Ancient relicts in the limelight: an evolutionary study of diversity and demographic history in species of the broad-leaved temperate forest tree genus *Tilia.*

Thesis submitted to the

School of Biology, Newcastle University

for the degree of Doctor of Philosophy

Samuel A. Logan

April 2016

Declaration

I hereby certify that this thesis is my own original work. No part of it has been previously submitted to any University for obtaining a higher degree. All information derived from other sources has been acknowledged.

Samuel A. Logan
April 2016

Dissemination of the findings from this study

Publications:

Logan, S.A.*, Phuekvilai, P.[#] and Wolff, K.[$] (2015). Ancient woodlands in the limelight: delineation and genetic structure of ancient woodland species *Tilia cordata* and *T. platyphyllos* (Tiliaceae) in the UK. *Tree Genetics and Genomes.* 11:52

Logan, S.A. and Wolff, K. Genetic diversity and demographic history of the Siberian lime (*Tilia sibirica*); an important conservation unit. *Manuscript.*

Logan, S.A. and Wolff, K. Clonal architecture and diversity of three *Tilia* species. *Manuscript.*

Logan, S.A. and Wolff, K. RNA-Seq analysis of *Tilia* (Malvaceae): *De Novo* Transcriptome assembly, annotation, and SSR marker identification. *Manuscript.*

* S.A.L. collected leaf samples, extracted all DNA, analysed all data, and wrote the article. [#] P.P. designed the microsatellite markers. [$] K.W. supervised the study.

Conference Presentations:

Logan, S.A. & Wolff, K. (2015). Ancient woodlands in the limelight: genetic structure of *Tilia* in the UK. [Oral presentation]. *Plant Sci. 2015*. Harper Adams University.

Logan, S.A. & Wolff, K. (2014). Clonal structure in UK *Tilia* populations. [Poster presentation]. *Ecological Genetics Group*. Newcastle University.

Logan, S.A. & Wolff, K. (2013). Molecular Ecology of UK *Tilia* populations. [Oral presentation]. *Ecological Genetics Group*. Queens University Belfast.

Logan, S.A**.** & Wolff, K. (2012). Historical ecology of lime (*Tilia*) populations inferred using modern molecular techniques. [Poster presentation]. *Trees beyond the Woods Conference.* Sheffield Hallam University.

Reports:

Logan, S.A. (2015). Białowieża Park Narodowy Research Unit Fieldwork Report

Logan, S.A. (2014). Sampling *Tilia* in the Białowieża National Park, Poland. Society of Biology Travelling Fellowship Report.

Abstract

Tilia L. is a temperate-forest tree genus with a wide northern hemisphere distribution. Several species within the genus are affected by forest fragmentation. Three species were the focus of this study, *T. cordata* Mill. (small-leaved lime) and *T. platyphyllos* Scop. (large-leaved lime) from the UK, Austria, Poland, and western Siberia and *T. sibirica* Bayer (Siberian lime) from southern Siberia.

Tilia specific microsatellite markers were used to assess various population genetics indices. Genetic diversity and structure of UK *T. cordata* and *T. platyphyllos* populations were estimated. To determine the genetic and demographic history of *T. sibirica* and *T. cordata*, Approximate Bayesian Computation (ABC) analyses were used. An investigation into the clonal architecture of the three species was carried out to assess the level of clonality and the impact of clonal reproduction on genetic diversity. In addition, Next Generation Sequencing of the *Tilia* leaf transcriptome was carried out using direct RNA sequencing

Results confirm that the three species are diploid and outcrossing. Although hybridisation occurs among *T. cordata* and *T. platyphyllos*, the two are distinct biological units with high genetic diversity and intra-specific population structure.

Significant genetic differentiation was observed between *T*. *sibirica* and *T. cordata* and low genetic diversity in the Siberian lime was revealed. ABC analysis suggests a relatively recent (Early Holocene) divergence between the Siberian lime and the small-leaved lime. The Holocene split coincided with a westerly migration of *Tilia* genotypes that may have contributed to the recolonization of *T. cordata* in Europe.

Fewer clones were observed in *T. platyphyllos* than the other two species and range-edge populations experience greater clonality than central European populations. Clonal occurrence does not appear to have had a negative effect on genetic diversity.

A method for the *de novo* assembly and annotation of the leaf transcriptome from *T. cordata* and *T. platyphyllos* is provided. Potentially thousands of simple sequence repeats (SSRs) from each species have been identified.

Acknowledgements

This study would not have been possible without the help and assistance from many people. First of all I would like to thank my supervisor Dr Kirsten Wolff for her guidance and support throughout this project. She has given me the freedom to develop this study from the onset and continuously encouraged my professional, academic and research development. Many thanks also to Dr Gordon Port.

I would like to thank Dr Paul Ashton, Dr Milan Chytrý and Seb Mankelow for providing leaf samples and Dr Rebecca Roseff and Tim Laurie for their assistance with sample collection and fieldwork. I would like to thank the National Trust, Forestry Commission, Local Wildlife Trusts and the private landowners who granted me permission to collect leaf material from their land. In particular, my appreciation goes to the Białowieża National Park authorities for permitting access to the Białowieża Special Reserve. The area is a truly inspiring place and having the opportunity to spend a few days there collecting *Tilia* samples was the highlight of my Ph.D. experience. Many thanks to Jade Lauren Cawthray for assisting with field work in Poland and for her botany, navigation and wild boar surveillance skills.

Many thanks to Dr Chris Wright from the High-Throughput Sequencing and Genotyping Unit at the University of Illinois for the RNA-Seq library construction and Dr Ben Allen from the Bioinformatics Support Unit at Newcastle University for his assistance with the transcriptome assemblies. I am grateful to Professor Donald Pigott and the anonymous reviewers for reviewing an earlier version of a manuscript now published from this study and for their helpful comments and suggestions. Many thanks to the thesis examiners, Dr Stephen Cavers and Prof. John Bythell. I am thankful to Dr Nicholas Levsen, Dr Prattana Phuekvilai and Dr Shreya Wani for their thoughts and helpful discussions and to Helen Martin for lab assistance. Special thanks to Barry, Matt, Dan, Tom and Ali for useful ‘non-Ph.D.’ distractions. This study was funded by the RB Cooke Studentship from Newcastle University, for which I am truly grateful.

And finally, I have to acknowledge the amazing patience, love and support from my family, and give a heartfelt thank you in particular to my partner Siobhan Ryan, our beautiful daughter Autumn Belle Logan and to my parents, without whom none of this would have been possible. I would like to dedicate this accomplishment to them.

Table of contents

List of figures

List of tables

Chapter 1: General Introduction

Population genetics is the study of allele frequencies, and allelic variation within and among populations. It quantifies the diversity of species at the molecular level following basic principles and assumed processes such as Mendelian inheritance, mutation, assorted mating systems, gene flow, and natural selection. Carrying out such studies can subsequently inform conservation action, either directly by concentrating on a particular species of animal or plant, or indirectly by focusing on habitat structure and fragmentation (*e.g.* forest systems).

Fragmentation is a continuous threat to woodlands in the UK and across the globe. Agricultural practices have been steadily fragmenting forest systems for some 6000 years (Watts, 2006; Fyfe *et al.*, 2015). The global population increase means that modern farming techniques, food security, infrastructural development, and the increase in building houses will inevitably mean a further reduction in forest habitats. The United Nations declared 2011 as the International Year of Forests to bring attention to the global importance and general trends of woodlands. Although Europe's woodland habitats, presently covering 45% of land area, are slowly increasing relative to the rest of the world, (approximately 0.26 million *ha* increase in two decades from 1990 – 2010), the UK woodland cover is just 13%. Relative to country size, UK woodlands are some of the smallest wooded areas worldwide (FAO, 2011).

While our use of woodlands as an integral part of our existence is less than it once was – at least in many parts of the UK and Europe, globally, they still play an important role to the daily lives and wellbeing of many people, either directly for recreational use, and living environments (1.6 billion people rely on woodlands as a direct resource, FAO, 2011) or indirectly for ecosystem services such as carbon sink, preventing land erosion and for the control of water runoff. Currently, over 31% of the Earth's land mass is woodland and over 80% of terrestrial organisms use woodland environments as habitat. However, the continuous fragmentation of wooded areas leads to genetic and ecological isolation, which in the short term can have negative effects on forest's core components *i.e.* tree species (Bailey *et al.*, 2002; Petit *et al.*, 2004; Riutta *et al.*, 2014) and in the long term may ultimately intensify the effects of climate change on biodiversity (Opdam and Wascher, 2004).

Fragmentation can reduce population size, and increase isolation. Isolation is a "major species forming factor" (Wright, 1931) and being isolated has the potential to

result in genetically diverse endemics (Mayr, 1942). If an endemic population meets the criteria of a measurable unit of conservation *i.e.* Evolutionary Significant Unit (ESU) or Management Unit (MU), then they should be managed separately (Crandall *et al.*, 2000). Fragmentation can also lead to changes in mating systems, and can affect genetic diversity. This current study focuses on the non-model and highly fragmented, but ecologically important tree genus *Tilia.* The outcome of this study will immediately broaden our understanding of the molecular ecology of *Tilia*, in the UK and throughout the rest of Europe and worldwide.

1.1 The genus Tilia

Tilia L. (Malvaceae), also called lime, linden or basswood is an ecologically important genus of broad-leaved, temperate forest trees (Elwes and Henry, 1913) with an almost circumboreal northern hemisphere distribution (Pigott, 2012). Although *Tilia* was present in western North America throughout the early Tertiary, it is now no longer present there. *Tilia* trees are typically insect pollinated and seed dispersal can be wind facilitated (Pigott, 2012).

Taxonomic uncertainty had resulted in more than 100 putative species being described in the past. This was largely due to morphological variation. This lead Pigott (2012) to undertake a revision of the genus, suggesting 23 species with 14 subspecies. Following Browicz (1968) and Pigott (2012), four species are generally recognized as native in Europe, *Tilia cordata,* Mill., *T. platyphyllos* Scop., *T. dasystyla* Steven, and *T. tomentosa* Moench. Two of these, *T. cordata* (small-leaved lime), and *T. platyphyllos* (large-leaved lime), are widely distributed and sympatric in central and western Europe, including England and Wales (Fig. 1.1), where their hybrid, *T. x europaea* L. (syn. *T. x vulgaris*) occurs naturally and has been widely planted in towns and parks (Pigott, 1969, 1981a).

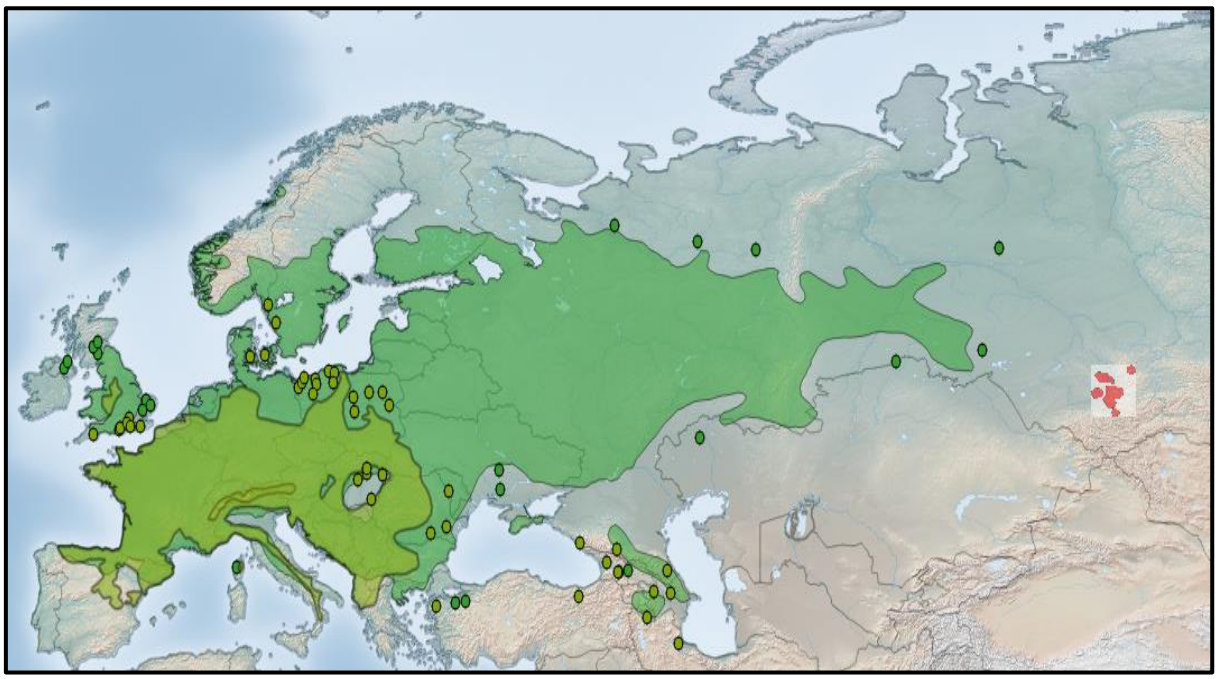


Figure 1.1 The distribution of *Tilia cordata* (large darker green area), *T. platyphyllos* (smaller green area) and *T. sibirica* (orange) throughout their natural ranges in Europe and Siberia. Map constructed in QGIS v2.14 with distribution of *T. cordata* and *T. platyphyllos* from EUFORGEN (www.euforgen.org) and *T. sibirica* distribution data georeferenced from Novák *et al.*, 2014.

It is thought that *Tilia* reproduce asexually in many parts of their range by clonal reproduction (Radoglou *et al.*, 2009). Clonal growth may have contributed to survival of individual trees at the extremes of their geographical range (Pigott and Huntley, 1978). Clonal spread is achieved in the genus through vegetative growth from epicormic shoots and branch extensions from the root collar, which is particularly apparent on steep terrain where soil erosion occurs and from fallen tree trunks (*pers. obs.,* Appendices 1.1 and 1.2)

The genus once dominated European forests (Pigott, 2012) and still dominates stands in parts of Poland, Czech Republic and Siberia (Pigott, 1975; Chytry and Sádlo, 1997; Novák *et al.*, 2014). However, in the current climate and with past human impact, *Tilia* trees in Europe and Siberia are now often subordinate and grow as an understorey to more dominant and/or commercial trees such as *Quercus* spp. (Pigott, 2012), *Abies sibirica* and *Populus tremula* (Novák *et al.*, 2014), although *T. cordata* are grown commerically on a larger scale in some parts of Siberia (Pigott, 2012).

Although *T. cordata*, and *T. platyphyllos* were once dominant species throughout UK and European woodlands, pollen records have revealed significant reductions for some 5,000 years, (Godwin, 1975). Today, *Tilia* are found mainly on ancient woodland sites (c1600, Peterken, 1977) across England and Wales (Appendix 1.3). The decline of *Tilia* was likely due to a combination of cooling climate which affected the development of seeds (Pigott and Huntley, 1978, 1981) and human manipulation of the landscape, from coppicing (Appendix 1.4 and 1.5a) and agricultural practices, which prevented flowering and cleared the landscape (Turner, 1962; Rackham, 2003). While numbers of *Tilia* are much lower than they once were, the general distribution is similar to their post-glacial colonization range (Pigott and Huntley, 1978).

Members of the genus have similar morphological traits, which has caused ambiguity in species designation (Pigott, 2012). The native UK species can, in most circumstances, be distinguished by their leaves. *T. cordata* leaves are generally small and cordate (heart shaped), with few hairs that are usually red or brown in colour at the main vein on the underside of the leaf, while the leaves of *T. platyphyllos* subsp. *cordifolia* (Besser) C.K. Schneid., (the subspecies native to England and Wales) are much larger, with many light hairs on the underside. Other subspecies of *T. platyphyllos* have very little or no hairs on the sun-leaves. Leaves of *T. cordata* have a sharply pointed acumen (tapered point of leaf) and their marginal teeth are rounded. However, the two species are most clearly distinguished by their inflorescences and fruit. Inflorescences of *T. cordata* grow upright while in *T. platyphyllos* they hang down. Fruit of *T. cordata* are small with a thin wall while *T. platyphyllos* are large and ridged with a thick wall. The petioles of *T. cordata* are typically less than 1.2 mm in diameter while *T. platyphyllos* have a petiole diameter of more than 1.5 mm (Pigott, 1969; 2012). However, these distinguishing features may not always be easily observable. Leaves, when regularly cut, browsed or damaged may look different (*pers. obs.*) from their type specimen. Similarly, the inflorescences may not be immediately diagnostic, as those of the hybrid also hang down. Furthermore, many of the inflorescences from woodland individuals tend to be found in the upper canopy level and so may not always be observable from the ground and given that these are only available for a few months of the year, they are often not useful as a diagnostic identifier. Moreover, considering that the two UK species naturally hybridize and display

intermediate characteristics (Pigott, 1969), identifying features can become less useful.

Tilia cordata extends its range eastwards to its limits in western Siberia. Approximately 1,500 km further east is another member of the genus – *T. sibirica* (Siberian lime, Fig. 1.1). It is endemic to the western foothills of the Altai Mountains and exists only in small fragmented forests (Novak et al., 2014), which are under constant threat from open-cast coal mining (Pigott, 2012). Pigott (2012) recognizes the taxon to be one of two subspecies, namely *T. cordata* subsp. *sibirica.* The other subspecies recognized as *T. cordata* subsp. *cordata* (Pigott, 2012). However, other authors recognize the Siberian lime as a separate species (Novák et al., 2015). For the purpose of this study, *T. sibirica* was named at the species level.

Tilia sibirica is morphologically similar to *T. cordata* subsp. *cordata* with subtle differences as pointed out by Pigott (2012). The twigs are typically reddish brown in colour, slim and hairless. Buds in winter are similar to subsp. *cordata.* The leaves of *T. sibirica* tend to be hairless while subsp. *cordata* has occasional hairs. The leaf blade of *T. cordata* is easily recognised from its cordate shape but this is not observed in *T. sibirica.* The inflorescences of subsp. *cordata* typically have more flowers (5 – 8) than *T. sibirica* (2 – 4). The flowers themselves are similar with the exception of the ovary and stigmas. In *T. sibirica* the ovary is covered in short hairs while subsp. *cordata* has long hairs and the stigmas of *T. sibirica* appear to vary in shape *i.e.* short lobes or long and narrow lobes, depending on the location of specimens (Pigott, 2012). However, these measurements were based on few samples.

While ecological studies have been carried out on the genus (Pigott, 1969; Pigott, 1975; Pigott, 1981a; Pigott, 1981b; Pigott, 1991; Wicksell and Christensen, 1999; Pigott, 2000), few studies have utilized molecular methods to address ecological and evolutionary questions. Compared to other forest trees, *Tilia* remain a genetically understudied genus. To aid in species identification, assess diversity, clonal reproduction and historical demographic scenarios of *Tilia*, genetic and genomic techniques should be applied.

1.2 Microsatellite markers

Microsatellites, also known as short tandem repeats (STRs) or simple sequence repeats (SSRs) are a series of repeated nucleotide bases *e.g.* AT_{11} is a sequence consisting of bases ‘AT’ repeated eleven times (*i.e.* ATATATATATATATATATATAT),

and GAG_9 is a repeated sequence (GAGGAGGAGGAGGAGGAGGAGGAG). Microsatellites can be classified as perfect, that is, not interrupted *e.g.* ATATATATATAT, imperfect (*i.e.* ATATATATGCATATAT), and composite *e.g.* ATATATATATATATGCGCGCGCGCGCGC (AT_7GC_8), which are sequences containing more than one set of repeats (Oliveira *et al.,* 2006). They are largely non-coding regions of DNA and arise through errors during recombination and mutational processes such as unequal crossover which can result in a sequence losing or gaining repeats (Strand *et al.,* 1993). Microsatellites tend to have high mutation rates and subsequent high polymorphism. They are co-dominantly inherited and generally found in abundance in eukaryote genomes (Tautz and Renz, 1984; Lagercrantz *et al.*, 1993; Morgante and Olivieri, 1993) including chloroplasts (Provan *et al.*, 2001) and mitochondria (Soranzo *et al.*, 1999).

It is still unclear of the purpose of microsatellites and why they are found in so many genomes. However, these short repeated sequences can be used as genetic markers and have been particularly useful in population genetics studies. Along with their ease of use (after initial development), and replicability, microsatellites (generally consisting of 1 – 6bp motifs repeated *n* number of times) have been one of the most commonly used molecular markers offering a wide utility in plant science. Studies have focused on genetic diversity and population structure (Streiff *et al.*, 1998; Cottrell *et al.*, 2003), species delineation (Gugerli *et al.*, 2007), clonal reproduction (Vaughan *et al.*, 2007), divergence dates and demographic history (Havrdová *et al.*, 2015).

Microsatellite markers have recently been developed for *T. platyphyllos* (Phuekvilai and Wolff, 2013), and were found to be highly polymorphic. Many of these markers can be successfully amplified in other *Tilia* species including *T. cordata*. It is expected therefore, that the markers will be also be useful in the genetic analyses of *T. sibirica*.

1.3 Taking Tilia into the next generation of genetic analyses

Having an informative set of molecular markers will inevitably bring about many more molecular based studies on species within the genus. However, genetic techniques have advanced very quickly in such a short period of time, and will likely continue doing so. Next Generation Sequencing (NGS) has proven to be a useful resource in generating a significant amount of high-throughput data for analysis of non-model organisms. Indeed, NGS of genomic DNA usually generates much more data than

can be efficiently analysed. Like many non-model organisms, *Tilia* lacks a reference genome. A solution to this, and an alternative to whole genome DNA sequencing, is direct RNA sequencing (RNA-Seq). RNA-Seq not only generates manageable genetic data, because only the coding regions (exons) are targeted, but it permits a *de novo* transcriptome assembly, either partial or complete, with enormous sequencing depth and coverage. Transcriptome characterisation is an important starting point for downstream analysis and could provide excellent opportunities for more widespread ecological and evolutionary applications such as analysis of adaptation and population processes (Ellegren, 2008). As mentioned, until the current set of microsatellite markers were developed, there was an obvious lack of resources available in the *Tilia* genetic toolkit. With NGS techniques more accessible and considering the potential of having access to a large amount of genomic/transcriptomic data, a high-throughput study on species within the genus should be initiated. This will provide a platform for further 'omics research and could generate many more molecular markers for future population genetic studies.

1.4 Aims of this study

This thesis has focused on using population and evolutionary genetic tools on largely fragmented populations of three species in the non-model, temperate forest, tree genus *Tilia.* The project aimed to use *Tilia*-specific microsatellite markers to address important ecological and evolutionary questions regarding genetic diversity, population structure, clonal incidence, and divergence times; factors that can be strongly influenced by forest fragmentation.

In Chapter 2, the genetic diversity and population structure of *Tilia cordata* Mill. (small-leaved lime) and *T. platyphyllos* Scop. (large-leaved lime) at ancient woodland sites across England and Wales has been described.

In Chapter 3, standard population genetic and Approximate Bayesian Computation (ABC) analyses has been used to determine the genetic and demographic history of *T. sibirica* (Siberian lime) and that of its closest congeneric, *T. cordata* from Siberia, Poland and Austria.

In Chapter 4, the clonal architecture of *Tilia cordata, T. platyphyllos* and *T. sibirica* across their respective ranges, and the extent (if any) to which clonality has affected the genetic diversity of the three species have been investigated.

Additionally, as NGS technologies have become a possibility for many labs, and no doubt will become an expectation for future molecular studies, an efficient and

effective method for the *de novo* assembly and annotation of the leaf transcriptome from *Tilia cordata* and *T. platyphyllos* has been described in Chapter 5. As well as generating an enormous amount of RNA-Seq data from Illumina NGS platform, potentially thousands of simple sequence repeats (SSRs) in each species were identified.

Chapter 2: Delineation and genetic structure of ancient woodland species *Tilia cordata* and *T. platyphyllos* in the UK.

2.1 Abstract

Ancient woodlands in Europe are currently subjected to fragmentation, which leads to ecological deterioration and genetic isolation. This is likely to intensify with further climate change and increased land use. *Tilia* are keystone, ancient woodland indicators, represented in the United Kingdom by two species (*Tilia cordata* and *T. platyphyllos*) and their hybrid (*T. x europaea*). For effective management of the species, genetic differentiation and population genetic structure must be investigated. Morphology can usually separate the two species but the distinguishing features are not always observable, particularly in less optimum, shady, conditions. Thirteen microsatellite markers have been used to discriminate the species and hybrid, and to assess the population genetic diversity of the two species. The markers successfully distinguished the two species and hybrid. *Tilia cordata* and *T. platyphyllos* show significant genetic differentiation, indicating two distinct biological units. Sub-structure within both species is loosely related to geographic location. Because *Tilia* populations in the UK are highly fragmented future conservation should consider provenance issues.

This chapter has been published as:

Logan, S. A.; Phuekvilai, P. and Wolff, K. (2015). Ancient woodlands in the limelight: delineation and genetic structure of ancient woodland species *Tilia cordata* and *T. platyphyllos* (Tiliaceae) in the UK. *Tree Genetics and Genomes.* 11:52 DOI 10.1007/s11295-015-0872-z.

2.2 Introduction

Ancient woodlands are areas of land with continuous woodland cover since at least c1600 (Peterken, 1977), and have been of considerable conservation concern within the UK and Europe (Pigott, 1969; Hermy *et al.*, 1999; Rackham, 2008). They may be recognized by the presence of indicator species and are regarded as remnants of old European forest (Pigott, 1969; Peterken, 1974; Rackham, 2003). Evidence of continuous cover at some sites can be found from pollen records (Pigott, 2012). While some ancient woodlands within the UK are protected as features of statutory designated sites (Natural England, 2014), many are highly fragmented and under threat of further fragmentation (Rackham, 2008). This can result in the loss of ecological interactions (Bailey *et al.*, 2002; Petit *et al.*, 2004; Riutta *et al.*, 2014) and genetic integrity (Cottrell *et al.*, 2009). It may also intensify the effects of climate change on biodiversity by slowing or obstructing responses from species or populations, such as biotic interactions, species growth rates, local extinctions and distribution rate (Opdam and Wascher, 2004 and references therein).

Fragmentation causes populations to break up into smaller groups and can lead to isolation (Young *et al.*, 1996). In remnant ancient woodlands, past genetic variation may therefore have been lost over time. However, Hamrick (2004), suggests that fragmentation may have a lesser effect on some tree species than on other organisms due to their long generation times, elevated genetic diversity and extensive pollen flow. Contrasting perspectives add an uncertainty about the genetic status of small fragmented woodlands. It is therefore important to understand the level of genetic diversity that remains in remnant stands and the level of gene flow between fragments.

The description and identification of species of plants has been primarily based on morphological traits. However, phenotypic plasticity and hybridization can confound identification and lead to misclassification (Duminil and Di Michele, 2009; Meimberg *et al.*, 2010). Studies in Oak (*Quercus*) have focused on the effects of hybridization, and population divergence on species delineation and have shown some incongruities following morphological and molecular analyses (Bacilieri *et al.*, 1995; 1996; Aldrich *et al.*, 2003). Incorrect identification could potentially have negative effects on conservation or management (Bateman *et al.*, 2008; Barrett and Freudenstein, 2009).

Lime trees (*Tilia* sp.) in the UK are ancient woodland indicator species, (Babington, 1862; Rackham, 2008). *Tilia cordata* Mill. (small-leaved lime), and *T.*

platyphyllos Scop. (large-leaved lime), are found across central and western Europe. In England and Wales they naturally hybridize, producing *T. x europaea* L. (syn. *T. x vulgaris*) and the hybrid has been widely planted in towns and parks (Pigott, 1969; 1981b). They are insect pollinated and the seeds can occasionally disperse in wind (Pigott, 2012).

Species within the genus have similar morphological traits, which led to ambiguity in species designation (Pigott, 2012). Distinguishing features that differentiate *T. cordata* and *T. platyphyllos* may not always be easily observable from the ground and given that *Tilia* can often grow to 30 – 40m (Pigott, 2012), and even taller in some parts of eastern Europe (Pigott, 1975; Wesolowski and Rowinski, 2006), collecting leaf samples from the upper canopy may be problematic. Moreover, considering that the two UK species naturally hybridize and display intermediate characteristics (Pigott, 1969), identifying features can become less useful. To confirm identification of the two species and hybrids, genetic techniques should be applied.

At present, there is a large amount of ecological information available regarding *Tilia* (Pigott, 2012). In contrast, there is very little genetic information due to the limited number of useful molecular markers available. As the two species are important ecological components of ancient woodland in the UK and in woodland throughout Europe, an understanding of their genetic diversity, as well as the level of hybridization between the two, is required. As UK woodlands have been fragmented for some time, it is expected that gene flow between populations will have been restricted, allowing populations to differentiate. Informative genetic markers will allow further invesitgation.

Earlier genetic studies using chloroplast PCR-RFLP markers (Fineschi *et al.*, 2003), RAPD markers, (Liesebach and Sinkó, 2008; Hosseinzadeh Colagar *et al.*, 2013); and the ITS regions (Yousefzadeh *et al.*, 2012) were successful in determining various levels of genetic diversity. However, RAPDs are dominant markers and therefore restrict population genetic analyses, while chloroplast and ITS regions are generally conserved and may not differentiate the two UK species. Other authors have used isozyme markers (Maurer and Tabel, 1995; Fromm and Hattemer, 2003), and to some extent were able to distinguish species and clones. However, the need for fresh tissue for isozyme/allozyme studies may be a disadvantage over DNA based approaches where samples can be stored until required. Fewer loci and lower diversity has also been observed using isozymes compared to microsatellite markers in some tree species (Pfeiffer *et al.*, 1997; Streiff *et al.*, 1998; Sun *et al.*, 2001).

Microsatellites reveal high levels of diversity and allow quantification of population genetic structure.

Microsatellites have recently been developed for *T. platyphyllos* (Phuekvilai and Wolff, 2013). Many of these markers can be successfully amplified in other *Tilia* species, including *T. cordata*. Phuekvilai and Wolff (2013) reported that one locus (*Tc918*) amplified in *T. platyphyllos* but not in *T. cordata* and so could be useful in species delineation.

This study aimed to test the resolution of 13 variable nuclear microsatellite markers to discriminate between the two UK species and to distinguish them from the hybrid. More specifically, given the fragmented nature of many UK *Tilia* populations, three hypotheses were tested; (1) although hybridization is known to occur between the two species, *T. cordata* and *T. platyphyllos* are two separate evolutionary units and have high genetic differentiation, (2) as remnant populations have undergone long-term fragmentation within the UK, a high degree of genetic structure is present and limited gene flow among populations occurs, and (3) small, isolated populations show low genetic diversity due to stochastic genetic effects and inbreeding. The outcome of this study broadens our understanding of the molecular ecology of *Tilia*, as it is the first population genetics study to be carried out on autochthonous species in the UK and highlights populations with high genetic diversity that may be of conservation and restoration interest to forest managers.

2.3 Material and Methods

2.3.1 Study sites and sample collection

Leaf samples were collected from 27 locations (Fig. 2.1, Appendix 2.1), spanning most of the UK range of the two species. In total 550 *Tilia* samples were collected from the two species and the hybrid across all locations. Population size ranged from five individuals in a small isolated Worcestershire site to 45 individuals from Chanstone Wood, a large mixed site. Most sites were documented to have *T. cordata* or *T. platyphyllos* present and all sites are considered to be of ancient and/or semi-natural origins. Many are presently designated as Sites of Special Scientific Interest (SSSI) or National Nature Reserves (NNR, Natural England, 2014). Samples from some Worcestershire and Herefordshire sites are believed to be remnants of parish boundaries and old access routes from before the Enclosures Act (R. Roseff *pers.*

comm. Appendices 1.5a and 1.5b). Although all sites are considered to be of ancient or natural origin, they would have been managed in the past (*e.g.* coppiced).

One leaf was taken from each tree, and a GPS point was recorded for most samples. For those that could not be given a GPS point due to canopy cover or topography, an eight digit grid reference was recorded. The grid references and GPS points were later transformed into a reference point and mapped in QGIS v2.14.2 (www.qgis.org, Fig. 2.1). All samples were dried immediately at room temperature between tissues. Once dried all leaves were stored at -20°C until DNA extraction.

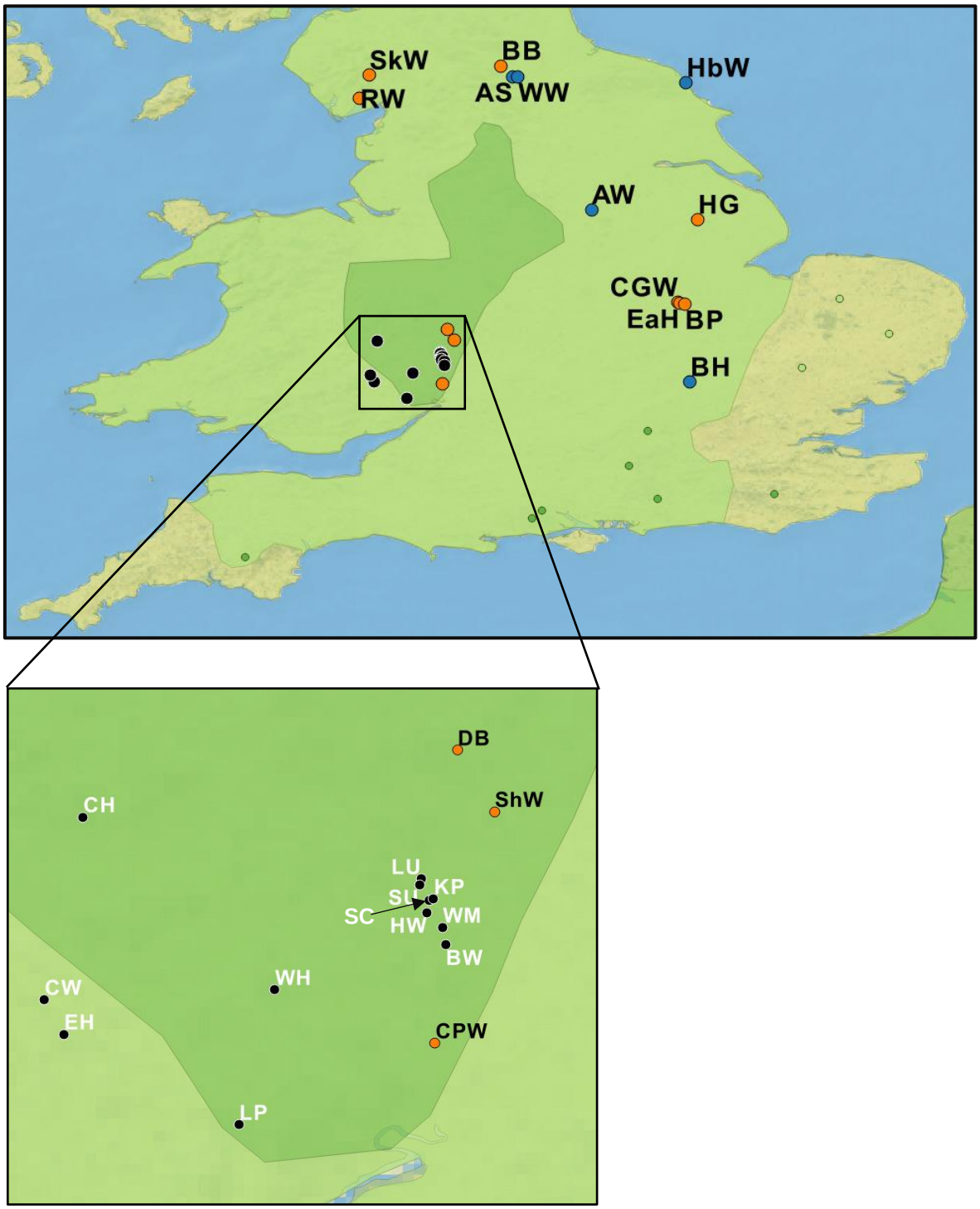


Figure 2.1 Locations of sampled *Tilia* populations. Blue points are *T. cordata*, orange points are *T. platyphyllos*, and black points (with white labels) are mixed populations. Map shows UK distribution of *T. cordata* (light green area and points) and *T. platyphyllos* (dark green area and points). Distribution data downloaded from EUFORGEN (www.euforgen.org).

2.3.2 DNA extraction and amplification

Genomic DNA was extracted using the CTAB (Cetyl Trimethyl Ammonium Bromide) method as described in Morgan-Richards and Wolff (1999). Extracted DNA was dissolved in 80 µl of TE (Tris-EDTA) buffer and stored at -20° C until required for DNA amplification. Four multiplex Polymerase Chain Reaction (PCR) procedures were carried out to amplify 13 microsatellite regions using primers developed for *Tilia* (Phuekvilai and Wolff, 2013). Amplification was carried out in a final volume of 10µl, consisting of 9µl of PCR master mix and 1µl (5ng) of DNA. Primer sequences, repeat motifs, allele size, fluorescent dyes, and primer concentration, as well as PCR conditions and parameters are described in Phuekvilai and Wolff (2013). PCR products were diluted to a 1/10 concentration with dH_2O, and stored at -20° C until required for genotyping.

2.3.3 Microsatellite genotyping

Electrophoresis was carried out with 1 µl of PCR product and a 10 µl mix of 10x Hi-Di formamide and 0.1x GeneScan™ 500 ROX size standard. Microsatellites were genotyped using an ABI 3130XL Genetic Analyser (Applied Biosystems), and allele fragment sizes were determined using GeneMapper® v4.0 (Applied Biosystems) software. Fragments were binned manually and checked for inconsistencies. To confirm consistent allele scoring, a selection of PCR products were re-scored and compared. GenAlEx v6.5 (Peakall and Smouse, 2012) was used to identify individuals sharing multi-locus genotypes. When more than one identical genotype (clone) was detected, only one individual from each clone was kept in the dataset. Genotyping errors were assessed in MICRO-CHECKER v2.2.3 (Van Oosterhout *et al.*, 2004). The program applies a Monte-Carlo approach to check for scoring errors due to stuttering and large allele dropout, and uses Hardy-Weinberg Equilibrium (HWE) to detect the presence of null alleles.

2.3.4 Species demarcation and hybrid identification

To determine the power of the markers in grouping the samples, a Principal Coordinates Analysis (PCoA) was carried out based on individual pairwise genetic distances in GenAlEx v6.5. The Bayesian clustering program STRUCTURE v2.3.4 (Pritchard *et al.*, 2000; Falush *et al.*, 2003) was used to assign individuals to clusters within the dataset. The program assigns individuals to a predefined number of clusters (*K*), based on the allele frequencies at each locus. In this case *K* = 2 was

chosen, with the assumption that the two species, *T. cordata* and *T. platyphyllos,* represent genetically distinct taxonomic units. STRUCTURE parameters were kept at the default settings, with a burn-in of 10^4 and MCMC iterations of 10^5. Runs with different values of *K* values (ranging 1 – 5) were also tested with the default settings and the same parameters. Each run was replicated ten times. Evanno's *ΔK* method (Evanno *et al.*, 2005), implemented in STRUCTURE HARVESTER (Earl and vonHoldt, 2012), was used to determine if the assumption of $K = 2$ was justified. DISTRUCT (Rosenberg, 2004) was used to visualize the data.

Values of *q*, ranging from 0 – 1, describe how individuals are proportionally assigned to a particular cluster. For example, individuals with $q = 0.50$ mean that 50% of their genotype belongs to one cluster and 50% to another cluster, *i.e.* a hybrid. However, individuals can be unequally proportioned within clusters, so to identify parental species and hybrids a cut-off value is required. Other authors have reported threshold *q*-values of 0.10 (Neophytou, 2014) and 0.20 (Duminil *et al.*, 2006; Vähä and Primmer, 2006; Larcombe *et al.*, 2014). A threshold *q*-value of 0.20 was used to identify the parental species and hybrids as this is a balance between accuracy and efficiency (Vähä and Primmer, 2006). Individuals with *q*-values between 0.20 – 0.80 were considered to be hybrids. They were removed from the dataset before population structure analyses. To be confident in a 0.20 threshold, a *q*-value of 0.10 was also tested and results compared.

2.3.5 Genetic diversity and differentiation

After identifying and removing hybrids and duplicated genotypes, standard population genetic diversity statistics were obtained from GenAlEx v6.5. Mean number of alleles (N_A), effective number of alleles (A_E), mean observed (H_O) and expected heterozygosity (H_E) were calculated from populations with five or more remaining individuals. The Inbreeding Coefficients (F_{IS}) and significance of each population was determined using FSTAT v2.9.3.2 (Goudet, 1995).

Genetic differentiation was assessed using two measures, pairwise F_{ST} (Weir and Cockerham, 1984) and D_{est} (Jost, 2008), calculated in GENEPOP on the web v4.2 (Raymond and Rousset, 1995) and SMOGD (Crawford, 2010), respectively. D_{est} is an estimated value of actual differentiation (Jost, 2008), and was calculated using 1000 bootstrap replicates to generate 95% confidence intervals (CI). An Analysis of Molecular Variance (AMOVA) was performed in the program Arlequin v3.5.1.3 (Excoffier and Lischer, 2010), to determine the distribution of genetic

variation (among species, among populations within species, and within populations). Statistical tests for differences between diversity measures were carried out in Minitab® v17.1 (2013 Minitab Inc.).

2.3.6 Population structure

Within-species population structure was assessed using the program STRUCTURE v2.3.4. For *T. cordata, K* was set to range from 1 – 18 and for *T. platyphyllos*, *K* ranged from 1 – 10. Both the default settings and the LOCPRIOR model (Hubisz *et al.*, 2009) were used, with a burn-in of 5×10^4 followed by 5×10^5 MCMC iterations and each run was replicated 20 times. The Evanno ΔK statistic, estimated in STRUCTURE HARVESTER, was used to determine optimal *K*. Different runs for the same *K* were averaged using the program CLUMPP (Jakobsson and Rosenberg, 2007) following the Greedy algorithm with 1000 repeats, and the data were visualised in DISTRUCT. To assess if genetic distance was correlated with geographic distance (Isolation by distance), a Mantel test was run with 10^4 permutations on each species, implemented in GenAlEx v6.5.

2.4 Results

In total 412 *Tilia* samples from 27 populations were analysed following the removal of 138 duplicated genotypes (clones). MICRO-CHECKER revealed no evidence of scoring errors due to stuttering with the exception of locus *Tc31* in population TcBB and no large allele dropout. The program revealed homozygote excess at some loci suggesting the possible presence of null alleles. However, loci showing homozygote excess were different in different populations, with the exception of locus *Tc963* which showed homozygote excess in four *T. cordata* and three *T. platyphlyllos* populations (Appendix 2.2). Therefore, all analyses were carried out with and without this marker to see if the results differed. The effect was negligible and so results including the marker are shown.

2.4.1 Species demarcation and hybrid identification

Two clear groups with some intermediate individuals were revealed by PCoA analysis based on pairwise genetic distance (Fig. 2.2). The orange group consists of *T. platyphyllos* individuals while the blue group consists of individuals thought to be *T. cordata* and largely confirm previously documented reports (Natural England, 2014).

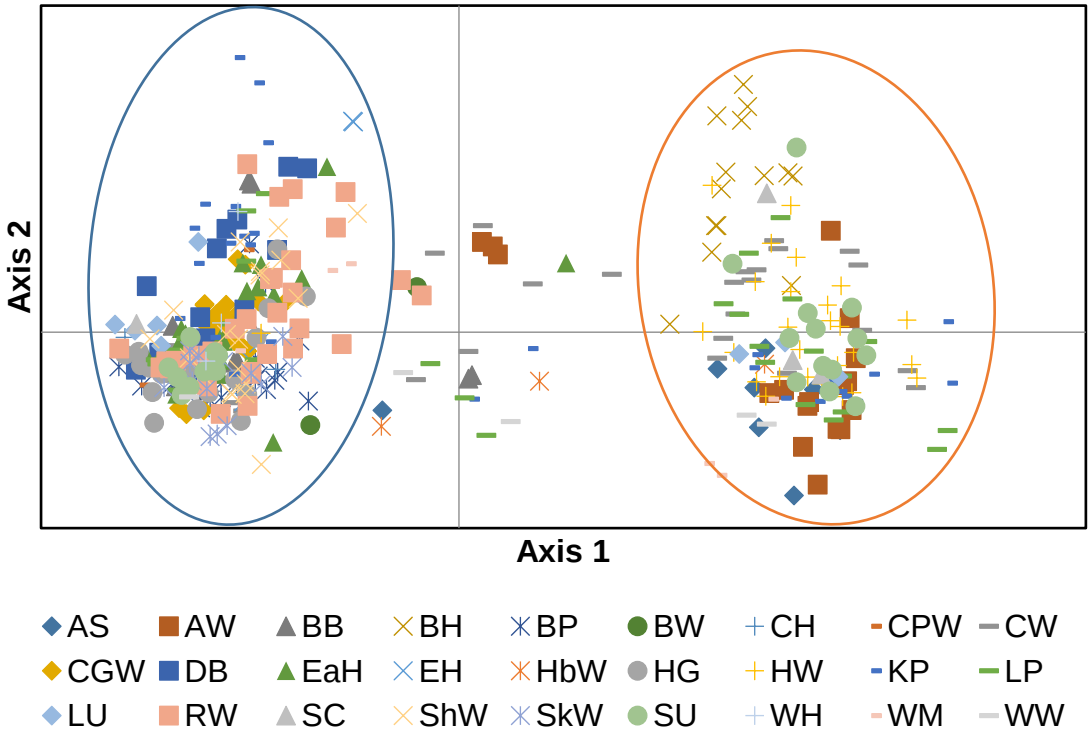


Figure 2.2 PCoA of *Tilia* individuals from 27 populations. Blue group individuals are *T. cordata* and the orange group are *T. platyphyllos*. Intermediate points are putative hybrids. Axes 1 and 2 explains 41% of the genetic variation (36% and 5%, respectively).

Two groups (K = 2) representing the two species were observed in STRUCTURE v2.3.4 (Fig. 2.3). Evanno's ΔK method also revealed that K = 2 was optimal (Appendix 2.3). Using a threshold q-value of 0.10, 32 individuals (7.8%), were determined hybrids, while using a q-value of 0.20, indicated 25 individuals to be hybrids (6.1%, Appendix 2.4). The 32 putative hybrid individuals had q-values for one parental species of between 0.105 and 0.866, while the 25 hybrid individuals had q-values for the same parental species of between 0.230 and 0.735 (Appendix 2.5). All loci were polymorphic with the exception of locus *Tc8* which was monomorphic in *T. cordata*. This locus was fixed in *T. cordata* at 141bp and ranged from 156 – 170bp in *T. platyphyllos* suggesting this locus has species specific alleles. Of the 32 putative hybrids using the 0.10 q-value threshold, 22 had the expected combination of *T. cordata* specific and *T. platyphyllos* specific alleles. Ten individuals were not a combination of *T. cordata* and *T. platyphyllos* alleles at this locus. Seven were homozygote with the *T. cordata* allele (141bp) and three were heterozygote *i.e.* had two of the *T. platyphyllos* alleles (156, 158, 160, 162,

164, 166 or 170bp). Of the 25 putative hybrids using the 0.20 threshold, two were homozygote at this locus with only the *T. cordata* allele and one was heterozygote with two *T. platyphyllos* alleles (Appendix 2.5). Two of the populations originally thought to be *T. platyphyllos* (WW and HbW), contained 19 (out of 20) and 11 (out of 13) *T. x europaea*, respectively.

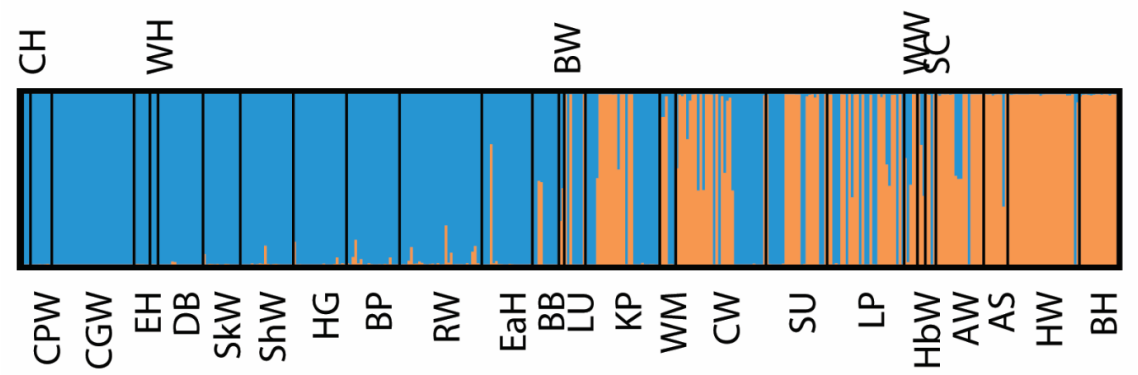


Figure 2.3 Assignment of 412 individuals from 27 populations with *K* = 2 (*T. cordata* – blue cluster and *T. platyphyllos* – orange cluster) inferred by Bayesian clustering analysis implemented in STRUCTURE, visualized in DISTRUCT.

2.4.2 Genetic diversity and differentiation

After removing hybrids, a total 380 (using *q*-value 0.10) and 387 (using *q*-value 0.20) *Tilia* individuals remained over 24 populations. The 13 microsatellite markers revealed high levels of polymorphism in both species, with the exception of locus *Tc8*. Diversity statistics were reported from populations with five or more individuals. The number of alleles per locus in the two species ranged from nine (loci *Tc8* and *Tc943*) to 33 (locus *Tc963*), a total of 206 alleles (Table 2.1). Overall *T. cordata* exhibited fewer alleles than *T. platyphyllos*, 127 and 149, respectively (Appendix 2.6). The number of alleles ranged from one (locus *Tc8*) to 26 (locus *Tc963*) in *T. cordata*, and from five (*Tc943*) to 18 (*Tc915*) in *T. platyphyllos*. More than twice as many private alleles were observed in *T. platyphyllos* than in *T. cordata* (21 and nine, respectively, Appendix 2.6). The average number of alleles per population (N_A) was significantly greater in *T. platyphyllos* (2 sample *t*-test, $P = 0.014$), and this was also reflected in the genetic diversity (H_E), which was higher in *T. platyphyllos* than in *T. cordata* (Table 2.2). Effective number of alleles (A_E) ranged from 1.93 (TcLu) to 3.54 (TcEaH) in *T. cordata* while observed (H_O) and expected (H_E) heterozygosity ranged from 0.34 (TcLu) to 0.59 (TcShW) and 0.35 (TcLU) to 0.57 (TcRW), respectively (Appendix 2.7). In *T. platyphyllos* these values were significantly higher ($P<0.05$),

with A_E ranging from 2.68 (TpAS) to 4.00 (TpSU), and H_O and H_E ranging from 0.66 (TpSU) to 0.78 (TpBH) and 0.60 (TpAS) to 0.73 (TpSU), respectively (Appendix 2.8). Inbreeding coefficients (F_{IS}) ranged from -0.14 to 0.14 and were not significantly different than zero (data not shown).

Table 2.1 Summary statistics within species and genetic differentiation within and between species. *A* - Number of alleles, H_O - Observed heterozygosity, H_E - Expected heterozygosity, F_{ST} and D_{est} values within *T. cordata* and *T. platyphyllos*, and differentiation values between species.

Locus	***T. cordata***					***T. platyphyllos***					***T. cordata and T. platyphyllos***		
	A	***H_O***	***H_E***	***F_{ST}***	***D_{est}***	***A***	***H_O***	***H_E***	***F_{ST}***	***D_{est}***	***A***	***F_{ST}***	***D_{est}***
Tc6	8	0.63	0.54	0.079	0.108	11	0.77	0.72	0.106	0.339	12	0.260	0.848
Tc937	6	0.44	0.41	0.077	0.054	9	0.68	0.61	0.121	0.253	13	0.404	0.935
Tc920	11	0.79	0.74	0.062	0.265	11	0.83	0.73	0.082	0.270	16	0.146	0.771
Tc8	1	0.00	0.00	0.000	0.000	8	0.71	0.63	0.087	0.194	9	0.734	0.987
Tc943	8	0.50	0.45	0.103	0.114	5	0.51	0.48	0.045	0.042	9	0.433	0.802
Tc31	6	0.27	0.33	0.104	0.055	10	0.65	0.61	0.142	0.313	13	0.416	0.890
Tc4	13	0.62	0.71	0.097	0.321	13	0.72	0.71	0.125	0.376	17	0.139	0.750
Tc927	4	0.03	0.03	0.533	0.128	15	0.74	0.72	0.089	0.282	17	0.620	0.978
Tc915	16	0.86	0.78	0.048	0.160	18	0.88	0.77	0.097	0.452	22	0.101	0.703
Tc963	26	0.77	0.86	0.046	0.439	15	0.65	0.69	0.133	0.486	33	0.088	0.816
Tc5	13	0.72	0.66	0.062	0.156	12	0.82	0.74	0.092	0.401	18	0.124	0.555
Tc951	7	0.56	0.53	0.095	0.131	9	0.52	0.67	0.093	0.267	11	0.234	0.687
Tc7	8	0.54	0.51	0.052	0.075	13	0.78	0.71	0.105	0.287	16	0.314	0.957
Mean	9.8	0.52	0.50	0.078	0.095	11.5	0.71	0.68	0.103	0.267	15.9	0.309	0.802

Table 2.2 Mean (and SE) values of within-population diversity measures from 16 *T. cordata* populations and eight *T. platyphyllos* populations. *N* – number of samples; N_A – Average number of alleles; A_E – Effective number of alleles; H_O – Observed heterozygosity; H_E – Expected heterozygosity.

Pop	***N***	N_A	A_E	H_O	H_E
T. cordata	15.38	4.36	2.90	0.52	0.51
	(1.93)	(0.23)	(0.10)	(0.01)	(0.01)
T. platyphyllos	14.5	5.45	3.46	0.71	0.68
	(1.84)	(0.33)	(0.15)	(0.01)	(0.15)

Genetic differentiation between populations within each species across all loci, measured as F_{ST} and D_{est} was higher in *T. platyphyllos* than in *T. cordata* (two-tailed Mann-Whitney *U* test, for F_{ST}: $P = 0.051$ and for D_{est}: $P = 0.006$, Table 2.2). Significant population pairwise F_{ST}, among the two species, was revealed by GENEPOP ranging from 0.305 (TcShW and TpBH) to 0.454 (TcLU and TpAS, Appendix 2.9).

There was high genetic differentiation between the two species, (28.54% of the total variation, AMOVA, $P < 0.001$). Among populations within species variation was 7.05% ($P < 0.001$) while the remaining variation (64.41%, $P < 0.001$) was found within populations (Table 2.4). Based on between species F_{ST} values loci*Tc8* and *Tc927* displayed the highest difference, 0.734 and 0.620, respectively (Table 2.1). While these loci also showed high differentiation based on D_{est}, the range of variation across D_{est} values was smaller than those across F_{ST} values.

Table 2.3 Analysis of Molecular Variance (AMOVA) showing the partitioning of genetic variation among species, populations and individuals.

Source of variation	**d.f.**	**Sum of Squares**	**Variance Components**	**Percentage of Variation**	***p* value**
Among species	1	862.15	1.77	28.54%	<0.001
Among populations within species	22	450.75	0.43	7.05%	<0.001
Within populations	922	3646.55	3.96	64.41%	<0.001
Total	945	4959.45	6.15		

2.4.3 Population sub-structure

Bayesian clustering analysis revealed population structure within *T. cordata*. When the default settings (data not shown) and the LOCPRIOR model was used, STRUCTURE v2.3.4 analysis and Evanno's *ΔK* method suggested *K* = 3 as optimal (Fig. 2.4). Populations CW, DB, EH, KP, LP, LU, ShW, and SU were assigned with large probability, to one group and population CPW also had high membership to this group. RW and SkW were assigned as a separate group with a high membership coefficient. Populations BP, EaH, and HG, although admixed, sharing membership with the first group, had a high membership coefficient with CGW, the third group. The remaining population, BB, is admixed with even probability of belonging to all three groups.

Likewise, there is clear population structure in *T. platyphyllos* (Fig. 2.5) When both models were used, STRUCTURE and Evanno's *ΔK* method suggested optimal *K* = 3. Populations AS, BH, and HW were assigned to different groups. Individuals from the other populations (AW, CW, KP, LP, and SU), were not assigned to a single cluster.

GenAlEx v6.5 revealed a non-significant negative correlation between genetic and geographic distance (Pairwise F_{ST}) in *T. cordata* (R^2 = 0.0291, *P* = 0.141). In *T. platyphyllos*, a non-significant positive correlation was observed (R^2 = 0.192, *P* = 0.06). This suggests no significant isolation by distance occurs in UK *Tilia*.

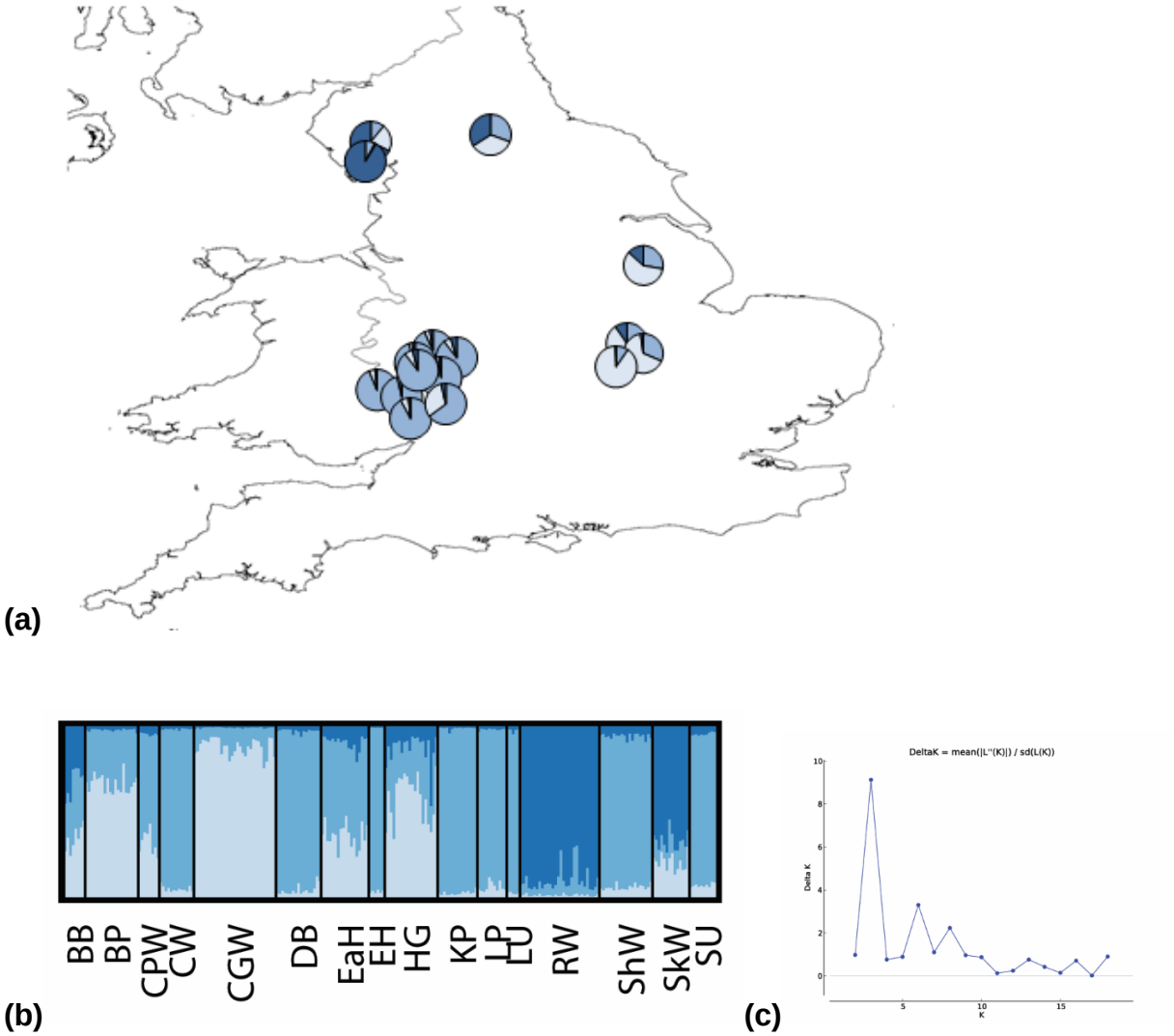


Figure 2.4 (a) Distribution of the genetic variation of 246 assigned individuals from 16 *T. cordata* populations with *K* = 3 inferred by Bayesian clustering analysis implemented in STRUCTURE, averaged in CLUMPP. (b) Assignment of *T. cordata* individuals when *K* = 3 visualized in DISTRUCT. (c) Evanno's ΔK revealing K = 3 in *T. cordata*, implemented in STRUCTURE HARVESTER.

