Relicts, Refugia and Reticulation: A Study of Population History, Hybrids and Phylogeny in the Long-Lived Flowering Tree Genus *Tilia*

Prattana Phuekvilai

Thesis submitted for the degree of Doctor of Philosophy

School of Biology
Newcastle University

July 2014

Declaration

I hereby certify that this thesis is the original work of my own. No part of the material has been submitted to the Newcastle University or any other University for obtaining a higher degree. All information derived from other sources has been acknowledged and indicated in the thesis.

Prattana Phuekvilai

July 2014

Abstract

Tilia L. (lime or basswood) is a genus of large trees that are widely distributed in the temperate regions of the Northern Hemisphere. *Tilia* is an under-investigated genus with unknown species relationships. Therefore, a phylogeny of the genus was reconstructed. This revealed disagreement of the phylogenetic placement of some species and also indicated extensive hybridization. To investigate this further, the most tractable and widely distributed species across Europe, *T. cordata* (Mill.) or small leaved lime and *T. platyphyllos* (Scop.) or large leaved lime, were selected for study. This study aims to increase the understanding of genetic diversity and hybridization between the two *Tilia* species. Also, to gain insight into postglacial recolonization in *Tilia* across Europe, the patterns of population genetic structure were investigated.

In order to achieve the goals, 15 microsatellite markers were developed for detailed genetic analysis. These loci clearly discriminated the two *Tilia* species. Cross-amplification results indicated that twelve microsatellite markers amplified polymorphic loci in 24 species in the genus. A high level of polymorphism was observed in twenty-five populations of *T. cordata* and 15 populations of *T. platyphyllos* from natural woods across Europe. The level of genetic diversity in *T. platyphyllos* is higher than in *T. cordata*. Both microsatellite and morphological analysis revealed that natural hybridisation and introgression have occurred between *T. cordata* and *T. platyphyllos* in sympatric UK populations, which could be of importance for adaptation and other evolutionary processes. The partial congruence of molecular and morphological analysis suggests that molecular markers are more reliable than morphological analysis for detecting hybridization.

The stronger genetic structure observed in *T. platyphyllos* than in *T. cordata* suggested that the migration and colonization in the northern areas of *T. cordata* occurred before those of *T. platyphyllos*. Microsatellite analysis suggested different possible colonization routes between the two *Tilia* species. However, *T. cordata* and *T. platyphyllos* seem to share the three main refugia in southern Europe (Iberia, Italy and the Balkans). In addition, *T. cordata* seems to have additional putative refugia in eastern areas. The haplotype network and some shared haplotypes of eight chloroplast regions indicate incomplete lineage sorting rather than recent hybridization.

Acknowledgements

Firstly, I would like to express my deep gratitude and thanks to my supervisor, Dr Kirsten Wolff for her supervision, encouragement and support throughout the completion of my study. Working with Kirsten for almost four years has been a great privilege. She has taught me how to be a population geneticist. Also, she gave me the opportunity to attend workshops and conferences, which helped me to broaden my perspective and provide me a great experience. She has supported me not only in the academic life, but also provided me opportunities to have experiences outside the University. She is an excellent supervisor who never hesitated to provide any kind of help. This project would never have been accomplished without her.

I would like to thanks Prof. Donald Pigott who helped in suggestion of species relationships and provided the samples of most *Tilia* species used in phylogenetic study. I would also like to express my gratitude to Dr. Paul Ashton and his colleages, who helped both in collecting the samples in the UK and writing the paper. I owe a great deal of thanks to people who helped me collecting the samples across Europe, without whom this study would not have been possible.

I am grateful to thanks all of the staff and friends in the School of Biology for their kindness, support and friendship. Special thanks go to Helen Martin for her help with cloning techniques and her advice. I appreciate the assistance of laboratory members, Dr. Nicholas Levsen and Samuel Logan for their useful discussion and assistance in my project.

I gratefully acknowledge the Royal Thai Government for providing me with a PhD scholarship. I would also like to thank to the staff at the Office of Educational Affairs (OEA) in London for their help and support. I also wish to thanks the members of the Thai Society in Newcastle for their sincere friendships.

Finally, I would like to express my deepest appreciation to my parents and sister for their love, support, understanding and encouragement. I would like to dedicate this achievement to them.

Publications Related to the Thesis

Publication:

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

Oral presentations:

Phuekvilai P (2012). Evolution and population genetics in the genus *Tilia*. *Ecological Genetics Group Conference*, Edinburgh, UK.

Phuekvilai P (2014). Population structure and phylogeography of *Tilia cordata* and *T. phythyllos* across Europe. *Ecological Genetics Group Conference*, Newcastle, UK.

Poster presentations:

Phuekvilai P, Wolff K (2011). Evolution and population genetics in the genus *Tilia*. *Ecological Genetics Group Conference*, London, UK.

Phuekvilai P, Wolff K (2012). Genetic diversity in ancient lime trees. *Trees Beyond the Wood Conference*, Sheffield, UK.

Phuekvilai P, Wolff K (2013). Genetics diversity and population structure in *T. cordata and T. platyphyllos*. *Ecological Genetics Group Conference*, Belfast, UK.

Table of contents

List of Figures

List of Tables

Chapter 1. General introduction

1.1 Genetic diversity

Genetic diversity is the fundament of biodiversity. It is commonly used to describe the amount of heritable variation within and between populations of organisms (Brown, 1983; Lowe *et al*, 2004). The processes that lead to genetic diversity are mutation and recombination. Selection, genetic drift and gene flow of alleles among different populations cause variation in the diversity within populations (Rao and Hodgkin, 2002). Genetic diversity is the basis for species survival and adaptation to environments. In addition, genetic diversity is important for species or populations to colonize new ecological niches (Crawford and Whitney, 2010; Dlugosch and Parker, 2008).

1.2 Factors influencing extent and distribution of genetic diversity

It is generally accepted that genetic diversity changes over time and in space (Loveless and Hamrick, 1984). The extent and distribution of genetic diversity in plants depends on various factors: ecological and geographical factors, breeding system, bottlenecks and also human factors (Rao and Hodgkin, 2002).

Different geographic locations usually differ in ecological characteristics such as latitude, altitude, temperature and moisture. These characteristics are significant in determining the distribution of genetic diversity and population structure. Climatic fluctuations during the post-glacial periods are believed to have had an important impact on genetic diversity, distribution ranges and differentiation in both plant and animal species that are living today (Hewitt, 2000). Under natural conditions, the habitats may define the characteristics of populations in which traits have evolved to survive. Even with small habitat differences, adaptive genetic variation often reacts with high sensitivity (Rao and Hodgkin, 2002).

The breeding system of a species is another factor that significantly affects the extent and distribution of genetic diversity (Loveless and Hamrick, 1984). Information about breeding system can provide an insight into the genotypic distribution of plants in populations. Outcrossing helps plant species to maintain a high level of genetic diversity

and results in gradual changes in allele frequency between populations. Outbreeding species usually display inbreeding depression in various forms, such as a decrease in seed set, germination, survival and growth. In contrast, inbreeding species usually exhibit a lower level of genetic diversity and a greater difference in alleles between different populations than outcrossing species (Rao and Hodgkin, 2002).

Bottlenecks can cause a dramatic loss of genetic variation. A population bottleneck occurs when population size is reduced, which could be the result of various events, such as an environmental disaster or human effects. In the population bottleneck, many alleles that were present in the original population are lost. Therefore, the remaining population has a low level of genetic diversity. The smaller the population size and the longer it remains small, the more genetic variation will be lost. The remaining population is faced with a high level of genetic drift, which can be described as the random change of allele frequencies in a population. Infrequently occurring alleles have a higher chance of being lost in a small population. The loss of genetic diversity in a new population can result in a population that is genetically distinct from the original population and this may accelerate the evolution of new species (William and Catton, 2009).

1.3 Tree diversity and their importance

Trees harbour the vast majority of the world's terrestrial biodiversity. An estimated 31% of earth surface is covered by forests (FAO, 2010) and trees make up approximately 90% of earth's biomass (Whittaker, 1975). In addition, forests also supply essential services, such as controlling water runoff, protecting soils and providing space for recreation. Estimates of global tree species range from 60,000 to 100,000 taxa. Over time, trees have developed various characteristics to compete with other organisms and have evolved to survive. These processes lead to the current diversity of trees. However, more than 10% of the world tree species are found to be under threat of extinction (Oldfield *et al*, 1998) due to various factors, for example, deforestation, overexploitation and impact of climate change. To avoid these harmful consequences and sustain plant biodiversity, knowledge of genetic diversity and evolution is required.

The assessment of genetic diversity and population structure provides basic information for increasing understanding of historical processes and evolution. Genetic diversity

also indicates whether populations can cope with changes in the environment, which may alter in an unpredictable way. This knowledge is a basic requirement for prioritizing populations or species that need conservation action to maintain the genetic resources.

One important issue that affects the evolutionary rates of plant species is their generation times. Plants with longer generation times, such as trees and shrub species, generally show slower rates of nucleotide substitution than those with shorter generation times. For example, in *Tilia*, the age of trees when they begin to flower and produce seed ranges from six to 40 years old (Pigott, 2012). The long generation time of tree species and low recruitment make them have far fewer generations compared to herbs and other smaller plants for the same periods of time. Therefore, the longer-lived trees and shrubs evolve more slowly. Thus, the study of evolution in some plants that have a long generation time and evolve slowly is hard, particularly regarding the differentiation of closely related species (Smith and Donoghue, 2008).

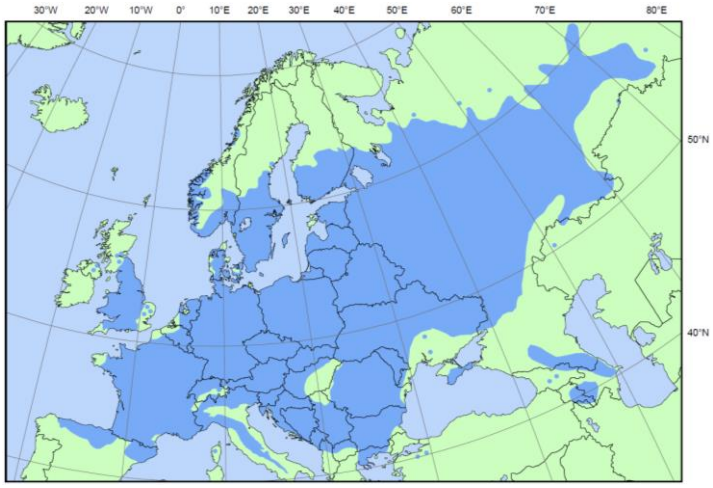
1.4 The genus *Tilia* and its distribution

Tilia L. is a genus of large trees with 23 species in the family Malvaceae of the order Malvales (Bremer *et al*, 2003). Most species are diploid (2n = 82), outcrossing and insect pollinated. These species readily reproduce vegetatively, both naturally and through human management (Pigott, 2012). *Tilia* species are distributed in the temperate regions of the Northern Hemisphere, throughout Europe, some parts of Asia and North America. *Tilia* is commonly known as lime in Britain and linden or basswood in North America. Within Europe, four species occur naturally, *T. cordata* (Mill.), *T. platyphyllos* (Scop.), *T. tomentosa* (Moench.) and *T. dasystyla* (Stev.). Among these species, *Tilia cordata* (small leaved lime) and *T. platyphyllos* (large leaved lime) are widely distributed throughout almost the whole of Europe. The distribution range of *T. platyphyllos* is more limited than that of *T. cordata* and it is rare in northern Europe (Figure 1.1 a and b). Based on pollen identification, *T. cordata* was a major constituent and even a dominant deciduous tree species of the woodlands until about 3000 BC (Pigott, 1991).

In the UK, *Tilia* was dominant in ancient woodland (Huntley and Birks, 1983). It has not been widely planted in woodlands and regenerates poorly from seed. Therefore, the

presence of *Tilia* indicates ancient woodlands (Pigott, 1969). *T. cordata* is common and naturally distributed in many parts of the UK, but only at low altitudes in the northern range in the Lake District and North Yorkshire (Pigott and Huntley, 1978). *T. platyphyllos* mostly occurs at the same latitudes as *T. cordata.* These two species can hybridise as they have overlapping flowering times. The hybrid, *T.* x *europaea* (common lime) has been planted in Europe including the UK since the Middle Ages (Pigott, 1991). They have been extensively planted in parklands and along roads in towns since the 17th century, therefore they are now widespread (Pigott, 2012). Figure 1.2 a-c shows the distribution of *T. cordata*, *T. platyphyllos* and their hybrid, *T.* x *europaea*, across the UK since 1930. This map shows all locations of planted as well as wild specimens, particularly inflating numbers of *T.* x *europaea*.

