forest Research, 1–10. https://doi.org/10.1093/forestry/cpz055

Nguyen, N. N., Kim, M., Jung, J. K., Shim, E. J., Chung, S. M., Park, Y., Lee, G. P., & Sim, S. C. (2020). Genome-wide SNP discovery and core marker sets for assessment of genetic variations in cultivated pumpkin (*Cucurbita* spp.). Horticulture Research, 7(1). https://doi.org/10.1038/S41438-020-00342-9

Northwest Natural Resource Group. (2014). Forest Inventory and Monitoring Guidelines. Northwest Natural Resource Group.

O'Connell, L. M., Ritland, K., & Thompson, S. L. (2008). Patterns of post-glacial colonization by western redcedar (*Thuja plicata*, Cupressaceae) as revealed by microsatellite markers. Botany, 86(2), 194–203. https://doi.org/10.1139/B07-124

O'Connell, L. M., Viard, F., Russell, J., & Ritland, K. (2001). The mating system in natural populations of western redcedar (*Thuja plicata*). Canadian Journal of Botany, 79(6), 753–756.

https://doi.org/10.1139/cjb-79-6-753

Parchman, T. L., Jahner, J. P., Uckele, K. A., Galland, L. M., & Eckert, A. J. (2018). RADseq approaches and applications for forest tree genetics. Tree Genetics & Genomes, 14(3), 39. https://doi.org/10.1007/s11295-018-1251-3

Pommerening, A. (2002). Approaches to quantifying forest structures. Forestry: An International Journal of Forest Research, 75(3), 305–324. https://doi.org/10.1093/forestry/75.3.305

Pommerening, A., & Murphy, S. T. (2004). A review of the history, definitions and methods of continuous cover forestry with special attention to afforestation and restocking. Forestry, 77(1), 27–44. https://doi.org/10.1093/forestry/77.1.27

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics, 81(3), 559–575. https://doi.org/10.1086/519795

QGIS. (2023). QGIS Geographic Information System. Open Source Geospatial Foundation Project. (3.32.3) [Computer software]. http://qgis.org

R Core Team. (2018). R: A language and environment for statistical computing (4.3.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

RStudio Team. (2020). RStudio: Integrated development for R (2023.06.0) [Computer software]. Posit. http://www.rstudio.com/

Schmidt, T. L., Jasper, M.-E., Weeks, A. R., & Hoffmann, A. A. (2021). Unbiased population heterozygosity estimates from genome-wide sequence data. Methods in Ecology and Evolution, 12(10), 1888–1898. https://doi.org/10.1111/2041-210X.13659

Shalev, T. J., Gamal El-Dien, O., Yuen, M. M. S., Shengqiang, S., Jackman, S. D., Warren, R. L., Coombe, L., van der Merwe, L., Stewart, A., Boston, L. B., Plott, C., Jenkins, J., He, G., Yan, J., Yan, M., Guo, J., Breinholt, J. W., Neves, L. G., Grimwood, J., … Bohlmann, J. (2022). The western redcedar genome reveals low genetic diversity in a self-compatible conifer.

Genome Research, 1952–1964. https://doi.org/10.1101/gr.276358.121

Shalev, T. J., Gamal El-Dien, O., Yuen, M. M. S., van der Merwe, L., Kirst, M., Yanchuk, A. D., Ritland, C., Russell, J. H., & Bohlmann, J. (2023). Genetic architecture of terpene chemistry and growth traits and the impact of inbreeding on these traits in western redcedar (Thuja plicata). Evolutionary Applications. https://doi.org/10.1111/EVA.13526

Shalev, T. J., Yuen, M. M. S., Gesell, A., Yuen, A., Russell, J. H., & Bohlmann, J. (2018). An annotated transcriptome of highly inbred Thuja plicata (Cupressaceae) and its utility for gene discovery of terpenoid biosynthesis and conifer defense. Tree Genetics and Genomes, 14(3). https://doi.org/10.1007/s11295-018-1248-y

Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27(4), 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Standard Biotools. (2015a). Genotyping with the 192.24 IFC Using SNP Type Assays. PN 100-3913 C1. https://fluidigm.my.salesforce.com/sfc/p/#700000009DAw/a/4u0000019jzg/UO9mKHykLnoutJDRGZUzhqrEEXCWtEMtXX84mNlCiYQ

Standard Biotools. (2015b). Genotyping with the Flex Six IFC Using SNP Type Assays. PN 100-7718 C1. https://fluidigm.my.salesforce.com/sfc/p/#700000009DAw/a/4u0000019jvZ/e7Cfyl_yhhUodOMoT_gmPYJatwDM._GkDhjjNetzMMM

Stirling, R., Sturrock, R. N., & Braybrooks, A. (2017). Fungal decay of western redcedar wood products– a review. International Biodeterioration & Biodegradation, 125, 105–115. https://doi.org/10.1016/j.ibiod.2017.09.001

Vaux, F., Dutoit, L., Fraser, C. I., & Waters, J. M. (2023). Genotyping-by-sequencing for biogeography. Journal of Biogeography, 50(2), 262–281. https://doi.org/10.1111/jbi.14516

Vitkova, L., & Dhubháin, Á. N. (2013). Transformation to continuous cover forestry—A review. Irish Forestry, 119–140.

Wang, T., & Russel, J. H. (2006). Evaluation of selfing effects on western redcedar growth and yield in operational plantations using the tree and stand simulator (TASS). Forest Science, 52(3),

281–289. https://doi.org/10.5849/forsci.15-042

Waples, R. K., Larson, W. A., & Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. Heredity, 117(4), 233–240. https://doi.org/10.1038/hdy.2016.60

Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. Evolutionary Applications, 3(3), 244–262. https://doi.org/10.1111/j.1752-4571.2009.00104.x

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer New York. https://doi.org/10.1007/978-0-387-98141-3

Wojacki, J., Eusemann, P., Ahnert, D., Pakull, B., & Liesebach, H. (2019). Genetic diversity in seeds produced in artificial Douglas-fir (Pseudotsuga menziesii) stands of different size. Forest Ecology and Management, 438(February), 18–24. https://doi.org/10.1016/j.foreco.2019.02.012

Zaytsev, D. A., Danilov, D. A., & Navalihin, S. V. (2019). The effect of tree distribution by diameter classes on the density of pine and spruce trees in mixed coniferous forest stands without forestry impact. IOP Conference Series: Earth and Environmental Science, 226(1), 012066. https://doi.org/10.1088/1755-1315/226/1/012066

Zheng, H., Hu, D., Wei, R., Yan, S., & Wang, R. (2019). Chinese Fir Breeding in the High-Throughput Sequencing Era: Insights from SNPs. Forests, 10(8), Article 8. https://doi.org/10.3390/f10080681

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	GBS-Based SNP Genotyping and Genetic Diversity Assessment in *Thuja plicata* Adults and Juveniles in UK Continuous Cover Forestry
Publication Status	□Published □ Accepted for Publication □Submitted for Publication ☑Unpublished and unsubmitted work written in a manuscript style
Publication Details	Guillardín, L.*, Glover, E., Kerr, G., MacKay, J.J.* (unplublished) GBS-Based SNP Genotyping and Genetic Diversity Assessment in *Thuja plicata* Adults and Juveniles in UK Continuous Cover Forestry.

Student Confirmation

Student Name:	Laura Guillardin Calvo	
Contribution to the Paper	Design the study, conceived and planned the experiments, coordinated sampling and laboratory experiments. Performed bioinformatic and statistical analyses and prepared the figures. Drafted and revised the manuscript.	
Signature 	Date	15/11/2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: John MacKay, Wood Professor of Forest Sciences		
Supervisor comments I agree with the above.		
Signature 	Date	05/01/2024

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 4. High Heterozygosity but low levels of Effective Population Size in UK Continuous Cover Forestry planted stands of *Pseudotsuga menziesii*

Laura Guillardin[1], Ella Glover[1], Gary Kerr[2], John J. MacKay[1]

1. Department of Biology, University of Oxford, Oxford OX1 3RB, United Kingdom

2. Forest Research, Alice Holt Lodge, Farnham, Surrey, GU10 4LH, United Kingdom

4.1 Abstract

Pseudotsuga menziesii is a conifer native to western North America that is commonly planted in the UK for its high timber value and is usually found in even-aged pure stands. The transformation towards more diversified woodlands using a Continuous Cover Forestry (CCF) approach may increase the ability of these forests to face ongoing environmental shifts. However, such plantations may lack genetic diversity or transfer a reduced gene pool to naturally regenerated offspring, which could result in future forests not adapted to face environmental changes. We performed a forest survey to analyse the species diversity and composition of four stands allowing us to categorise them among the different CCF transformation stages. To assess the genetic diversity of these CCF sites, we genotyped individuals from different strata using two locus-specific genotyping systems with an assay designed to genotype a total of 176 SNPs recovered from genetic markers datasets described previously in the literature. Our results showed a high level of genetic diversity within the *P. menziesii* populations in England, with an overall mean observed heterozygosity of 0.48. The genotypic comparison between generations revealed no differences in heterozygosity but indicated substantial differences in effective population size (N_e). This genetic diversity indicator also varies among sites, detecting lower N_e in sites with lower species diversity levels. These findings shed light on the potential vulnerability of such populations to changing conditions and highlight the need for genetic resource management strategies to ensure the continued viability of these forests in the face of ongoing environmental changes.

4.2 Background

Pseudotsuga menziesii (Mirb.) Franco was introduced in Europe in the 19th century for ornamental purposes (Holmes, 1975) but the use of non-native conifers for timber production such as *P. menziesii*, *Thuja plicata* and *Picea sitchensis*, rapidly increased in the UK (Raum & Potter, 2015). *P. menziesii* is a large evergreen conifer native to North Western America used by First Nations people before 1850 European settlements for timber (McDadi & Hebda, 2008) and traditionally managed to obtain a valuable sweet substance (Douglas maple) (Kuhnlein & Turner, 1991). It occurs in two different ecological varieties, *P. menziesii* var. *menziesii* (coastal variety), and *P. menziesii* var. *glauca* (interior variety, or Rocky-Mountain variety), which hybridise naturally in the wild (van Loo et al., 2015), although there are no records of the interior variety in the UK, as it has a lower growth potential than the coastal variety (Ferrell & Woodard, 1966). In the second half of the 20th century, new forestry policies were developed in the UK to explicitly include other aims than timber harvest (Holmes, 1975), leading to the current reference policy published by the Forestry Commission (2017) named the United Kingdom Forestry Standards (UKFS). The UKFS defines requirements and standards of sustainability and biodiversity as mandatory for every woodland in the UK to increase resilience to rapid environmental change. This policy encourages the transformation of even-aged pure populations towards mixed irregular ones where a continuous cover is guaranteed, as proposed by the Continuous Cover Forestry (CCF) approach (Helliwell & Wilson, 2012). Pommerening & Murphy (2004) developed the original structure diversity diagram which we modified and presented in Figure 1. This diagram shows four different woodland's structure types, from an even-aged plantation to a late stage of CCF, with multi and uneven-aged woodlands in between. Although not every woodland may go through every stage, this diagram together with the estimation of species diversity indexes allows us to broadly categorise the study sites stage of transformation towards CCF.

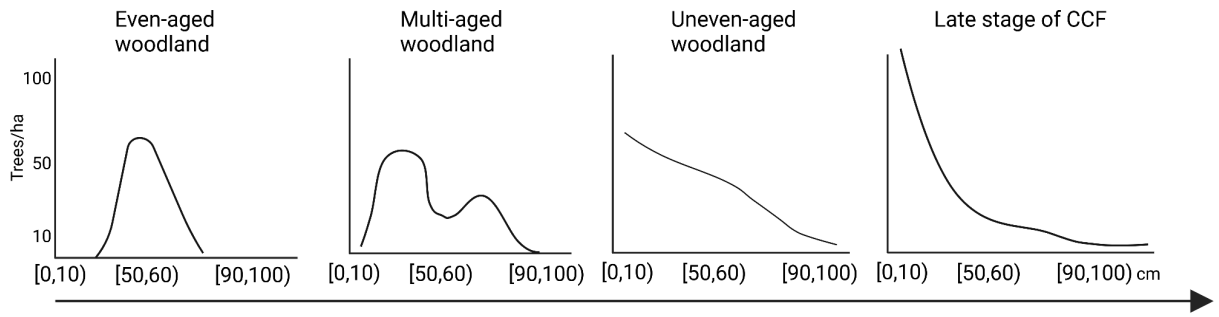


Figure 1. Structure diversity diagram transition from a plantation towards a CCF woodland. Modified from Pommerening & Murphy (2004). The x-axis represent the diametric classes in cm and the y-axis is the number of trees/ha.

Rapid environmental changes risk outpacing the ability of tree species to adapt (Rosenzweig et al., 2008; Taylor et al., 2015). Biodiversity loss, appearance of invasive species, potential physical damage from mamals, and an exponential increase in the number of invertebrated pests and pathogens, have been detected affecting UK forests regarding the current climate and biodiversity crisis (Tew et al., 2023). Thus, in order to anticipate the potential effects on forests, it is vital to understand both the phenotypic plasticity and adaptation ability of forest species to cope with observed and predicted conditions (Kremer et al., 2012, 2014; Peters, 1990). Researchers have outlined the potential benefits of the CCF approach for increasing woodland resilience, have defined silvicultural practises to transform even-aged plantations to an irregular woodland structure (Kerr et al., 2010; Mason & Kerr, 2000), and described methods to monitor the transition (Bennett et al., 2020; Kerr et al., 2017). However, there is a paucity of studies looking at the genetic context of CCF. Moreover, although several studies have investigated the genetic diversity of planted *P. menziesii* stands in central Europe (Eckhart et al., 2017; Pakull et al., 2021; Wojacki et al., 2019), including a comparison between adults and juveniles (Neophytou et al., 2019; van Loo et al., 2019), there are not yet studies investigating the genetic status of *P. menziesii* planted woodlands in the UK.