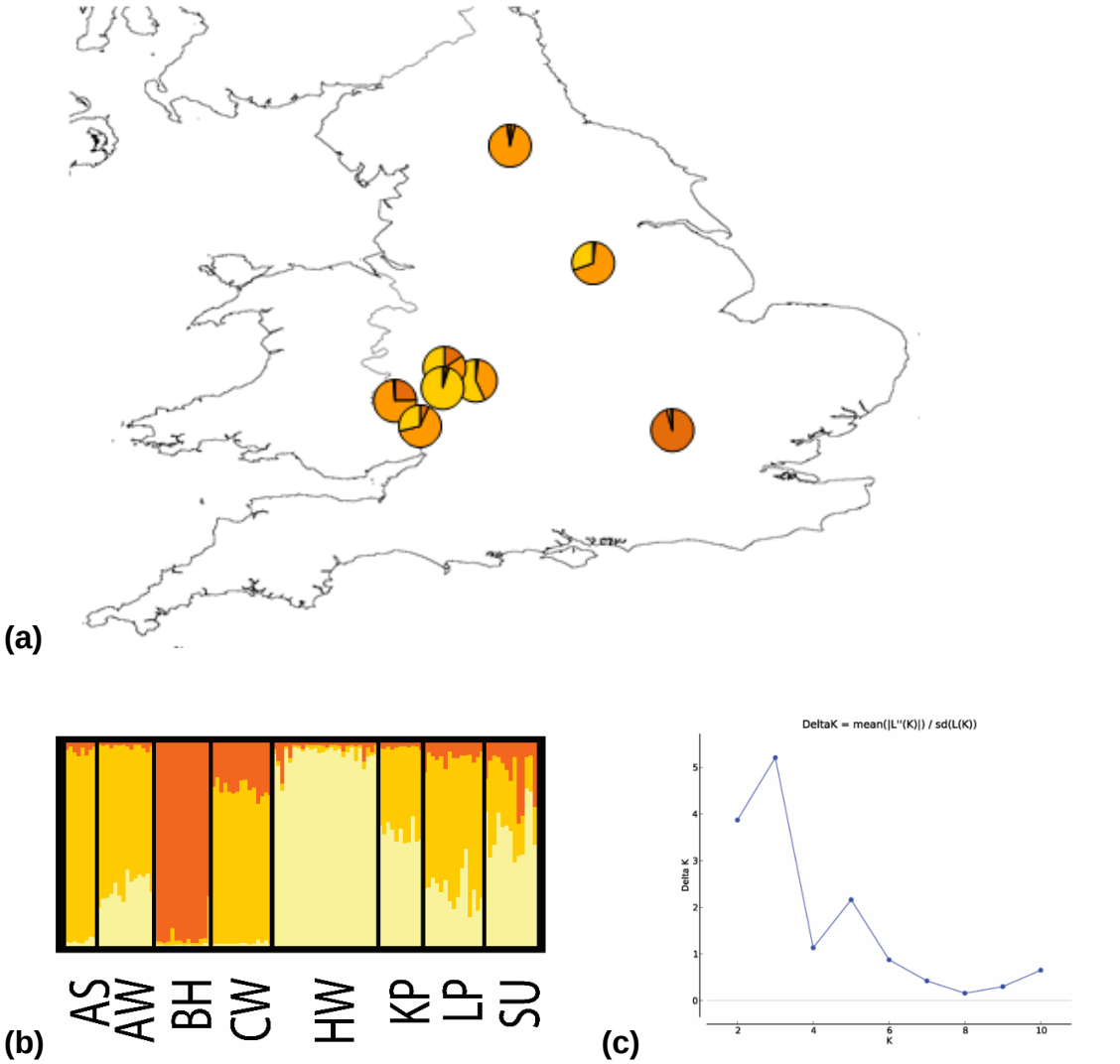


Figure 2.5 (a) Distribution of the genetic variation of 116 assigned individuals from eight *T. platyphyllos* populations when *K* = 3 inferred by Bayesian clustering analysis implemented in STRUCTURE, averaged in CLUMPP. (b) Assignment of *T. platyphyllos* individuals when *K* = 3 visualized in DISTRUCT. (c) Evanno's ΔK revealing K = 3 in *T. platyphyllos*, implemented in STRUCTURE HARVESTER.

2.5 Discussion

Genetic analysis of 246 *T. cordata* and 116 *T. platyphyllos* individuals showed that the two taxa are distinct species. While phylogenetic analysis shows that they are not sister species (Phuekvilai, 2014), this study has indicated that the two species naturally hybridise. This was clear in the PCoA and STRUCTURE analyses (Figs. 2.2 and 2.3), where several intermediate individuals were found.

2.5.1 Identification of pure species and hybrids from mixed populations

To determine whether an individual belongs to either species or is a hybrid two optimal threshold *q*-values were tested, 0.10 (Neophytou, 2014) and 0.20 (Duminil *et al.*, 2006; Larcombe *et al.*, 2014). In total, 92.2% and 93.9% of individuals could be assigned to either species (Appendix 2.4). The remaining 7.8% and 6.1% were considered to be hybrids or introgressed individuals. Although the difference between the two thresholds in this present study was negligible, a *q*-value of 0.20 might be a preferred option. Vähä and Primmer (2006), define three outcomes depending on threshold values chosen *i.e.* efficiency (highest proportion of correctly assigned pure species or hybrids), accuracy (the proportion of true species or true hybrids in the respective group) and performance (the result of considering both). The authors suggest that while *q*-values of 0.10 results in higher efficiency, when taking accuracy into consideration a *q*-value of 0.20 increases the overall performance.

STRUCTURE has generally been reported to do well in detecting hybrids from pure species (Duminil *et al.*, 2006; Larcombe *et al.*, 2014). For example, Larcombe *et al.* (2014), used a *q*-value of 0.20 and revealed 94-100% accuracy in detecting F_1 hybrids between six Eucalypts species. Similarly, Duminil *et al.* (2006), set a *q*-value of 0.20 to identify tropical rainforest trees (*Carapa* spp.) following a 'blind' survey *i.e.* without *a priori* information on taxonomy or morphology. STRUCTURE assigned over 90% of their samples to one of the *Carapa* species. To test the assignment power of STRUCTURE, Vähä and Primmer (2006), simulated a series of datasets consisting of different number of loci and divergence levels (F_{ST}). Their study revealed that hybrids can be identified with 12 – 24 loci and pairwise F_{ST} values of 0.21 – 0.12. This study used 13 variable markers analysed between species and observed F_{ST} values ranging from 0.305 to 0.454 (Appendix 2.9) and therefore should be sufficient for detecting hybrids in this instance.

2.5.2 High genetic diversity and differentiation in UK Tilia

Tilia cordata and *T. platyphyllos* showed a high degree of genetic diversity within and between species (Table 2.1). All samples showed a maximum of two alleles per locus, confirming that *T. cordata* and *T. platyphyllos* are both diploid species, 2n=2x=82 (Pigott, 2012). Diversity indices from the UK samples were marginally lower than those found in the two *Tilia* species from other European countries (Phuekvilai, 2014). This might be expected given that some UK populations are nearing the edge of their ecological range. Range edge populations often have less within-population genetic diversity due to genetic drift, limited gene flow, inbreeding and clonal reproduction (Caughley *et al.*, 1988; Eckert, 2002; Arnaud-Haond *et al.*, 2006) and show more genetic differentiation among populations than their central range counterparts (Beatty *et al.*, 2008).

With large distances between many of the fragmented populations, gene flow among populations is limited. Significant F_{ST} values confirm this. While the inbreeding coefficient (F_{IS}) values were not significant, suggesting little or no inbreeding, a high level of clonal trees (25%) was observed (data not shown). *T. cordata* is thought to reproduce more freely from asexual methods than production from seed (Radoglou et al., 2009 and references therein).

Overall, *T. cordata* exhibited less genetic diversity than *T. platyphyllos* (Tables 2.1 and 2.2). The difference in diversity might be a result of ascertainment bias as the markers were designed from *T. platyphyllos* individuals (Phuekvilai and Wolff, 2013). However, *T. cordata* does not freely regenerate from seed in many parts of its UK range and is rarely planted in native woodlands, while *T. platyphyllos* regularly has regenerated from seed and has been planted (Pigott, 1981b) and this may be an alternative reason why lower diversity in *T. cordata* is observed compared to *T. platyphyllos*. Another explanation is that cross-pollination from planted *T. platyphyllos* trees, originating from France and the Low Countries since c1500 (D. Pigott *pers. comm*.) may have increased diversity of UK *T. platyphyllos*. Furthermore, *T. platyphyllos* is thought to cross with planted *T. x europaea* as the two taxa flower simultaneously and share pollinators (D. Pigott *pers. comm*.).

Using the same set of markers, mean multi-locus H_E in *T. platyphyllos* is slightly lower than that Phuekvilai and Wolff (2013) found in two French populations with values of 0.74 compared to 0.68 (Table 2.2). In their study, locus *Tc943* was monomorphic in one population and showed only two alleles in the other. In this study, locus *Tc943* had the fewest alleles in *T. platyphyllos*, with only five found

across all eight populations. Mean multi-locus number of alleles within UK populations of *T. platyphyllos* was 11.5 (Table 2.1), compared to a mean of 9.1 in two French populations (Phuekvilai and Wolff, 2013).

MICRO-CHECKER showed UK populations are in Hardy-Weinberg equilibrium (HWE). Exhibiting high heterozygosity and non-significant F_{IS} values infers that both *Tilia* species within the UK are naturally outcrossing systems. Outcrossing woody species show high within-population genetic variation (Hamrick and Godt, 1996). While the AMOVA analysis revealed high differentiation between the species, a small but significant proportion of the total variation was found among populations within species and not surprisingly, the largest proportion of variation was found within populations (Table 2.3). While within-population variation was comparable, among-population variation was lower and among species considerably higher than for *Quercus petraea* and *Q. robur* (Kelleher *et al.*, 2005). While different markers reveal different estimates (Storfer *et al.*, 2010), greater species differentiation might be because the two *Tilia* species diverged earlier from each other than the two *Quercus* species. Significant F_{ST} was found both between and within species (Appendix 2.9) indicating two distinct biological units (Duminil *et al.*, 2006). Likewise, per locus F_{ST} and D_{est} were higher between species than within species (Table 2.1) revealing sufficient power of the markers in species delineation.

Phuekvilai and Wolff (2013), observed one locus (*Tc918,* not used in this study) did not amplify in *T.* cordata. So this marker is useful as a starting point in species delineation. However, amplification at this locus also occurs in the hybrids, therefore, it is less useful as a single identifying marker. This study has revealed other potential loci that can be used to delineate the species. Based on F_{ST}, locus *Tc8* is an obvious candidate. This locus is monomorphic within UK *T. cordata* and exhibits up to eight alleles in *T. platyphyllos*. Locus *Tc927* is another potential candidate because there are very few alleles at this locus within *T. cordata* and F_{ST} is high. Conversely, within *T. platyphyllos* the locus exhibits 15 alleles and F_{ST} is relatively low (Table 2.1). Large numbers of alleles at this locus have also been observed in two French *T. platyphyllos* populations (Phuekvilai and Wolff, 2013). Based on these values, loci *Tc8* and *Tc927* will be useful species discriminators. Other loci showing high between species variation are *Tc937, Tc943,* and *Tc31*, and will also be useful in distinguishing the two species.

2.5.3 Observed intra-specific structure in UK populations

Sub-structure within both *T. cordata* and *T. platyphyllos* was observed, (K = 3). All putative groups assigned by STRUCTURE within each of the species can loosely be split by geographic region and membership within each putative group is not unexpected due to the close proximity of the respective sites.

Within *T. cordata*, the first group (CPW, CW, DB, EH, KP, LP, LU, ShW, and SU) are west Midland sites, close to the Welsh border (Figs. 2.1 and 2.4). These populations are interesting because, with the exception of CPW, there is very little admixture within each population. This might be related to the regeneration and dispersal ability of *T. cordata* in this area. While the species can produce fertile seeds in this part of England, the conditions of the wood must be favourable for seedlings to grow *i.e.* open canopy, little competition and limited predation (Pigott, 2012).

The second group, (RW and SkW), are North West UK sites in relative isolation and there is evidence that RW has been managed as a coppiced wood for 300 years (Cottrell *et al.*, 2003). While this could effectively restrict the potential for sexual recruitment, *Tilia* has frequently set seed and regenerated at this site. Conversely, *Tilia* within SkW, due to ecological conditions, has not reproduced from seed for at least a century (Pigott, 2012). Populations RW and SkW might have ancestral genotypes that have been maintained due to isolation and although recruitment occurs at RW, the large number of *T. cordata* found there are continuing to maintain the ancestral genotypes and genetic drift does not have a significant effect on allele frequencies.

The remaining populations (BB, BP, CGW, EaH and HG) do not clearly constitute a single group. Three of these (BP, EaH, and CGW) are further to the east, but still within the central UK range. They are remnants of the old Royal Rockingham Forest (Peterken and Welch, 1975), and are within close proximity of one another. Population HG, which is part of the Bardney Limewoods complex, is also relatively close to the three sites, so genetic similarity is not unexpected. Populations CGW and EaH are adjacent woodlands and would be expected to be genetically similar, but EaH is admixed with similar genotypes from the other groups and significant F_{ST} between the two is observed (Appendix 2.9). Many large maiden trees once occurred at EaH in the early 20th Century but have been destroyed (Pigott, 1991). This may have provided the optimal conditions for young trees to grow and so may explain the difference in genetic composition.

Three interesting sites assigned to three genetic groups, are clear from the *T. platyphyllos* analysis. Populations AS, BH, and HW represent the three assigned groups with very little admixture. All the other sites are assigned to different clusters. These sites are at the north (AS), west (HW) and central (BH) of the UK *T. platyphyllos* range (Figs. 2.1 and 2.5). AS, which is in the first genetic group, is a relatively isolated population compared to the other sites. It is the most northerly native UK population, and is the most north westerly natural European *T. platyphyllos* population (Pigott, 2012). The limes are scattered across old coppice woodland on both the upper and lower part of a steep rocky limestone cliff (T. Laurie *pers. comm.*). Similar genotypes to those in AS were also observed in population CW which is close to the Welsh border. These may be ancestral genotypes that have been maintained within these populations.

BH, in the second genetic group, is old managed woodland. Pigott (2000), reported that a single *T. platyphyllos* tree at this site may be over 450 years old, and another over 200 years old. This population is more distinct than other *T. platyphyllos* populations with more private alleles (Appendix 2.6) and the highest F_{ST} values (Appendix 2.9), explaining the greatest genetic differentiation among populations. Although there were younger individuals in this population, many of the trees at this site are large and multi-stemmed. Multiple stems from the base of *Tilia* are usually a sign of past coppicing, either naturally or managed. Coppicing was, and still is, an important part of the management of *Tilia* species (Pigott, 2012), and may have resulted in the large present day differentiation from other *T. platyphyllos s*ites. Furthermore, BH is relatively isolated, much like AS, and so gene exchange with other populations is expected to be limited. However, sexual recruitment of *Tilia*, although restricted by roe deer, is more prevalent in southern populations *e.g.* BH, where temperatures during flowering can be up to 2°C higher than at northern populations *e.g.* AS (D. Pigott, *pers. comm.*) thus may be influenced by more optimum ecological conditions.

Population HW, in the third genetic group, is a large woodland close to the Welsh border. Although documented as having both species and natural hybrids present (Natural England, 2014), all but one sample collected and genotyped from this population were *T. platyphyllos*. Other populations (KP, LP and SU), have a genetic affinity with HW and this is not unexpected due to their close proximity. However, population AW is further north, and while these could be ancestral genotypes, human assistance, due to the planting of the species in parts of its

range might be a factor in the genetic similarity of these *T. platyphyllos* populations.

2.6 Conservation implications

Although some populations are small and isolated, genetic diversity remains high. However, if these ancient woodland sites further fragment, diversity could eventually be eroded. Intra-specific structure, possibly from historical fragmentation, has resulted in three genetic groups within each species. At present, the two UK species are considered Least Concern on the IUCN Red List of Threatened Species. However, some sites are regarded as important UK *Tilia* populations, *e.g.* BH (Natural England, 2014) and *Tilia cordata* is currently listed as a high priority species for forestry on the SilviFuture Database (SilviFuture, 2015). Therefore, efforts to maintain stands should be considered. The loss of ancient woodlands within the UK is a growing concern and many *Tilia* populations are facing further size reductions. In wake of future climate change, having an understanding of the status of the genus will now enable us to inform effective decision making regarding their conservation by concentrating on populations with the highest diversity, and detecting truly endemic genotypes. Restoration of small populations should introduce lime trees from woodlands that group together *i.e.* trees with similar genotypes. Introducing genotypes from woodlands that group together rather than from genetically distinct populations may reduce outbreeding depression. This is when progeny, derived from crossed individuals originating from genetically different populations, exhibit lower fitness than the parent stock. Cutting back or removing other tree species that are in abundance or encroaching shrubs will provide suitable areas for *Tilia* and will increase the biodiversity at certain sites. These efforts will ensure that the unique and ecologically important role of *Tilia* remains well established in the UK landscape.

2.7 Conclusion

This study showed the power of microsatellite markers in successfully distinguishing between the two UK *Tilia* species and estimating genetic diversity and differentiation. Using 13 variable markers, the two species can be confidently defined and distinguished from the hybrid. Although, it was not the focus of this study to explain the extent of hybridisation in *Tilia*, it is clear that some level of genetic introgression does occur, or at least has occurred in the past.

The genetic data confirm that *T. cordata* and *T. platyphyllos* are outcrossing and diploid. Both species have high genetic diversity. The two species show high genetic differentiation from one another, indicating two distinct biological units.

Evidence of intraspecific structure and inter-population differentiation in both species is apparent. Based on the genetic markers used in this study, this structure is loosely related to location. Although no significant isolation by distance occurs, provenance issues must be considered. Future conservation and management of certain sites should avoid restocking lime woodlands with *Tilia* trees that are genetically different to the standing crop.

Chapter 3: Genetic diversity and demographic history of the Siberian lime (*Tilia sibirica* Bayer), an important conservation unit, and its congeneric *T. cordata* (Mill.).

3.1 Abstract

Tilia sibirica Bayer (Siberian lime), is endemic to the Altai Mountain region of southern Siberia, approximately 1,500 km to the east of the natural range of *T. cordata.* Two centuries of intense logging may have had significant genetic consequences on populations of the species. However, to date no genetic analyses have been carried out, leaving questions unanswered regarding its genetic status.

This study uses standard population genetic and Approximate Bayesian Computation (ABC) analyses to determine the genetic and demographic history of the Siberian lime and that of its closest congeneric, the small-leaved lime (*T. cordata*) from Siberia, Poland and Austria.

The results show that the two taxa are distinct species with significant genetic differentiation and further suggests their relatively recent (early Holocene) divergence. Further ABC analyses of European and Siberian *T. cordata* suggest that the Holocene split coincided with a westerly migration of *Tilia* genotypes that may have contributed to the recolonization of *T. cordata* in Europe.

While the study suggests that logging since the 19th Century appears not to have had a significant effect on the current effective population size, it has revealed low within-population genetic diversity. Therefore, future conservation and management efforts should consider maintaining populations with the highest genetic diversity and restoring less diverse populations by reseeding and/or replanting from other locations. This could be carried out *in situ* or *ex situ* but will need to be a collaborative effort to include contributions from geneticists, ecologists, forest managers, and government parties.

3.2 Introduction

Tilia sibirica Bayer (Siberian lime) is considered to be an endemic relict species of southern Siberia (Novák *et al.*, 2014), but is regarded as a subspecies of *T. cordata*, namely, *T. cordata* subsp. *sibirica* by Pigott (2012). However, the assignment as a subspecies is based on few specimens Pigott (2012). This current study recognizes the taxon as a separate species to *T. cordata.*

The two species do not occur in sympatry naturally and while their morphologies are broadly similar, there are subtle differences in their leaves, ovaries, and stigma (Pigott, 2012). Intense logging throughout the 19th and 20th Century resulted in the fragmentation of the southern Siberian forests (Novák *et al.*, 2014) but little is known of the consequences this may have had on genetic diversity in *T. sibirica* populations because, to date, no genetic analyses have been carried out on the Siberian lime.

Krylov (1891, 1902) and other references cited in Novák *et al.* (2014) regard Siberian *Tilia* as part of the Tertiary relict flora, based on phytogeographical evidence. So the ancestral lineage of the current linden species within the region may date back several millions of years. *Tilia*-type pollen has been recorded from deposits dating back to the Tertiary period (Mai 1961 in Pigott 2012). The genus is well represented in the Siberian pollen record from the last interglacial (125 kya) to before the Last Glacial maximum (LGM) which ended ~18 – 19 kya (Peterson, 1983; Krementski, 1997; Tarasov *et al.*, 2005; Kuneš *et al.*, 2008; Markova *et al.*, 2009; Tarasov *et al.*, 2009; Bolikhovskaya and Shunkov, 2014). Much of the *Tilia* pollen found in Siberia has been identified at latitudes 60° - 65°N. A single *Tilia* pollen grain has even been found as for north as the Laptev Sea (74°N), in deposits dated 7.2 kya (Naidina and Bauch, 2001). Members of the genus may therefore have once persisted closer to this latitude in Siberia. While long distance dispersal cannot be ruled out, pollen spores are unlikely to have travelled far, as *Tilia* pollen is poorly dispersed (Pigott, 2012).

While *Tilia* is frequently mentioned in palynology studies, it is usually only identified to genus. Whereas *T. cordata* and *T. platyphyllos* pollen are distinguishable (Andrew, 1971), it can generally be difficult to identify pollen to species (Nakazawa *et al.*, 2013). However, based on their palynological study, Bolikhovskaya and Shunkov (2014), have reported *Tilia* species being present in southern Siberia during glacial and interglacial periods. Although there is little detail on how pollen spores were distinguished to species level, it can be accepted

that *Tilia* has had a long persistence in (various regions of) Siberia and that it has indeed existed there since the Tertiary period. However, we do not know how or when species got to where they presently are.

The accepted paradigm is that species retreated or were lost as ice advanced throughout glacial periods (Bennett *et al.*, 1991). Many of those species that exist today at higher latitudes are believed to have survived in southern refugia – the southern refugia hypothesis – and then expanded their range along particular migratory routes when the ice retreated. The post-glaciation migratory routes of broad-leaved forest trees are well documented and currently three main southern refugia are accepted, namely the Balkans, Italy and Iberia (Huntley and Birks, 1983; Bennett *et al.*, 1991). Species were able to survive at these sites during the maximum extent of glacial periods, as the climate was more suitable there (Bennett et al., 1991). A consequence for the surviving species following long-term isolation, expansion and subsequent recolonization is greater genetic differentiation among populations at different latitudes and greater genetic diversity in refugia compared to that of the newly colonized areas (Petit *et al.*, 2003), the signal of which can be found in phylogeographic studies (Avise *et al.*, 1987; Cavers *et al.*, 2003; Provan and Bennett, 2008; Beatty and Provan, 2013; Beatty and Provan, 2014).

Based on pollen analysis, Huntley and Birks (1983), reported the expansion of *Tilia* across Europe from Italy and Greece beginning ~18 kya, near the end of the LGM, and by 13 kya several species had expanded from the east (Fig. 3.1). While broad-leaved trees disappeared from much of Europe during the LGM, some species, including *Tilia*, may have survived in small pockets of suitable habitat outside the main southern refugia. Pollen spores have been recorded in deposits of this time (18 – 21 kya) in Moldova, the Azor region north of the Black Sea, and the Dnieper in northern Ukraine (Markova *et al.*, 2009).

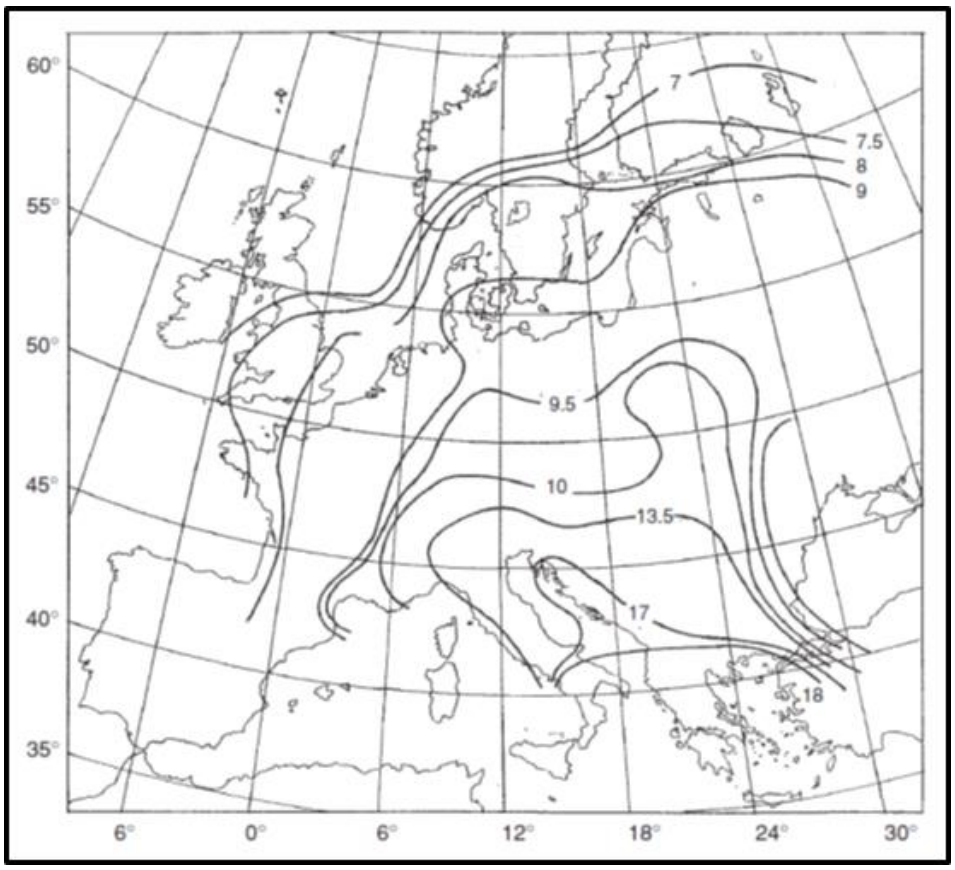


Figure 3.1 Post-LGM migration of *Tilia* across Europe from Italy and Greece. Contour lines represent kilo-years ago (kya). Map sourced from Huntley and Birks (1983).

A recent phylogeographic study using both microsatellite and chloroplast DNA showed that *Tilia* likely followed known post-glacial colonisation routes. However, no molecular dating was performed (Phuekvilai, 2014). In addition, based on the microsatellite markers, the study also revealed possible refugia in eastern Europe. This is not surprising given that Markova *et al.* (2009) reported *Tilia* pollen deposits from the intercontinental borders of eastern Europe and Middle east Asia. Furthermore, the Caucasus and the Caspian Sea have also been considered to be putative ice-age refugia where several plant and tree species expanded their range (Hewitt, 1999).

To date, no genetic analyses have been carried out on Siberian *Tilia cordata* or the Siberian lime *T. sibirica*. To infer the extent of genetic diversity and differentiation of the two, *Tilia*-specific microsatellite markers (Phuekvilai and Wolff, 2013) were used in this study. These markers revealed high genetic diversity and population structure in UK *T. cordata* and *T. platyphyllos* (Chapter 2)

and in central European stands (Phuekvilai, 2014). Moreover, they were suitable for distinguishing between the two species and their hybrid.

Microsatellite markers have a wide utility in population, evolutionary and conservation genetics. Therefore, using such markers in this study three hypotheses were tested; (1) given their geographic and ecological isolation from one another, *T. cordata* and *T. sibirica* show significant genetic differentiation and are two distinct species; (2) the divergence of the two species from a common ancestor predates the LGM; and (3) given its small and fragmented nature, *T. sibirica* show relatively low genetic diversity in comparison to *T. cordata* populations. In addition, effective population sizes (N_e) of *T. sibirica* and *T. cordata* were estimated and putative recolonization routes of *T. cordata* from Eastern Europe – Western Siberia explored. Answering these fundamental questions greatly advances our understanding of two Siberian relicts and could be used to advise on suitable conservation measures.

3.3 Material and Methods

3.3.1 Study sites and sample collection

Two *Tilia cordata* populations from the Vagay region and six *T. sibirica* populations from Kuzedeevo, Kemerovo region in Russia were sampled for genetic analyses (Table 3.1). Details of the sampling procedure used for the collection of the Russian samples are described in Novák *et al.* (2014).

T. cordata samples were collected from the Białowieża National Park (BNP), Poland. Three sites that had previously been identified to have *Tilia* present were chosen (Pigott, 1975). To reduce activity and minimise disturbance under the canopy, a 30x30m quadrat was set out within each site and almost all *Tilia* within that area was sampled. *Tilia* in Eastern Europe can reach heights of 35-50m (Pigott, 1975; Jaworski *et al.*, 2005; Wesolowski and Rowinski, 2006). With trees of this height, branches and leaves are often difficult to reach. For some of the trees within the quadrats, leaf samples could not be collected. Additionally, one *T. cordata* population from Austria was also sampled for comparative genetic analysis. All *Tilia* samples were dried at room temperature and stored at -20°C until required for DNA extraction.

Table 3.1 Species, country, sites sampled, population codes, latitude and longitude coordinates of *Tilia cordata* and *T. sibirica*.

Species	Country	Locations	Code	Latitude (°N)	Longitude (°E)
T. cordata	Austria	Stams[1]	AST	47.2757	10.9772
	Poland	Białowieża[2]	B69	52.7289	23.8328
			B99	52.7186	23.8443
			B40	52.7336	23.8319
	Russia	Vagay Region[3]	V20	57.5097	69.1956
			V25	57.9304	68.9099
T. sibirica	Russia	Kuzedeevo[3]	K12	53.17-53.21	87.13-87.21
			K21	53.17-53.21	87.13-87.21
			K22	53.17-53.21	87.13-87.21
			K28	53.17-53.21	87.13-87.21
			K29	53.17-53.21	87.13-87.21
			K38	53.17-53.21	87.13-87.21

[1] Collected in 2013 by Dr Kirsten Wolff; [2] Collected in July 2014 by Samuel Logan; [3] Collected in July and August 2012 by Dr Milan Chytrý and colleagues (Masaryk University, Czech Republic).

3.3.2 DNA extraction and amplification

Genomic DNA was extracted from leaf tissue using the CTAB method (Morgan-Richards and Wolff, 1999). A multiplex Polymerase Chain Reaction (PCR) procedure was carried out to amplify twelve microsatellite regions (Phuekvilai and Wolff, 2013). PCR conditions and parameters were as described in Phuekvilai and Wolff (2013). Microsatellites were genotyped using an ABI 3130XL Genetic Analyser (Applied Biosystems), and scored using Genemapper (Applied Biosystems). Microsatellite fragments were binned manually and checked for inconsistencies.

3.3.3 Standard population genetic analyses

Samples were checked for deviations from Hardy-Weinberg Equilibrium (HWE) using GENEPOP on the web v4.2 (Raymond and Rousset, 1995; Rousset, 2008) and tested for scoring errors, stuttering and large allele dropout in MICRO-CHECKER v2.2.3 (Van Oosterhout *et al.*, 2004). Null alleles rates were estimated per locus using FreeNA (Chapuis and Estoup, 2007). Summary statistics such as expected heterozygosity (H_E), total number of alleles (A), number of private alleles

(*Ap*) and average number of alleles (N_A) were calculated in GenAlEx 6.5 (Peakall and Smouse, 2012). Allelic richness (A_R), based on a minimum of six diploid individuals, was calculated in FSTAT v2.9.3.2 (Goudet, 1995; Goudet, 2001). Statistical tests of differences were carried out in R v3.1.3 (R Development Core Team 2015, www.r-project.org).

Genetic differentiation between the two species was estimated using a number of different diversity measures *i.e.* F_{ST} (Weir and Cockerham, 1984), D_{est} (Jost, 2008), G_{ST_est} (Nei and Chesser, 1983), and G'_{ST_est} (Hedrick, 2005), in Genepop on the web v4.2 and SMOGD v1.2.5 (Crawford, 2010) respectively. D_{est} is an estimate of actual differentiation, G_{ST_est} is a nearly unbiased estimate of relative differentiation, and G'_{ST_est} is a standardised estimate of genetic differentiation. These were calculated using 1000 bootstrap replicates generating 95% confidence intervals.

Population differentiation (pairwise F_{ST} and their significance) was calculated in FSTAT v2.9.3.2. Adjusted values were subsequently used to re-calculate expected heterozygosity and F_{ST} (without significance level – both sets of values are reported). To determine the distribution of genetic variation – among species, among populations within species, and within populations –an analysis of molecular variance (AMOVA), implemented in Arlequin v3.5.1.3 (Excoffier and Lischer, 2010), was used.

To test whether *T. cordata* and *T. sibirica* individuals form distinct genetic groups, a Principal Coordinates Analysis (PCoA) based on individual pairwise genetic distance, was performed in GenAlEx and a Bayesian genetic clustering analysis performed in the program STRUCTURE v2.3.4 (Pritchard *et al.*, 2000; Falush *et al.*, 2003). STRUCTURE assigns individuals to a predefined number of clusters (*K*), based on allele frequencies at each locus. In this case *K* was set to range from 1 – 6 (*T. sibirica*, *T. cordata* Siberia, Poland, and Austria plus two). STRUCTURE parameters were kept at the default settings, with a burn-in of 10^4 and MCMC iterations of 10^5. Each run was replicated ten times. Model selection relied on the Evanno ΔK statistic (Evanno *et al.*, 2005) estimated in STRUCTURE HARVESTER (Earl and vonHoldt, 2012). DISTRUCT (Rosenberg, 2004) was used to visualise the data.

3.3.4 Approximate Bayesian Computation (ABC) analyses

An Approximate Bayesian Computation (ABC) approach was implemented in DIYABC v2.1.0 (Cornuet *et al.*, 2008; Cornuet *et al.*, 2014) to infer historical demographic scenarios of *Tilia* in Siberia. Effective population size (N_e) and the time of divergence (*td*) were estimated for *T. cordata* and *T. sibirica,* as well as testing possible refugia and migration route hypotheses. The program uses a suite of summary statistics to test genetic data based on a set of user defined historical parameters. DIYABC has been used to test various evolutionary and demographic hypotheses (Poudel *et al.*, 2014; Bagnoli *et al.*, 2015; Havrdová *et al.*, 2015). It is also able to test un-sampled or hypothetical populations as mentioned in the DIYABC v2.1.0 manual, available at: http://www1.montpellier.inra.fr/CBGP/diyabc (Estoup *et al.*, 2015).

For the first analysis, the simple divergence model (SDM, Fig. 3.2), *i.e.* the divergence of *T. cordata* and *T. sibirica* from a common ancestor, four historical parameters were set (Table 3.2). The effective population sizes (*N1* and *N2*) for each species ranged from 10 – 20,000. One condition was set regarding N_e of *Tilia sibirica* (*N1*) and *T. cordata* (*N2*), that is, both were smaller than or equal to the ancestral population size (*Na*). Time of divergence (*td*), was set as minimum 10 generations to maximum 100,000 generations. Assuming a generation time of 25 years (Collingham and Huntley, 2000; Pigott, 2012), that translates to 0.25 kya – 2.5 Mya. A Generalised Stepwise Mutation (GSM) model was assumed and a mean microsatellite mutation rate of 10^{-4} to 10^{-3}, was drawn from a uniform distribution. All default settings remained in place with the exception of the mean single nucleotide insertion (SNI) rate, which was changed to uniform distribution.

Table 3.2 Historic parameters used for the simple divergence model (SDM), bottleneck model (BM) and the expansion/migration model (EMM). *N#* – Effective population size during respective time period (*i.e. t3, t2, t1*); *t#* – time of historic event (divergence/bottleneck/expansion/migration); *Na/Na#* – Effective population size of ancestral population during respective time period (*i.e. t3, t2, t1*); *ra* – admixture rate.

Taxa/Parameter	SDM (min – max)	BM (min – max)	EMM (min – max)
T. cordata N# (Siberia)	10 – 10,000	–	10 – 10,000
T. cordata N# (Poland)	–	–	10 – 10,000
T. cordata N# (Austria)	–	–	10 – 10,000
T. sibirica N#	10 – 10,000	10 – 20,000	–
Ancestral population *Na*	10 – 50,000	10 – 50,000	10 – 50,000
td (in generations)	10 – 100,000	–	–
t3 (in generations)	–	720 – 100,000	10 – 10,000
t2 (in generations)	–	10 – 720	10 – 10,000
t1 (in generations)	–	1 – 10	10 – 10,000
Nai	–	10 – 50,000	–
Naii	–	10 – 50,000	–
Naiii	–	10 – 50,000	–
ra	–	–	0.001 – 0.999

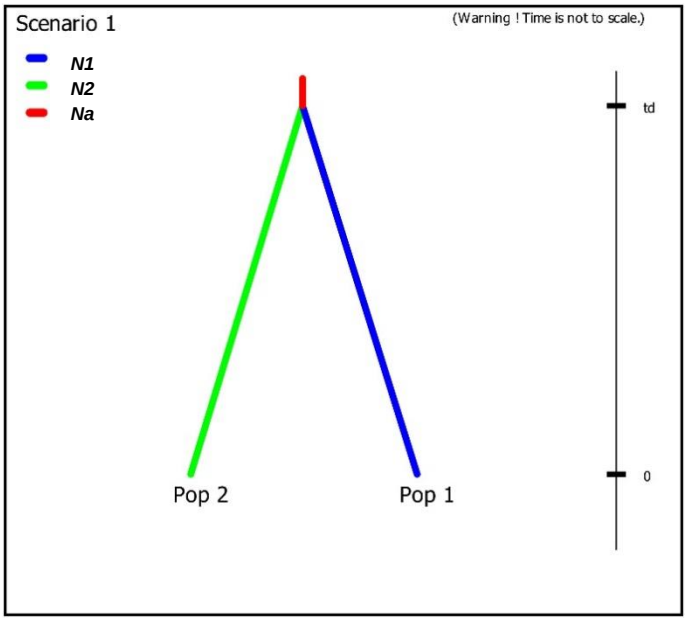


Figure 3.2 The simple divergence model tested in DIYABC: *0* – time of sampling; *td* – time of divergence; *N1* – Effective population size of pop 1 (*T. sibirica*) during the time period *td* - 0; *N2* – Effective population size of pop 2 (*T. cordata*) during the time period *td* – 0; *Na* – Effective population size of ancestral population (pre-time *td*).

An option within the DIYABC program (pre-evaluate scenario prior combinations) allows scenarios and priors to be initially tested. A principal component analysis (PCA) is performed on 100,000 simulated data sets and a method that ranks each statistic of the observed data set against that of the simulated data. If the simulated data are significantly different to the observed (Appendix 3.1 and 3.2) then that set of summary statistics may not be best suited in describing the priors for the scenario (Estoup *et al.*, 2015). An initial run using all 16 available summary statistics was used to determine the optimum set. The summary statistics that showed significant differences between the simulated datasets and the observed were subsequently omitted (Appendix 3.2). This resulted in a total of ten summary statistics recorded to build the SDM reference table (Appendix 3.3). One-sample statistics included: mean number of alleles (*A*), genic diversity (*H* - Nei, 1987), and allele size variance across loci (*V*). Two-sample statistics included the mean number of alleles (*A2P*), allele size variance (*V2P*), F_{ST} (Weir and Cockerham, 1984) and $(\delta\mu)^2$ distance between two samples (Goldstein *et al.*, 1995). These statistics are referred to as the optimum set throughout this study. A total of 1 x 10^6 datasets were simulated. The option 'model check' was used to assess the goodness-of-fit. The analysis is similar to

the ‘pre-evaluate option’, in that it simulates data sets and compares those to the observed data. However, on this occasion the posterior distributions of parameters are used rather than the prior distributions. As before, a PCA (this time performed on 1 x 10^6 data sets) and a ranked approach are used. The goodness-of-fit PCA graphs presented visualise 10,000 (1%) simulated data sets.

Following the SDM a bottleneck model (BM) was run with only the *T. sibirica* samples to determine if the species experienced historical population size fluctuations. Assuming a reduction in effective population size (N_e) following a split from a common ancestor, eight historic parameters (Table 3.2) and four scenarios (Fig. 3.3) were tested.

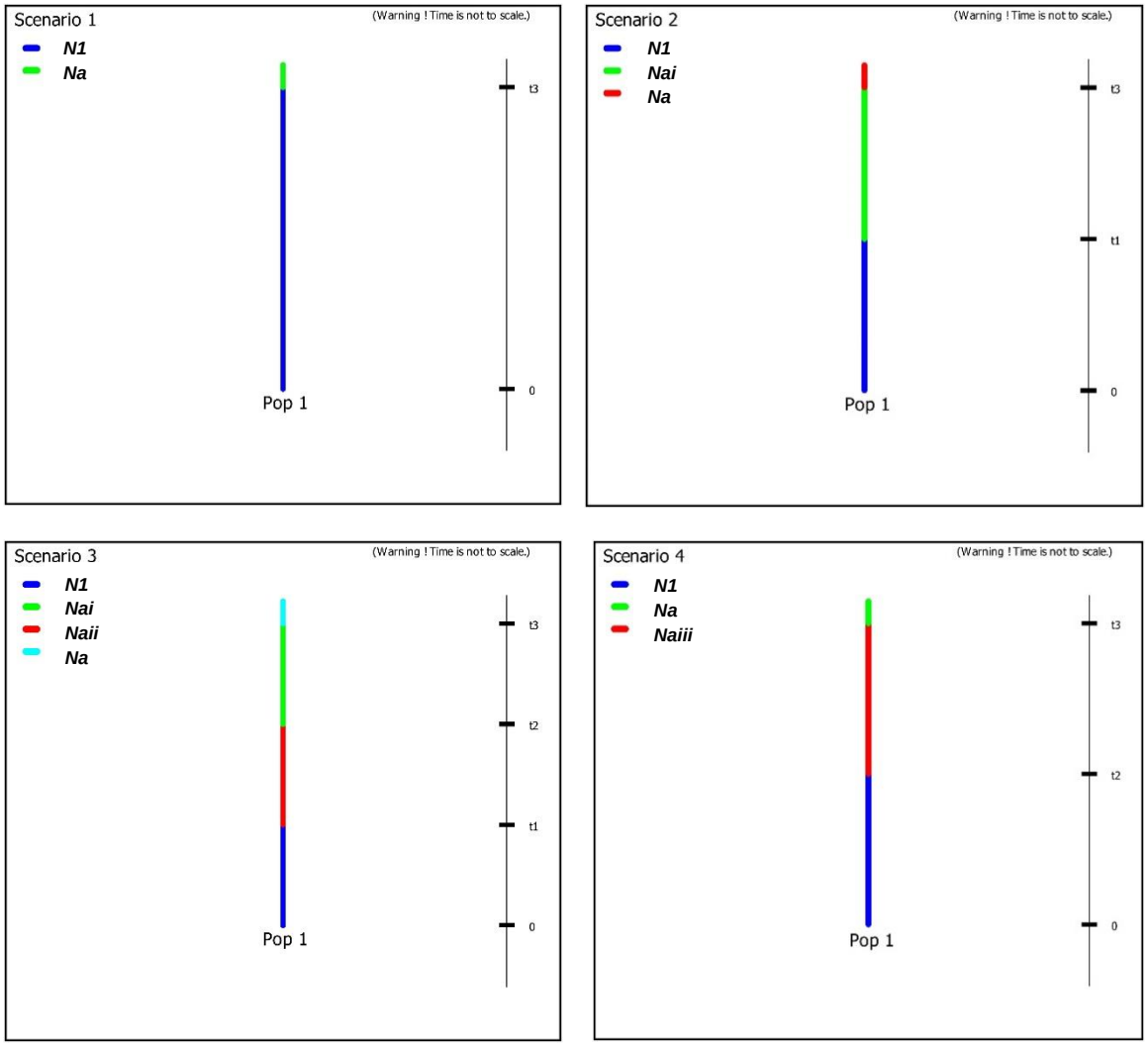


Figure 3.3 The bottleneck model (BM) tested in DIYABC: *0* – time of sampling; *t#* – time of demographic event; *N1* – Effective population size of pop 1 (*T. sibirica*) during the time period *t1* - 0; *Na*/*Na#* – Effective population size of ancestral populations during the time period *t3* - *t1.*

Scenario 1: A time *t3* (720 – 100,000 generations ago *i.e.* 18 kya – 2.5 Ma), there was a reduction in N_e but has remained constant since. **Scenario 2:** Same reduction at time *t3* with a further reduction at time *t1* (1 – 10 generations ago *i.e.* 19th Century logging (Novák *et al.*, 2014). **Scenario 3:** Same N_e reduction at time *t3* but an increase of N_e at time *t2* after the LGM (10 – 720 generation ago *i.e.* 0.25 – 18 kya), and at time *t1* another reduction occurred following intensive logging. **Scenario 4:** Same N_e reduced at time *t3* but increased at time *t2,* (after the LGM) and then remained constant despite intense logging in the 19th Century.

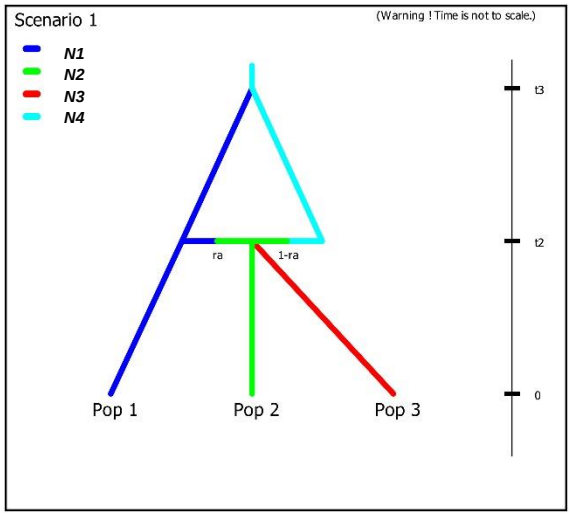
Eleven conditions were set for this model regarding population sizes and historical events (Appendix 3.4). In total 4×10^6 data sets were simulated, that is 1×10^6 for each scenario. As this is a single population analysis, only the one-sample statistics from the 'optimum set of summary statistics' (see above) were used. A GSM model was assumed and all the same settings as described above were used. Posterior probabilities were estimated to compare scenarios using two methods; a direct estimate and a logistic regression. For the direct approach, 1,000 simulated data sets (those closest to the observed) were selected, while for the logistic regression, 40,000 data sets were selected. The scenario and respective parameters with the highest posterior probability was subsequently considered to be the best scenario. Confidence in scenario choice was determined by evaluating type I and type II errors (Cornuet *et al.*, 2010), *i.e.* the number of times the 'best' scenario did not show the highest posterior probability compared to the other scenarios when it is the true scenario (type I) and the number of times the 'best' scenario did show the highest posterior probability compared to the others when it is not the true scenario (type II). To assess the goodness-of-fit of the scenario, the PCA and ranked approaches were implemented on 1% of simulated data sets using the model check option.

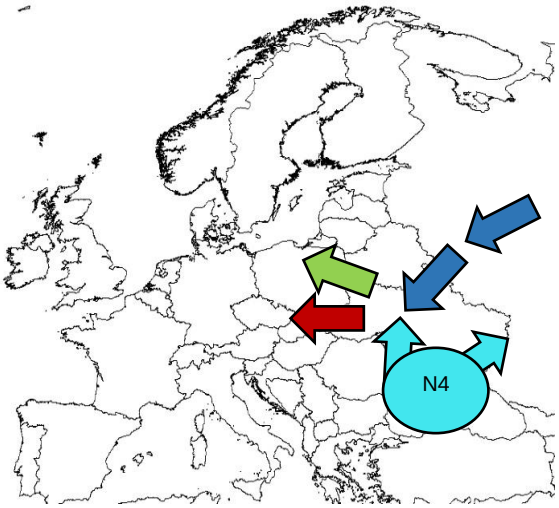
In addition to the SDM and BM, an expansion/migration model (EMM) was run using only *T. cordata* data from Siberia (*N1*), Poland (*N2*) and Austria (*N3*). Phuekvilai (2014), recognised a possible colonization route from a putative unknown refugium. So for this particular model, a hypothetical population (*N4*) was included to test possible recolonization/expansion/migration routes of *T. cordata.* Nine historic parameters (Table 3.2) and eleven scenarios (Fig. 3.4) were tested. The time of events (*t3, t2, t1*) were set at 10 – 10,000 generations (Table 3.2) for all scenarios.

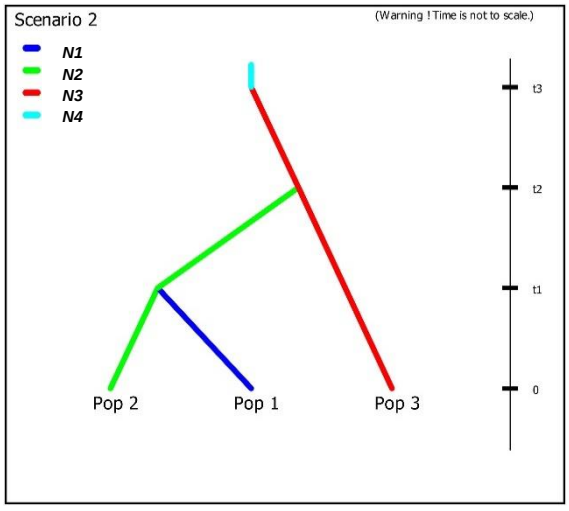
Scenario 1: *Tilia cordata* expanded its range from the hypothetical population (*N4*) into Siberia (*N1*) at time *t3* and the two subsequently became isolated from one another. At time *t2,* secondary contact between the two – newly recolonizing – populations (*N1* and *N4*) resulted in the recolonization of central Europe via Poland (*N2*) and Austria (*N3*). **Scenario 2:** *T. cordata* expanded its range from the hypothetical population (*N4*) to Austria (*N3*) at time *t3* then at time *t2* expanded to Poland (*N2*). From there a further expansion occurred at time *t1* to Siberia (*N1*). **Scenario 3:** A panmictic ancestral *T. cordata* population (*Na*) split at time *t3* resulting in two isolated populations – Siberia (*N1*) and the hypothetical population (*N4*). At time *t2* secondary contact led to admixture between the two resulting in the range expansion of Polish populations (*N2*). At time *t1* the hypothetical population (*N4*) expanded its range to Austria (*N3*). **Scenario 4:** At time *t3* the hypothetical population (*N4*) expanded its range to Siberia (*N1*), Poland (*N2*) and Austria (*N3*). **Scenario 5:** The hypothetical population (*N4*) split into two isolated populations – Siberia (*N1*) and Poland (*N2*) – at time *t3*. At time *t2* the Polish populations (*N2*) expanded their range to Austria (*N3*). **Scenario 6:** The hypothetical population (*N4*) expanded its range to Siberia (*N1*) then at time *t2* it expanded its range to Austria (*N3*) then onto Poland (*N2*) at time *t1.* **Scenario 7:** The hypothetical population (*N4*) expanded its range to Siberia (*N1*) then at time *t2* to Poland (*N2*) and Austria (*N3*). **Scenario 8:** A panmictic ancestral population (*Na*) underwent a reduction of N_e (bottleneck or founder event) at time *t3* resulting in an isolated Siberian population (*N1*). At time *t2* the Siberian population (N1) expanded its range to the hypothetical population (*N4*) then at time *t1* expanded to Poland (*N2*) and Austria (*N3*). **Scenario 9:** A panmictic ancestral population (*Na*) underwent a reduction of N_e (bottleneck or founder event) at time *t3* resulting in an isolated Siberian population (*N1*). At time *t2* the Siberian population (N1) expanded its range to Poland (*N2*), Austria (*N3*) and the hypothetical population (*N4*). **Scenario 10:** At time *t3 T. cordata* expanded its range from Austria (*N3*) to Poland (*N2*). At time *t2* the Austrian population then expanded to the hypothetical populations (*N4*) then onto Siberia at time *t1*. **Scenario 11:** At time *t3 T. cordata* expanded its range from Poland (*N2*) to Austria (*N3*). At time *t2* the Polish population then expanded to the hypothetical population (*N4*) then on to Siberia at time *t1*.

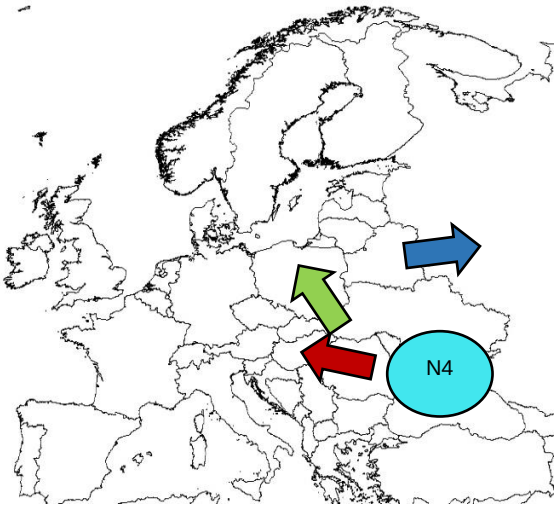
Priors were set as; *N1, N2*, *N3* and the hypothetical population *N4* ranging from 10 – 10,000; *t#* ranging from 10 – 10,000 generations (considering a minimum of 25 years generation time this translates to 0.25 – 250 kya); ancestral populations (*Na*)

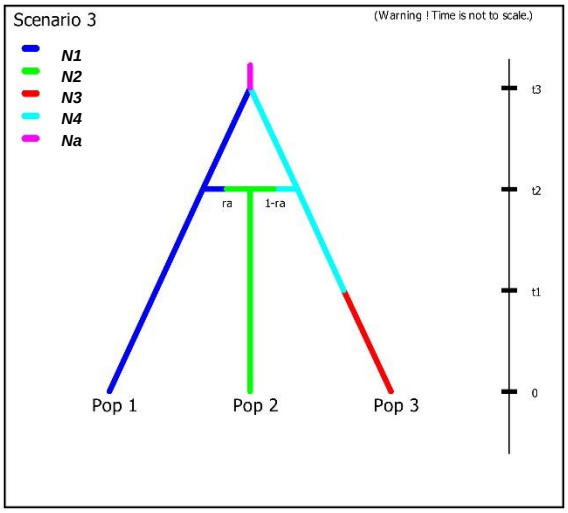
ranging from 10 – 50,000; and *ra* (rate of admixture) ranged from 0.001 – 0.999 (Table 3.2). The default mutation settings with the exception of the SNI (as above) were kept. Several conditions were set regarding N_e of *Tilia cordata* populations and historical time of events (Appendix 3.4). In total 21 summary statistics were recorded to build the reference table. The optimum set of summary statistics were used. A total of 11 x 10^6 datasets were simulated (1 x 10^6 for each scenario). Posterior probabilities were estimated using the direct and logistic approaches. Type I and type II errors were assessed using the 'confidence in scenario' option. The goodness-of-fit of the best scenario was assessed with the 'model check' option. The PCA and ranked approaches were used on 1% simulated data sets.