(a)



(b)

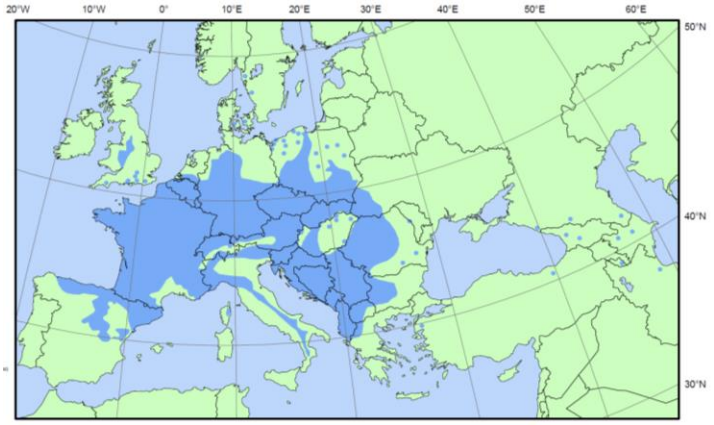


Figure 1.1 The natural distribution area of *T. cordata* (a) and *T. platyphyllos* (b) in 2008. Map downloaded from EUFORGEN, http://www.euforgen.org

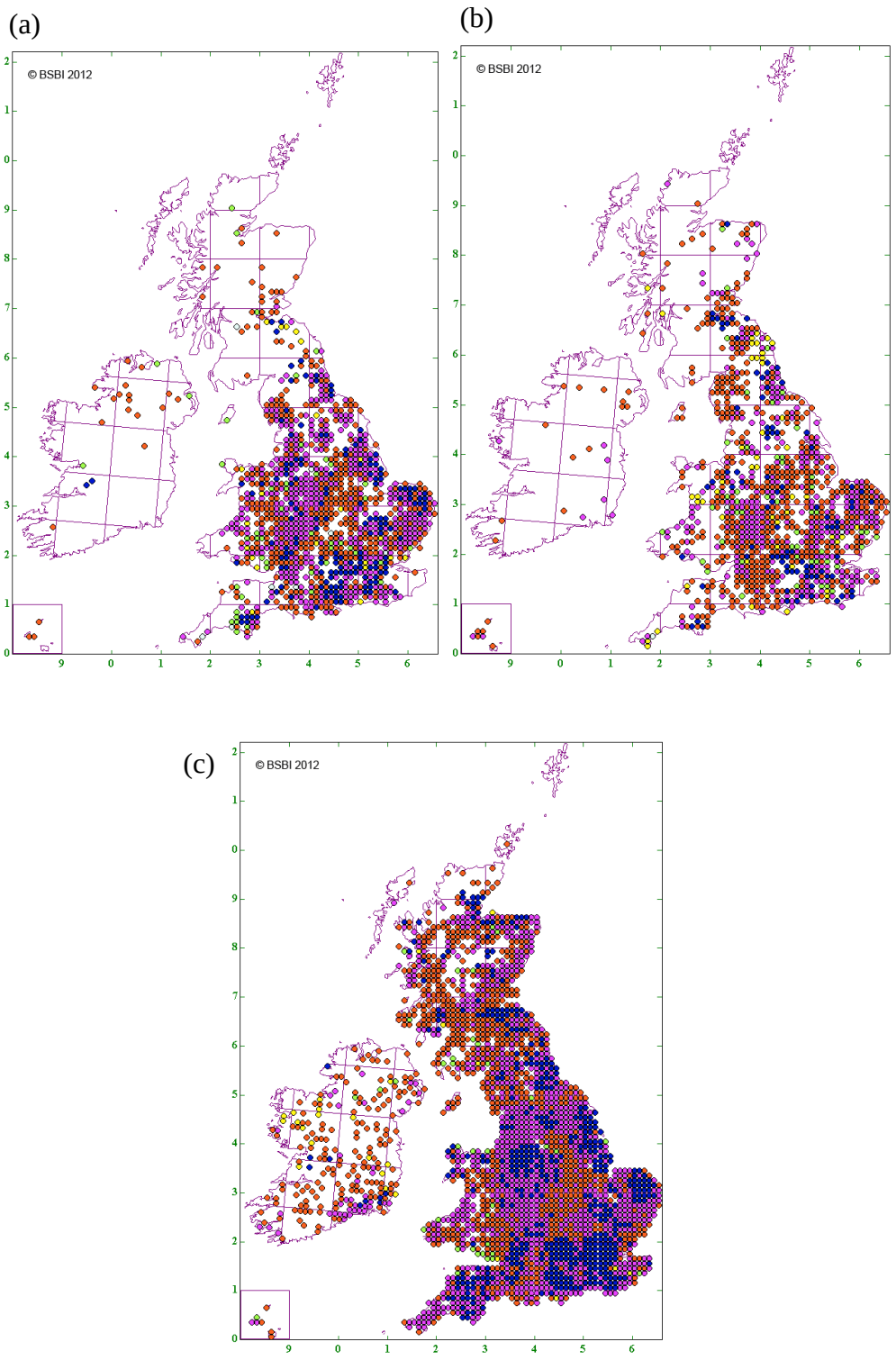


Figure 1.2 Distribution of *T. cordata* (a), *T. platyphyllos* (b) and *T.* x *europaea* (c) in Great Britain and Ireland. Different colours of dots indicate recorded date ranges; ○ - 1930, ○ 1930-1969, ○ 1970-1986, ○ 1987-1999, ○ 2000-2009 and ○ 2010 – now. Map downloaded from Online Atlas of the British and Irish Flora, http://www.brc.ac.uk

The distribution of *Tilia* across Europe has decreased due to human impact and climatic conditions. *Tilia* pollen disappeared from many sites from about 3000 BC, when humans began farming the areas, deforested large areas and managed the woodlands, for example by coppicing or pollarding for fodder. The analysis of pollen evidence indicated the significant decrease in *Tilia* pollen while the pollen of grasses and weeds were increasing during those periods (Pigott, 1991). In addition, Pigott (1991) stated that *T. cordata* is sensitive to low temperature during seed set and germination. It is suggested that the colder climate since the mid-Holocene has been the reason for the reduced ability of *Tilia* to set seed regularly (Huntley and Birks, 1983). In northern Europe *Tilia* are thought to have disappeared in many places because of the low seed fertility (Svejgaard Jensen, 2003).

1.5 Importance and uses of *Tilia*

1.5.1 Biodiversity value

Tilia is an important species for biodiversity, in particular in mixed woods where it is accompanied by diverse plant, insect and fungal species. *Tilia* leaves are popular food for several species of moth caterpillars, such as the lime hawk, peppered, vapourer, triangle and scarce hook-tip moth. The flowers of *Tilia* also provide nectar for honey bees and pollen for pollinators. Old, hollow trunks and fallen trees provide habitats for many wood-boring insects, such as stag beetles (Royal Forestry Society, 2014). The scarce lime bark beetle (*Ernoporus tiliae)* is a lime dependent species as it is believed to feed almost exclusively on *T. cordata*. This beetle is recorded as an endangered species in the Red Data Book Category 1 and has been included on the priority UK Biodiversity Action Plan (UKBAP) list (Broome *et al*, 2004). *Tilia* can also be planted for soil improvement. Its leaves contain many minerals and when they have fallen in autumn they are favoured by earthworms. The leaves feed the soil organisms and take short periods (a few months) to decompose and turn into humus, compared to other tree species. For example, beech leaves take up to two years to decompose. Thus, *Tilia* improves the nutrient status of soils (Royal Forestry Society, 2014).

1.5.2 Human value

Tilia trees, especially *T.* x *europaea*, are important as ornamental trees and are frequently used when deep shade is desired, particularly in parks and streets (Pigott, 2012). The wood of *Tilia* is soft and light coloured with fine texture. These properties make it popular for model building, intricate carving and furniture making. Almost all parts of the tree can be used for fodder (Svejgaard Jensen, 2003). In the past, *Tilia* was used as a coppice species as it can produce long straight poles and the inner bark (bast) was used for making ropes. *Tilia* flowers were also important for honey production. These products have probably led to the wide spread of *Tilia* species during the Middle Ages. The flowers of *Tilia* are still important for producing a richly flavoured honey and its dried flowers can also be used for herbal tea. *Tilia* flower tea has a pleasant taste, due to the aromatic volatile oil. The flowers contain various active ingredients, such as flavonoids that act as antioxidants and mucilaginous constituents that reduce inflammation (Bradley, 1992).

1.6 Approach and aims of research

Tilia is an interesting genus to study because very little research has been carried out to evaluate the evolutionary history and genetic diversity of the genus. Also, species relationships within the genus are unknown and are complicated by frequent hybridization. So far only morphological characters and karyotypes of the whole species in the genus have been studied (Pigott, 2012). In addition, few molecular studies have been carried out, and only with a restricted number of species. In this study, phylogeny is analysed to increase understanding of the relationships and evolutionary history among the species in this genus.

In order to clarify relationships of *Tilia* species, both chloroplast and nuclear regions were sequenced to provide data. The chloroplast genome is encoded in circular DNA and is non-recombinant and maternally inherited in most angiosperms. The multiple copies of chloroplast DNA and the availability of universal PCR primers facilitate PCR amplification of chloroplast DNA regions. It is commonly used in studies assessing variation at various levels from population, species, genera and higher taxonomic levels (Small *et al*, 2004). Over the last decade, nuclear genes have increasingly been used for assessing genetic variation as well as chloroplast DNA. One of the primary advantages

of nuclear genes for phylogenetic analysis is the higher rate of sequence evolution than in organellar genomes. The substitution rates of nuclear genes can be up to five times greater than those of chloroplast genes (Gaut, 1998). This advantage provides a useful source of DNA sequence data for understanding the relationships among low taxonomic levels, particularly when universal markers, such as chloroplast DNA primers, are unable to provide phylogenetic resolution. More details of chloroplast and nuclear DNA are discussed in Chapter 2.

The resulting phylogeny reveals disagreement of the phylogenetic placement of some species. This indicates extensive hybridization among those *Tilia* species. To investigate this further, the most tractable and widely distributed species across Europe, *T. cordata* and *T. platyphyllos*, were selected for study. The hybridization and introgression between these two species was assessed. Also, their genetic diversity and population structure were investigated to gain insight into postglacial recolonization in *Tilia* across Europe. In order to achieve these goals, microsatellite markers needed to be developed, as they were not available for *Tilia* species.

Microsatellites are sets of non-coding repetitive DNA sequences found abundantly in the genome of most taxa. Microsatellite markers are one of the most popular markers used for genetic studies. The high mutation rate of microsatellites makes them more informative among closely related species compared to other markers. In addition, microsatellite markers are co-dominant, so that homozygotes and heterozygotes can be identified. This allows the use of microsatellite markers to genotype individuals, assess parentage or detect hybrids. The combination of microsatellite markers produces a DNA fingerprint which is individual specific. As microsatellite markers provide the highest level of polymorphism, they are commonly used for the study of local patterns (Ellegren, 2004). Microsatellites are discussed in more detail in Chapter 3.

The general details and aims of each chapter are as below:

In Chapter 2, the phylogeny of the genus *Tilia* was reconstructed using chloroplast and nuclear regions to improve understanding of the relationships and evolutionary history among the species.

In Chapter 3, details of microsatellite development were given. In addition, these markers were tested for polymorphism and their transferability to other species in the genus.

In Chapter 4, microsatellite markers developed in Chapter 3 were used to assess the genetic diversity between and within *T. cordata* and *T. platyphyllos* from natural populations across Europe.

In Chapter 5, patterns of population genetic structure of *T. cordata* and *T. platyphyllos* across Europe were investigated using microsatellite markers and chloroplast regions. Also, the observed geographical patterns in terms of recolonisation after postglacial periods were interpreted and possible refugia of the two *Tilia* species were identified.

In Chapter 6, microsatellite markers and morphology were used to examine the extent of hybridisation between *T. cordata* and *T. platyphyllos* in sympatric populations in the UK.

In Chapter 7, the findings of this study are discussed in a wider context.

Chapter 2. Phylogenetic analysis of the genus *Tilia*

2.1 Abstract

Tilia L. (Malvaceae) is a genus of twenty-three species of trees with unknown phylogenetic relationships. Fifty-five accessions, including twenty-two species, as well as subspecies and hybrids were sequenced for seven chloroplast regions and three low-copy nuclear regions. Phylogenetic analysis revealed little resolution because of low sequence divergence, particularly of chloroplast DNA. Of the three nuclear regions, *Grx* was the most variable and provided the best resolution of the phylogeny among *Tilia* species. Although many species relationships are not well resolved, several relationships between species were apparent. The two American species, *T. americana* and *T. caroliniana*, formed a monophyletic clade, while European species are intermixed with Asian species. Interestingly, the two species that hybridise readily and have a similar geographical distribution in Europe, *T. cordata* and *T. platyphyllos*, seem not to be sister species. Some incongruence between phylogenetic trees suggested incomplete lineage sorting and supported extensive hybridisation among *Tilia* species.

2.2 Introduction

Phylogenetics is the study of evolutionary relationships among groups of organisms. Phylogenetic relationships are the foundation of evolutionary biology and other disciplines, such as biodiversity and biogeography. During recent decades, molecular sequence data have been widely used for examining plant evolutionary history, ranging from the close relationships between individuals to the relationships among genera in angiosperms (land plants) (Zhang *et al*, 2012). Among the molecular sequence data, three classes of DNA sequence data (chloroplast regions, nuclear ribosomal DNA and low-copy nuclear genes) have been widely reported for resolving plant relationships. The inclusion of DNA sequence data from both maternally inherited genome and biparentally inherited genome often provides robust phylogenetic relationships (Mort *et al*, 2007; Small *et al*, 2004).

2.2.1 Phylogenetic utility of chloroplast DNA

Chloroplast DNA (cpDNA) sequences are reliable sources of molecular markers for phylogenetic studies. They are widely used for phylogeny reconstruction in plants. There are many properties of chloroplast DNA that make it an attractive tool. The presence of multiple copies of chloroplast DNA is a significant advantage to facilitate PCR amplification of specific regions. In addition, the structural stability of chloroplast has aided the design and use of universal PCR primers. Although there are some mutation during evolutionary divergence, such as insertion, deletion and inversion, the overall structure of gene order and content are consistent (Small *et al*, 2004).