Finkeldey (2002) discussed the genetic impact of forests being transformed into CCF associated with a change in the population density due to selective cutting. A decrease in the number of individuals that potentially could contribute genetically to the offspring (effective population size

(Ne)) leads to an increase in inbreeding and selfing, limiting the gene pool of future natural generations. Subsequently, future stands may have reduced standing genetic diversity and be less able to adapt to emerging climate stressors (Finkeldey, 2002). When primary trees hold a reduced gene pool, the vigour of the future forest may be further compromised. N_e has become an important parameter for assessing genetic diversity and has been estimated in several studies (Capblancq et al., 2020; Slavov et al., 2012) including in *P. menziesii* (Wojacki et al., 2019) with genetic diversity monitoring and conservation aims.

Over the years, extensive research efforts have yielded essential genetic resources available for *P. menziesii*, including the assembled genome (Neale et al., 2017), a comprehensive transcriptome (Howe et al., 2013), and informative Single Nucleotide Polymorphism (SNP) datasets (Howe et al., 2013, 2020; Krutovsky & Neale, 2005; Müller et al., 2012a, 2012b). These resources provide the ability to analyse *P. menziesii* populations, enabling insights into different genetic analyses such as population structure, adaptation, and diversity. Additionally, targeted genotyping and high-throughput platforms such as Illumina Infinium iSelect HD, (Illumina, San Diego, CA) the Sequenom MassARRAY platform (Agena Bioscience, San Diego, CA), and Standard Biotools Dynamic Arrays (San Francisco, CA, USA) are a useful alternative to sequencing due to their lower cost and time-efficiency when analysing large sample sizes and, their low levels of missing data (Choi et al., 2020; Davey et al., 2011). In this study, we developed a targeted genotyping assay using SNP datasets available (Howe et al., 2020) for *P. menziesii* to assess the genetic diversity of a large number of samples.

Currently, CCF is being deployed in some UK woodlands to increase biodiversity, mitigate climate change and increase woodland resilience while maintaining timber production. *P. menziesii* is widely planted in the UK mostly under clear-felling management but it has become a predominant species in UK CCF-managed forests. However, the levels of genetic diversity are uncharacterised in the planted canopy trees, and silvicultural practices are applied to these woodlands, such as the selective harvesting selection and removal of trees for timber production. Thus the hypothesis of this work is based on a potential loss of genetic diversity, specifically a reduction of N_e from one

generation to the next, as natural regeneration is encouraged under this management approach. To test this hypothesis, we analysed *P. menziesii* trees using PCR-based genotyping methods. The main goal of this study is to assess genetic diversity differences between planted trees and their offspring from four UK woodlands in transition to CCF. These specific objectives were: 1) To characterise the level of transformation of the CCF stands based on species diversity and stand structure, 2) To design and validate a reliable SNP genotyping assay and 3) To assess and compare the genetic diversity in planted adults and naturally regenerated juveniles within each stand.

4.3 Material and Methods

4.3.1 Study sites characterisation: structure and species diversity

We sampled four CCF sites located in England (UK) that included *P. menziesii* as part of the canopy and are identified as Ringwood, Mortimer, Stourhead2, and Longleat2. To avoid small sample size biases that may affect genetic diversity estimations, each study site had at least 100 adult trees as recommended by Wojacki et al. (2019). The estimated year of planting (Table 2) was provided by owners and managers. However, no information regarding the source of the planted material was found. We assessed every site in order to ensure sufficient viable natural regeneration required for our study. We used the National Forest Inventory Woodland public dataset (Forestry Commission, 2019) and private land cover shapefiles (supplied by landowners) to calculate each site area using QGIS software (QGIS, 2023) and evaluate the level of isolation from other *P. menziesii* woodlands within a 2 Km radius. Additionally, we defined the stand distribution as patchy, if the stand is fragmented or as continuous. We performed a forest survey between September 2020 and June 2021 following the recommended methodology described in the Forest Inventory and monitoring guidelines (Northwest Natural Resource Group, 2014) to estimate the level of transformation from a plantation towards an old-growth CCF woodland (Figure 2).

We established up to 25 plots in each site randomly located using QGIS (QGIS, 2023) and exported them into a GPX file format to upload to a GPS device. The number of plots varied in each site due to its size and the density of adult trees, ranging from 16 in Ringwood to 25 in Stourhead2. The total number of adult trees measured per site varied, ranging from 67 in Stourhead2 to 91 in Longleat2. Overall, the average number of adult trees within one plot ranges from 2.7 in Stourhead2 to 4.6 in Mortimer. These data can be found in the Supplementary online repository (https://github.com/LauGuillardin/Thesis_Supplementary_Information/tree/main/Inventory). To exemplify the study design, we created a diagram that showed a simulated site with 16 plots (Figure 3).

We used 8 or 6 m radius plots, depending on the site, and every natural regeneration plot was of 3m radius (Figure 2). We measured the tree dimensions of each individual over 10 cm in diameter inside the tree plot: diameter at breast height (dbh) using a dbh tape, and height, using the Bitterlich Relascope method (Bitterlich, 1984). We measured the dimensions of regenerated seedlings and saplings within the regeneration plot using a digital calliper and a metre stick and took a census of each species (Figure 2). We estimated the overall canopy density of each plot by using a densiometer and determined the *P. menziesii* average height. Additionally, we identified the species of shrubs, its average height and evaluated the proportion of each species ground coverage within the regeneration plot. We calculated both the overall and specific tree density using the site area and number of trees per plot following the Forest Inventory and monitoring guidelines (Northwest Natural Resource Group, 2014).

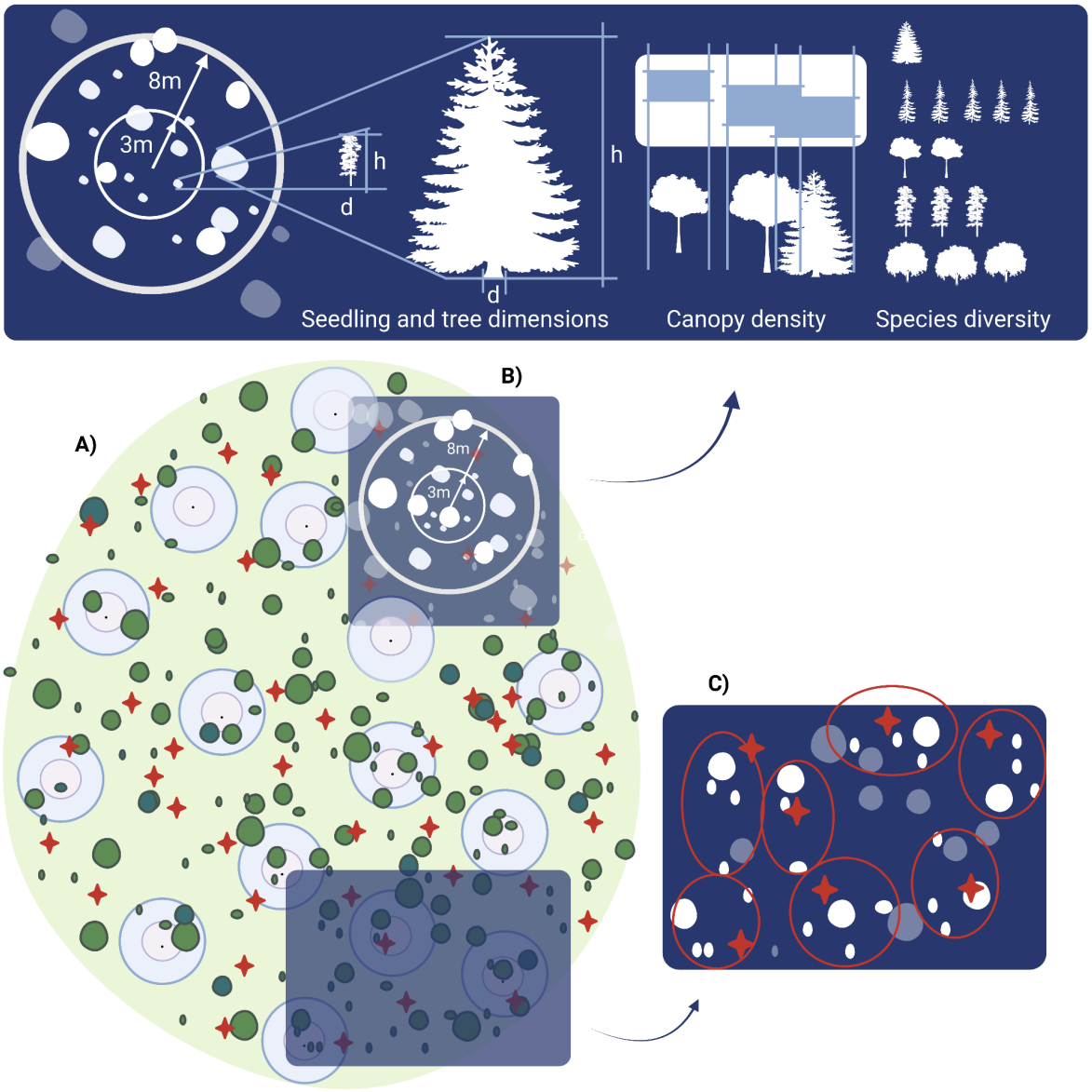


Figure 2. Diagram of field work carried out for both, the forest survey and the sampling of tissue for DNA extraction. A) Represents a simplified site where trees are the green shapes of different sizes, the larger are adult trees and the smaller are seedlings or saplings. The circular concentric circles represent the tree and natural regeneration plots for the forest survey, and the red stars are the GPS random points for the tissue sampling. B) Forest inventory plots description and parameters measured. Seedling and tree dimensions: d = diameter at breast high (cm), h= high (m), We included the context of the location of the plots within a site. C) Example of the selection of seven adult trees (white) and the three seedlings or saplings (white) assigned per adult for tissue sampling.

To measure the canopy, offspring snd shrub species diversity, we estimated the following indexes: 1) the Species richness (S) or actual number of different species, 2) The Shannon index, *H'*, (Shannon, 1948), which considers both species richness and evenness, 3) the Species evenness (E) (Moore, 2013), which describes the commonness or rarity of a species within a stand, and 4) the Effective number of species (Moore, 2013), considered the number of equally-common species required to give a particular value of a Shannon index (H').

The Shannon entropy index was calculated by counting the number of different species and their frequency of appearances (Figure 2), and then applying the following formula (1) where S = species richness and pi = relative abundance of species i.

$$H' = -\sum_{i=1}^{S} p_i \cdot log\, p_i \qquad (1)$$

The Species evenness, which ranges from 0 to 1, and the Effective number of species were calculated using the formulas (2) and (3), respectively:

$$E = \frac{H'}{ln(S)} \qquad (2)$$

$$Effective\ number\ of\ species = EXP(H') \qquad (3)$$

The raw inventory data for each study site, the estimated canopy density and the data needed to calculate the species diversity indexes can be found in the Supplementary online repository (https://github.com/LauGuillardin/Thesis_Supplementary_Information/tree/main/Inventory).

4.3.2 Sampling and DNA isolation

Foliage was collected from canopy trees following a modified method proposed by Youngentob et al., (2016) using an arborist slingshot to sample trees with very high branches unreachable with a telescopic pruner. We also collected foliage from three naturally regenerated seedlings near every adult that was sampled as exemplified in Figure 2 panel C. We recorded the GPS location of each sampled adult and assigned the same waypoint to the three adjacent juveniles. The number of sampled adults ranged from 47 to 50 and from 133 to 150 juveniles in each of the sites

(Table 1). These sampling design is independent to the forest survey described in section *4.3.1 Study sites characterisation: structure and species diversity,* therefore, the trees measured for dbh and height are not the ones we selected to sample leaves from. The foliage was dried by storing each sample in separate paper envelopes which were placed in sealed plastic bags or boxes with silica gel.

The DNA was isolated using the DNeasy Plants mini kit (Qiagen, USA, Valencia, CA) modified method described in Chapter 2 (see section *DNA isolation procedures and improvements*). DNA samples that were used for genotyping had a minimum of 20 ng/ul of DNA determined with a Qubit 4 Fluorometer (Invitrogen, Thermo Fisher Scientific, Cleveland, OH, USA), and a minimum absorbance ratio of 1.8 for both A260/A280 and A260/A230 using NanoDrop One Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE).

4.3.3 Overview of genetic analyses

We designed a reduced SNP assay that offers cost-effective genotyping of informative SNPs, making it an efficient choice for large sample sizes in non-model tree species, facilitating reliability and data analysis. We used two PCR-based genotyping systems developed separately to investigate the genetic diversity of four UK sites: 1) Standard Biotools microfluidics SNP genotyping platform (Standard Biotools, San Francisco, CA, USA) (Std Biotools assay) and 2) Sequenom MassARRAY® iPLEX GOLD technology (Agena Bioscience, San Diego, California, USA) (Sequenom assay). The initial objective was to develop an Std biotools assay for use in our laboratory; however, the opportunity arose to validate and complement our results with data from an independent external platform, i.e. Sequenom, which has been used in other research on conifers (Godbout et al., 2017; Trembizki et al., 2014; Vidal et al., 2017).

For the Std biotools assay, we selected 96 SNPs from available resources to design the Standard Biotools panel based on markers characteristics (Figure 3). The Sequenom assay includes 80 SNPs previously developed and tested by Howe et al (personal communication), and was implemented as a collaboration between the Conifer Breeding Co-operative, the University of Oxford,

Forest Research, and the UK Centre for Hydrology and Ecology. This collaboration allowed us to increase the number of markers to provide better statistical power in our population genetics analysis.