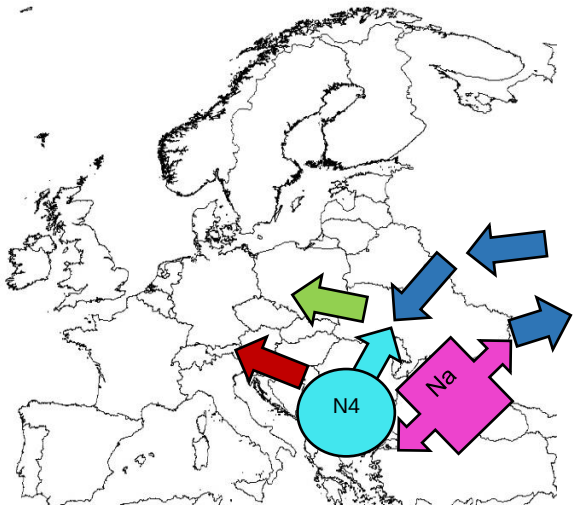


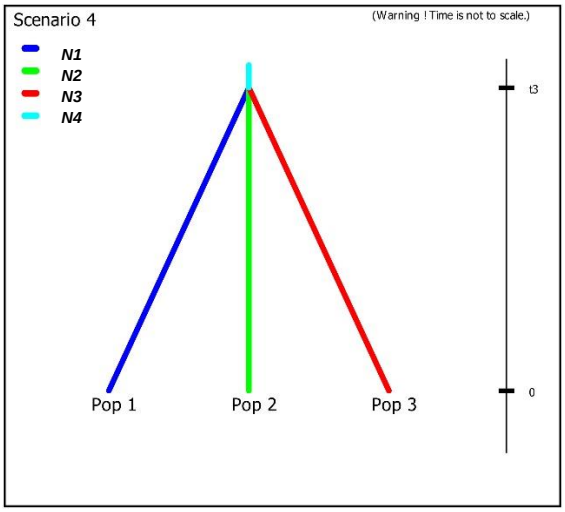


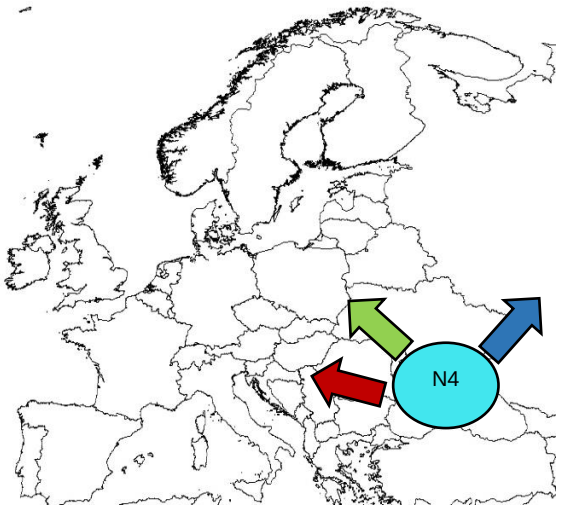


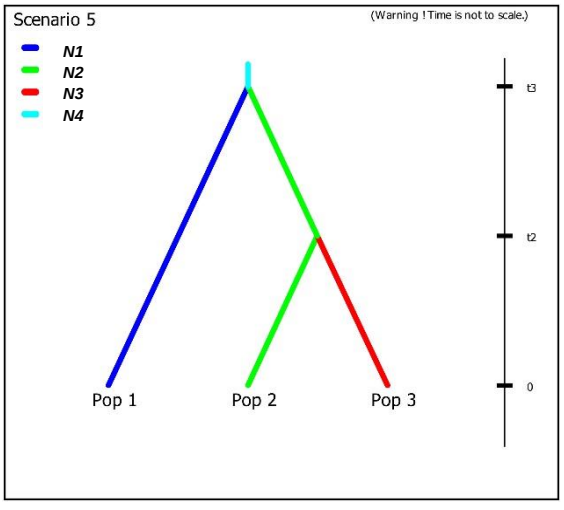


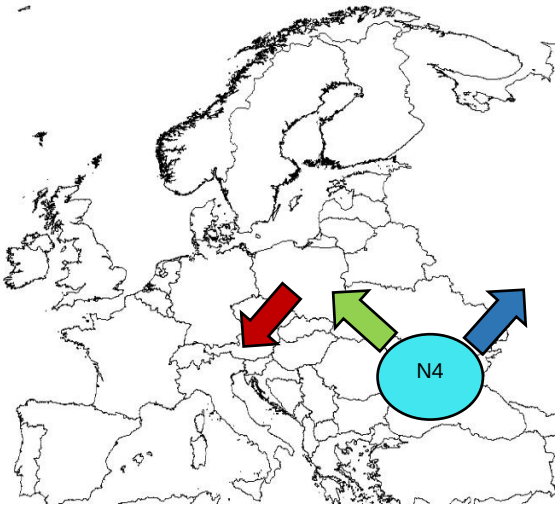

Scenario 3
(Warning ! Time is not to scale.)
N1
N2
N3
N4
Na
ra
1-ra
t3
t2
t1
0
Pop 1
Pop 2
Pop 3

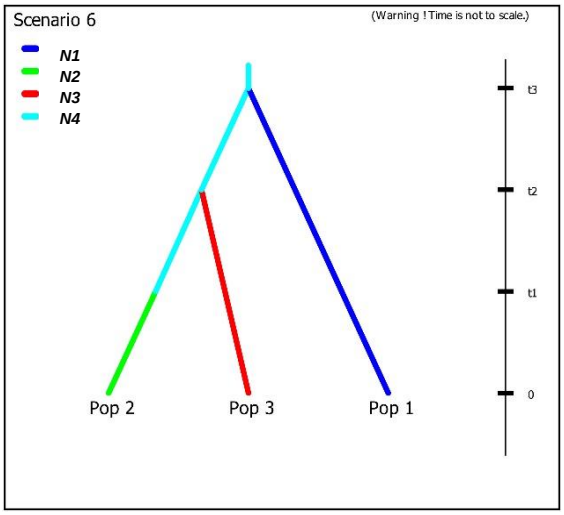

N4
Na

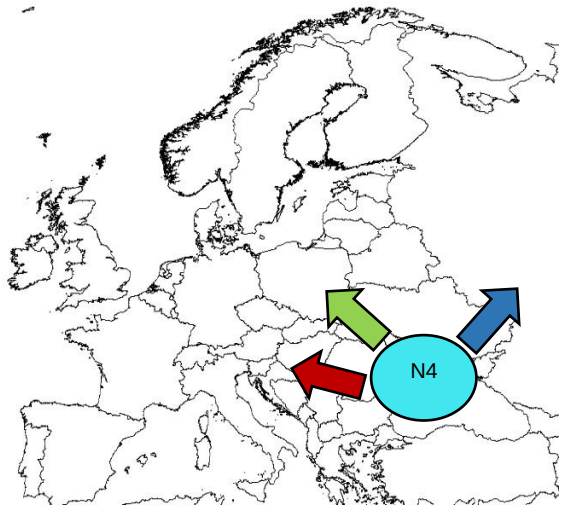

Scenario 4
(Warning ! Time is not to scale.)
N1
N2
N3
N4
t3
0
Pop 1
Pop 2
Pop 3

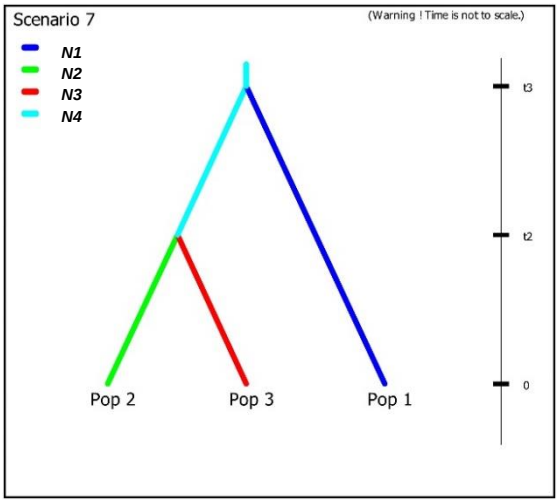

N4

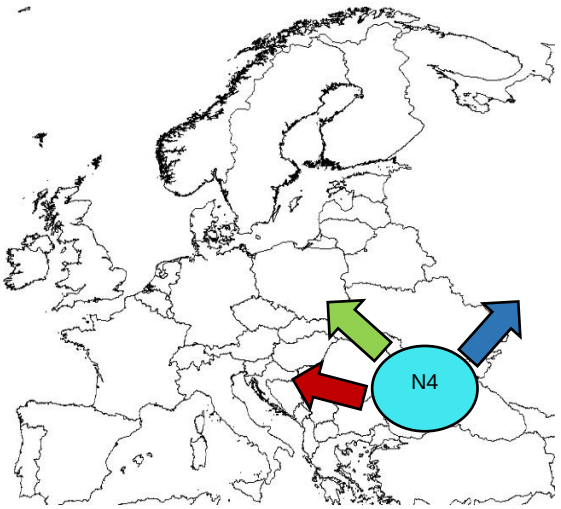

Scenario 5
(Warning ! Time is not to scale.)
N1
N2
N3
N4
t3
t2
0
Pop 1
Pop 2
Pop 3

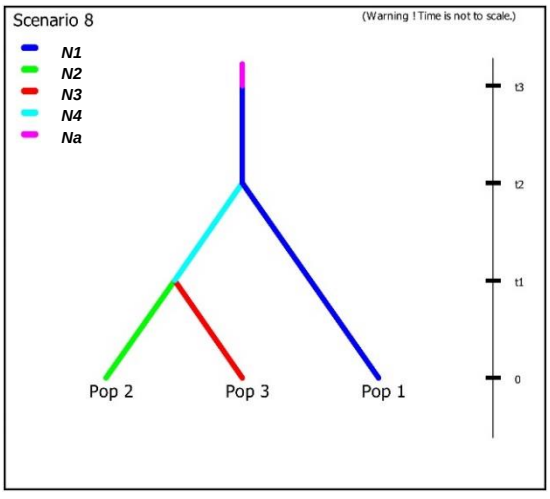

N4

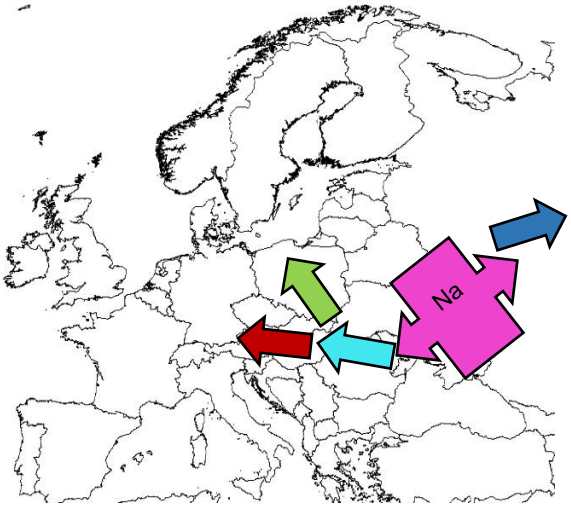

Scenario 6
(Warning ! Time is not to scale.)
N1
N2
N3
N4
t3
t2
t1
0
Pop 2
Pop 3
Pop 1


N4


Scenario 7
(Warning ! Time is not to scale.)
N1
N2
N3
N4
t3
t2
0
Pop 2
Pop 3
Pop 1


N4


Scenario 8
(Warning ! Time is not to scale.)
N1
N2
N3
N4
Na
t3
t2
t1
0
Pop 2
Pop 3
Pop 1


Na

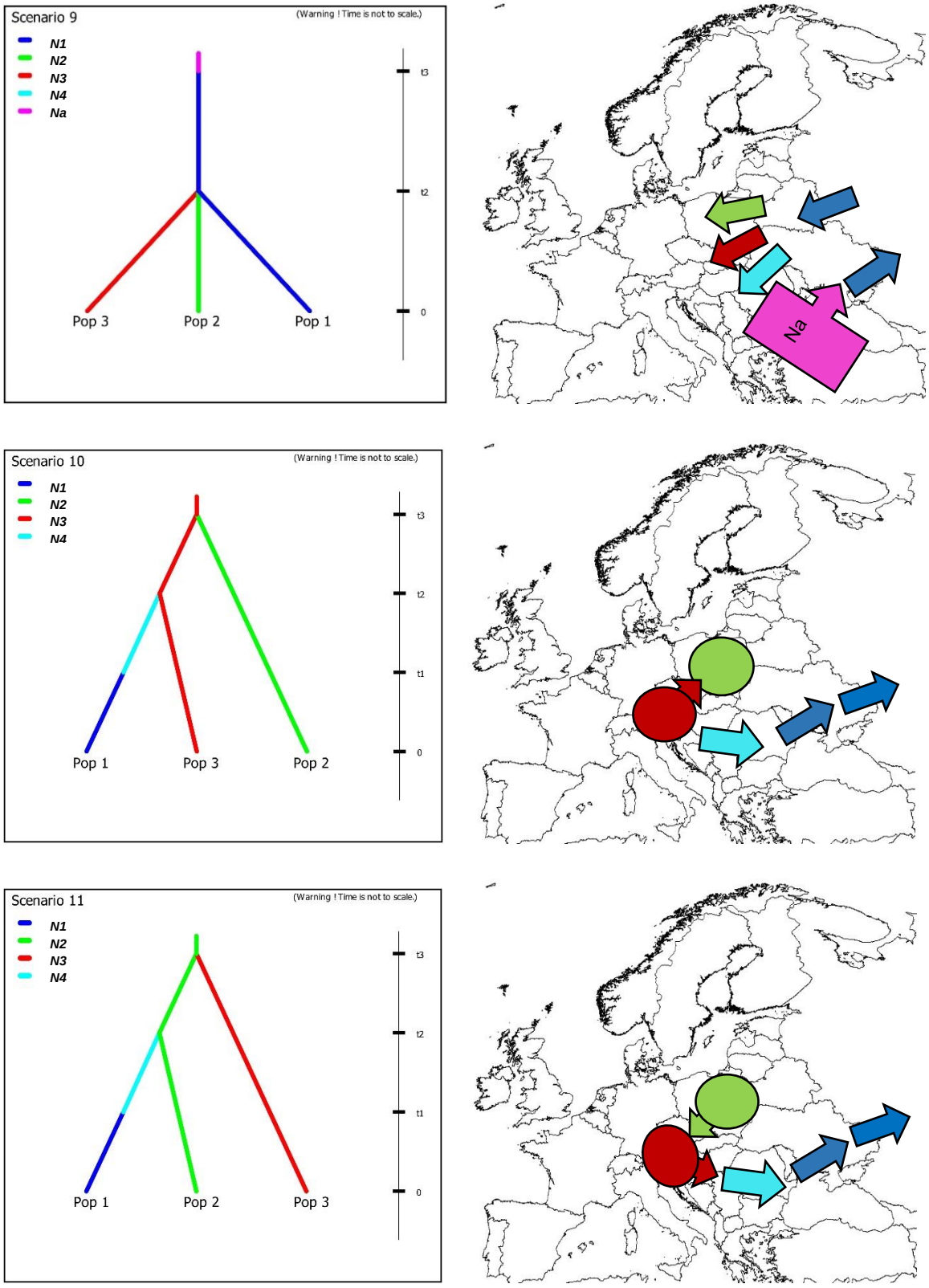


Figure 3.4 The Expansion/Migration Model (EMM) tested in DIYABC: *0* – time of sampling; *t#* – time of expansion/migration; *N1* –*T. cordata* Siberia (dark blue); *N2* – *T. cordata* Poland (green); *N3* –*T. cordata* Austria (red); *N4* –*T. cordata* un-sampled hypothetical population (pale blue); *Na* – Ancestral populations (*t3* – 0). Maps show putative movement of genotypes corresponding to its respective scenario.

3.4 Results

In total, 79 *Tilia sibirica* samples and 27 *T. cordata* from western Siberia were genetically analysed. Additionally, 42 *T. cordata* samples from Poland and 22 from Austria were also analysed following the removal of identical multi-locus genotypes. MICRO-CHECKER revealed no evidence of scoring error due to stuttering and no large allele dropout. Homozygote excess was observed at some loci but was not consistent across populations. Some deviation from Hardy-Weinberg Equilibrium (HWE) was observed. More specifically, GENEPOP revealed significant heterozygote deficit at many loci within *T. sibirica*. While this could be a result of inbreeding, Chapter 2 shows the two UK species to be outcrossing, and the inbreeding coefficient (F_{IS}) values for *T. sibirica* were not significantly different from zero in three of the five populations (Table 3.3). Therefore, the heterozygote deficit is likely due to the presence of null alleles at some loci. FreeNA estimated null allele frequencies (Appendix 3.5) and adjusted H_E and F_{ST} values accordingly *i.e.* excluding null alleles (Table 3.3 and 3.4).

3.4.1 Genetic diversity and differentiation

Two groups were identified following the PCoA analysis based on pairwise genetic distance (Fig. 3.5). The blue cluster are *Tilia sibirica* while the orange cluster are *T. cordata* from Siberia, Poland and Austria. Two groups (*K=2*) were observed following STRUCTURE analysis (Fig. 3.6a). Evanno's *ΔK* method also revealed *K=2* to be optimal (Fig. 3.6b). Based on the molecular markers used in this study, there is a clear a genetic divide of the two taxa. *Tilia cordata* in Europe and west Russia and *T. sibirica* further east in southern Siberia (Fig. 3.7).

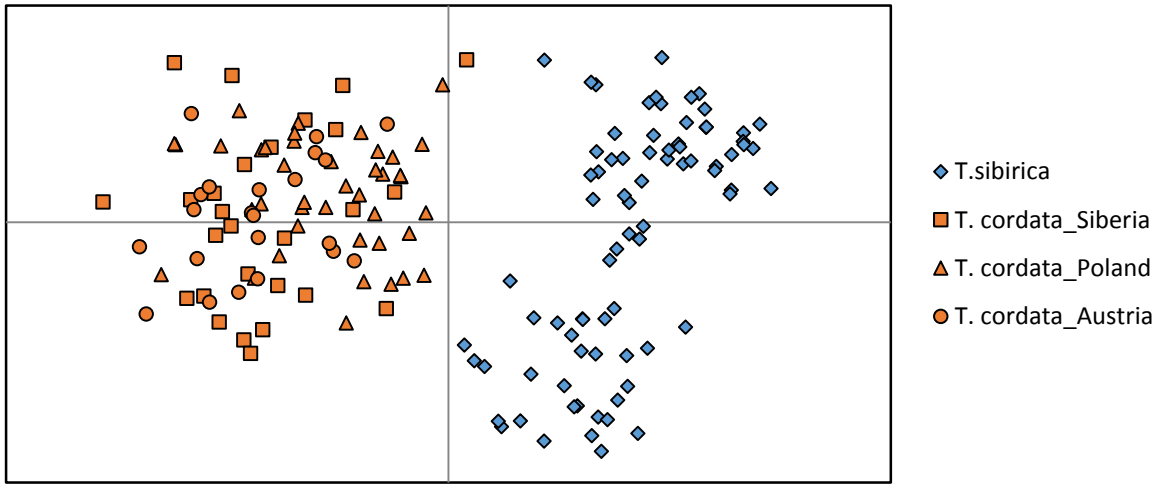
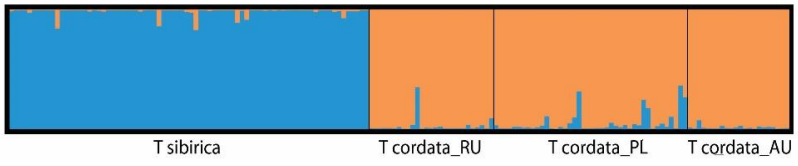


Figure 3.5 PCoA of 170 *Tilia* individuals from four regions. Blue cluster are *Tilia sibirica* and the orange cluster are *T. cordata.* Axes 1 and 2 explains 55% of the genetic variation (38 and 17% respectively).

a)



b)

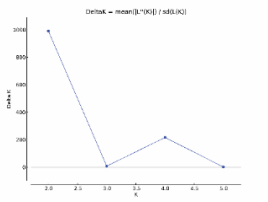


Figure 3.6 a) Assignment of 79 *Tilia sibirica* individuals (blue cluster) and 91 *T. cordata* individuals from three regions (Siberia, Poland and Austria – orange cluster) inferred by Bayesian clustering analysis implemented in STRUCTURE, visualized by DISTRUCT. b) Evanno's *ΔK* showing *K=2* to be optimal, implemented in STRUCTURE HARVESTER.

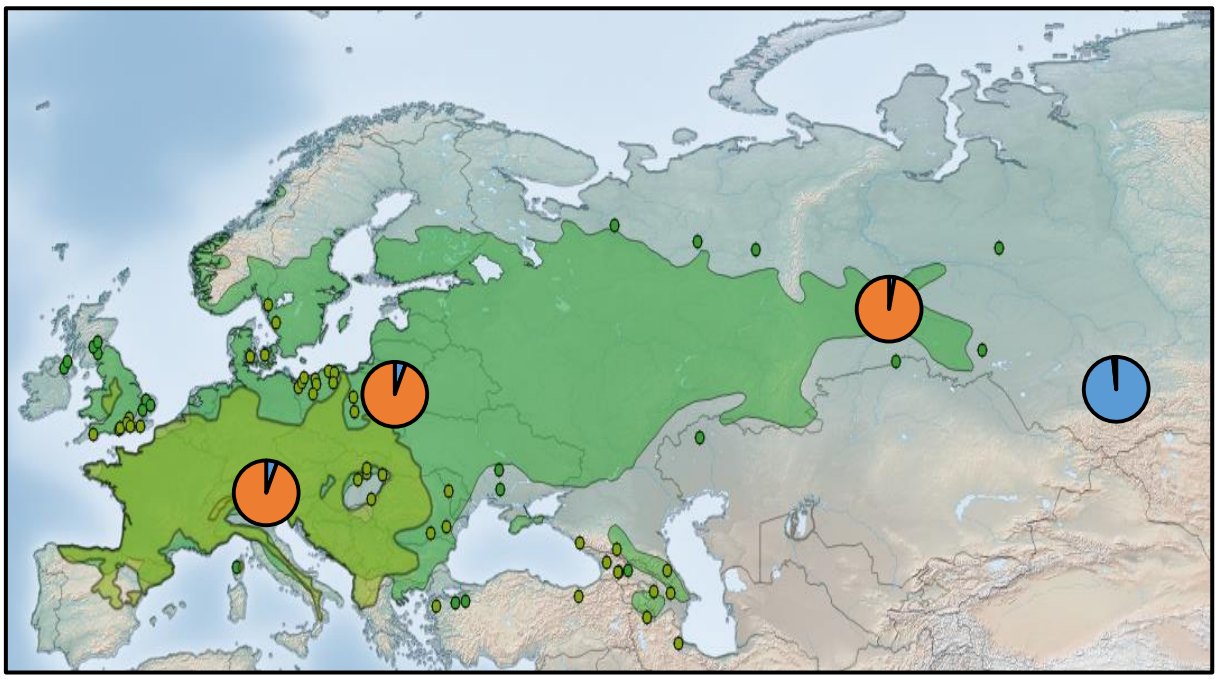


Figure 3.7 The distribution of genetic variation in 91 *T. cordata* individuals (orange group) and 79 *T. sibirica* individuals (blue group) inferred by Bayesian clustering analysis implemented in STRUCTURE, showing two distinct groups (*K=2*).

The microsatellite markers revealed various levels of polymorphism within the *Tilia* samples. Summary statistics are only presented for populations with more than five individuals. The total number of individuals analysed, following the removal of duplicates, ranged from 6 – 22 in *T. sibirica* and from 12 – 22 in *T. cordata.* It is clear that *T. sibirica* has lower genetic diversity than *T. cordata*, based on the markers and populations used in this study, as revealed by expected heterozygosity (H_E) and adjusted expected heterozygosity (H_{E_NULL}). Values ranged from 0.272 – 0.379 in *T. sibirica* and were significantly higher (Mann Whitney *U* test, *W* = 30, *P* = 0.004) in *T. cordata*, ranging from 0.523 – 0.609. Values were also significantly different (Mann Whitney *U* test, *W* = 30, *P* = 0.004), when adjusted for null alleles (Table 3.3). Estimates of null allele frequencies for each locus within each population are provided as supplementary material (Appendix 3.5). The average number of alleles (N_A) in *T. cordata* ranged from 4.92 – 5.83 and was significantly higher (Mann Whitney *U* test, *W* = 30, *P* = 0.008) than in *T. sibirica* which ranged from 2.00 – 3.08. Allelic richness (A_R) in *T. cordata*, ranging from 3.78 – 3.98 was also significantly higher (Mann Whitney *U* test, *W* = 30, *P* = 0.004) than in *T. sibirica* where A_R was no higher than 2.38. The inbreeding coefficient (F_{IS}) ranged from -0.194 to 0.448 in *T. sibirica* and were

significantly different from zero (P<0.05) in two populations. F_{IS} values within *T. cordata* ranged from -0.047 to 0.158 and were not significant (Table 3.3).

Table 3.3 Summary statistics of five *T. sibirica,* two Siberian *T. cordata,* three Polish *T. cordata* and one Austrian *T. cordata* population. *N* – Number of individuals; N_A – Average number of alleles; A_R – Allelic Richness; H_E – Nei's unbiased Expected Heterozygosity; H_{E_NULL} – Expected heterozygosity adjusted for null alleles; F_{IS} – Inbreeding coefficient (**P<0.05*).

Population	Species	*N*	N_A	A_R	H_E	H_{E_NULL}	F_{IS}
*K*12	*T. sibirica*	22	3.08	2.376	0.346	0.377	-0.006
*K*22	*T. sibirica*	6	2.00	2.000	0.274	0.266	0.095
*K*28	*T. sibirica*	21	2.92	2.307	0.320	0.355	0.448*
*K*29	*T. sibirica*	20	2.75	2.128	0.272	0.292	0.348*
*K*38	*T. sibirica*	6	2.25	2.250	0.379	0.368	-0.194
*V*20	*T. cordata*	14	5.08	3.878	0.597	0.577	0.158
*V*25	*T. cordata*	13	4.92	3.777	0.566	0.548	0.062
*B*40	*T. cordata*	13	5.00	3.934	0.583	0.569	-0.047
*B*69	*T. cordata*	17	5.83	3.835	0.523	0.510	-0.033
*B*99	*T. cordata*	12	5.08	3.981	0.575	0.543	-0.015
A01	*T. cordata*	22	5.42	3.869	0.609	0.587	0.055

In total, 136 alleles were identified between the two taxa. The total number of alleles per locus ranged from 1 (*Tc8*) to 29 (*Tc963*). There were more private alleles found in *T. cordata* (81) than in *T. sibirica* (14). Genetic differentiation between the two taxa varied depending on which measure was applied (Table 3.4). Based on F_{ST} values, loci *Tc920, Tc943* and *Tc7* showed the highest differentiation, followed by *Tc6, Tc4, Tc915* and *Tc951.* These changed slightly when adjusted for null alleles. Based on D_{est} *Tc6, Tc920, Tc943, Tc4, Tc915,* and *Tc963* showed the highest actual differentiation. G_{ST_est} and G'_{ST_est} values varied from 0.023 – 0.248 and 0.056 – 0.883 respectively, with locus *Tc920* showing the highest differentiation from the two measures. Indeed this locus showed the highest differentiation among all measures (Table 3.4) and may be useful as a species identifier. In *T. sibirica* allele size at this locus was generally fixed at 223bp and 230bp with a rare allele size of 218bp, while allele size ranged from 221 – 242bp in *T. cordata.*

Table 3.4 Total number of alleles and private alleles within *T. sibirica* and *T. cordata* and per locus genetic differentiation between species. A_P - Private alleles among taxa; F_{ST} and F_{ST_NULL} (adjusted), D_{est}, $G_{ST\ est}$ and $G'_{ST\ est}$ between taxa.

Locus	No. of Alleles	A_P *T. sibirica*	A_P *T. cordata*	F_{ST}	$F_{ST\ NULL}$	D_{est}	$G_{ST\ est}$	$G'_{ST\ est}$
Tc6	13	-	7	0.17	0.15	0.63	0.09	0.66
Tc937	8	1	4	0.04	0.11	0.03	0.02	0.06
Tc920	15	2	10	0.39	0.29	0.85	0.25	0.88
Tc8	1	-	-	-	0.04	0.00	-	-
Tc943	8	2	4	0.31	0.26	0.23	0.20	0.38
Tc4	10	-	6	0.17	0.16	0.38	0.10	0.43
Tc927	3	-	2	0.08	0.08	0.01	0.05	0.06
Tc915	19	1	12	0.15	0.17	0.52	0.08	0.56
Tc963	29	3	18	0.08	0.08	0.53	0.04	0.55
Tc5	19	3	12	0.05	0.07	0.10	0.02	0.12
Tc951	6	2	2	0.14	0.09	0.13	0.08	0.20
Tc7	5	-	4	0.24	0.27	0.10	0.15	0.23
Total/Mean	136	14	81	0.17	0.15	0.29	0.10	0.38

Significant population pairwise F_{ST} between the two species (Table 3.5a) ranged from 0.002 (populations V25 and V20) to 0.303 (populations V25 and K29). These values changed only slightly following adjustment for null alleles (0.005 and 0.286, Table 3.5b).There was high genetic differentiation between the two species revealed by an AMOVA (15.86% of the total variation, P<0.001). Among populations within taxa variation was 6.83% (P<0.001), while the remaining variation (77.31%, P<0.001) was found within populations (Table 3.6).

Table 3.5a Population pairwise F_{ST} values and significance of *T. sibirica* and *T. cordata.* K# – *T. sibirica* populations, V# – Siberia *T. cordata* populations, B# – Poland *T. cordata* populations, A# – Austria *T. cordata* population. (* - 0.05, ** - 0.01, *** - 0.001, NS - Not significant).

	K22	K28	K12	K29	K38	V20	V25	B69	B99	B40	A01
K22	-	NS	*	NS	NS	***	**	***	**	**	***
K28	0.011	-	***	**	NS	***	***	***	***	***	***
K12	0.078	0.107	-	***	*	***	***	***	***	***	***
K29	0.073	0.088	0.120	-	**	***	***	***	***	***	***
K38	0.172	0.109	0.168	0.211	-	**	**	***	**	**	***
V20	0.222	0.242	0.247	0.291	0.194	-	NS	***	***	***	***
V25	0.227	0.234	0.256	0.303	0.175	0.002	-	***	***	***	***
B69	0.222	0.238	0.227	0.258	0.215	0.094	0.117	-	**	**	***
B99	0.195	0.207	0.192	0.260	0.175	0.062	0.083	0.046	-	NS	***
B40	0.178	0.190	0.182	0.222	0.151	0.046	0.058	0.031	0.033	-	***
A01	0.222	0.234	0.243	0.263	0.181	0.050	0.049	0.099	0.085	0.051	-

Table 3.5b F_{ST_NULL} - Adjusted population pairwise F_{ST} values (*i.e.* excluding null alleles).

	K22	K28	K12	K29	K38	V20	V25	B69	B99	B40	A01
K22	-										
K28	0.042	-									
K12	0.088	0.093	-								
K29	0.078	0.065	0.109	-							
K38	0.174	0.112	0.158	0.210	-						
V20	0.232	0.231	0.240	0.279	0.197	-					
V25	0.231	0.223	0.244	0.286	0.175	0.005	-				
B69	0.214	0.214	0.209	0.231	0.206	0.100	0.115	-			
B99	0.196	0.190	0.181	0.238	0.174	0.067	0.086	0.046	-		
B40	0.173	0.166	0.167	0.200	0.143	0.052	0.059	0.031	0.037	-	
A01	0.220	0.214	0.230	0.245	0.176	0.052	0.050	0.099	0.085	0.051	-

Table 3.6 Analysis of Molecular Variance (AMOVA) of *T. sibirica* and *T. cordata*, showing the partitioning of genetic variation among species, among and within populations.

Source of variation	Sum of squares	Variance components	Percentage variation	*P* value
Among species	103.993	0.565	15.86%	<0.001
Among populations within species	88.622	0.243	6.83%	<0.001
Within populations	882.185	2.753	77.31%	<0.001
Total	1074.80	3.561	100%	

3.4.2 Historical scenarios

Following the 'pre-evaluate scenario and priors' option with all available summary statistics, the PCA graph – which visualises the distribution of the data – clearly showed that the observed data set (yellow dot) was not clustered within the 1,000 (1%) simulated data sets (green dots). This suggests that using all available statistics is not suited for the priors and may not fit the model (Appendix 3.1). The statistics with significant differences (*P* <0.01) between the simulated data sets and the observed (Appendix 3.2) were omitted. All other statistics were close enough to the observed data to accept the parameters and scenario priors (Estoup *et al.*, 2015). In total, ten summary statistics were used to build the reference table for the simple divergence model (Appendix 3.3) and these statistics were used throughout this study for further analyses.

3.4.3 Tilia sibirica and T. cordata Simple Divergence Model (SDM)

Using the optimum set of summary statistics, the 'pre-evaluate option' showed the observed data set (yellow dot) situated within 1% of the simulated data sets from the simple divergence scenario suggesting that the priors fit the data (Appendix 3.6). The SDM (Fig.3.2) revealed a higher effective population size for *T. cordata* (*N2*) than for *T. sibirica* (*N1*). Mode (and 95% CI) values (Appendix 3.7) were estimated to be 9,490 (5,480 – 9,860) and 1,000 (444 – 2,460) respectively. The time of divergence (*td*), was estimated at 369 generations ago (163 – 4,680 gens). Considering a minimum generation time of 25 years, this infers a split occurred around 9.2 kya (4.1 – 117 kya). Model checking assessed the goodness-of-fit and produced a PCA graph showing a large cluster of simulated data from the prior and a smaller cluster of data

from the posterior predictive distribution with the observed data set placed well within (Fig. 3.8). This suggests that the observed data fits the scenario well (Estoup *et al.*, 2015). However, posterior checking using the rank approach showed a significant difference at one statistic (Appendix 3.8).

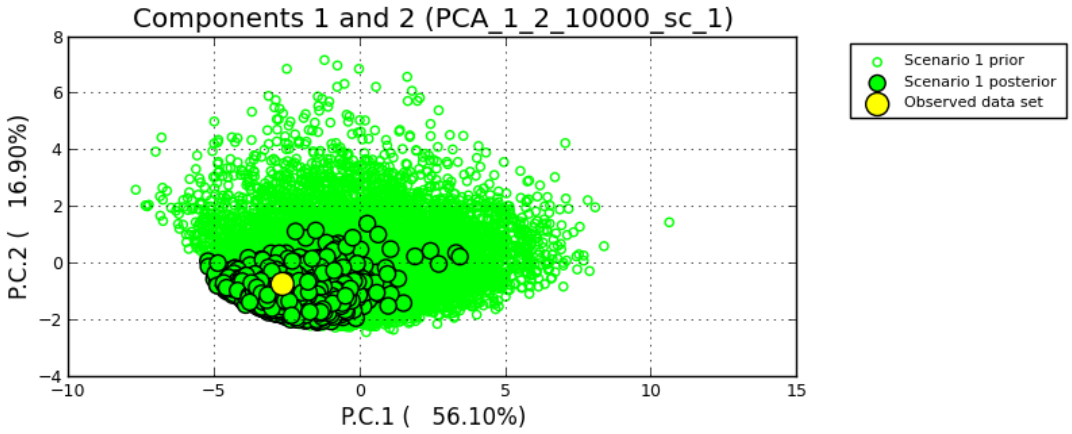


Figure 3.8 Goodness-of-fit of the SDM, assessed by model check within DIYABC. The PCA shows the observed data set (yellow dot) nestled within the posterior predictive distribution (large green dots) and the large cloud of simulated data from the prior.

3.4.4 Tilia sibirica Bottleneck Model (BM)

The 'pre-evaluate option' showed the observed data set situated within 1% of the simulated data sets from each of the four scenarios suggesting that the priors fit at least one of the scenarios tested (Appendix 3.9). The BM scenario (Fig. 3.3) with highest posterior probability, using both the direct and logistic approach, was scenario 4 (0.553 and 0.772 respectively – Figs. 3.9 and 3.10, Appendix 3.10a). Using the mode (and 95% CI) values (Appendix 3.11), this suggests that DIYABC analysis of the microsatellite data support a reduction in N_e of the ancestral lineage of *T. sibirica* approximately 3,710 generations ago (2,340 – 83,000 generations) *i.e.*, 92.8 kya (58.5 kya – 2.1 mya) with a population expansion approximately 73 generations ago (35 – 629 gens) *i.e.* 1.9 kya (0.9 kya – 15.7 kya) and has remained constant. Type I error (true scenario rejected) for scenario 4 was relatively high – 0.318 and 0.294 – while type II error (false scenario not rejected) was low at 0.113 and 0.132 for the direct and logistic approaches respectively. The 'goodness-of-fit' of the scenario was assessed with the model check option. The PCA and the ranked

approaches were performed on 1% simulated data sets and suggested that the observed data fits the scenario (Fig. 3.11 and Appendix 3.8).

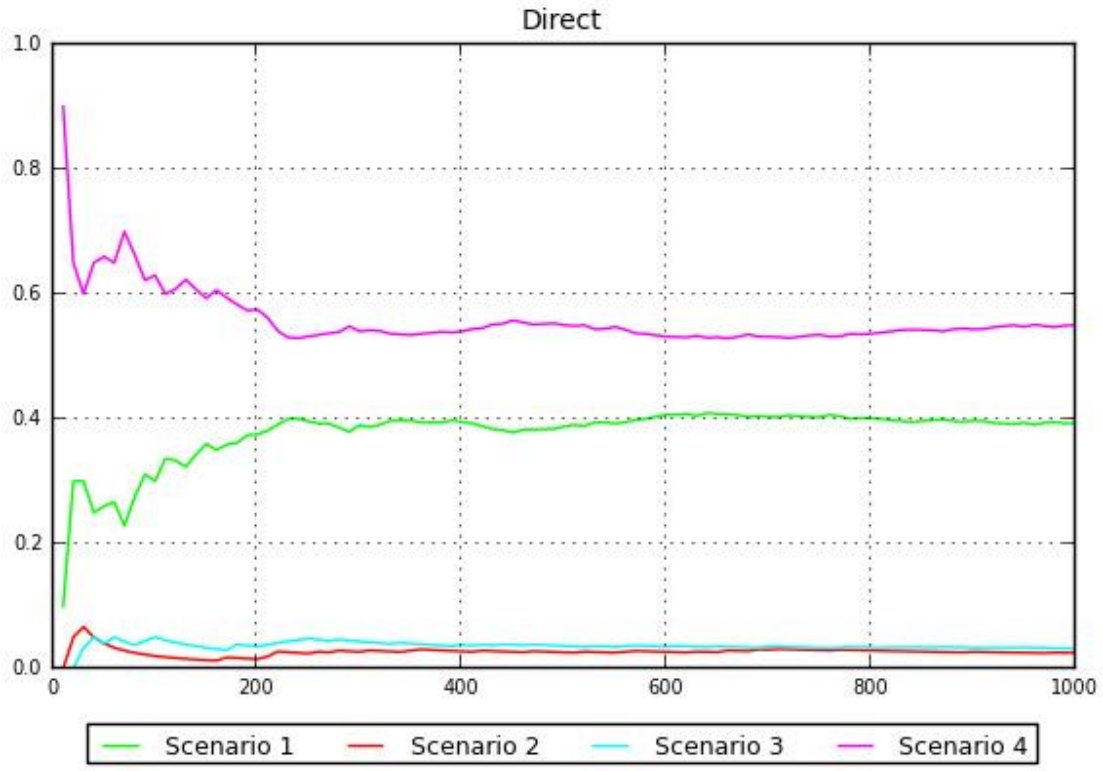


Figure 3.9 Direct approach assessing the posterior probabilities of each scenario from the BM using 1000 data sets. Scenario 4 is shown to have the highest posterior probability values.

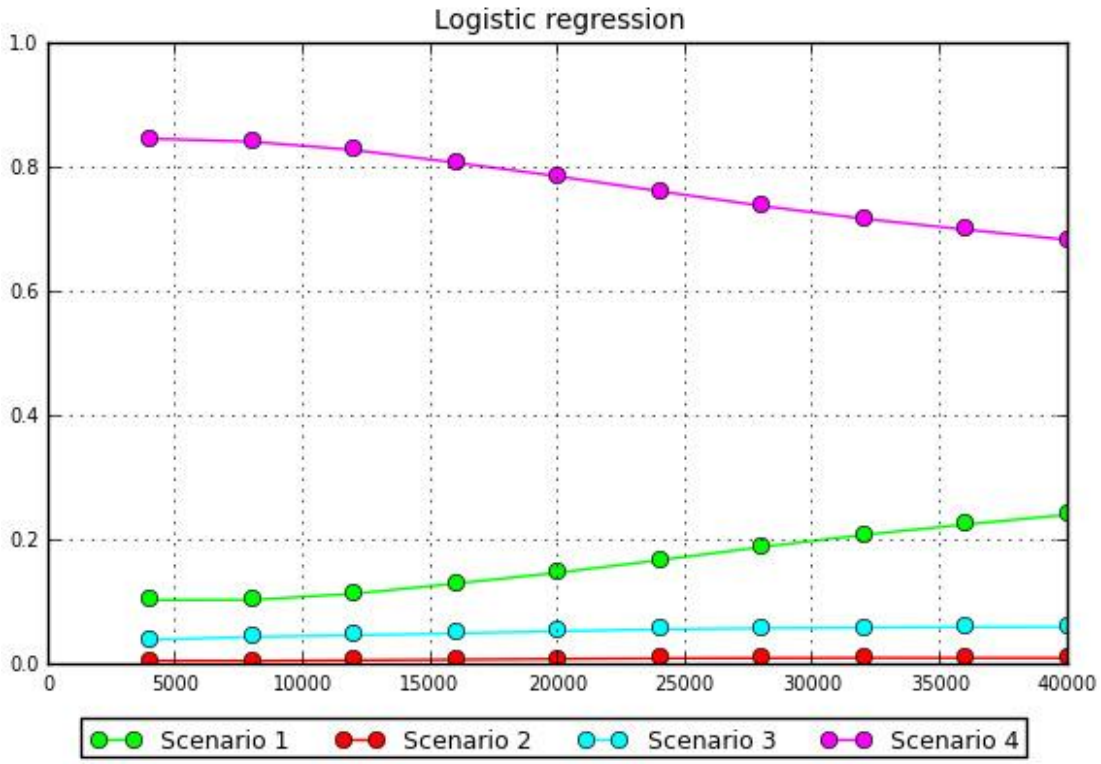


Figure 3.10 Logistic approach assessing the posterior probabilities of each scenario from the BM using 40,000 data sets. Once again scenario 4 is shown to have the highest posterior probability values.

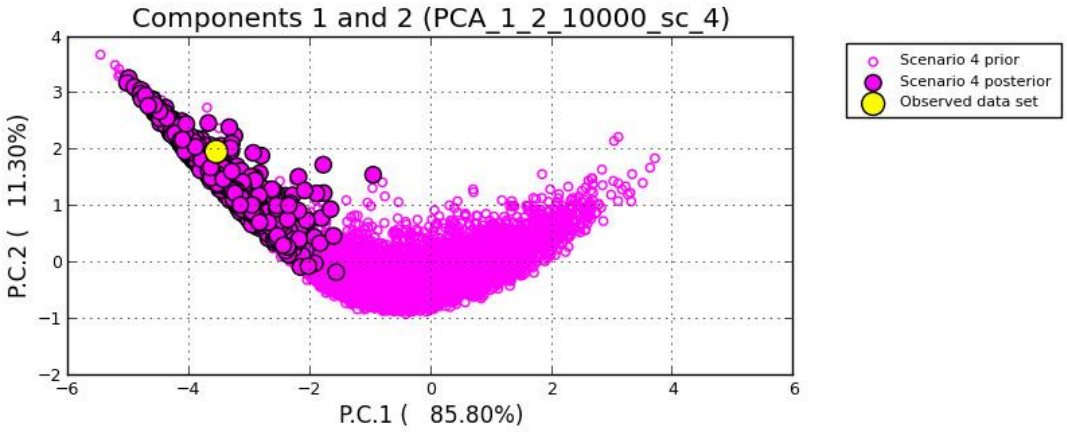


Figure 3.11 Goodness-of-fit of scenario 4 from the BM assessed by model check within DIYABC. The PCA shows the observed data set (yellow dot) placed within the posterior predictive distribution (large lilac dots) and the large cloud of simulated data form the prior (small lilac dots).

3.4.5 Tilia cordata Expansion/Migration Model (EMM)

The PCA graph produced from the 'pre-evaluate scenario and priors' option showed the observed data (large yellow dot) positioned within 1% of the simulated data suggesting that the priors fit at least one proposed scenario (Appendix 3.12). The EMM scenario (Fig. 3.4) with highest posterior probabilities, using the direct approach, was scenario 9 (0.347, Fig. 3.12, Appendix 3.10b). The logistic approach did not provide a clear 'best' scenario (data not shown). Using the mode (and 95% CI) values (Appendix 3.13), the results support a reduction in N_e of ancestral *Tilia* around 9,370 generations ago (3,240 – 9,780 generations) *i.e.* 234.3 kya (81 – 244.5 kya) with contribution to the lineage's range expansion of eastern and central Europe from the central Russian plain around 418 generations ago (210 – 1,080 generations) *i.e*. 10.5 kya (5.3 – 27 kya). The hypothetical population (*N4*) was un-sampled at time *0*, so it does not appear in the scenario illustration but the model suggests that a third population (*i.e.* Austria, Poland and one other) had diverged and expanded from the east. Type I error (true scenario rejected) for scenario 9 was 0.293 and 0.713, while type II error (false scenario not rejected) was 0.240 and 0.138 for the direct and logistic approaches, respectively. The 'goodness-of-fit' of the scenario was assessed by the model check option. The PCA and the ranked approaches were used on 1% simulated data sets and suggest that the observed data fits the scenario (Fig. 3.13 and Appendix 3.8).

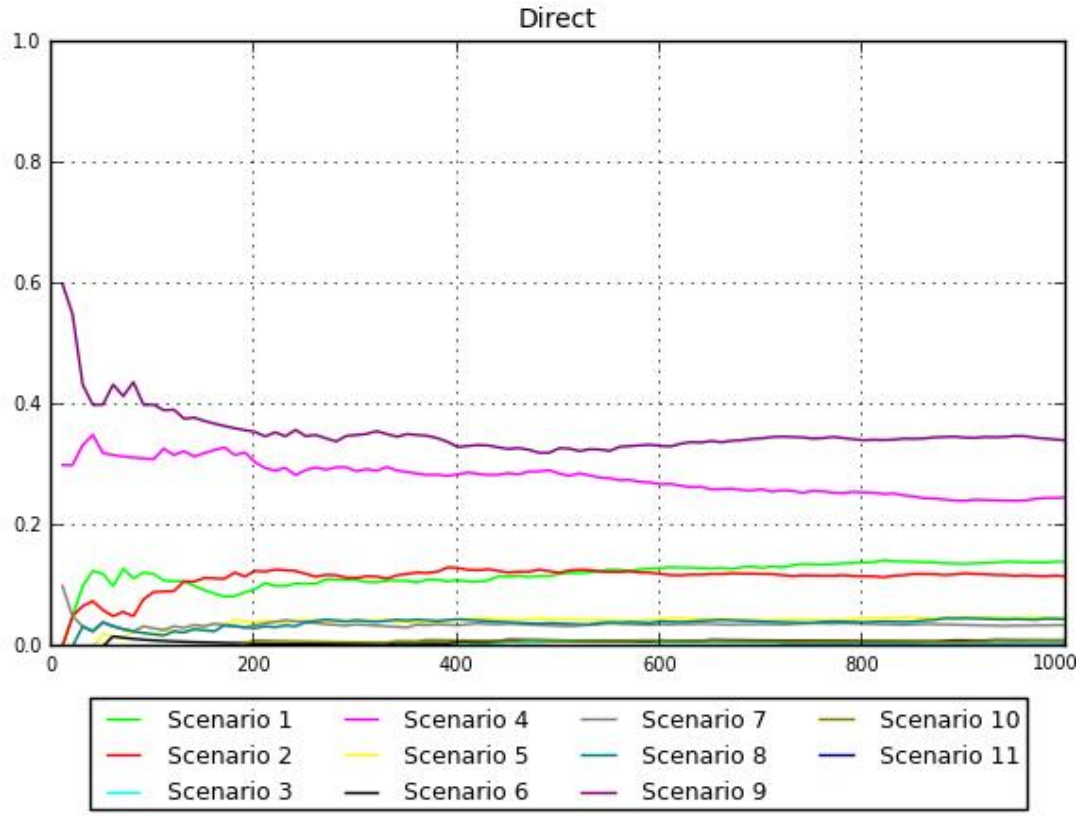


Figure 3.12 Direct approach assessing the posterior probabilities of each scenario from the EMM using 1000 data sets. Scenario 9 is shown to have the highest posterior probability values.

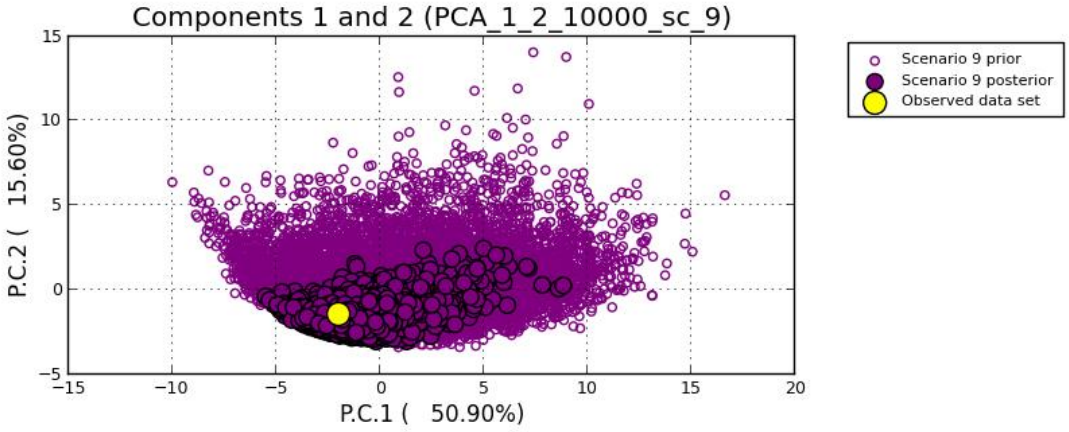


Figure 3.13 Goodness-of-fit of scenario 9 from the EMM assessed by model check within DIYABC. The PCA shows the observed data set (yellow dot) nestled within the posterior predictive distribution (large purple dots) and the large cloud of simulated data form the prior (small purple dots).

3.5 Discussion

Genetic analyses of *T. sibirica* and *T. cordata* suggests that the two taxa are distinct biological units. This is apparent from the PCoA and STRUCTURE analyses (Figs. 3.5 and 3.6), and from strong genetic differentiation (Table 3.5). The estimated divergence time obtained from the DIYABC analysis suggests that the two diverged from a common ancestor relatively recently (post-LGM). While this is not impossible, it does seem rather recent for species divergence despite the strong and sudden effects that continuously retreating and expanding ice sheets can have on taxa (Hewitt, 2000; Hewitt, 2004). However, given the level of genetic difference between the two and the ecological and geographical isolation of *Tilia sibirica* in southern Siberia, a population split or indeed, a subspecies split approximately 369 – 4,680 generations ago (sometime between 4.1 – 117 kya, based on a generation time of 25 years), cannot be ruled out.

3.5.1 Low genetic diversity in Tilia sibirica

Not surprisingly, *Tilia cordata* showed a high degree of genetic diversity while *T. sibirica* did not (Table 3.3). The results show that Siberian, Polish, and Austrian *T. cordata* has slightly higher diversity levels than UK populations (Logan *et al.*, 2015). *T. cordata* were confirmed to be diploid and this study confirms that *T. sibirica* are also diploid (Pigott, 2012), because all samples analysed showed a maximum of two alleles per locus. Ascertainment bias might be a reason why *T. sibirica* shows such low diversity (Ellegren *et al.*, 1995). The microsatellite markers used in this study were designed from *T. platyphyllos* (Phuekvilai and Wolff, 2013). However, stochastic genetic drift can also lower genetic diversity of rare species and small populations (Ellstrand and Elam, 1993) and this could be an alternative reason for the low levels of expected heterozygosity seen here in the Siberian lime. Another alternative is that *T. sibirica* is not an obligatory outcrossing system, as outcrossing woody species tend to show high within-population genetic diversity (Hamrick and Godt, 1996). However, only two of the five populations showed F_{IS} values significantly higher than zero (Table 3.3). Some deviation from HWE were observed which was likely caused by a heterozygote deficit in some *T. sibirica* populations. Past logging of populations may partly be responsible for low heterozygosity seen in *T. sibirica*. However, DIYABC analyses (BM Scenario 4, Fig. 3.3, and Appendix 3.11) suggest that 19th Century logging had no effect on effective population size based on the microsatellite markers used in this study (see below).

3.5.2 Relatively recent divergence between T. cordata and T. sibirica

Assuming a generation time of 25 years and considering the summary statistics, the parameters set and the microsatellite markers used, this study can infer a Late Pleistocene/ early Holocene split between *T. cordata* and *T. sibirica* 369 generations ago (approximately 9.2 kya). Considering some of the earliest putative *Tilia* species have been dated from the Tertiary period (Wolfe and Wehr, 1987; Pigott, 2012) and a recently estimated age of the Tilioideae was inferred to be ~17 million years (Richardson *et al.*, 2015), the estimated divergence date from this study seems rather more recent than expected. However, DIYABC analyses provided credible intervals – the highest value being 117 kya (Appendix 3.7). A coalescent based study using nuclear loci from *Populus balsamifera* and *P. trichocarpa* inferred a divergence date of ~75 kya (Levsen *et al.*, 2012) and while this is much more recent than *Quercus* spp. (Bagnoli *et al.*, 2015) it does suggest that climatic fluctuations of the Late Pleistocene had a strong effect on tree species causing a lineage divergence.

While no particular geological or environmental events occurred in the region that would have initiated a split at around 9 kya, the Altai Mountains and western Siberia were experiencing a general rise in temperatures with subsequent climatic changes from the Late Pleistocene to the early Holocene. This was interspersed with periods of cooling (Younger Dryas – 12.5 kya) and warming (Preboreal – 11.5 kya). Genetic differences between the two taxa may have been strongly influenced by regular expansion and retraction of *Tilia* populations in the Altai region (Bolikhovskaya and Shunkov, 2014) and by climatic fluctuations (Groisman *et al.*, 2013) which may have subsequently led to a recent divergence.

3.5.3 Tilia sibirica holding its ground despite human disturbance?

The reduction in N_e of the *T. sibirica* lineage around 92.8 kya, as suggested by DIYABC, coincides with the Early Weichselian glaciation and while model reconstructions suggest that the extent of the ice reached far into the Eurasian continent, it did not seem to reach as far south as the Altai mountain region (Siegert *et al.*, 2001; Svendsen *et al.*, 2004; Larsen *et al.*, 2006; Peyaud *et al.*, 2007). However, consequences from this sudden period of glaciation and further Quaternary temperature fluctuations were ice-dammed lakes forming at lower latitudes and causing severe flooding in southern Siberia including the Altai Mountains (Rudoy and Baker, 1993; Reuther *et al.*, 2006; Carling *et al.*, 2010). Indeed, Carling *et al.* (2010) state that at least three catastrophic flooding events have occurred in the region

during the Late Pleistocene and modelled the rate and route of the largest which is believed to have occurred around 40 kya. They concluded that water flowed from an ice-dammed lake and completely flooded the western edge of the Altai mountain region (more or less exactly where *T. sibirica* persists today), at a rate of millions of cubic meters per second ($Mm^3\ s^{-1}$). While many tree species, particularly those at higher altitudes, may have survived the flooding it is likely that such events had a direct effect on smaller sized populations of some tree species.

The population size expansion, that the ABC analyses suggests took place 73 generations ago (approximately 1.8 kya, based on a generation time of 25 years), occurred at a time of general climatic fluctuations in the Altai region (Groisman *et al.*, 2013). Despite logging of the Siberian forests throughout the 19th Century, effective population size has remained constant. While this might seem incongruous, one possible explanation is that at some point prior to the intense logging, the balance between sexual and asexual reproduction had shifted (Price and Marshall, 1999). If much of the southern Siberian forest was logged during the 19th Century and only small fragmented – but largely clonally reproducing – populations remained, then effective population size may not have been affected (or at least not as significantly as would be presumed). This is because genetic diversity is not lost in clonal organisms (Balloux *et al.*, 2003). The authors show that effective population size rises to almost infinity when populations reached full clonality, but that genotype diversity decreases as clonal reproduction increases, although this fall is slow until full clonality is reached.

Stoeckel *et al.* (2006) showed significant negative F_{IS} values in *Prunus avium* and deduced that this was due to clonal occurrence. While two of the five *T. sibirica* populations from this study show negative F_{IS} values, these were not significantly different from zero (Table 3.3). A high level of clonality (25% of the samples analysed) was noted in the two UK *Tilia* species (Chapter 2), and although this current study did not address clonal reproduction of *T. sibirica*, it was apparent that some ‘duplicate’ genotypes do exist in some populations. A further investigation into the clonal dynamics of the genus is therefore warranted.

3.5.4 Eastern recolonization of Tilia from the Russian plains?

A hypothetical population was used in the EMM to determine if a putative refugium, perhaps from the Caucasus, Caspian Sea region (Hewitt, 1999) or indeed the central Russian plains (Svenning *et al.*, 2008; Markova *et al.*, 2009), may have contributed to

the recolonization of *T. cordata* into the rest of Europe. Phuekvilai (2014), suggested a possible colonization route from a more easterly refugium. DIYABC analyses of the EMM did indeed show a possible contribution to the recolonization of Europe from a putative refugium in Russia. Possible cryptic refugia at northern latitudes (northern refugia hypothesis) have been reviewed and tested (Willis *et al.*, 2000; Willis and Whittaker, 2000; Stewart and Lister, 2001; Willis and Van Andel, 2004; Bhagwat and Willis, 2008; Svenning *et al.*, 2008; Normand *et al.*, 2011; Väliranta *et al.*, 2011). This has led many authors to suggest that the current paradigm – that central Europe was largely treeless during the LGM – should be revised, but see Tzedakis *et al.,* (2013) for a counter-argument. Nonetheless, the results from this current study suggest at least some contribution from a possible refugium in western Russia into Europe 418 generations ago (around 10.5 kya, based on a generation time of 25 years). This fits with the European post-glacial pollen distribution of *Tilia* (Huntley and Birks, 1983), and with a 7 – 10°C rise in temperature around 10.3 kya, following the Younger Dryas (11.7 kya) which saw a spread of deciduous tree species including *Tilia* (Borzenkova *et al.*, 2015). The date also coincides with the SDM divergence time of 9.2 kya (Appendix 3.7) suggesting an Eastern European/Western Russian LGM refugium. Regarding the likelihood of *T. cordata* surviving at higher latitudes during the maximum extent of ice, there is some evidence that this may have been possible. Svenning *et al.* (2008) modelled the LGM distribution of several boreal and nemoral tree species and concluded that while nemoral species were largely confined to traditional southern refugia, *T. cordata* and a few others may have found suitable conditions further north. Furthermore, Jørgensen *et al.* (2012) provide evidence that suggests areas of Siberia remained ecologically stable during the Late Pleistocene and Bhagwat and Willis (2008) suggest that species with a current northern range of >60°N may have been able to survive at northern latitudes during the LGM. *Tilia cordata* is regarded as a hardy species with a widespread geographical range (65°N in Norway and Finland and 63°N in Russia) and can tolerate severe climates. The species is more tolerant than *Quercus robur* or *Q. petraea* (Pigott, 2012). Indeed, its tolerance for the cold increases during winter in order to protect shoots and they can survive undamaged at -48°C (Pigott, 2012). *Tilia cordata* can clonally reproduce through vegetative spread (Radoglou *et al.*, 2009; Pigott, 2012) so under suboptimum conditions and in the absence of sexual reproduction, the species may reproduce asexually. Bhagwat and Willis (2008), suggest that having the ability to vegetatively propagate in extreme environments can prolong the existence of woody

species at northern locations. It is likely therefore, that if *Tilia cordata* had persisted in northern refugia during the glacial periods, then clonal propagation may have played an important role in its survival.

3.5.5 Model choice and confidence

Posterior probabilities for scenario 4 in the BM (0.553 and 0.772, Figs. 3.9 and 3.10, Appendix 3.10a) were comparable to the average highest probability value (0.685) found across five other DIYABC studies (Bodare *et al.*, 2013; Poudel *et al.*, 2014; Bagnoli *et al.*, 2015; Soliani *et al.*, 2015; Tsuda *et al.*, 2015). However, the posterior probability for scenario 9 in the EMM (0.347, Fig. 3.12, Appendix 3.10b) was lower. Although there is no suggested or recommended value for a high probability, based on the direct approach, the posterior probability for EMM's 'best' scenario was not much higher than the next closest scenario (Fig. 3.12). The 'western migration' interpretation therefore should be taken with caution. More samples from a wider European range could be added to this current study to further investigate the potential contribution of western *Tilia cordata* genotypes to its European distribution.

The goodness-of-fit model checks for each of the 'best' scenarios showed the observed data set positioned within the cluster of data points that represent the posterior predictive distribution. These in turn were positioned within a larger cluster of data points representing the simulated data. The option implemented in DIYABC, provides an opportunity to statistically check scenarios (Estoup *et al.*, 2015), and suggests that the combined parameters are a good fit for each of the respective models.

Likewise, confidence in scenario choice is improved by estimating type I and type II errors. Type I error (true scenario rejected) for scenario 4 of the BM was relatively high at 0.318 and 0.294, while type II error (false scenario not rejected) was low at 0.113 and 0.132, for the direct and logistic approaches, respectively. Type I error for scenario 9 of the EMM was also high at 0.293 and 0.713, while Type II was 0.240 and 0.138, for the direct and logistic approaches, respectively. It has been suggested that lower Type II errors provide good confidence in scenario choice even if Type I errors are large (Bermond *et al.* 2012). However, the present study's results may need to be interpreted with caution.

3.5.6 Constraints of the ABC analyses

Our results from the ABC analyses should be interpreted with caution. There are several factors that may contribute uncertainties from the ABC inference *e.g.* summary statistics used, generation times, mutation rates, models tested, central tendency used and wide credible intervals. In this study, a generation time of 25 years (Collingham and Huntley, 2000; Pigott, 2012) was applied, but as with many forest trees, the age of first flowering in *Tilia* varies considerably, not to mention long living trees tend to cause generations to overlap. *Tilia cordata* can first flower around 12 – 20 years old in unshaded conditions and in shaded conditions this rises to 30 – 40 years old (Pigott, 2012). Using the conservative age of 25 years suggests a rather recent divergence between *T. cordata* and *T. sibirica.* However, using the older generation time of 40 years, the split is still recent with a wide credible interval range (mode value: 14.8 kya, 95% CI: 6.5 – 187.2 kya). *Tilia* can live for many hundreds of years and so generation times could be >100 – 300 years (Pigott, 2012). When taking this into consideration, the divergence time between *T. cordata* and *T. sibirica* extends to 36.9 – 110.7 kya. Species divergence dates for American *Quercus* spp. are much earlier, in the order of millions of years (Cavender-Bares et al., 2015). Bagnoli et al., (2015), estimated divergence times between *Quercus cerris* populations – using a minimum generation time of 30 years – of 261 kya and 1.67 mya. Their study used chloroplast microsatellites (cpSSRs), which are uni-parentally inherited and non-recombinant. Nuclear microsatellites (nSSRs) are multi-allelic and heterozygous. While both cpSSRs and nSSRs can cause size homoplasy (*i.e.* when different alleles evolve independently into the same character state), the latter tends to have higher mutation rates and together (Estoup et al., 2002) these can affect estimations of ancient evolutionary or demographic events (Provan et al., 2001). Hedrick *et al.* (2006) have shown theoretically that divergence can occur more extensively in microsatellite markers due to high polymorphisms, than other molecular markers, such as mitochondrial DNA (mtDNA), which can take thousands of generations (Neigel and Avise, 1986). So the divergence date estimated here, even considering the largest generation times, may be somewhat underestimated. Cornuet et al., (2010), suggest using a combination of molecular markers when testing complex evolutionary scenarios. Further work using DIYABC analyses could combine cpSSRs and nSSRs and sequence data for improved estimation and narrower credible intervals.

3.6 Conclusion

This is the first genetic study carried out on the Siberian lime (*T. sibirica*). It is endemic to only a small number of sites in southern Siberia (Novák et al., 2014). Significant genetic differences were observed between *T. sibirica* and its nearest congeneric *T. cordata.* Combined with ecological and morphological differences (Pigott, 2012), the taxa are therefore considered to be two separate species, supporting Novák et al., (2014) and this study's first hypothesis. The estimated time of divergence was very recent, and when considering a 25 year generation time, contradicts the second hypothesis of a pre-glacial split. However, using a larger generation time *e.g.* 100 years, then the estimates fit a pre-glacial split. The results have shown that the estimated divergence time coincided with a westerly migration of genotypes which may reflect the post-glacial recolonization of Europe.

A phylogeographic study combining the use of cpDNA and single copy nDNA as well as the current set of SSR markers could be carried out using populations of *Tilia* from across its range. This will fully explain the status of *T. sibirica.* It will be interesting to see how similar or dissimilar the results of this model based study compare to future studies.

The molecular markers revealed low genetic diversity within the species, confirming the third hypothesis. It is recommended that an *in situ* and *ex situ* conservation approach, to restore populations in the area and within its natural range, should be considered and defined as soon as possible to ensure that these relict populations are preserved.

Chapter 4: Clonal architecture and diversity of *Tilia cordata*, *T. platyphyllos* and *T. sibirica*

4.1 Abstract

Adopting a clonal reproductive strategy can prolong the lifespan of an individual. Having the ability to clonally reproduce can be advantageous when environmental conditions are not optimal or when sexual reproduction is restricted. However, when clonal growth occurs more than sexual reproduction, the demography and genetics of natural populations may be affected and genotypic diversity may change.

While it is known that *Tilia* regularly propagate asexually, the extent of this occurrence throughout the range of three *Tilia* species, *T. cordata*, *T. platyphyllos* and *T. sibirica,* is unclear. This study addresses the level of clonality and the impact of clonal reproduction on genetic diversity of the three species. Results reveal fewer clones in *T. platyphyllos* than the other two species and that edge-range populations experience greater clonality than central European populations. Clonal occurrence does not appear to have had a negative effect on diversity. However, given the low genetic diversity observed in the rare Siberian lime (*T. sibirica*), it is recommended that an immediate conservation effort to restore populations is undertaken.

4.2 Introduction

The term ‘ancient woodland’ has been used to describe small pockets of important, sometimes protected, and usually isolated UK and European woods (Rackham, 2008). The biodiversity found within these sites is usually richer than in younger woodlands (Peterken, 1983), but rarely are the individuals truly ancient. Individual organisms typically have a finite life span and exist only as long as their form and function remains healthy. When this fails to provide protection or sustenance it will generally result in the death of the individual. However, an organism’s genetic material can persist for thousands of years through continuous clonal reproduction (Parks and Werth, 1993; Jónsdóttir *et al.*, 2000; Wesche *et al.*, 2005; Brundu *et al.*, 2008) and with the exception of somatic mutations, produce genetically identical individuals.

A clonal organism reproduces asexually and by doing so reduces, or in some cases, completely suspends the exchange and recombination of genetic material

with its conspecifics. Clonality is common in plants and it is achieved through vegetative growth (*e.g.* root suckering, root collar growth, epicormic shoots) and agamospermy *i.e.* clonal seed development (Ellstrand and Roose, 1987; Widén *et al.*, 1994). Asexual reproduction has both advantages and disadvantages. Having the ability to reproduce clonally not only prolongs the lifespan of an individual organism but it can lead to a single genet spreading greater distances horizontally by producing multiple, genetically identical ramets (de Witte and Stöcklin, 2010). Furthermore, when sexual reproduction is restricted due to stressful or less optimum conditions, clonal reproduction can ensure greater longevity of an individual (Eckert, 2002; Silvertown, 2008). However, when asexual reproduction occurs more than sexual reproduction, the demography and genetics of natural populations can be strongly affected, for example changes in genotypic diversity (Balloux *et al.*, 2003; Halkett *et al.*, 2005). These changes may be elevated in fragmented woodlands, but may not necessarily be negative. Heinken and Weber (2013) and references therein, suggest that clonal species may cope better with the effects of habitat fragmentation, due to less reliance on pollen development and seed dispersal and due to their potential longevity. So it is important to understand the level of vegetative reproduction in putative clonal and partially clonal species that are subjected to fragmentation.