The other important features of the chloroplast DNA molecule are its uniparental inheritance and the fact that it is assumed to be non-recombinant as it is a haploid genome. The latter property simplifies the phylogenetic analysis because of the absence of allelic variation within individuals. However, due to its uniparental inheritance, chloroplast DNA sequence reveals only one parent, usually the maternal in angiosperms. Therefore, chloroplast DNA analysis of polyploids or hybrids may not reveal their full history. Hybrids or polyploid species will be in one of the two parental clades (Small *et al*, 2004).

In angiosperms, a large number of different chloroplast regions have been investigated. Generally, the plant chloroplast genome evolves at a slower rate than the nuclear genome (Gaut, 1998). The variation rate varies among different regions of the chloroplast genome. Chloroplast protein coding regions, such as *rbcL*, *matK* and *atpB*, evolve more slowly than non-coding regions (introns and intergenic spacers) (Small *et al*, 2004). Therefore, coding DNA regions are best suited to reconstruct phylogenetic relationships among high level taxa, at family level or above, but may not useful for closely related species (Chase *et al*, 1993; Soltis *et al*, 1999). Among coding regions of the chloroplast genome, *matK* is one of the most rapidly evolving sections and has ubiquitous presence in plants (Hilu and Liang, 1997), thus it has been widely used as a marker for plant phylogeny (Vijayan and Tsou, 2010). The non-coding sequences of introns, such as in *trnL*, *rpL16*, *trnK*, and intergenic spacers, such as in *psbD-trnT*, *psbJ-petA*, *trnL-trnF*, and *trnS-trnG*, are frequently used to reconstruct the relationships at lower taxonomic levels (Taberlet *et al*, 1991). However, due to the relatively slow rate of chloroplast evolution, sometimes even non-coding sequences fail to provide enough phylogenetic information in closely related taxa (Small *et al*, 1998).

2.2.2 Phylogenetic utility of nuclear ribosomal DNA

Since the 1990s, one of the most commonly used markers for phylogenetic studies of plant taxa is the internal transcribed spacers (ITS1 and ITS2) of the ribosomal DNA (rDNA) (Baldwin *et al*, 1995). The faster evolution in the rDNA region can overcome the lack of informative characters due to the slow evolution of cpDNA and rDNA is still as commonly used as cpDNA (Small *et al*, 2004). The rDNA is a region in the nuclear genome, so it is biparentally inherited. This property can overcome the limitation of uniparental inherited chloroplast genome and is useful for unravelling hybrid formation and parentage of polyploids (Baldwin, 1992; Fehrer *et al*, 2007). In addition, the fast rate of evolution of ITS rDNA compared to organellar genes is advantageous for resolving phylogenetic relationships of closely related species (Mort *et al*, 2007).

There are several other beneficial properties of the ITS region to use for phylogenetic inference. The nuclear ribosomal gene is composed of 18S-5.8S-26S rDNA and the ITS region is located in between these coding genes. The ITS priming sites are highly conserved. This allows amplification using universal primers for plants and fungi (White *et al*, 1990). In addition, this ribosomal gene exists in tandem arrays with hundreds to thousands of copies in plant genomes. Also, the ITS sequences in angiosperms are small (500-700 bp). These properties make the ITS region easy to isolate, even from ancient materials, and it can amplified by a standard PCR reaction (Alvarez and Wendel, 2003).

The copy numbers of ITS are expected to be homogenized through a phenomenon termed concerted evolution. This phenomenon occurs when the different sequences of individual copies in the genome are homogenised to the same sequence type by a mechanisms of unequal-crossing over or gene conversion (Alvarez and Wendel, 2003; Elder and Turner, 1995). If the process of concerted evolution is complete, PCR amplification and subsequent sequencing can be easily performed. However, concerted evolution may also be a confounding effect. In principle, concerted evolution would act to eliminate paralogous sequences, which facilitate the inference of homology among taxa. If concerted evolution is incomplete, different rDNA copies can be present across the genome and these constitute orthologs and paralogs. This phenomenon makes the analysis of phylogeny more complicated or inconclusive (Alvarez and Wendel, 2003; Mort *et al*, 2007).

2.2.3 Phylogenetic utility of nuclear genes

Over the last decade, nuclear genes are increasingly being used for phylogenetic reconstruction, particularly when universal markers, such as, cpDNA and ITS rDNA are unable to provide phylogenetic resolution. One of the primary advantages of nuclear genes for phylogenetic analysis is the higher rate of sequence evolution than organellar genomes. The substitution rates of nuclear genes can be up to five times greater than those of chloroplast genes (Gaut, 1998). This advantage provides a useful source of DNA sequence data for understanding the relationships among low taxonomic levels. For example, nuclear-encoded alcohol dehydrogenase (*adhC*) sequences were fully resolved the relationship among *Gossypium* L. species, whereas non-coding cpDNA showed incomplete resolution (Small *et al*, 1998). Another advantage of nuclear genes is that they are biparentally inherited, unlike the maternal inherited chloroplast genome. Also, nuclear genes are not subject to concerted evolution like the ITS rDNA, thus they are ideal candidates for detection of the origin of hybrids or polyploids (Calonje *et al*, 2009; Small *et al*, 2004).

However, there are also some difficulties in using nuclear regions for phylogenetic studies. The identification of DNA regions that are easily amplifiable and evolving fast enough to provide sufficient variation are a challenge. In eukaryotes, nuclear genes tend to exist in gene families, which consist of multiple copies of homologous genes due to duplication. This makes it difficult to identify orthologous from paralogous genes. Thus, nuclear genes with a low copy number or ideally a single copy are preferred. The lack of universal PCR primers of nuclear genes, as opposed to chloroplast marker and ITS rDNA, also hamper their phylogenetic utility (Calonje *et al*, 2009; Sang, 2002). Although nuclear regions usually require additional work, such as, primer design, PCR cloning and optimisation, it is worth the effort to increase the quality of phylogenetic reconstruction and to clarify the evolutionary processes of the nuclear genome.

2.2.4 Nuclear genes used in this study

Among nuclear genes, the nitrate reductase (*Nr*) gene is one of the low copy nuclear genes found in plant species, fungi and algae (Zhou and Kleinhofs, 1996). This gene catalyzes the reduction of nitrate to nitrite, facilitating the uptake of nitrogen from soil in plants. This gene has four exons and three introns in conserved positions in all

angiosperms studied (Salanoubat and Ha, 1993). The sequence of the *Nr* gene has been used for studies of phylogenetic relationships among closely related wild potatoes and for inferring the relationships of diploids and polyploids (Rodriguez and Spooner, 2009). Also, the relationships among *Pistacia* species have been resolved using this nuclear region (Yi *et al*, 2008). The comparison of *Nr* sequence data and granule-bound starch synthase I (GBSSI), ITS and the cpDNA sequence *trnT-trnF* in American Lycieae (Solanaceae) revealed that *Nr* sequence data had the most parsimony informative characters (Levin *et al*, 2009). The high rate of evolution of this low copy nuclear gene may make it a useful region for evolutionary studies of relationships among species.

The *WRKY* genes are a gene family of transcription factors. They are involved in several pathways, such as regulating starch metabolism, seed development and the responses of biotic and abiotic stress in plants (Luo *et al*, 2005). The *WRKY* genes contain one or two highly conserved DNA binding domains with an intron. The independent loci of the *WRKY* genes have been clearly identified in various species, such as *Theobroma cacao* and *Persea americana* P. Mill. using locus-specific primers. These *WRKY* genes are easy to isolate and provide a high level of variation. They have been used for phylogenetic reconstruction in some species of Malvaceae and can be candidate regions of single and low-copy nuclear genes for phylogenetic reconstruction of low taxonomic levels (Borrone *et al*, 2007). Thus some *WRKY* loci are used in this study.

Another nuclear gene used in this study is the ATP synthase subunit β (*atpB*) gene. This gene encodes the β-subunit of plastid ATP synthase (ATPase), which is an important enzyme that provides energy for the cells by catalysing ATP synthesis. From the identification of informative markers among 141 low copy nuclear regions in two groups of rosids, *atpB* was identified as a phylogenetically informative region in one of the two rosid groups studied, *Psiguria* (Steele *et al*, 2008). The other nuclear gene is the glutaredoxin (*Grx*) gene. This is a gene family coding for the oxidoreductase enzymes involved in a variety of cellular processes. One of the most documented functions of this gene in plants is its involvement in oxidative stress responses (Rouhier *et al*, 2006). The phylogenetic utility of this gene is unknown as this study is the first time the *Grx* gene is used for phylogenetic reconstruction.

The analysis of these multiple evolutionarily independent regions of nuclear genes should provide a picture of relationships between the species (species tree) rather than

using a single gene, which reveals how that gene has evolved in a group of species (gene tree).

2.2.5 Taxonomic history of *Tilia*

Tilia species are widely distributed across all continents. The genus consists of 23 species and is divided into three groups based on geographical distribution; European and western Asian species, north and central American species and eastern Asian species (Pigott, 2012) (Table 2.1). Four species are native to Europe and western Asia (*T. cordata*, *T. dasystyla*, *T. platyphyllos* and *T. tomentosa*). Some species are divided into subspecies based on the variation in morphological characters between populations and altitudinal and ecological distribution. For example, *T. platyphyllos* has been divided into three subspecies with a different geographical distribution and some different morphology of hairs on leaf surface and size and shape of fruit (Pigott, 2012). In North America, there are two *Tilia* species, *T. americana* and *T. caroliniana* with three subspecies. Seventeen species are recognized in eastern Asia. Three of these species, *T. amurensis*, *T. paucicostata* and *T. tuan*, are divided into subspecies as described in Table 2.1. Some species are native to and found in a specific location, such as *T. kiusiana* , which is found only in southern Japan, while *T. maximowicziana* occures in northern Japan (Pigott, 2012).

In the genus *Tilia*, the basic chromosome number is 41. Almost all European and American species are diploid (2n = 82), except for one species, *T. dasystyla*, which is tetraploid (2n = 164). The eastern Asian species are diploid, tetraploid and also octaploid. *T. amurensis* is both diploid and tetraploid and *T. nobilis* is octaploid (2n = 328).

Extensive hybridisation among species within the genus *Tilia* has been recognized and reported (Pigott, 2012). The evidence of intermediate morphology of trees in mixed populations of two *Tilia* species has been investigated. Among European and western Asian species, the hybrid (*T.* x *europaea*) between *T. cordata* and *T. platyphyllos* has been widely studied. This hybrid was extensively cloned and frequently planted in England. In addition, *T. cordata* can hybridise naturally with the other species native in Europe, *T. tomentosa* (*T.* x *jaranyana* Simonk.) and *T. dasystyla* subsp. *caucasica*. Hybridisation with species in the US (*T. americana*) has been investigated as well. Another example of hybridisation between a European and western Asian species is between *T. platyphyllos* and *T. tomentosa*. This hybrid (*T.* x *haynaldiana* Simonkai) has

been observed in Hungary and in the Balkans (Pigott, 2012). All taxa of the US species can hybridise, both between species (*T. americana* and *T. caroliniana*) and between subspecies of *T. caroliniana*. Hybridisation has also been recognised in numerous species in eastern Asia. For example, in Japan hybrids of *T. japonica* with both *T. maximowicziana* and *T. kuisiana* were reported. In China, trees with intermediate morphology were investigated in mixed populations of *T. chinensis* and *T. paucicostata* and of *T. collidonta* and *T. nobilis* (Pigott, 2012). Some hybrids were treated as a new species due to the formation of a species with higher ploidy level. For example, the hybrid between *T. oliveri* (diploid, 2n = 82) and *T. miqueliana* (diploid, 2n = 82) is tetraploid (2n = 164). The hybrid would be allotetraploid, thus this hybrid is treated as a new species (*T. concinna*) (Table 2.1) (Pigott, 2012). Polyploidisation could lead to the blurring of species boundaries and make the species identification even more complicated. In the genus *Tilia*, there are two classes of polyploid species. Autopolyploidy is found in some species, such as *T. maximowicziana* (2n = 164) and *T. nobilis* (2n = 328), while allopolyploidy is found in some hybrids, such as *T.* x *euchlora* (2n = 164) (Pigott, 2012).

Table 2.1 Currently recognized species and subspecies of *Tilia* and their ploidy level, divided by geographical regions, based on Pigott (2012).