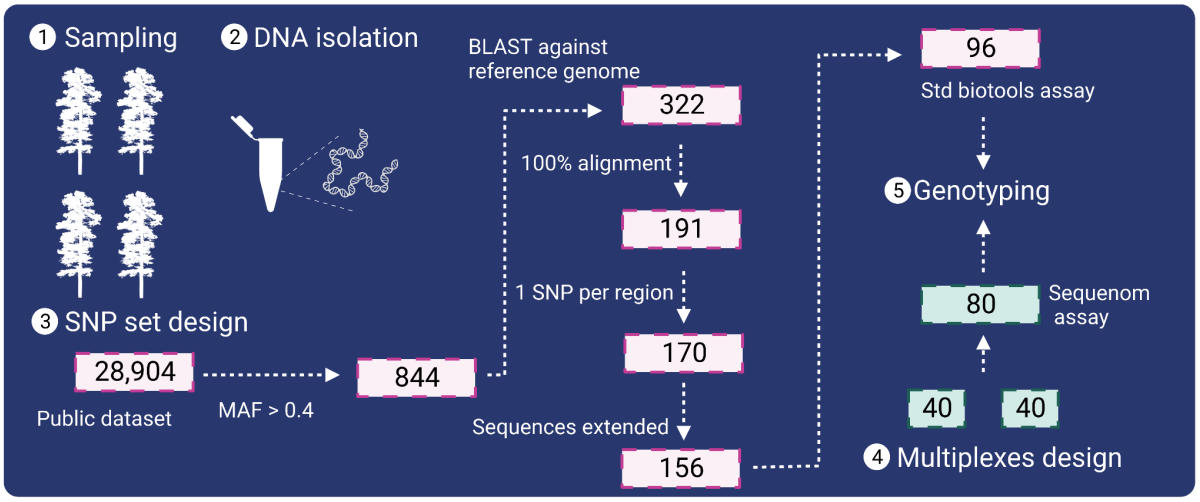


Figure 3. Genetic analysis workflow. A detailed description of each step performed from the sampling (1) and DNA isolation (2) to the design of both SNP sets (3 and 4) to use in the genotyping step (5). MAF = minor allele frequency.

4.3.4 SNP assay design & Genotyping

For the Std biotools assay, we selected SNPs available discovered by Howe et al., (2020). We filtered by minor allele frequency (MAF) > 0.4 aiming to detect the presence of the variants in our study populations as the SNPs were discovered in other populations and a high MAF may represent common variants. Then, we aligned the SNP sequences to the *P. menziesii* genome assembled (Neale et al., 2017), to assess the accuracy and reliability of the SNP data and, only considered the SNPs with a complete alignment (Figure 3) using BLAST (Camacho et al., 2009). We retained only one SNP per region to maximise the genome coverage and we used the efetch e-utility tool from NCBI (Kans, 2023) to find the SNP location in the genome and download only the scaffolds that contained the SNPs of interest. This also allowed us to extend the flanking sequence to 100 bp using the grep bash tool, to enable primer design. We submitted the SNP sequences to the D3 assay primer designer (Standard Biotools, San Francisco, CA, USA) and kept those with 40-60% GC content to ensure stable binding to the template DNA. We tested 156 SNPs on ten samples and used the call rate and

level of polymorphisms to select the final 96 SNPs. This test was performed using FlexSix IFC plates on the Standard Biotools microfluidics SNP genotyping platform using the Juno system to load and combine the microfluidics, the Biomark to perform the PCR, and the FluidiGM SNP Genotyping software to digitalised and analyse the image outputs (all from Standard Biotools, San Francisco, CA, USA). We followed a modified preparation protocol (Standard Biotools, 2015b) using Fast Probe Master Mix (Biotium Inc. Fremont, CA) and avoided the pre-amplification step improving the assay performance. To genotype all samples we followed the same procedure but we used the 192.24 ICF plates and followed the manufacturer protocol (Standard Biotools, 2015a). We used four plates per study site to genotype each sample using the selected 96 SNPs, each plate performed 4608 PCR reactions so we ran a total of 73,728 PCR reactions. The results were visualised using the FluidiGM SNP genotyping software, we set the confidence threshold to 50 which had been tested on replicate samples to ensure a high sensitivity allowing more genotype calls, without introducing a higher risk of false positives. We removed the SNPs with >50% of missing data.

The Sequenom chip was designed following the recommendation of Howe (personal communication), which included one multiplex of 40 SNPs already tested and produced reliable results and high levels of polymorphism (not published), while the other multiplex of 40 SNPs were selected from a ranked list of SNPs provided by Howe (not published) (Figure 3). The samples were genotyped at the Genome Transcriptome Facility of Bordeaux (INRAE, University of Bordeaux) using the Sequenom MassARRAY® iPLEX GOLD technology (Agena Bioscience, San Diego, California, USA).

We estimated the heterozygosity levels from both platforms across sites to validate the Std biotools assay and combined the results to genotype all the individuals from the study sites. We analysed the genetic diversity within sites using the data from both platforms; however, the kinship analysis only used the genotype data from Std biotools platform as we incorporated all the juveniles in the analysis, including those juveniles that were not genotyped by the Sequenom platform (Table 1) as this data was part of an additional project. We retained only the genotypes that showed < 20% of missing data.

Table 1. Description of which genotyping platform, number of samples and number of SNPs used in each genetic analysis. N = range of number of samples analysed, detailed numbers can be found in Table 5 for Genetic analysis and in Table 7 for Kinship analysis.

Source	Sample type	Genetic diversity (N)	Kinship analysis (N)
Study sites	Adults	47-49	48-50
Study sites	Juveniles	48-55	133-150
Genotyping platform		Sequenom + Std biotools	Std biotools
Number of SNPs		176	96

4.3.5 Population genetics analysis

Genetic diversity and genetic variation

We estimated different genetic diversity indicators such as heterozygosity (observed and expected), the inbreeding coefficient (F_{is}) and the effective population size (N_e) using both genotyping assays (Table 1). We analysed all the adults in each site but we selected one of the three juveniles per adult to avoid biases due to different sample sizes in each group. We used the gl.report.heterozygosity and gl.test.heterozygosity functions from dartR package (Mijangos et al., 2022) implemented in R studio (R Core Team, 2018; RStudio Team, 2020) to estimate the heterozygosity (observed and expected) levels and significant differences of both adults and juveniles. We used the R package diveRsity (Keenan et al., 2013) to estimate the F_{is} and 95% confidence intervals executing 1000 bootstraps. We estimated the contemporary N_e of adults and juveniles using the linkage disequilibrium method implemented in NeEstimator V2.1 software (Do et al., 2014) and used the recommended Jackknife method to estimate the confidence intervals (Jones et al., 2016). We estimated the population census size per site (N_c) using the tree/density and site size to calculate the N_e/N_c ratio.

To evaluate the amount of genetic variation among sites and within sites defining two categories, adults and juveniles, we performed a hierarchical Analysis of Molecular Variance (AMOVA) using Arlequin (Excoffier & Lischer, 2010) with P-value = 0.05.

Kinship analysis

We estimated parent-offspring relationships found in each of the study sites including all the samples collected, including all the juveniles, using the software CERVUS (Kalinowski et al., 2007). We only used the SNPs from the Standard Biotools (Table 1) and filtered those that exhibited departure from Hardy-Weinberg equilibrium (HWE) per site to focus on relationships that are less influenced by biological processes like inbreeding or selection, which can lead to deviations from HWE. The final number of SNPs used per site is shown in Table 7. We used all the adults as candidate parents of all the juveniles sampled, including those discarded in the genetic diversity and genetic variation analyses. To perform the test, we simulated a random-mating population randomising 10,000 mating events using the observed allele frequency of the stands to assign relaxed (80%) confidence levels. To determine the distance between related parents and offspring we used the software (QGIS, 2023) and geographic location data collected when the foliage was sampled.

4.4 Results

4.4.1 Study sites characterisation: structure and species diversity

Mortimer, Ringwood and Longleat2 shared a similar surface area and Stourhead2 is the largest and also has the tallest *P. menziesii* individuals on average although the tallest individual tree was found in Longleat2 (55 m) (Table 2). The canopy cover ranged from 63.5% in Longleat2 to 85.9% in Mortimer indicating a large coverage that could compromise the development of natural regeneration. There is a large difference between the overall and specific tree density shown in Ringwood, Stourhead2 and Longleat2 revealing that these woodlands are composed of a mix of species in various densities (Table 2 and Table 4).

Table 2. Site characterisation results, including overall site estimations and specific estimations for *Pseudotsuga menziesii* individuals.

		Overall estimations					*Pseudotsuga menziesii*		
Site	Year planted	Site area (ha)	Isolated (> 2Km)	Stand distribution	Canopy cover (%)	Tree density (trees/ha)	Tree density (trees/ha)	Avg. Height (m)	Max. Height (m)
Mortimer	1980	5.33	No	Continuous	85.9	274.6	265.2	27.9	48
Ringwood	1961	5.77	No	Continuous	75.5	195.8	79.6	25.1	38.5
Stourhead2	1939	18.57	No	Continuous	73.5	226.4	92	29.3	44.5
Longleat2	1925	6.87	No	Continuous	63.5	213.2	137.4	26.3	55

A list of the different adult, offspring and shrub species found in each site is shown in Table 3. Every site had more tree species in its offspring than in the canopy strata, which may indicate a colonisation from neighboring sites or that the adult trees of these species were outside every measured plot. Mortimer is the only site without broadleaves species in the canopy. Due to the variety of species' abundance, the effective number of species is smaller than the actual number of species (S) in every strata. Longleat2 is the site with the highest S in the canopy (S = 9) but had an effective number of species of 2.7 (Table 3 and Table 4), additionally, its species evenness is the lowest among the canopy strata of all the sites (E = 0.45). Stourhead2 had the highest effective number of species (S = 4.5) and the highest species eveness (E = 0.73) in the canopy strata. Both Longleat2 and Stourhead2

showed high canopy species diversity (*H'* > 1.5) in both adults and juveniles, Mortimer had very low species diversity in adults (*H'* < 0.5) and Ringwood was intermediate (Table 4).

The diameter distribution profiles from Stourhead2 showed a reverse-j profile indicating a late stage of CCF, although had a low number of trees in the smallest diametric class (Figure 4). Therefore, Stourhead2 profile could be considered an uneven-aged woodland (Figure 1). Longleat2 seems to be in a transition into a reverse-j found in uneven-aged woodlands or in late stage of CCF (Figure 4 and Figure 1) but exhibited a rotated sigmoid curve found in multi-aged woodlands (Figure 1). Mortimer had a pure plantation distribution profile while Ringwood had two nodes which may indicate an intermediate stage of transformation (multi-aged woodland) into CCF (Figure 4 and Figure 1) according to Pommerening & Murphy (2004).

Table 3. List of species found in each stand separated by strata. The code represents the scientific name of the species found as follows: Pm: *Pseudotsuga menziesii*, Ps: *Picea sitchensis*, Ld: *Larix decidua,* Sa: *Sorbus aucuparia,* Tb: *Taxus baccata*, Pa: *Picea abies*, Th: *Tsuga heterophylla*, Ia: *Ilex aquifolium*, Ag: *Abies grandis*, Paq: *Pteridium aquilinum*, Rf: *Rubus fructicosa*, Ud: *Urtica dioica*, Bsp: *Betula* spp., Rp: Rhododendron ponticum, Qp: *Quercus petraea*, Fe: *Fraxinus excelsior*, Fs: *Fagus sylvatica*, Tp: *Thuja plicata*, Ap: *Acer pseudoplatanus*, Qr: *Quercus robur*, Ca: *Corylus avellana*, Bp: *Betula pubescens*, Cs: *Castanea sativa*, Cv: *Calluna vulgaris*, Jsp: *Juncus* spp., Csp: *Carex* spp., Ea: *Equisetum arvense,* Bs: *Buxus sempervirens,* Hh: *Hedera helix,* Lp: *Lonicera periclymenum.*

	Mortimer			Ringwood		
	Canopy	Offspring	Shrubs	Canopy	Offspring	Shrubs
	Pm Ps Ld	Sa Pm Tb Pa Th Ia Ps Ag	PaQ Rf Ud	Pm Th Ps Bsp	Pm Th Ps Ia Sa Bsp	Paq Rp Usp Rf
Species richness (S)	3	8	3	4	6	4

	Stourhead2			Longleat2		
	Canopy	Offspring	Shrubs	Canopy	Offspring	Shrubs
	Th	Fs	Paq	Pm	Ia	Paq
	Ps	Ia	Rf	Th	Bp	Rf
	Pm	Th	Ud	Cs	Th	Rp
	Ld	Pm	Jsp	Ap	Sa	Hh
	Tp	Ps	Csp	Ia	Pm	Lp
	Qp	Tp	Ea	Bp	Fs	Ud
	Pa	Ap		Ssp	Ld	
	Fe	Ag		Bs	Ap	
		Ld		Pa	Bs	
		Qr			Fe	
		Ca				
		Bp				
Species richness (S)	8	12	6	9	10	6

Table 4. Species diversity indexes estimated for each strata and study site.