Tilia (lime or linden), are found in many parts of the northern hemisphere (Pigott, 2012). They now mainly exist in native or 'ancient' woodlands throughout UK and Europe and many of these are fragmented. Species within the genus tend to cope well with cyclic management regimes, such as coppicing (Pigott, 2012; Buckley *et al.*, 2015). Coppicing encourages clonal sprouting, which can result in individual trees (genotypes) dominating local areas. For example, Vaughan *et al.* (2007) observed almost 50% of *Prunus avium* (wild cherry) on managed sites and 65% on unmanaged sites to be clonal, and De Woody *et al.* (2009), observed a clonal group of *Populus tremuloides* (trembling aspen) with 43,000 stems covering an area of 43ha.

In the UK, *Tilia* show significant population differentiation and high genetic diversity (Chapter 2). However, from observations, many natural populations within the UK seem to experience some level of clonal reproduction (*pers. obs.*). At some sites this is aided by land erosion, particularly on steep banks and cliffs (see Appendix 1.3b and 1.6). Individual trees can spread through root collars, epicormic shoots, low hanging branches and fallen trunks (*pers. obs.*). Although *Tilia* are not

regarded as a clonal organism *s.str,* genotyping of *Tilia* stands from the UK and Siberia, revealed some evidence of clonality (see Chapters 2 and 3).

Vegetative growth in *Tilia* has been documented elsewhere by other researchers. For example, references cited in Radoglou *et al.,* (2009) suggest that *T. cordata* has a greater incidence of clonal reproduction than sexual reproduction. The study states that up to 80% of young trees in south-western Russia and nearly all *Tilia* in the north-east of Europe originated from vegetative reproduction. They point out that 90% of the lime trees in the Białowieża National Park (BNP) sprout from root collars, and that this is a range-edge phenomenon in the UK, Finland, and Siberia. However, the references cited are Polish and Russian articles and difficult to access, so at this time, these suggestions cannot be verified. Conversely, it has been noted that *T. cordata* regenerates freely from seedlings in the Białowieża Forest (Pigott, 1975) and indeed in other areas of Poland (Jaworski *et al.*, 2005). Vegetative reproduction has also been proposed as an important survival strategy for *T. sibirica* in *Tilia*-dominated stands of southern Siberia. Adhering to this form of regeneration permits the species to outcompete other forest trees such as *Abies,* other conifers, and *Betula* (Novák *et al.*, 2014). The extent of clonal reproduction and structure of *Tilia* trees across their range is yet to be tested empirically. Therefore, further investigation into the clonal dynamics of the genus is required.

To infer the clonal architecture of *Tilia* stands, microsatellite markers (Phuekvilai and Wolff, 2013) were used. These markers revealed high polymorphism and population structure in UK *T. cordata* and *T. platyphyllos* (see Chapter 2) and in central European stands (Phuekvilai, 2014). Moreover, they proved to be ideal for discriminating between individual genotypes and will be useful in inferring the clonal dynamics of the genus. The aim of this study was to investigate the clonal architecture of *Tilia* stands across its European range, (UK and mainland Europe, Poland, Scandinavia and Russia). In particular, this study investigated; [1] the level of clonality in *T. platyphyllos, T. cordata, T. x europaea* and *T. sibirica* and [2] the impact of clonal reproduction on genetic diversity of the three species.

The expectations were that *T. cordata* will be more clonal at locations on the margins of its natural range *e.g.* UK, Danish, Finnish, Russian, and possibly Poland sites due range-edge effects at the northern limit, competition and predation (Pigott, 1981a; Pigott and Huntley, 1981; Radoglou *et al.*, 2009). In contrast, central European sites *e.g.* Germany, Austria, and France, were

expected to show less clonal reproduction as sexual regeneration occurs more frequently and seeds tend to be viable (Pigott, 1981a). The same trend was expected expected to be found in *T. platyphyllos, i.e.* clonality was expected to be higher in Denmark than in central ranges such as Germany and Austria. In the UK *T. platyphyllos* frequently produce fertile seeds and freely regenerate (Pigott, 2012), so clones were expected to be minimal at these sites. Regarding *T. sibirica,* Novák *et al.* (2014) points out that reproduction via clonal propagation is likely an important survival strategy for the species. Given the rarity of sexual recruitment within *T. sibirica*, as a result of intensive logging, regular competition and disease (Novák *et al.*, 2014), the expectation was that there would be greater clonality observed in *T. sibirica i.e.* genotypic richness would be low compared to the other species.

4.3 Material and Methods

4.3.1 Study sites and sample collection

To examine the extent of clonal structure, three *Tilia* species from across Europe and Russia and the hybrid of *T. cordata* and *T. platyphyllos* were the focus of this current study (Fig. 4.1). *Tilia cordata* from thirteen locations in the UK (N=261), three sites from the Białowieża National Park, Poland (N=42), and two sites from the Vagay region, Russia (N=33), as well as *T. platyphyllos* from nine UK locations (N=160) and *T. sibirica* from six Kuzedeevo sites in the Kemerovo region of Russia were sampled (N=113, Table 4.1). Additionally, raw microsatellite data from *Tilia cordata* and *T. platyphyllos* were accessed from the *Tilia* SSR database, here at Newcastle University. These included samples from seven countries either at the centre of the *Tilia* range or on the range margins.

T. x europaea occurs naturally in the UK and Europe where *T. cordata* and *T. platyphyllos* are sympatric (or at least were once sympatric). In natural populations hybrids are infrequent but like the pure species are capable of clonal propagation. The hybrid populations were identified from the true species following Logan *et al.,* (2015) and analysed as a separate taxon.

UK samples were mainly collected from Sites of Special Scientific Interest (SSSI) and National Nature Reserves (NNR). These sites have been documented to have *T. cordata* or *T. platyphyllos* present (Natural England, 2014). Leaf samples from the two species and the hybrid were also sampled from other locations across England and Wales. *Tilia sibirica* and *T. cordata* leaf samples were obtained from Russia. When possible, the Russian samples were collected at up to 20m apart. However,

some sites were small and so the distance between each sample was less (M. Chytrý *pers. comm.*).

Most of the hybrid samples were collected from sites near both sides of the Welsh border where either, or both, of the UK species occurs. Naturally occurring hybrids were also collected from sites further north (Table 4.1). As *Tilia* populations in the UK are generally small and fragmented, in most cases samples were collected from as many individual trees available from any particular site, with the exception of St. Pierre Great Wood (SP) in Wales. This woodland had both species of *Tilia* and the hybrid present. At this site, more than 50 trees were sampled at the north east corner of the woodland from two plots, one of approximately 50x100m, the other, just over 250m to the west, was approximately 20x70m. It is possible that not all trees were sampled from within the particular sampling area at all sites.

All *T. cordata* samples from Poland were collected from the Białowieża National Park (BNP), from three plots that had previously been identified to have *Tilia* present (Pigott, 1975). A 30x30m quadrat was set out within each plot and leaves within reaching distance on trees, were collected. *Tilia* in Poland can reach heights of 35-50m (Pigott, 1975; Jaworski *et al.*, 2005; Wesolowski and Rowinski, 2006). With trees of these heights, no samples could be collected. Additionally, there were a large number of seedlings and saplings at each of the Polish sites. Only adult trees were considered for comparative clonal analyses. However, juvenile trees from the Polish sites were also assessed for clonality to provide an estimate of clonal recruitment at this part of the *T. cordata* range. All samples were dried at room temperature and stored at -20°C until required for DNA extraction.

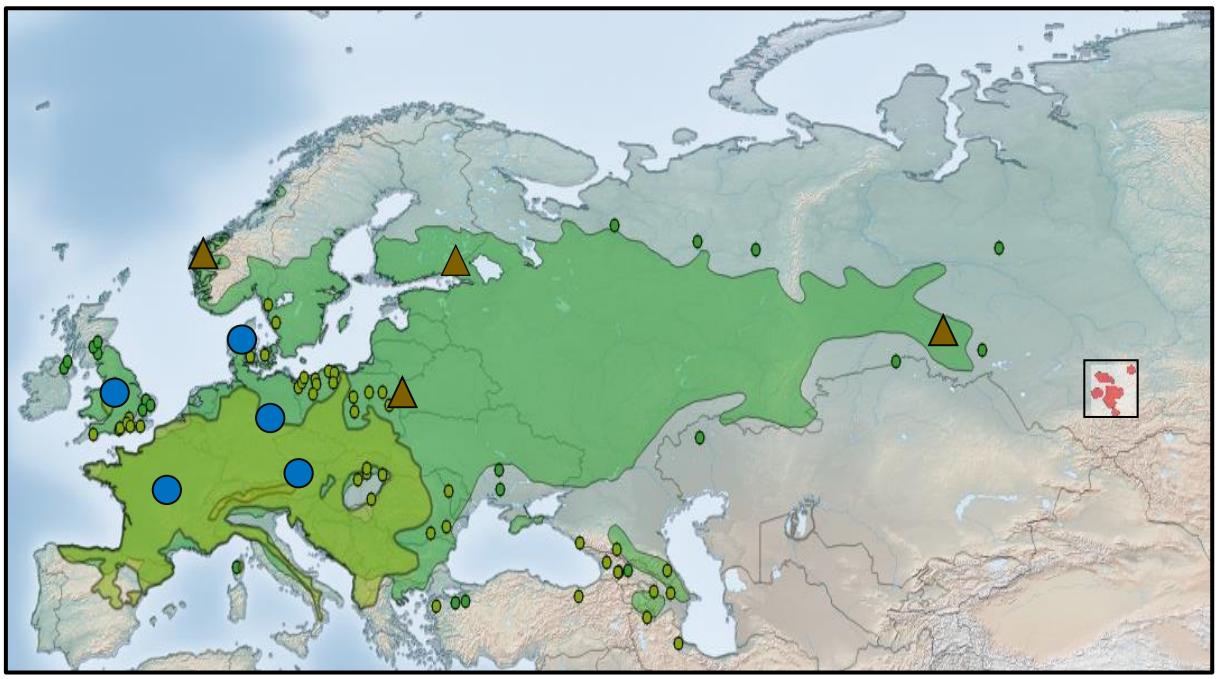


Figure 4.1 Countries and regions where samples of *Tilia* were collected and analysed for clonality in this study. Blue circles are countries where *T. platyphyllos* and *T cordata* were sampled, brown triangles are countries where only *T. cordata* were sampled, and the orange area is sites where only *T. sibirica* were sampled.

Table 4.1 Species, country, sites, population codes and coordinates of samples used for clonal analyses.

Species	Country	Location	Code	Latitude (°N)	Longitude (°E)
T. platyphyllos	United Kingdom (UK)	Anston Stone Wood[1]	AW	53.3402	-1.1985
		Applegarth Scar[1]	AS	54.4093	-1.8165
		Barton Hills[1]	BH	51.9562	-0.4220
		Braitwaite[2]	BR	54.2736	-1.8076
		Hudswell[2]	HD	54.4012	-1.7573
	France (FR	Cessieres	CE	49.5588	3.4888
		Gorges de la Caran[3]	CC	43.3743	2.4521
		Issole[3]	IS	43.2721	0.1142
	Denmark (DK)	Bolderskev[3]	BO	55.0168	9.3890
	Austria (AU)	Dobra[3]	DO	47.9860	16.6956
		Leopoldsberg[3]	LE	48.2771	16.3547
		Sommerein[3]	SO	48.5906	15.3974
		Thayatal Park[3]	TH	48.8475	15.8800
	Germany (GE)	Lichtenstein[3]	LC	51.3028	13.0188
T. x europaea	United Kingdom (UK)	Whitcliffe Wood[1]	WW	54.4097	-1.7792
T. platyphyllos and *T. cordata* mixed[5] sites	United Kingdom (UK)	Brockhill Wood[1]	BW	52.0928	-2.3538
		Chanstone Wood[1]	CW	52.0111	-2.9422
		Halesend Wood[1]	HW	52.1409	-2.3815
		Knapp & Papermill[1]	KP	52.1615	-2.3713
		Lady Park Wood[1]	LP	51.8245	-2.6568
		St. Pierre[1]	SP	51.6348	-2.7195
		West Malvern[1]	WM	52.1189	-2.3581

[1] Collected 2012 – 2014 by Samuel Logan. [2] Collected in 2014 by Seb Mankelow (National Trust). [3] Collected 2012 – 2014 by Dr Kirsten Wolff and colleagues. [4] Collected in 2012 by Dr Paul Ashton (Edge Hill University). [5] Mixed - two species occurring with or without the hybrid or one species with the hybrid.

Table 4.1. cont.

T. cordata	United Kingdom (UK)	Bedford Purlieus[4]	BP	52.5833	-0.4646
		Brignall Banks[1]	BB	54.4968	-1.9114
		Collyweston Wood[4]	CG	52.5974	-0.5176
		Dowles Brooke[4]	DB	52.3834	-2.3364
		Easton Hornstock[4]	EH	52.5912	-0.4981
		Hardy Gang Wood[4]	HG	53.2607	-0.3614
		Roudsea Wood[4]	RW	54.2332	-3.0255
		Shrawley Wood[4]	SH	52.2917	-2.2820
		Skelghyll Wood[4]	SK	54.4197	-2.9519
	France (FR)	Foret Dom de Mountiers[3]	FM	48.9139	4.9144
		Cessieres[3]	CE	49.5588	3.4888
	Denmark (DK)	Åbybjerg[3]	AB	57.1024	9.3531
		Bolderslev[3]	BO	55.0168	9.3890
	Austria (AU)	Dobra[3]	DO	47.9860	16.6956
		Stams[3]	ST	47.2757	10.9772
		Sommerein[3]	SO	48.5906	15.3974
		Thayatal Park[3]	TH	48.8475	15.8800
	Germany (GE)	Colbitz[3]	CO	52.3303	11.5572
		Wasserwerk[3]			
	Norway (NO)	Sogn og Fjordane[3]	SO	61.8505	6.13754
	Finland (FI)	Muukonsaari[3]	MU	61.1600	28.4820
		Niinisaar[3]	NI	61.8171	29.3895
	Poland (PL)	Białowieża Park Narodowy[1]	B69	52.7289	23.8328
			B99	52.7186	23.8443
			B40	52.7336	23.8319
	Russia (RU)	Vagay Region[6]	V20	57.5097	69.1956
			V25	57.9304	68.9099
T. sibirica	Russia (RU)	Kuzedeevo, Kemerovo Region[6]	K22	53.17-53.21	87.13-87.21
			K28	53.17-53.21	87.13-87.21
			K12	53.17-53.21	87.13-87.21
			K21	53.17-53.21	87.13-87.21
			K29	53.17-53.21	87.13-87.21
			K38	53.17-53.21	87.13-87.21[7]

[6] Collected in July and August 2012 by Dr Milan Chytrý and colleagues (Masaryk University, Czech Republic). [7] Kuzedeevo - approximate range of sites.

4.3.2 DNA extraction and microsatellite genotyping:

Genomic DNA was extracted from leaf tissue using the CTAB method. A multiplex Polymerase Chain Reaction (PCR) procedure was carried out to amplify eleven microsatellite regions (Phuekvilai and Wolff, 2013). PCR conditions and parameters were as described in Phuekvilai and Wolff (2013). Microsatellites were genotyped using an ABI 3130XL Genetic Analyser (Applied Biosystems), and scored using Genemapper (Applied Biosystems). Microsatellite fragments were binned manually and checked for inconsistencies.

4.3.3 Clonal structure and genetic analyses:

As clonal determination and analyses can be affected by samples with missing data, only those that did not have missing values were analysed. Samples from species within each country were pooled for clonal analyses and reported here (clonal diversities for populations from UK, Russia, Poland, Denmark and Finland are provided separately Appendix 4.3). To determine if in fact clonality is observed within each species, the presence of clones, *i.e.* individuals with identical multi-locus genotypes (MLG) within each species/region, was checked in GenAlEx v6.5 using the '*Find Clones'* function (Peakall and Smouse, 2012). The 'proportion distinguishable' (*Pd*) is presented as *G/N*, where *G* is the number of genotypes and *N* is the number of sampled individuals (Ellstrand and Roose, 1987).

The Probability of Identity (*PI*)*, i.e.* the average probability of two unrelated individuals within a randomly mating population sharing the same MLG by chance was also estimated. This statistic is often used to assess the power of molecular markers (Peakall and Smouse, 2012) and if low values are observed, will permit the justification of their use (Waits *et al.*, 2001).

Individuals sharing MLG were also identified in GenClone v2.0 (Arnaud-Haond and Belkhir, 2007). The program implements a resampling procedure (1000 permutations in this study), to determine the number of MLG within each population and provides a modified measure of *Pd* that is less affected by genet size, *i.e.* genotypic richness *R* = (*G-1/N-1*). These values will always show '0' when stands consist of a single clone and '1' when all sampled individuals are separate genets (Dorken and Eckert, 2001).

A useful function provided by the program is the identification of somatic mutations and/or scoring errors. Samples with single allele differences were identified as putative somatic mutations or scoring errors. If the latter was

determined these samples were re-extracted and re-analysed following the above methods (see section 4.3.2).

To further describe clonal heterogeneity, an adapted estimate of the Simpson's complement index (*i.e. D**), independent of sample size (Pielou, 1969) was estimated in GenClone. The Simpson index (*D*) represents the probability that two random samples will belong to the same species (Simpson, 1949) and is widely used in ecology. The Simpson's complement index (1-*D*) of diversity is commonly reported in clonal studies as *D** (Arnaud-Haond *et al.*, 2007) and ranges from '0' when all individuals within a population are clonal to '1' when all individuals are unique.

Additionally, the statistics P_{gen}, the probability of samples having the same MLG by chance following the methods of Parks and Werth (1993) and P_{sex}, the probability that a repeated MLG originates from sexual reproduction at the first reencounter, were also estimated. P_{gen} assumes populations are in Hardy-Weinberg Equilibrium (HWE). An adjusted measure P_{gen} (*f*), that takes into account HW departure can also be estimated, providing a more conservative estimate of P_{sex} (Arnaud-Haond *et al.*, 2007). To determine if there were deviations from HWE, data were checked using GENEPOP on the web v4.2 (Raymond and Rousset, 1995; Rousset, 2008) and the presence of null alleles tested in MICRO-CHECKER v2.2.3 (Van Oosterhout *et al.*, 2004).

Standard population genetic diversity statistics, such as observed heterozygosity (H_O), expected heterozygosity (H_E), and percentage of polymorphic loci (*P*), were calculated with and without repeated MLG, in GenAlEx, for only the UK, Poland, and Siberian samples. Statistical tests were carried out in R v3.1.3 (R Development Core Team 2015, www.r-project.org). The inbreeding coefficient (F_{IS}) is a measure of inbreeding within populations. F_{IS} and their significance, were calculated in FSTAT v2.9.3.2 (Goudet, 1995; Goudet, 2001).

4.4 Results

The microsatellite markers used in this study successfully identified the presence of clones in all three species. Four of the five countries where *T. platyphyllos* were sampled and five of the nine countries where *T. cordata* were sampled, as well as *T. sibirica* and the hybrid showed evidence of clonal occurrence.

4.4.1 Clone identification and structure:

The Probability of Identity (*PI*) for each species within each country sampled were very low (*PI* <0.001, Table 4.2). This provides a reasonable justification that samples with the same MLG were not identical by chance and allows the assignment of samples to genets to be accepted (Waits *et al.*, 2001).

In total, 663 trees were genotyped at eleven microsatellite loci. With the raw data from central European populations included, this made a total of 1110 individuals analysed for clonal diversity and structure. From this, 263 individuals were clonal (24% of samples) making up a total of 931 MLG, of which 84 were clonal (9%). This is considered a conservative estimate of clonality within the three species as possible somatic mutations were treated here as unique genotypes. No identical genotypes were found between different locations or countries.

Reaching an asymptotic curve following the resampling procedure in GenClone (Figure 4.2), indicates that enough loci (*i.e.* sampling effort) have been screened to adequately identify MLG (Arnaud-Haond *et al.*, 2005). Indeed, it is clear that genotypic resolution can be achieved using fewer than eleven molecular markers: in *T. cordata* five to seven, depending on the region of the samples and in *T. platyphyllos* five to ten markers, depending on the region (Appendix 4.1).

Deviations from Hardy-Weinberg equilibrium were observed at some loci (data not shown). Therefore, the probability of identical genotypes arising by chance in a randomly mating population (P_{gen}) and the probability that those repeated genotypes (at the first re-encounter), were the result of sexual reproduction (P_{sex}) – taking into account deviations from HWE – were calculated. P_{gen} (f) and P_{sex} (f) were <0.001 at all locations (Table 4.2 and Appendix 4.2), suggesting that repeated genotypes are in fact clones.

Table 4.2 The probability of identity (*PI*) in *Tilia* taxa across their European and Siberian range and average P_{gen} *(f)* and P_{sex} *(f)* for each sampled region.

Species	Region	*PI*	P_{gen} *(f)*[8]	P_{sex} *(f)*[9]
T. platyphyllos UK	UK	5.3E-14	5.18E-14	1.29E-12
	FR	7.0E-15	8.23E-15	–
	DK	3.8E-09	4.89E-08	8.36E-07
	AU	3.1E-15	3.09E-15	4.00E-15
	GE	6.3E-13	1.92E-13	–
T. x. europaea	UK	2.5E-13	3.30E-12	2.45E-11
T. cordata	UK	1.1E-09	1.66E-09	2.98E-07
	FR	3.3E-15	6.42E-09	3.62E-08
	DK	1.2E-09	7.73E-10	3.87E-09
	AU	1.8E-10	8.11E-11	–
	GE	1.1E-08	1.76E-08	–
	NO	1.6E-07	1.84E-07	–
	FI	3.7E-08	3.56E-08	2.62E-07
	PL	2.1E-09	6.82E-10	–
	RU	5.6E-09	3.14E-09	2.18E-08
T. sibirica	RU	5.6E-05	4.82E-05	2.21E-04

[8] P_{gen} *(f)* - the probability of identical genotypes arising by chance in a randomly mating population.

[9] P_{sex} *(f)* - the probability that repeated genotypes (at the first re-encounter), were the result of sexual reproduction. Both values take into account deviations from HWE.

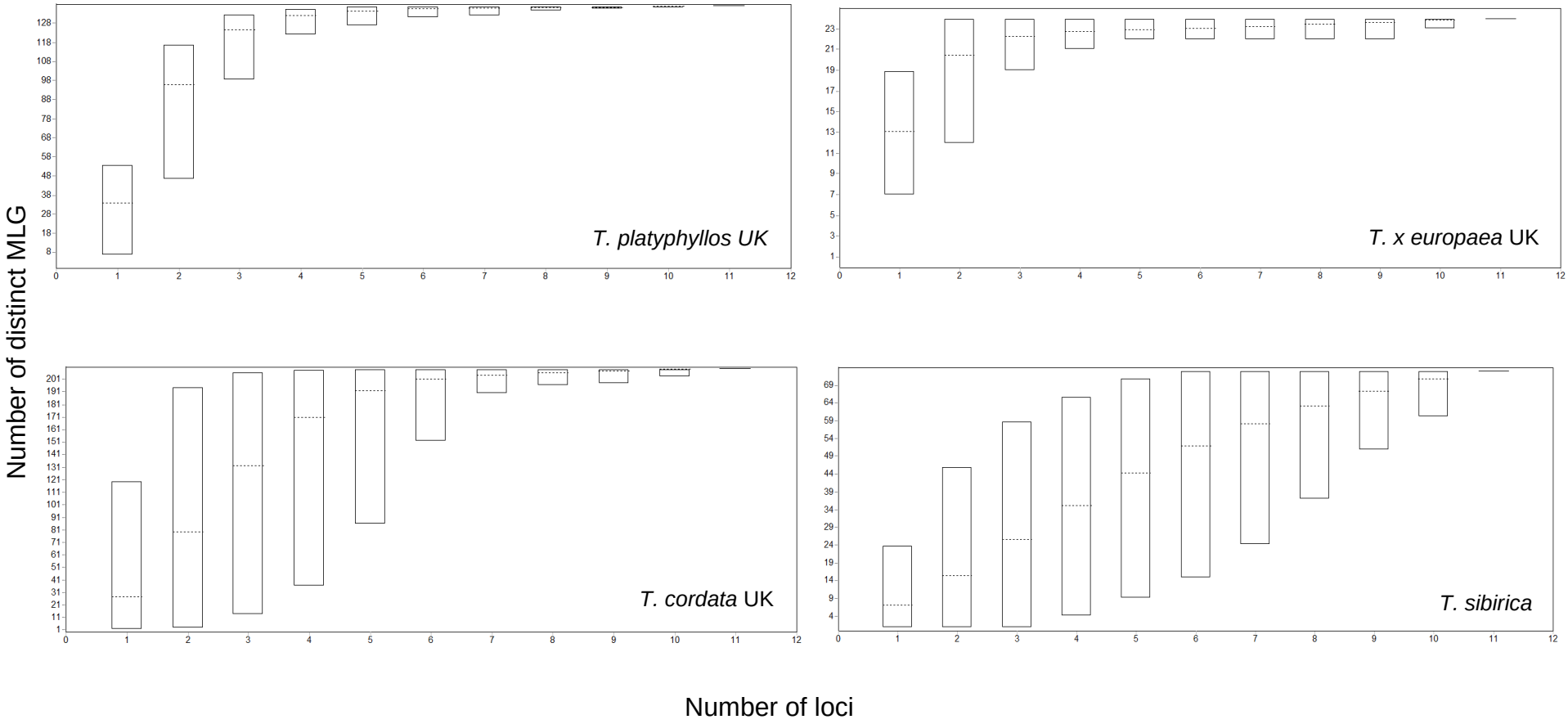


Figure 4.2 Distinct number of MLG and number of loci required to describe the genotypic resolution of *Tilia platyphyllos* UK, *T. x europaea* UK, *T. cordata* UK, and *T. sibirica*. Box plots show the minimum, maximum and average MLG detected within each taxon. (Other regions are shown in Appendix 4.1).

Sample sizes ranged from 11 – 160 across sample sites in *T. platyphyllos* and 16 – 261 in *T. cordata*. Sample size was 54 in *T. x. europaea,* and 113 in *T. sibirica*. MLG identified by GenClone ranged from nine to 137 in *T. platyphyllos*, and 16 to 210 in *T. cordata*. There were 24 MLG in *T. x. europaea* and 73 in *T. sibirica* (Table 4.3 and Appendix 4.3). The hybrid was considered for clonal analysis as the results provide an estimation of clonal structure within the taxon and can be broadly compared with the three species, particularly the UK *Tilia cordata* and *T. platyphyllos.*

Genotypic diversity (inversely measures clonal occurrence), was estimated as the proportion distinguishable (*Pd*) and genotypic richness (*R*). Total values across each region within taxa are reported in Table 4.3. Clonal heterogeneity measured as the Simpson's complement index for genotypic diversity (*D**) across the whole sampling area was high, ranging from 0.904 to 1.000 (Table 4.3), suggesting a high probability that two random samples were genetically different.

The total number of putative somatic mutations (μ_{SOM}) *i.e.* single loci with allele size difference, across all species/regions ranged from 0 - 24 (Table 4.3). In total 44 loci with single allele differences were observed across all sampled taxa; five in *T. platyphyllos*, one in *T. x europaea*, 14 in *T. cordata*, and 24 in *T. sibirica* (Table 4.3). From the 44 loci, 25 differed by just one mutational step (*i.e.* 2bp difference). At the other 19 loci, the differences in allele size were greater (up to 13 mutational steps *i.e.* 26bp difference). Most allele differences were found at marker Tc963 (18 in total). Five of these could have resulted from the presence of possible null alleles. Null alleles were observed at different loci across different populations. In total 24 single allele differences across all sampled taxa were found at loci with possible null alleles. However, 20 putative somatic mutations ranging from 1 to 13 mutational step differences (2bp – 26bp) were observed at eight loci in the combined dataset.

Table 4.3 Estimates of clonal occurrence of adult trees (except [10]), presented as *Pd, R,* and *D** in *Tilia platyphyllos, T. cordata;* the hybrid *T. x europaea* and *T. sibirica* across all locations. *N* - total number of samples; *G* - number of genotypes; *Pd* - proportion distinguishable; *R* - genotypic richness; μ_{SOM} - the number of putative somatic mutations; *D** - Simpson's complement index for genotypic diversity.

Species	Region	*N*	*G*	μ_{SOM}	*Pd* (*G/N*)	*R* (*G-1/N-1*)	*D**
T. platyphyllos	UK	160	137	3	0.856	0.855	0.998
	FR	52	52	0	1.000	1.000	1.000
	DK	11	9	1	0.818	0.800	0.945
	AU	56	53	0	0.946	0.945	0.998
	GE	31	31	0	1.000	1.000	1.000
T. x europaea	UK	54	24	1	0.444	0.434	0.904
T. cordata	UK	261	210	12	0.805	0.804	0.997
	FR	28	27	0	0.964	0.963	0.997
	DK	60	49	2	0.817	0.814	0.985
	AU	83	83	0	1.000	1.000	1.000
	GE	20	20	0	1.000	1.000	1.000
	NO	16	16	0	1.000	1.000	1.000
	FI	40	30	0	0.750	0.744	0.973
	PL	42	42	0	1.000	1.000	1.000
	PL[10]	140	135	0	0.964	0.964	0.999
	RU	33	27	0	0.818	0.813	0.985
T. sibirica	RU	113	73	24	0.646	0.643	0.977

[10] Both adult and juvenile trees sampled for clonal analyses.

4.4.2 Genetic diversity within species/region

Summary statistics of UK, Poland and Russian populations without the repeated genotypes – meaning that every genotype was only represented once – are presented (Table 4.4). Omitting the repeated genotypes inevitably reduced the number of samples within some populations. The number of alleles ranged from 3.64 – 7.09 in *T. platyphyllos*, 2 – 6 in *T. x europaea*, 1.91 – 5.45 in *T. cordata* UK, 5 – 6 in Polish *T. cordata*, 5 – 5.09 in Russian *T. cordata* and 1.45 - 2.91 in *T. sibirica* (Table 4.4 and Appendix 4.5). There were no significant differences in the number of alleles between *T. cordata* populations; between *T. x europaea* and *T. cordata*; or between *T. platyphyllos* and *T. cordata* from Poland or Russia (Kruskal-Wallis with pairwise Mann-Whitney U Post-Hoc test, $P > 0.05$). However, significant differences were observed between *T. sibirica* and the two UK species (*T. platyphyllos* $P = 0.002$, *T. cordata* $P = 0.003$) and both the Russian and Polish *T. cordata* populations ($P = 0.049$ and $P = 0.027$ respectively). *T. x europaea* was significantly different to *T. cordata* Poland ($P = 0.026$), *T. platyphyllos* ($P = 0.045$), and *T. sibirica* ($P = 0.037$).

Genetic diversity, measured as observed and expected heterozygosity was significantly higher (Kruskal-Wallis with pairwise Mann-Whitney U Post-Hoc test, $P < 0.05$) in *T. platyphyllos* UK (0.72 and 0.70) than in *T. cordata* UK (0.53 and 0.53) and both were significantly lower ($P = 0.002$ and $P = 0.038$, $P < 0.001$ and $P < 0.001$) than the hybrid (0.91 and 0.79). There were no significant differences between *T. cordata* from Poland (0.57 and 0.55) and those in the UK or Russia (0.52 and 0.58, $P > 0.05$) nor between Russian and UK populations ($P > 0.05$). *T. sibirica* showed significantly lower genetic diversity (0.29 and 0.29) than *T. platyphyllos* ($P = 0.002$) and *T. cordata* in UK and Poland ($P < 0.05$). *T. x europaea* showed significantly higher genetic diversity than *T. cordata* in UK ($P < 0.001$) and Poland ($P < 0.05$); *T. platyphyllos* ($P < 0.05$); and *T. sibirica* ($P < 0.01$).

The proportion of polymorphic loci was high in *T. platyphyllos* and *T. cordata* but low in *T. sibirica*. Mean F_{IS} values ranged from -0.267 in *T. x europaea* to 0.097 in *T. cordata* Russia (Table 4.4). Two *T. sibirica* populations (K28 and K29) showed positive F_{IS} values significantly different from zero (Appendix 4.5).

Table 4.4 Average diversity indices (and SE) for UK *T. platyphyllos* and *T. x europaea*, and for *T. cordata* from three regions and *T. sibirica*, excluding repeated genotypes (clones). *N* - number of samples, *A* - number of alleles, *P* - Proportion of polymorphic loci, H_O - Observed heterozygosity, H_E - Nei's unbiased Expected heterozygosity, F_{IS} – Inbreeding coefficient.

Species/Region	*N*	*A*	*P*	H_O	H_E	F_{IS}[12]
T. platyphyllos	15.22	5.43 (0.18)	0.99 (0.01)	0.72 (0.02)	0.70 (0.01)	-0.041
T. x europaea	3.83	3.86 (0.21)	0.99 (0.02)	0.91 (0.03)	0.79 (0.02)	-0.267
T. cordata UK	16.15	4.42 (0.28)	0.83 (0.02)	0.53 (0.03)	0.53 (0.03)	-0.011
T. cordata PL	14.00	5.33 (0.58)	0.91 (0.00)	0.57 (0.06)	0.55 (0.05)	-0.049
T. cordata RU	13.50	5.05 (0.69)	0.91 (0.00)	0.52 (0.06)	0.58 (0.05)	0.097
T. sibirica	12.17	2.38 (0.21)	0.56 (0.07)	0.29 (0.04)	0.29 (0.04)	-0.078

4.5 Discussion

This study examined the clonal architecture of three species of *Tilia* (*T. cordata* and *T. platyphyllos*) from Europe and *T. sibirica* from Russia. Results revealed varying levels of clonality (Table 4.3), which both confirmed and contradicted some prior expectations.

4.5.1 Resolution of the markers

This is the first molecular study to investigate clonal occurrence of the three *Tilia* species from across much of their range. The markers used in this study effectively distinguished between *Tilia cordata* and *T. platyphyllos* within the UK (see Chapter 2) and across European populations (Phuekvilai, 2014), and between *T. cordata* and *T. sibirica* (see Chapter 3), so were considered ideal for assessing clonal analysis. The microsatellites were enough to adequately resolve the identification of MLG within the three species. All MLG are unique to each population, indicating no inter-population movement, neither naturally nor artificial. However, two similar genotypes – with a 7bp single allele difference – were observed in two *T. sibirica* populations. It is not expected that this is a somatic mutation as the two genotypes are found in trees at different locations. In all but *T. sibirica* it is clear that fewer molecular markers are required (Figure 4.2 and Appendix 4.1). Further studies using these microsatellite markers to investigate clonality in other *Tilia cordata* or *T. platyphyllos* populations may only require between five and ten of the most polymorphic markers.

Departure from HWE was observed at some loci in some locations. So the adjusted measures P_{gen} (*f*) and P_{sex} (*f*) were presented. The probability of trees having the same MLG by chance was extremely low P_{gen} (*f*) <0.001 and equally low that any identical MLG within a location arose via sexual reproduction P_{sex} (*f*) <0.001 (Table 4.2 and Appendix 4.2). This provides evidence that individuals sharing MLG are indeed clones arising from clonal reproduction (Arnaud-Haond *et al.*, 2005).

4.5.2 Overall low clonality in lime

In general, clonality was low in *T. cordata* and *T. platyphyllos* but relatively high in *T. sibirica*. This was indicated by high and low genotypic richness respectively (Table 4.3 and Appendix 4.3). Genotypic richness or the Proportion distinguishable (*R* or *Pd*) are indirect measures for quantifying sexual versus asexual reproduction (Silvertown, 2008) and can be broadly compared across taxa. Clonal occurrence varies greatly across many species of plants (Ellstrand and Roose, 1987) and populations (Dorken and Eckert, 2001). As deduced from published literature, clonality varies within the model tree genus *Populus,* and across other tree species *e.g. Prunus avium, Prunus ssiori, Ulmus minor, Sorbus torminalis, Cryptomeria japonica* and *Bashania fangiana*, using presented *R* values, or calculating them from those studies that did not directly present them (Shimizu *et al.*, 2002; Nagamitsu *et al.*, 2004; Suvanto and Latva‐Karjanmaa, 2005; Stoeckel *et al.*, 2006; Vaughan *et al.*, 2007; Mock *et al.*, 2008; Rasmussen and Kollmann, 2008; Van Loo *et al.*, 2008; De Woody *et al.*, 2009; Chenault *et al.*, 2011; Ma *et al.*, 2013; Santos-del-Blanco *et al.*, 2013; Fuentes-Utrilla *et al.*, 2014).

Clonal contribution in *Tilia* measured as *Pd* (*G*/N) and *D** (0.444 – 1.000 and 0.904 – 1.000, respectively) was much lower than the average clonal contribution calculated from 21 clonal plant species *i.e. Pd* = 0.167,(0.002 – 1.000), and *D* = 0.620, (0.130 – 1.000, (Ellstrand and Roose, 1987). Interestingly, the only tree species included in their study was *Populus tremuloides* and sampling 222 individuals from seven populations revealed no clonality (*Pd* and *D* were both 1.000). Genotypic richness (*R*), averaged across all *T. platyphyllos*, *T. cordata*, *T. x europaea*, and *T. sibirica* populations was 0.920, 0.904, 0.434, and 0.643, respectively (Table 4.3). Whereas, these values revealed relatively low levels of clonality within *Tilia platyphyllos* and *T. cordata* compared to other forest tree species, clonality within *T. sibirica* was comparable to *e.g. Populus* sp. (Suvanto and Latva‐Karjanmaa, 2005; Brundu *et al.*, 2008; Mock *et al.*, 2008; Van Loo *et al.*,

2008; De Woody *et al.*, 2009; Chenault *et al.*, 2011; Santos-del-Blanco *et al.*, 2013; Cristóbal *et al.*, 2014), *Prunus* (Nagamitsu *et al.*, 2004; Stoeckel *et al.*, 2006; Vaughan *et al.*, 2007), and *Ulmus* (Fuentes-Utrilla *et al.*, 2014). This suggests that clonality may be an important reproductive strategy for *T. sibirica.* Despite being isolated and highly fragmented, clonal reproduction might maintain genetic diversity, albeit at low levels observed within the species. With 24 putative somatic mutations found across locations (Table 4.3), the level of clonality described here may not be truly indicative of the species as genotypes with somatic mutations were reported here as different MLG. Somatic mutations within the species may contribute to higher genetic diversity in the future.

The highest occurrence of clones in *T. cordata* was observed at two locations namely West Malvern (WM) and Brignal Banks (BB), as revealed by lower genotypic richness (R = 0.111 and 0.389, respectively, Appendix 4.3). Coppicing has been and still is performed as a management technique in the UK and this type of management produces multiple trunks. Multiple trunks can initiate clonal spread and increase clonal propagation when low hanging branches take root several metres from the trunk (*pers. obs.* See appendices). Past coppicing and branch spread is likely the case at WM, but is not apparent at BB. At WM, a small cluster of lime trees are located on the Malvern Hills in Herefordshire and past management at this site almost certainly included coppicing. The clones at WM were aggregated in a ringed shape, one of which was more than four metres across. The numerous trees that made up each circle were identified in this study as clonal. It is likely that an older coppiced *T. cordata* tree produced multiple stems that eventually spread out and subsequently created several independent, viable trees. Alternatively, this could have been a very large tree, several hundred years old that subsequently died and the trees that exist now are remnants of growth from the root collar. Although clone size does not necessarily correlate to clone age (Ally *et al.*, 2008), if this was the case in this particular group, this clone could have been ~800 years old (Pigott, 2012). Very large and old *Tilia* trees are found in various parts of Europe. In the UK, one of the largest, possibly oldest *T. cordata* individual can be found in Herefordshire. It is an old gnarly, almost hollow tree, ~10m circumference (*pers. obs.* Appendix 1.5b) and likely to be an Enclosures Act boundary tree (Dr R. Roseff *pers. comm.*). The tree is probably ~800 years old. A large clonal group, estimated to be approximately 2000 years old can be found at Westonbirt Arboretum, (Forestry Commission, 2015). Although, the age of the

Westonbirt group cannot be confirmed at present, we have confirmed that it is indeed a very large clone of some 22 individual trees with a number of possible somatic mutations (Logan and Wolff, *unpub. data*).

Self-coppicing also occurs at some sites, in particular on steep embankments and at these locations vegetative spread is more apparent (*pers. obs*.). This is likely the reason clonality occurs at BB. Many of the *T. cordata* trees grow on steep banks and clones occur from root collars spreading down the slope. This is also observed among *T. x europaea* trees at BB and at other locations where the hybrid occurs *e.g.* Whitclife Wood (WW).

Tilia platyphyllos in the UK showed various levels of clonality. *R* ranged from 0.600 – 0.958, mean *R* = 0.814 (Appendix 4.3). Genotypic richness in the UK was similar to *T. platyphyllos* in Denmark and lower than populations in the central range of mainland Europe *i.e.* Austria, Germany and France (Table 4.3). Pigott (2012) states that sexual regeneration of *T. platyphyllos* occurs naturally and frequently in the UK and this current study has confirmed that clonal reproduction does not appear to have an influential role compared to some *T. cordata* and *T. sibirica* sites and that of other forest trees.

Outside the UK, relatively high clonality in *T. cordata* was observed at one of the Danish locations (BO) as revealed by lower genotypic richness (*R* = 0.621, Appendix 4.3). This is a mature ancient woodland with both species of *Tilia* present (Lawesson, 2004). Indeed, *T. platyphyllos* at this site also shows evidence of clonal growth, albeit to a lesser degree than *T. cordata*, but slightly higher than that found in central European populations (Table 4.3. and Appendix 4.3). The two Finnish *T. cordata* populations also show relatively high clonality (*R* = 0.744*,* Table 4.3). Clonal growth is not surprising given the location of these sites at the edge of the *T. cordata* range and the infrequent production of fertile ovules (Pigott, 1981a).

Although references in Pigott (1975) state that fertile fruit are produced in Finland during very warm summers, it is clear that asexual reproduction is an influential mechanism maintaining Finnish – and to a certain extent, the Danish – populations compared to other locations. In Finland, *Tilia* compete with *Picea abies* at the northern edge of its range (Pigott, 2012), while in Denmark, *Tilia* are scarce in many woodlands and where they do occur, they compete with a rich structure at both the tree and herb layer (Lawesson, 2004). Clonal reproduction may therefore improve the survival of *Tilia* in these marginal and competitive areas.

Clonality at the two Russian *T. cordata* sites varied greatly *i.e.* 0.684 and 1.000 (Appendix 4.3). Whereas, one population appears to adhere to 'clonal edge-range effects' as expected, the other population does not show this. The influence of climate at range boundaries can have a prolonged effect on seed production, tree growth and mode of reproduction (Pigott, 1968; Pigott, 1981a; Pigott and Huntley, 1981; Parsons, 1991). Monthly temperatures at the eastern limits of *T. cordata* range from -20°C to 25°C and while this is well within the tolerance range for *T. cordata* (Pigott, 2012), it is possible that climate affects some *Tilia* populations and not others. Alternatively, clonal propagation at only one site could be due to the topography, as observed at some UK sites *e.g.* Brignal Banks.

While clonality was observed at the Finnish and Danish sites, there were no clones found in the Norwegian population, which is also regarded to be close to the range edge (Pigott, 2012). However, only one population was sampled, and so inference about the Norwegian reproductive strategy is limited. Although, the Norwegian and one Danish population showed no clonal growth following genetic analysis, the findings from this study suggest that clonality is greater at edge range populations, at least in Finland and Denmark and to a lesser degree the UK, than central populations in mainland Europe.

Radoglou *et al.* (2009) cited authors stating Polish *T. cordata* populations to be highly clonal. However, the current study found no clonal individuals in the adult trees sampled from three plots within the Białowieża Forest Strict Reserve. Indeed when samples were collected there were numerous young trees present, many of uniform size in roughly three cohorts *i.e.* seedlings, saplings <0.5m and ~1.5m tall (*pers. obs*.). Following Pigott (1981a) the small saplings would be approximately 3 – 4 years old, while the taller trees would be approximately 10 – 13 years old. Many of these young trees were aggregated in tight clusters that had possibly become established due to a recent gap (~10 years ago) in the canopy (Bobiec, 2007). Among these young trees and at various patches throughout the forest *Tilia* seedlings were present. When the seedlings and saplings were analysed along with adult trees, there was very little clonality observed (R = 0.964, Table 4.3) suggesting that clonal recruitment is limited. Poland is not considered to be at the edge of the *T. cordata* range (Pigott, 2012) and so putative 'range-edge' effects are not expected to have an influential role in the reproduction strategy of *Tilia* stands. Furthermore, the Białowieża Forest Strict Reserve has had less human management, compared to other European forests, for many centuries (Miścicki,

2012). The presence of numerous seedlings which had originated largely from sexual reproduction suggests that *Tilia* in Poland reproduces readily from seed, confirming observations from Pigott (1975) in the Białowieża National Park Strict Reserve and that of Jaworski *et al.* (2005) in other protected forest areas of southern Poland.

Novák *et al.* (2014), reported that while seedlings do occur at the Kuzedeevo site in southern Siberia in some years, recruitment from seedlings is limited due to competition, grazing and fungal attack. The Siberian lime appears to be predominantly clonal, reproducing via vegetative growth. In the measures of clonal diversity (*i.e. R* and *D**), very low values were observed (Table 4.3 and Appendix 4.3), indicating high clonality, confirming that clonal propagation is indeed an important strategy in this species.

Although based on few samples, clonal propagation appears to be greater in *T. x europaea* than the two UK species (Table 4.3). The findings of low clonality in parent species and high in the hybrid have been documented in *Populus* (Santos-del-Blanco *et al.*, 2013). Their study showed higher clonal structure in the hybrid *Populus x canescens* than that of *P. alba*, one of its parent species. Further analysis of *T. x europaea* should be conducted to investigate if asexual reproduction is an important reproductive strategy to the taxon. With limited (if any) sexual reproduction occurring in *T. x europaea,* clonal propagation could be an important mechanism for its continued existence and spread. With potentially greater incidence of clonal occurrence, the hybrid may out-compete *T. cordata* or *T. platyphyllos*. Particularly in small fragmented populations where the two species may become vulnerable to localised extinctions.

4.5.3 Genetic effects of clonality

Genetic diversity within *T. cordata* and *T. platyphyllos* was similar to that found in other partially clonal trees (Stoeckel *et al.*, 2006; Vaughan *et al.*, 2007; Santos-del-Blanco *et al.*, 2013). Conversely, diversity levels in *T. sibirica* were very low (*i.e.* <0.30, Table 4.4). Heterozygosity within the three species, was not significantly different before and after clones were removed (data not shown). This suggests that while clonal reproduction is occurring or has occurred in the past, it has little negative effect on diversity of each species and population. This may be expected as clonal reproduction can maintain or indeed increase genetic diversity (Balloux *et al.*, 2003; Silvertown, 2008). Non-significant differences, before and after clones were removed, was also reported in *Prunus avium* (Stoeckel *et al*., 2006).

Diversity within Polish and Russian *T. cordata* is comparable to UK sites (Table 4.4), and similar to other European stands (Phuekvilai, 2014). No significant difference between genetic diversity of the young and adult trees of the three Polish stands was observed (Logan et al., *unpub. data*) suggesting that stochastic genetic drift is not currently having an adverse effect and that heterozygosity is being maintained in a viable, sexually reproducing population.

Although Balloux *et al.* (2003) predicted that asexual reproduction can increase heterozygosity, it is clear that this is not apparent in the endemic Siberian species *T. sibirica*. Significantly lower genetic diversity was observed in the Siberian lime than the other *Tilia* species although ascertainment bias cannot be ruled out (Tables 4.4 and Appendix 4.5). Indeed, according to Ellstrand and Elam (1993) rare species are expected to have lower heterozygosity due to stochastic genetic drift. While the Siberian lime has been treated as a subspecies of *T. cordata* (Pigott, 2012) it is clear that the two are distinct biological units based on allele frequencies and population genetic analyses (see Chapter 3). For this reason, the species should be the target of future conservation efforts.

The level of clonality reported in this study, is most likely an under-estimation as many of the MLG differed at one allele. While scoring errors may exist, somatic mutations are likely, suggesting ancestral multi-locus lineages and persistent clonality as these mutations can easily become fixed in populations when clonal propagation occurs. Evidence of somatic mutations has been found in *Prunus avium* (Vaughan *et al.*, 2007). Somatic mutations may cause an excess of heterozygotes when asexual reproduction occurs more frequently than sexual reproduction due to fixed mutations. Some observed heterozygosity values in this study were greater than expected, *i.e.* heterozygote excess (Appendix 4.5). While this is generally uncommon, it does occur in small outcrossing clonal populations that experience over-dominance or heterozygote advantage. A heterozygote excess could suggest that little if any inbreeding has occurred within *Tilia* populations as inbreeding would show a reduction in heterozygotes. With the exception of the two *T. sibirica* sites (K28 and K29), no significant F_{IS} values were observed throughout this study (Appendix 4.5).

4.6 Conclusion

This is the first study to assess the level of clonality within members of the genus *Tilia.* Three species were targeted, *T. platyphyllos, T. cordata* and *T. sibirica.* While it

is known that species within the genus regularly propagate asexually, the extent to which this occurs throughout the region of the three species was unclear.

The study has revealed that the markers used were sufficient to identify clones and that fewer than eleven markers may be required in future studies into clonal structure of *Tilia*. The probability of identical genotypes arising by chance was very low suggesting that clones derived from asexual reproduction.

Tilia sibirica populations were shown to be more clonal than *T. cordata* and *T. platyphyllos*, indicating a greater emphasis on asexual reproduction of this ancient relict. As expected, clonal reproduction was generally greater at edge-range sites of *T. cordata* and *T. platyphyllos*, than central European locations. However, the extent of clonal occurrence in *Tilia* was not as great as other partially clonal forest trees.

Tilia x europaea appears to have greater incidence of clonal growth than either parent species. If the hybrid is able to sexually reproduce (and it is not yet clear whether this is indeed the case or not), then a greater ability to spread clonally could, in theory, threaten the already highly fragmented and restricted occurrence of *T. cordata* or *T. platyphyllos*, where sympatry occurs.

The occurrence of clonal growth in *Tilia* does not appear to be having any negative effect on genetic diversity of UK and European populations. However, very low levels of genetic diversity were observed in the Siberian lime (*T. sibirica*). This was equally low before and after clones were removed suggesting that while clonality is not affecting diversity, the species may have experienced low genetic diversity for a long period. It is recommended that efforts are put in place for the preservation of this fragmented species.

Chapter 5: *De novo* Transcriptome assembly, annotation, and SSR marker identification using RNA-Seq data from *Tilia platyphyllos* and *T. cordata* (Malvaceae).

5.1 Abstract

Standard population genetic techniques have been useful in addressing important ecological and evolutionary questions regarding *Tilia*. However, the 'genetic toolkit' currently available for the genus is limited in comparison to other important tree species. Next generation sequencing (NGS) has become a common tool for exploration of genomic and transcriptomic data of non-model organisms. Like many non-model organisms, *Tilia* lacks a reference genome. NGS of the *Tilia* leaf transcriptome was carried out using direct RNA sequencing (RNA-Seq). More than 335,000,000 reads were generated from twelve *Tilia* individuals by the Illumina HiSeq 2500 platform. Two programs, (CLC Genomics Workbench and Trinity), separately assembled the transcriptomes of *T. platyphyllos* and *T. cordata* and produced a similar number of transcripts. BLAST identified transcripts from both assemblies that matched with homologous sequences from *Theobroma cacao* (cocoa), with considerably different results. A similarity search against the Swiss-Prot database with the Trinity assemblies produced several thousand hits. BLAST2GO annotated the two species with Gene Ontology (GO) terms and Enzyme Commission (EC) numbers, identifying enzyme activity. Potentially thousands of simple sequence repeats (SSRs) have been identified in each species that could be useful for further population genetic studies of the genus. This study is a first step towards a robust understanding of *Tilia* genetics, providing a platform for further genomic and transcriptomic research.

5.2 Introduction

Next generation sequencing (NGS) has become a common tool for large scale exploration of genomic and transcriptomic data. Studies of species' conservation, ecology, population biology and evolutionary genetics have all greatly benefited from new and improved, not to mention, easily accessible and affordable technologies. Though the technology has advanced quickly, in devising NGS strategies contemporary researchers can rely on a wealth of detailed reviews (Martin and Wang, 2011; Neale and Kremer, 2011; De Wit *et al.*, 2012; Ekblom and Wolf, 2014)

and novel research articles (Zalapa *et al.*, 2012; Han *et al.*, 2015; Wachowiak *et al.*, 2015). It is clear that NGS technologies have a wide utility in plant science, having been employed in marker development, hybridization and introgression, transcriptome analyses, phylogenetic and ecological studies, and polyploidy (Egan *et al.*, 2012).

As expected, a large number of plant studies using NGS techniques have focused on crops and model organisms, such as *Arabidopsis*, cocoa, cotton and wheat (Varshney *et al.*, 2009; Argout *et al.*, 2011; Austin *et al.*, 2011; Berkman *et al.*, 2012; Wang *et al.*, 2012; Motamayor *et al.*, 2013; Xu *et al.*, 2013). However, the apparent ease and relative low cost of designing and implementing a NGS project has now made it possible to generate large amounts of genetic data from non-model organisms (Ekblom and Galindo, 2011; Strickler *et al.*, 2012; Ward *et al.*, 2012). This is particularly applicable and beneficial to those taxa with important ecological functions where only limited genetic information is available.

The study of non-model organisms offers depth to population and evolutionary biology, not only by expanding our current knowledge of less studied species but by confirming, and in some cases improving the utility of methods to a wider range of taxa. The opportunity to generate high-throughput genomic and transcriptomic data, through various platforms *e.g.* Illumina (Illumina Inc.), SOLiD (Life Technologies), and 454 sequencing (Roche), and the ability to carry out in-house analyses of these data, through commercial or open source software *e.g.* CLC Genomics (CLC-Bio, Denmark), Trinity (Grabherr *et al.*, 2011), Oases (Schulz *et al.*, 2012), Velvet (Zerbino and Birney, 2008), Rnnotator (Martin *et al.*, 2010), ABySS (Simpson *et al.*, 2009), transABySS (Robertson *et al.*, 2010), SOAPdenovo (Luo *et al.*, 2012) and SOAPdenovo-Trans (Xie *et al.*, 2014) has encouraged application of bioinformatics to many ‘non-bioinformatics’ labs. In particular, applications of NGS techniques to economically and ecologically important tree species have gained much interest, for example *Pinus* (Parchman *et al.*, 2010), *Populus* (Rai *et al.*, 2013), and *Quercus,* (Torre *et al.*, 2014). These studies include transcriptome assembly, characterisation and gene identification and marker development using RNA sequencing (RNA-Seq) data.

RNA-Seq is NGS of the transcriptome. It has some clear advantages over other transcriptomics methods such as Tiling microarrays and cDNA or EST sequencing. Only small quantities of RNA are required for good quality coverage. Moreover, RNA-Seq is a relatively inexpensive, high-throughput sequencing technique with

potentially very high expression levels (Wang *et al.*, 2009), meaning good coverage of expressed genes at the time of sampling. This allows a direct comparison and quantification of genes that exist in different organic tissue, or that were sampled at different times. However, RNA-Seq reads obtained from commonly used NGS platforms often produce short reads in the region of 35 – 500bp (Metzker, 2010). This means an assembly of the transcriptome is required and with non-model organisms, unlikely to have a reference genome, this could be problematic given the computation power and knowledge required (Martin and Wang, 2011). Once the assemblies of more than one species are available the opportunity for a comparative analyses is possible (Dassanayake et al., 2009; Wachowiak *et al.*, 2015).

De Novo assembly does not require a reference genome/transcriptome. The method relies on overlapping short reads and then assembling these into transcripts (Martin and Wang, 2011). A De Bruijn graph is used in many of the currently available assemblers to reconstruct transcripts or to assemble isoforms (Wang *et al.*, 2010; Martin and Wang, 2011). Although assembly without a reference genome/transcriptome can be difficult, it does have some advantages. A preliminary set of transcripts, following a *de novo* assembly of the transcriptome, can be obtained and used for expression studies. Moreover, it does not require alignment to splice-sites and is not hindered by long introns. Spliced transcripts and chromosomal transcripts can be assembled using a *de novo* method (Martin and Wang, 2011).

Simple Sequence Repeats (SSRs or microsatellites), are nucleotide motifs 1 – 6bps in length and at least five tandem repeats long. While SSRs have been the ‘marker of choice’ in population genetic studies because of their codominant, highly polymorphic and multi-allelic nature, their development is laborious, expensive, and time consuming. By using NGS techniques, the development of SSRs has become both cost-effective and relatively quick (Ekblom and Galindo, 2011; Zalapa et al., 2012). Microsatellite development from transcriptome analysis, offers huge advantages over whole genome sequencing, as SSRs in the coding region may be associated with functional genes and gene expression (Li et al., 2002). This makes them likely to be better suited for cross-species amplification and so will be an important genetic resource for future studies.

Tilia is an ecologically important, temperate forest tree genus, consisting of ~23 species and 17 subspecies, most of which are diploid (2n=2x=82). Commonly

used molecular markers have been applied to species within the genus (see Chapters 2 – 4) and have successfully estimated genetic diversity, population differentiation and structure, the extent of clonality, effective population sizes and a divergence time.

To date, no NGS studies have been carried out on species within the genus, so no reference genome is currently available. The closest available genomes from the Malvales are cotton, *Gossypium* sp. (Argout et al., 2011; Wang et al., 2012), and cocoa, *Theobroma* sp. (Montamayor et al, 2013). The genome size of *Tilia* is C1.06pg (1,032Mb R. Buggs *pers. comm.*), and is smaller in size than cotton, C1.74pg (1706Mb), but generally larger than cocoa, C0.46pg (452Mb) (Bennett and Leitch, 2012). Compared with other forest trees, *Tilia* genome size is larger than other angiosperm trees (484Mb – 861Mb), but much smaller than the 10 – 40Gb genome size of many conifers (Ahuja and Neale, 2005; Bennett and Leitch, 2012).

The two European *Tilia* species (*T. platyphyllos* and *T. cordata*) have morphological and ecological differences (Pigott, 2012) and are genetically distinct (see Chapter 2), yet they naturally hybridise. They have a wide distribution and although they can occupy similar environments, *T. cordata* has a higher tolerance for colder climates and is found at higher latitudes (Pigott, 2012). This makes it a good system for NGS study and further downstream comparative analyses, where genes specifically related to temperature, light or environmental stresses can be compared. RNA has been chosen for this current study over DNA, because only exon regions will be targeted, thus greater coverage of expressed genes will be generated. The Illumina platform was chosen for sequencing as it can generate 100bp paired-end reads, which produce better assemblies than single-end reads because they will overlap (De Wit *et al.*, 2012). Therefore, the aim of this study is to report an efficient method for a *de novo* assembly of the *Tilia* leaf transcriptome using RNA-Seq data, and to identify potential new microsatellite markers. The markers developed from this study are from coding regions and so could contribute greatly to our understanding of *Tilia* genetics and evolution.

5.3 Materials and Methods

5.3.1 Plant material and RNA isolation

Mature leaves were collected from four *T. platyphyllos* and seven *Tilia cordata* individuals from populations across England and Wales (Table 5.1). Leaves, with no damage, were collected from low hanging branches. Approximately 100mg of leaf

tissue was fully submerged into 800µl of RNA*later*® (Ambion), an RNA stabilisation reagent. Samples were left for up to 60 minutes, to allow for the RNAlater to infuse in leaf cells. The samples were kept on dry ice for 1-2 days before being transferred into a -80°C freezer where they were stored until required for extraction.

Table 5.1 Species, sites sampled for RNA extractions, codes, latitude and longitude coordinates.

Species	**Sites**[1]	**Code**	**Latitude (°N)**	**Longitude (°E)**
T. platyphyllos	Lady Park Wood	T1.1	51.8245	-2.6568
	St. Pierre Great Wood	T3.1	51.6348	-2.7195
	Anston Stone Wood	T11.1	53.3402	-1.1985
	Barton Hills	T13.1	51.9562	-0.4220
T. cordata	Lady Park Wood	T2.1	51.8245	-2.6568
	Weston Big Wood	T4.2	51.4709	-2.7884
	Chedder Wood	T5.1	51.2906	-2.7994
	Collin Park Wood	T6.1	51.9463	-2.3695
	Ivy Wood	T7.1	53.2466	-0.2872
	Roudsea Wood	T10.1	54.2332	-3.0255
	Groton Wood	T12.1	52.0520	0.8830

[1] All samples collected 7th – 14th of July 2013 by Samuel Logan.