Region	Species	Ploidy level
Europe and western Asia	*T. cordata* Mill.	2n = 82
	- subsp. *sibirica* (Bayer) Pigott	2n = 82
	T. dasystyla Steven	2n = 164
	- subsp. *caucasica* (v. Engl.) Pigott	2n = 164
	- subsp. *multiflora* (Ledeb.) Pigott	2n = 164
	T. platyphyllos Scop.	2n = 82
	- subsp. *cordifolia* (Besser) C.K. Schneid.	
	- subsp. *pseudorubra* C.K. Schneid.	
	- subsp. *corinthiaca* (Bosc ex K. Koch) Pigott	
	T. tomentosa Moench	2n = 82
North America	*T. americana* L.	2n = 82
	T. caroliniana Mill.	2n = 82
	- subsp. *floridana* (Small) E. Murray	2n = 82
	- subsp. *heterophylla* (Vent.) Pigott	2n = 82
	- subsp. *occidentalis* (Rose) Pigott	2n = 82
Eastern Asia	*T. endochrysea* Hand.-Mazz.	2n = 82
	T. henryana Szyszyl.	2n = 164
	T. amurensis Rupr.	2n = 82, 164
	- subsp. t*aquetii* (C.K. Schneid.) Pigott	2n = 82
	T. japonica (Miquel) Bayer.	2n = 82
	T. kiusiana Shiras.	2n = 82
	T. mongolica Maxim.	2n = 164
	T. paucicostata Maxim.	2n = 164
	- subsp. *dictyoneura* (V. Engler) Pigott	
	- subsp. *yunnanensis* (Diels) Pigott	
	T. collidonta Chang H.-t.	2n = 164
	T. chinensis Maxim.	2n = 164
	T. chingiana Hu&Cheng	2n = 82
	T. concinna Pigott	2n = 164
	T. mandshurica Maxim.	2n = 82
	T. maximowicziana Shiras.	2n = 164
	T. miqueliana Maxim.	2n = 82
	T. nobilis Rehder&Wilson	2n = 328
	T. oliveri Szyszyl.	2n = 82
	T. tuan Szyszyl.	2n = 164
	- subsp. *tristis* Pigott	
	- subsp. *oblongifolia* (Rehder) Pigott	

There has been little taxonomic work done on the genus *Tilia* using molecular techniques. RAPD analysis was used for assessing the systematic relationships within some *Tilia* species, mainly those native in Europe and including one American species (*T. americana*) and one eastern Asian species (*T. henryana*). Also, hybrid clones (*T.* x *europaea* and *T.* x *euchora*) were investigated to clarify the parental species (Liesebach

and Sinko, 2008). Chloroplast regions (*rpL32-trnL* and *ndhF-rpL32*) have been used to construct the phylogeny of a subset of species in the genus *Tilia*. The results revealed that the US species appear to be monophyletic, while European and Asian species are intermixed within the same clade (McCarty, 2012). Analysis of ITS and 5.8S rDNA gene sequences and their secondary structure has been carried out to infer the phylogeny of *Tilia* species growing in northern Iran (Yousefzadeh *et al*, 2012). An unpublished study by Li *et al* (2002) of 12 diploid taxa of *Tilia* species based on a non-coding intron of the nitrate reductase (*NIA*) gene has shown that *T. platyphyllos* is not closely related to both *T. cordata* and any American species. This is the only study using the low copy sequences in the genus *Tilia*. In addition, all these previous studies have been done with a limited number of species. This study is the first in which almost all *Tilia* species (only one species has been missed out, *T. paucicostata*) have been analysed using both chloroplast and nuclear regions. The objective of this study is to reconstruct the phylogeny of the genus *Tilia* to improve understanding of the relationships and evolutionary history among the species.

2.3 Materials and methods

2.3.1 Species sampling

A total of twenty-two species including subspecies and some hybrids of *Tilia* (55 individuals) was analysed (Table 2.2). These species cover almost all species in this genus based on species identification by Pigott (2012). Only one species, *T. paucicostata* Maxim is not included in this study due to unsuccessful DNA extraction from herbarium specimen and no live specimen was located. Fresh leaves of most species were collected from a living collection in Cartmel, Cumbria, United Kingdom (Pigott, 2002). Two species (3 samples) of *T. chinensis* and *T. miqueliana* were obtained from Peasmarsh Place Arboretum, Rye, UK and one species of *T. tomentosa* 'Petiolaris' was obtained from the Royal Botanical Garden Edinburgh, UK (Table 2.2). No voucher specimens have been deposited because only leaves from ground level were available. Among 55 individuals, 31 are European and western Asian species, eight are the US species and the remainders are native to eastern Asia. *Theobroma cacao* was selected as an outgroup due to the placement of *Tilia* in the same clade in the family Malvaceae, based on the analysis of *rbcL* sequence data (Alverson *et al*, 1998) and the availability of an annotated genome sequence of *Theobroma cacao*.

Table 2.2 List of *Tilia* species, provenance, region and label used in this study.
* indicates samples obtained from Peasmarsh Place Arboretum, Rye, UK, ** indicates samples obtained from the Royal Botanical Garden Edinburgh, UK.

Species	Label	Provenance
T. cordata	Pg35	Brigsteer, Kendal, UK
T. cordata	Pg36	Borrowdale, Cumbria, UK
T. cordata	Pg41	Coniston Water, Cumbria, UK
T. cordata	Pg12	Burton Wood, Aughton, Lancs, UK
T. cordata	Pg46	Hesketh Wood, Cartmel, UK
T. cordata	Pg01	Earls Colne, Essex, UK
T. dasystyla subsp. *caucasica*	Pg03	Lagodekhi, Georgia, UK
T. dasystyla subsp. *dasystyla*	Pg04	Yalta, Krim, UK
T. platyphyllos	Pg45	Longlands, Cartmel, UK
T. platyphyllos	Pg48	Yewbarrow, Grange, UK
T. platyphyllos	Pg50	Wood Broughton old park, UK
T. platyphyllos	Pg37	Bohinj, Slovenia
T. platyphyllos subsp. *cordifolia*	Pg02	Cambridge, UK
T. platyphyllos subsp. *cordifolia*	Pg15	Sutton, Sussex, UK
T. platyphyllos subsp. *corinthiaca*	Pg14	Peloponnese, Greece
T. tomentosa	Pg13	Crna Gora, Montenegro
T. tomentosa	Pg40	Chepelarska gorge, Rodopi, Bulgaria
T. tomentosa 'Orbicularis'	Pg39	Wisley, Surrey, UK
T. tomentosa 'Petiolaris'	Pg38	Cambridge, UK
T. tomentosa 'Petiolaris'**	RBGE	-
T. x europaea 'Hatfield'	Pg44	Longlands, Cartmel, UK
T. x europaea 'Hatfield'	Pg47	Yewbarrow, Grange, UK
T. x europaea 'Pallida'	Pg43	Longlands, Cartmel, UK
T. x europaea 'Pallida'	Pg53	Longlands, Cartmel, UK
T. x euchlora	Pg49	Longlands, Cartmel, UK
T. endochrysea	Pg51	Guangdong, China
T. henryana	Pg24	Arboretum national des Barres, France
T. amurensis	Pg08	Arnold Arboretum, UK
T. amurensis	Pg31	Amur valley, Primorsky Krai, Siberia
T. amurensis subsp. *amurensis*	Pg10	Ussuriysk, Primorsky Krai, Siberia
T. amurensis subsp. *Taquetii*	Pg06	Ussuriysk, Primorsky Krai, Siberia
T. amurensis subsp. *Taquetii*	Pg32	Ussuriysk, Primorsky Krai, Siberia
T. japonica	Pg33	Hokkaido, Japan
T. japonica	Pg25	Kew, originally from Japan
T. kiusiana	Pg23	Kyushu, Japan
T. monogolica	Pg05	Beijing area, China
T. callidonta	Pg29	Longxi, Sichuan, China
*T. chinensis**	960	-
*T. chinensis**	986	-
T. chingiana	Pg26	Lashan, Jiangxi, China
T. concinna	Pg30	Xuzhou, Jiansu, China
T. mandshurica	Pg11	Beijing area, Hebei, China
T. maximowicziana	Pg28	Sendai, Honshu, Japan
T. maximowicziana	Pg34	Saporo, Hokaido, Japan
*T. miqueliana**	1752	-
T. nobilis	Pg07	Emei shan, Sichuan, China
T. oliveri	Pg27	Hubei, China

Table 2.2 Continued.

Species	Label	Provenance
T. tuan	Pg09	Hubei, China
T. tuan	Pg22	Yunnan, China
T. americana	Pg19	Kane Co, Illinois, US
T. americana	Pg20	Whately, Massachusetts , US
T. caroliana subsp. *floridana*	Pg17	Florida Caverns, Mariana, Florida US
T. caroliana subsp. *floridana*	Pg18	Monterey, Sierra Madre, Mexico
T. caroliniana subsp. *caroliniana*	Pg16	Carolina border, US
T. caroliniana subsp. *occidentalis*	Pg52	Cinconchiaco, Mexico

2.3.2 DNA extraction

Total genomic DNA was extracted from dried *Tilia* leaves using a CTAB (cetyltrimethyl ammonium bromide) procedure (Morgan-Richards and Wolff, 1999). In short, approximately 1 cm^2 leaves were ground in a mortar containing 800 µl of 2X CTAB, 1% of β-mercaptoethanol and a pinch of insoluble PVP (polyvinylpyrrolidone). The DNA solution was then transferred into a tube and incubated at 60 ^oC in heating block for 15-30 minutes with occasional mixing. After that, the tube was placed at room temperature for cooling before adding 600 µl of chloroform: isoamyl alcohol (24:1). The solution was mixed with a vortex mixer before being centrifuged at 13,000 rpm for 5 minutes. The supernatant was transferred into a new tube and chloroform: isoamyl alcohol was added again. After centrifugation, supernatant was transfered into a fresh tube, and 0.7 volume of cold isopropanol (approximately 560 µl) was added and incubated at -20 ^oC for 15-30 minutes to precipitate DNA. The tube was centrifuged at 13,000 rpm for 10 minutes and the supernatant was discarded. The pellet was washed with 1 ml of 70% ethanol and again centrifuged for 10 minutes (13,000 rpm). After this, the pellet was dried at 30 ^oC in the incubator and resuspended in 100 µl of TE buffer. The extracted DNA was stored at -20 ^oC until use.

2.3.3 Chloroplast amplification and sequencing

For all 22 *Tilia* species in this investigation, a total of seven regions (Table 2.3) were polymerase chain reaction (PCR) amplified and sequenced. PCR amplifications were carried out in 15 µl reactions containing 5 ng of template DNA, 1x reaction buffer, 2mM $MgCl_2$, 0.2 mM of each dNTP, 0.4 mM of each primer and 0.5 U of *Taq* DNA polymerase (Bioline). The PCR reactions were amplified as follows: an initial pre-

denaturation step at 94 °C for 3 min, followed by 30-35 cycles of denaturation step at 94 °C for 15 s, annealing temperature and time as described in Table 2.3, and extension at 72 °C for 60 s, and a final extension step of 72 °C for 4 min.

PCR products were cleaned with Exo-SAP (USB Corporation, Cleveland, Ohio, USA). The following amounts of reagents were added into each sample: 1 µl of Shrimp Alkaline Phosphatase (SAP), 1.35 µl of Shrimp Alkaline Phosphatase buffer and 0.15 µl of Exonuclease I. The mixed reactions were incubated in a thermocycler at 37 °C for 40 minutes and 80 °C for 15 minutes. The purified PCR products were sequenced using ABI Prism BigDye® Terminator Version 3.1 Cycle Sequencing Kits (Applied Biosystem, Foster City, California, USA). Sequencing reaction were performed in 10 µl containing 1.5 µl of 5X sequencing buffer, 0.5 pmol of chloroplast primer, 0.5 µl of BigDye terminator sequencing mix and 0.5 µl of purified PCR product. The sequencing reactions were amplified by 35 cycles of 96 °C for 10 min, 51 °C for 5 s and 60 °C for 4 min. The sequences were visualized on an ABI 3100 automated sequencer (Applied Biosystems). All individuals were sequenced in one direction only.

Table 2.3 Primer sequences and PCR details of seven chloroplast regions

Primer name	Sequence	Length (bp)	Annealing Temp (ºC)	Annealing time (second)	Number of PCR cycles	Reference
MatK-390F	CGATCTATTCATTCAATATTTC	920	53	30	35	Cuénoud *et al* (2002)
MatK-1326R	TCTAGCACACGAAAGTCGAAGT					
psbD	CTCCGTARCCAGTCATCCATA	1218	53	30	35	Shaw *et al* (2007)
trnT-R	CCCTTTTAACTCAGTGGTAG					
psbJ	ATAGGTACTGTARCYGGTATT	1269	53	30	35	Shaw *et al* (2007)
petA	AACARTTYGARAAGGTTCAATT					
rbcL-724R	TCGCATGTACCTGCAGTAGC	743	53	30	35	Fay *et al* (1997) Kress and Erickson (2007)
rbcL-F	ATGTCACCACAAACAGAGACTAAAGC					
trnF-f	ATTTGAACTGGTGACACGAG	447	53	30	35	Taberlet *et al* (1991)
trnL-e	GGTTCAAGTCCCTCTATCCC					
trnG (UCC)	GAACGAATCACACTTTTACCAC	844	53	30	35	Hamilton (1999)
trnS (GCU)	GCCGCTTTAGTCCACTCAGC					
trnL-c	CGAAATCGGTAGACGCTACG	456	53	30	35	Taberlet *et al* (1991)
trnL-d	GGGGATAGAGGGACTTGAAC					

2.3.4 Nuclear amplification and sequencing

Twelve nuclear regions, including ITS rDNA (White *et al*, 1990), obtained from species in the same genus were tested (Appendix 1). Among these nuclear regions, DNA was successfully amplified with three primer pairs (*TcWRKY*-13, *atpB* and ITS rDNA). Primers *TcWRKY*-13 were designed for *Theobroma cacao* (Borrone *et al*, 2007) and

primers *atpB* were designed for rosids (Steele *et al*, 2008). Two primers were developed for two other nuclear regions, nitrate reductase (*Nr*) and glutaredoxin (*Grx*) genes from *T. platyphyllos* sequences available in GenBank (Accession number AY138811.1 and AF406809.1, respectively). The sequences were aligned with the *Theobroma cacao* genome to identify exons before designing primers using Primer3 software version 0.4.0. However, the *Nr* primers amplified multiple PCR products, thus they were not further used. Therefore, four nuclear regions were used for phylogenetic analysis (Table 2.4).