	Mortimer			Ringwood		
	Canopy	Offspring	Shrubs	Canopy	Offspring	Shrubs
Shannon entropy index (H)	0.172	1.260	0.805	0.720	1.286	1.001
Species evenness (E)	0.156	0.606	0.732	0.520	0.718	0.722
Effective number of species	1.187	3.527	2.236	2.055	3.617	2.722
	Stourhead2			Longleat2		
	Canopy	Offspring	Shrubs	Canopy	Offspring	Shrubs
Shannon entropy index (H)	1.51	1.81	1.03	0.99	1.67	1.43
Species evenness (E)	0.73	0.73	0.58	0.45	0.72	0.80
Effective number of species	4.53	6.09	2.81	2.70	5.29	4.18

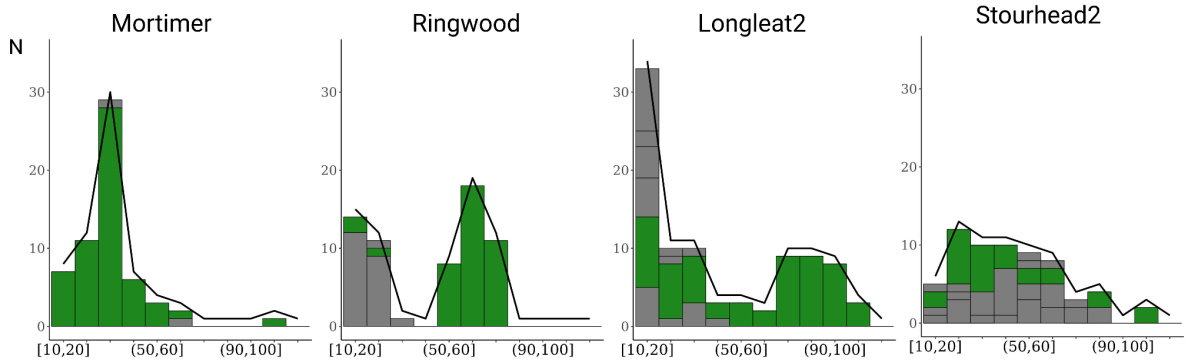


Figure 4. Structural diversity per site shown as the number of trees (from the canopy strata) in each diametrical class in cm. *P. menziesii* individuals coloured in blue, other species are shown in grey.

4.4.2 Genotyping assay development

We designed the Std biotools assay with the 96 SNPs most polymorphic and with the highest call rate to ensure robustness across populations and low levels of missing data after testing 10 samples on 156 SNPs. After genotyping the total sample of 777 individuals using the 96 SNPs we visualised the results using FluidiGM SNP genotyping software and removed 24 markers with >50% of missing data. The overall call rate of the remaining 72 SNPs across sites is 91.2% and no private monomorphic nor polymorphic SNPs were found.

We validated the assay by comparing the expected heterozygosity levels when using the Sequenom platform, which ranged from 0.47 to 0.48 across sites, to the Std biotools results ranging from 0.48 to 0.49, revealing the same patterns in the genetic diversity levels captured when using both platforms. We integrated both datasets to increase the statistical power of the genetic diversity and genetic variation analyses. The Std biotools and Sequenom assays can be found in the Supplementary online repository. (https://github.com/LauGuillardin/Thesis_Supplementary_Information/tree/main/Genomic_data/PM).

4.4.3 Genetic diversity and genetic variation

Every study site had high levels of heterozygosity (observed and expected), ranging from H_o = 0.437 in Longleat2 juveniles to H_o = 0.456 in Stourhead2 juveniles and no significant differences were found between adults and juveniles across sites (Table 5). Overall, the inbreeding coefficient showed low levels across all sites ranging from 0.015 to 0.061 revealing low variability of F_{is} across sites and no significant differences between adults and juveniles within sites were found.

The effective population size (N_e) levels varied among sites and within sites (Table 5) and only Ringwood adults showed high levels (N_e = 428.6), although substantially lower N_e was found in juveniles (N_e = 271.9). The lowest levels of N_e were found in Mortimer both adults (Ne = 112.5) and juveniles (N_e = 154.2). The Ne/Nc ratio overall average is 0.32, it varied from 0.08 in Mortimer to 0.60 in Ringwood.

Table 5. Genetic diversity indicators. Differences between adults (A) and juveniles (J) within each site are reported as significant p-value <0.05 (*). N_c = population census size, N_s = number of samples collected, N_e = effective population size, C.I = confidence intervals, S.D = standard deviation, H_o = observed heterozygosity, H_e = expected heterozygosity, F_{is} = Inbreeding index.

	Stourhead2		Longleat2		Mortimer		Ringwood	
	A	J	A	J	A	J	A	J
N_c	1707.6		943.7		1413.8		717.4	
N_s	49	53	49	56	47	48	48	55
N_e	271.9	259.7	395.4	302.7	112.5	154.2	428.6	271.9
lower C.I	173.8	172.7	219.9	196.3	90	116.6	229.6	182.2
upper C.I	586.2	491.4	1630.9	623.9	147.6	222.2	2434	510.7
N_e / N_c	0.16		0.42		0.08		0.60	
H_o	0.443	0.456	0.454	0.437	0.444	0.449	0.446	0.444
H_o S.D	0.088	0.082	0.091	0.086	0.099	0.092	0.087	0.089
H_e	0.481	0.480	0.480	0.481	0.477	0.479	0.482	0.482
H_e S.D	0.025	0.027	0.024	0.025	0.030	0.028	0.025	0.022
F_{is}	0.029	0.022	0.015	0.065	0.025	0.042	0.061	0.059
lower C.I	0.001	-0.001	-0.017	0.041	0.000	0.016	0.036	0.034
upper C.I	0.055	0.044	0.045	0.087	0.047	0.065	0.084	0.083

We found statistical genetic variance among sites but not between adults and juveniles within sites. The variation explained by differences among sites only explained 0.59%, and most of the variation is explained by differences within groups, indicating high variability among individuals from the same sites (Table 6).

Table 6. Analysis of the Molecular Variance (AMOVA) table statistics results. Significant P-value <0.05. d.f = degrees of freedom

Source of variation	d.f	Sum of Squares	Variance components	Percentage of variation (%)	P-value
Among sites	3	158.035	0.15749	0.59	0.01369
Among groups within sites	4	99.632	-0.01632	-0.06	0.84653
Within groups	698	18388.227	26.34416	99.47	<0.01
Total	705	18645.894			

4.4.4 Kinship analysis

Between 5 to 15% of parent-offspring relations were found across sites (Table 7). We only found 5% of assigned offspring to parents in Ringwood, and one single parent was linked to seven juveniles, ranging from 3 to 570.30 metres apart. Stourhead2 showed a high number of links (14%) and the maximum distance between a parent and offspring (718.49 m). In all four sites, the minimum assigned kinship distance was 3 m which reveals that the juveniles sampled in the surroundings of the adults may be its offspring while other seedlings travel further as the average assigned parent-offspring distance was 181.175 m.

Table 7. Kinship analysis results, including the number of adults and juveniles included in the analysis (A-J), the number of SNPs used per site (N SNPs), the percentage of parent-offspring related found (Average %), the mean, maximum and minimum distance these relationships were found and the maximum number of offspring related to one parent found (Max offspring from one parent). N: number, Max: maximum, Min: minimum.

	A - J	N SNPs	Assigned (%)	Average distance (m)	Max distance (m)	Min distance (m)	Max offspring from one parent
Stourhead2	49-133	56	14	191.47	718.49	3	5
Longleat2	49-149	43	9	220.15	670.41	3	6
Mortimer	48-149	57	15	126.10	242.01	3	4
Ringwood	50-150	40	5	186.98	570.30	3	7

4.5 Discussion

In this Chapter we investigated genetic diversity among *P. menziesii* adults and offspring across four UK woodlands transitioning to CCF. We found different levels of transformation to CCF across sites and observed higher levels of species diversity in the sites with a more diverse structure. Our population genetics findings varied based on the different genetic diversity indicators. The overall heterozygosity levels were high across every site and no significant differences were found between the planted adults and their naturally regenerated offspring. However, we found high variability in N_e both among and within sites and the N_e levels were generally considered low to ensure the maintenance of long-term genetic diversity. Here we discussed how these findings may inform current forestry management practices in CCF woodlands and contribute to the discussion on sustainable genetic conservation strategies.

4.5.1 Sites biodiversity and levels of transformation

Our four *P. menziesii* sites were at different stages of transformation toward CCF according to the evolution of structural diversity proposed by Pommerening & Murphy (2004) (see Figure 1 and Figure 4). Various levels of species diversity estimated as Species evenness (S), effective number of species and The Shannon index (*H'*) were found across our sites, and higher levels were found in more structurally complex sites (Longleat2 and Stourhead2), which is consistent with other studies (Bravo-Oviedo et al., 2014; Gärtner & Reif, 2004; Kerr et al., 2010, 2017). In our study, three of the sites had $1 < H' < 2$, and only Mortimer had substantially lower canopy species diversity (H'=0.17). Pommerening, (2002) reported the Shannon index value ($H' = 0.42$) in a young plantation (24 years old) composed of *P. menziesii* in transformation to CCF in Germany. These results contrast with the lower Shannon index found in our youngest stand (40 years old) Mortimer ($H' = 0.17$), suggesting that factors other than age which may contribute to the number of species in a stand, such as different silvicultural practises or the diversity of species surrounding the stand in transformation. A study of two *P. menziesii* plantations in its natural range (Vancouver Island, BC) showed species diversity ranging from $H' = 0.77$ to 1.01, which was estimated before commercial thinning (El-Kassaby & Benowicz, 2000). This revealed how different plantations established from a single species, even

those in the species' natural distribution may have different levels of species diversity which may affect the future of the woodland if the aim is to transition to CCF.

A study of three forest types dominated by pine in northern Europe reported a higher Shannon index ($H' = 2.9$) including sites classified as near-natural, selected logged and managed (De Quesada & Kuuluvainen, 2020), which is higher than observed in our four sites ($H' < 2$). A reverse-j profile was reported in the near-natural pine woodlands, which is similar to our observations at Stourhead2, one of the closest stands to complete CCF transformation (uneven-aged woodland). However, we found substantially fewer trees in Stourhead2 smallest diameter class, which makes the profile incomplete making possible to link it to earlier stages of transformation. The managed pine forest had a rotated sigmoidal curve (De Quesada & Kuuluvainen, 2020) similar to Longleat2, although its largest diameter class is only 50 cm diameter (De Quesada & Kuuluvainen, 2020) while in Longleat2 we found trees over 100 cm in diameter, which indicates a more advanced transition to CCF.

Overall, Longleat2 had the highest diversity among mature trees and also a diameter curve that indicates a transition to complete CCF. In contrast, Stourhead2 had both a high mature tree diversity and the highest offspring diversity ($H' > 1.5$), which may ensure a high future mature tree diversity composition in line with the objectives of CCF. Different levels of structure and species diversity were found across the study sites and compared to other woodlands and plantations, this revealed the complexity of each forest and the needs and requirements to transition into a close-to-nature forest may differ site by site.

4.5.2 Population genetics analysis

The genetic diversity levels among and within sites varied depending on the genetic diversity indicator estimated. The heterozygosity levels were high across sites ($H_o > 0.44$ and $H_e > 0.48$) but other studies reported even higher levels of heterozygosity ($H_o > 0.60$ and $H_e > 0.88$) in *P. menziesii* stands (van Loo et al., 2019) and seed orchards ($H_o > 0.64$ and $H_e > 0.86$) (Pakull et al., 2021) planted in Europe as estimated using microsatellites markers. Similarly, in central Europe Neophytou et al. (2019) also used microsatellite markers to compare the genetic diversity between the adults ($H_o >$

0.703 and H_e > 0.905) planted and their offspring (H_o > 0.714 and H_e > 0.885). These authors reported a significant reduction in the offspring H_e, while the H_o remained constant, which contrasts with our results where no significant differences were found between adults and natural regeneration. The heterozygosity results from genotyping using microsatellites differ from SNP markers; Tereba & Konecka (2021) found higher levels of heterozygosity (H_o > 0.74 and H_e > 0.85) when using microsatellites compared to SNPs (H_o > 0.34 and H_e > 0.37) in two *Pinus sylvestris* populations. Their findings indicate that our sites may have heterozygosity levels similar to those reported in other parts of Europe, despite the relative differences in estimates.

The N_e varied across and within our sites, and ranged from 112.5 to 428.6 in adults and from 154.2 to 302.7 in juveniles. None of the sites had a N_e > 500, which has been recently accepted as a genetic diversity indicator to monitor genetic conservation by the Convention on Biological Diversity (CBD) (CBD, 2022; O'Brien et al., 2022). We found substantially lower N_e in Ringwood juveniles (271.9) compared to the adults (428.6), which indicated that the outcrossing in the planted trees is limited to a low number of individuals. This observation suggests inefficient transfer of genetic diversity to the offspring and an increased risk of inbreeding and genetic drift. Wojacki et al. (2019) estimated different genetic diversity estimators on different *P. menziesii* stand sizes and found that N_e increased in larger stand sizes; however, we found that the sites with the largest census size (N_c), i.e. Stourhead2 (1,707.6) and Mortimer (1,413.8), had the lowest N_e, 271.9 and 112.5, respectively. Studies estimating N_e in planted *P. menziesii* stands are scarce; however, Pakull et al., (2021) reported that *P. menziesii* seed orchards in Europe had a N_e < 41, but over the N_e minimum threshold of 10 recommended in seed orchards (Stoehr, 2004).

The N_e / N_c is the effective-to-census size ratio and was proposed as a useful indicator for inferring N_e from N_c data for conservation purposes (Frankham, 1995; Luikart et al., 2010), allowing to monitor genetic diversity when genetic data is not available (Hoban et al., 2022). Hoban et al., (2020) reviewed studies that estimated N_e / N_c and they found that values ranged from 0.07 to 0.855 in plant species. The ratio calculated in our studies varied from 0.08 to 0.60. Only Mortimer and Stourhead were close to 0.1, which is the N_e / N_c value recommended to infer N_e across species. This

indicates the importance of further investigating N_e / N_c in different species to allow more accurate guidance and estimation of N_e for conservation practices.