Total RNA was isolated from leaf samples of the two species. Species delineation was determined following Logan *et al.* (2015). Extraction was carried out using QIAGEN RNeasy Plant Mini Kit (QIAGEN) following the manufacturer's guidelines with only slight modifications. Approximately 50mg of leaf tissue was fully dried to remove excess RNA*later*®. The sample was then ground in liquid nitrogen using an RNase-free, liquid nitrogen (LN_2) cooled pestle and mortar before being transferred into a LN_2 cooled, Eppendorf tube. Buffer RLC was used instead of buffer RLT (RNeasy Plant Mini Kit). Although both produced equal quality RNA products, buffer RLC produced a higher yield. Samples were incubated at 65°C rather than 56°C as this improved the quality and quantity of the RNA. DNase treatment was carried out on all samples following the guidelines from the RNeasy Plant Mini Kit. On completion of the remaining steps in the protocol, 35µl of RNase-free water was added to the spin column to elute the RNA. This step was repeated using the elution

from the sample. By doing this rather than simply adding a further 35µl of water, the RNA yield was 18 - 22% less but final concentration was higher. The quality and quantity of each RNA sample was assessed on a nanodrop spectrometer. Confirmation of quality was determined by the RNA Integrity Number (RIN) using the Agilent RNA 6000 Nano Kit in an Agilent Bio-analyser (Agilent Technologies). RNA products with a RIN >7 were considered useful for downstream analyses. RNA samples were then stored in a -80°C freezer until required for sequencing.

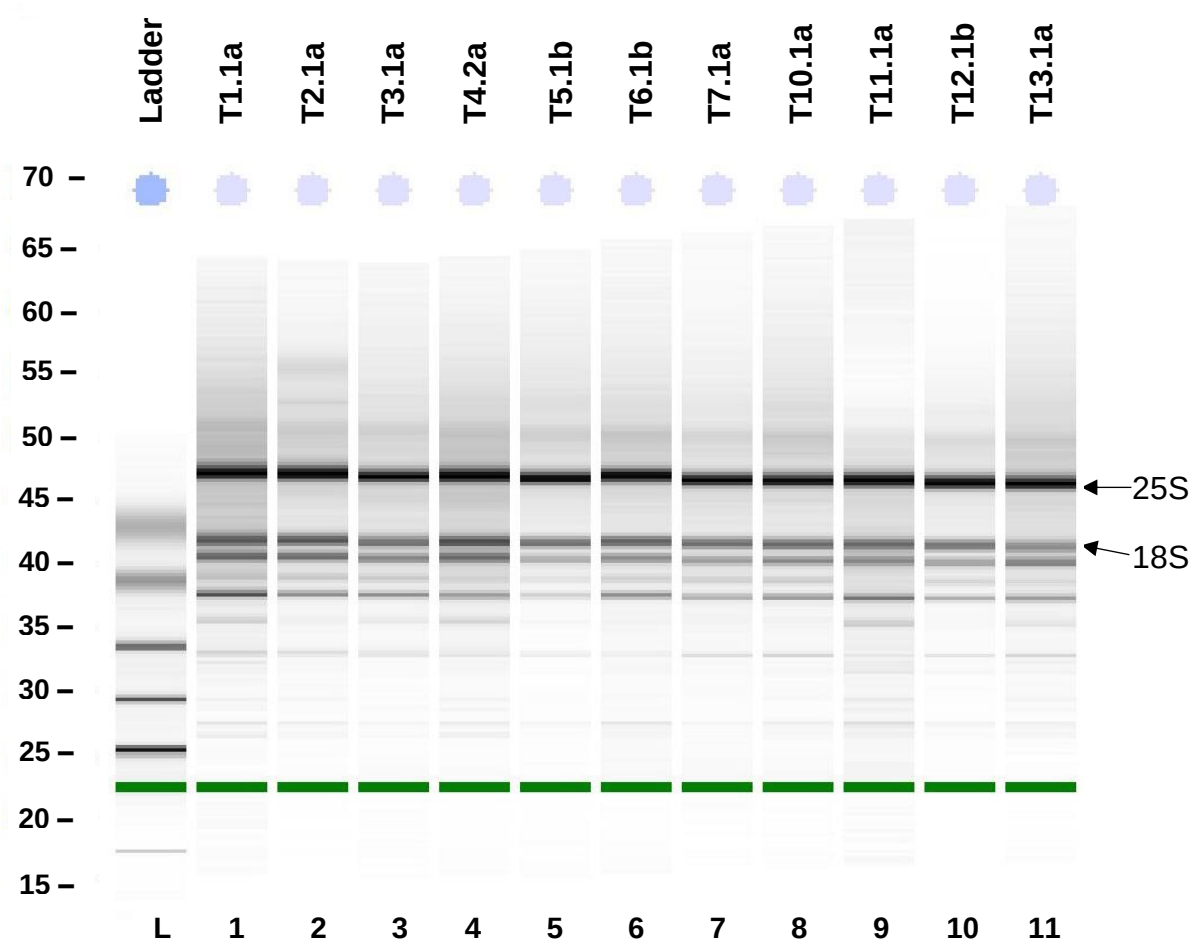


Figure 5.1 Gel image from the Agilent Bio-analyser of the 11 *Tilia* RNA samples. Top two bands are the 25S and 18S RNA genes.

5.3.2 Next Generation Sequencing and Transcriptome assembly

RNA samples were transferred into labelled RNAstable® columns (Biomatrica), placed into zip locked, heat sealed moisture barrier bags and sent for sequencing. Next generation sequencing was carried out using the Illumina Hi-Seq platform by the High-Throughput Sequencing and Genotyping Unit (HTSGU) at the University of Illinois, Chicago.

5.3.3 Library construction and Illumina HiSeq sequencing

RNA-Seq libraries were constructed using the TruSeq Stranded RNA Sample Preparation Kit (Illumina San Diego, CA). Total RNA was quantified by Qubit (Life Technologies, Grand Island, NY) and checked for integrity on a 1% eGel (Life Technologies). Constructed libraries were quantified on Qubit and the average size determined on an Agilent bioanalyzer DNA7500 DNA chip (Agilent Technologies, Wilmington, DE), diluted to 10nM final concentration and quantified by qPCR on a BioRad CFX Connect Real-Time System (Bio-Rad Laboratories, Inc. CA).

The eleven libraries were loaded onto one lane of an 8-lane flowcell and sequenced on an Illumina HiSeq2500 to generate reads 100 bases long. Fastq files were generated using Casava 1.8.4 (Illumina, CA). A detailed account of the library construction process was provided by the HTSGU and is included as an appendix (Appendix 5.1).

5.3.4 Transcriptome assembly and functional annotation

The *de novo* assembly of the leaf transcriptome from two *Tilia* species was carried out on two separate platforms, namely CLC Genomics Workbench v8.5.1 (CLC-Bio, Denmark) and Trinity v2.06 (Grabherr *et al.*, 2011).

Raw Illumina paired-end (forward and reverse) reads (fastq.gz) were imported into and assembled in CLC Genomics Workbench. Assembly was constructed on a single DELL PRECISION 7910 HPC running Linux, Ubuntu 14.04. Failed reads were removed during the import process. Quality trimming was carried out to remove low quality reads, ambiguous reads, and adaptors. Trimming parameters were kept at the default values, *i.e.* trim using quality scores = 0.05 (this is an error probability score in CLC similar to the phred quality scoring in Trinity), and maximum number of ambiguities allowed = 2. The trimmed reads from each species were pooled and assembled as an initial set of transcripts. Assembly parameters were as follows: bubble size = 98, word size = 23, minimum contig length = 100bp, scaffolding was permitted, and paired-end distances were auto detected.

In the CLC Genomics Workbench, word size refers to (user defined) *k*-mer size (*i.e.* the size of assembled sequences of length *k*). The word size is used in CLC to construct De Bruijn graphs. Bubble size refers to the (user defined) maximum size permitted to resolve a dispute at a single nucleotide site due to a heterozygous SNP or a sequencing error. When this happens and the distance between sites is smaller

than the word size, two or more assembled reads will be produced and will re-join further along the assembly path (De Bruijn graph), creating a 'bubble'. CLC resolves a graph dispute (bubble) by assembling the sequences that has the highest read coverage. Smaller word sizes will overlap better and will produce more reliable graphs. However, some information can be lost if word size is too small. Larger word size could lead to fewer paths along the graph and so could help construct transcripts. However, more ‘transcripts’ may be created when larger word sizes do not overlap. While a larger bubble size could resolve sequence differences by permitting more space between unresolved sites, large bubble sizes can create incorrect assemblies (CLC Genomics Workbench manual, available at http://www.clcbio.com/files/usermanuals/CLC_Genomics_Workbench_User_Manual.pdf). A balance between word size and bubble size is required and was determined by repeating assemblies from a sample using both forward and reverse reads, with different values (*i.e.* word size, bubble size to a maximum of 100bp, and minimum length) as well as the default parameters.

For comparison of programs, the raw Illumina reads were also assembled in Trinity. Assemblies were constructed in Newcastle University’s High Performance Computing (HPC) cluster by the Bioinformatics Support Unit (BSU). Adapter sequences were trimmed using the Trimmomatic v0.33 (http://www.usadellab.org/cms/?page=trimmomatic). The program was also used to trim poor quality bases from the end of reads using a sliding window algorithm based on the phred quality scores. Before the assembly, raw reads from *Tilia platyphyllos* were pooled as were reads from *T. cordata*. Reads were assembled using the Trinity transcript assembly pipeline (Haas *et al.*, 2013), described in full at: https://github.com/trinityrnaseq/trinityrnaseq/wiki.

Trinity consists of three main programs: Inchworm, Chrysalis and Butterfly (Grabherr *et al.*, 2011). Inchworm performs an initial assembly of the raw Illumina reads into contigs by extracting overlapping *k*-mers and using a ‘greedy’ *k*-mer extension. Once an inchworm contig is found it is saved and the next *k*-mer extension begins and the next contig is found. This is repeated until all *k*-mers have been processed. Chrysalis constructs de Bruijn graphs by taking associated Inchworm contigs and clusters them into groups of sequences that originated from the same ‘gene’. The graphs are recombined with the original reads and processed by Butterfly. Butterfly produces a set of full length transcripts of alternative spliced isoforms by processing individual graphs in parallel for each cluster (Haas *et al.*,

2013). A detailed account of the Trinity assembly procedure provided by the BSU is included in Appendix 5.2.

Once assembled, a BLASTn search algorithm, implemented in CLC Genomics Workbench, was used to search for homologous gene sequences with the CLC and Trinity assembled transcripts of both *Tilia* species against the coding sequences of *Theobroma cacao* (Motamayor *et al.*, 2013). The coding regions were downloaded from the NCBI RefSeq database (Pruitt *et al.*, 2005). BLASTn parameters for the search were as follows: expectation value = 0.0001, word size = 6, filter low complexity = yes, match/mismatch = 2, gap cost = 2, maximum number of hits = 50.

As Trinity was specifically designed for *de novo* mRNA transcript assembly (Grabherr *et al.*, 2011), and the assembled transcripts were available on the BSU's computer cluster, the Trinity assemblies were further used in a BLASTx search for homologous sequences against the Swiss-Prot database (Boeckmann *et al.*, 2003). The computational requirements to BLAST all the transcripts generated by the Trinity assembly step is demanding and hugely time consuming for a single machine. Therefore, the whole transcript from each species was split into groups of 1,000 and submitted as an individual 'job' on the University's HPC cluster. The HPC cluster is able to parallelize (run simultaneously across multiple machines) due to the complete independence of one alignment to another. The command line BLASTx program v2.2.3 was used, which produced an XML formatted BLAST output. The "--outfmt 5" option was used to generate the required alignment files. All other BLAST options were left at default values. The output files were merged using a utility Python script, downloaded from https://bitbucket.org/peterjc/galaxy-central/src/5cefd5d5536e/tools/ncbi_blast_plus/blast.py. The script was then imported into the BLAST2GO plugin (Conesa *et al.*, 2005) on the CLC Genomics Workbench and converted into a B2G (BLAST2GO) project for analyses and functional GO (Gene Ontology) annotation *i.e.* biological process, molecular function and cellular component. BLAST2GO functionally annotates and analyses gene and protein sequences by finding homologs using BLAST alignments (Conesa *et al.*, 2005). GO terms and Enzyme Commission (EC) codes from the BLAST hits were determined through the 'mapping' and the 'annotation' options implemented in the CLC Genomics Workbench. All default settings were kept; e-value hit filter = 1.0e-6, annotation cutoff = 55, GO weight = 5 and HSP hit coverage cutoff = 0.

5.3.5 Simple Sequence Repeat (SSR) detection

Identification of SSRs was performed in the Perl script program MISA (MIcroSAtellite identification tool): available at http://pgrc.ipk-gatersleben.de/misa/ (Thiel *et al.*, 2003). Default parameters were used to search for perfect microsatellites – definition (unit size-min repeats) = 1–10, 2–6, 3–5, 4–5, 5–5, 6–5, and compound microsatellites – interruptions (maximum difference between two SSRs) = 100. SSRs will be available for further downstream population genetic study analyses.

5.4 Results

In total >335,000,000 paired-end reads, each of 100bp long, were generated from the Illumina sequencing platform (Table 5.2). The *de novo* assembly from the CLC Genomics Workbench produced 335,853 and 385,859 contigs (predicted transcripts) for *Tilia platyphyllos* and *T. cordata*, respectively. Length of transcripts ranged from 100 – 20,679bp in *T. platyphyllos* and 100 – 17,133bp in *T. cordata.* The *de novo* assembly carried out in Trinity produced 277,097 predicted transcripts for *T. platyphyllos* and 316,004 predicted transcripts for *Tilia cordata*. Length of the transcripts ranged from 224 – 17,200bp in both species.

5.4.1 Homologous search, Functional annotation and GO classifications

Theobroma cacao (cocoa) coding regions (Motamayor *et al.*, 2013) were used to search for homologous sequences with the CLC and Trinity assembled *Tilia* transcriptomes, using BLAST. Of the 335,853 transcripts (contigs) assembled for *T. platyphyllos* using CLC, 139,796 hits (42%) were retrieved from the cocoa coding sequences. From the 385,856 transcripts assembled for *T. cordata,* 136,696 (35%) were retrieved. Of the 277,097 predicted transcripts assembled for *T. platyphyllos* using Trinity, 41,113 hits (15%) were retrieved from the cocoa coding sequences. From the 316,004 predicted transcripts assembled for *T. cordata,* 43,264 hits (13%) were retrieved (Table 5.3).

Table 5.2 Sample codes, number of reads, read length, number of transcripts from both the CLC and Trinity assemblers.

Sample	No. of reads	Read length	CLC transcripts	Trinity transcripts
T1.1	28,115,420	100		
T3.1	30,087,400	100		
T11.1	29,536,648	100		
T13.1	36,091,062	100		
	123,830,530	–	335,853	277,097[2]
T2.1	31,563,864	100		
T4.2	27,526,610	100		
T5.1	31,124,824	100		
T6.1	27,834,958	100		
T7.1	30,718,994	100		
T10.1	31,870,900	100		
T12.1	31,287,240	100		
	211,927,240	–	385,859	316,004[3]
Total	335,757,920			

[2] *Tilia platyphyllos* samples; [3] *T. cordata* samples.

Table 5.3 Number of BLASTn hits with *Theobroma cacao* using both the CLC and Trinity assemblies.

Species	CLC BLAST hits	Trinity BLAST hits
T. platyphyllos	139,796	41,113
T. cordata	136,696	43,264

Gene Ontology (GO) terms and Enzyme Commission (EC) codes for *Tilia platyphyllos* and *T. cordata* leaf transcripts were obtained from BLAST2GO using the Trinity assemblies. Of the 277,097 and 316,004 predicted transcripts, 40.47% and 22.65%, respectively, were successfully annotated (Table 5.4). GO terms were summarized into three main categories, biological process (BP), molecular function (MF) and cellular component (CC). In *T. platyphyllos,* a total of 729,368 GO terms relate to biological processes, 185,033 to molecular function and 382,069 to cellular components (Fig. 5.2). In *T. cordata,* 547,391 GO terms relate to biological processes, with 118,788 assigned to molecular function and 243,770 to cellular components (Fig. 5.3). The highest proportion of GO terms, retrieved from BLAST2GO for *T. platyphyllos* (290,974) and *T. cordata* (186,518), categorised as BP were for sub-categories cellular process (GO:0009987), single-organism process (GO:0044699), and metabolic processes (GO:0008152); 145,044 and 92,921, respectively, categorised as MF were binding (GO:0005488) and catalytic activity (GO:0003824); and 263,068 and 168,415, respectively, categorised as CC were cell (GO:0005623), organelle (GO:0043226), and membrane (GO:0016020). The main GO categories can also be explained as sub-categories. For both species, biosynthetic process (BP), protein binding (MF), and plasma membrane (CC) were the sub-categories with the largest number of transcripts associated (Figs. 5.4 and 5.5).

Table 5.4 Total number of predicted transcripts that were successfully blasted, mapped and annotated using BLAST2GO from the Trinity assemblies.

Species	Transcripts	Blasted	Mapped	Annotated
T. platyphyllos	277,097	213,520	209,509	112,142
T. cordata	316,006	241,847	128,533	71,571

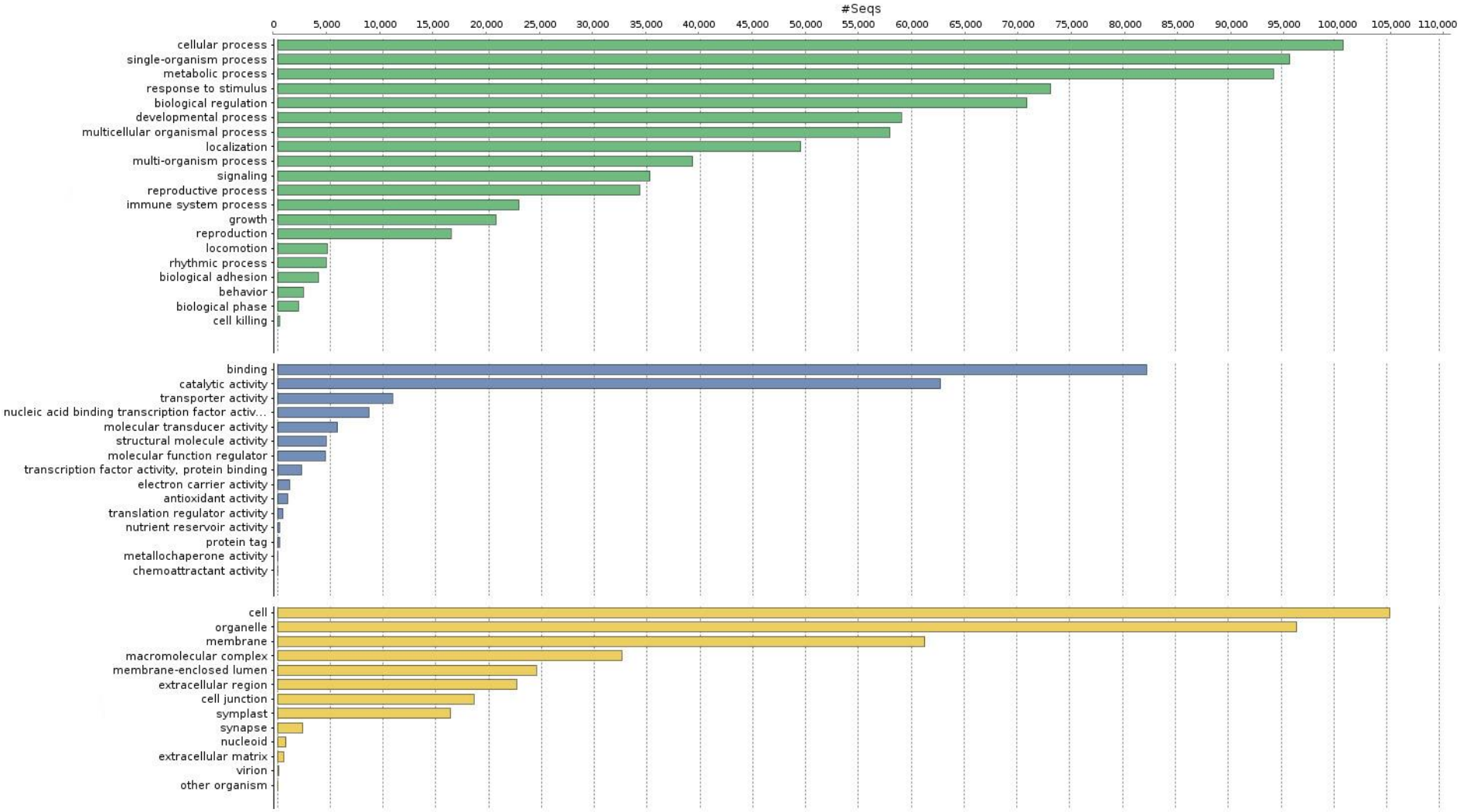


Figure 5.2 Gene Ontology (GO) terms for *Tilia platyphyllos* transcripts assembled in Trinity. Green – Biological process; Blue – Molecular Function; Yellow – Cellular Component.

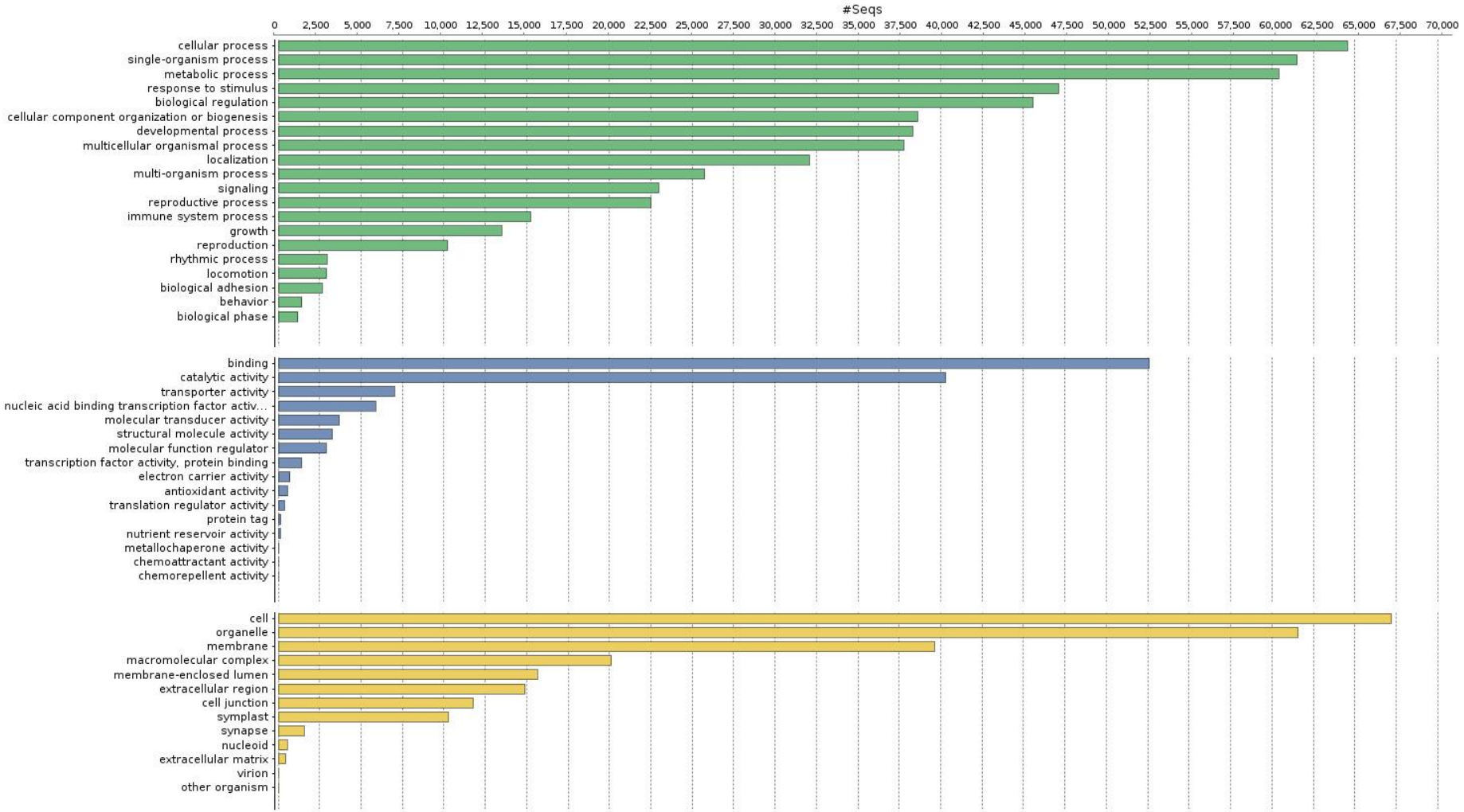


Figure 5.3 Gene Ontology (GO) terms for *Tilia cordata* transcripts assembled in Trinity. Green – Biological process; Blue – Molecular Function; Yellow – Cellular Component.

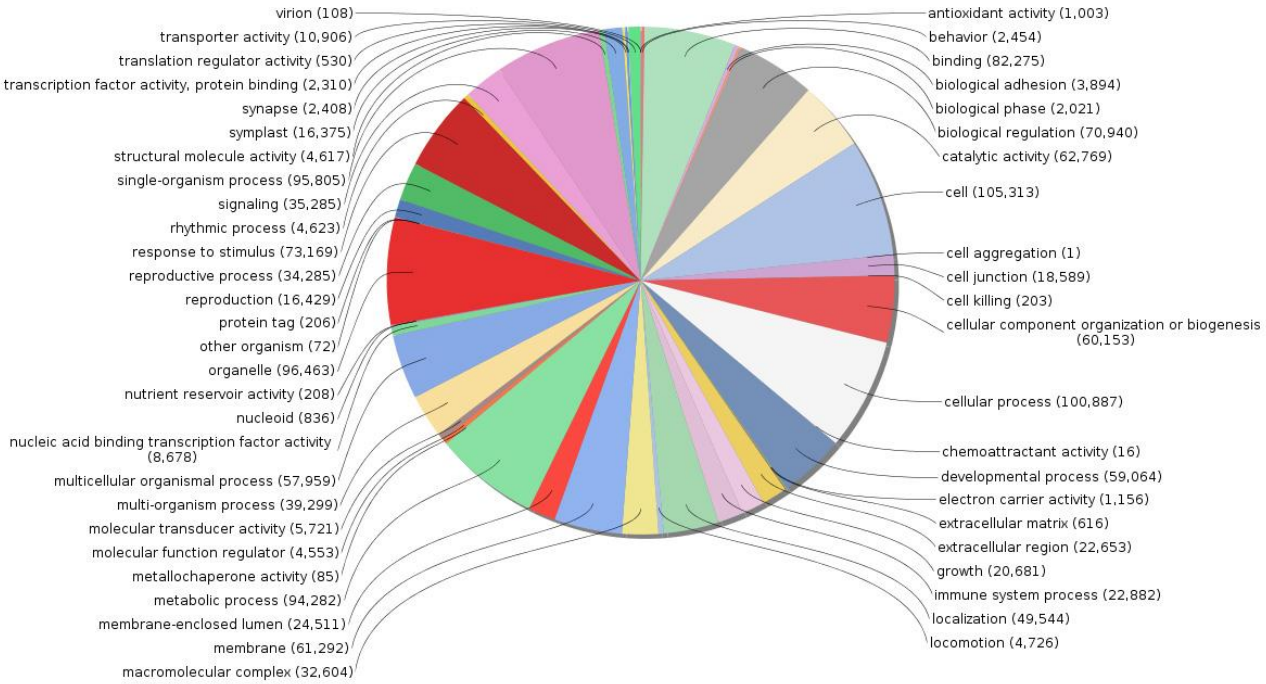


Figure 5.4 Combined distribution of the assembled sequences of *Tilia platyphyllos* in three main GO categories *i.e.* Biological Process, Molecular Function, and Cellular Component.

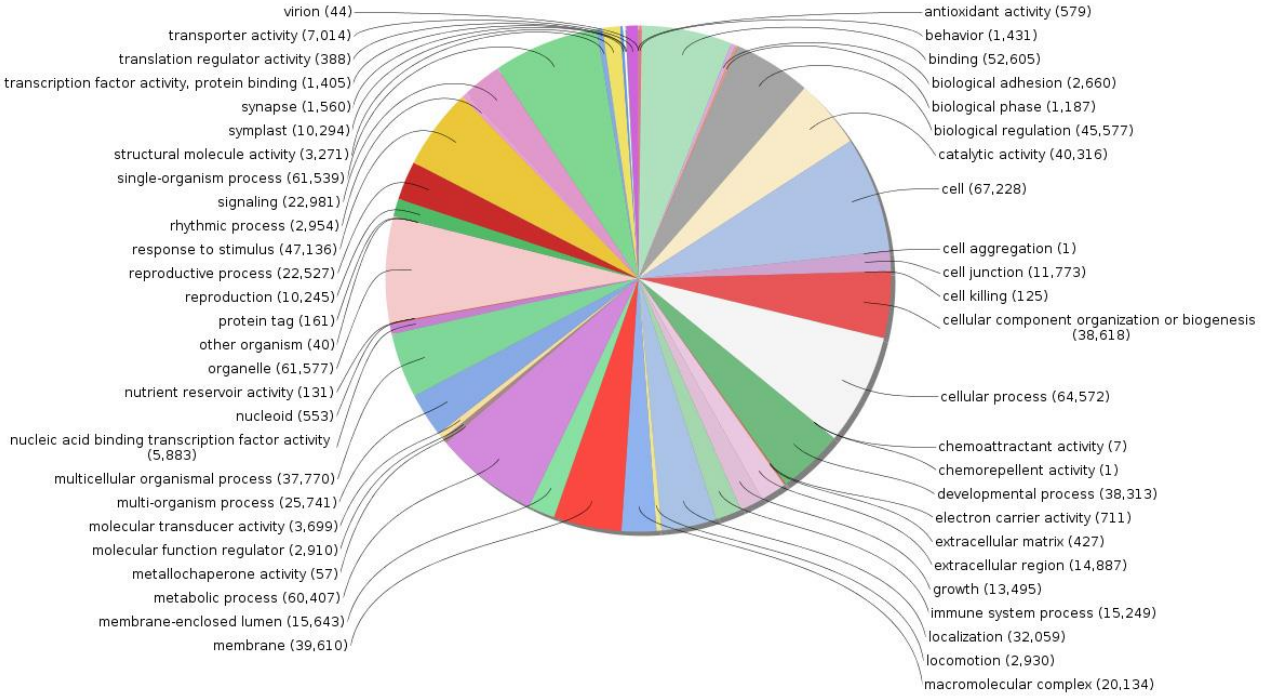


Figure 5.5 Combined distribution of the assembled sequences of *Tilia cordata* in three main GO categories *i.e.* Biological Process, Molecular Function, and Cellular Component.

Enzyme Commission (EC) codes for six categories (Oxidoreducrases, Transferases, Hydrolases, Lyases, Isomerases and Ligases), were associated with 42,515 *T. platyphyllos* and 27,183 *T. cordata* transcripts following GO mapping and annotation (Figs 5.6 and 5.7). For both species, most hits were linked with Transferase activity (18,486 and 11,934, respectively), hydrolase activity (12,840 and 8,067, respectively), and oxidoreductase activity (8,516 and 5,418, respectively).
The six main enzyme classes also consist of sub-classes. In both *Tilia platyphyllos* and *T. cordata*, the greatest number of hits linked to Transferases were related to enzymes responsible for transferring phosphorus-containing groups EC numbers 2.7 (10,415 and 6,788, respectively, Figs 5.8 and 5.9).

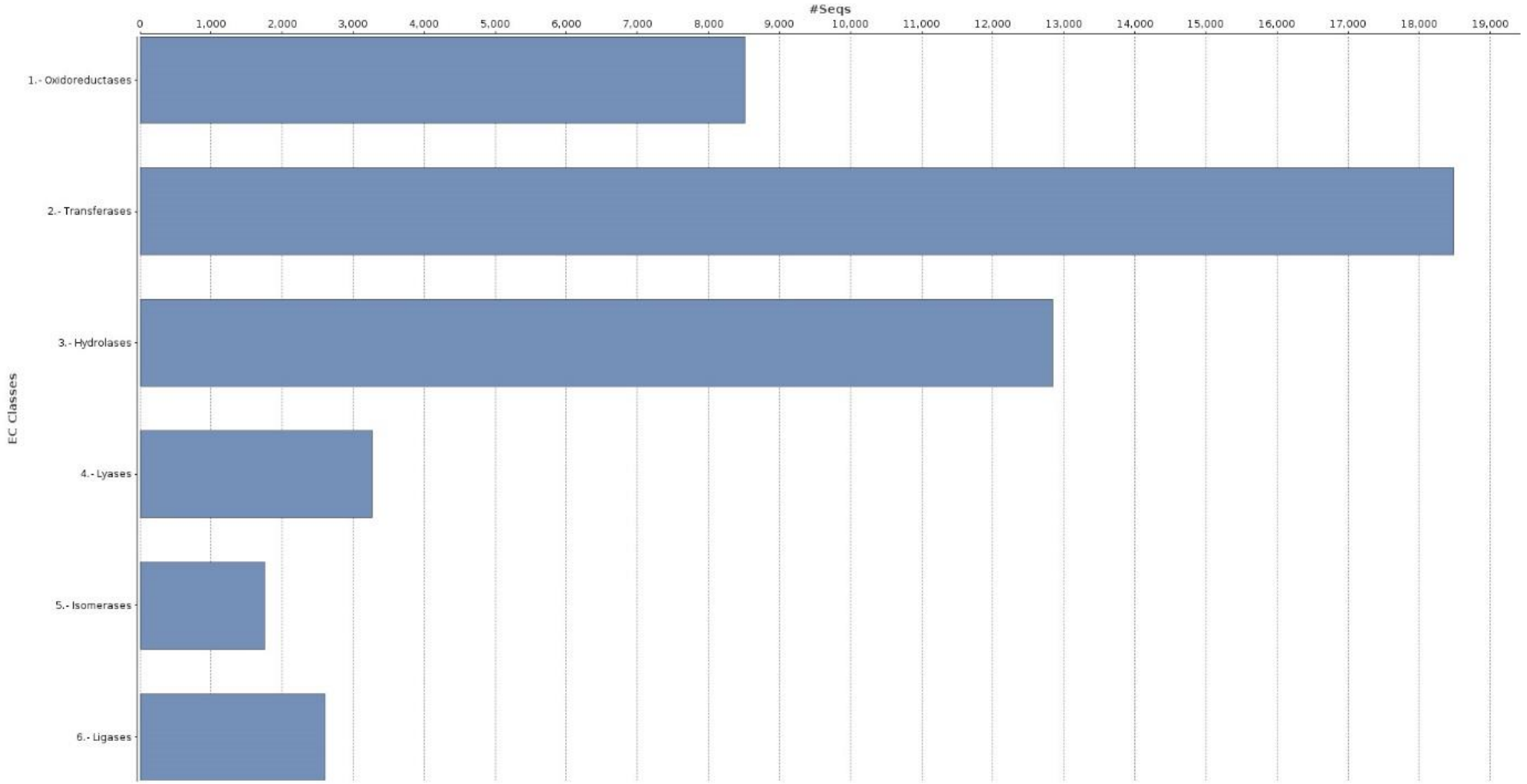


Figure 5.6 Enzyme Commission (EC) classes associated with 42,515 *T. platyphyllos* transcripts assembled using Trinity. Most transcripts (seqs) were associated with Transferases (18,486).

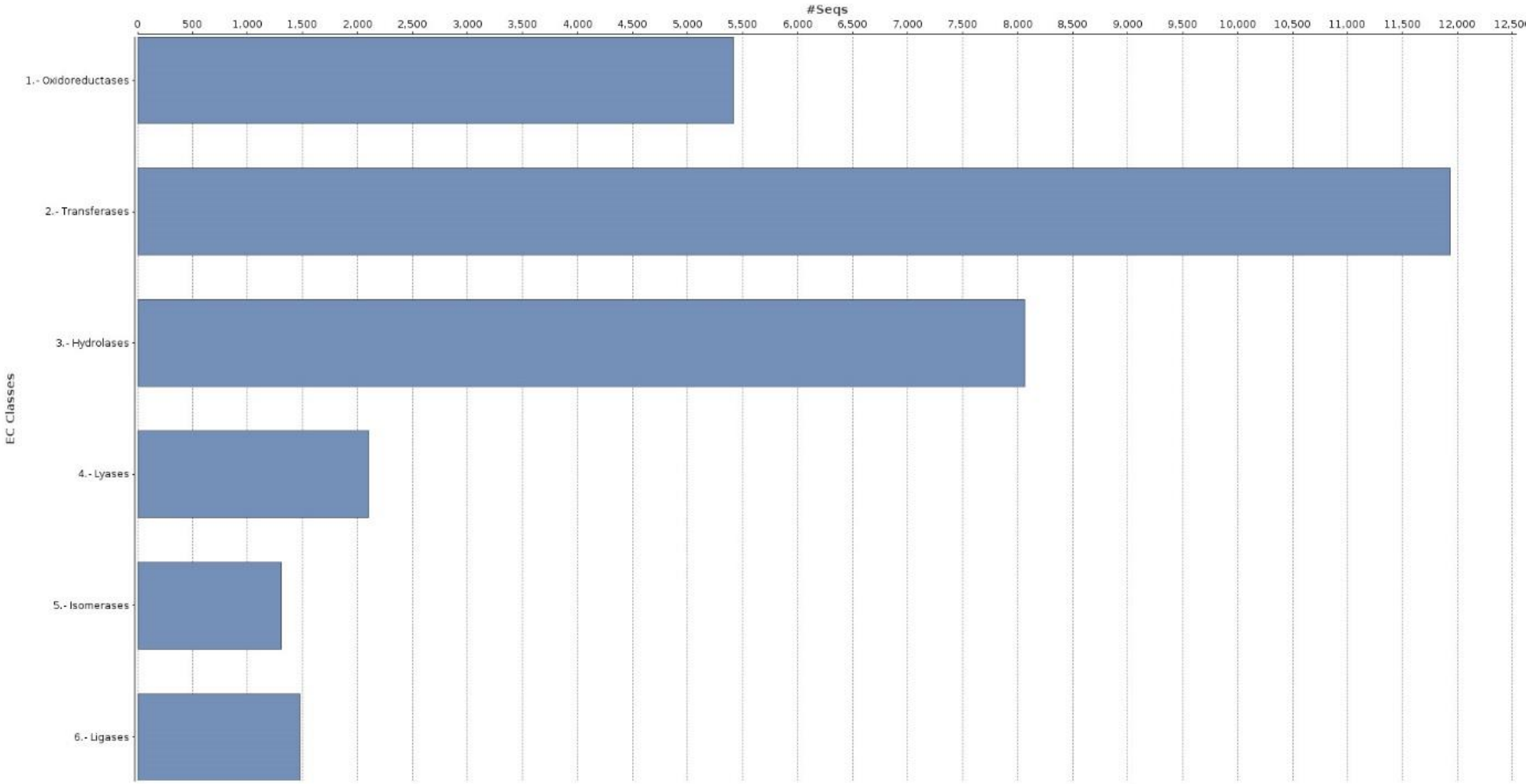


Figure 5.7 Enzyme Commission (EC) classes associated with 27,183 *T. cordata* transcripts assembled using Trinity. Most transcripts (seqs) were associated with Transferases (11,934).

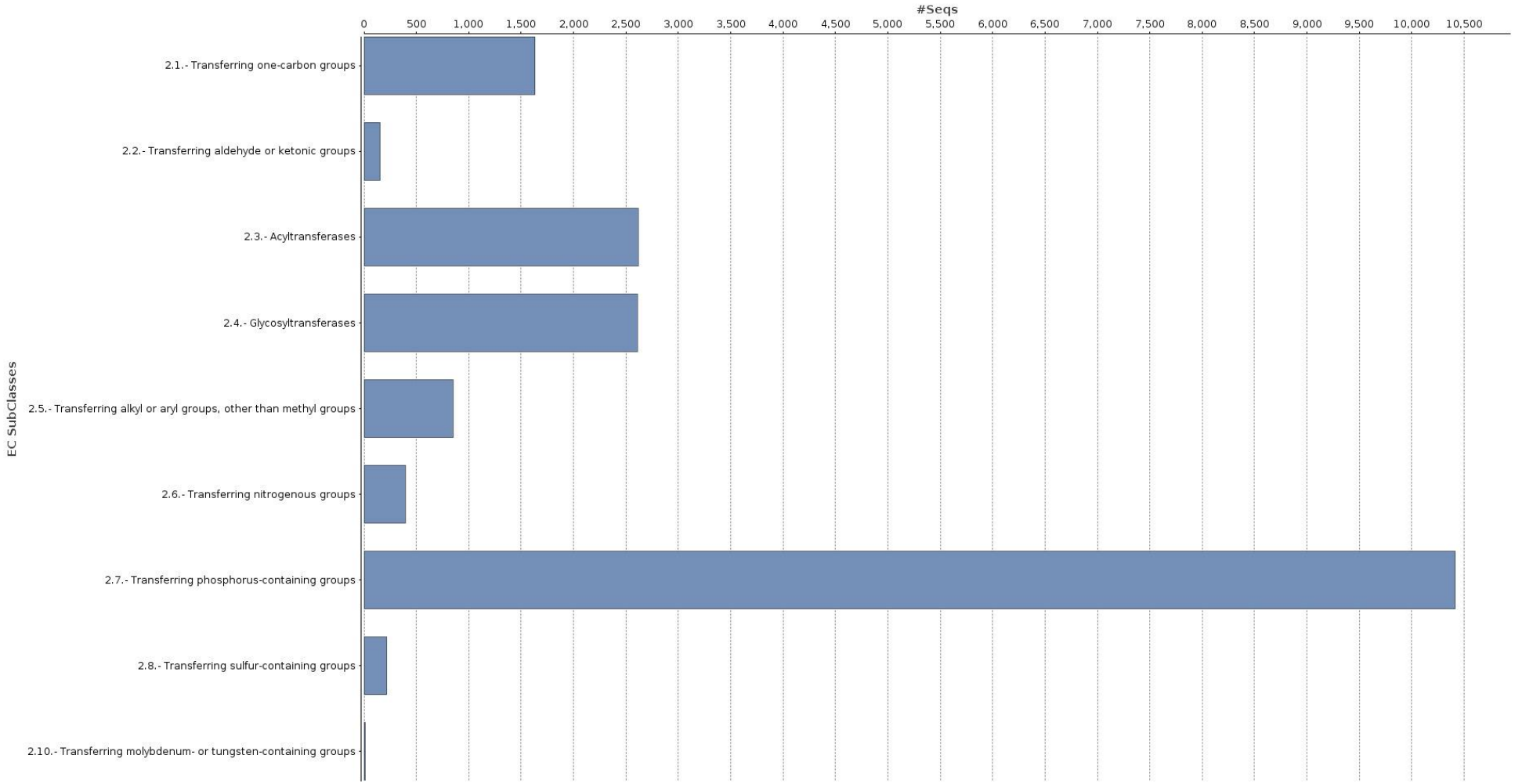


Figure 5.8 Sub-division of Transferase enzyme class from *T. platyphyllos* showing the number of hits (seqs) classified to a particular sub-class. Most transcripts (seqs) were related to Transferring phosphorous-containing groups (10,415).

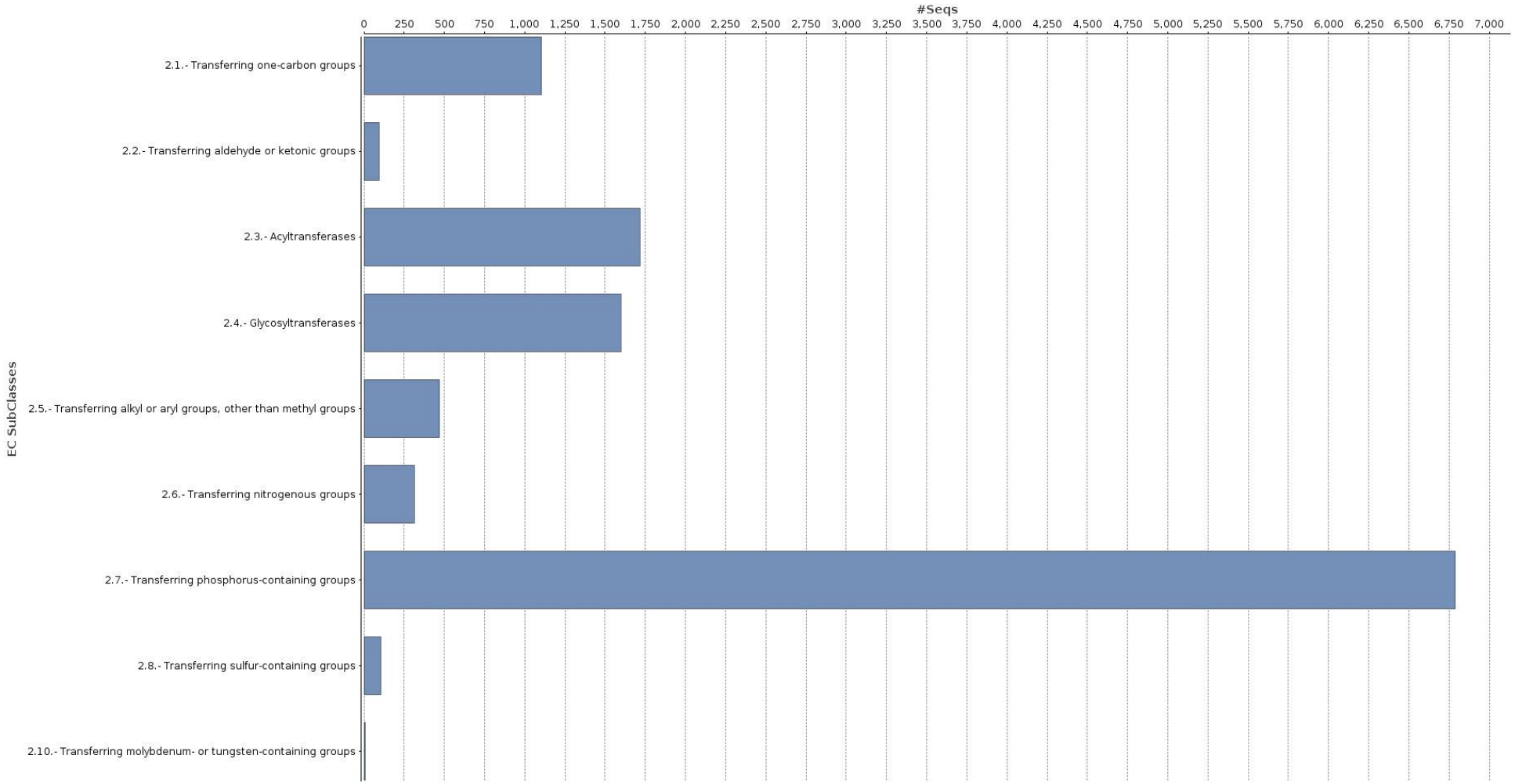


Figure 5.9 Sub-division of Transferase enzyme class from *T. cordata* showing the number of hits (seqs) classified to a particular sub-class. Most transcripts (seqs) were related to Transferring phosphorous-containing groups (6,788).

5.4.2 SSR marker identification

Simple Sequence Repeats (SSRs) were identified in the 277,097 and 316,004 predicted transcripts from *T. platyphyllos* and *T. cordata*, respectively. In total, 80,732 perfect SSRs were identified in *T. platyphyllos* ranging from 1 – 6 bps and repeating more than five times, with an additional 8,051 compound SSRs. In *T. cordata,* 102,195 perfect SSRs ranging from 1 – 6 bps and repeating more than five times were identified, with an additional 10,853 compound SSRs. Mono-nucleotides were the most common in both species with 59,458 (73.7%) and 75,490 (73.8%) in *T. platyphyllos* (blue) and *T. cordata* (orange), respectively (Fig. 5.10). In total, 12,305 (15.2%) and 15,991 (15.6%) di-nucleotides and 7,803 (9.67%) and 9,298 (9.09%) tri-nucleotides were identified in the two species. The most abundant di-nucleotide repeat motifs were TC/CT and AT/TA in both species (Fig. 5.11). The most abundant tri-nucleotide repeat motifs were GAA (573) and TTC (508) in *T. platyphyllos* and TTA (733) and TTC (691) in *T. cordata* (Fig. 5.12). All *T. platyphyllos* SSRs were found within 75.6% transcripts; 24.4% of which contained more than one SSR. In *T. cordata*, SSRs were found within 74.0% of the transcripts; 26.3% of which contained more than one SSR (Table 5.5).

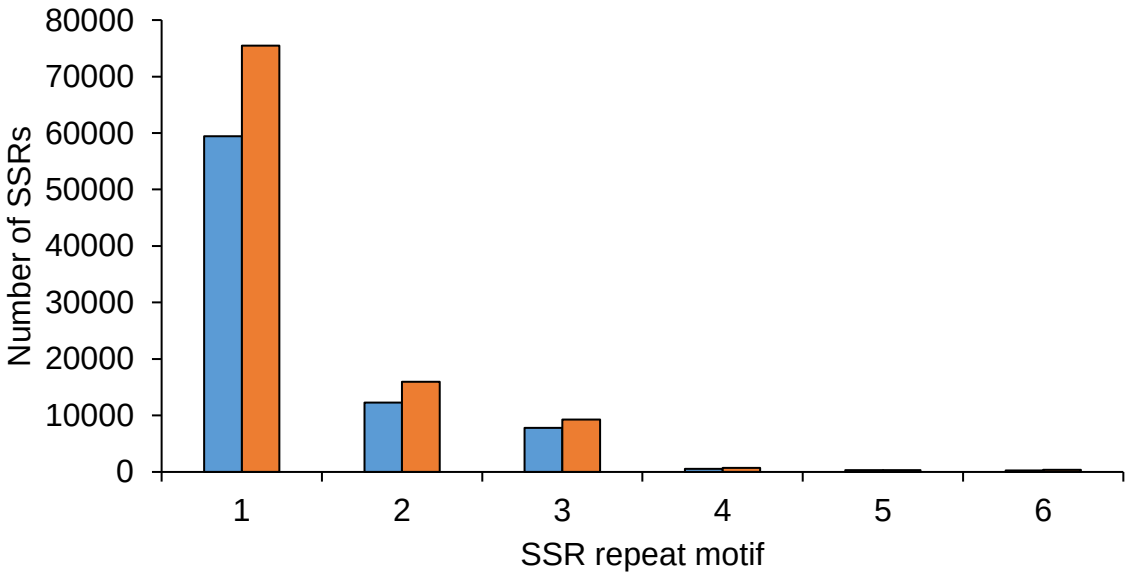


Figure 5.10 Distribution of microsatellite repeat motifs in *T. platyphyllos* (blue) and *T. cordata* (orange).

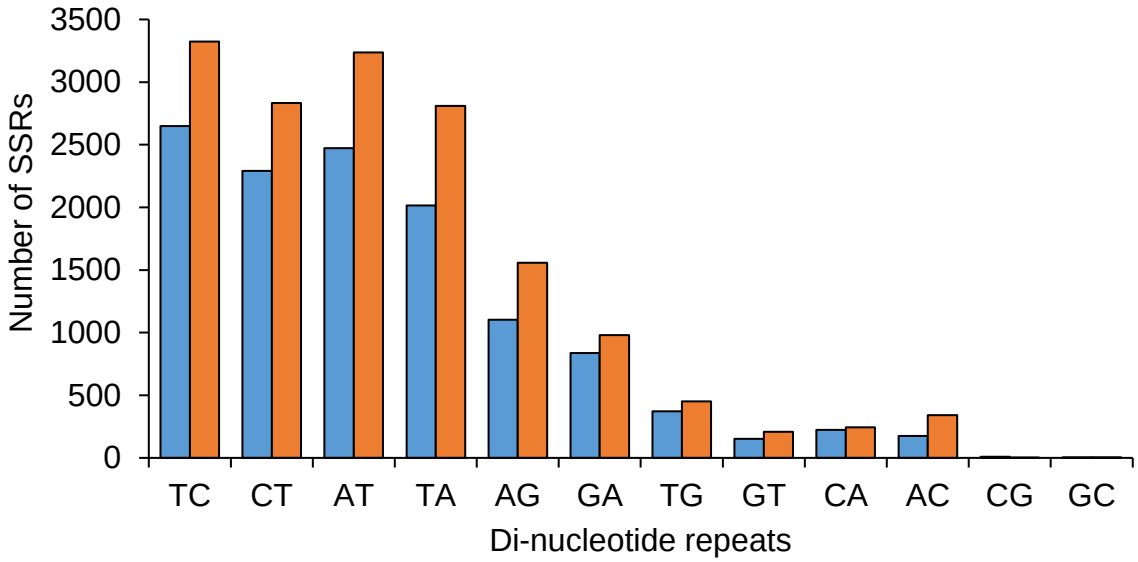


Figure 5.11 Number of di-nucleotide repeat motifs in *T. platyphyllos* (blue) and *T. cordata* (orange).

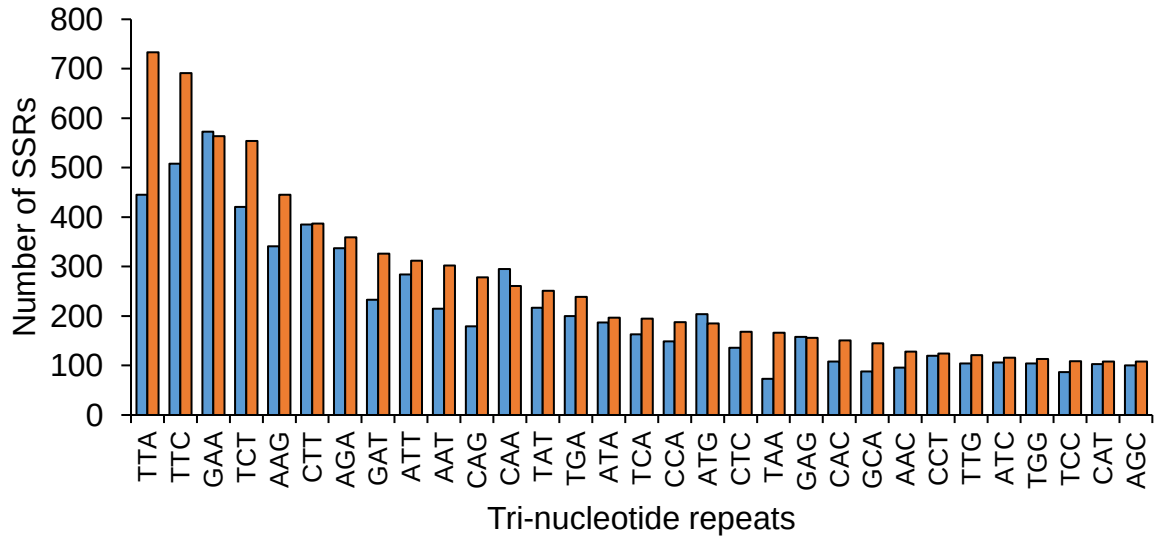


Figure 5.12 Number of tri-nucleotide repeat motifs (most abundant >100 repeats) in *T. platyphyllos* (blue) and *T. cordata* (orange).

Table 5.5 Number of perfect and compound microsatellites found within transcripts of both species.

Species	No. of sequences searched	Perfect SSRs	Compound SSRs	Sequences containing SSRs	Sequences with >1 SSR
T. platyphyllos	277,097	80,732	8,051	61,078	14,919
T. cordata	316,004	102,195	10,853	75,590	19,880

5.5 Discussion

Tilia are considered non-model organisms and as such fewer genetic studies have been carried out on species within the genus than on so-called model organisms. To date, most molecular studies carried out on species of *Tilia*, in general, have included the assessment of genetic diversity, species determination, clonal analyses, cultivar identification, hybridisation, and phylogeny construction *e.g.* Maurer and Tabel (1995), Fineschi *et al.* (2003), Fromm and Hattemer (2003), Liesebach and Sinkó (2008), McCarthy (2012), Yousefzadeh *et al.* (2012), Hosseinzadeh Colagar *et al.* (2013), Hansen *et al.* (2014), Phuekvilai (2014), and this present thesis (see Chapters 2 – 4). These studies have used various chloroplast and nuclear markers, and while they have all revealed useful and informative results, many of the molecular methods and approaches used were time consuming, laborious and expensive and generated only a small amount of data specifically targeted to address particular questions. In contrast, NGS technologies provide a relatively fast and cost effective approach in generating large amounts of molecular data that can be useful for many different downstream analyses. In particular, relevant to the present study, NGS provides the means to characterise the transcriptome of non-modal organisms (Strickler *et al.*, 2012; Ward *et al.*, 2012). This is the first RNA-Seq study to be carried out on species within the genus *Tilia* and has provided a gateway for further high-throughput molecular studies.

5.5.1 A de novo approach for an omics future for Tilia

Neale and Kremer (2011), point out five important areas to consider when using NGS technologies in forest tree genetic research; (1) a greater number of reference genome sequences are required for more tree species; (2) more NGS studies are needed in ecologically, as well as economically, important trees; (3) the need to focus on adaptation and divergence of forest trees at the species and population

levels; (4) the necessity to make newly derived genomic data from forest tree systems available for further advancements, and (5) the need to develop more phenotyping technologies for forest tree systems. This study has directly addressed three of the five priority areas suggested by the authors, and indirectly has provided a platform for further addressing the other two.

Firstly, an efficient and effective method for the de novo assembly and analyses of the *Tilia* leaf transcriptome from Illumina generated paired-end, short reads, using both the CLC Genomics Workbench and Trinity assemblers has been provided for two *Tilia* species. This has added a genetically under-studied forest tree genus to the 'omics arena and has provided an opportunity for further advancements in genomic/transcriptomic studies for the genus. Secondly, the two species - *T. cordata* and *T. platyphyllos* - are ecologically important UK and European forest trees but have little economic value, at least not in the UK or parts of Europe. *Tilia* are associated with several species of fungus, lichen, and mammals, not to mention many species of insect from orders including Diptera, Hymenoptera, Hemiptera, Lepidoptera, Coleoptera, and Acarina (Pigott, 2012). By providing an initial set of reference transcripts for the host tree species, the present study offers a starting point for further research into the ecological interactions between *Tilia* and its symbionts and other associated species. And thirdly, the necessary next step forward in *Tilia* genetics has been taken and a novel RNA-Seq study carried out. The raw Illumina reads and *Tilia* reference transcriptome will be deposited into an online genetic database and made available for further 'omics studies involving the genus. The availability of a reference transcriptome will greatly increase the *Tilia* 'genetic toolkit', by providing opportunities for further studies. Research into adaptation, divergence, or phenotyping of the two ancient woodland relicts, or any *Tilia* species, can begin here. It is only a matter of time until more genomic and transcriptomic data are available, adding to the toolkit and expanding the ecological and evolutionary understanding of the genus.

5.5.2 A tale of two assemblers

Although it is difficult to directly compare RNA-Seq studies, the assemblies constructed in this study have produced a greater number of transcripts than similar studies using the Illumina platform and either CLC or Trinity assemblers e.g. *Quercus pubescens* (Torre et al., 2014), *Pinus halepensis* (Pinosio et al., 2014), *Ginkgo bilboa* (Han et al., 2015), and *Pinus sylvestris* (Wachowiak et al., 2015 – Appendix 5.3).

Reported here are the predicted transcripts assembled from both programs. It is likely that these other studies, particularly the two that used Trinity, did not report all the predicted transcripts (*i.e.* different isoforms or alternative spliced transcripts). Considering the two studies that used CLC, using different parameters and organic tissue are likely to produce different results across studies. Either way, the transcripts assembled here will be of particular benefit to further molecular studies of *Tilia*.

The total number of contigs/transcripts assembled by the two assemblers were similar (Table 5.2). However, the Trinity output file following assembly, better defines each contig by identifying different isoforms of the same transcript while the CLC assembly does not do this. While the number of (predicted) transcripts were similar, the subsequent BLASTn homologous search against *Theobroma cacao* results varied considerably. The CLC assemblies produced more aligned hits with cocoa than the Trinity assemblies (Table 5.3). However, the CLC assemblies contained a greater number of short contigs, in the region of 100 – 200bp long, and so it is likely that a single cocoa sequence matched to the different assembly fragments of the same CLC transcripts. This may be a reason why more 'hits' were observed. For future assemblies using the CLC Genomics Workbench, different assembly parameters could be used which might provide fewer but longer contigs/transcripts.

Although this study has used two assemblers for a general comparison, it has not concluded that one is better over the other, nor was it the intention to do so. While the CLC assemblies produced more similarity hits with *Theobroma cacao* coding sequences, the Trinity assemblies were used for further BLASTx and GO annotations. The reason for this was purely based on computational requirements needed for a full BLASTx search. One potential drawback for a non-bioinformatics lab initiating a large bioinformatics study is likely to be the lack of computational resources, as was evident with this current study. Although a high performance computer was available for the analyses, a single standing machine is still restricted when using such large datasets when there is no direct link to a computer cluster or server (Martin and Wang, 2011). The Trinity assemblies were constructed using a HPC cluster in a bioinformatics unit and as such these transcripts were available for similarity searches with public databases. What took almost one month to run on the single standing HPC, took less than a day to run on the HPC cluster. Even with the extra computer power, the Trinity assemblies were only run against one database, namely Swiss-Prot. Other such databases *e.g.* NCBI Non-Redundant (NR), RefSeq,

and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) could be used in future to provide more information.