PCR reactions of all nuclear regions were the same as carried out in chloroplast regions. PCR cycling conditions were different for each region. For *Grx* region, the amplification was as follows: an initial denaturation step at 94 °C for 4 min, followed by 35 cycles of denaturation step at 94 °C for 30 s, annealing temperature at 57 °C for 1 min and extension step at 72 °C for 30 s, and a final extension step of 72 °C for 7 min. The *TcWRKY*-13 and *atpB* sequences were amplified as follows: an initial pre-denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 15s, annealing at 56°C for 20 s and extension at 72 °C for 60 s, and a final extension of 72 °C for 4 min. The PCR cycling conditions for ITS rDNA were pre-denaturation at 94 °C for 3 min, followed by 30 cycles of denaturation at 94 °C for 30 s, primer annealing at 60°C for 30 s and primer extension at 72 °C for 60 s, and a final extension at 72 °C for 4 min. After amplification, PCR products were purified with Exo-SAP and sequenced using the same reaction as for chloroplast regions.

Table 2.4 Primer sequences and PCR details of nuclear regions used in this study

Primer name	Sequence	Reference
*Grx*F	TTTCAGCAAGTCGTACTGTCC	Newly designed
*Grx*R	AATCAGCTTCCCATTCTTGTG	
TcWRKY-13F	AAGCAAGTGAAAGGAAGTGAG	Borrone *et al* (2007)
TcWRKY-13R	TGAAAGCTCTTGGATCATCCGATGC	
atpB-51F	CCTAGCTTGATGACACCAC	Steele *et al* (2008)
atpB-51R	CTTGGACGTATCCTGAAT	
ITS4	TCCTCCGCTTATTGATATGC	White *et al* (1990)
ITS5	GGAAGTAGAAGTCGTAACAAGG	

2.3.5 Nuclear gene cloning

After sequencing, some individuals that showed multiple overlapping sequences were cloned. The PCR products were purified using QIAquick PCR Purification Kit (Qiagen). Cleaned PCR products were ligated in pJET1.2 vector using CloneJET™ PCR cloning kit (Thermo Scientific) and transformed into α-Select Competent Cells (Bioline). Eight colonies per individual were collected in 15 µl autoclaved water before they were heated at 99°C for 5 min in thermo cycle. Colonies were PCR amplified to check the insertion in 15 µl reactions containing 3 µl of heated colony solution, 1x reaction buffer, 2mM $MgCl_2$, 0.2 mM of each dNTP, 0.4 mM of each pJET1.2 primer and 0.5 U of *Taq* DNA polymerase (Bioline). The thermal cycle program had an initial pre-denaturation step at 94 °C for 3 min, followed by 35 cycles of denaturation step at 94 °C for 30s, annealing temperature at 60°C for 30s and extension step at 72 °C for 60s, and ending with 72 °C for 4 min. The PCR products were cleaned and sequenced as above.

2.3.6 Sequence alignment and phylogenetic analysis

Sequences from all regions were double-checked using the chromatograms, edited and aligned using Geneious version 6.1.4 (Drummond *et al*, 2011). The nuclear sequences were aligned with *Theobroma cacao* genome to identify intron and exon. The percentages of introns present in each nuclear gene sequence were calculated. All seven chloroplast regions of each sample were concatenated and regarded as a single sequence using Mesquite Version 2.75 (Maddison and Maddison, 2001). For clones of nuclear genes, all different sequences of each individual were used for phylogenetic analysis.

Phylogenies were estimated using maximum parsimony (MP) and maximum likelihood (ML). For each data set, parsimony analysis was conducted in PAUP* version 4.0b10 (Swofford, 2003) with the heuristic search option, random taxon addition sequence limited to 10 and tree bisectional reconnection (TBR) branch-swapping. Bootstrap support (Felsenstein, 1985) was estimated with 1000 replications. All trees of each data set were combined and produced as a 50% majority-rule consensus tree. Both consistency and retention index (CI and RI, respectively) were calculated to indicate the amount of homoplasy in the phylogenetic tree. Maximum likelihood analysis was performed as an alternative means of phylogenetic reconstruction using GARLI 2.0 (Zwickl, 2006). The analysis was performed through heuristic search with TBR branch

swapping. 1000 bootstrap replications were performed to assess the robustness of branches. A 50% majority-rule consensus tree of each data set was constructed using PAUP*.

2.4 Results

2.4.1 Characteristics of chloroplast, nuclear DNA and ITS rDNA sequences

Seven regions (*MatK-390, psbD, psbJ, rbcL724, trnF, trnG and trnLc*, Table 2.3) were sequenced for the phylogenetic analysis of 22 *Tilia* species. For nuclear genome analysis, four nuclear genes (*Grx*, *TcWRKY*-13, *atpB* and ITS rDNA) were successfully amplified in *Tilia* species (Table 2.4) and were further sequenced. After sequencing, 18 individuals for *Grx* region and 15 individuals for *atpB* region showed heterozygote sites/ indels. PCR products of these individuals were cloned to obtain the sequences of individual copies of the regions. Diploid species showed two main copies of the nuclear regions, thus two sequences of cloned PCR products were used for further phylogenetic analysis. For the species that are polyploid, all different copies of the clones were analysed. The sequences of the clones of the hybrid (*T.* x *europaea* Hatfield) indicated two different sequences. One sequence is similar to *T. cordata* and the other is similar to *T. platyphylos* (Table 2.5). This is also observed in *T.* x *euchlora*, the hybrid of *T. cordata* and *T. dasystyla* (data not shown).

For the ITS rDNA region, DNA from 26 individuals was cloned. Most of the individuals showed a large number of different ITS haplotypes. This makes phylogenetic inference difficult, also because variable sites were shared between individuals and between species, and thus did not reveal species relationships. Therefore the ITS region was not used for phylogeny reconstruction. Table 2.6 illustrates one of the diploid species (*T.* x *europaea* Hatfield) that had seven different copies of the ITS region.

Table 2.5 Variable sites of *T. cordata*, *T. platyphyllos* and *T.* x *europaea* Hatfield (hybrid) based on *Grx* region indicate the hybrid contained two gene copies from the parental species.

Species	Variable sites																						
	29	33	45	61	131	195	199-222	243	280	288	296	318	322	347	348	357	359	380	386	390	442	469	579
T. cordata	A	A	C	A	C	A	-	T	T	C	T	G	A	T	-	A	A	A	A	A	C	T	C
T. platyphyllos	T	G	T	T	A	G	+	C	G	T	C	A	G	A	+	C	G	C	G	T	T	A	T
T. x europaea Hatfield clone 1	A	A	C	A	C	A	-	T	T	C	T	G	A	T	-	A	A	A	A	A	C	T	C
*T. x europae*a Hatfield clone 2	A	A	C	A	C	A	-	T	T	C	T	G	A	T	-	A	A	A	A	A	C	T	C
T. x europaea Hatfield clone 3	T	G	T	T	A	G	+	C	G	T	C	A	G	A	+	C	G	C	G	T	T	A	T
*T. x europae*a Hatfield clone 4	T	G	T	T	A	G	+	C	G	T	C	A	G	A	+	C	G	C	G	T	T	A	T
T. x europaea Hatfield clone 5	T	G	T	T	A	G	+	C	G	T	C	A	G	A	+	C	G	C	G	T	T	A	T
*T. x europae*a Hatfield clone 6	T	G	T	T	A	G	+	C	G	T	C	A	G	A	+	C	G	C	G	T	T	A	T
T. x europaea Hatfield clone 7	T	G	T	T	A	G	+	C	G	T	C	A	G	A	+	C	G	C	G	T	T	A	T

Table 2.6 Variable sites in clones of *T.* x *europaea* Hatfield (diploid species) based on based on ITS region

T. x europaea Hatfield	Variable sites																													
	12	43	54	113	121	129	170	195	213	216	224	243	371	412	413	416	436	440	442	444	496	541	542	548	580	581	590	591	593	598
Clone 1	G	G	C	G	A	T	A	T	-	C	A	T	T	T	C	A	A	A	A	A	G	A	G	T	G	T	A	G	C	G
Clone 2	A	G	C	G	G	T	C	G	T	C	A	C	T	T	T	A	G	C	G	G	G	G	A	T	C	T	G	G	C	A
Clone 3	G	G	C	G	G	T	C	G	T	T	G	T	C	T	T	G	G	C	G	G	G	G	G	C	G	C	G	A	A	A
Clone 4	A	G	C	G	G	T	C	G	T	T	G	T	T	T	T	G	G	C	G	G	G	G	G	C	G	C	G	A	A	A
Clone 5	A	G	C	G	G	T	C	G	T	T	G	T	T	T	T	G	G	C	G	G	G	G	G	C	G	C	G	A	A	A
Clone 6	A	A	C	A	G	T	A	T	T	C	A	T	T	T	T	G	G	C	G	G	G	G	G	C	G	C	G	A	A	A
Clone 7	G	G	C	G	A	T	C	G	T	T	G	T	T	T	T	A	A	A	A	A	G	G	G	C	G	C	G	G	A	A
Clone 8	A	G	T	G	G	C	C	G	T	T	G	T	T	C	T	G	G	C	G	G	A	G	G	C	G	C	G	A	A	A

The characteristics of the chloroplast and nuclear DNA sequences of 55 individuals in 22 *Tilia* species are shown in Table 2.7. For seven combined chloroplast regions, the total number of aligned nucleotides was 2,843 bp of which 139 sites were variable and 27 were parsimony informative (0.95%), with high consistency index (CI) and retention index (RI) of 0.96. There were 17 insertions-deletions (indels) ranging from 1 bp in *trnLc* to 13 bp in *psbD*. For nuclear regions, the aligned positions of *Grx*, *TcWRKY*-13 and *atpB* data sets were 562, 386 and 410 bp, respectively. The parsimony informative characters were highest in *Grx* at 61 sites (10.85%), while those of *TcWRKY*-13 and *atpB* were 24 (6.22%) and 32 (7.80%), respectively. Among these nuclear regions, *Grx* had the highest percentage of intron presence in the sequence data (85.04%), while those of *atpB* and *TcWRKY*-13 were 59.04% and 17.36%, respectively. The CI of all nuclear regions were similar at approximately 0.83 and the RI were 0.95 in both *Grx* and *TcWRKY*-13 and 0.90 in *atpB*. Five and six indels were observed in *Grx* and *atpB*, respectively. No indel was detected in *TcWRKY*-13 regions. The longest indel, 23 bp, was found in the *Grx* region of all *T. cordata* and some clones of *T. dasystyla*, *T. chinensis*, *T.* x *europaea* Hatfield and *T.* x *euchlora*.

Table 2.7 Characteristics of seven combined chloroplast and three nuclear regions of 56 individuals of *Tilia* species. CI = consistency index, RI = retention index

Data set	Aligned positions	No. of parsimony uninformative sites	No. of parsimony informative sites	Indels	% of intron in the sequence	CI	RI
Seven chloroplast regions	2,843	112	27 (0.95%)	17	-	0.962	0.956
Grx	562	107	61 (10.85%)	5	85.04	0.837	0.951
TcWRKY-13	386	27	24 (6.22%)	0	17.36	0.834	0.953
atpB	410	81	32 (7.80%)	6	59.04	0.837	0.904

2.4.2 Phylogenetic analysis

Chloroplast DNA tree

Maximun parsimony searches of combined chloroplast DNA yielded 380,063 parsimonious trees. This topology of maximum parsimony (MP) tree was congruent with that of the maximum likelihood (ML) tree. Moreover, the bootstrap values for

branch robustness obtained from the maximum parsimony criterion were similar to those from likelihood criterion. Therefore, only the MP tree is presented, while presenting bootstrap values from both criteria (Figure 2.1).

The 50% majority-rule consensus tree revealed that *Tilia* species can be divided into three major clades (A, B and C). Two species, *T. endochrysea* and *T. kiusiana*, did not group with other species. Clade A, which had the lowest bootstrap support (MP=58%, ML=75%), is the major clade comprising all four European species (*T. cordata*, *T. platyphyllos*, *T. dasystyla* and *T. tomentosa*), their hybrids (*T.* x *europaea* and *T.* x *euchlora*) and 13 species of eastern Asia (Figure 2.1). The relationships of the species within the clade are weakly resolved due to lack of sequence divergence. Clade B with high bootstrap support (MP=99%, ML= 96%) contained three eastern Asia species. Two out of three species (*T. amurensis* and *T. tuan*) in this clade were also present in clade A. *Tilia* species from the US (*T. americana* and *T. caroliniana*) formed a distinct group (clade C) supported by high bootstrap values (MP=78%, ML=79%).