Overall, the low N_e found across all sites may compromise the future genetic variation of the woodlands and adaptability to environmental changes. Some of the sites (Stourhead2) have a planting schedule that may counteract these effects although it would require monitoring to establish its potential. Our data on genetic diversity including both heterozygosity and N_e estimators may suggest that the latter is a more informative indicator to monitor genetic diversity in the short-term. N_e is predictive of evolutionary change in a population that may result from genetic drift (Waples & Do, 2010) and it may decrease more rapidly between generations than heterozygosity (H_e). Overall, N_e may be a more suitable indicator than H_e to inform on the genetic adaptability of populations faced with rapid environmental change (Hoban et al., 2023).

Variable results have been reported from studies of planted *P. menziesii* in Europe compared to its natural range, from no significant difference in genetic diversity (van Loo et al., 2019) to a decrease in allelic richness and H_e (Eckhart et al., 2017; Neophytou et al., 2019). The disparities could be due to differences in the seed source collection method, the region of origin or gene flow from individuals in the plantation surroundings. The material used in our study sites is of unknown provenance and other *P. menziesii* stands are found in their surroundings which may affect our results in the natural regeneration. Determining the seed source of our sites by comparative analysis to populations in the natural range may shed insights into the patterns found in this study.

The low level of parent-offspring assignments found across sites (5 - 15%) suggests that either the parents were present within the stands but not sampled or that neighbouring stands are contributing to the reproductive process. Van Loo et al., (2019) observed an overall higher frequency of parent-offspring relations (26 - 50%) in *P. menziesii* stands in central Europe and a maximum seed/pollen dispersion of 92 m. In contrast, we estimated 718 m as the maximum dispersal distance (Stourhead2) and the average distance of all the relationships found across the study sites was 181 m. Other studies have shown shorter pollination distances, usually around 50 m or less (Pakull et al., 2022; Wojacki et al., 2019). Our results indicate that neighbour stands, even at a 700 m distance, may

be contributing to the gene flow, further research looking at *P. menziesii* woodlands surrounding our sites may shed light on the reproductive process and genetic makeup of the study sites and its effects in the genetic diversity of our stands.

4.6 Conclusion

Understanding the genetic makeup and the extent of genetic diversity of the planted trees and its transfer to the offspring is vital for predicting how these woodlands may face current and future climate change disruptions. Although the levels of heterozygosity remained high across sites, the four study sites had a $N_e < 500$, indicating low levels of genetic diversity, which may compromise the adaptation potential to rapid environmental change. We also observed a substantially lower N_e in juveniles compared to adults at one site (Ringwood), indicating a reduced number of individuals contributed to the gene flow between generations, which could result in genetic drift and a loss of genetic diversity. Assisted gene flow in these sites may be recommendable to avoid further effects of low genetic diversity levels such as inbreeding depression and genetic drift. The lowest N_e was found in Mortimer, a site in the earliest stage of transformation into a CCF with low structural and species diversity. Low levels of N_e in both adults and juveniles were also found at Stourhead2 and Longleat2, which had high levels of species diversity and stand structure. These results revealed the importance of understanding and monitoring the genetic makeup of stands transformed or in transformation to close-to-nature woodlands due to the potential lack of genetic diversity which may compromise the stand's adaptability. This knowledge may be crucial to assess the need for genetic diversification as part of a resilience strategy.

4.7 Acknowledgements

The authors acknowledge the forestry estates where the samples came from: Stourhead (Western), Longleat (Wiltshire), Mortimer and Ringwood Forest. We also thank Glenn Howe for providing advice and key information on *P. menziesii* genetic markers (SNPs) and Barley Rose Collier Harris for helping with sampling. We want to thank Will Hoare, George MacKay and Emily Farley for

their assistance in both laboratory and field work. LG received financial support from the Oxford-John Oldacre Graduate Scholarship and from Mr Henry Hoare. The project was funded in part by the OxLEP fund for the Oxford Plant Sciences Centre for Innovation.

4.8 References

Bennett, G., Hardy, A., Bunting, P., Morgan, P., & Fricker, A. (2020). A Transferable and Effective Method for Monitoring Continuous Cover Forestry at the Individual Tree Level Using UAVs. Remote Sensing 2020, Vol. 12, Page 2115, 12(13), 2115. https://doi.org/10.3390/rs12132115

Bravo-Oviedo, A., Pretzsch, H., Ammer, C., Andenmatten, E., Barbati, A., Barreiro, S., Brang, P., Bravo, F., Coll, L., Corona, P., Den Ouden, J., Ducey, M. J., Forrester, D. I., Giergiczny, M., Jacobsen, J. B., Lesinski, J., Löf, M., Mason, B., Matovic, B., … Zlatanov, T. (2014). European mixed forests: Definition and research perspectives. Forest Systems, 23(3), 518–533. https://doi.org/10.5424/fs/2014233-06256

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10, 421. https://doi.org/10.1186/1471-2105-10-421

Capblancq, T., Butnor, J. R., Deyoung, S., Thibault, E., Munson, H., Nelson, D. M., Fitzpatrick, M. C., & Keller, S. R. (2020). Whole‐exome sequencing reveals a long‐term decline in effective population size of red spruce (*Picea rubens*). Evolutionary Applications, eva.12985. https://doi.org/10.1111/eva.12985

CBD. (2022). Decision adopted by the conference of the parties to the Convention on Biological Diversity. 15/15 Monitoring framework for the Kunming-Montreal Global Biodiversity Framework. In Convention of Biological Diversity. https://www.cbd.int/doc/decisions/cop-15/cop-15-dec-05-en.pdf

Choi, S. R., Oh, S. H., Dhandapani, V., Jang, C. S., Ahn, C. H., Rameneni, J. J., Kim, H., Jeon, I., & Lim, Y. P. (2020). Development of SNP markers for marker-assisted breeding in Chinese cabbage using Fluidigm genotyping assays. Horticulture Environment and Biotechnology, 61(2), 327–338. https://doi.org/10.1007/S13580-019-00211-Y

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics, 12(7), Article 7. https://doi.org/10.1038/nrg3012

De Quesada, G., & Kuuluvainen, T. (2020). Tree diametric-species diversity is affected by human impact in old Scots pine dominated forest in boreal Fennoscandia. For. Ecosyst. 7, 8. https://doi.org/10.1186/s40663-020-0219-6

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources, 14(1), 209–214. https://doi.org/10.1111/1755-0998.12157

Eckhart, T., Walcher, S., Hasenauer, H., & van Loo, M. (2017). Genetic diversity and adaptive traits of European versus American Douglas-fir seedlings. European Journal of Forest Research, 136(5–6), 811–825. https://doi.org/10.1007/s10342-017-1072-1

El-Kassaby, Y. A., & Benowicz, A. (2000). Effects of commercial thinning on genetic, plant species and structural diversity in second growth Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) stands. Forest Genetics, 7(3), 193–203.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, 10(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Ferrell, W. K., & Woodard, E. S. (1966). Effects of Seed Origin on Drought Resistance of Douglas‐Fir (*Pseudotsuga Menziesii*) (Mirb.) Ecology Wiley Online Library. 47(3), 499–503. https://doi.org/10.2307/1932994

Finkeldey, R. (2002). Reproduction in Continuous Cover Forests—The Geneticist's Perspective. In Continuous Cover Forestry (pp. 67–79). https://doi.org/10.1007/978-94-015-9886-6_6

Forestry Commission. (2019). National Forest Estate Subcompartments England. https://data-forestry.opendata.arcgis.com/datasets/3993555ec8124b1e91b55a4a8b84567c_0/explore?location=52.881100%2C-1.770299%2C8.24

Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: A review. Genetics Research, 66(2), 95–107. https://doi.org/10.1017/S0016672300034455

Gärtner, S., & Reif, A. (2004). The impact of forest transformation on stand structure and ground vegetation in the southern Black Forest, Germany. Plant and Soil, 264(1–2), 35–51.

https://doi.org/10.1023/B:PLSO.0000047751.25915.77

Godbout, J., Tremblay, L., Levasseur, C., Lavigne, P., Rainville, A., Mackay, J., Bousquet, J., & Isabel, N. (2017). Development of a traceability system based on a SNP array for large-scale production of high-value white spruce (Picea glauca). Frontiers in Plant Science, 8(July), 1–13. https://doi.org/10.3389/fpls.2017.01264

Helliwell, R., & Wilson, E. R. (2012). Continuous Cover Forestry in Britain. Quarterly Journal of Forestry.

Hoban, S., Archer, F. I., Bertola, L. D., Bragg, J. G., Breed, M. F., Bruford, M. W., Coleman, M. A., Ekblom, R., Funk, W. C., Grueber, C. E., Hand, B. K., Jaffé, R., Jensen, E., Johnson, J. S., Kershaw, F., Liggins, L., MacDonald, A. J., Mergeay, J., Miller, J. M., … Hunter, M. E. (2022). Global genetic diversity status and trends: Towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. Biological Reviews, 97(4), 1511–1538. https://doi.org/10.1111/brv.12852

Hoban, S., Bruford, M., D'Urban Jackson, J., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., Paz-Vinas, I., Sjögren-Gulve, P., Segelbacher, G., Vernesi, C., Aitken, S. N., Bertola, L. D., Bloomer, P., Breed, M., Rodríguez-Correa, H., Funk, W. C., Grueber, C. E., Hunter, M. E., Jaffe, R., Liggins , L., Mergeay , J., Moharrek , F., O'Brien , D., Ogden , R., Palma-Silva , C., Pierson , J., Ramakrishnan , U., Simo-Droissart , M., Tani , N., Waits , L. & Laikre, L. (2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biological Conservation, 248, 108654. https://doi.org/10.1016/j.biocon.2020.108654

Holmes, G. D. (1975). History of Forestry and Forest Management. Philosophical Transactions of the Royal Society of London, 271(911), 69–80.

Howe, G. T., Jayawickrama, K., Kolpak, S. E., Kling, J., Trappe, M., Hipkins, V., Ye, T., Guida, S., Cronn, R., Cushman, S. A., & McEvoy, S. (2020). An Axiom SNP genotyping array for Douglas-fir. BMC Genomics, 21(1), 9. https://doi.org/10.1186/s12864-019-6383-9

Howe, G. T., Yu, J., Knaus, B., Cronn, R., Kolpak, S., Dolan, P., Lorenz, W. W., & Dean, J. F. D. (2013). A SNP resource for Douglas-fir: De novo transcriptome assembly and SNP detection

and validation. BMC Genomics, 14(1). https://doi.org/10.1186/1471-2164-14-137

Jones, A. T., Ovenden, J. R., & Wang, Y. G. (2016). Improved confidence intervals for the linkage disequilibrium method for estimating effective population size. Heredity, 117(4), 217–223. https://doi.org/10.1038/hdy.2016.19

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Molecular Ecology, 16(5), 1099–1106. https://doi.org/10.1111/J.1365-294X.2007.03089.X

Kans, J. (2023). Entrez Direct: E-utilities on the Unix Command Line. In Entrez Programming Utilities Help [Internet]. National Center for Biotechnology Information (US). https://www.ncbi.nlm.nih.gov/books/NBK179288/

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution, 4(8), 782–788. https://doi.org/10.1111/2041-210X.12067

Kerr, G., Morgan, G., Blyth, J., & Stokes, V. (2010). Transformation from even-aged plantations to an irregular forest: The world's longest running trial area at Glentress, Scotland. Forestry, 83(3), 329–344. https://doi.org/10.1093/forestry/cpq015

Kerr, G., Snellgrove, M., Hale, S., & Stokes, V. (2017). The Bradford-Hutt system for transforming young even-aged stands to continuous cover management. Forestry, 90(4), 581–593. https://doi.org/10.1093/forestry/cpx009

Kremer, A., Potts, B. M., & Delzon, S. (2014). Genetic divergence in forest trees: Understanding the consequences of climate change. Functional Ecology, 28(1), 22–36. https://doi.org/10.1111/1365-2435.12169

Kremer, A., Ronce, O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., Bridle, J. R., Gomulkiewicz, R., Klein, E. K., Ritland, K., Kuparinen, A., Gerber, S., & Schueler, S. (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecology Letters, 15(4), 378–392. https://doi.org/10.1111/j.1461-0248.2012.01746.x

Krutovsky, K. V., & Neale, D. B. (2005). Nucleotide diversity and linkage disequilibrium in

cold-hardiness- and wood quality-related candidate genes in douglas fir. Genetics, 171(4), 2029–2041. https://doi.org/10.1534/genetics.105.044420

Kuhnlein, H. V., & Turner, N. J. (1991). Traditional Plant Foods of Canadian Indigenous Peoples: Nutrition, Botany, and Use. Taylor & Francis.

Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., & Allendorf, F. W. (2010). Estimation of census and effective population sizes: The increasing usefulness of DNA-based approaches. Conservation Genetics, 11(2), 355–373. https://doi.org/10.1007/s10592-010-0050-7

Mason, B., & Kerr, G. (2000). Transforming Even-aged Conifer Stands to Continuous Cover Management. 40, 1–8.

McDadi, O., & Hebda, R. J. (2008). Change in historic fire disturbance in a Garry oak (*Quercus garryana*) meadow and Douglas-fir (*Pseudotsuga menziesii*) mosaic, University of Victoria, British Columbia, Canada: A possible link with First Nations and Europeans. Forest Ecology and Management, 256(10), 1704–1710. https://doi.org/10.1016/j.foreco.2008.03.012

Mijangos, J. L., Gruber, B., Berry, O., Pacioni, C., & Georges, A. (2022). dartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. Methods in Ecology and Evolution, 13(10), 2150–2158. https://doi.org/10.1111/2041-210X.13918

Moore, J. C. (2013). Diversity, Taxonomic versus Functional. In S. A. Levin (Ed.), Encyclopedia of Biodiversity (Second Edition) pp. 648–656. Academic Press. https://doi.org/10.1016/B978-0-12-384719-5.00036-8

Müller, T., Ensminger, I., & Schmid, K. J. (2012a). A catalogue of putative unique transcripts from Douglas-fir (*Pseudotsuga menziesii*) based on 454 transcriptome sequencing of genetically diverse, drought stressed seedlings. BMC Genomics, 13(1). https://doi.org/10.1186/1471-2164-13-673

Müller, T., Ensminger, I., & Schmid, K. J. (2012b). Assembly , annotation and SNP detection using twelve Douglas-fir cDNA libraries. SMBE Meeting Dublin 2012 June, 3–4.