5.5.3 Gene Ontology (GO) and Enzyme Commission (EC)

GO terms were summarized into three main categories, biological process (BP), molecular function (MF) and cellular component (CC). Each category consists of several sub-categories and these terms can be extracted from the CLC Genomics Workbench explaining a more detailed ontology (Appendices 5.4 and 5.5). Most transcripts were associated with BP (729,368 GO terms in *T. platyphyllos* and 547,391 in *T.* cordata). Cellular process (GO:0009987), single-organism process (GO:0044699), and metabolic processes (GO:0008152) were the top three sub-categories with the most represented associated transcripts (Figs. 5.2 to 5.4). This may suggest that, at time of sampling, considerable metabolic and cellular activity were taking place in the *Tilia* leaf tissue. Metabolic processes occur more in plants than in other organisms (Aharoni and Galili, 2011), and this is likely the reason for such a large association of transcripts.

Transferases and Hydrolases, were the most represented enzymes associated with the transcripts of both species as suggested by the Enzyme Commission (EC) codes (Figs. 5.6 and 5.7). Transferase enzymes catalyse transfer of chemicals from one compound to another, while hydrolase enzymes catalyse the hydrolysis of compounds. The large number of transcripts related to these two enzyme activities suggest an association to pathways that synthesis secondary metabolites (Li *et al.*, 2010; Blanca *et al.*, 2011). Secondary metabolites that are being produced in *Tilia* are likely to be restricted to the genus (Pichersky and Gang, 2000). It would be interesting to see if these differ among species, and is a possible area for further research.

5.5.4 Simple Sequence Repeat (SSR) identification

This study has identified potentially thousands of simple sequence repeats (SSRs) using Illumina sequences and Trinity assemblies. The advantages of using NGS technologies to identify microsatellites is that a greater number of SSRs will be discovered than using traditional Sanger sequencing (Zalapa *et al.*, 2012). Furthermore, ascertainment bias will not be a concern in further analyses as specific SSRs can be directly derived and identified from both species. While the number of potential SSRs identified in this study is much more than the average (4648) found in

Zalapa et al. (2012) using Illumina sequencing, the microsatellites from the current study are yet to be validated. SSR validation must be carried out in order to confirm useful microsatellites that are available in both species. Of course, not all the potential microsatellites identified here will be of use but it is expected that there will be several tens if not hundreds of potentially useful markers available for future population genetic/genomic studies, thus greatly increasing the *Tilia* genetic toolkit.

5.6 Conclusion

This study reports a simple and efficient protocol for an RNA-Seq study of the *Tilia* leaf transcriptome, describing a suitable method for collecting and storing leaf tissue, RNA extraction, library preparation, transcriptome assembly and analyses. The raw reads and reference transcriptome assemblies of the *Tilia* leaf will be made available in a public depository for use in further downstream analyses, such as phylogenetic studies, gene expression, SNP discovery (nucleotide substitution), and alternative splicing and will greatly contribute to the future of *Tilia* genomic and transcriptomic studies.

Chapter 6: General Discussion

This thesis has addressed basic evolutionary and ecological questions concerning three highly fragmented species of *Tilia.* Prior to the development of *Tilia* specific microsatellite markers (Phuekvilai and Wolff, 2013), there was an obvious lack of knowledge regarding genetic diversity of the genus, in particular in comparison to other important forest trees, such as oak and ash. This study has advanced our understanding of the molecular ecology of *Tilia cordata, T. platyphyllos* and *T. sibirica* and has provided a means for genomics and transcriptomic studies, allowing *Tilia* to become more of a focus in future research.

6.1 The main findings/conclusions of this study

6.1.1 Chapter 2

- The *Tilia* microsatellite markers successfully distinguished between the two UK species and the hybrid
- *Tilia cordata* and *T. platyphyllos* are two distinct species with high genetic differentiation
- Both species are diploid and outcrossing
- UK *Tilia* populations show high genetic diversity
- Intra-specific structure and inter-population differentiation is apparent in both species
- Restoration of woodlands with lime should consider replanting trees from woodlands with similar genotypes

6.1.2 Chapter 3

- *Tilia sibirica* and *T. cordata* show significant genetic differentiation
- The two are therefore considered distinct species, clarifying taxon status of *T. sibirica*
- The split of the two species from a common ancestor occurred post-glacial (early Holocene)
- DIYABC analyses suggest a westerly migration of *Tilia cordata* genotypes from Eastern Europe/West Russian plains
- Genetic diversity in *T. sibirica* appears to be significantly lower than *T. cordata*
- An immediate conservation programme should be considered for *T. sibirica*

6.1.3 Chapter 4

- In general, clonal occurrence in *Tilia* was less than other partially clonal forest trees
- *Tilia sibirica* shows more incidence of clonality than *T. cordata* and *T. platyphyllos*
- Clonality was generally greater at the range-edge of *T. cordata* and *T. platyphyllos* than at central European locations
- The hybrid (*T. x europaea*) appears to have greater clonal occurrence than either *T. cordata* and *T. platyphyllos*
- Clonal growth does not appear to be having negative (or positive) effects on genetic diversity of UK and European *Tilia*

6.1.4 Chapter 5

- A simple and efficient protocol for an RNA-Seq study of the *Tilia* leaf transcriptome is reported
- A method for assembling transcripts using two assemblers is provided
- Several thousand SSRs have been identified and following validation will provide a useful genetic resource for further population genetic studies
- Raw reads and assembled transcripts will be made available for future genomic/transcriptomic studies and as such has provided an opportunity to expand the *Tilia* genetic toolkit

6.2 Preserving the historical and ecological importance of *Tilia*

The reason that there have been so few genetic studies of lime is likely due to their non-commercial value in the UK and many parts of Europe (Pigott, 2012). As it is not regarded as a model organism, carrying out molecular studies or creating genetic resources for the genus was not of immediate concern. While lime wood is generally thought to be good timber, other species grow quicker and are more suited for harvesting (Pigott, 2012) and studies have been typically focused on those species. However, *Tilia* dominated the landscape until approximately 5,000 years ago (Turner, 1962; Birks *et al.*, 1975; Godwin, 1975), both in the UK and throughout much of Europe where the pollen records suggest a general reduction in many parts of its range occurred. Today, *Tilia* is still an important component of English, Welsh and many European woodlands. During medieval times, lime trees had many more

practical uses that were once exploited to a much greater degree than they are today. Pigott (2012), provides a detailed account of the many uses of lime wood throughout pre-historic times. Mesolithic people used lime wood to make boats, both small dug-out canoe-like boats and larger vessels clearly made to carry several people; bast fibres from lime were used for fishing nets and fish traps, cord, rope, thread and were regularly used in fabrics; lime's strong but pliable wood was used for pots and jugs; and of course lime produces large amounts of pollen and nectar and the honey was extensively used (and still is today) by many people (Pigott, 2012).

Above all else, like all forest tree systems, *Tilia* have an important ecological value, both for their ecosystem services and for their symbiotic relationships with various insect species, in particular the Red Data Book species, *Ernoporus tiliae* (the scarce lime bark beetle), fungi (including specific mycorrhiza species), lichen and mammals. *Tilia* populations in the UK and Europe are fragmented and isolated and at risk of further reduction due to human activities (and needs), such as farming and housing, and at potential risk of stochastic events. Presently, *Tilia* are generally resistant to common tree pathogens, and in the UK and Europe, the genus are not facing any immediate, species-specific, threat. Some trees can be affected by *Phytophthora,* a fungus-like micro-organism that causes root killing and bleeding cankers in mature trees (Forestry Commission, 2016). Different bacterial and fungal pathogens have been associated with *Tilia* in other parts of Europe, in particular on nursery and planted urban trees, such as leaf and stem spots and sooty mould (Kunneman and Albers, 1991; Stravinskienė *et al.*, 2015). However, these are not generally thought to be of serious concern to woodland trees. With the sudden loss of many ash trees across Europe from ash dieback (*Chalara fraxinea*), *Tilia* could potentially be a replacement tree for ash in native UK and European woodlands, and so genetic research should be carried out on lime trees and additional genetic resources should be developed. Being resistant to disease and having the genetic potential to remain resilient will be an important factor to future management of UK woodlands in the wake of climatic changes (Cavers and Cottrell, 2015). Therefore, it was essential to increase our understanding of the molecular ecology of *Tilia* species and provide an opportunity for further research of the genus. By maintaining our ancient woodlands with healthy populations of *Tilia,* we will not only preserve biodiversity, now and in the future, but will subsequently maintain important historical aspects of the British countryside.

6.3 Introgression of cultivar genotypes? A note to landowners and forest managers

This study has used microsatellite markers and traditional population genetic analyses to assess the fine-scale genetic diversity and structure of UK populations of *Tilia cordata* and *T. platyphyllos* (Chapter 2). It has revealed populations with high genetic diversity and highlighted small, isolated populations that may require future management. Many of the UK sites, where leaf samples were collected, are currently NNRs or SSSIs and so the presence of *Tilia* had been recorded. However, at one site, *T. platyphyllos* was noted to be present but following genetic analyses, some of the trees sampled from this particular site were identified as the hybrid. This may not seem unusual, considering the high number of hybrids sampled from woodlands in Herefordshire, Worcestershire and Monmouthshire, where the two species overlap. However, these particular hybrids were genetically more similar to cultivar types rather than natural hybrids (Logan and Wolff, *unpub. data*), suggesting cultivar planting in natural woodlands. Another occurrence of possible cultivar planting was observed at three other sites involving *T. cordata* trees (data not included in this study). Cultivars being planted in or close to ancient or semi-natural woodlands (see for example Appendix 1.7) could cause even more deterioration of the naturally occurring species (and their relict genotypes), that grow close by. As pointed out in Chapter 4, the hybrid appears to have a greater ability for clonal growth than both parent species and this could further threaten the already small and fragmented populations of *T. cordata* and *T. platyphyllos.* Genetic consequences should be taken into consideration when replanting any woodland with sourced trees.

Molecular techniques have been used in studies of declining or important forest species (*e.g.* Awad et al. 2014), and can provide advice on suitable sources for translocation and/or replanting (Boshier and Buggs, 2015; Wagner *et al.*, 2015). This present study offers guidance by highlighting sites that may be more suitable for sourcing *Tilia* trees for replanting. For example, sites with few trees or with relatively low genetic diversity can be restocked with trees from other locations where trees are genetically similar. Landowners and forest managers should consult this study and other studies of this type, that carry out genetic research of forest trees, particularly if that research involved trees from their land or is focused on species that grow in their woodland.

Areas of ancient woodland are deteriorating across Europe, becoming smaller and more isolated. As a result many forest tree species, and the organisms that depend

on them, are becoming more fragmented. These forests are usually under some degree of management but it is unclear exactly how they are managed. For example, what particular species do forest managers or landowners focus on and why? Which species are of particular concern within the woodlands? To what degree are sites being replanted with sourced stock from other woodlands or indeed from commercial growers? How many of these woodlands are connected to one another? How can forest connectivity be improved - creating green corridors? Who ultimately decides on which species, and how many of each, to plant? Is it the landowners or is it the woodland management organisation? The uncertainty is more pronounced in eastern European and Siberian woodlands, where records of management systems are more difficult to obtain than in the UK.

In reality, private landowners (or forest managers) may not require anything more than a good canopy cover for their woodland, and so might challenge the need to be specific about their choice of tree cover or from where they source the trees. This may be occurring in UK woodlands at present (*pers. obs.* and personal opinion). Many landowners may not even be aware of the potential negative genetic effects (*e.g.* outbreeding or inbreeding depression), that replanting incompatible trees *i.e.* those that are genetically and/or geographically different, can cause.

6.4 Sites to consider for future conservation management and restoration

Despite being highly fragmented and isolated in parts of their UK range, this present study has revealed that *Tilia cordata* and *T. platyphyllos* have high levels of genetic diversity (see Chapter 2). Given that they rarely reproduce sexually (certainly at northern UK locations) and are planted less than other tree species (at least *T. cordata,* D. Pigott, *pers. comm.*), the diversity found here is likely to be due to remnants of ancient diversity and, as such, the surviving genotypes could have adaptive genetic traits that are worth preserving. For example, trees that are locally adapted to a particular environment or specific abiotic stresses, will likely survive better in those conditions than trees adapted to different environments and stresses. So sourcing trees from elsewhere (that might be adapted to different conditions) could be detrimental. With some conservation effort and suitable management, populations of *Tilia* may be restored, particularly at sites where numbers of individual lime trees are low, *e.g.* Eywas Harold, Suckley Hills and Lulsley. *Tilia* grow in small, sometimes mixed (*i.e.* one or both species present, with or without the hybrid), scattered groups at these sites and at others close by. Intra-specific structure has

been observed in UK *Tilia* populations and so replanting of trees from woodlands that have similar genotypes should be considered. This study has shown that *T. cordata* in the Herefordshire and Worcestershire sites are genetically similar to one another (*i.e.* they group together), and so trees could be sourced locally to restock fragmented and isolated areas. Additionally, greater connectivity between these woodlands (*e.g.* green corridors) should be created.

Green corridors are wooded areas that connect fragmented pockets of woodland. Their utility in "reversing the effects of fragmentation on biodiversity" has been addressed by the Forestry Commission (FC) in a report published online [available at: http://www.forestry.gov.uk/fr/urbangreenspace]. Suggested approaches in quantifying the effects of fragmentation, considered by the FC, have included 'Landscape ecology' and 'Focal species modelling' (Watts *et al.*, 2005; Eycott *et al.*, 2007; Watts *et al.*, 2007). These approaches are useful in identifying actual species or 'generic' groups of conservation concern but seem to be more focused on species that use the woodlands rather than the tree species that make up the woodland. The need to 'protect', 'restore', 'improve' and 'manage' Ancient Semi-Natural Woodland, Plantations on Ancient Woodland Sites, and broadleaved woodlands are highlighted in these reports and the need to 'replant' areas (broad-leaved, coniferous and mixed woodlands) are addressed. The FC have also produced several reports highlighting the need for genetic analyses of their forest trees (Ennos *et al.*, 2000; Wilson and Samuel, 2002; Lowe, 2004), and state that genetic erosion, particularly at ancient woodland sites, must not occur. Although, only one FC owned woodland was sampled in this present study, the FC own several small pockets of forests at sites in Counties on either side of the Welsh border and some of those sites presently have *Tilia.* Therefore, landowners (FC and others), considering restoring *Tilia* stands in the Herefordshire and Worcestershire sites should consider the findings of this present study and source trees locally from nearby woodlands thus keeping the genetic integrity of the nearby ancient woodlands intact.

Another UK site that should be considered for restoration is Applegarth Scar in Swaledale, North Yorkshire. This small but genetically diverse relict population of *T. platyphyllos* are potentially centuries old, and scattered across an old coppice woodland on both the upper and lower part of a steep rocky limestone cliff (see Appendix 1.6). It is unlikely that these trees were planted but were probably managed at some point, although self-coppicing could have occurred as trees become damaged due to the loose, rocky terrain. The population is the most northerly native

UK population and is the most north westerly natural European *T. platyphyllos* population (Pigott, 2012). The site is relatively isolated compared to the other sites across England but shares similar genotypes with trees in Anston Stones Wood (AW) in south Yorkshire and Chanstone Wood (CW) further south in Herefordshire. Genetically similar populations, hundreds of kilometres apart are likely to possess ancestral genotypes that have been maintained within these populations. The trees at the northern site (AS) might be locally adapted to less optimum conditions (for example colder climate, less light or less fertile soils) and so there should be an effort to ensure that these genotypes are preserved. Few seedlings were observed at the site so natural regeneration may be restricted. However, clonal occurrence is relatively high at AS. This is likely due to the woodland's topography *i.e.* steep rocky outcrop. Clonality within *Tilia* populations appears to be maintaining genetic diversity (see Chapter 4). Careful management of AS could include removing fertile seeds, germinating them *ex situ* then replanting young trees at the site. This will ensure that the diverse, locally adapted, northern population is maintained.

Likewise, *T. platyphyllos* in Barton Hills (BH) in Bedfordshire are genetically distinct from other *T. platyphyllos* populations sampled in this study. The site is a SSSI and is currently managed by Natural England. It is recommended that the lime trees here undergo continued careful management. While the site has many old *Tilia* trees, there were several seedlings scattered throughout but few young recruits. An *in situ* or *ex situ* recruitment of large-leaved lime from the trees that already grow in this woodland will maintain the genetic diversity and integrity of the site.

With *Tilia* currently present and naturally reproducing from seed at some sites, it suggests that the habitat and the climatic conditions (at these particular sites) are suitable for the two species, respectively. With climate change expected to contribute to a rise of UK temperatures, it is likely that some species may be affected more than others. Some may benefit by increasing or shifting their range, while some species may not have time for their phenology to adapt to the much quicker change in climatic conditions (Walther *et al.*, 2002; Parmesan, 2006).

Tilia cordata rarely produce viable seeds in northern parts of England but with a warming climate we may see more natural regeneration, from seed, of this species at higher latitudes, perhaps even as far north as Scotland, where the species does not naturally occur. Ironically, while a warming climate may benefit *Tilia* species in the UK by permitting them to migrate further north, a sudden migration of *Tilia* genotypes from southern populations may not be so beneficial for ancient woodlands in northern

England. However, it is clear that some attention must be directed to tree species migration *e.g.* assisted migration (Vitt *et al.,* 2010), and more suitable habitats for *Tilia* may need to be created. For species that have the capability to track environmental and ecological changes brought on by climate change, increasing suitable habitat will permit those species to establish themselves. Collingham and Huntley (2000), showed a significant relationship between migration rates and available habitat. Their simulated study used the biological parameters (*i.e.* reproductive and dispersal characteristics) of *Tilia cordata,* as a model to investigate migration rates across heterogeneous landscapes. They conclude that any migration of *T. cordata* will be affected by the lack of suitable habitat if availability falls to less than 25% across the landscape, and for the species to successfully migrate, habitat 'patches' – if they must be fragmented – should be small and close together rather than large and further apart. *Tilia* populations in the UK are generally small and fragmented with large distances between one another in some parts of their range. While corridors could connect these better (as mentioned above), the results from Collingham and Huntley (2000) and references therein, suggest that perhaps creating more smaller habitats but closer together will be better suited for species migration rather than creating thin corridors connected to larger patchy areas. However, one caveat from their study (regarding the potential spread of *Tilia*), is that their findings may be more applicable to *T. cordata* rather than *T. platyphyllos* as the latter's seeds are much larger and heavier. Grashof-Bokdam (1997), suggests that species with light seeds are less affected by habitat connectivity than those with large seeds. So the management of the large-leaved lime may not necessarily be the same as that of the small-leaved lime. While the notion that small, separated populations are sufficient for future range expansion has been addressed before (Bennett, 1985; Birks, 1989), it is important to ensure that genetic diversity at such sites is also sufficient to encourage species to take hold, adapt, and remain competitive and subsequently to contribute to their range distribution (*e.g.* MacDonald and Cwynar, 1985). This way species can use such 'patches' as a 'stepping stone' for further migration (Collingham and Huntley, 2000) and the success of that migration will not be restricted by confounding genetic factors.

While this study has described the genetic diversity of UK *Tilia* populations, it has also reported the extent of clonality in the two UK species and the hybrid (see Chapter 4). In general, clonality in UK *Tilia* was less than other partially clonal forest trees, but UK lime did appear to be more clonal than central European lime

populations, suggesting a range-edge effect. Clonality has its advantages and disadvantages, and it seems that genetic diversity is being maintained in UK populations. However, forest managers and landowners need to know that the hybrid appears to have a greater ability for clonal growth than the two species and address this in their future management plans.

6.5 How this study can help preserve lime populations outside the UK

The Siberian lime (*T. sibirica*) is endemic to the south-western corner of the Altai Mountains and populations are highly fragmented and isolated (Novák *et al.,* 2014). This was the first genetic study to be carried out on the species and the study has revealed genetic differences between it (the Siberian lime) and the small-leaved lime and further suggests that the two are distinct biological units that split from a common ancestor in the early Holocene. During this period, climatic fluctuations (Groisman *et al.,* 2013) and continuous retraction and expansion of *Tilia* in the region (Bolikhovskaya and Shunkov, 2014) may have caused the isolation and subsequent split. The study also suggests that despite logging in the region, the effective population size of *T. sibirica* has not changed and this may have resulted from a shift from sexual to asexual reproduction. In Chapter 4, it was revealed that the species has high clonal incidence compared to *T. platyphyllos* and *T. cordata.* Furthermore, genetic diversity in the Siberian lime is significantly lower than that of the small-leaved lime from Siberia, Poland and Austria. Therefore, an immediate conservation programme should be considered for the species.

Using the microsatellite markers, there was no intra-specific structure observed in *T. sibirica* populations and low to moderate genetic differentiation (see Chapter 3). Further research using DNA sequences are currently ongoing to confirm or further explain the genetic status of the Siberian lime (Logan and Wolff, *unpub. data*). Restoration of the species could include introducing genotypes (*i.e.* trees from one of the larger, more genetically diverse populations *e.g.* K12 or K28) into sites with fewer individuals or sites that predominantly consist of more clones (*e.g.* K22 or K38).

Tilia cordata from the Białowieża Forest Special Reserve, Poland are largely unmanaged and have been for many generations. Trees sampled for this area show high levels of genetic diversity and appear to be reproducing sexually and recruiting young trees, naturally (see Chapter 4). The forest is without a doubt a testimony to the concept of permitting forests to grow, reproduce, die and regenerate in a natural, unmanaged state, and one that would surely benefit many other European countries.

However, it is unlikely that this will become a reality in the UK or indeed other parts of Europe and it is only a matter of time until the last remaining pocket of European primeval forest will slowly begin to make way for other land uses such as agriculture and housing.

6.6 From population genetics to population genomics

As well as using traditional population genetic approaches, this study has used Next Generation Sequencing techniques to assemble the leaf transcriptome of *T. cordata* and *T. platyphyllos* (Chapter 5). Studies of non-model organisms, using Next Generation Sequencing (NGS) technologies, are increasing, and from the number of recent publications using such techniques, seem to be becoming standard practice. This present study has taken advantage of the current paradigm, that is, that every organism can now be a 'model' species, where large amounts of genomic and/or transcriptomic data can be generated and analysed at a relatively low cost. The advancement of NGS has provided opportunities for more genetic laboratories to concentrate their research on systems that would not normally have been at the forefront of genetic research, such as *Tilia.*

Recognition of *Tilia* as an important forest resource has to be taken into consideration and as such, genomic and transcriptomic information from the genus must be available for future studies. So, while a step towards omics was ambitious it was a necessary step forward for *Tilia* research. The availability of *Tilia* transcripts will now further lead to an increased understanding of the molecular ecology of the two UK species and will contribute to future genomic research and a more comprehensive understanding of the genus as whole.

6.7 Recommended additional work

6.7.1 Chapter 3: Tilia sibirica

Approximate Bayesian Computation (ABC) analyses were used in the program DIYABC. The program is gaining popularity in population and evolutionary genetic studies and provides estimates on various historical and recent demographic events. However, results of DIYABC analyses require careful interpretation. Here only microsatellite markers, which can be affected by ascertainment bias and size homoplasy, were used. Therefore, the study could be expanded using a combination of different molecular markers *e.g.* chloroplast and single copy nuclear genes and/or more SSRs (or SNPs) retrieved from the RNA-Seq study (Chapter 5). This may

provide additional and better estimates of effective population size and dates/time of demographic events (Cornuet *et al.,* 2010). Additionally, a traditional phylogeographical approach combined with molecular dating could provide a more robust divergence date and may better explain the migration of *Tilia* genotypes from Eastern Europe and the western Russian plains to central and Western Europe.

6.7.2 Chapter 4: Clonal incidence

While only a small number of *T. x europaea* samples were assessed in this study, it appears that clonal reproduction occurs more in the hybrid than either of the parent species. If the hybrid competes better in natural conditions than *T. cordata* and *T. platyphyllos* by propagating clonally at a greater rate, this could result in further deterioration of already fragmented populations of each species. Additional work should be carried out on the hybrid in natural locations across the UK and Europe, as the hybrid is extensively planted in parks and towns (Pigott, 2012; Hansen *et al.,* 2014) and possibly even in woodlands (as mentioned above). Local authorities, councils and landowners should take care when planting the hybrid close to ancient woodland sites and be aware of potential genetic consequences of mixing genotypes.

Clonal growth within the *Tilia* species, reported here, may be an underestimation as a number of putative somatic mutations were observed. Somatic mutations are an interesting phenomenon. They differ to germ cell mutations in that they are not passed on to offspring. However, in clonally reproducing trees these mutations will persist. An accumulation of somatic mutations within a population, or indeed within a species, could suggest the persistence of clonal reproduction as these mutations will generally become fixed. To expand on this study, further research into somatic mutations within the genus should be carried out. Possible areas of interest would be to estimate a somatic mutation rate (*e.g.* Cloutier *et al.,* 2003; O'Connell and Ritland, 2004). Large clonal groups from natural stands (*e.g.* a *T. sibirica* population) or a clonal group of cultivars (*e.g.* the large clone at Westonbirt Arboretum) could be investigated. Likewise, assessing the contribution to genetic variation from somatic mutations versus sexual reproduction *e.g.* Mock et al. (2008), is another potentially interesting area to investigate. Somatic mutations within clonal groups of trees introduces genetic diversity and so could potentially be important to an organism's evolution and adaptation.

6.7.3 Chapter 5: RNA-Seq analysis

Areas to focus on, using the NGS transcriptomic data, is single nucleotide polymorphism (SNP) development. SNPs are a useful and abundant source of genetic variation. SNPs discovered using NGS techniques provide non-neutral markers that can be used to investigate genes under selection (Ouborg *et al.,* 2010; Kirk and Freeland, 2011). Variations observed in transcripts of trees from different environments or from fragmented habitats could reveal useful information about gene expression and gene activity under optimum and stressful conditions *e.g.* (Qiu *et al.*, 2011; Somervuo *et al.*, 2014). The trees sampled in this present study were collected from different sites across England and Wales. Although only twelve sites in total were analysed, geographical variation could be investigated using SNP markers. Additional samples should be added for a more comprehensive focus of UK *Tilia*, but a simple and efficient method is provided. Additionally, nucleotide substitution rates and synonymous vs non-synonymous substation ratio can be determined across species *e.g.* Pinosio et al. (2014). This may offer further insight into the evolutionary history and divergence of the two UK *Tilia* species.

6.8 Conclusion

This study has greatly advanced our understanding of population genetics and structure of three species of *Tilia.* It has highlighted several important *Tilia* populations within the UK and has provided some recommendations for future management of those sites. The study has also described for the first time, the genetic diversity of a highly fragmented and putative rare species of *Tilia* from Siberia, and suggests that an immediate conservation programme should be initiated. Clonal structure of the three species, across their respective ranges, has also been addressed for the first time, and has revealed that genus may not be as clonal as first thought but edge-range effects (more clonal occurrence at margins of the range) are apparent. Finally, this study has provided a platform and an opportunity to further expand the field of *Tilia* genomics research by assembling the leaf transcriptome of the two UK species.

Appendices

Appendix 1.1 A fallen *T. platyphyllos* tree at Barton Hills, Bedfordshire (top) and a *T. cordata* tree in Herefordshire (bottom). Fallen trees and branch extensions contribute to the clonal spread of *Tilia* across woodlands. Each upright branch from the fallen tree will eventually develop into separate trees once the main trunk is covered with soil and leaves. Branches growing from the base of *T. cordata* trunk are touching down and taking hold several metres away from the tree where they develop into separate trees. Photo: Samuel A. Logan ©

Appendix 1.2a Clonal spread can also occur from growth from root collars (top and bottom). The tree in the bottom photo is growing on steep topography at a site in Yorkshire. Root collars and self-coppiced branches (see below A1.2b) spread down the banks. Over time, a single tree can spread great distances. Photo: Samuel A. Logan ©

Appendix 1.2b A large self-coppiced *T. cordata,* on unstable, loose terrain. This tree and the tree in the top picture (A1.2a), as well as several others close by, were on the top ridge of old quarries at sites in Worcestershire and Herefordshire. Photo: Samuel A. Logan ©

Appendix 1.3a An old *T. platyphyllos* tree growing in Barton Hill, Bedfordshire. The site is regarded as ancient woodland and is managed by Natural England as a SSSI. Photo: Samuel A. Logan ©

Appendix 1.3b A *T. platyphyllos* tree growing on an overhang of a cliff in at Anston Stones Wood, Yorkshire. The site is regarded as ancient woodland and is managed by Natural England as a SSSI. Photo: Samuel A. Logan ©

Appendix 1.4 Old coppiced wood in Worcestershire, no longer managed (top) and coppiced wood in Herefordshire, currently still managed by coppicing (bottom). Photo: Samuel A. Logan ©

Appendix 1.5a A very old multi-stemmed (~10 stems) *T. cordata* tree growing in Herefordshire. This was likely an-Enclosures Act (R. Roseff *pers. comm*) boundary tree. Photo: Samuel A. Logan ©

Appendix 1.5b A very old, gnarly, almost hollow *T. cordata* tree still growing in Herefordshire. This was likely an Enclosures Act (R. Roseff *pers. comm*) or village boundary tree. It is likely to be ~800 years old or more. Photo: Samuel A. Logan ©

Appendix 1.6 Applegarth Scar, Yorkshire. *T. platyphyllos* grow naturally on steep rocky terrain. Trees self-coppice as trunks become damaged. Clonal spread occurs from damaged branches and root collars. Photo: Samuel A. Logan ©

Appendix 1.7 *T. cordata* trees growing in a woodland in Herefordshire. These trees are likely to have been planted. The sites is near an old natural *T. cordata* population. Landowners should be aware of the genetic consequences of planting cultivars near woodland sites where trees are likely to be quite old and. Photo: Samuel A. Logan ©

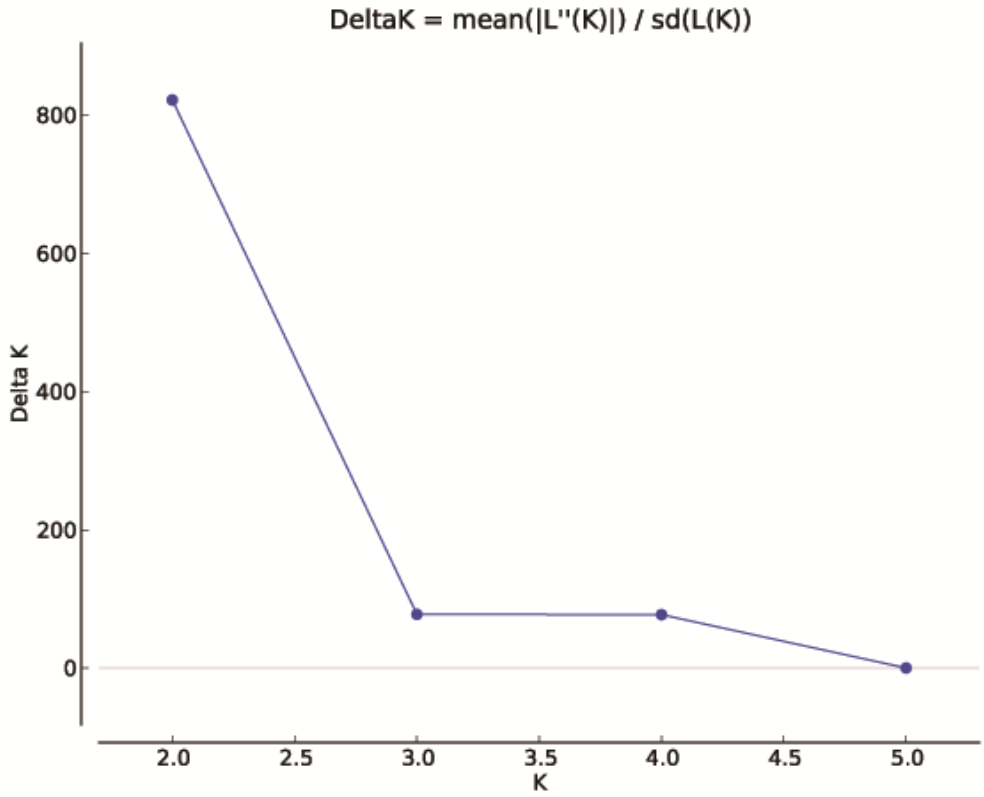
Appendix 2.1 Putative species present, sites sampled, codes, latitude and longitude coordinates. Secondary coding of populations in text are denoted with Tp, Tc or Txe representing *T. platyphyllos, T. cordata* and Txe – *T. x europaea* respectively.

Species	Population	Code	Latitude (°N)	Longitude (°E)
T. platyphyllos	Applegarth Scar[1]	AS	54.4093	-1.8165
(Tp)	Whitcliffe Wood[1]	WW	54.4097	-1.7792
	Hayburn Wyke[1]	HbW	54.3600	-0.4518
	Anston Stone Wood[1]	AW	53.3402	-1.1985
	Barton Hills[1]	BH	51.9562	-0.4220
Mixed[3]	Lulsley[1]	LU	52.1914	-2.3888
	Suckley Hills[1]	SU	52.1826	-2.3921
	Sheepshill Coppice[1]	SC	52.1595	-2.3772
	Knapp & Papermill[1]	KP	52.1615	-2.3713
	Halesend Wood[1]	HW	52.1409	-2.3815
	West Malvern[1]	WM	52.1189	-2.3581
	Brockhill Wood[1]	BW	52.0928	-2.3538
	Lady Park Wood[1]	LP	51.8245	-2.6568
	Woolhope[1]	WH	52.0259	-2.6042
	Eywas Harold[1]	EH	51.9596	-2.9136
	Chanstone Wood[1]	CW	52.0111	-2.9422
	Covenhope[1]	CH	52.2837	-2.8855
T. cordata	Brignall Banks[1]	BB	54.4968	-1.9114
(Tc)	Roudsea Wood[2]	RW	54.2332	-3.0255
	Skelghyll Wood[2]	SkW	54.4197	-2.9519
	Hardy Gang Wood[2]	HG	53.2607	-0.3614
	Collyweston Wood[2]	CGW	52.5974	-0.5176
	Easton Hornstock[2]	EaH	52.5912	-0.4981
	Bedford Purlieus[2]	BP	52.5833	-0.4646
	Dowles Brooke[2]	DB	52.3834	-2.3364
	Shrawley Wood[2]	ShW	52.2917	-2.2820
	Collin Park Wood[2]	CPW	51.9463	-2.3695

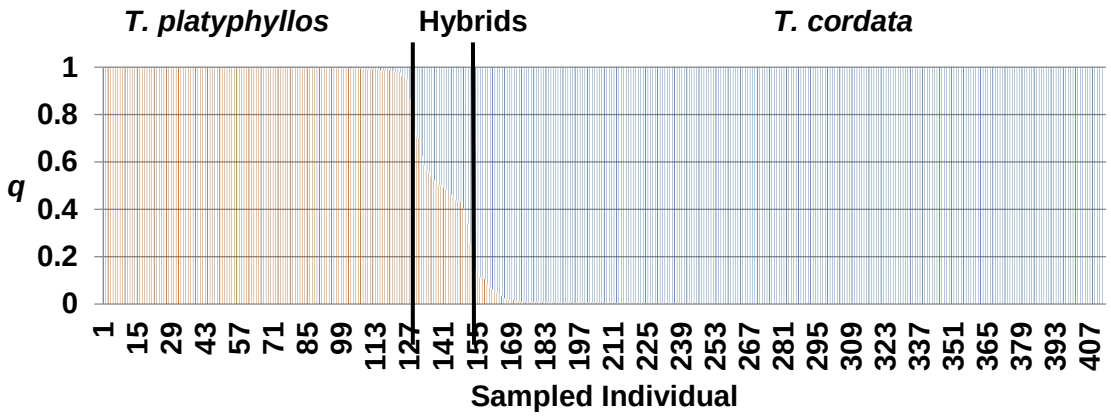
[1] Samples collected by Samuel Logan in August 2012 and between June-July 2013.
[2] Samples collected by Dr Paul Ashton (Edge Hill University), June-August 2012.
[3] Mixed sites had either both species present and/or one species and the hybrid.

Appendix 2.2 Results from Micro-Checker showing possible null alleles at each locus in each species

	Tc6	Tc937	Tc920	Tc8	Tc943	Tc31	Tc4	Tc927	Tc915	Tc963	Tc5	Tc951	Tc7
Pop													
TcLP	-	-	-	-	-	-	null	-	-	-	-	-	-
TcBB	-	-	-	-	-	null	-	-	-	-	-	-	-
TcKP	null	-	-	-	-	-	-	-	-	null	-	-	-
TcCW	-	-	-	-	-	-	-	-	-	-	-	-	-
TcDB	-	-	-	-	-	-	-	-	-	-	-	-	-
TcSkW	-	-	-	-	-	-	-	-	-	-	-	-	-
TcEaH	-	-	-	-	-	-	-	-	-	null	-	-	-
TcBP	-	-	-	-	-	-	-	-	-	-	-	-	-
TcCPW	-	-	-	-	-	-	-	-	-	-	-	-	-
TcHG	-	-	-	-	-	-	-	-	-	null	-	-	-
TcCGW	-	-	-	-	-	-	-	-	-	-	-	-	null
TcShW	-	-	-	-	-	-	-	-	-	-	-	-	-
TcRW	-	-	-	-	-	-	-	-	-	null	-	-	-
TcLu	-	-	-	-	-	-	-	-	-	-	-	-	-
TcSu	-	-	-	-	-	-	-	-	-	-	-	-	-
TcEH	-	-	-	-	-	-	-	-	-	-	-	-	-
TpAS	-	-	-	-	-	-	-	-	-	-	-	-	-
TpBH	-	-	-	-	-	-	-	-	-	-	-	-	-
TpAW	-	-	-	-	-	-	-	-	-	-	-	-	-
TpLP	-	-	-	-	-	-	-	-	-	-	-	-	-
TpHW	-	-	-	-	-	-	-	-	-	null	-	null	-
TpKP	-	-	-	-	-	-	-	-	-	null	-	null	-
TpCW	-	-	-	-	-	-	-	-	-	-	-	-	-
TpSu	-	-	-	-	-	-	-	-	-	null	-	null	-



Appendix 2.3 Evanno's ΔK revealing K = 2, implemented in Structure harvester



Appendix 2.4 Membership coefficient (*q*) values of all individuals. Individuals (with *q*-value of between 0.20 and 0.80), shown between the two black lines, are considered hybrids, edited in Microsoft Excel 2010

Appendix 2.5 Allele fragment size at locus *Tc8*, and *q* – values of 32 (including shaded individuals) and 25 putative hybrid individuals following threshold 0.10 and 0.20, respectively. Cluster 1 is *T. platyphyllos* and cluster 2 is *T. cordata*

Allele size		*q*- value	
Locus *Tc8*	Tree ID	Cluster 1	Cluster 2
156,158	WM03	0.866	0.134
156,158	WM01	0.865	0.135
141,158	CW06	0.735	0.265
141,158	EaH04	0.706	0.294
158,164	HbW02	0.701	0.299
141,158	WW01	0.625	0.375
141,158	LP27	0.586	0.414
141,158	CW_sd2	0.564	0.436
141,158	KP18	0.556	0.444
141,158	CW25	0.538	0.462
141,160	AW11	0.521	0.479
141,158	KP06	0.508	0.492
141,160	AW12	0.501	0.499
141,158	AW13	0.501	0.499
141,156	BB03	0.492	0.508
141,156	BB05	0.483	0.517
141,166	WW03	0.468	0.532
141,158	LP28	0.461	0.539
141,156	BW002	0.451	0.549
141,158	CW13	0.437	0.563
141,160	CW11	0.434	0.566
141,158	CW28	0.433	0.567

Appendix 2.5 continued

Allele size		*q*- value	
Locus *Tc8*	Tree ID	Cluster 1	Cluster 2
141,164	HbW04	0.404	0.596
141,162	LP10	0.394	0.606
141,158	AS16	0.341	0.659
141,141	BW001	0.257	0.743
141,141	RW26	0.230	0.770
141,141	BP04	0.147	0.853
141,141	HG01	0.134	0.866
141,141	ShW10	0.113	0.887
141,141	RW37	0.112	0.888
141,141	RW05	0.105	0.895

Appendix 2.6 Number of alleles per locus and populations with private alleles at each locus

Locus	*T. cordata*		*T. platyphyllos*	
	No. of alleles	**Pop. Private alleles**	**No. of alleles**	**Pop. Private alleles**
Tc6	8	–	11	TpCW, TpLP
Tc937	6	–	9	TpHW, TpCW
Tc920	11	TcSU	11	TpLP
Tc8	1	–	8	TpBH
Tc943	8	TcEaH	5	–
Tc31	6	–	10	TpHW, TpCW
Tc4	13	TcCPW	13	TpBH, TpKP
Tc927	4	TcEH	15	TpBH, TpHW, TpLP
Tc915	16	TcCPW	18	–
Tc963	26	TcEH	15	TpCW, TpHW
Tc5	13	TcCGW	12	TpBH, TpCW
Tc951	7	TcEaH	9	TpHW
Tc7	8	TcCGW	13	TpAW, TpBH, TpSU
Total	127	9	149	21

Appendix 2.7 Diversity measures of 16 T. cordata populations. N – number of samples; N_A – Average number of alleles; A_E – Effective number of alleles; H_O – Observed heterozygosity; H_E – Expected heterozygosity.

Pop	***N***	***N_A***	***A_E***	***H_O***	***H_E***
TcBB	*8*	*3.69*	*2.76*	*0.58*	*0.51*
TcBP	*20*	*5.15*	*3.06*	*0.54*	*0.53*
TcCPW	*8*	*4.23*	*3.01*	*0.53*	*0.49*
TcCW	*13*	*3.69*	*2.67*	*0.50*	*0.49*
TcCGW	*31*	*4.62*	*2.91*	*0.58*	*0.52*
TcDB	*17*	*4.92*	*3.09*	*0.53*	*0.52*
TcEaH	*18*	*5.38*	*3.54*	*0.53*	*0.56*
TcEH	*6*	*3.31*	*2.61*	*0.51*	*0.50*
TcHG	*20*	*5.46*	*3.20*	*0.49*	*0.51*
TcKP	*15*	*4.23*	*2.57*	*0.46*	*0.51*
TcLP	*11*	*4.15*	*2.98*	*0.51*	*0.53*
TcLU	*5*	*2.46*	*1.93*	*0.34*	*0.35*
TcRW	*30*	*5.38*	*3.24*	*0.56*	*0.57*
TcShW	*20*	*5.46*	*3.47*	*0.59*	*0.56*
TcSkW	*14*	*4.31*	*2.79*	*0.52*	*0.48*
TcSU	*10*	*3.38*	*2.73*	*0.52*	*0.45*
Mean	*15.38*	*4.36*	*2.90*	*0.52*	*0.51*

Appendix 2.8 Diversity measures of eight T. platyphyllos populations. N – number of samples; N_A – Average number of alleles; A_E – Effective number of alleles; H_O – Observed heterozygosity; H_E – Expected heterozygosity.

Pop	***N***	***N_A***	***A_E***	***H_O***	***H_E***
TpAS	*8*	*3.69*	*2.68*	*0.73*	*0.60*
TpAW	*14*	*5.08*	*3.36*	*0.69*	*0.66*
TpBH	*14*	*6.31*	*3.80*	*0.78*	*0.71*
TpCW	*15*	*5.62*	*3.27*	*0.75*	*0.66*
TpHW	*26*	*6.69*	*3.91*	*0.72*	*0.71*
TpKP	*11*	*4.85*	*3.38*	*0.67*	*0.67*
TpLP	*15*	*5.62*	*3.24*	*0.70*	*0.66*
TpSU	*13*	*5.77*	*4.00*	*0.66*	*0.73*
Mean	*14.5*	*5.45*	*3.46*	*0.71*	*0.68*

Appendix 2.9 Among species population pairwise F_{ST} and their significance. * - 0.05; ** - 0.01; NS – not significant

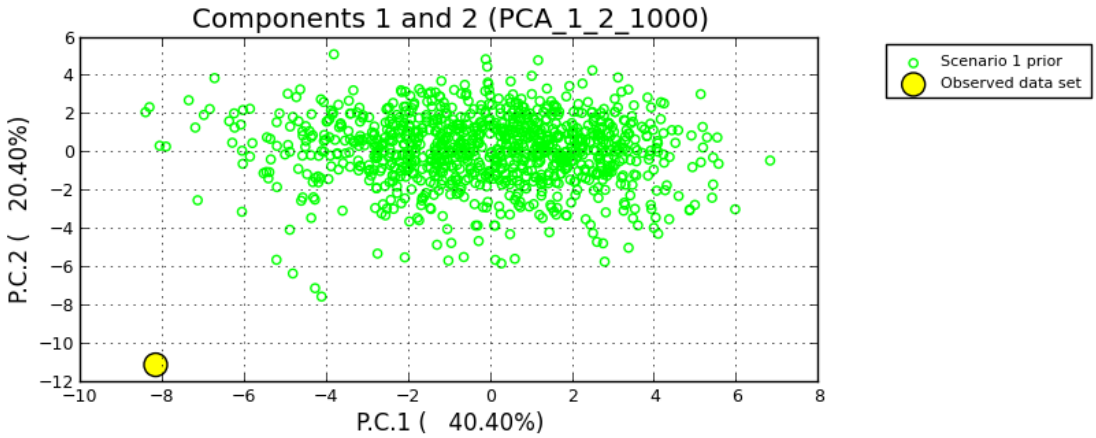
pop	TcBB	TcBP	TcCPW	TcCW	TcCGW	TcDB	TcEaH	TcEH	TcHG
TcBB	–	NS	NS	**	**	*	NS	NS	NS
TcBP	0.053	–	NS	**	**	**	*	**	**
TcCPW	0.044	0.018	–	NS	*	NS	NS	NS	NS
TcCW	0.112	0.065	0.046	–	**	**	**	*	**
TcCGW	0.091	0.039	0.054	0.093	–	**	**	**	**
TcDB	0.083	0.070	0.021	0.070	0.101	–	**	**	**
TcEaH	0.026	0.029	0.012	0.048	0.051	0.047	–	**	NS
TcEH	0.173	0.186	0.161	0.176	0.204	0.148	0.156	–	**
TcHG	0.053	0.024	0.026	0.072	0.054	0.095	0.021	0.171	–
TcKP	0.093	0.102	0.072	0.123	0.113	0.045	0.059	0.171	0.104
TcLP	0.038	0.045	-0.009	0.025	0.074	0.031	0.009	0.140	0.039
TcLU	0.146	0.127	0.052	0.077	0.163	0.048	0.073	0.185	0.123
TcRW	0.046	0.083	0.040	0.096	0.113	0.081	0.054	0.167	0.090
TcShW	0.069	0.047	0.024	0.040	0.059	0.035	0.025	0.151	0.061
TcSkW	0.115	0.079	0.067	0.114	0.098	0.108	0.083	0.240	0.083
TcSU	0.106	0.094	0.035	0.054	0.130	0.092	0.066	0.189	0.079
TpAS	**0.387**	**0.374**	**0.396**	**0.403**	**0.397**	**0.399**	**0.364**	**0.378**	**0.397**
TpAW	**0.361**	**0.360**	**0.368**	**0.382**	**0.386**	**0.373**	**0.348**	**0.354**	**0.383**
TpBH	**0.318**	**0.329**	**0.332**	**0.350**	**0.348**	**0.331**	**0.306**	**0.307**	**0.355**
TpCW	**0.351**	**0.351**	**0.359**	**0.374**	**0.375**	**0.352**	**0.335**	**0.339**	**0.373**
TpHW	**0.324**	**0.329**	**0.335**	**0.346**	**0.358**	**0.343**	**0.319**	**0.318**	**0.353**
TpKP	**0.354**	**0.352**	**0.361**	**0.387**	**0.386**	**0.372**	**0.347**	**0.347**	**0.379**
TpLP	**0.340**	**0.347**	**0.358**	**0.375**	**0.380**	**0.362**	**0.340**	**0.338**	**0.375**
TpSU	**0.320**	**0.335**	**0.329**	**0.348**	**0.360**	**0.346**	**0.314**	**0.307**	**0.352**

Appendix 2.9 Among species population pairwise F_{ST}. continued

pop	TcKP	TcLP	TcLU	TcRW	TcShW	TcSkW	TcSU
TcBB	**	NS	NS	**	**	**	*
TcBP	**	**	**	**	**	**	**
TcCPW	NS	NS	NS	NS	NS	**	NS
TcCW	**	NS	NS	**	**	**	*
TcCGW	**	**	**	**	**	**	**
TcDB	*	NS	NS	**	**	**	**
TcEaH	**	NS	NS	**	*	**	**
TcEH	**	*	NS	**	**	*	NS
TcHG	**	*	*	**	**	**	**
TcKP	–	NS	NS	**	**	**	**
TcLP	0.048	–	NS	**	NS	**	NS
TcLU	0.099	0.034	–	*	NS	*	NS
TcRW	0.095	0.029	0.101	–	**	**	**
TcShW	0.062	0.030	0.091	0.067	–	**	**
TcSkW	0.131	0.083	0.155	0.090	0.061	–	**
TcSU	0.142	0.051	0.134	0.100	0.074	0.149	–
TpAS	**0.398**	**0.380**	**0.454**	**0.366**	**0.358**	**0.387**	**0.421**
TpAW	**0.374**	**0.356**	**0.414**	**0.347**	**0.342**	**0.380**	**0.403**
TpBH	**0.316**	**0.314**	**0.371**	**0.314**	**0.305**	**0.355**	**0.365**
TpCW	**0.347**	**0.344**	**0.395**	**0.348**	**0.336**	**0.375**	**0.396**
TpHW	**0.338**	**0.322**	**0.375**	**0.309**	**0.316**	**0.344**	**0.357**
TpKP	**0.372**	**0.356**	**0.411**	**0.345**	**0.342**	**0.376**	**0.394**
TpLP	**0.360**	**0.343**	**0.401**	**0.333**	**0.337**	**0.370**	**0.388**
TpSU	**0.340**	**0.318**	**0.371**	**0.313**	**0.316**	**0.350**	**0.364**

Appendix 2.9 Among species population pairwise F_{ST}. continued

pop	TpAS	TpAW	TpBH	TpCW	TpHW	TpKP	TpLP	TpSU
TcBB	*	**	**	**	**	*	**	*
TcBP	**	**	**	**	**	**	**	**
TcCPW	*	**	**	**	**	*	**	*
TcCW	**	**	**	**	**	**	**	**
TcCGW	**	**	**	**	**	**	**	**
TcDB	**	**	**	**	**	**	**	**
TcEaH	**	**	**	**	**	**	**	**
TcEH	**NS**	**	**	*	**	**NS**	**	**NS**
TcHG	**	**	**	**	**	**	**	**
TcKP	**	**	**	**	**	**	**	**
TcLP	**	**	**	**	**	**	**	**
TcLU	**NS**	*	*	*	**	**NS**	*	**NS**
TcRW	**	**	**	**	**	**	**	**
TcShW	**	**	**	**	**	**	**	**
TcSkW	**	**	**	**	**	**	**	**
TcSU	*	**	**	**	**	**	**	*
TpAS	–	**	**	**	**	*	**	*
TpAW	0.122	–	**	**	**	**	**	**
TpBH	0.196	0.164	–	**	**	**	**	**
TpCW	0.139	0.101	0.129	–	**	**	**	**
TpHW	0.139	0.086	0.118	0.105	–	**	**	**
TpKP	0.106	0.059	0.149	0.079	0.079	–	**	*
TpLP	0.108	0.077	0.136	0.084	0.083	0.046	–	**
TpSU	0.108	0.095	0.119	0.090	0.074	0.075	0.073	–



Appendix 3.1 The PCA graph of the SDM using all available summary statistics shows the observed dataset (yellow dot) clearly out of position following 1000 simulated datasets (green dots) suggesting that some statistics were not suited for the priors.

Appendix 3.2 SDM using all available summary statistics: The proportion of simulated data sets that have values below the observed data set are presented. One-sample statistics – mean number of alleles (*A*), genic diversity (*H*), allele size variance (*V*), Garza-Williamson *M* (*MGW*); Two-sample statistics - mean number of alleles (*A2P*), genic diversity (*H2P*), allele size variance (*V2P*), genetic differentiation (F_{ST}), the classification index (*LIK*), shared allele distance (*DAS*) and genetic distance between two samples $(\delta\mu)^2$. (*) = 0.05. (**) = 0.01, (***) = 0.001. Those at the 0.01 and 0.001 level were omitted from further analyses.

Summary statistic	Observed value	*P* (sim<obs)
*A*_1_1	4.6667	0.0623
*A*_1_2	10.1667	0.3011
*H*_1_1	0.3541	0.0279(*)
*H*_1_2	0.6096	0.1074
*V*_1_1	7.0080	0.2527
*V*_1_2	9.3921	0.3254
*MGW*_1_1	0.6054	0.0053(**)
*MGW*_1_2	0.8299	0.1583
*A2P*_1_1&2	11.3333	0.0532
*H2P*_1_1&2	0.5412	0.0004(***)
*V2P*_1_1&2	8.9478	0.0196(*)
F_{ST}_1_1&2	0.1672	0.5481
*LIK*_1_1&2	1.1762	0.0000(***)
*LIK*_1_2&1	2.2838	0.0030(**)
*DAS*_1_1&2	0.4196	1.0000(***)
$(\delta\mu)^2$_1_1&2	2.7112	0.0089(*)

Appendix 3.3 SDM using the optimum set of summary statistics: One-sample statistics – mean number of alleles (*A*), genic diversity (*H*), allele size variance (*V*); Two-sample statistics - mean number of alleles (*A2P*), genic diversity (*H2P*), allele size variance (*V2P*), genetic differentiation (F_{ST}) and genetic distance between two samples $(\delta\mu)^2$. (*= 0.05).

Summary statistic	observed value	*P* (sim<obs)
*A*_1_1	4.6667	0.0958
*A*_1_2	10.1667	0.4130
*H*_1_1	0.3541	0.0448 (*)
*H*_1_2	0.6096	0.1612
*V*_1_1	7.0008	0.2419
*V*_1_2	9.3921	0.3036
*A2P*_1_1&2	11.3333	0.1615
*V2P*_1_1&2	8.9478	0.1258
*FST*_1_1&2	0.1672	0.5454
$(\delta\mu)^2$_1_1&2	2.7112	0.0920

Appendix 3.4 Conditions set for each of the historical models analysed in DIYABC (SDM, BM and EMM).

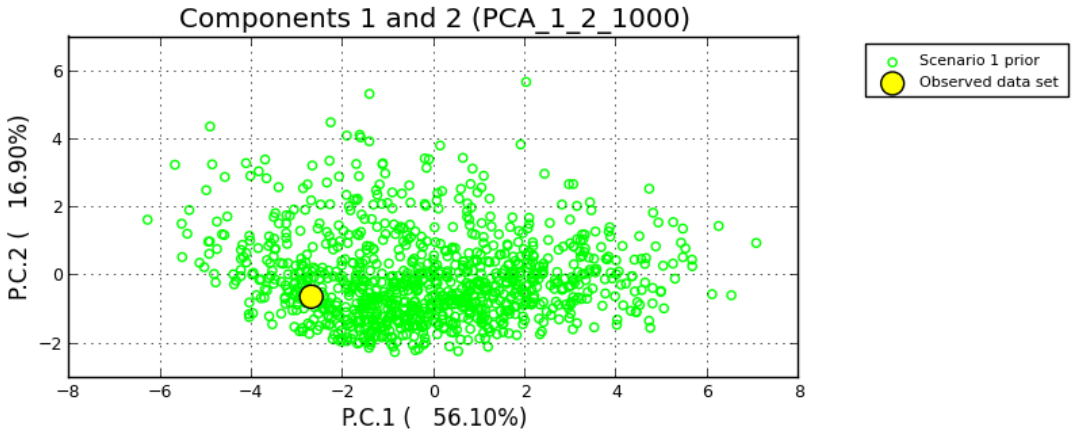
SDM	BM	EMM
Na>=N1	t1<t2	t3>=t2
Na>=N2	t1<t3	t3>=t1
	t2<t3	t2>=t1
	Na>=N1	Na>=N1
	Nai>Na	
	Nai>=N1	
	Naii>=Nai	
	Naii>=N1	
	Na>=Naiii	
	Na>=Naii	
	Naiii<N1	

Appendix 3.5 Estimated null allele frequencies at each locus within each population.

Locus	Pop	Estimate of null allele frequency	Locus	Pop	Estimate of null allele frequency
1	1	0.00000	4	1	0.00100
1	2	0.02456	4	2	0.00100
1	3	0.00000	4	3	0.00100
1	4	0.05909	4	4	0.00100
1	5	0.00000	4	5	0.00100
1	6	0.00000	4	6	0.00100
1	7	0.01465	4	7	0.00100
1	8	0.00000	4	8	0.00100
1	9	0.08851	4	9	0.00100
1	10	0.00000	4	10	0.00100
1	11	0.06628	4	11	0.00100
2	1	0.00100	5	1	0.00100
2	2	0.14980	5	2	0.00001
2	3	0.00000	5	3	0.00100
2	4	0.00100	5	4	0.00100
2	5	0.00000	5	5	0.00000
2	6	0.08611	5	6	0.04159
2	7	0.00002	5	7	0.00001
2	8	0.00000	5	8	0.00005
2	9	0.00000	5	9	0.00000
2	10	0.00000	5	10	0.09021
2	11	0.00000	5	11	0.11803
3	1	0.00100	6	1	0.33333
3	2	0.06006	6	2	0.31478
3	3	0.07644	6	3	0.14131
3	4	0.12766	6	4	0.27846
3	5	0.00001	6	5	0.25122
3	6	0.01141	6	6	0.13654
3	7	0.00000	6	7	0.00000
3	8	0.00000	6	8	0.02914
3	9	0.00000	6	9	0.00000
3	10	0.00000	6	10	0.08336
3	11	0.00000	6	11	0.00001

Appendix 3.5 (cont.) Estimated null allele frequencies at each locus within each population

Locus	Pop	Estimate of null allele frequency	Locus	Pop	Estimate of null allele frequency
7	1	0.00100	10	1	0.00000
7	2	0.00100	10	2	0.34704
7	3	0.00100	10	3	0.22950
7	4	0.00100	10	4	0.15511
7	5	0.00100	10	5	0.00000
7	6	0.00000	10	6	0.15313
7	7	0.00002	10	7	0.03213
7	8	0.00001	10	8	0.00000
7	9	0.00000	10	9	0.00000
7	10	0.00002	10	10	0.00000
7	11	0.00001	10	11	0.00000
8	1	0.00000	11	1	0.00000
8	2	0.14980	11	2	0.00001
8	3	0.00000	11	3	0.09875
8	4	0.03529	11	4	0.00001
8	5	0.05136	11	5	0.00000
8	6	0.02828	11	6	0.04728
8	7	0.05138	11	7	0.00004
8	8	0.00000	11	8	0.00000
8	9	0.00000	11	9	0.00000
8	10	0.00000	11	10	0.00000
8	11	0.03224	11	11	0.01048
9	1	0.00000	12	1	0.00100
9	2	0.21021	12	2	0.00100
9	3	0.00000	12	3	0.00100
9	4	0.13892	12	4	0.00100
9	5	0.00000	12	5	0.00100
9	6	0.01352	12	6	0.14709
9	7	0.00000	12	7	0.00836
9	8	0.00000	12	8	0.00003
9	9	0.19397	12	9	0.00000
9	10	0.00000	12	10	0.00000
9	11	0.04006	12	11	0.00001



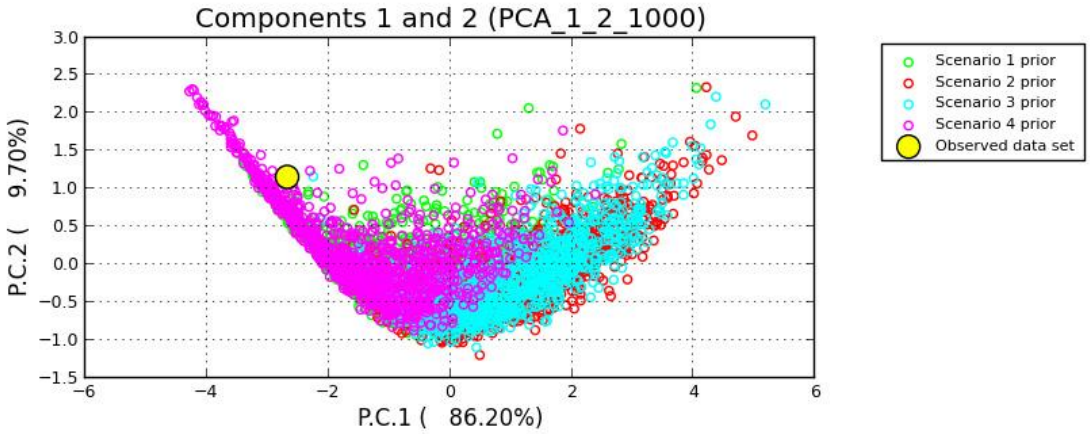
Appendix 3.6 SDM pre-evaluate scenario and priors using the optimum set of summary statistics. The PCA graph shows the observed dataset (yellow dot) positioned within 1000 simulated datasets (green dots) suggesting the model is good.

Appendix 3.7 Parameter estimates of the simple divergence model: *N1* – effective population size of *T. sibirica*; *N2* - effective population size of *T. cordata*; *Na* – effective population size of ancestral population; *td* – time of divergence in generations; *Âµmic* – mean SSR mutation rate; *pmic* – Mean *P* (parameter of geometric distribution); *snimic* – mean single nucleotide insertion (SNI).