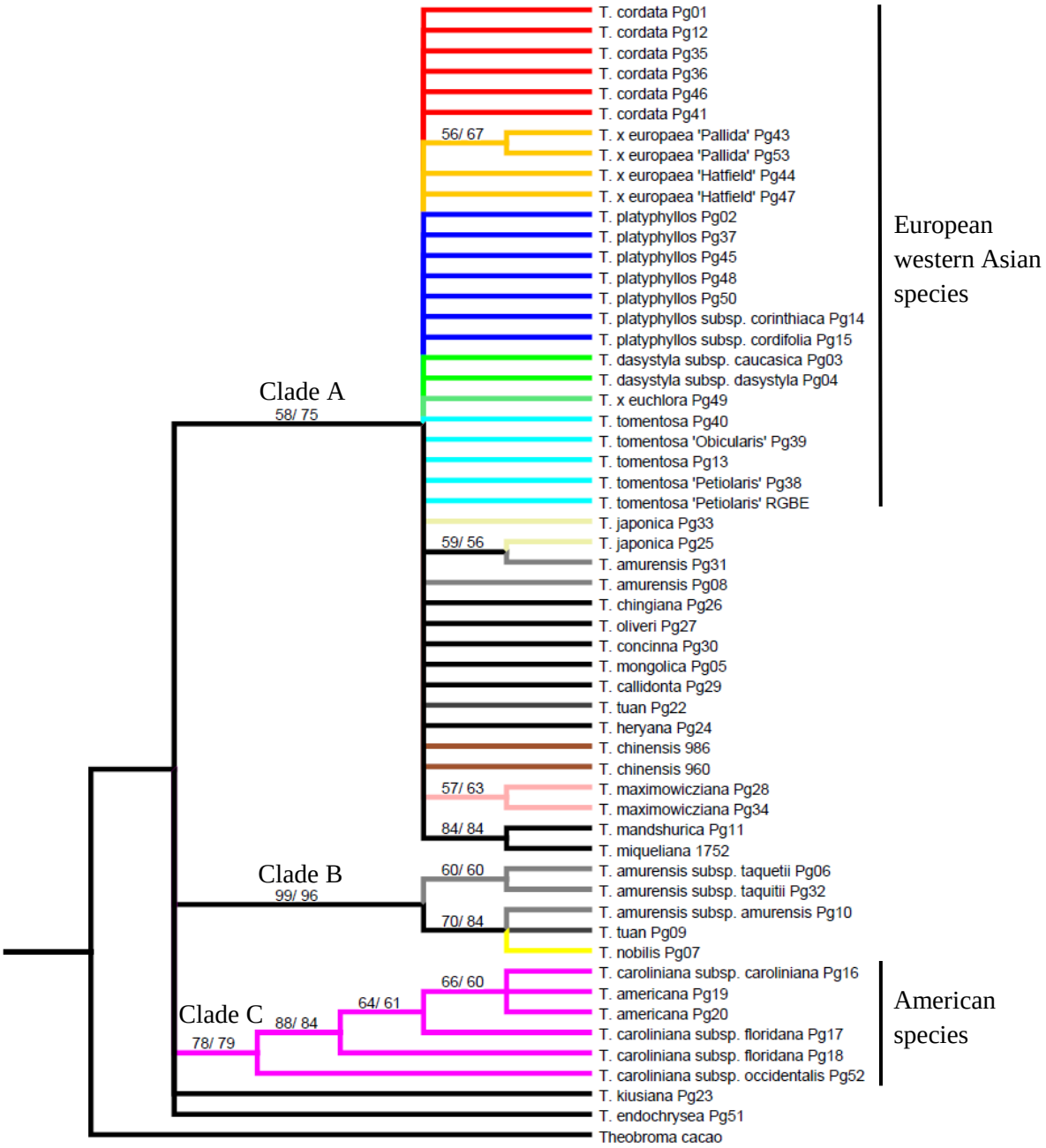


Figure 2.1 The 50% majority rule consensus tree of maximum parsimony (MP) based on seven combined chloroplast DNA sequences from 22 *Tilia* species. Numbers above branches indicate bootstrap values of maximum parsimony and maximum likelihood analysis, respectively. Names after species represent the sample labels. Branch colours indicate the same species except that the two North American species are the same colour.

Nuclear DNA trees

The phylogenetic analysis of the three nuclear regions, *Grx*, *TcWRKY*-13 and *atpB*, revealed that the topologies of 50% majority rule consensus trees obtained from maximum parsimony (MP) analysis in the three regions were very similar to the topology obtained from maximum likelihood (ML) analysis. Thus, only MP trees are presented (Figure 2.2, 2.3 and 2.4). The *Grx* region, which has the highest percentage of parsimony informative characters (10.85%), revealed the best resolved phylogenetic tree (Figure 2.2). The other two nuclear regions, with low numbers of parsimony informative sites, showed little phylogenetic resolution among 22 *Tilia* species. However, trees were largely congruent. In addition, the phylogenetic tree of each nuclear region was congruent with its unrooted phylogenetic networks (Appendix 2, 3 and 4).

The parsimony analysis revealed a close relationship of *T. cordata*, *T. amurensis*, *T. japonica* and *T. chinensis.* These four species grouped in the same clade for both *Grx* and *TcWRKY*-13 regions with strong support, a high MP bootstrap value (80% and 82%, respectively). The other two Eastern Asian species, *T. kiusiana* and *T. mongolica*, were also clustered closely to those four species. *T. platyphyllos*, and most individuals of *T. dasystyla* and *T. tomentosa* (all are European species) were grouped together in *Grx* region with 72% bootstrap support. It is clearly observed in all nuclear regions that the two main European species, *T. cordata* and *T. platyphyllos*, were separated in different groups. Their hybrids (*T.* x *europaea*) were polyphyletic, and clustered into both parental species clades. Also, *T.* x *euchlora* grouped into two clades of the parental species (*T. cordata* and *T. dasystyla*). The US species (*T. americana* and *T. caroliniana*) formed a monophyletic group for both *Grx* and *atpB* with 79% and 73% bootstrap support. These species were closely related to an eastern Asian clade comprised of *T. oliveri, T. concinna* and *T. miqueliana*, as well as the species *T. chingiana*, with low bootstrap support (58%) in *Grx*. The remaining species showed little to no phylogenetic resolution, particularly in the tree constructed from *atpB* sequences.

In this study, some incongruence of the tree topology between chloroplast and nuclear regions was observed in some species, for example, some individuals of *T. amurensis*, *T. tuan* and *T. nobilis*. These species formed a clade (clade B, Figure 2.1) with high bootstrap support (99%) based on combined chloroplast regions but this relationship

was not observed in nuclear phylogeny. Also, in the tree obtained from nuclear regions, *T. nobilis* and *T. tuan* largely formed unrooted relationships. In addition, conflicting placement between *TcWRKY*-13 and *Grx* was observed in some species. *T. maximowicziana* was clustered with *T. mandshurica* based on *TcWRKY*-13. On the other hand, one sequence of *T. maximowicziana* was placed with *T. tomentosa* and the other was placed in the clade of *T. cordata*, based on the *Grx* phylogenetic tree.

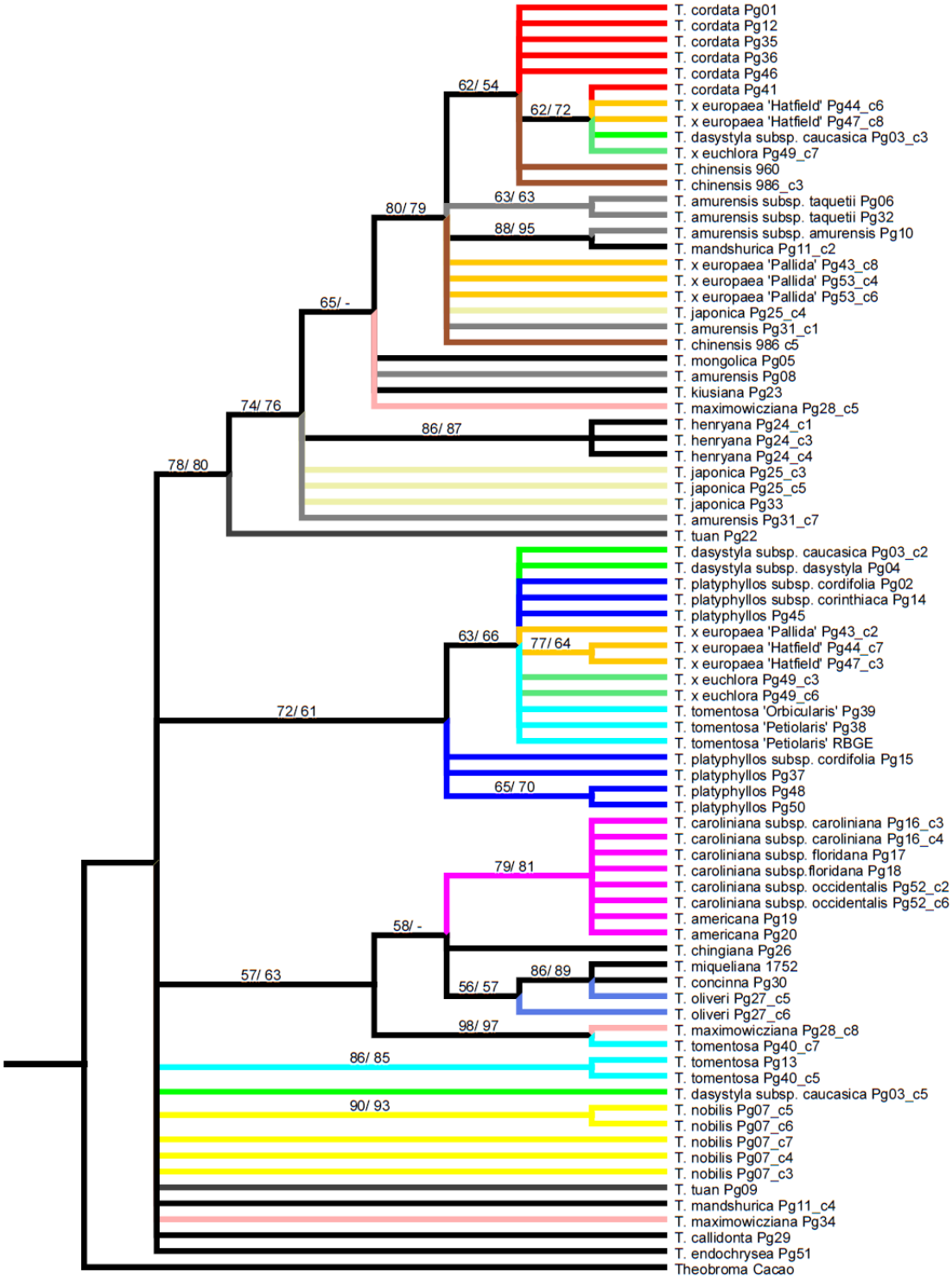


Figure 2.2 The 50% majority rule consensus tree of maximum parsimony (MP) based on *Grx* nuclear sequences from 22 *Tilia* species. Numbers above branches indicate bootstrap values of maximum parsimony and maximum likelihood analysis, respectively. Names after species represent the sample labels and the numbered cloned sequence. Branch colours indicate the same species except that the two North American species are the same colour.

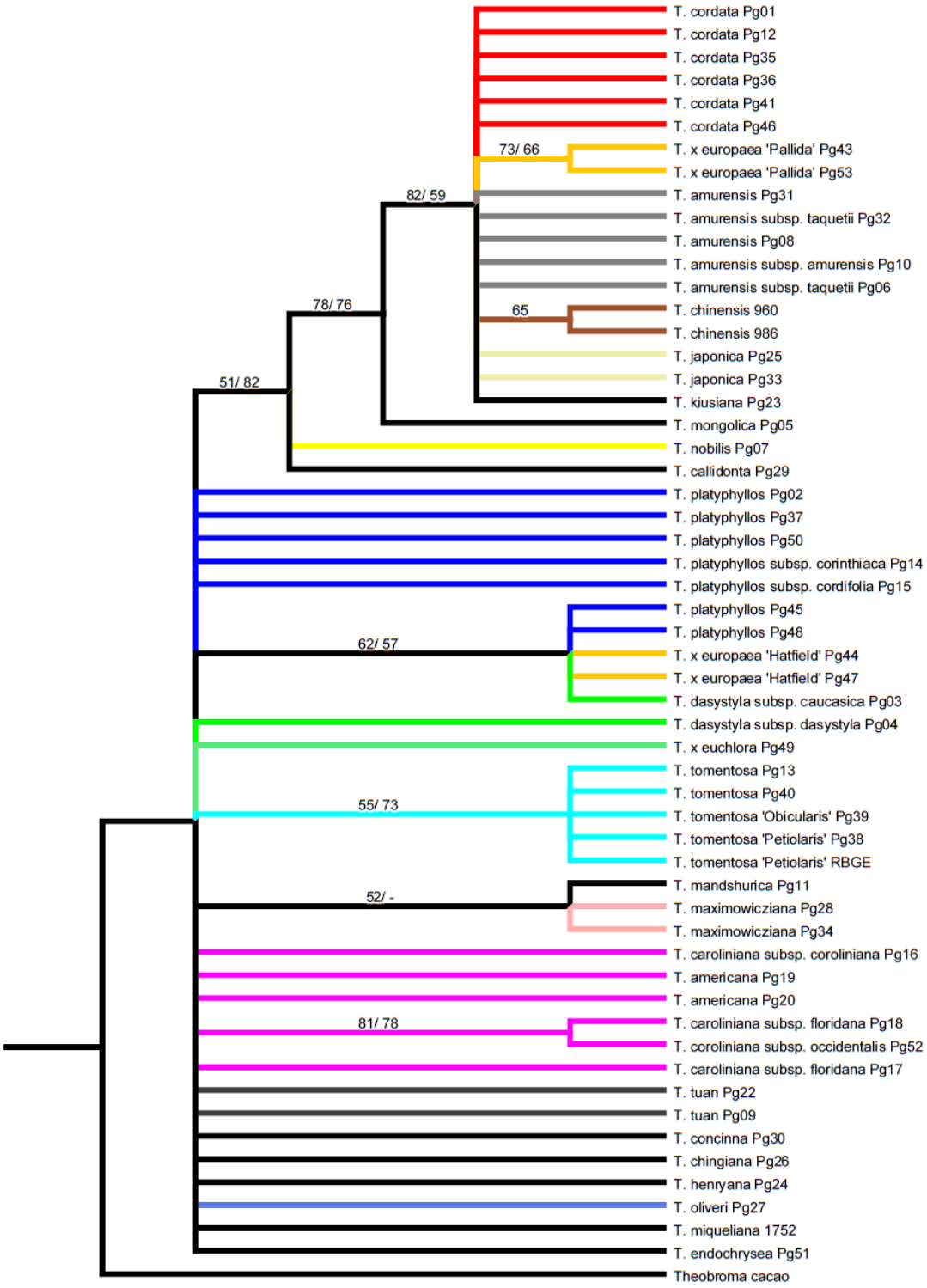


Figure 2.3 The 50% majority rule consensus tree of maximum parsimony (MP) based on *TcWRKY*-13 nuclear sequences from 22 *Tilia* species. Numbers above branches indicate bootstrap values of maximum parsimony and maximum likelihood analysis, respectively. Names after species represent the sample labels. Branch colours indicate the same species except that the two North American species are the same colour.