Neale, D. B., McGuire, P. E., Wheeler, N. C., Stevens, K. A., Crepeau, M. W., Cardeno, C., Zimin, A. V., Puiu, D., Pertea, G. M., Sezen, U. U., Casola, C., Koralewski, T. E., Paul, R., Gonzalez-Ibeas, D., Zaman, S., Cronn, R., Yandell, M., Holt, C., Langley, C. H., … Wegrzyn,

J. L. (2017). The Douglas-Fir Genome Sequence Reveals Specialization of the Photosynthetic Apparatus in Pinaceae. G3 Genes|Genomes|Genetics, 7(9), 3157–3167. https://doi.org/10.1534/g3.117.300078

Neophytou, C., van Loo, M., & Hasenauer, H. (2019). Genetic diversity in introduced Douglas-fir and its natural regeneration in Central Europe. Forestry: An International Journal of Forest Research, 1–10. https://doi.org/10.1093/forestry/cpz055

Northwest Natural Resource Group. (2014). Forest Inventory and Monitoring Guidelines. Northwest Natural Resource Group.

O'Brien, D., Laikre, L., Hoban, S., Bruford, M. W., Ekblom, R., Fischer, M. C., Hall, J., Hvilsom, C., Hollingsworth, P. M., Kershaw, F., Mittan, C. S., Mukassabi, T. A., Ogden, R., Segelbacher, G., Shaw, R. E., Vernesi, C., & MacDonald, A. J. (2022). Bringing together approaches to reporting on within species genetic diversity. Journal of Applied Ecology, 59(9), 2227–2233. https://doi.org/10.1111/1365-2664.14225

Pakull, B., Eusemann, P., Wojacki, J., Ahnert, D., & Liesebach, H. (2021). Genetic diversity of seeds from four German Douglas fir (*Pseudotsuga menziesii*) seed orchards. European Journal of Forest Research, 140(6), 1543–1557. https://doi.org/10.1007/s10342-021-01419-3

Pakull, B., Wojacki, J., Eusemann, · Pascal, Fussi, B., Ahnert, · Diana, & Liesebach, H. (2022). Sexual reproduction in two mixed stands of coastal and interior Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) in Germany. European Journal of Forest Research, 3. https://doi.org/10.1007/s10342-022-01514-z

Peters, R. L. (1990). Effects of global warming on forests. Forest Ecology and Management, 35(1–2), 13–33. https://doi.org/10.1016/0378-1127(90)90229-5

Pommerening, A. (2002). Approaches to quantifying forest structures. Forestry: An International Journal of Forest Research, 75(3), 305–324. https://doi.org/10.1093/forestry/75.3.305

Pommerening, A., & Murphy, S. T. (2004). A review of the history, definitions and methods of continuous cover forestry with special attention to afforestation and restocking. Forestry, 77(1), 27–44. https://doi.org/10.1093/forestry/77.1.27

QGIS. (2023). QGIS Geographic Information System. Open Source Geospatial Foundation Project.

(3.32.3) [Computer software]. http://qgis.org

R Core Team. (2018). R: A language and environment for statistical computing (4.3.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Raum, S., & Potter, C. (2015). Forestry paradigms and policy change: The evolution of forestry policy in Britain in relation to the ecosystem approach. Land Use Policy, 49, 462–470. https://doi.org/10.1016/j.landusepol.2015.08.021

Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Menzel, A., Root, T. L., Estrella, N., Seguin, B., Tryjanowski, P., Liu, C., Rawlins, S., & Imeson, A. (2008). Attributing physical and biological impacts to anthropogenic climate change. Nature, 453(7193), 353–357. https://doi.org/10.1038/nature06937

RStudio Team. (2020). RStudio: Integrated development for R (2023.06.0) [Computer software]. Posit. http://www.rstudio.com/

Slavov, G. T., DiFazio, S. P., Martin, J., Schackwitz, W., Muchero, W., Rodgers-Melnick, E., Lipphardt, M. F., Pennacchio, C. P., Hellsten, U., Pennacchio, L. A., Gunter, L. E., Ranjan, P., Vining, K., Pomraning, K. R., Wilhelm, L. J., Pellegrini, M., Mockler, T. C., Freitag, M., Geraldes, A., … Tuskan, G. A. (2012). Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. The New Phytologist, 196(3), 713–725. https://doi.org/10.1111/j.1469-8137.2012.04258.x

Standard Biotools. (2015a). Genotyping with the 192.24 IFC Using SNP Type Assays. PN 100-3913 C1. https://fluidigm.my.salesforce.com/sfc/p/#700000009DAw/a/4u0000019jzg/UO9mKHykLnoutJDRGZUzhqrEEXCWtEMtXX84mNlCiYQ

Standard Biotools. (2015b). Genotyping with the Flex Six IFC Using SNP Type Assays. PN 100-7718 C1. https://fluidigm.my.salesforce.com/sfc/p/#700000009DAw/a/4u0000019jvZ/e7Cfyl_yhhUodOMoT_gmPYJatwDM._GkDhjjNetzMMM

Stoehr, M. (2004). Protocol for rating seed orchard seedlots in British Columbia: Quantifying genetic gain anddiversity. Forestry, 77(4), 297–303. https://doi.org/10.1093/forestry/77.4.297

Taylor, S. A., Larson, E. L., & Harrison, R. G. (2015). Hybrid zones: Windows on climate change. Trends in Ecology and Evolution, 30(7), 398–406. https://doi.org/10.1016/j.tree.2015.04.010

Tereba, A., & Konecka, A. (2021). Comparison of Microsatellites and SNP Markers in Genetic Diversity Level of Two Scots Pine Stands. 4. https://doi.org/10.3390/iecf2020-07776

Tew, E. R., Ambrose-Oji, B., Beatty, M., Büntgen, U., Butterworth, H., Clover, G., Cook, D., Dauksta, D., Day, W., Deakin, J., Field, A., Gardiner, B., Harrop, P., Healey, J. R., Heaton, R., Hemery, G., Hill, L., Hughes, O., Khaira-Creswell, P. K., … Sutherland, W. J. (2023). A horizon scan of issues affecting UK forest management within 50 years. Forestry: An International Journal of Forest Research, cpad047. https://doi.org/10.1093/forestry/cpad047

Trembizki, E., Smith, H., Lahra, M. M., Chen, M., Donovan, B., Fairley, C. K., Guy, R., Kaldor, J., Regan, D., Ward, J., Nissen, M. D., Sloots, T. P., & Whiley, D. M. (2014). High-throughput informative single nucleotide polymorphism-based typing of *Neisseria gonorrhoeae* using the Sequenom MassARRAY iPLEX platform. Journal of Antimicrobial Chemotherapy, 69(6), 1526–1532. https://doi.org/10.1093/JAC/DKT544

van Loo, M., Hintsteiner, W., Pötzelsberger, E., Schüler, S., & Hasenauer, H. (2015). Intervarietal and intravarietal genetic structure in Douglas-fir: Nuclear SSRs bring novel insights into past population demographic processes, phylogeography, and intervarietal hybridization. Ecology and Evolution, 5(9), 1802–1817. https://doi.org/10.1002/ece3.1435

van Loo, M., Lazic, D., Chakraborty, D., Hasenauer, H., & Schüler, S. (2019). North American Douglas-fir (*P. menziesii*) in Europe: Establishment and reproduction within new geographic space without consequences for its genetic diversity. Biological Invasions, 21(11), 3249–3267. https://doi.org/10.1007/s10530-019-02045-2

Vidal, M., Plomion, C., Raffin, A., Harvengt, L., & Bouffier, L. (2017). Forward selection in a maritime pine polycross progeny trial using pedigree reconstruction. Annals of Forest Science, 74(1). https://doi.org/10.1007/s13595-016-0596-8

Wojacki, J., Eusemann, P., Ahnert, D., Pakull, B., & Liesebach, H. (2019). Genetic diversity in seeds produced in artificial Douglas-fir (Pseudotsuga menziesii) stands of different size. Forest Ecology and Management, 438(February), 18–24.

https://doi.org/10.1016/j.foreco.2019.02.012

Youngentob, K. N., Zdenek, C., & van Gorsel, E. (2016). A simple and effective method to collect leaves and seeds from tall trees. Methods in Ecology and Evolution, 7(9), 1119–1123. https://doi.org/10.1111/2041-210X.12554

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	High Heterozygosity but low levels of Effective Population Size in UK Continuous Cover Forestry planted stands of *Pseudotsuga menziesii*
Publication Status	□Published □ Accepted for Publication □Submitted for Publication ☑Unpublished and unsubmitted work written in a manuscript style
Publication Details	Guillardín, L.*, Glover, E., Kerr, G., MacKay, J.J.* (unplublished) High Heterozygosity but low levels of Effective Population Size in UK Continuous Cover Forestry planted stands of *Pseudotsuga menziesii*.

Student Confirmation

Student Name:	Laura Guillardin Calvo	
Contribution to the Paper	Design the study, conceived and planned the experiments, coordinated sampling and laboratory experiments. Performed bioinformatic and statistical analyses and prepared the figures. Drafted and revised the manuscript.	
Signature 	Date	15/11/2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: John MacKay, Wood Professor of Forest Sciences		
Supervisor comments I agree with the above comments		
Signature 	Date	05/01//2024

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 5. General Discussion

Changing environmental conditions and extreme weather disruptions are affecting ecosystems worldwide (Rosenzweig et al., 2008) including forests and woodlands (Aitken et al., 2008). Continuous Cover Forestry (CCF) has been applied in many woodlands in the UK (Pommerening et al., 2021) and several studies have developed systems to implement it through silvicultural treatments (Kerr et al., 2017). It has become a popular approach to enhance resilience in woodlands by increasing the number of species and diversifying its structure (Pommerening & Murphy, 2004). Our study is the first to assess the genetic diversity of these CCF woodlands in the UK, emphasising the differences between genetic diversity in planted trees and their offspring, which is linked to the potential adaptation and long-term viability of these ecosystems.

In this chapter, we will discuss the major findings of this thesis and how they may inform forest managers to enhance woodland resilience. We will discuss the implications and future research on sustainable practices, regarding both, laboratory plastic reduction and the increase of genetic diversity within CCF woodlands.

5.1 An effective DNA isolation system that reduces carbon footprint

Genetic and genomics experiments require the isolation of nucleic acids (RNA or DNA), which can be challenging depending on the study species (Rezadoost et al., 2016). Lately, the number of samples studied per project is increasing due to lower sequencing and genotyping costs per sample (Fumagalli, 2013) and to increasing statistical power (Hong & Park, 2012). Traditional DNA isolation methods use large amounts of single-use plastic and lengthy hands-on procedures, therefore, finding an efficient DNA isolation method for each study species that recovers high-quality DNA while reducing the time and plastic needed is of great interest. This thesis presents a comparison between a silica column-based method, DNeasy plant mini kit (Qiagen USA, Valencia, CA) (QIAGEN), widely used to isolate DNA from conifer trees and a temperature-controlled enzymatic cocktail method, Plant DNA Extraction (MicroGEM International PLC, 2019) (MicroGEM).

One of the main challenges of this thesis was to deal with very large sample sizes in both species, so finding a DNA isolation method to reduce both time and single-use plastic while recovering high-quality DNA was crucial to streamline the experimental work. Overall, we found that both systems recovered sufficient and reliable DNA to use on both species, in addition to *T. heterophylla,* although MicroGEM showed better suitability to *T. plicata* due to the higher yield of DNA. However, MicroGEM had the advantage of using half the amount of plastic and time than QIAGEN, indicating that this DNA isolation system was an option to reduce the CO_2 footprint of the project while decreasing the time needed to process the samples.

Some molecular techniques, such as the preparation of libraries to perform sequencing experiments have been developed to reduce processing time (Torkamaneh et al., 2020) and some laboratories have estimated their plastic use (Alves et al., 2021) to decrease their plastic footprint; however, to date, there is a lack of studies estimating time and plastic used in DNA isolation protocols. Currently, the optimisation of DNA isolation methods in forest species is primarily aimed at overcoming the difficulties encountered when isolating DNA from tree species such as the presence of contaminants (Rezadoost et al., 2016), to enhance yield and DNA quality (Green & Sambrook, 2017) and to reduce time (FU et al., 2017). Marengo et al., (2019) developed a one-step DNA isolation method that used less single-use plastic although it was not quantified and the method was not tested in tree species. Our study presents original findings regarding the estimation of time and plastic use linked to DNA isolation experiments in forest species and will encourage researchers to seek low impact DNA isolation protocols to increase sustainable laboratory practices.

5.2 Woodlands potential adaptation to environmental changes under CCF

The potential adaptation of species to changing climatic conditions depends on the standing genetic diversity carried within the populations. Species within woodlands managed under CCF may experience a decrease in genetic diversity from generation to generation due to a reduction of N_e caused by selective harvesting. This thesis presents a first comparison of the genetic diversity between

generations of two conifer species found in CCF woodlands. The aim is to narrow the existing knowledge gap around standing genetic diversity and provide useful information to design effective forestry management strategies.