SDM model				Quartile					
Parameter	mean	median	mode	2.5%	5%	25%	75%	95%	97.5%
N1	1270	1150	1000	363	444	808	1580	2400	2780
N2	8240	8540	9490	4820	5480	7480	9320	9860	9930
td	1420	890	369	163	216	496	1700	4680	6460
Na	2320	1450	574	101	177	710	2820	7040	9730
Âµmic	3.44E-04	2.95E-04	1.83E-04	1.20E-04	1.31E-04	2.03E-04	4.41E-04	7.34E-04	8.28E-04
pmic	0.268	0.284	0.300	0.16	0.178	0.251	0.300	0.300	0.300
snimic	7.20E-06	7.84E-06	1.00E-05	1.21E-06	2.04E-06	5.74E-06	9.24E-06	1.00E-05	1.00E-05

Appendix 3.8 Model check (ranked approach) on 10,000 simulated data from each tested models (SDM, BM, and EMM). Mean number of alleles (*A*), genic diversity (*H*), allele size variance (*V*), mean number of alleles (*A2P*), genic diversity (*H2P*), allele size variance (*V2P*), genetic differentiation (F_{ST}) and genetic distance between two samples $(\delta\mu)^2$. (*=0.05, **=0.01).

Summary statistic	SDM observed value	*P* (sim<obs)	BM observed value	*P* (sim<obs)	EMM observed value	*P* (sim<obs)
*A*_1_1	6.0833	0.4860	4.5833	0.5621	6.0833	0.3679
*A*_1_2	4.6667	0.8224	-	-	8.1667	0.5319
*A*_1_3	-	-	-	-	5.4167	0.3277
*H*_1_1	0.5841	0.0840	0.3470	0.3647	0.5841	0.1015
*H*_1_2	0.3541	0.4001	-	-	0.5696	0.0292 (*)
*H*_1_3	-	-	-	-	0.6085	0.2311
*V*_1_1	7.0208	0.8833	6.9969	0.9314	7.0208	0.3493
*V*_1_2	7.0080	0.8764	-	-	10.5710	0.5243
*V*_1_3	-	-	-	-	7.9801	0.4273
*A2P*_1_1&2	8.1667	0.8483	-	-	9.3333	0.5115
*A2P*_1_1&3	-	-	-	-	7.6667	0.3781
*A2P*_1_2&3	-	-	-	-	9.2500	0.5318
*V2P*_1_1&2	7.6963	0.8713	-	-	9.5182	0.4564
*V2P*_1_1&3	-	-	-	-	8.1335	0.3884
*V2P*_1_2&3	-	-	-	-	9.9052	0.4807
*FST*_1_1&2	0.2355	0.7477	-	-	0.0673	0.7866
*FST*_1_1&3	-	-	-	-	0.0492	0.1671
*FST*_1_2&3	-	-	-	-	0.0719	0.7014
$(\delta\mu)^2$_1_1&2	3.7333	0.7840	-	-	1.6581	0.5962
$(\delta\mu)^2$_1_1&3	-	-	-	-	3.0401	0.6779
$(\delta\mu)^2$_1_2&3	-	-	-	-	1.2961	0.4545



Appendix 3.9 The PCA graph of the BM using the optimum set of summary statistics shows the observed dataset (large yellow dot) is positioned within the simulated datasets (small coloured dots) suggesting that the priors fit the model.

Appendix 3.10a Posterior probabilities (and CI) of the best scenarios from the BM using the direct and logistic approaches.

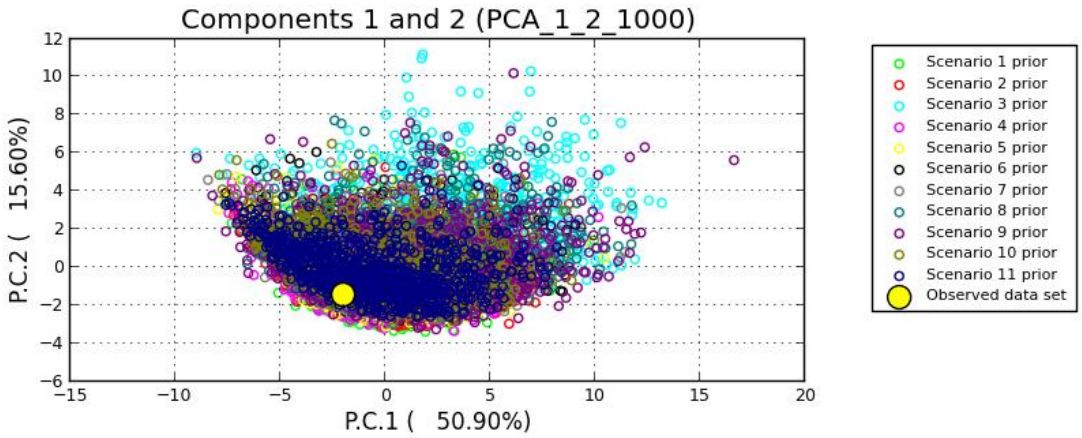
BM Scenario 4			
Direct approach		Logistic approach	
100	0.6300 (0.3308, 0.9292)	4000	0.8476 (0.8206, 0.8746)
200	0.5750 (0.2686, 0.8814)	8000	0.8427 (0.8191, 0.8663)
300	0.5400 (0.2311, 0.8489)	12000	0.8292 (0.8055, 0.8529)
400	0.5400 (0.2311, 0.8489)	16000	0.8087 (0.7838, 0.8336)
500	0.5500 (0.2417, 0.8409)	20000	0.7869 (0.7604, 0.8133)
600	0.5317 (0.2224, 0.8165)	24000	0.7626 (0.7344, 0.7908)
700	0.5314 (0.2221, 0.8407)	28000	0.7392 (0.7094, 0.7690)
800	0.5363 (0.2272, 0.8453)	32000	0.7185 (0.6874, 0.7496)
900	0.5433 (0.2346, 0.8521)	36000	0.7010 (0.6690, 0.7330)
1000	0.5500 (0.2417, 0.8583)	40000	0.6847 (0.6521, 0.7174)

Appendix 3.10b Posterior probabilities (and CI) of the best scenarios from the EMM using the direct approaches.

EMM Scenario 9	
Direct approach	
100	0.4000 (0.0964, 0.7036)
200	0.3550 (0.0584, 0.6516)
300	0.3500 (0.0544, 0.6456)
400	0.3300 (0.0286, 0.6214)
500	0.3280 (0.0370, 0.6190)
600	0.3317 (0.0399, 0.6235)
700	0.3443 (0.0498, 0.6388)
800	0.3412 (0.0474, 0.6351)
900	0.3456 (0.0508, 0.6403)
1000	0.3410 (0.0472, 0.6348)

Appendix 3.11 Parameter estimates of the bottleneck model scenario 4: *N1* – effective population size of *T. sibirica*; *Na* & *Naiii* – effective population size of ancestral populations; *t#* – time of event in generations; *Âµmic* – mean SSR mutation rate; *pmic* – Mean *P* (parameter of geometric distribution); *snimic* – mean single nucleotide insertion (SNI).

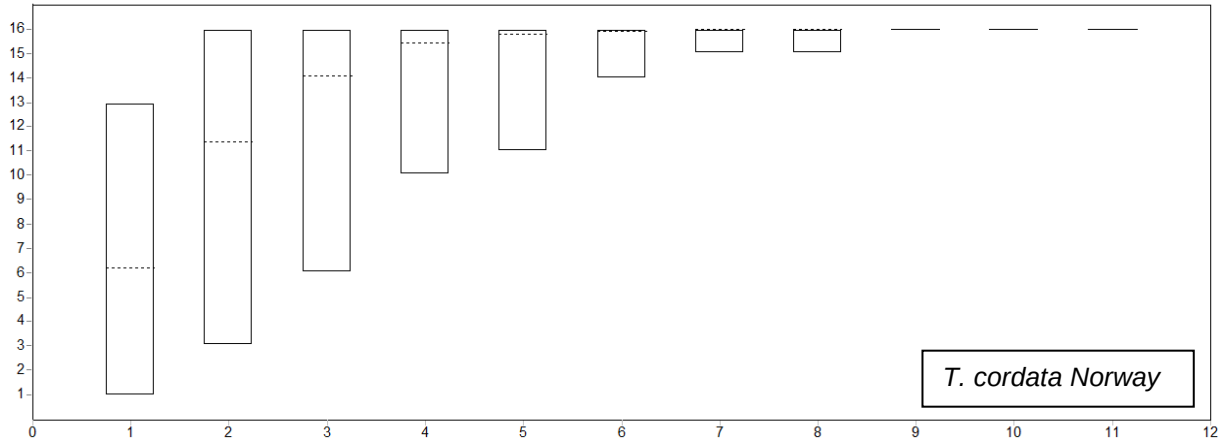
SDM model				Quartile					
Parameter	mean	median	mode	2.5%	5%	25%	75%	95%	97.5%
N1	8450	7470	2860	1130	1510	3890	12500	18100	19000
t2	255	204	73.3	25.5	34.4	97.8	381	629	677
t3	990	776	315	79.2	132	400	1370	2530	3020
Na	29500	21200	3710	1680	2340	8450	44700	83000	90600
Naiii	30000	30500	44300	7030	9630	20400	40400	48100	49000
Âµmic	2.66E-04	2.02E-04	1.00E-04	1.02E-04	1.06E-04	1.40E-04	3.27E-04	6.75E-04	7.85E-04
pmic	0.245	0.259	0.300	0.125	0.141	0.213	0.288	0.300	0.300
snimic	4.69E-06	4.5E-06	2.84E-06	2.37E-07	4.66E-07	2.26E-06	7.09E-06	9.39E-06	9.67E-06

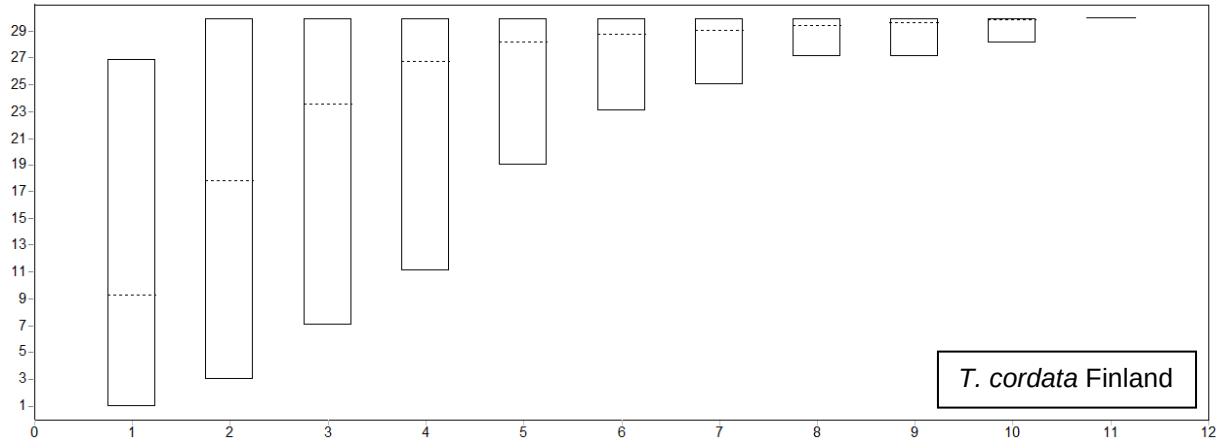


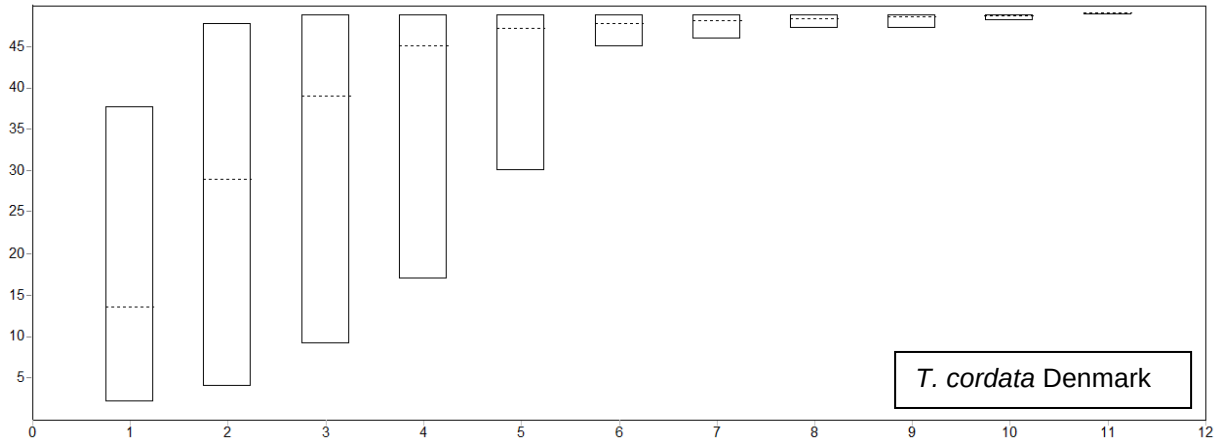
Appendix 3.12 The PCA graph of the EMM using the optimum set of summary statistics shows the observed dataset (large yellow dot) is positioned within the simulated datasets (small coloured dots) suggesting that the priors fit the model.

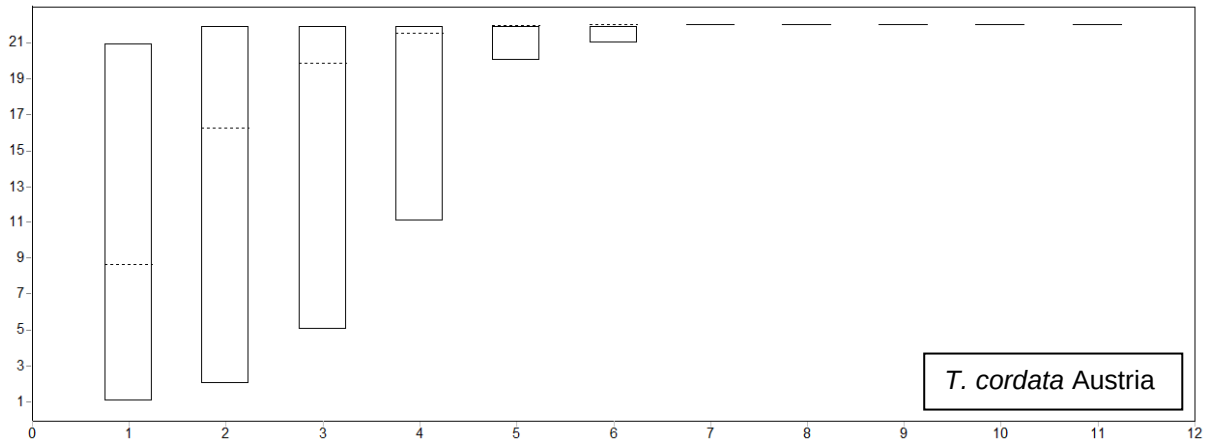
Appendix 3.13 Parameter estimates of the expansion/migration model: *N1* – effective population size of *T. cordata* Siberia; *N2* - effective population size of *T. cordata* Poland; *N3* - effective population size of *T. cordata* Austria; *N4* - effective population size of the theoretical *T. cordata* population; *Na* – effective population size of ancestral population; *td* – time of expansion/migration in generations; *Âµmic* – mean SSR mutation rate; *pmic* – Mean *P* (parameter of geometric distribution); *snimic* – mean single nucleotide insertion (SNI).

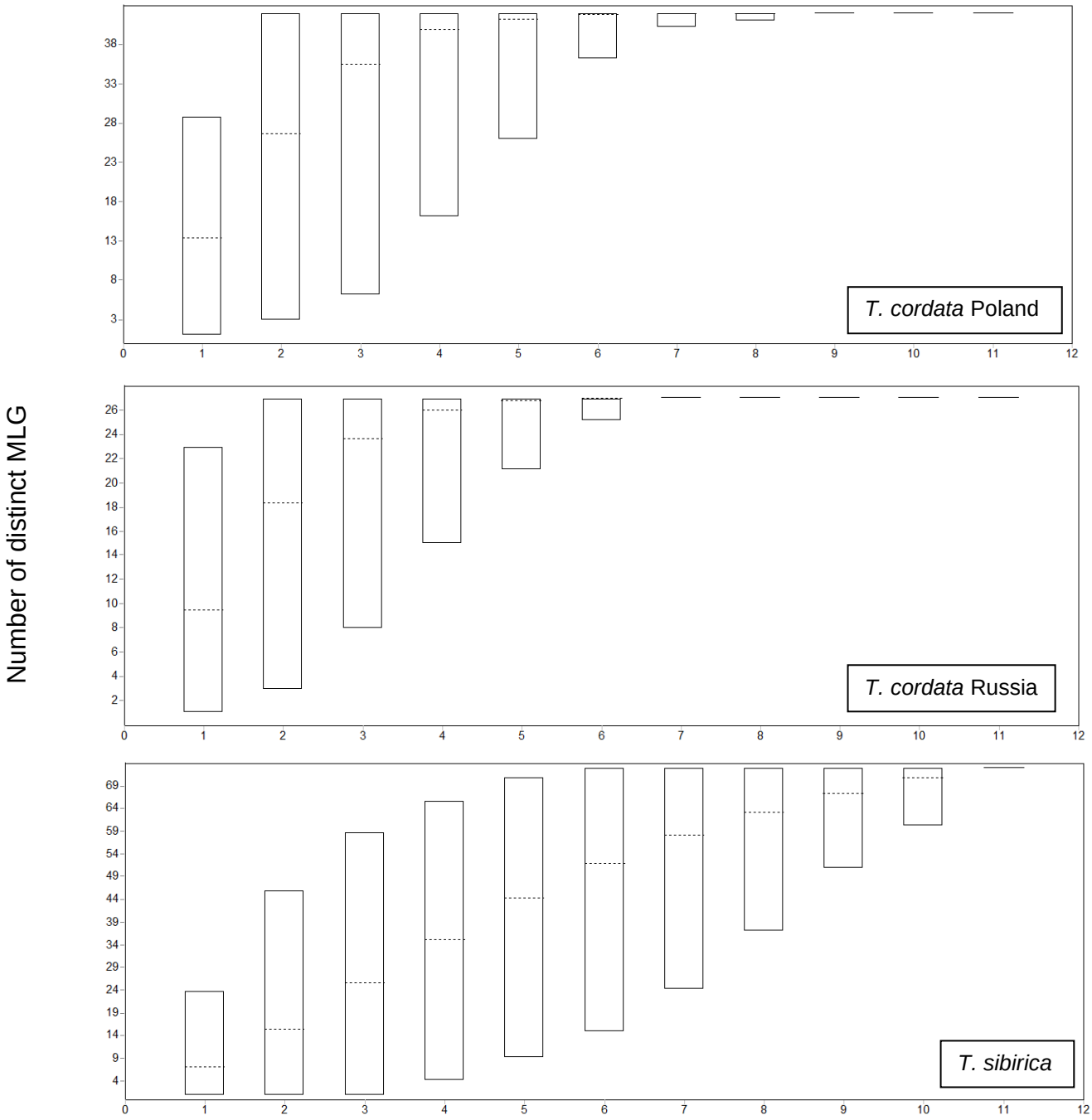
SDM model				Quartile					
Parameter	mean	median	mode	2.5%	5%	25%	75%	95%	97.5%
N1	3430	3210	2980	1140	1360	2290	4290	6260	7110
N2	6930	7150	7550	2900	3540	5600	8450	9640	9810
N3	3580	3090	1700	706	891	1890	4830	8150	8950
N4	5000	4980	4480	257	517	2570	7440	9480	9740
t2	571	526	418	169	210	373	724	1080	1220
t3	7140	7520	9370	2460	3240	5730	8880	9780	9890
Na	21200	18600	9650	3010	4090	10300	30900	45400	47700
Âµmic	3.44E-04	2.89E-04	1.77E-04	1.15E-04	1.25E-04	1.92E-04	4.44E-04	7.63E-04	8.50E-04
pmic	0.246	0.259	0.300	0.13	0.145	0.215	0.288	0.300	0.300
snimic	6.08E-06	6.51E-06	1.00E-05	5.3E-07	9.89E-07	3.92E-06	8.49E-06	9.9E-06	1.00E-05


T. cordata Norway


T. cordata Finland


T. cordata Denmark


T. cordata Austria



Appendix 4.1 Box plot of *Tilia cordata* Norway, *T. cordata* Finland, *T. cordata* Denmark, *T. cordata Poland*, and *T. cordata* Russia showing distinct number of MLG and number of loci required to describe the genotypic resolution.

Appendix 4.2 P_{gen} F_{IS} and P_{sex} F_{IS} values at the first reencounter, for each MLG.

Tilia platyphyllos UK

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	1.83E-15	1		25	7.19E-17	1	
2	2.15E-15	1		26	9.96E-14	1	
3	7.17E-17	1		27	3.88E-15	1	
4	1.46E-14	1		28	7.62E-17	1	
5	8.11E-18	1		29	5.75E-23	1	
6	1.72E-17	1		30	5.42E-14	2	8.67E-12
7	2.57E-14	1		31	1.29E-16	1	
8	1.76E-16	1		32	7.86E-16	1	
9	2.90E-16	1		33	2.19E-16	1	
10	1.04E-17	2	1.66E-15	34	8.94E-17	1	
11	4.61E-15	2	7.38E-13	35	6.25E-15	1	
12	8.19E-16	1		36	6.09E-14	1	
13	8.62E-14	2	1.38E-11	37	3.98E-16	1	
14	2.42E-14	2	3.88E-12	38	2.39E-16	1	
15	1.28E-13	1		39	4.48E-15	1	
16	7.77E-16	1		40	8.90E-15	1	
17	6.26E-23	3	1.00E-20	41	9.58E-18	1	
18	9.42E-15	1		42	3.13E-16	1	
19	4.77E-18	1		43	5.32E-15	1	
20	1.05E-13	1		44	7.33E-19	1	
21	2.10E-17	1		45	4.57E-17	1	
22	1.45E-12	1		46	6.16E-15	3	9.85E-13
23	1.98E-15	1		47	6.98E-13	1	
24	3.39E-24	1		48	7.44E-17	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
49	5.65E-16	1		75	2.40E-16	1	
50	1.01E-13	1		76	1.44E-15	2	2.31E-13
51	4.96E-14	1		77	5.24E-18	1	
52	3.35E-17	1		78	1.26E-22	1	
53	9.68E-16	3	1.55E-13	79	7.46E-24	1	
54	5.05E-17	1		80	8.32E-18	1	
55	1.42E-16	2	2.27E-14	81	7.95E-19	1	
56	4.55E-16	1		82	2.54E-18	1	
57	1.34E-15	1		83	2.71E-17	1	
58	7.52E-16	2	1.20E-13	84	1.52E-15	1	
59	2.67E-12	1		85	1.46E-17	3	2.34E-15
60	3.84E-14	1		86	6.42E-16	1	
61	3.33E-15	1		87	1.53E-16	1	
62	5.37E-17	1		88	1.21E-17	2	1.94E-15
63	7.30E-15	1		89	1.34E-15	1	
64	1.58E-15	1		90	3.76E-15	1	
65	1.04E-15	1		91	1.17E-14	1	
66	8.98E-17	1		92	1.44E-16	1	
67	1.67E-15	1		93	7.48E-16	1	
68	1.37E-15	1		94	2.54E-19	1	
69	3.02E-16	2	4.83E-14	95	2.09E-14	1	
70	2.11E-15	1		96	3.80E-13	1	
71	5.90E-15	3	9.43E-13	97	2.11E-15	1	
72	2.12E-17	1		98	6.66E-20	3	1.07E-17
73	2.11E-16	1		99	3.37E-13	1	
74	3.55E-19	1		100	6.98E-14	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
101	1.97E-18	1		127	5.68E-17	1	
102	1.61E-14	1		128	1.00E-18	1	
103	2.33E-15	1		129	1.13E-14	1	
104	6.82E-14	1		130	4.68E-13	1	
105	7.48E-13	1		118	7.96E-24	1	
106	1.21E-16	1		119	3.78E-16	1	
107	1.75E-17	1		120	9.97E-19	1	
108	1.85E-15	1		121	2.02E-21	1	
109	4.60E-16	1		122	2.43E-18	1	
110	1.95E-14	1		123	4.39E-15	1	
111	1.18E-14	1		124	1.48E-17	1	
112	2.31E-14	1		125	2.15E-15	1	
113	5.53E-14	1		126	1.81E-16	1	
114	7.78E-14	1		127	5.68E-17	1	
115	1.09E-17	1		128	1.00E-18	1	
116	7.97E-17	1		129	1.13E-14	1	
117	6.68E-16	2	1.07E-13	130	4.68E-13	1	
118	7.96E-24	1		131	1.83E-15	1	
119	3.78E-16	1		132	3.22E-17	1	
120	9.97E-19	1		133	1.43E-14	1	
121	2.02E-21	1		134	1.22E-14	1	
122	2.43E-18	1		135	1.86E-16	1	
123	4.39E-15	1		136	1.50E-14	1	
124	1.48E-17	1		137	1.02E-15	1	
125	2.15E-15	1		133	1.43E-14	1	
126	1.81E-16	1		134	1.22E-14	1	

T. platyphyllos FR

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	2.41E-19	1		25	1.70E-17	1	
2	8.61E-17	1		26	3.58E-16	1	
3	5.70E-19	1		27	2.46E-17	1	
4	7.05E-16	1		28	2.84E-14	1	
5	9.57E-17	1		29	1.62E-17	1	
6	4.56E-21	1		30	6.45E-16	1	
7	5.91E-14	1		31	5.19E-22	1	
8	1.57E-16	1		32	7.82E-16	1	
9	6.04E-14	1		33	1.33E-15	1	
10	3.42E-20	1		34	4.03E-16	1	
11	1.61E-16	1		35	1.07E-13	1	
12	7.40E-19	1		36	4.31E-17	1	
13	8.85E-16	1		37	1.21E-20	1	
14	5.54E-18	1		38	1.10E-17	1	
15	2.98E-18	1		39	1.57E-14	1	
16	3.09E-21	1		40	3.68E-18	1	
17	6.14E-16	1		41	6.31E-14	1	
18	1.15E-15	1		42	1.48E-17	1	
19	1.58E-17	1		43	2.73E-17	1	
20	2.18E-17	1		44	1.95E-16	1	
21	1.33E-16	1		45	2.59E-14	1	
22	8.13E-19	1		46	5.31E-18	1	
23	1.28E-17	1		47	1.60E-17	1	
24	4.78E-14	1		48	3.41E-18	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
49	2.71E-18	1		51	5.57E-18	1	
50	5.87E-19	1		52	1.42E-14	1	

T. platyphyllos DK

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	3.16E-10	1		6	4.16E-09	1	
2	2.52E-09	1		7	1.22E-10	1	
3	5.31E-08	1		8	2.97E-12	1	
4	1.52E-07	3	1.67E-06	9	1.27E-10	1	
5	2.18E-08	1					

T. platyphyllos AU

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	4.10E-19	1		12	2.36E-16	1	
2	2.52E-16	1		13	1.06E-14	1	
3	1.98E-17	1		14	1.39E-15	1	
4	7.88E-17	1		15	3.43E-16	1	
5	1.72E-16	2	9.66E-15	16	3.93E-16	1	
6	1.55E-16	1		17	5.07E-22	1	
7	3.76E-18	1		18	2.12E-20	1	
8	8.25E-16	1		19	8.96E-19	1	
9	1.65E-18	1		20	1.17E-17	1	
10	3.18E-19	1		21	1.04E-15	1	
11	1.18E-17	1		22	9.42E-20	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
23	4.76E-18	1		38	3.91E-15	1	
24	1.93E-18	1		39	1.71E-17	1	
25	1.40E-17	1		40	3.82E-15	1	
26	2.98E-18	1		41	2.60E-16	1	
27	1.76E-16	1		42	1.02E-14	1	
28	2.32E-16	1		43	1.15E-13	1	
29	4.98E-18	1		44	7.40E-15	1	
30	4.21E-17	2	2.36E-15	45	3.55E-15	1	
31	4.62E-21	2	2.59E-19	46	4.55E-16	1	
32	5.97E-17	1		47	1.84E-19	1	
33	7.13E-15	1		48	2.03E-18	1	
34	4.08E-15	1		49	1.06E-16	1	
35	1.31E-18	1		50	1.04E-18	1	
36	1.51E-18	1		51	1.52E-18	1	
37	1.29E-19	1		52	4.49E-16	1	

T. platyphyllos GE

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	1.03E-14	1		8	5.92E-15	1	
2	1.74E-13	1		9	3.65E-15	1	
3	6.17E-14	1		10	8.57E-18	1	
4	3.41E-15	1		11	1.26E-14	1	
5	3.42E-17	1		12	1.48E-13	1	
6	2.14E-13	1		13	2.36E-15	1	
7	2.20E-14	1		14	1.08E-15	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
15	5.86E-14	1		24	3.72E-15	1	
16	9.10E-16	1		25	4.85E-16	1	
17	7.24E-13	1		26	3.21E-16	1	
18	7.05E-14	1		27	8.00E-14	1	
19	1.51E-12	1		28	8.12E-13	1	
20	8.12E-13	1		29	1.11E-16	1	
21	4.92E-17	1		30	1.21E-12	1	
22	4.36E-17	1		31	5.06E-16	1	
23	4.63E-16	1					

T. x europaea UK

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	8.82E-14	7	4.76E-12	13	2.80E-18	1	
2	1.81E-18	1		14	1.63E-16	1	
3	1.44E-14	5	7.77E-13	15	1.22E-11	14	6.57E-10
4	1.03E-15	1		16	8.88E-15	4	4.80E-13
5	4.81E-15	1		17	2.37E-16	1	
6	-5.48E-20	1		18	6.72E-14	1	
7	1.17E-17	1		19	1.05E-14	1	
8	1.32E-12	5	7.14E-11	20	1.14E-18	1	
9	5.10E-13	1		21	4.61E-17	1	
10	1.35E-16	1		22	1.43E-14	1	
11	4.85E-19	1		23	1.50E-14	1	
12	1.44E-16	1		24	3.07E-18	1	

T. cordata UK

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	3.53E-12	1		26	4.93E-12	1	
2	3.01E-09	1		27	3.45E-10	1	
3	1.07E-09	3	2.80E-07	28	1.52E-10	1	
4	6.65E-12	1		29	6.48E-10	1	
5	3.89E-12	1		30	6.51E-12	1	
6	1.41E-11	1		31	3.77E-11	1	
7	2.23E-10	1		32	5.09E-11	1	
8	7.34E-11	1		33	8.87E-13	1	
9	9.91E-11	2	2.59E-08	34	4.02E-12	1	
10	1.57E-10	1		35	1.95E-11	1	
11	3.01E-09	1		36	1.90E-13	1	
12	4.66E-12	1		37	1.54E-12	1	
13	6.93E-12	1		38	7.27E-11	1	
14	1.81E-11	2	4.73E-09	39	1.89E-11	1	
15	9.93E-12	1		40	3.54E-11	1	
16	2.32E-11	1		41	1.86E-12	1	
17	1.03E-10	1		42	8.40E-14	9	2.19E-11
18	1.26E-09	1		43	2.14E-11	1	
19	9.79E-13	1		44	1.20E-08	1	
20	8.81E-12	1		45	4.66E-09	1	
21	3.33E-12	1		46	4.12E-09	1	
22	3.14E-11	1		47	1.15E-08	1	
23	2.91E-10	3	7.59E-08	48	4.34E-11	1	
24	2.19E-10	1		49	8.61E-14	1	
25	4.34E-12	1		50	7.95E-12	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
51	2.39E-11	1		77	1.02E-12	1	
52	5.58E-09	1		78	4.27E-13	1	
53	6.07E-10	1		79	8.33E-09	1	
54	1.36E-11	1		80	2.06E-10	1	
55	3.38E-09	2	8.82E-07	81	1.34E-09	1	
56	2.61E-11	2	6.82E-09	82	1.56E-10	1	
57	4.68E-12	1		83	1.46E-11	1	
58	3.12E-11	1		84	4.05E-11	1	
59	7.13E-11	1		85	8.36E-13	1	
60	9.93E-12	1		86	1.39E-08	2	3.62E-06
61	2.94E-11	1		87	2.00E-10	1	
62	3.69E-09	1		88	3.92E-14	1	
63	1.07E-09	1		89	2.39E-10	1	
64	2.16E-08	3	5.64E-06	90	1.21E-10	1	
65	8.31E-10	1		91	5.64E-09	1	
66	9.80E-10	1		92	1.80E-14	1	
67	2.05E-11	1		93	6.35E-11	2	1.66E-08
68	2.23E-09	1		94	8.79E-12	2	2.29E-09
69	7.92E-08	1		95	3.31E-13	2	8.63E-11
70	7.95E-09	3	2.07E-06	96	2.58E-12	1	
71	3.30E-11	1		97	1.68E-13	1	
72	2.04E-13	1		98	1.95E-12	1	
73	6.80E-09	1		99	1.38E-11	1	
74	1.40E-09	1		100	1.48E-12	1	
75	1.29E-12	1		101	1.24E-09	1	
76	5.81E-10	1		102	1.14E-09	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
103	1.54E-13	1		128	1.80E-09	1	
104	1.50E-13	1		129	1.45E-11	1	
105	3.76E-12	2	9.81E-10	130	1.42E-10	1	
106	9.25E-12	1		131	3.26E-09	2	8.51E-07
107	1.35E-12	1		132	2.15E-10	1	
108	1.89E-10	1		133	1.54E-09	1	
109	2.64E-10	1		134	1.06E-09	1	
110	2.74E-10	1		135	1.03E-08	1	
111	1.85E-09	1		136	2.84E-12	1	
112	1.68E-08	1		137	7.97E-10	1	
113	2.80E-08	1		138	1.01E-10	2	2.65E-08
114	1.21E-08	1		139	7.15E-10	1	
115	4.05E-10	1		140	3.42E-14	1	
116	5.60E-09	1		141	5.80E-09	1	
117	4.42E-09	1		142	1.19E-12	1	
118	9.91E-10	1		143	9.78E-11	1	
119	2.49E-09	1		144	6.27E-11	1	
120	5.40E-13	1		145	5.83E-09	1	
121	1.12E-10	1		146	5.20E-10	1	
122	3.13E-09	1		147	8.34E-11	1	
123	5.11E-11	1		148	1.66E-09	1	
124	2.71E-11	1		149	5.28E-10	1	
125	2.71E-11	1		150	3.69E-10	5	9.62E-08
126	9.95E-12	1		151	1.10E-11	1	
127	2.12E-14	1		152	2.64E-10	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
153	1.24E-09	1		178	1.85E-11	1	
154	5.53E-10	1		179	4.17E-11	1	
155	1.72E-11	1		180	7.01E-13	1	
156	1.66E-14	1		181	4.24E-11	1	
157	2.60E-10	1		182	1.29E-12	2	3.38E-10
158	3.41E-09	2	8.89E-07	183	1.91E-12	1	
159	5.07E-15	1		184	3.29E-11	1	
160	9.36E-13	1		185	2.40E-11	1	
161	2.06E-11	1		186	2.34E-12	2	6.11E-10
162	4.15E-10	7	1.08E-07	187	5.42E-12	2	1.41E-09
163	1.85E-14	1		188	7.29E-15	1	
164	3.30E-15	2	8.60E-13	189	2.89E-10	1	
165	3.32E-10	1		190	3.26E-11	1	
166	9.95E-10	1		191	2.57E-12	1	
167	7.37E-10	1		192	1.57E-11	1	
168	1.12E-09	1		193	1.81E-12	1	
169	2.84E-12	1		194	9.00E-10	1	
170	2.34E-10	1		195	1.81E-12	1	
171	2.04E-10	2	5.33E-08	196	1.66E-10	1	
172	1.65E-11	1		197	3.70E-11	1	
173	3.73E-12	1		198	3.57E-10	1	
174	9.63E-13	1		199	2.11E-12	1	
175	8.27E-13	1		200	7.20E-10	1	
176	9.65E-13	1		201	1.07E-09	2	2.78E-07
177	2.99E-10	3	7.80E-08	202	1.11E-12	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
203	7.68E-14	1		207	2.88E-10	1	
204	3.10E-11	2	8.10E-09	208	4.26E-10	1	
205	6.16E-10	1		209	8.15E-11	2	2.13E-08
206	6.49E-10	4	1.69E-07	210	3.96E-13		

T. cordata FR

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	2.07E-09	1		15	2.56E-11	1	
2	2.21E-10	1		16	7.74E-12	1	
3	4.22E-10	1		17	2.83E-09	1	
4	1.07E-10	1		18	1.01E-10	1	
5	1.49E-07	1		19	4.64E-14	1	
6	4.51E-10	1		20	2.49E-09	1	
7	8.18E-10	1		21	1.27E-11	1	
8	5.03E-09	1		22	1.44E-09	1	
9	1.29E-09	2	3.62E-08	23	1.50E-10	1	
10	2.45E-10	1		24	5.97E-11	1	
11	9.87E-11	1		25	9.13E-10	1	
12	3.15E-09	1		26	7.15E-09	1	
13	1.11E-14	1		27	1.91E-12	1	
14	7.16E-10	1					

T. cordata DK

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	4.10E-10	1		26	2.84E-12	1	
2	4.31E-11	1		27	2.32E-10	1	
3	8.19E-11	1		28	1.97E-09	1	
4	7.92E-11	1		29	1.21E-09	1	
5	1.40E-10	1		30	2.34E-10	1	
6	8.99E-11	1		31	2.79E-11	1	
7	2.38E-14	1		32	3.00E-11	1	
8	7.60E-15	1		33	1.19E-12	1	
9	2.27E-10	1		34	1.18E-09	1	
10	1.67E-10	1		35	2.00E-10	1	
11	7.25E-10	1		36	3.74E-12	1	
12	3.67E-12	1		37	5.74E-12	3	3.45E-10
13	2.68E-08	1		38	4.16E-12	1	
14	9.04E-12	1		39	1.62E-13	1	
15	3.90E-11	1		40	2.54E-11	1	
16	6.37E-13	1		41	1.14E-11	2	6.85E-10
17	8.24E-11	1		42	3.91E-12	1	
18	4.18E-10	1		43	2.74E-10	1	
19	3.93E-09	1		44	1.79E-11	1	
20	2.33E-09	1		45	7.99E-11	2	4.80E-09
21	8.62E-10	1		46	3.02E-12	1	
22	3.43E-11	1		47	8.17E-13	1	
23	6.79E-11	2	4.07E-09	48	9.95E-14	1	
24	4.20E-10	1		49	8.12E-13	1	
25	5.44E-10	7	3.26E-08				

T. cordata AU

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	3.65E-11	1		26	9.04E-15	1	
2	3.45E-12	1		27	5.75E-13	1	
3	1.67E-13	1		28	4.34E-11	1	
4	1.99E-10	1		29	1.78E-11	1	
5	9.19E-13	1		30	1.29E-11	1	
6	1.53E-12	1		31	7.24E-13	1	
7	5.60E-11	1		32	8.97E-11	1	
8	1.23E-13	1		33	2.55E-13	1	
9	4.32E-15	1		34	3.06E-10	1	
10	1.75E-12	1		35	8.64E-12	1	
11	1.35E-12	1		36	4.44E-11	1	
12	3.82E-11	1		37	4.82E-11	1	
13	5.92E-16	1		38	1.83E-12	1	
14	1.16E-14	1		39	1.98E-11	1	
15	1.46E-11	1		40	1.72E-11	1	
16	1.52E-11	1		41	1.69E-11	1	
17	4.56E-13	1		42	9.12E-12	1	
18	2.16E-10	1		43	1.04E-10	1	
19	3.97E-10	1		44	5.43E-14	1	
20	8.01E-11	1		45	7.70E-12	1	
21	1.08E-10	1		46	6.58E-16	1	
22	3.43E-13	1		47	8.42E-13	1	
23	4.75E-13	1		48	1.88E-12	1	
24	1.44E-10	1		49	9.80E-11	1	
25	1.84E-11	1		50	2.41E-11	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
51	6.44E-14	1		68	2.45E-09	1	
52	8.04E-11	1		69	8.09E-12	1	
53	3.13E-15	1		70	1.67E-10	1	
54	2.05E-12	1		71	1.84E-11	1	
55	1.40E-15	1		72	5.07E-14	1	
56	4.47E-15	1		73	1.03E-11	1	
57	8.06E-11	1		74	2.34E-11	1	
58	1.27E-09	1		75	5.35E-12	1	
59	3.18E-12	1		76	7.62E-11	1	
60	7.47E-11	1		77	1.14E-13	1	
61	1.78E-12	1		78	4.66E-16	1	
62	7.69E-14	1		79	1.31E-11	1	
63	8.58E-11	1		80	1.19E-10	1	
64	3.00E-13	1		81	8.49E-16	1	
65	4.24E-14	1		82	1.64E-12	1	
66	8.09E-12	1		83	1.98E-11	1	
67	3.21E-13	1					

T. cordata GE

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	3.11E-10	1		6	1.03E-07	1	
2	3.27E-09	1		7	2.25E-09	1	
3	5.05E-08	1		8	7.21E-08	1	
4	4.87E-08	1		9	5.36E-09	1	
5	2.70E-08	1		10	1.25E-11	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
11	6.41E-09	1		16	8.88E-10	1	
12	3.23E-12	1		17	2.61E-08	1	
13	4.82E-10	1		18	1.29E-11	1	
14	3.46E-12	1		19	5.96E-10	1	
15	4.71E-09	1		20	2.02E-11	1	

T. cordata NO

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	1.18E-09	1		9	1.28E-06	1	
2	6.43E-09	1		10	1.46E-07	1	
3	1.77E-08	1		11	7.42E-07	1	
4	3.27E-09	1		12	4.19E-08	1	
5	6.37E-10	1		13	2.65E-08	1	
6	1.68E-07	1		14	5.62E-10	1	
7	1.04E-09	1		15	2.84E-07	1	
8	1.69E-10	1		16	2.18E-07	1	

T. cordata FI

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	1.21E-09	1		7	2.16E-07	1	
2	1.48E-09	1		8	5.80E-08	1	
3	5.10E-11	2	2.04E-09	9	6.87E-12	2	2.75E-10
4	1.36E-07	1		10	1.17E-10	3	4.70E-09
5	6.74E-08	1		11	2.69E-10	1	
6	5.39E-08	1		12	3.06E-07	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
13	2.18E-08	1		22	1.10E-09	1	
14	1.06E-08	1		23	6.13E-08	1	
15	6.52E-08	6	2.61E-06	24	5.08E-11	2	2.03E-09
16	1.23E-10	1		25	1.22E-11	1	
17	1.39E-08	1		26	6.10E-08	1	
18	6.56E-09	1		27	2.10E-11	1	
19	8.22E-09	1		28	2.85E-10	1	
20	3.32E-11	1		29	2.64E-13	1	
21	8.59E-09	1		30	1.10E-12	1	

T. cordata PL

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	6.10E-11	1		14	3.77E-13	1	
2	5.43E-11	1		15	5.52E-11	1	
3	2.92E-10	1		16	1.23E-10	1	
4	5.98E-12	1		17	2.04E-09	1	
5	1.75E-11	1		18	2.78E-12	1	
6	7.80E-11	1		19	2.00E-10	1	
7	6.75E-10	1		20	2.24E-09	1	
8	3.65E-12	1		21	2.88E-11	1	
9	5.43E-11	1		22	2.18E-10	1	
10	3.48E-11	1		23	5.31E-10	1	
11	2.52E-09	1		24	4.94E-13	1	
12	1.40E-12	1		25	5.90E-09	1	
13	8.75E-09	1		26	2.99E-12	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
27	1.26E-09	1		35	2.22E-10	1	
28	2.48E-10	1		36	4.26E-10	1	
29	2.88E-10	1		37	7.87E-12	1	
30	3.47E-11	1		38	1.38E-10	1	
31	1.67E-09	1		39	1.89E-13	1	
32	2.39E-10	1		40	6.63E-13	1	
33	3.34E-11	1		41	1.61E-10	1	
34	1.45E-11	1		42	5.23E-12	1	

T. cordata RU

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	9.36E-14	1		15	8.17E-10	1	
2	1.44E-09	1		16	4.72E-10	1	
3	1.57E-11	1		17	1.02E-12	1	
4	2.35E-11	1		18	1.04E-09	2	3.43E-08
5	1.28E-12	1		19	1.32E-11	1	
6	1.47E-09	3	4.87E-08	20	7.50E-10	3	2.48E-08
7	1.36E-11	1		21	2.06E-11	1	
8	1.07E-08	1		22	4.02E-10	1	
9	2.03E-09	1		23	1.71E-09	1	
10	5.66E-10	1		24	1.87E-10	1	
11	3.30E-10	1		25	1.80E-09	1	
12	7.09E-08	1		26	5.13E-10	1	
13	1.63E-09	1		27	6.92E-10	2	2.28E-08
14	4.34E-11	1					

T. sibirica RU

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
1	8.03E-08	1		26	9.75E-06	1	
2	1.05E-05	1		27	1.55E-08	1	
3	8.38E-09	1		28	1.37E-06	2	1.55E-04
4	1.84E-06	1		29	8.06E-08	1	
5	6.53E-06	2	7.38E-04	30	2.67E-07	2	3.02E-05
6	1.50E-06	1		31	5.42E-05	1	
7	1.32E-05	3	0.001492	32	5.70E-06	1	
8	1.10E-04	1		33	1.41E-07	1	
9	8.65E-07	6	9.77E-05	34	3.45E-07	1	
10	1.73E-07	1		35	1.41E-04	1	
11	2.38E-05	9	0.002681	36	1.81E-06	1	
12	1.23E-07	1		37	1.07E-04	1	
13	2.58E-05	1		38	1.29E-04	1	
14	1.07E-05	1		39	4.77E-04	1	
15	5.88E-07	1		40	1.26E-05	1	
16	4.52E-07	1		41	2.66E-08	1	
17	1.47E-06	1		42	1.78E-05	1	
18	1.10E-08	2	1.25E-06	43	1.86E-05	1	
19	3.03E-07	7	3.42E-05	44	1.64E-05	3	0.00185
20	3.52E-09	1		45	2.23E-07	1	
21	4.14E-05	1		46	4.87E-06	1	
22	2.14E-05	1		47	8.95E-08	1	
23	2.13E-05	1		48	7.01E-06	1	
24	2.97E-06	1		49	3.87E-05	1	
25	2.09E-05	1		50	1.41E-06	1	

MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)	MLG	P_{gen} (*f*)	No. of ramets	P_{sex} (*f*)
51	7.04E-07	1		63	5.35E-05	1	
52	9.15E-06	1		64	5.66E-07	1	
53	6.68E-06	2	7.55E-04	65	5.92E-06	3	2.21E-07
54	6.19E-04	1		66	2.96E-06	11	6.68E-04
55	1.57E-04	1		67	9.86E-08	1	
56	1.02E-04	1		68	1.89E-07	1	
57	2.63E-04	1		69	9.06E-05	1	
58	0.002095	1		70	7.09E-07	1	
59	3.50E-05	1		71	5.81E-05	1	
60	1.47E-07	1		72	0.000134	1	
61	3.12E-05	1		73	1.13E-05	1	
62	2.23E-07	1					

Appendix 4.3 Estimates of clonal diversity in 34 populations of *Tilia.* Number of samples in each population (*N*), number of genotypes (G), the Proportion distinguishable (*Pd*), genotypic richness (*R*), the number of putative somatic mutations, Simpson's index for genotypic diversity (*D**).

T. platyphyllos			*Pd*	*R*		
UK	*N*	*G*	(*G/N*)	(*G-1/N-1*)	*Somatic*	*D**
AS	12	8	0.667	0.636	0	0.909
BH	17	13	0.764	0.750	0	0.956
AW	15	14	0.933	0.928	1	0.990
LP	16	13	0.813	0.800	0	0.967
HW	24	23	0.958	0.956	0	0.996
CW	15	12	0.800	0.785	0	0.971
BR	11	7	0.636	0.600	1	0.909
HD	25	24	0.960	0.958	0	0.997
SP	25	23	0.920	0.916	1	0.993
Mean	17.78	15.22	0.828	0.814	0.33	0.965

T. x europaea			*Pd*	*R*		
UK	*N*	*G*	(*G/N*)	(*G-1/N-1*)	*Somatic*	*D**
BB	6	2	0.333	0.200	1	0.333
BW	7	1	0.143	0.000	0	0.000
CW	5	5	1.000	1.000	0	1.000
KP	6	2	0.333	0.200	0	0.333
LP	6	3	0.500	0.400	0	0.600
SP	9	9	1.000	1.000	0	1.000
WW	15	2	0.133	0.071	0	0.133
Mean	7.71	3.43	0.491	0.410	0.14	0.486

T. cordata UK	*N*	*G*	*Pd* (*G/N*)	*R* (*G-1/N-1*)	*Somatic*	*D**
LP	10	8	0.800	0.778	0	0.933
BB	19	8	0.421	0.389	2	0.813
KP	21	15	0.714	0.700	0	0.962
CW	16	12	0.750	0.733	0	0.985
DB	17	16	0.941	0.938	1	0.993
SkW	14	14	1.000	1.000	0	1.000
EaH	16	16	1.000	1.000	0	1.000
BP	20	20	1.000	1.000	0	1.000
HG	18	18	1.000	1.000	0	1.000
CGW	40	30	0.750	0.744	7	0.982
ShW	20	20	1.000	1.000	0	1.000
RW	40	31	0.775	0.769	2	0.988
WM	10	2	0.200	0.111	0	0.200
Mean	20.08	16.15	0.796	0.782	0.92	0.912

T. cordata FL	*N*	*G*	*Pd* (*G/N*)	*R* (*G-1/N-1*)	*Somatic*	*D**
MU	20	14	0.700	0.684	0	0.916
SN	20	16	0.800	0.789	0	0.974
Mean	20	15.00	0.750	0.737	0	0.945

T. cordata DK	*N*	*G*	*Pd* (*G/N*)	*R* (*G-1/N-1*)	*Somatic*	*D**
AB	30	30	1.000	1.000	1	1.000
BO	30	19	0.633	0.621	1	0.938
Mean	30	24.50	0.817	0.811	1	0.969

T. cordata PL	*N*	*G*	*Pd* (*G/N*)	*R* (*G-1/N-1*)	*Somatic*	*D**
B69	17	17	1.000	1.000	0	1.000
B99	12	12	1.000	1.000	0	1.000
B40	13	13	1.000	1.000	0	1.000
Mean	14	14	1.000	1.000	0	1.000

T. cordata RU	*N*	*G*	*Pd* (*G/N*)	*R* (*G-1/N-1*)	*Somatic*	*D**
K20	20	14	0.700	0.684	0	0.958
K25	13	13	1.000	1.000	0	1.000
Mean	16.50	13.50	0.850	0.842	0	0.979

T. sibirica	*N*	*G*	*Pd* (*G/N*)	*R* (*G-1/N-1*)	*Somatic*	*D**
K12	20	19	0.950	0.947	3	0.995
K21	15	3	0.200	0.142	2	0.448
K22	19	6	0.316	0.278	5	0.701
K28	22	20	0.909	0.905	5	0.987
K29	21	19	0.905	0.900	5	0.990
K38	16	6	0.375	0.333	3	0.783
Mean	18.83	12.17	0.609	0.584	3.83	0.817

Appendix 4.4 Average diversity indices (and SE) for each population from UK, Poland, and Russia, excluding repeated genotypes; *N* - number of samples, *A*- number of alleles, *P* - Proportion of polymorphic loci, H_O - Observed heterozygosity, H_E - Nei's unbiased Expected heterozygosity, F_{IS} Inbreeding coefficient,
* - Significant at the 5% level

T. platyphyllos	*N*	*A*	*P*	H_O	H_E	F_{IS}
AS	8	3.64 (0.28)	1.00	0.74 (0.43)	0.65 (0.29)	-0.146
BH	13	5.82 (0.54)	1.00	0.78 (0.04)	0.74 (0.03)	-0.068
AW	14	5.18 (0.52)	1.00	0.71 (0.06)	0.69 (0.04)	-0.029
LP	13	5.27 (0.36)	1.00	0.70 (0.05)	0.69 (0.04)	-0.018
HW	23	6.36 (0.61)	1.00	0.70 (0.07)	0.72 (0.04)	0.024
CW	12	5.45 (0.48)	1.00	0.75 (0.04)	0.69 (0.04)	-0.091
SP	23	5.27 (0.70)	0.91	0.69 (0.09)	0.64 (0.08)	-0.082
BR	7	4.82 (0.38)	1.00	0.68 (0.07)	0.71 (0.04)	0.052
HD	24	7.09 (0.39)	1.00	0.75 (0.05)	0.74 (0.03)	-0.014
Mean	15.2	5.43 (0.18)	0.99 (0.01)	0.72 (0.02)	0.70 (0.01)	-0.041

T. x europaea	*N*	*A*	*P*	H_O	H_E	F_{IS}
WW	2	3.18 (0.12)	1.00	1.00 (0.00)	0.86 (0.02)	-0.257
LP	3	3.46 (0.28)	1.00	0.91 (0.07)	0.78 (0.03)	-0.224
BB	2	2.00 (0.14)	0.91	0.91 (0.09)	0.62 (0.06)	-0.905
KP	2	3.36 (0.24)	1.00	0.91 (0.09)	0.89 (0.04)	-0.026
CW	5	5.18 (0.44)	1.00	0.95 (0.03)	0.84 (0.03)	-0.149
SP	9	6.00 (0.59)	1.00	0.79 (0.06)	0.76 (0.04)	-0.037
Mean	3.83	3.87 (0.21)	0.99 (0.02)	0.91 (0.03)	0.79 (0.02)	-0.267

T. cordata UK	*N*	*A*	*P*	H_O	H_E	F_{IS}
LP	8	3.64 (0.62)	0.82	0.55 (0.09)	0.54 (0.09)	-0.004
BB	8	3.73 (0.80)	0.82	0.60 (0.11)	0.54 (0.09)	-0.136
KP	15	4.00 (0.67)	0.82	0.44 (0.09)	0.50 (0.09)	0.127
CW	12	3.55 (0.67)	0.82	0.53 (0.10)	0.53 (0.09)	-0.007
DB	16	5.00 (1.00)	0.91	0.55 (0.10)	0.53 (0.10)	-0.035
SkW	14	4.27 (0.69)	0.82	0.52 (0.10)	0.50 (0.10)	-0.036
EaH	16	5.27 (1.26)	0.82	0.55 (0.09)	0.57 (0.09)	0.039
BP	20	5.18 (1.23)	0.91	0.55 (0.09)	0.54 (0.09)	-0.002
HG	18	5.36 (1.28)	0.91	0.53 (0.08)	0.54 (0.09)	0.022
CGW	30	4.82 (1.08)	0.82	0.54 (0.11)	0.51 (0.10)	-0.050
ShW	20	5.45 (1.37)	0.82	0.60 (0.09)	0.57 (0.10)	-0.064
RW	31	5.27 (1.24)	0.91	0.56 (0.09)	0.56 (0.09)	0.001
WM	2	1.91 (0.25)	0.64	0.41 (0.11)	0.41 (0.11)	0.000
Mean	16.15	4.42 (0.28)	0.83 (0.02)	0.53 (0.03)	0.52 (0.03)	-0.011

T. cordata PL	*N*	*A*	*P*	H_O	H_E	F_{IS}
B69	17	6.00 (1.19)	0.91	0.55 (0.11)	0.52 (0.10)	-0.052
B99	12	5.00 (0.89)	0.91	0.57 (0.08)	0.55 (0.08)	-0.017
B40	13	5.00 (0.95)	0.91	0.61 (0.11)	0.57 (0.10)	-0.077
Mean	14.00	5.33 (0.58)	0.91	0.57 (0.06)	0.55 (0.05)	-0.049

T. cordata RU	*N*	*A*	*P*	H_O	H_E	F_{IS}
K20	14	5.09 (0.94)	0.91	0.51 (0.08)	0.58 (0.08)	0.133
K25	13	5.00 (1.04)	0.91	0.54 (0.09)	0.57 (0.09)	0.061
Mean	13.50	5.05 (0.69)	0.91	0.52 (0.06)	0.58 (0.06)	0.097

T. sibirica	*N*	*A*	*P*	H_O	H_E	F_{IS}
K12	19	2.91 (0.73)	0.64	0.36 (0.12)	0.33 (0.09)	-0.069
K21	3	1.45 (0.28)	0.27	0.27 (0.14)	0.19 (0.10)	-0.636
K22	6	2.00 (0.43)	0.46	0.27 (0.11)	0.25 (0.10)	-0.104
K28	20	2.91 (0.44)	0.73	0.20 (0.07)	0.30 (0.08)	0.364*
K29	19	2.73 (0.66)	0.55	0.19 (0.07)	0.25 (0.09)	0.266*
K38	6	2.27 (0.27)	0.73	0.49 (0.12)	0.39 (0.09)	-0.290
Mean	12.17	2.38 (0.21)	0.56 (0.07)	0.29 (0.04)	0.29 (0.04)	-0.078

Appendix 5.1 Detailed procedure of library construction and Illumina HiSeq sequencing (provided by High-Throughput Sequencing and Genotyping Unit).

Construction of Stranded RNAseq libraries

RNAseq libraries were constructed using the TruSeq Stranded RNA Sample Preparation Kit (Illumina San Diego, CA). Briefly, total RNA was quantitated by Qubit (Life Technologies, Grand Island, NY) and checked for integrity on a 1% eGel (Life Technologies). PolyA+RNA was selected from 1 µg of total RNA, then first-strand synthesis was synthesized with a random hexamer and SuperScript II (Life Technologies). Double stranded DNA was blunt-ended, 3'-end A-tailed and ligated to indexed adaptors. The adaptor-ligated double-stranded cDNA was amplified by PCR for 10 cycles with the Kapa HiFi polymerase (Kapa Biosystems, Woburn, MA). The final libraries were quantitated on Qubit and the average size determined on an Agilent bioanalyzer DNA7500 DNA chip (Agilent Technologies, Wilmington, DE) and diluted to 10nM final concentration. The 10nM dilution was further quantitated by qPCR on a BioRad CFX Connect Real-Time System (Bio-Rad Laboratories, Inc. CA).

Sequencing on an Illumina HiSeq2500

The 12 pooled libraries were loaded onto one lane of an 8-lane flowcell for cluster formation on the cBOT and then sequenced on an Illumina HiSeq2500 with version 4 sequencing reagents from one end of the molecules for a total read length of 100nt from that end. The run generated .bcl files which were converted into demultiplexed compressed fastq files using Casava 1.8.4 (Illumina, CA). A secondary pipeline decompressed the fastq files, generated plots with quality scores using FastX Tool Kit, removed perfect matches to reads that contain only adaptor and generated a report with the number of reads per sample/library. Demultiplexed fastq files were .tgz compressed and posted to a password-secured FTP site.

Appendix 5.2 Detailed procedure of the Trinity assembly (provided by the Bioinformatics Support Unit).

Cleaning and trimming of the raw Illumina reads

Adapter sequences specified by the sequencing company using the "Trimmomatic" program (http://www.usadellab.org/cms/?page=trimmomatic) version 0.33. Poor quality bases were also trimmed from the end of reads using a sliding window algorithm (minimum phred quality of 30 over 5 bases), again using Trimomatic. Reads corresponding to the same species were then pooled together prior to assembly with Trinity.

Trinity transcript assembly

Reads were assembled using the Trinity transcript assembly pipeline (version 2.06) Designed specifically for de-novo mRNA transcript assembly, Trinity consists of three main programs. Inchworm performs an initial assembly of the reads by using a "greedy kmer extension" algorithm. Initially the reads are processed to determine the unique "kmers" present (default length 25), after which the most abundant kmer is selected as a "seed". The seed is extended by finding the most abundant remaining kmer with a k-1 overlap. The process is repeated until no further kmers can form a k-1 overlap with the extended seed. At this point, an inchworm "contig" has been found, is saved, and a new seed is started (the next most abundant kmer still unused in the pool). This is continued until all kmers have been processed.

The Chrysalis program takes the unique Inchworm contigs and clusters them together to form collections of sequences that likely formed from the same "gene". The output of this step is a collection of cluster graphs which are recombined with the original reads and processed by "Butterfly" to produce a set of full length transcripts, grouped by "gene".

The Trinity pipeline itself is executed from the Linux command line, with the only required parameters being the name of the files containing the reads. However for the present study two extra options were specified. Firstly, it is important to properly specify the "strandedness" of sequenced reads (which strand of the source RNA/DNA is sequenced by which read). The Illumina reads used in this study were sequenced using a protocol that generates first-in-pair reads aligning to the anti-sense strand, with the second-in-pair reads aligning to the sense strand of the RNA. Thus the --SS_lib_type RF option was specified to correctly describe the read orientation.