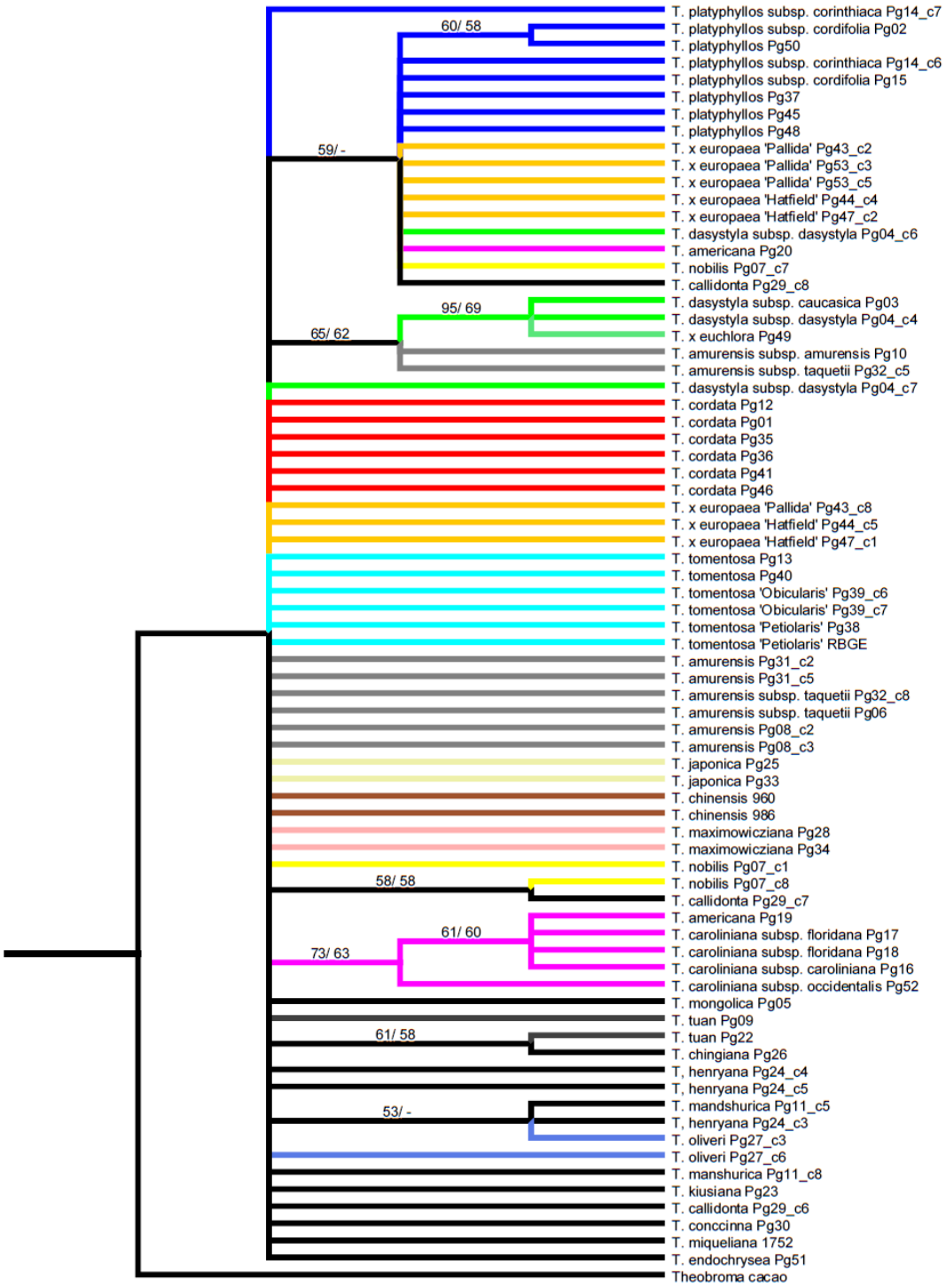


Figure 2.4 The 50% majority rule consensus tree of maximum parsimony (MP) based on *atpB* nuclear sequences from 22 *Tilia* species. Numbers above branches indicate bootstrap values of maximum parsimony and maximum likelihood analysis, respectively. Names after species represent the sample labels and the numbered clone. Branch colours indicate the same species except that the two North American species are the same colour.

2.5 Discussion

2.5.1 Molecular evolution of chloroplast, nuclear DNA and ITS rDNA in *Tilia*

For the study of phylogeny at the low level of a taxonomic group, it is necessary to have sufficient informative characters. When using the chloroplast genome, which evolves relatively slowly, a multi-locus analysis is essential for sufficient phylogenetic resolution (Shaw and Small, 2004). The seven chloroplast regions (*MatK-390, psbD, psbJ, rbcL724, trnF, trnG and trnLc*) showed low variation within each particular region. Therefore, these chloroplast regions were combined. However, the number of parsimony informative sites of seven combined regions is still low (0.95%), compared to three nuclear regions (7-11%) as shown in Table 2.7. This is expected as generally the evolutionary rate of the chloroplast genome is slower than the nuclear genome (Gaut, 1998). The lack of sufficient variation of the chloroplast regions leads to low resolution in phylogeny reconstruction. Most of the species relationships are not resolved. This limitation of chloroplast DNA, failing to provide a sufficient phylogenetic resolution at species level, is not unexpected and has been observed in many plant species (Small *et al*, 2004).

Of the three nuclear regions examined in this study, *Grx* is the most variable region, with the greatest number of parsimony informative (PI) characters and almost twice the percentage of PI sites (10.85%) compared to the other two nuclear regions (Table 2.7). These could be due to the sequences of *Grx* region containing the highest percentage of introns compared to the other two regions, as the variations of the DNA sequences tend to occur in introns. Therefore, it is not unexpected that the *Grx* region provided a better phylogenetic resolution than the *TcWRKY*-13 and *atpB* regions.

However, some species relationships are still not well resolved. The low resolution of phylogenetic trees in the genus *Tilia* could be the result of the long generation time. The age at which *Tilia* trees begin to flower and produce seed ranges from six to 40 years old depending on the environmental conditions, particularly related to light exposure. A tree in shady conditions may not produce flowers. In addition, human management (e.g. coppicing) has reduced sexual regeneration. Even though stems will mostly not survive longer than 200 – 300 years and are often hollow, it is well known that *Tilia* tree individuals can survive for many hundreds of years. The genotype may simply survive through sprouting near the base of the tree or through regeneration from branches touching the ground (Pigott, 1991; Pigott, 2012). Therefore, current genotypes may well

be ancient and seemingly separate trees could actually be the same genotype. This low recruitment leads to slow turnover of trees. The small number of generations may be insufficient for an individual gene to accumulate sufficient variation for resolving informative phylogenetic tree. Thus, the species within this genus *Tilia* seem to be genetically fairly close to each other, which is reflected in the fact that large numbers of species can hybridise. This evidence is also supported by a relatively small number of species within the genus (approximately 23 species) (Pigott, 2012) compared to other tree species in the same group of rosids, such as the genus *Quercus* (oak), which contains approximately 600 species (Hogan, 2012).

The internal transcribed spacer (ITS) of nuclear ribosomal DNA is one of the most popular regions used for phylogenetic study among plant species. This rDNA region provides many advantages for phylogenetic reconstruction (Alvarez and Wendel, 2003). The high copy number of ITS rDNA in the genome facilitates PCR amplification, compared to the low-copy nuclear loci. However, the multiple copies of ITS per genome are subject to a process called concerted evolution. Due to this process, the multiple sequences in the genome will be uniform or largely one unique sequence within the individual (Alvarez and Wendel, 2003; Buckler *et al*, 1997). In the absence of sequence homogenisation, multiple copies of ITS rDNA sequences in one individual can result in misleading phylogenetic inference and distorted evolutionary history. Alvarez and Wendel (2003) stated that multiple ITS rDNA sequences are quite common in polyploid species and hybrids. From the analysis of ITS sequences of *Tilia* in this study, multiple ITS sequences were observed not only in polyploid species, but also in the diploid species. For example, the sequences of the diploid species *T.* x *europaea* Hatfield showed seven different sequences within an individual, while a maximum of two would be expected in a diploid hybrid. These results clearly indicate incomplete concerted evolution in the genus *Tilia*. Therefore, ITS rDNA sequences are not used here for phylogenetic inference.

2.5.2 Relationships of *Tilia* species

Phylogeny reconstruction from independent genome regions can increase the confidence of the relationships. Phylogenetic topologies of the different data sets can provide insight into evolutionary processes particularly with little genetic divergence between species (Steele *et al*, 2008). In this study although low resolutions of

phylogenetic trees were observed and some species relationships are still not well resolved, considering the tree topologies of all regions together did help to raise support for some species relationships. Among the reconstruction of phylogenetic trees, the tree obtained from the *Grx* DNA sequence has provided the highest resolution. Therefore, in order to discuss the relationships in the genus *Tilia*, I will concentrate mainly on this tree.

Tilia are widely distributed species across all continents based on geographical distribution. The species were divided into three groups; European and western Asian species, eastern Asian species and north and central American species (Pigott, 2012). In the present study, some correlation between species groups and their geographical distribution was observed. American species, which are geologically separated, appear to be monophyletic in most phylogenetic trees. Most European species are intermixed with Asian species, although some of the European species did group together. These results supported the phylogenetic analysis based on two chloroplast regions (*rpL*32-*trnL* and *ndhF-rpL32*) by McCarty (2012).

North American species

The two species from North America, *T. caroliniana* and *T. americana*, appear to form a monophyletic clade. This US clade is congruent with most of the phylogenetic trees (chloroplast DNA, *Grx* and *atp*B) except for *TcWRKY*-13. The placement of *T. caroliniana* and *T. americana* in the same clade suggested that intercontinental disjunction has evolved in the genus and separated the two US species from European and Asian species. The close relationship between the two species supports the view that *T. caroliniana* and *T. americana* are closely related and can hybridise (Pigott, 2012). Within the clade, the relationships of the species based on chloroplast sequences indicated that specimens collected from different regions of the American continent had genetic differences, particularly *T. caroliniana* subspecies *occidentalis*, which occupied the terminal position of the clade. This subspecies was collected from Mexico and used to be treated as a separate species, *T. mexicana* (Pigott, 2012). *T. caroliniana* subspecies *floridana* (Pg18) has the sister position to *T. caroliniana* subspecies *occidentalis* and was also collected from Mexico. The rest of the taxa in this clade were collected from the US. These results suggest a biogeographic relationship of northern and central American taxa.

Europe and western Asian species

The phylogeny of European and western Asian species based on *Grx* sequences grouped them into two main clades. The first clade comprises only European and western Asian species (*T. platyphyllos, T. dasystyla, T. tomentosa* and some hybrids). *T. cordata*, which is the most common species in Europe, is in another clade with some eastern Asian species. The separation of *T. cordata* and *T. platyphyllos* into different clades was also observed in the analysis of other nuclear regions. This suggests that although these two species have a similar geographic distribution and can hybridise naturally (*T.* x *europaea*), they seem not to be sister species. The result is in agreement with the study of the relationships of twelve *Tilia* taxa based on the sequences of the nitrate reductase (*NIA*) gene, which revealed that *T. platyphyllos* is not closely related to *T. cordata* (Li *et al*, 2002). The clearly separated clades of *T. cordata* and *T. platyphyllos* also support the difference in morphology (Chapter 6).

The main clade of European and western Asian species based on *Grx* comprises *T. platyphyllos, T. dasystyla, T. tomentosa* and some hybrids (*T.* x *europaea* and *T.* x *euchlora*) suggesting a biogeographic relationship among these European species. The close relationship of *T. platyphyllos* and *T. dasystyla* supports the similarity in morphological characters of the species. These two species differ only in a few morphological characters, such as the marginal teeth of the leaves, the distribution and type of hairs on the leaves and the size and thickness of the fruit wall (Pigott, 2012). Loria (1967) proposed that *T. dasystyla* should be treated as a subspecies of *T. platyphyllos*. In addition, the placement of *T. dasystyla* subspecies *caucasica* in the same clade as *T.* x *europaea* confirms the evidence based on morphology that *T. dasystyla* subspecies *caucasica* is very similar to some variants of *T.* x *europaea*. Pigott (2012) suggested that the tetraploid of *T. dasystyla* (2n=164) possibly originated from this hybrid (2n=82). Moreover, *T. platyphyllos* can hybridise with *T. tomentosa* and form a hybrid (*T.* x *haynaldiana*). *T. tomentosa* can hybridise with *T.* x *euchlora* and their hybrid was named *T.* x *orbicularis* (Pigott, 2012). The extensive hybridisation among these species was supported by the close relationships observed in the phylogenetic tree.