5.2.1 High levels of heterozygosity and monomorphic SNPs found in *T. plicata* stands

T. plicata is known to have relatively high selfing rates, and, although Shalev et al., (2023) found low levels of inbreeding depression in growth traits, Shalev et al., (2022) observed low genetic diversity levels across its natural range. Therefore, investigating genetic diversity levels of planted trees and comparing them to their offspring is crucial for understanding the potential adaptability of *T. plicata* to climate change in a CCF context. This thesis presented the first population genetics study on *T. plicata* woodlands managed under CCF in the UK. In this project, we designed a SNP genotyping assay based on data from a Genotyping by Sequencing (GBS) step to discover high quality and reliable SNPs.

We found low levels of heterozygosity in Longleat1 (H_o = 0.275), while in the other sites (Bagley Wood, Stourhead1 and Stourhead2) it was relatively high (average H_o = 0.381) which contrasts with the low levels of heterozygosity (H_o = 0.219) within the species natural distribution reported by Shalev et al., (2022). Only Stourhead1 juveniles had a $N_e \approx 100$, considered a threshold for short-term genetic diversity maintenance (Frankham et al., 2014) while all the other sites had even lower N_e indicating a poor potential adaptability to face environmental changes (Hoban et al., 2020). Additionally, we found a high number of monomorphic SNPs private to Longleat1 adults (30.3%), which may indicate the use of different genetic sources to establish the plantation compared to the other stands. Furthermore, these monomorphic sites could be explained by the fixation of alleles due to selection to specific conditions in the provenance of origin of the material used.

5.2.2 Low levels of N_e in *P. menziesii* stands but high heterozygosity

P. menziesii is one of the most important forest species regarding its commercial potential both, within and outside its natural distribution (Hermann & Lavender, 1999; Isaac-Renton et al., 2014). St Clair et al., (2005) reported its partial resistance to summer droughts suggesting that *P. menziesii* may be a potential candidate to resist some climate change impacts and it may improve growth of broadleaf species when planted in mixed stands (Thomas et al., 2022). Hence, *P. menziesii* may play an important role in CCF woodlands; however, studies investigating its suitability from a genetic perspective are lacking. Additionally, CCF prioritises natural regeneration as the succession method (Mason et al., 1999), but selective harvesting may affect the reproduction process in these woodlands (Finkeldey, 2002). The reduction of candidate parents may decrease the genetic diversity of the offspring, so the potential adaptability to overcome climate change impacts may decline (Finkeldey, 2002). Therefore reporting the genetic diversity levels of planted *P. menziesii* CCF stands and comparing it to its offspring is crucial to understanding the effect of logging in CCF woodlands. This thesis presents the first population genetics study investigating *P. menziesii* CCF woodlands.

We found high levels of heterozygosity across and within sites (average $H_o = 0.447$) and no differences between adults and juveniles. These high genetic diversity results follow other reports in planted *P. menziesii* stands in Europe (Pakull et al., 2021; van Loo et al., 2019); however, a significant reduction in H_e juveniles was reported by Neophytou et al. (2019), potentially suggesting site to site variation. In contrast, we found different N_e across and within sites, indicating variability of genetic diversity regarding this estimator. In fact, none of the study sites had a $N_e > 500$, selected threshold to maintain genetic diversity in the long-term (CBD, 2022; O'Brien et al., 2022). Furthermore, Ringwood adults had the highest N_e (428.6) but the juveniles had substantially lower N_e (271.9), suggesting a potential poor transfer of genetic diversity to the offspring.

5.2.3 Diversity in species, structure and genes

CCF aims to enhance the biodiversity of woodlands by increasing the number of species, promoting different ages and heights, while providing several forest services (Mason et al., 1999).

Although several methods have been proposed to categorise CCF levels of transformation (e.g. Bennett et al., 2020; Kerr et al., 2003, 2010), the complexity and uniqueness of woodlands may require other estimators for its assessment. This study presents a detailed characterisation of the study stands to evaluate their level of transformation to CCF. Furthermore, this thesis performed the first combined study of diversity in species and structure while looking at each CCF stand's genetic diversity focusing on two conifer species.

A trend between species and structure diversity was found in all the study stands, including those composed primarily of *T. plicata* (Chapter 3) and *P. menziesii* (Chapter 4). Stourhead2 stand was included in both Chapters, as both species were found in the area, and was described to be at an advanced stage of transformation to CCF. However, the N_e estimated in both species (*P. menziesii* adults = 271.9 and juveniles = 259.7, *T. plicata* adults = 9.7 and juveniles = 7.4) indicated low levels of genetic diversity. This may suggest a lack of potential adaptability to rapidly changing climate from two of the main species of this woodland. In a similar way, Longleat2 was the most diverse in species and canopy structure but the main species (*P. menziesii*) had a low N_e (adults = 395.4 and juveniles = 302.7). In contrast, the largest N_e found for *T. plicata* was in the very early stage of transformation stand, Bagley Wood (adults = 55.1 and juveniles = 48.7), which compared to the $N_e > 500$ threshold is still very low, however, it could indicate a better potential adaptation than the other sites studied. These results indicate that the seemingly diverse and close-to-natural woodlands may lack sufficient genetic diversity and resilience to withstand the challenges posed by current and future climate variations, suggesting that genetic diversity monitoring is key to ensuring these forests' persistence.

Previous studies have examined the effects of management systems on the genetic diversity of different forest species revealing evidence of potential impacts on the genetic characteristics of the stands (reviewed by Ratnam et al., 2014). A study on *Abies alba* woodlands under different management systems in Germany and Switzerland showed a slightly lower genetic diversity in stands managed under a selective harvesting system, represented by uneven-aged populations compared to even-aged plantations (Konnert & Hussendörfer, 2001). In another study, the genetic diversity variation was compared between adults and seedlings from two species (*Abies amabilis* and *Tsuga heterophylla*) commonly found in mixed species forest populations of British Columbia managed by

using either shelterwood (selective harvesting), patch cut, green tree-retention or clear-cutting (El-Kassaby et al., 2003). The results differed between species, *Abies amabilis* did not show significant differences between silvicultural treatments while *Tsuga heterophylla* had a lower genetic diversity under shelterwood management (El-Kassaby et al., 2003). Both of the above studies suggest that selective harvesting, commonly used under CCF management, may reduce the N_e and subsequently, the genetic diversity of the species leading to a lack of potential adaptation to different climatic conditions.

The pattern that emerges from these results is that genetic diversity differs among forests differently managed and among species, as reviewed by Ratnam et al. (2014). Furthermore, the variable results from studies on woods managed by a CCF approach, together with our results, suggest that the genetic diversity may be affected by the reduction of effective population size, the potential gene flow from surrounding areas and may be dependent on the pollen dispersal method of each species.

5.2.4. Assessment of the thesis objectives and hypothesis

Overall, the objectives of the thesis were reached and the hypothesis was partially confirmed as some of the stands showed diverse results. We optimised the procedure to isolate DNA (Objective 1) and developed reliable genotyping assays (Objective 2) for both study species which allowed us to estimate genetic diversity parameters on each study stand. We performed a forest survey to characterise the level of transformation of the stands to CCF (Objective 3) and assessed the differences between the genetic diversity levels present in adults and juveniles (Objective 4). The hypothesis of the thesis is that in the long-term, the natural regeneration resulting from planted trees and managed under CCF may show a low genetic diversity. This can result when: (i) the planted canopy trees hold a reduced gene pool, (ii) the gene pool is transferred poorly to the offspring, and (iii) the woodlands are logged and trees are selected to be removed decreasing the effective population size, which may result in a decrease of diversity over time.

Therefore, the low N_e of some of the *P. menziesii* adult groups tested (Stourhead2, Longleat2 and Mortimer), compared to the similarly low N_e in the juveniles, may indicate that current

generations may already hold a narrow gene pool to withstand changes in the climate, confirming the first potential explanation of the hypothesis (i). The *P. menziesii* adults at Ringwood had high N_e (428.6) but in the juveniles, the N_e was substantially lower (271.9). Together with the low levels of assigned parent-offspring relationships, this may indicate a potentially poor transfer of the gene pool to the offspring, confirming the second potential explanation of the hypothesis (ii). Likewise, the N_e results from *T. plicata* adults suggest a rather narrow gene pool across all the stands, following the first potential explanation of the hypothesis suggested (i); however, in some cases the juveniles showed a larger N_e compared to the adults (at Longleat1 and Stourhead1), partially invalidating the hypothesis in these stands. The N_e increase from generation to generation may be due to gene flow, external individuals contributing to the reproduction process. This may increase the genetic diversity of the stands, but a genetic assessment of the neighbour material followed by genetic diversity monitoring may be needed to ensure the adaptation potential of the stands to face environmental changes. Finally, stands that were at a late stage of transformation (Stourhead2 including both species) showed lower N_e compared to some of the stands at an earlier transformation stage (Ringwood, *P. menziesii* and Bagley wood, *T. plicata*) in both Chapters 3 and 4. This may be due to a larger reduction of potential parent trees in later stages of transformation to CCF decreasing the effective size of the species in the woodlands, which confirms the third potential explanation of the hypothesis (iii).

5.3 Future research directions

A provenance assignment study may shed light and confirm the suitability of the source material used to establish the study sites in woodlands. In forestry species, it is a widespread practice to select populations from across the native distribution range as seed sources to then be used to establish new populations (Alizoti et al., 2022). Several studies have validated the use of specific provenances over others by developing trials evaluating the performance under different environmental conditions to select the optimal provenance to use in the target plantation region (Park & Rodgers, 2023). Public and private woodland estates may hold records of the material used for the

establishment of their plantations; however, this information may be lacking in woodlands planted in the early and mid-1900s. We intended to assess the origin of the source material used in the *P. menziesii* stands by analysing a subsample of adults together with individuals from provenance trials but data were not available on time for inclusion in the thesis. There is limited information on *T. plicata* provenance tests in the UK, although the New Forest holds a provenance trial which could be sampled and used for this matter. Regarding *P. menziesii*, there are several provenance trials established across the UK, including the Forest of Dean and the Radnor Forest. We selected nine provenances evenly distributed through its natural range and collected 20 individuals per provenance. The DNA was isolated and sent to be genotyped using the Sequenom assay. Genetic analyses to assess the similarity between the provenances and the study sites will be used to assign the potential origin of the material used.

This research would also benefit from additional studies looking at the potential gene flow across the study sites and neighbouring plantations. Other studies have analysed the pollen contamination in seed orchards from neighbouring plantations in conifer species (Bouffier et al., 2023; Galeano et al., 2021). Bouffier et al. (2023) found a 50% pollen contamination average across *P. pinaster* seed orchards while (Pakull et al., 2021) found a lower contamination rate in *P. menziesii* seed orchards (a maximum of 8%). Additionally, Song et al. (2018) reported 28% of pollen contamination in *P. menziesii* stands under natural conditions. In our study, we found potential pollen exchange with neighbouring woodlands, thus, a more detailed gene flow analysis may shed light on the reproduction process and the potential contribution of external pollen to a stand's genetic diversity. Additionally, long-term monitoring projects may be useful to capture the changes in genetic diversity in response to CCF progress and to assess the potential need to supplement genetic diversity using different strategies.

5.4 Impacts and potential applications of the research

Minimising laboratory plastic footprint is crucial to reducing CO_2 emissions and therefore tackling the climatic crisis (Alves et al., 2021). Our results showed that using a sustainable method

(MicroGem), laboratories could reduce by half their single-plastic use when isolating DNA compared to a commonly used kit (QIAGEN). This method had to be optimised to provide sufficient quality DNA, which added some time to the project; however, MicroGem needs half of the time to process one sample compared to QIAGEN, which makes the optimisation time worth it. Our findings may encourage researchers to select sustainable DNA isolation methods and apply sustainable laboratory practices to reduce their plastic footprint. However, further research is needed to explore the performance of this method on a broader range of species, together with the development and optimisation of other low-impact molecular experiments on specific species.

Understanding the genetic makeup of the study woodlands may hold significant implications to inform forest management practices, particularly in the context of non-native conifers under CCF management. The thesis findings can guide strategies to monitor the potential adaptability of planted conifer populations to be transformed or in transformation to CCF. The genotyping assays that were designed together with a simple, cost-efficient DNA isolation and genotyping service may be useful to landowners and forest managers to integrate genetic diversity results into their management plan. Insights on the levels of genetic diversity of both adults and natural regeneration may inform the potential need to supplement genetic diversity (Heuertz et al., 2023) by providing assisted gene flow (Aitken & Bemmels, 2016).

Moreover, the insights gained from this research can contribute to broader discussions on the new genetic diversity indicators that were approved by the Convention on Biological Diversity (CBD, 2022) at COP15 to assess the status and trends in genetic diversity (Hoban et al., 2023). An effective population size of at least 500 is the recommended target to ensure long-term genetic diversity maintenance (O'Brien et al., 2022); however, none of our study sites met this minimum threshold, suggesting a lack of potential adaptability to changing climate conditions unless there is a significant influx of genetically diverse individuals.

Additionally, in some cases, there is a limitation of resources to perform DNA-based analysis to monitor the levels of genetic diversity for conservation urges. Hoban et al., (2023) proposed the use of the N_e / N_c ratio to calculate the effective population size (N_e) by estimating the actual population size (N_c); however, the N_e / N_c ratio currently recommended is 0.1 and it is generalised to every

species which may limit the accuracy of the N_e estimation. Therefore, our project provides accurate results of N_c and N_e for *P. menziesii* and may contribute to the development of a specific N_e / N_c database to be used in conservation genetics projects.