The --normalize was also applied during the assembly, in order to simultaneously speed up the assembly pipeline as well as reduce its memory requirements. Read normalization is applied to the raw kmer

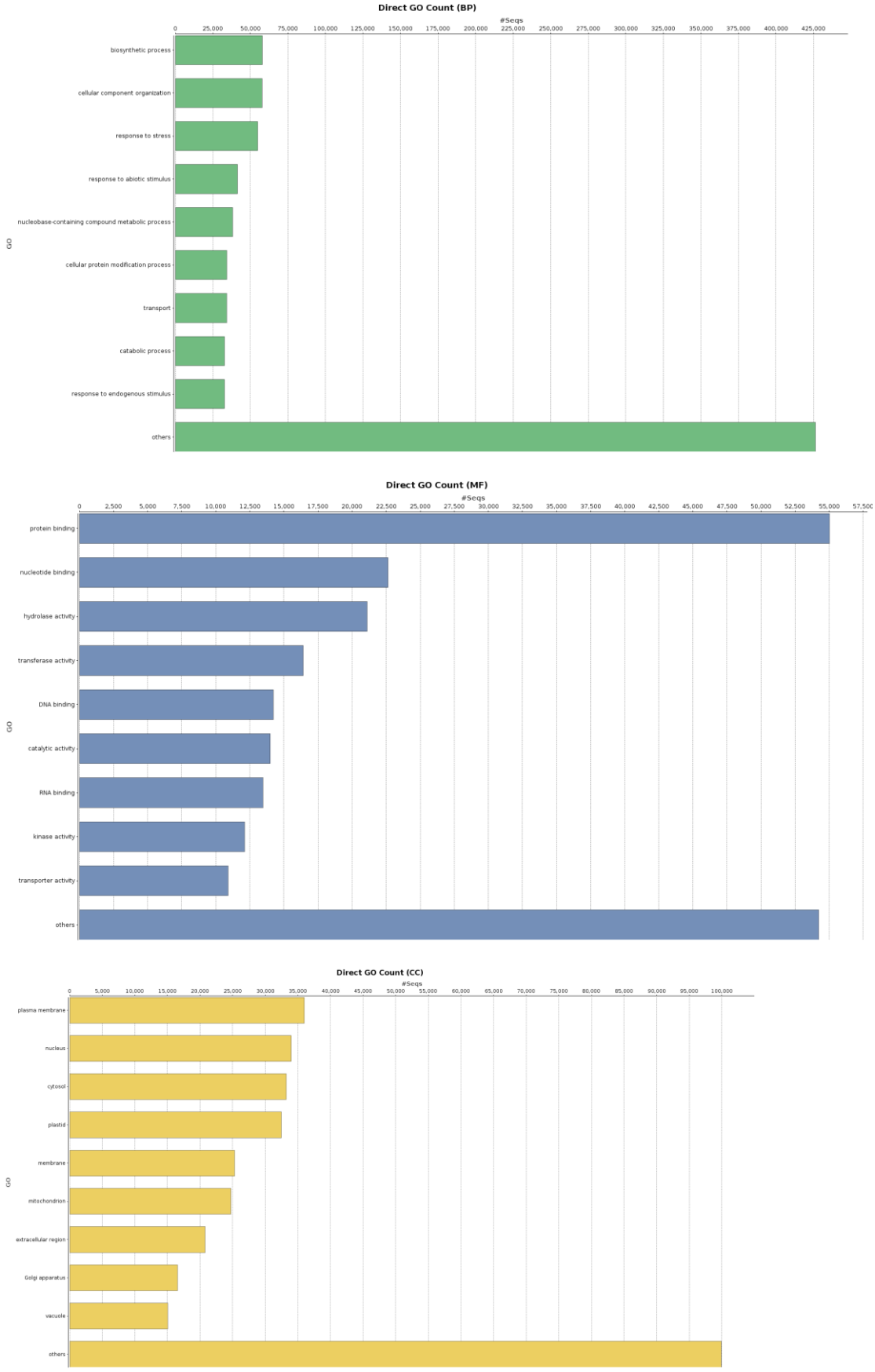
BLASTx search and BLAST2GO

The transcripts were split into groups of 1000, and submitted as an individual job on the university's High Performance Computing (HPC) cluster. The command line BLASTx program (version 2.2.3) was used. The "-outfmt 5" option was used to generate the alignment files required for BLAST2GO. All other BLAST options were left at default values. All the output files were subsequently merged using a utility Python script, downloaded from https://bitbucket.org/peterjc/galaxy-central/src/5cefd5d5536e/tools/ncbi_blast_plus/blast.py.

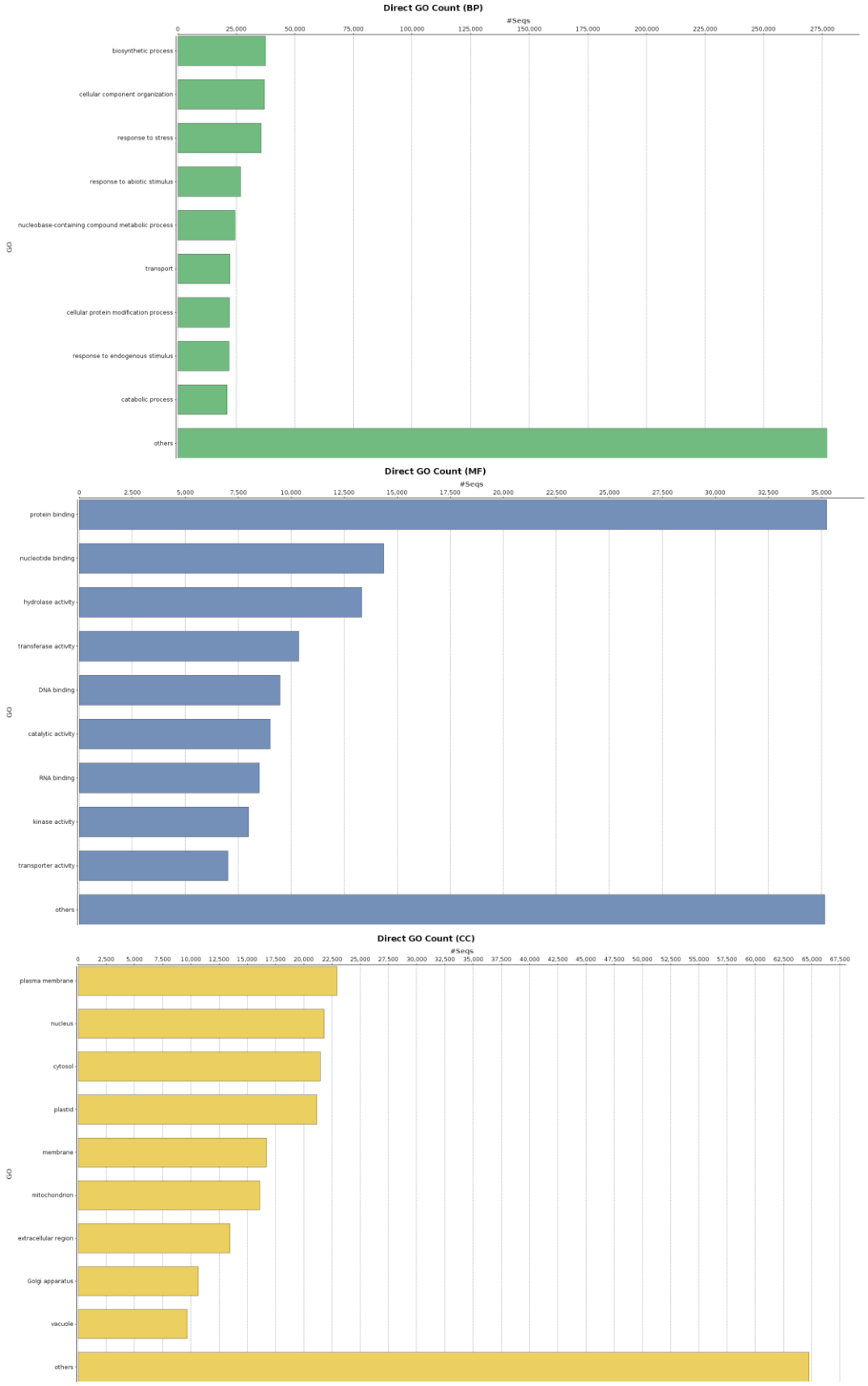
Appendix 5.3 Comparison of similar RNA-Seq studies using *de novo* CLC or Trinity assemblies from Illumina paired-end reads.

Species	**Tissue sampled**	**Raw reads**	**Length (bp)**	**CLC/Trinity**	**Reference**
Tilia platyphyllos	Leaf	123,830,530	100	335,853 CLC 277,097 Trinity	This study
Tilia cordata	Leaf	211,757,920	100	385,859 CLC 316,004 Trinity	This study
Quercus pubescens	Leaf	310,521,410	100	96,006 CLC	Torre *et al.,* 2014
Pinus halepensis	Needles	89,922,352	101	48,629 CLC	Pinosio *et al.,* 2014
Ginkgo bilboa	Leaf[1]	63,680,000	75	70,752 Trinity	Han *et al.,* 2015
Pinus sylvestris.	Needles	258,401,512	100	151,935 Trinity	Wachowiak *et al.,* 2015

[1] Accession number: SRX087425 – Submitted by the Buell Lab, Michigan State.



Appendix 5.4 Sub-categories of GO terms for *T. platyphyllos*: BP – Biological Process; MF – Molecular Function; CC – Cellular Component



Appendix 5.5 Sub-categories of GO terms for *T. cordata*: BP – Biological Process; MF – Molecular Function; CC – Cellular Component

References

Aharoni, A. and Galili, G. (2011) 'Metabolic engineering of the plant primary-secondary metabolism interface', *Curr Opin Biotechnol*, 22(2), 239-244.

Aldrich, P.R., Parker, G.R., Michler, C.H. and Romero-Severson, J. (2003) 'Whole-tree silvic identifications and the microsatellite genetic structure of a red oak species complex in an Indiana old-growth forest', *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 33(11), pp. 2228-2237.

Ally, D., Ritland, K. and Otto, S.P. (2008) 'Can clone size serve as a proxy for clone age? An exploration using microsatellite divergence in Populus tremuloides', *Molecular Ecology*, 17(22), pp. 4897-4911.

Andrew, R. (1971) 'Exine pattern in the pollen of British species of Tilia', *New Phytologist*, 70(4), pp. 683-686.

Arnaud-Haond, S., Alberto, F., Teixeira, S., Procaccini, G., Serrao, E.A. and Duarte, C.M. (2005) 'Assessing genetic diversity in clonal organisms: Low diversity or low resolution? Combining power and cost efficiency in selecting markers', *Journal of Heredity*, 96(4), pp. 434-440.

Arnaud-Haond, S. and Belkhir, K. (2007) 'genclone: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization', *Molecular Ecology Notes*, 7(1), pp. 15-17.

Arnaud-Haond, S., Duarte, C.M., Alberto, F. and Serrao, E.A. (2007) 'Standardizing methods to address clonality in population studies', *Molecular Ecology*, 16(24), pp. 5115-5139.

Arnaud-Haond, S., Teixeira, S., Massa, S.I., Billot, C., Saenger, P., Coupland, G., Duarte, C.M. and SerrÃO, E.A. (2006) 'Genetic structure at range edge: low diversity and high inbreeding in Southeast Asian mangrove (*Avicennia marina*) populations', *Molecular Ecology*, 15(12), pp. 3515-3525.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. and Saunders, N.C. (1987) 'Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics', *Annual review of ecology and systematics*, pp. 489-522.

Awad, L., Fady, B., Khater, C., Roig, A. and Cheddadi, R. (2014) 'Genetic structure and diversity of the endangered fir tree of Lebanon (Abies cilicica Carr.): implications for conservation', *PloS one*, 9(2), p. e90086.

Babington, C.C. (1862) *Manual of British botany, containing the flowering plants and ferns arranged according to the natural orders*. John Van Voorst.

Bacilieri, R., Ducousso, A. and Kremer, A. (1995) 'Genetic, morphological, ecological and phenological differentiation between *Quercus petraea* (Matt) Liebl. and *Quercus robur* L. in a mixed stand of northwest of France', *Silvae Genetica*, 44(1), pp. 1-10.

Bacilieri, R., Ducousso, A. and Kremer, A. (1996) 'Comparison of morphological characters and molecular markers for the analysis of hybridization in sessile and pedunculate oak', *Annales Des Sciences Forestieres*, 53(1), pp. 79-91.

Bagnoli, F., Tsuda, Y., Fineschi, S., Bruschi, P., Magri, D., Zhelev, P., Paule, L., Simeone, M.C., González-Martínez, S.C. and Vendramin, G.G. (2015) 'Combining molecular and fossil data to infer demographic history of Quercus cerris: insights on European eastern glacial refugia', *Journal of Biogeography*, 43(4), 679-690.

Bailey, S.A., Haines-Young, R.H. and Watkins, C. (2002) 'Species presence in fragmented landscapes: modelling of species requirements at the national level', *Biological Conservation*, 108(3), pp. 307-316.

Balloux, F., Lehmann, L. and de Meeûs, T. (2003) 'The population genetics of clonal and partially clonal diploids', *Genetics*, 164(4), pp. 1635-1644.

Barrett, C.F. and Freudenstein, J.V. (2009) 'Patterns of morphological and plastid DNA variation in the *Corallorhiza striata* Species Complex (Orchidaceae)', *Systematic Botany*, 34(3), pp. 496-504.

Bateman, R.M., Smith, R.J. and Fay, M.F. (2008) 'Morphometric and population genetic analyses elucidate the origin, evolutionary significance and conservation implications of *Orchis xangusticruris* (*O. purpurea* x *O. simia*), a hybrid orchid new to Britain', *Botanical Journal of the Linnean Society*, 157(4), pp. 687-711.

Beatty, G.E., McEvoy, P.M., Sweeney, O. and Provan, J. (2008) 'Range-edge effects promote clonal growth in peripheral populations of the one-sided wintergreen *Orthilia secunda*', *Diversity and Distributions*, 14(3), pp. 546-555.

Beatty, G.E. and Provan, J. (2013) 'Post-glacial dispersal, rather than in situ glacial survival, best explains the disjunct distribution of the Lusitanian plant species Daboecia cantabrica (Ericaceae)', *Journal of Biogeography*, 40(2), pp. 335-344.

Beatty, G.E. and Provan, J. (2014) 'Phylogeographical analysis of two cold-tolerant plants with disjunct Lusitanian distributions does not support in situ survival during the last glaciation', *Journal of Biogeography*, 41(11), pp. 2185-2193.

Bennett, K.D. (1985) 'The spread of Fagus grandifolia across eastern North America during the last 18 000 years', *Journal of Biogeography*, pp. 147-164.

Bennett, K.D., Tzedakis, P.C. and Willis, K.J. (1991) 'Quaternary refugia of north European trees', *Journal of Biogeography*, pp. 103-115.

Bermond, G., Ciosi, M., Lombaert, E., Blin, A., Boriani, M., Furlan, L., Toepfer, S. and Guillemaud, T. (2012) 'Secondary contact and admixture between independently invading populations of the western corn rootworm, Diabrotica virgifera virgifera in Europe', *PloS one*, 7(11), p. e50129.

Bhagwat, S.A. and Willis, K.J. (2008) 'Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits?', *Journal of Biogeography*, 35(3), pp. 464-482.

Birks, H.J.B. (1989) 'Holocene isochrone maps and patterns of tree-spreading in the British Isles', *Journal of Biogeography*, pp. 503-540.

Birks, H.J.B., Deacon, J. and Peglar, S. (1975) 'Pollen maps for the British Isles 5000 years ago', *Proceedings of the Royal Society of London B: Biological Sciences*, 189(1094), pp. 87-105.

Blanca, J., Canizares, J., Roig, C., Ziarsolo, P., Nuez, F. and Pico, B. (2011) 'Transcriptome characterization and high throughput SSRs and SNPs discovery in Cucurbita pepo (Cucurbitaceae)', *BMC Genomics*, 12(1), p1.

Bobiec, A. (2007) 'The influence of gaps on tree regeneration: a case study of the mixed lime-hornbeam (Tilio-Carpinetum Tracz. 1962) communities in the Białowieża Primeval Forest', *Polish journal of Ecology*, 55(3), pp. 441-455.

Bodare, S., Tsuda, Y., Ravikanth, G., Uma Shaanker, R. and Lascoux, M. (2013) 'Genetic structure and demographic history of the endangered tree species Dysoxylum malabaricum (Meliaceae) in Western Ghats, India: implications for conservation in a biodiversity hotspot', *Ecology and evolution*, 3(10), pp. 3233-3248.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C. and Phan, I. (2003) 'The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003', *Nucleic Acids Res*, 31, pp. 365 - 370.

Bolikhovskaya, N.S. and Shunkov, M.V. (2014) 'Pleistocene Environments of Northwestern Altai: Vegetation and Climate', *Archaeology, Ethnology and Anthropology of Eurasia*, 42(2), pp. 2-17.

Borzenkova, I., Zorita, E., Borisova, O., Kalniņa, L., Kisielienė, D., Koff, T., Kuznetsov, D., Lemdahl, G., Sapelko, T. and Stančikaitė, M. (2015) 'Climate Change During the Holocene (Past 12,000 Years)', in *Second Assessment of Climate Change for the Baltic Sea Basin*. Springer, pp. 25-49.

Boshier, D. and Buggs, R.J.A. (2015) 'The potential for field studies and genomic technologies to enhance resistance and resilience of British tree populations to pests and pathogens', *Forestry*, 88(1), pp. 27-40.

Brundu, G., Lupi, R., Zapelli, I., Fossati, T., Patrignani, G., Camarda, I., Sala, F. and Castiglione, S. (2008) 'The origin of clonal diversity and structure of Populus alba in Sardinia: evidence from nuclear and plastid microsatellite markers', *Annals of botany*, 102(6), pp. 997-1006.

Buckley, G.P., Mills, J., Kirby, K.J. and Watkins, C. (2015) 'Coppice silviculture: from the Mesolithic to the 21st century', *Europe's changing woods and forests*, pp. 77-92.

Carling, P., Villanueva, I., Herget, J., Wright, N., Borodavko, P. and Morvan, H. (2010) 'Unsteady 1D and 2D hydraulic models with ice dam break for Quaternary megaflood, Altai Mountains, southern Siberia', *Global and Planetary Change*, 70(1–4), pp. 24-34.

Caughley, G., Grice, D., Barker, R. and Brown, B. (1988) 'The edge of the range', *Journal of Animal Ecology*, 57(3), pp. 771-785.

Cavers, S. and Cottrell, J.E. (2015) 'The basis of resilience in forest tree species and its use in adaptive forest management in Britain', *Forestry*, 88(1), pp. 13-26.

Cavers, S., Navarro, C. and Lowe, A.J. (2003) 'Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, Cedrela odorata L., in Mesoamerica', *Molecular Ecology*, 12(6), pp. 1451-1460.

Chapuis, M.-P. and Estoup, A. (2007) 'Microsatellite Null Alleles and Estimation of Population Differentiation', *Molecular Biology and Evolution*, 24(3), pp. 621-631.

Chenault, N., Arnaud-Haond, S., Juteau, M., Valade, R., Almeida, J.-L., Villar, M., Bastien, C. and Dowkiw, A. (2011) 'SSR-based analysis of clonality, spatial genetic structure and introgression from the Lombardy poplar into a natural population of Populus nigra L. along the Loire River', *Tree genetics & genomes*, 7(6), pp. 1249-1262.

Chytry, M. and Sádlo, J. (1997) 'Tilia-dominated calcicolous forests in the Czech Republic from a Central European perspective', *Annali di Botanica*, 55.

Collingham, Y.C. and Huntley, B. (2000) 'Impacts of habitat fragmentation and patch size upon migration rates', *Ecological Applications*, 10(1), pp. 131-144.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. (2005) 'Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research', *Bioinformatics*, 21, pp. 3674 - 3676.

Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.-M. and Estoup, A. (2014) 'DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data', *Bioinformatics*, p.763.

Cornuet, J.-M., Ravigné, V. and Estoup, A. (2010) 'Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1. 0)', *Bmc Bioinformatics*, 11(1), p. 1.

Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.-M., Balding, D.J., Guillemaud, T. and Estoup, A. (2008) 'Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation', *Bioinformatics*, 24(23), pp. 2713-2719.

Cottrell, J.E., Munro, R.C., Tabbener, H.E., Milner, A.D., Forrest, G.I. and Lowe, A.J. (2003) 'Comparison of fine-scale genetic structure using nuclear microsatellites within two British oakwoods differing in population history', *Forest Ecology and Management*, 176(1–3), pp. 287-303.

Cottrell, J.E., Vaughan, S.P., Connolly, T., Sing, L., Moodley, D.J. and Russell, K. (2009) 'Contemporary pollen flow, characterization of the maternal ecological neighbourhood and mating patterns in wild cherry (*Prunus avium* L.)', *Heredity*, 103(2), pp. 118-128.

Couch, S. (1992) 'The practice of avenue planting in the 17th and 18th Centuries.'. *Garden History,* 20(2), pp173-200.

Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M. and Wayne, R.K. (2000) 'Considering evolutionary processes in conservation biology', *Trends in ecology & evolution*, 15(7), pp. 290-295.

Crawford, N.G. (2010) 'smogd: software for the measurement of genetic diversity', *Molecular Ecology Resources*, 10(3), pp. 556-557.

Cristóbal, D., Martínez-Zurimendi, P., Villamediana, I., Ciriza, J., Villar, J., Nanos, N. and Sierra-de-Grado, R. (2014) 'Clonal structure and dynamics of peripheral Populus tremula L. populations', *iForest-Biogeosciences and Forestry*, 7(3), p. 140.

de Witte, L.C. and Stöcklin, J. (2010) 'Longevity of clonal plants: why it matters and how to measure it', *Annals of Botany*, 106(6), pp. 859-870.

De Woody, J., Rickman, T.H., Jones, B.E. and Hipkins, V.D. (2009) 'Allozyme and microsatellite data reveal small clone size and high genetic diversity in aspen in the southern Cascade Mountains', *Forest ecology and management*, 258(5), pp. 687-696.

Dorken, M.E. and Eckert, C.G. (2001) 'Severely reduced sexual reproduction in northern populations of a clonal plant, Decodonverticillatus (Lythraceae)', *Journal of Ecology*, 89(3), pp. 339-350.

Duminil, J., Caron, H., Scotti, I., Cazal, S.-O. and Petit, R.J. (2006) 'Blind population genetics survey of tropical rainforest trees', *Molecular Ecology*, 15(12), pp. 3505-3513.

Duminil, J. and Di Michele, M. (2009) 'Plant species delimitation: A comparison of morphological and molecular markers', *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology*, 143(3), pp. 528-542.

Earl, D. and vonHoldt, B. (2012) 'STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method', *Conservation Genetics Resources*, 4(2), pp. 359-361.

Eckert, C. (2002) 'The loss of sex in clonal plants', *Evolutionary Ecology*, 15(4-6), pp. 501-520.

Ellegren, H. (2008) 'Sequencing goes 454 and takes large-scale genomics into the wild', *Molecular Ecology*, 17(7), pp. 1629-1631.

Ellegren, H., Primmer, C.R. and Sheldon, B.C. (1995) 'Microsatellite Evolution - Directionality or Bias', *Nature Genetics*, 11(4), pp. 360-362.

Ellstrand, N.C. and Elam, D.R. (1993) 'Population genetic consequences of small population size: implications for plant conservation', *Annual review of Ecology and Systematics*, pp. 217-242.

Ellstrand, N.C. and Roose, M.L. (1987) 'Patterns of Genotypic Diversity in Clonal Plant Species', *American Journal of Botany*, 74(1), pp. 123-131.

Elwes, H.J. and Henry, A. (1913) 'The Trees of Great Britain and Ireland', *Privately printed, Edinburgh*.

Ennos, R., Worrell, R., Arkle, P. and Malcolm, D. (2000) *Genetic variation and conservation of British native trees and shrubs: current knowledge and policy implications*. Forestry Commission.

Estoup, A., Cornuet, J.M., Dehne-Garcia, A. and Loire, E. (2015) 'New version of the computer program DIYABC (DIYABC v2. 1.0): a user-friendly approach to approximate Bayesian computation for inference on population history using molecular markers'.

Evanno, G., Regnaut, S. and Goudet, J. (2005) 'Detecting the number of clusters of individuals using the software structure: a simulation study', *Molecular Ecology*, 14(8), pp. 2611-2620.

Excoffier, L. and 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), pp. 564-567.

Eycott, A.E., Watts, K., Moseley, D.G. and Ray, D. (2007) *Evaluating biodiversity in fragmented landscapes: the use of focal species* (FCIN089). Edinburgh: Forestry_Commission.

Falush, D., Stephens, M. and Pritchard, J.K. (2003) 'Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies', *Genetics*, 164(4), pp. 1567-87.

Fineschi, S., Salvini, D., Taurchini, D., Carnevale, S. and Vendramin, G.G. (2003) 'Chloroplast DNA variation of *Tilia cordata* (Tiliaceae)', *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 33(12), pp. 2503-2508.

Forestry_Commission (2016) 'Small-leaved lime (SLI)'. 25th March 2016. Available at: http://www.forestry.gov.uk/fr/infd-8czbch.

Fromm, M. and Hattemer, H.H. (2003) 'Inheritance of allozymes and hybridization in two European *Tilia* species', *Heredity*, 91(3), pp. 337-344.

Fuentes-Utrilla, P., Valbuena-Carabaña, M., Ennos, R. and Gil, L. (2014) 'Population clustering and clonal structure evidence the relict state of Ulmus minor Mill. in the Balearic Islands', *Heredity*, 113(1), pp. 21-31.

Fyfe, R.M., Woodbridge, J. and Roberts, N. (2015) 'From forest to farmland: pollen-inferred land cover change across Europe using the pseudobiomization approach', *Global change biology*, 21(3), pp. 1197-1212.

Godwin, H. (1975) 'History of the natural forests of Britain: establishment, dominance and destruction', *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 271(911), pp. 47-67.

Goldstein, D.B., Ruiz Linares, A., Cavalli-Sforza, L.L. and Feldman, M.W. (1995) 'An evaluation of genetic distances for use with microsatellite loci', *Genetics*, 139(1), pp. 463-471.

Goudet, J. (1995) 'FSTAT (Version 1.2): A computer program to calculate F-statistics', *Journal of Heredity*, 86(6), pp. 485-486.

Goudet, J. (2001) 'FSTAT, a program to estimate and test gene diversities and fixation indices version 2.9. 3.2, updated from Goudet 1995'. FSTAT.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. and Zeng, Q. (2011) 'Full-length transcriptome assembly from RNA-Seq data without a reference genome', *Nature biotechnology*, 29(7), pp. 644-652.

Grashof-Bokdam, C. (1997) 'Forest species in an agricultural landscape in the Netherlands' Effects of habitat fragmentation', *Journal of Vegetation Science,* 8(1), pp21-28.

Groisman, P.Y., Blyakharchuk, T.A., Chernokulsky, A.V., Arzhanov, M.M., Marchesini, L.B., Bogdanova, E.G., Borzenkova, I.I., Bulygina, O.N., Karpenko, A.A. and Karpenko, L.V. (2013) 'Climate changes in Siberia', in *Regional Environmental Changes in Siberia and Their Global Consequences*. Springer, pp. 57-109.

Gugerli, F., Walser, J.C., Dounavi, K., Holderegger, R. and Finkeldey, R. (2007) 'Coincidence of small-scale spatial discontinuities in leaf morphology and nuclear microsatellite variation of Quercus petraea and Q-robur in a mixed forest', *Annals of Botany*, 99(4), pp. 713-722.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B. and Lieber, M. (2013) 'De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis', *Nature protocols*, 8(8), pp. 1494-1512.

Halkett, F., Simon, J.C. and Balloux, F. (2005) 'Tackling the population genetics of clonal and partially clonal organisms', *Trends in Ecology & Evolution*, 20(4), pp. 194-201.

Hamrick, J.L. (2004) 'Response of forest trees to global environmental changes', *Forest Ecology and Management*, 197(1–3), pp. 323-335.

Hamrick, J.L. and Godt, M.J.W. (1996) 'Effects of life history traits on genetic diversity in plant species', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1345), pp. 1291-1298.

Han, S., Wu, Z., Jin, Y., Yang, W. and Shi, H. (2015) 'RNA-Seq analysis for transcriptome assembly, gene identification, and SSR mining in ginkgo (Ginkgo biloba L.)', *Tree Genetics & Genomes*, 11(3), pp. 1-10.

Hansen, O.K., Thomsen, P. and Rasmussen, C.W. (2014) 'DNA markers provide insight about common lime in historical plantings–An example from the Royal Danish Gardens', *Urban Forestry & Urban Greening*, 13(3), pp. 543-552.

Havrdová, A., Douda, J., Krak, K., Vít, P., Hadincová, V., Zákravský, P. and Mandák, B. (2015) 'Higher genetic diversity in recolonized areas than in refugia of Alnus glutinosa triggered by continent-wide lineage admixture', *Molecular Ecology*, 24(18), pp. 4759-4777.

Hedrick, P.W. (2005) 'A standardized genetic differentiation measure', *Evolution*, 59(8), pp. 1633-1638.

Hedrick, P.W., Lee, R.N. and Hurt, C.R. (2006) 'The endangered Sonoran topminnow: examination of species and ESUs using three mtDNA genes', *Conservation Genetics*, 7(4), pp. 483-492.

Heinken, T. and Weber, E. (2013) 'Consequences of habitat fragmentation for plant species: Do we know enough?', *Perspectives in Plant Ecology, Evolution and Systematics*, 15(4), pp. 205-216.

Hermy, M., Honnay, O., Firbank, L., Grashof-Bokdam, C. and Lawesson, J.E. (1999) 'An ecological comparison between ancient and other forest plant species of Europe, and the implications for forest conservation', *Biological Conservation*, 91(1), pp. 9-22.

Hewitt, G. (2000) 'The genetic legacy of the Quaternary ice ages', *Nature*, 405(6789), pp. 907-913.

Hewitt, G.M. (1999) 'Post-glacial re-colonization of European biota', *Biological Journal of the Linnean Society*, 68(1-2), pp. 87-112.

Hewitt, G.M. (2004) 'Genetic consequences of climatic oscillations in the Quaternary', *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 359(1442), pp. 183-195.

Hosseinzadeh Colagar, A., Yusefi, M., Zarei, M. and Yousefzadeh, H. (2013) 'Assessment of genetic diversity of *Tilia rubra* DC. by RAPD analysis in the Hyrcanian forests, north of Iran', *Polish Journal of Ecology*, 61(2), pp. 341 - 348.

Hubisz, M.J., Falush, D., Stephens, M. and Pritchard, J.K. (2009) 'Inferring weak population structure with the assistance of sample group information', *Molecular ecology resources*, 9(5), pp. 1322-1332.

Huntley, B. and Birks, H.J.B. (1983) 'An atlas of past and present pollen maps for Europe, 0-13,000 years ago'.

IUCN (2015) *The IUCN Red List of Threatened Species*. Available at: http://www.iucnredlist.org/ (Accessed: 16th December 2015).

Jakobsson, M. and Rosenberg, N.A. (2007) 'CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure', *Bioinformatics*, 23(14), pp. 1801-1806.

Jaworski, A., Kołodziej, Z. and Bartkowicz, L. (2005) 'Structure and dynamics of stands of primeval character composed of the little-leaf linden (Tilia cordata Mill.) in the “Las lipowy Obrożyska” reserve (southern Poland)', *Journal of Forest Science*, 51(7), pp. 283-304.

Jónsdóttir, I.S., Augner, M., Fagerström, T., Persson, H. and Stenström, A. (2000) 'Genet age in marginal populations of two clonal Carex species in the Siberian Arctic', *Ecography*, 23(4), pp. 402-412.

Jørgensen, T., Haile, J., Möller, P.E.R., Andreev, A., Boessenkool, S., Rasmussen, M., Kienast, F., Coissac, E., Taberlet, P. and Brochmann, C. (2012) 'A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability', *Molecular Ecology*, 21(8), pp. 1989-2003.

Jost, L. (2008) 'GST and its relatives do not measure differentiation', *Molecular Ecology*, 17(18), pp. 4015-4026.

Kelleher, C.T., Hodkinson, T.R., Douglas, G.C. and Kelly, D.L. (2005) 'Species distinction in Irish populations of *Quercus petraea* and *Q. robur*: morphological versus molecular analyses', *Annals of Botany*, 96(7), pp. 1237-1246.

Kirk, H. and Freeland, J.R. (2011) 'Applications and implications of neutral versus non-neutral markers in molecular ecology', *International journal of molecular sciences*, 12(6), pp. 3966-3988.

Krementski, C.V. (1997) 'The Late Holocene environmental and climate shift in Russia and surrounding lands', in *Third millennium BC climate change and Old World collapse*. Springer, pp. 351-370.

Kuneš, P., Pelánková, B., Chytrý, M., Jankovská, V., Pokorný, P. and Petr, L. (2008) 'Interpretation of the last-glacial vegetation of eastern-central Europe using modern analogues from southern Siberia', *Journal of Biogeography*, 35(12), pp. 2223-2236.

Kunneman, B.P.A.M. and Albers, M.R.J. (1991) 'Linden Trees (Tilia spp.)', in Bajaj, Y.P.S. (ed.) *Trees III*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 152-163.

Lagercrantz, U., Ellegren, H. and Andersson, L. (1993) 'The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates', *Nucleic Acids Research*, 21(5), pp. 1111-1115.

Larcombe, M.J., Vaillancourt, R., xe, E, Jones, R.C. and Potts, B.M. (2014) 'Assessing a Bayesian approach for detecting exotic hybrids between plantation and native Eucalypts', *International Journal of Forestry Research*, 2014, p. 13.

Larsen, E., KjæR, K., Demidov, I.N., Funder, S., Grøsfjeld, K., Houmark-Nielsen, M., Jensen, M., Linge, H. and Lysa, A. (2006) 'Late Pleistocene glacial and lake history of northwestern Russia', *Boreas*, 35(3), pp. 394-424.

Lawesson, J.E. (2004) *Ancient forests in Denmark and the importance of Tilia*. pp97-116. Wallingford: CABI.

Levsen, N.D., Tiffin, P. and Olson, M.S. (2012) 'Pleistocene speciation in the genus Populus (Salicaceae)', *Systematic biology*, 61(3), pp. 401-412.

Li, Y., Luo, H., Sun, C., Song, J., Sun, Y., Wu, Q., Wang, N., Yao, H., Steinmetz, A. and Chen, S. (2010) 'EST analysis reveals putative genes involved in glycyrrhizin biosynthesis', *BMC Genomics*, 11.

Liesebach, H. and Sinkó, Z. (2008) 'A contribution to the systematics of the genus *Tilia* with respect to some hybrids by RAPD analysis', *Dendrobiology*, 59, pp. 13-22.

Loeschcke, V., Tomiuk, J. and Jain, S.K. (2013) *Conservation genetics*. Birkhäuser.

Logan, S.A., Phuekvilai, P. and Wolff, K. (2015) 'Ancient woodlands in the limelight: delineation and genetic structure of ancient woodland species Tilia cordata and Tilia platyphyllos (Tiliaceae) in the UK', *Tree Genetics & Genomes*, 11(3), pp. 1-12.

Lowe, A. (2004) 'Restoration of wooded landscapes–Informing assumptions with microsatellite technology', *Report Produced Under Commission for Forestry Commission, Edinburgh*.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q. and Liu, Y. (2012) 'SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler', *GigaScience*, 1(1), pp. 1-6.

Ma, Q.Q.., Song, H.-x., Zhou, S.-q., Yang, W.-q., Li, D.-s. and Chen, J.-s. (2013) 'Genetic Structure in Dwarf Bamboo (Bashania fangiana) Clonal Populations with Different Genet Ages'. *PLoS one,* 8(11), e78784.

MacDonald, G.M. and Cwynar, L.C. (1985) 'A fossil pollen based reconstruction of the late Quaternary history of lodgepole pine (Pinus contorta ssp. latifolia) in the western interior of Canada', *Canadian Journal of Forest Research*, 15(6), pp. 1039-1044.

Markova, A.K., Simakova, A.N. and Puzachenko, A.Y. (2009) 'Ecosystems of Eastern Europe at the time of maximum cooling of the Valdai glaciation (24–18kyr BP) inferred from data on plant communities and mammal assemblages', *Quaternary International*, 201(1), pp. 53-59.

Martin, J., Bruno, V.M., Fang, Z., Meng, X., Blow, M., Zhang, T., Sherlock, G., Snyder, M. and Wang, Z. (2010) 'Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads', *BMC genomics*, 11(1), p. 663.

Martin, J.A. and Wang, Z. (2011) 'Next-generation transcriptome assembly', *Nature Reviews Genetics*, 12(10), pp. 671-682.

Maurer, W.D. and Tabel, U. (1995) 'A methodological study to improve Isozyme analysis for identification of clones of *Tilia* (Linden syn. Lime tree)', *Silvae Genetica*, 44(5-6), pp. 351 - 356.

Mayr, E. (1942) *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

McCarthy, D. (2012) *Systematics and phylogeography of the genus Tilia in North America*. Doctoral Dissertation, University of Hawaii.

Meimberg, H., Rischer, H., Turini, F., Chamchumroon, V., Dreyer, M., Sommaro, M., Bringmann, G. and Heubl, G. (2010) 'Evidence for species differentiation within the *Ancistrocladus tectorius* complex (Ancistrocladaceae) in Southeast Asia: a molecular approach', *Plant Systematics and Evolution*, 284(1-2), pp. 77-98.

Miścicki, S. (2012) 'Structure and dynamics of temperate lowland natural forest in the Biaowiea National Park, Poland', *Forestry*, 85(4), pp. 473-483.

Mock, K.E., Rowe, C.A., Hooten, M.B., Dewoody, J. and Hipkins, V.D. (2008) 'Clonal dynamics in western North American aspen (Populus tremuloides)', *Molecular Ecology*, 17(22), pp. 4827-4844.

Morgan-Richards, M. and Wolff, K. (1999) 'Genetic structure and differentiation of *Plantago major* reveals a pair of sympatric sister species', *Molecular Ecology*, 8(6), pp. 1027-1036.

Morgante, M. and Olivieri, A.M. (1993) 'PCR-amplified microsatellites as markers in plant genetics', *The Plant Journal*, 3(1), pp. 175-182.

Motamayor, J.C., Mockaitis, K., Schmutz, J., Haiminen, N., Livingstone Iii, D., Cornejo, O., Findley, S., Zheng, P., Utro, F. and Royaert, S. (2013) 'The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color', *Genome biology*, 14(6)r53

Nagamitsu, T., Ogawa, M., Ishida, K. and Tanouchi, H. (2004) 'Clonal diversity, genetic structure, and mode of recruitment in a Prunus ssiori population established after volcanic eruptions', *Plant Ecology*, 174(1), pp. 1-10.

Naidina, O.D. and Bauch, H.A. (2001) 'A Holocene pollen record from the Laptev Sea shelf, northern Yakutia', *Global and Planetary Change*, 31(1–4), pp. 141-153.

Nakazawa, F., Uetake, J., Suyama, Y., Kaneko, R., Takeuchi, N., Fujita, K., Motoyama, H., Imura, S. and Kanda, H. (2013) 'DNA analysis for section identification of individual Pinus pollen grains from Belukha glacier, Altai Mountains, Russia', *Environmental Research Letters*, 8(1), p. 014032.

Natural_England (2014) [Online]. Available at: www.gov.uk/government/organisations/natural-england (Accessed: May).

Neale, D.B. and Kremer, A. (2011) 'Forest tree genomics: growing resources and applications', *Nature Reviews Genetics*, 12(2), pp. 111-122.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia university press.
Nei, M. and Chesser, R.K. (1983) 'Estimation of fixation indices and gene diversities', *Annals of human genetics*, 47(3), pp. 253-259.

Neigel, J.E. and Avise, J.C. (1986) 'Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation', *Evolutionary processes and theory*, pp. 515-534.

Neophytou, C. (2014) 'Bayesian clustering analyses for genetic assignment and study of hybridization in oaks: effects of asymmetric phylogenies and asymmetric sampling schemes', *Tree Genetics & Genomes*, 10(2), pp. 273-285.

Normand, S., Treier, U.A. and Odgaard, B.V. (2011) 'commentary: Tree refugia and slow forest development in response to post-LGM warming in North-Eastern European Russia', *Frontiers of Biogeography*, 2(4), pp91-93

Novák, J., Trotsiuk, V., Sýkora, O., Svoboda, M. and Chytrý, M. (2014) 'Ecology of Tilia sibirica in a continental hemiboreal forest, southern Siberia: An analogue of a glacial refugium of broad-leaved temperate trees?', *The Holocene*, DOI: 10.1177/0959683614534744.

Oliveira, E.J., Pádua, J.G., Zucchi, M.I., Vencovsky, R. and Vieira, M.L.C. (2006) 'Origin, evolution, and genome distribution of microsatellites'. *Genetics and Molecular Biology*, 29(2), pp. 294-307.

Opdam, P. and Wascher, D. (2004) 'Climate change meets habitat fragmentation: linking landscape and biogeographical scale levels in research and conservation', *Biological Conservation*, 117(3), pp. 285-297.

Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R.K. and Hedrick, P.W. (2010) 'Conservation genetics in transition to conservation genomics', *Trends in genetics*, 26(4), pp. 177-187.

Parks, J.C. and Werth, C.R. (1993) 'A study of the spatial features of clones in an population of Bracken fern, *Pteridium aquilinum* (Dennstaedtiaceae)', *American Journal of Botany*, 80(5), pp. 537-544.
Parmesan, C. (2006) 'Ecological and evolutionary responses to recent climate change', *Annual Review of Ecology, Evolution, and Systematics*, pp. 637-669.

Parsons, P.A. (1991) 'Evolutionary rates: stress and species boundaries', *Annual Review of Ecology and Systematics*, pp. 1-18.

Peakall, R. and Smouse, P.E. (2012) 'GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update', *Bioinformatics (Oxford, England)*, 28(19), pp. 2537-2539.

Peterken, G.F. (1974) 'A method for assessing woodland flora for conservation using indicator species', *Biological Conservation*, 6(4), pp. 239-245.

Peterken, G.F. (1977) 'Habitat conservation priorities in British and European woodlands', *Biological Conservation*, 11(3), pp. 223-236.

Peterken, G.F. (1983) 'Woodland conservation in Britain', *Conservation in perspective*, pp. 83-100.

Peterken, G.F. and Welch, R.C. (1975) 'Bedford Purlieus: its history, ecology and management'. In: Proceddings of the Symposium, Vol. 7. Monks Wood Experimental Station, Abbots Ripton.

Peterson, G.M. (1983) 'Recent pollen spectra and zonal vegetation in the western USSR', *Quaternary Science Reviews*, 2(4), pp. 281-321.

Petit, R.J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D. and Lascoux, M. (2003) 'Glacial refugia: hotspots but not melting pots of genetic diversity', *science*, 300(5625), pp. 1563-1565.

Petit, S., Griffiths, L., Smart, S.S., Smith, G.M., Stuart, R.C. and Wright, S.M. (2004) 'Effects of area and isolation of woodland patches on herbaceous plant species richness across Great Britain', *Landscape Ecology*, 19(5), pp. 463-471.

Peyaud, V., Ritz, C. and Krinner, G. (2007) 'Modelling the Early Weichselian Eurasian Ice Sheets: role of ice shelves and influence of ice-dammed lakes', *Clim. Past*, 3(3), pp. 375-386.

Pfeiffer, A., Olivieri, A.M. and Morgante, M. (1997) 'Identification and characterization of microsatellites in Norway spruce (Picea abies K.)', *Genome*, 40(4), pp. 411-419.

Phuekvilai, P. (2014) *Relicts, refugia, and reticulation: a study of population history, hybrids and phylogeny in the long-lived flowering tree genus Tilia.* Ph.D. Thesis thesis. Newcastle University.

Phuekvilai, P. and Wolff, K. (2013) 'Characterization of microsatellite loci in *Tilia platyphyllos* (Malvaceae) and cross-amplification in related species', *Applications in Plant Sciences*, 1(4), p. 1200386.

Pichersky, E. and Gang, D.R. (2000) 'Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective', *Trends in Plant Science*, 5(10), pp. 439-445.

Pigott, C.D. (1968) 'Cirsium Acaulon (L.) Scop', *The Journal of Ecology*, pp. 597-612.

Pigott, C.D. (1969) 'The status of *Tilia Cordata* and *T. Platyphyllos* on the Derbyshire limestone', *Journal of Ecology*, 57(2), pp. 491-504.

Pigott, C.D. (1975) 'Natural regeneration of *Tilia cordata* in relation to forest structure in the Forest of Białowieża, Poland', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 270(904), pp. 151-179.

Pigott, C.D. (1981a) 'Nature of seed sterility and natural regeneration of Tilia cordata near its northern limit in Finland', *Annales Botanici Fennici*, 18(4), pp. 255-263.

Pigott, C.D. (1981b) 'The status, ecology and conservation of *Tilia platyphyllos* in Britain', in *The biological aspects of rare plant conservation. Chichester etc., Wiley*. pp. 305-317.

Pigott, C.D. (1991) '*Tilia cordata* MILLER', *Journal of Ecology*, 79(4), pp. 1147-1207.

Pigott, C.D. (2000) 'A probably native and regenerating population of *Tilia platyphyllos* Scop.', *Bedfordshire. Watsonia*, 23(2), pp. 344-345.

Pigott, C.D. and Huntley, J.P. (1978) 'Factors controlling the distribution of Tilia cordata at the northern limits of its geographical range', *New Phytologist*, 81(2), pp. 429-441.

Pigott, C.D. and Huntley, J.P. (1981) 'Factors controlling the distribution of Tilia cordata at the northern limits of its geographical range III. Nature and causes of seed sterility', *New phytologist*, 87(4), pp. 817-839.

Pigott, D. (2012) *Lime-trees and Basswoods*. Cambridge University Press.

Pinosio, S., González-Martínez, S.C., Bagnoli, F., Cattonaro, F., Grivet, D., Marroni, F., Lorenzo, Z., Pausas, J.G., Verdú, M. and Vendramin, G.G. (2014) 'First insights into the transcriptome and development of new genomic tools of a widespread circum-Mediterranean tree species, Pinus halepensis Mill', *Molecular Ecology Resources*, 14(4), pp. 846-856.

Poudel, R.C., Möller, M., Li, D.-Z., Shah, A. and Gao, L.-M. (2014) 'Genetic diversity, demographical history and conservation aspects of the endangered yew tree Taxus contorta (syn. Taxus fuana) in Pakistan', *Tree Genetics & Genomes*, 10(3), pp. 653-665.

Price, E.A.C. and Marshall, C. (1999) 'Clonal plants and environmental heterogeneity–An introduction to the proceedings', *Plant Ecology*, 141(1-2), pp. 3-7.

Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) 'Inference of population structure using multilocus genotype data', *Genetics*, 155(2), pp. 945-959.

Provan, J. and Bennett, K.D. (2008) 'Phylogeographic insights into cryptic glacial refugia', *Trends in ecology & evolution*, 23(10), pp. 564-571.

Provan, J., Powell, W. and Hollingsworth, P.M. (2001) 'Chloroplast microsatellites: new tools for studies in plant ecology and evolution', *Trends Ecol Evol*, 16(3), pp. 142-147.

Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2005) 'NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins', *Nucleic acids research*, 33(suppl 1), pp. D501-D504.

Qiu, Q., Ma, T., Hu, Q., Liu, B., Wu, Y., Zhou, H., Wang, Q., Wang, J. and Liu, J. (2011) 'Genome-scale transcriptome analysis of the desert poplar, Populus euphratica', *Tree Physiology*, p. tpr015.

Rackham, O. (2003) *Ancient Woodland: it's history, vegetation and uses in England.* 2nd edn. Dalbeattie, UK: Castlepoint Press.

Rackham, O. (2008) 'Ancient woodlands: modern threats', *New Phytologist*, 180(3), pp. 571-586.

Radoglou, K., Dobrowolska, D., Spyroglou, G. and Nicolescu, V. (2009) 'A review on the ecology and silviculture of limes (*Tilia cordata* Mill., *Tilia platyphyllos* Scop. and *Tilia tomentosa* Moench.) in Europe', *Die Bodenkultur*, 60(3), pp. 9-19.

Rasmussen, K.K. and Kollmann, J. (2008) 'Low genetic diversity in small peripheral populations of a rare European tree (Sorbus torminalis) dominated by clonal reproduction', *Conservation genetics.*, 9(6), pp. 1533-1539.

Raymond, M. and Rousset, F. (1995) 'Genepop (version 1.2) - population genetics software for exact tests and ecumenicism', *Journal of Heredity*, 86(3), pp. 248-249.

Reuther, A.U., Herget, J., Ivy-Ochs, S., Borodavko, P., Kubik, P.W. and Heine, K. (2006) 'Constraining the timing of the most recent cataclysmic flood event from ice-

dammed lakes in the Russian Altai Mountains, Siberia, using cosmogenic in situ 10Be', *Geology*, 34(11), pp. 913-916.

Richardson, J.E., Whitlock, B.A., Meerow, A. and Madriñán, S. (2015) 'The Age of Chocolate: a biogeographic history of Theobroma and Malvaceae', *Frontiers in Ecology and Evolution*, 3, p. 120.

Riutta, T., Slade, E.M., Morecroft, M.D., Bebber, D.P. and Malhi, Y. (2014) 'Living on the edge: quantifying the structure of a fragmented forest landscape in England', *Landscape Ecology*, 29(6), pp. 949-961.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M. and Qian, J.Q. (2010) 'De novo assembly and analysis of RNA-seq data', *Nature methods*, 7(11), pp. 909-912.

Rosenberg, N.A. (2004) 'distruct: a program for the graphical display of population structure', *Molecular Ecology Notes*, 4(1), pp. 137-138.

Rousset, F. (2008) 'genepop'007: a complete re-implementation of the genepop software for Windows and Linux', *Molecular ecology resources*, 8(1), pp. 103-106.

Rudoy, A.N. and Baker, V.R. (1993) 'Sedimentary effects of cataclysmic late Pleistocene glacial outburst flooding, Altay Mountains, Siberia', *Sedimentary Geology*, 85(1), pp. 53-62.

Santos-del-Blanco, L., de-Lucas, A.I., González-Martínez, S.C., Sierra-de-Grado, R. and Hidalgo, E. (2013) 'Extensive Clonal Assemblies in Populus alba and Populus x canescens from the Iberian Peninsula', *Tree genetics & genomes*, 9(2), pp. 499-510.

Schulz, M.H., Zerbino, D.R., Vingron, M. and Birney, E. (2012) 'Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels', *Bioinformatics*, 28(8), pp. 1086-1092.

Shimizu, Y., Ando, M. and Sakai, F. (2002) 'Clonal structure of natural populations of Cryptomeria japonica growing at different positions on slopes, detected using RAPD markers', *Biochemical Systematics and Ecology*, 30(8), pp. 733-748.

Siegert, M.J., Dowdeswell, J.A., Hald, M. and Svendsen, J.-I. (2001) 'Modelling the Eurasian Ice Sheet through a full (Weichselian) glacial cycle', *Global and Planetary Change*, 31(1), pp. 367-385.

Silvertown, J. (2008) 'The Evolutionary Maintenance of Sexual Reproduction: Evidence from the Ecological Distribution of Asexual Reproduction in Clonal Plants', *International Journal of Plant Sciences*, 169(1), pp. 157-168.

SilviFuture (2015). Available at: http://www.silvifuture.org.uk/about.

Simpson, E.H. (1949) 'Measurement of diversity', *Nature,* 163, p688

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M. and Birol, I. (2009) 'ABySS: a parallel assembler for short read sequence data', *Genome research*, 19(6), pp. 1117-1123.

Soliani, C., Tsuda, Y., Bagnoli, F., Gallo, L.A., Vendramin, G.G. and Marchelli, P. (2015) 'Halfway encounters: Meeting points of colonization routes among the southern beeches Nothofagus pumilio and N. antarctica', *Molecular phylogenetics and evolution*, 85, pp. 197-207.

Somervuo, P., Kvist, J., Ikonen, S., Auvinen, P., Paulin, L., Koskinen, P., Holm, L., Taipale, M., Duplouy, A. and Ruokolainen, A. (2014) 'Transcriptome analysis reveals signature of adaptation to landscape fragmentation', *PloS one*, 9(7), p. e101467.

Soranzo, N., Provan, J. and Powell, W. (1999) 'An example of microsatellite length variation in the mitochondrial genome of conifers', *Genome*, 42(1), pp. 158-61.

Stewart, J.R. and Lister, A.M. (2001) 'Cryptic northern refugia and the origins of the modern biota', *Trends in Ecology & Evolution*, 16(11), pp. 608-613.

Stoeckel, S., Grange, J., Fernandez-Manjarres, J.F., Bilger, I., Frascaria-Lacoste, N. and Mariette, S. (2006) 'Heterozygote excess in a self-incompatible and partially clonal forest tree species - Prunus avium L', *Molecular Ecology*, 15(8), pp. 2109-2118.

Storfer, A., Murphy, M.A., Spear, S.F., Holderegger, R. and Waits, L.P. (2010) 'Landscape genetics: where are we now?', *Molecular Ecology*, 19(17), pp. 3496-3514.

Strand M., Prolla, T.A., Liskey, R.M. and Petes, T.D. (1993) 'Destabilization of tracks of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair'. *Nature* 365. pp. 274-276.

Stravinskienė, V., Snieškienė, V. and Stankevičienė, A. (2015) 'Health condition of Tilia cordata Mill. trees growing in the urban environment', *Urban Forestry & Urban Greening*, 14(1), pp. 115-122.

Streiff, R., Labbe, T., Bacilieri, R., Steinkellner, H., Glössl, J. and Kremer, A. (1998) 'Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites', *Molecular Ecology*, 7(3), pp. 317-328.

Strickler, S.R., Bombarely, A. and Mueller, L.A. (2012) 'Designing a transcriptome next-generation sequencing project for a nonmodel plant species', *American Journal of Botany*, 99(2), pp. 257-266.

Sun, G.L., Díaz, O., Salomon, B. and von Bothmer, R. (2001) 'Genetic diversity and structure in a natural *Elymus caninus* population from Denmark based on microsatellite and isozyme analyses', *Plant Systematics and Evolution*, 227(3-4), pp. 235-244.

Suvanto, L.I. and Latva-Karjanmaa, T.B. (2005) 'Clone identification and clonal structure of the European aspen (Populus tremula)', *Molecular Ecology*, 14(9), pp. 2851-2860.

Svendsen, J.I., Alexanderson, H., Astakhov, V.I., Demidov, I., Dowdeswell, J.A., Funder, S., Gataullin, V., Henriksen, M., Hjort, C. and Houmark-Nielsen, M. (2004) 'Late Quaternary ice sheet history of northern Eurasia', *Quaternary Science Reviews*, 23(11), pp. 1229-1271.

Svenning, J.C., Normand, S. and Kageyama, M. (2008) 'Glacial refugia of temperate trees in Europe: insights from species distribution modelling', *Journal of Ecology*, 96(6), pp. 1117-1127.

Tarasov, P., Granoszewski, W., Bezrukova, E., Brewer, S., Nita, M., Abzaeva, A. and Oberhänsli, H. (2005) 'Quantitative reconstruction of the last interglacial vegetation and climate based on the pollen record from Lake Baikal, Russia', *Climate Dynamics*, 25(6), pp. 625-637.

Tarasov, P.E., Bezrukova, E.V. and Krivonogov, S.K. (2009) 'Late Glacial and Holocene changes in vegetation cover and climate in southern Siberia derived from a 15 kyr long pollen record from Lake Kotokel', *Climate of the Past*, 5(3), pp. 285-295.

Tautz, D. and Renz, M. (1984) 'Simple sequences are ubiquitous repetitive components of eukaryotic genomes', *Nucleic Acids Research*, 12(10), pp. 4127-4138.

R Core Development Team (2015) 'R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing'.

Thiel, T., Michalek, W., Varshney, R. and Graner, A. (2003) 'Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.)', *Theoretical and Applied Genetics*, 106(3), pp. 411-422.

Tsuda, Y., Nakao, K., Ide, Y. and Tsumura, Y. (2015) 'The population demography of Betula maximowicziana, a cool-temperate tree species in Japan, in relation to the last glacial period: its admixture-like genetic structure is the result of simple population splitting not admixing', *Molecular ecology*, 24(7), pp. 1403-1418.

Turner, J. (1962) 'The Tilia decline: an anthropogenic interpretation', *New Phytologist*, 61(3), pp. 328-341.

Vähä, J.P. and Primmer, C.R. (2006) 'Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci', *Molecular Ecology*, 15(1), pp. 63-72.

Väliranta, M., Kaakinen, A., Kuhry, P., Kultti, S., Salonen, J.S. and Seppä, H. (2011) 'Scattered late-glacial and early Holocene tree populations as dispersal nuclei for forest development in north-eastern European Russia', *Journal of Biogeography*, 38(5), pp. 922-932.

Van Loo, M., Joseph, J.A., Heinze, B., Fay, M.F. and Lexer, C. (2008) 'Clonality and spatial genetic structure in Populus× canescens and its sympatric backcross parent P. alba in a Central European hybrid zone', *New Phytologist*, 177(2), pp. 506-516.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. and Shipley, P. (2004) 'MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data', *Molecular Ecology Notes*, 4(3), pp. 535-538.

Vaughan, S.P., Cottrell, J.E., Moodley, D.J., Connolly, T. and Russell, K. (2007) 'Clonal structure and recruitment in British wild cherry (Prunus avium L.)', *Forest Ecology and Management*, 242(2-3), pp. 419-430.

Vitt, P., Havens, K., Kramer, A.T., Sollenberger, D. and Yates, E. (2010) 'Assisted migration of plants: Changes in latitudes, changes in attitudes', *Biological Conservation,* 143(1), 18-27.

Wagner, S., Liepelt, S., Gerber, S. and Petit, R.J. (2015) 'Within-range translocations and their consequences in European larch', *PloS one*, 10(5), p. e0127516.

Waits, L.P., Luikart, G. and Taberlet, P. (2001) 'Estimating the probability of identity among genotypes in natural populations: cautions and guidelines', *Molecular ecology*, 10(1), pp. 249-256.

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.-M., Hoegh-Guldberg, O. and Bairlein, F. (2002) 'Ecological responses to recent climate change', *Nature*, 416(6879), pp. 389-395.

Ward, J.A., Ponnala, L. and Weber, C.A. (2012) 'Strategies for transcriptome analysis in nonmodel plants', *American Journal of Botany*, 99(2), pp. 267-276.
Watts, K. (2006) 'British Forest Landscapes:The Legacy of Woodland Fragmentation', *Quarterly Journal of Forestry*, 100(4), pp. 273-279.

Watts, K., Humphrey, J.W., Griffiths, M., Quine, C.P. and Ray, D. (2005) *Evaluating biodiversity in fragmented landscapes: principles.* (FCIN073). Edinburgh: Forestry_Commission.

Watts, K., Ray, D., Quine, C.P., Humphrey, J.W. and Griffiths, M. (2007) *Evaluating biodiversity in fragmented landscapes: applications of landscape ecology tools* (FCIN085). Edinburgh: Forestry_Commission.

Weir, B.S. and Cockerham, C. (1984) 'Estimating F-statistics for the analysis of population structure', *Evolution*, 38(6), pp. 1358-1370.

Wesche, K., Ronnenberg, K. and Hensen, I. (2005) 'Lack of sexual reproduction within mountain steppe populations of the clonal shrub Juniperus sabina L. in semi-arid southern Mongolia', *Journal of Arid Environments*, 63(2), pp. 390-405.

Wesolowski, T. and Rowinski, P. (2006) 'Timing of bud burst and tree-leaf development in a multispecies temperate forest', *Forest Ecology and Management*, 237(1-3), pp. 387-393.

Wicksell, U. and Christensen, K.I. (1999) 'Hybridization among Tilia cordata and T. platyphyllos (Tiliaceae) in Denmark', *Nordic Journal of Botany*, 19(6), pp. 673-684.

Widén, B., Cronberg, N. and Widén, M. (1994) 'Genotypic diversity, molecular markers and spatial distribution of genets in clonal plants, a literature survey', *Folia Geobotanica*, 29(2), pp. 245-263.

Willis, K.J., Rudner, E. and Sümegi, P. (2000) 'The full-glacial forests of central and southeastern Europe', *Quaternary Research*, 53(2), pp. 203-213.

Willis, K.J. and Van Andel, T.H. (2004) 'Trees or no trees? The environments of central and eastern Europe during the Last Glaciation', *Quaternary Science Reviews*, 23(23), pp. 2369-2387.
Willis, K.J. and Whittaker, R.J. (2000) 'The refugial debate', *Science(Washington)*, 287(5457), pp. 1406-1407.

Wilson, S.M. and Samuel, C.J.A. (2002) 'Genetic conservation of native trees', *Forest Research Annual Report and Accounts for*, 2003, pp. 56-61.

Wolfe, J.A. and Wehr, W. (1987) *Middle Eocene dicotyledonous plants from Republic, northeastern Washington*.

Wright, S. (1931) 'Evolution in Mendelian populations', *Genetics*, 16(2), pp. 97-159.
Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S. and Li, S. (2014) 'SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads', *Bioinformatics*, 30(12), pp. 1660-1666.

Young, A., Boyle, T. and Brown, T. (1996) 'The population genetic consequences of habitat fragmentation for plants', *Trends in Ecology & Evolution*, 11(10), pp. 413-418.

Yousefzadeh, H., Hosseinzadeh Colagar, A., Tabari, M., Sattarian, A. and Assadi, M. (2012) 'Utility of ITS region sequence and structure for molecular identification of *Tilia* species from Hyrcanian forests, Iran', *Plant Systematics and Evolution*, 298(5), pp. 947-961.

Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R. and Simon, P. (2012) 'Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences', *American Journal of Botany*, 99(2), pp. 193-208.

Zerbino, D. and Birney, E. (2008) 'Velvet: Algorithms for de novo short read assembly using de Bruijn graphs', *Genome Research*, 18(5), pp. 821 - 829.