5.5 References

Aitken, S. N., & Bemmels, J. B. (2016). Time to get moving: Assisted gene flow of forest trees. *Evolutionary Applications*, *9*(1), 271–290. https://doi.org/10.1111/eva.12293

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evolutionary Applications*, *1*(1), 95–111. https://doi.org/10.1111/J.1752-4571.2007.00013.X

Alizoti, P., Bastien, J. C., Chakraborty, D., Klisz, M. M., Kroon, J., Neophytou, C., Schueler, S., van Loo, M., Westergren, M., Konnert, M., Andonovski, V., Andreassen, K., Brang, P., Brus, R., Cvjetković, B., Đodan, M., Fernández, M., Frýdl, J., Karlsson, B., … Tsvetkov, I. (2022). Non-Native Forest Tree Species in Europe: The Question of Seed Origin in Afforestation. *Forests*, *13*(2), 273. https://doi.org/10.3390/F13020273/S1

Alves, J., Sargison, F. A., Stawarz, H., Fox, W. B., Huete, S. G., Hassan, A., McTeir, B., & Pickering, A. C. (2021). A case report: Insights into reducing plastic waste in a microbiology laboratory. *Access Microbiology*, *3*(3). https://doi.org/10.1099/acmi.0.000173

Bennett, G., Hardy, A., Bunting, P., Morgan, P., & Fricker, A. (2020). A Transferable and Effective Method for Monitoring Continuous Cover Forestry at the Individual Tree Level Using UAVs. *Remote Sensing 2020, Vol. 12, Page 2115*, *12*(13), 2115. https://doi.org/10.3390/rs12132115

Bouffier, L., Debille, S., Alazard, P., Raffin, A., Pastuszka, P., & Trontin, J.-F. (2023). Pollen contamination and mating structure in maritime pine (*Pinus pinaster* Ait.) clonal seed orchards revealed by SNP markers. *Peer Community Journal*, *3*. https://doi.org/10.24072/pcjournal.302

CBD. (2022). Decision adopted by the conference of the parties to the Convention on Biological Diversity. 15/15 Monitoring framework for the Kunming-Montreal Global Biodiversity Framework. In *Convention of Biological Diversity*. https://www.cbd.int/doc/decisions/cop-15/cop-15-dec-05-en.pdf

El-Kassaby, Y. A., Dunsworth, B. G., & Krakowski, J. (2003). Genetic evaluation of alternative silvicultural systems in coastal montane forests: Western hemlock and amabilis fir.

Theoretical and Applied Genetics, *107*(4), 598–610. https://doi.org/10.1007/s00122-003-1291-3

Finkeldey, R. (2002). Reproduction in Continuous Cover Forests—The Geneticist's Perspective. In *Continuous Cover Forestry* (pp. 67–79). https://doi.org/10.1007/978-94-015-9886-6_6

Frankham, R., Bradshaw, C. J. A., & Brook, B. W. (2014). Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, *170*, 56–63. https://doi.org/10.1016/j.biocon.2013.12.036

FU, Z. yu, SONG, J. cheng, & Jameson, P. E. (2017). A rapid and cost effective protocol for plant genomic DNA isolation using regenerated silica columns in combination with CTAB extraction. *Journal of Integrative Agriculture*, *16*(8), 1682–1688. https://doi.org/10.1016/S2095-3119(16)61534-4

Fumagalli, M. (2013). Assessing the Effect of Sequencing Depth and Sample Size in Population Genetics Inferences. *PLOS ONE*, *8*(11), e79667. https://doi.org/10.1371/JOURNAL.PONE.0079667

Galeano, E., Bousquet, J., & Thomas, B. R. (2021). SNP-based analysis reveals unexpected features of genetic diversity, parental contributions and pollen contamination in a white spruce breeding program. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-84566-2

Green, M. R., & Sambrook, J. (2017). Isolation of High-Molecular-Weight DNA Using Organic Solvents. *Cold Spring Harbor Protocols*, *2017*(4), pdb.prot093450. https://doi.org/10.1101/pdb.prot093450

Hermann, R. K., & Lavender, D. P. (1999). Douglas-fir planted forests. *New Forests*, *17*(1/3), 53–70. https://doi.org/10.1023/A:1006581028080

Heuertz, M., Carvalho, S. B., Galindo, J., Rinkevich, B., Robakowski, P., Aavik, T., Altinok, I., Barth, J. M. I., Cotrim, H., Goessen, R., González-Martínez, S. C., Grebenc, T., Hoban, S., Kopatz, A., McMahon, B. J., Porth, I., Raeymaekers, J. A. M., Träger, S., Valdecantos, A., Vella, A., Vernsesi, C., Garnier-Géré, P. (2023). The application gap: Genomics for biodiversity and ecosystem service management. Biological Conservation, 278, 109883.

https://doi.org/10.1016/J.BIOCON.2022.109883

Hoban, S., Bruford, M., D'Urban Jackson, J., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., Paz-Vinas, I., Sjögren-Gulve, P., Segelbacher, G., Vernesi, C., Aitken, S. N., Bertola, L. D., Bloomer, P., Breed, M., Rodríguez-Correa, H., Funk, W. C., Grueber, C. E., Hunter, M. E., Jaffe, R., Liggins , L., Mergeay , J., Moharrek , F., O'Brien , D., Ogden , R., Palma-Silva , C., Pierson , J., Ramakrishnan , U., Simo-Droissart , M., Tani , N., Waits , L. & Laikre, L. (2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biological Conservation, 248, 108654. https://doi.org/10.1016/j.biocon.2020.108654

Hoban, S., Bruford, M. W., da Silva, J. M., Funk, W. C., Frankham, R., Gill, M. J., Grueber, C. E., Heuertz, M., Hunter, M. E., Kershaw, F., Lacy, R. C., Lees, C., Lopes-Fernandes, M., MacDonald, A. J., Mastretta-Yanes, A., McGowan, P. J. K., Meek, M. H., Mergeay, J., Millette, K. L., Mittan-Moreau, C.S., Navarro, L.M., O'Brien, D., Ogden, R., Segelbacher, G., Paz-Vinas, I., Vernesi, C. & Laikre, L. (2023). Genetic diversity goals and targets have improved, but remain insufficient for clear implementation of the post-2020 global biodiversity framework. Conservation Genetics. https://doi.org/10.1007/s10592-022-01492-0

Hong, E. P., & Park, J. W. (2012). Sample Size and Statistical Power Calculation in Genetic Association Studies. *Genomics & Informatics*, *10*(2), 117. https://doi.org/10.5808/GI.2012.10.2.117

Isaac-Renton, M. G., Roberts, D. R., Hamann, A., & Spiecker, H. (2014). Douglas-fir plantations in Europe: A retrospective test of assisted migration to address climate change. *Global Change Biology*, *20*(8), 2607–2617. https://doi.org/10.1111/gcb.12604

Kerr, G., Boswell, R., & Mason, B. (2003). A sampling system to monitor the transformation from even-aged stands to continuous cover. *Forestry*, *76*(4), 425–435. https://doi.org/10.1093/forestry/76.4.425

Kerr, G., Morgan, G., Blyth, J., & Stokes, V. (2010). Transformation from even-aged plantations to an irregular forest: The world's longest running trial area at Glentress, Scotland. *Forestry*, *83*(3), 329–344. https://doi.org/10.1093/forestry/cpq015

Kerr, G., Snellgrove, M., Hale, S., & Stokes, V. (2017). The Bradford-Hutt system for transforming young even-aged stands to continuous cover management. *Forestry*, *90*(4), 581–593. https://doi.org/10.1093/forestry/cpx009

Konnert, M., Hussendörfer, E. (2001). Genetic variation of silver fir (*Abies alba*) in unevenaged forests ("Plenter" forest) in comparison with evenaged forests (Altersklassenwald). In: Müller-Starck, G., Schubert, R. (eds) Genetic Response of Forest Systems to Changing Environmental Conditions. Forestry Sciences, vol 70. Springer, Dordrecht. https://doi.org/10.1007/978-94-015-9839-2_25

Marengo, A., Cagliero, C., Sgorbini, B., Anderson, J. L., Emaus, M. N., Bicchi, C., Bertea, C. M., & Rubiolo, P. (2019). Development of an innovative and sustainable one-step method for rapid plant DNA isolation for targeted PCR using magnetic ionic liquids. *Plant Methods*, *15*(1), 1–11. https://doi.org/10.1186/S13007-019-0408-X/FIGURES/7

Mason, B., Kerr, G., & Simpson, J. (1999). What is Continuous Cover Forestry ? *Forestry Commission*. 1–8.

Neophytou, C., van Loo, M., & Hasenauer, H. (2019). Genetic diversity in introduced Douglas-fir and its natural regeneration in Central Europe. *Forestry: An International Journal of Forest Research*, Volume 93, Issue 4, July 2020, Pages 535–544, https://doi.org/10.1093/forestry/cpz055

O'Brien, D., Laikre, L., Hoban, S., Bruford, M. W., Ekblom, R., Fischer, M. C., Hall, J., Hvilsom, C., Hollingsworth, P. M., Kershaw, F., Mittan, C. S., Mukassabi, T. A., Ogden, R., Segelbacher, G., Shaw, R. E., Vernesi, C., & MacDonald, A. J. (2022). Bringing together approaches to reporting on within species genetic diversity. *Journal of Applied Ecology*, *59*(9), 2227–2233. https://doi.org/10.1111/1365-2664.14225

Pakull, B., Eusemann, P., Wojacki, J., Ahnert, D., & Liesebach, H. (2021). Genetic diversity of seeds from four German Douglas fir (*Pseudotsuga menziesii*) seed orchards. *European Journal of Forest Research*, *140*(6), 1543–1557. https://doi.org/10.1007/s10342-021-01419-3

Park, A., & Rodgers, J. L. (2023). Provenance trials in the service of forestry assisted migration: A review of North American field trials and experiments. *Forest Ecology and Management*,

537, 120854. https://doi.org/10.1016/J.FORECO.2023.120854

Pommerening, A., Maleki, K., & Haufe, J. (2021). Tamm Review: Individual-based forest management or Seeing the trees for the forest. *Forest Ecology and Management*, *501*, 119677. https://doi.org/10.1016/j.foreco.2021.119677

Pommerening, A., & Murphy, S. T. (2004). A review of the history, definitions and methods of continuous cover forestry with special attention to afforestation and restocking. *Forestry*, *77*(1), 27–44. https://doi.org/10.1093/forestry/77.1.27

Ratnam, W., Rajora, O. P., Finkeldey, R., Aravanopoulos, F., Bouvet, J.-M., Vaillancourt, R. E., Kanashiro, M., Fady, B., Tomita, M., & Vinson, C. (2014). Genetic effects of forest management practices: Global synthesis and perspectives. *Forest Ecology and Management*, *333*, 52–65. https://doi.org/10.1016/j.foreco.2014.06.008

Rezadoost, M. H., Kordrostami, M., & Kumleh, H. H. (2016). An efficient protocol for isolation of inhibitor-free nucleic acids even from recalcitrant plants. *3 Biotech*, *6*(1), 1–7. https://doi.org/10.1007/S13205-016-0375-0

Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Menzel, A., Root, T. L., Estrella, N., Seguin, B., Tryjanowski, P., Liu, C., Rawlins, S., & Imeson, A. (2008). Attributing physical and biological impacts to anthropogenic climate change. *Nature*, *453*(7193), 353–357. https://doi.org/10.1038/nature06937

Shalev, T. J., Gamal El-Dien, O., Yuen, M. M. S., Shengqiang, S., Jackman, S. D., Warren, R. L., Coombe, L., van der Merwe, L., Stewart, A., Boston, L. B., Plott, C., Jenkins, J., He, G., Yan, J., Yan, M., Guo, J., Breinholt, J. W., Neves, L. G., Grimwood, J., … Bohlmann, J. (2022). The western redcedar genome reveals low genetic diversity in a self-compatible conifer. *Genome Research*, 1952–1964. https://doi.org/10.1101/gr.276358.121

Shalev, T. J., Gamal El-Dien, O., Yuen, M. M. S., van der Merwe, L., Kirst, M., Yanchuk, A. D., Ritland, C., Russell, J. H., & Bohlmann, J. (2023). Genetic architecture of terpene chemistry and growth traits and the impact of inbreeding on these traits in western redcedar (Thuja plicata). *Evolutionary Applications*. https://doi.org/10.1111/EVA.13526

Song, J., Ratcliffe, B., Kess, T., Lai, B. S., Korecký, J., & El-Kassaby, Y. A. (2018). Temporal

quantification of mating system parameters in a coastal Douglas-fir seed orchard under manipulated pollination environment. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-30041-4

St Clair, J. B., Mandel, N. L., & Vance-Borland, K. W. (2005). Genecology of Douglas fir in western Oregon and Washington. *Annals of Botany*, *96*(7), 1199–1214. https://doi.org/10.1093/aob/mci278

Thomas, F. M., Rzepecki, A., & Werner, W. (2022). Non-native Douglas fir (Pseudotsuga menziesii) in Central Europe: Ecology, performance and nature conservation. *Forest Ecology and Management*, *506*. https://doi.org/10.1016/J.FORECO.2021.119956

Torkamaneh, D., Laroche, J., & Belzile, F. (2020). Fast-GBS v2.0: An analysis toolkit for genotyping-by-sequencing data. *Genome*, *63*(11), 577–581. https://doi.org/10.1139/gen-2020-0077

van Loo, M., Lazic, D., Chakraborty, D., Hasenauer, H., & Schüler, S. (2019). North American Douglas-fir (*P. menziesii*) in Europe: Establishment and reproduction within new geographic space without consequences for its genetic diversity. *Biological Invasions*, *21*(11), 3249–3267. https://doi.org/10.1007/s10530-019-02045-2

Wang, T., & Russel, J. H. (2006). Evaluation of selfing effects on western redcedar growth and yield in operational plantations using the tree and stand simulator (TASS). *Forest Science*, *52*(3), 281–289. https://doi.org/10.5849/forsci.15-042