T. tomentosa, one of the four European species, clustered together with other European species based on *Grx* sequence analysis. Some individuals in this species were placed outside the European clade. Considering the location of the collection, *T. tomentosa*

Pg13 and some clones of *T. tomentosa* Pg40, which clustered outside the European clade, were collected from western Asia (Montenegro and Bulgaria, respectively). The other *T. tomentosa* were collected from the UK. This indicated that *T. tomentosa* from different geographical regions may have a different evolutionary history. However, all individuals within this species still clustered in the same group based on *TcWRKY*-13. This may reflect the little genome difference among these individuals for this gene.

The polyphyletic origin of the hybrid *T.* x *europaea* was observed. The hybrid was placed into two clades of *T. cordata* and *T. platyphyllos*. This finding confirmed that hybridization occurred between these two species. The analysis of the sequences of the hybrid clones revealed that hybrids contained two genomes from the parental species. This is agreement with the intermediate morphology between *T. cordata* and *T. platyphyllos* observed in *T.* x *europaea* (Pigott, 2012). Moreover, the placement of *T.* x *euchlora* close to *T. dasystyla* and *T. cordata* also supports the evidence of hybridisation between these parental species (*T. dasystyla* and *T. cordata*) (Pigott, 2012).

Eastern Asian species

The phylogenetic relationships among eastern Asian species are not well resolved. Some intermixing between eastern Asian and European species were observed. In general, *T. cordata*, which is the most widely spread species across Europe, was clustered with three eastern Asian species, *T. amurensis*, *T. japonica* and *T. chinensis*, in the well resolved phylogenetic trees, particularly in *TcWRKY*-13. This suggests relatively close evolutionary relationships among these species. This finding is consistent with historical evidence that both *T. amurensis* and *T. japonica* have been treated as varieties of *T. cordata* based on their morphology. These three species are morphologically very similar and they are all variable. This causes a difficulty in species identification, particularly when no flowers are present (Pigott, 2012). Some morphological characters have been used to differentiate these species. For example, the consistent present of five staminodes distinguished *T. japonica* from *T. cordata* and *T. amurensis*. Inflorescence is another character that is used for separation of *T. japonica* from *T. cordata*. The number of branches and of flowers in *T. japonica* is much larger than those of *T. cordata.* The inflorescence of *T. cordata* is compact and erect compared to that of *T. amurensis*, which is usually widely branched and hangs down (Pigott, 2012).

Surprisingly, *T. chinensis* also grouped in this clade. There is no evidence in terms of morphological affinities or biogeographic location to support the placement of this species closely related to those three species (*T. cordata*, *T. japonica* and *T. amurensis*). Prof. Pigott, an expert in *Tilia* identification, stated (personal communication) that a *T. cordata* tree at the Royal Botanic Garden, Kew, was treated as *T. chinenesis* and the same tree had been propagated and placed in other arboreta. In this study, the specimens of *T. chinensis* were collected from living species in Peasmarsh Place Arboretum, Rye, which could be propagated from the *T. cordata* tree in Kew. Therefore, it occupied the same clade as *T. cordata* in the phylogenetic trees. To clarify relationships of *T. chinensis*, a new sample is needed. Trees of *T. chinensis* in Peasmarsh Place Arboretum need to be identified by an expert and may need to be renamed.

In the phylogeny based on *Grx* sequence data, the three eastern Asian species, *T. miqueliana*, *T. concinna* and *T. oliveri*, group in a sister relationship with *T. chingiana* and the US species. The close relationship among these three Asian species is congruent with the similarity in morphology (Pigott, 2012). The morphology of *T. concinna* resembles that of *T. miqueliana* in the form of leaves including leaf-shape and the marginal teeth. They differ only in the type of hairs at the underside of the leaves. Also, the morphology of stems, petiole and peduncles of *T. concinna* is similar to those of *T. oliveri*. In addition, it was suggested that *T. concinna* could be the product of hybridisation between the two diploid species, *T. miqueliana* and *T. oliveri* (Pigott, 2012), which may have formed this allotetraploid species. Therefore, the taxonomic position among these species supported the close affinity between them. The sister relationship between *T. chingiana* and the cluster containing *T. oliveri* supports the evidence of hybridisation between these species (Pigott, 2012). The close relationship of these four eastern Asian species with the US species suggests some evolutionary relationships, although the bootstrap value is low (58%).

Among *Tilia* species, *T. endochrysea* was placed separately from other member species in all phylogenetic trees. This is in agreement with the study of *Tilia* species based on the sequencing of the intron of the nitrate reductase gene (*NIA*) that separated *T. endochrysea* at the first division (Li *et al*, 2002). In addition, the result also supports the evidence from the fossil record, which suggest this species to be primitive. The morphology of bract and fruits based on fossil records of *T. endochrysea* is very similar to those of the living species in western USA and in central Europe (Pigott, 2006).

2.5.3 Data incongruence

It has been suggested that a disagreement of some species in phylogenetic placement could be the result of incomplete lineage sorting or hybridisation among taxa (Mort *et al*, 2007; Pigott, 2012; Small *et al*, 2004). Natural hybridisation between two *Tilia* species in mixed populations has been widely reported (Pigott, 2012) and seems to be a fairly common phenomenon in this genus. The presence of trees that have an intermediate morphology between two *Tilia* species has been investigated in many mixed populations not only between the two widely distributed species, *T. cordata* and *T. platyphyllos*, but also between numerous species in eastern Asia, for example, the hybrids between *T. chinensis* and *T. paucicostata*, between *T. callidonta* and *T. nobilis*, and between *T. chingiana* and *T. oliveri*. Also introgression was observed between the hybrids of *T. japonica* and *T. maximowicziana* (Pigott, 2012).

In this study, conflicting placements between nuclear regions were observed in some species, such as *T. maximowicziana* with *T. mandshruica* based on *TcWRKY*-13. The placement of these species in the same group supports the similar morphology of these species. In addition, Pigott (2012) suggested that *T. maximowicziana* is possibly an autopolyploid of *T. mandshurica*. However, considering the analysis of *Grx* sequences, one clone of *T. maximowicziana* was placed with *T. tomentosa* and the other was placed in the clade of *T. cordata*. The conflict among phylogenetic trees in these species could be the results of the extensive hybridisation among species.

2.6 Conclusion

Phylogenetic analysis of 22 species in the genus *Tilia* revealed a fairly low resolution because of the lack of sequence divergence, particularly in chloroplast DNA. Of the three nuclear regions, *Grx* had the highest number of variable characters and provided more resolution of the phylogenetic reconstruction among *Tilia* species. The low resolution of phylogenetic trees could be the result of long generation time and low recruitment of *Tilia* species. Although some species relationships are not well resolved, several relationships between species were suggested. *Tilia* species in the US, *T. americana* and *T. caroliniana*, formed a monophyletic clade, while most European species are intermixed with Asian species. Including *T. paucicostata* in the analysis

would increase our understanding of species relationships of all species in the genus *Tilia*. The two main species in Europe, *T. cordata* and *T. platyphyllos*, seem to be evolutionarily distant. Some incongruence between phylogenetic trees suggested incomplete lineage sorting and supported extensive hybridisation among *Tilia* species.

Chapter 3. Isolation and characterization of microsatellite loci in *Tilia platyphyllos* (*Malvaceae*) and cross-amplification in related species

3.1 Abstract

Little molecular research has been carried out on the genus *Tilia*. There are currently no nuclear markers available for population genetics in this genus. In this study, 15 microsatellite markers were developed from *T. platyphyllos* by using a microsatellite enrichment protocol. Most loci show a high level of polymorphism in two *T. platyphyllos* populations from France. Cross-amplification results indicated that 12 out of 15 loci amplified polymorphic loci in 20 species and two hybrids in the genus. These microsatellite markers will be useful tools for the study of genetic diversity and population structure and increase our understanding of their phylogeography and of the hybridization between *Tilia* species.

This work has been published as the following: Phuekvilai P, Wolff K (2013). Characterization of microsatellite loci in *Tilia platyphyllos* (Malvaceae) and cross-amplification in related species. *Applications in Plant Sciences* 1(4)**:** 1200386.

3.2 Introduction

Microsatellites or Simple Sequence Repeats (SSR) or Short Tandem Repeats (STR) are sets of non-coding repetitive DNA sequences found abundantly in the genome of most taxa. Microsatellites are composed of motifs of 1- 6 nucleotide tandem repeats (Selkoe and Toonen, 2006; Tautz and Renz, 1984). The length of microsatellites typically varies between 5-40 repeats (Selkoe and Toonen, 2006). The variation of microsatellite repeats is mainly due to slippage and proofreading error during DNA replication. The slippage in replication occurs more frequently than point mutations, thus microsatellites tend to be highly variable. Generally, microsatellites with longer repeats are more polymorphic than those containing the shorter repeats (Ellegren, 2004). The mutation rate of microsatellites varies among different species, ranging from 10^{-6} in Drosophila (Schug *et al*, 1998) to 10^{-3} in humans (Brinkmann *et al*, 1998) with an average of 5×10^{-4} per locus per generation (Schlotterer, 2000). The mutation rates of microsatellite also differ drastically among repeat types. The estimated mutation rate in *Arabidopsis thaliana*

(Brassicaceae) was 2.03 x 10^{-3} per allele per generation for AT repeats, 3.31 x 10^{-5} for CT repeats, and 4.96 x 10^{-5} for CA repeats (Marriage *et al*, 2009). Unlike microsatellite regions, the DNA sequences surrounding microsatellites (flanking regions) are generally conserved. Thus, these regions can be used for designing specific primers to amplify polymorphic microsatellites, which can generate fragments through the polymerase chain reaction (PCR).

3.2.1 DNA markers: microsatellites

Microsatellite markers are one of the most popular markers used for genetic studies since their introduction in the late eighties. The first use of microsatellite markers in natural populations was reported more than 20 years ago (Ellegren, 1991; Schlotterer *et al*, 1991). Many advantages of microsatellites markers make them attractive and they remain popular. One of the advantages is their high reproducibility. As microsatellite analysis is based on PCR, only small amount of tissue sample are needed. In addition, due to the short length of the microsatellite PCR product, DNA samples that are partly degraded can still be analysed. Also, microsatellite analysis does not require very pure DNA. This allows microsatellite analysis after a simple DNA extraction and use for the analysis of ancient DNA (Ishida *et al*, 2012). Furthermore, the high mutation rate of microsatellites makes them more informative among closely related species compared to other markers. Morgante *et al* (1994) revealed that the genetic variation observed in 61 soybean accessions (both wild soybean and elite germplasm) using microsatellites was almost two times higher than the genetic variation observed by AFLP. The co-dominant inheritance of microsatellites is an advantage over other dominant markers. Microsatellites have been used increasingly since their introduction for various applications, such as in population genetics, molecular ecology, the construction of genetic maps, population structure, DNA fingerprinting, hybrid detection and parentage analysis (Ellegren, 2004; Guichoux *et al*, 2011; Jones *et al*, 2010).

3.2.2 Development of microsatellite markers

The standard method for microsatellite isolation requires the construction of a genomic library and sequencing (Squirrell *et al*, 2003). There are various protocols to isolate microsatellites from the genome (Weising *et al*, 2005). In the traditional method,

microsatellites were isolated from genomic libraries by hybridisation of colonies with microsatellite probes (Chase *et al*, 1996). However, this method is inappropriate for the species with large genomes as the detection of clones containing microsatellite repeats is relatively low (Park *et al*, 2009). The microsatellite enrichment method was introduced to facilitate the isolation. This protocol incorporates one step of microsatellite hybridisation prior to constructing the library. This can increase the number of clones containing microsatellites up to 80% (Kandpal *et al*, 1994). The recent development of next generation sequencing (NGS) technologies can facilitate the isolation of microsatellite markers (Kang *et al*, 2012; McEwen *et al*, 2011; Setsuko *et al*, 2012).

3.2.3 Transferability of microsatellites

Microsatellite primers developed in one species can sometimes be used in closely related species of the same genus or family, which is known as transferability or cross-species amplification. Microsatellites are transferable if they have conserved DNA sequences in their flanking regions across taxa. This is an important factor as it can save time and cost for developing a new set of microsatellite markers (Barbara *et al*, 2007).

The cross-species transferability of microsatellites has been studied in various organisms. The success of amplification across species decreases if the divergence between species increases (Primmer and Merila, 2002). In plants, the success rate of transferability of polymorphic markers between species within a genus is approximately 60% in eudicots and close to 40% in monocots. The transfer rate between genera reduces dramatically to 10% in eudicots (Barbara *et al*, 2007). Tang *et al* (2010) showed a high transferability rate (66.7%) of *Phyllostachys pubescens* microsatellites derived from GenBank database in six related *Phyllostachys* species. Microsatellite markers developed from chokecherry (*Prunus virginiana* L.) can transfer across *Prunus* species and eleven other rosaceous species (63.2% and 58.7%, respectively). An average of 42.7% of amplifiable primers can amplify DNA from other rosaceous species (Wang *et al*, 2012).

3.2.4 Background of *Tilia*

Tilia L. (lime or basswood) is a genus of large trees with approximately 20-25 species in the family Malvaceae of the order Malvales (Bremer *et al*, 2003). These species are distributed in the temperate climate of the Northern Hemisphere, throughout Europe, some parts of Asia and North America. *Tilia cordata* (small leaved lime) and *T. platyphyllos* (large leaved lime) are the two species that are widely distributed in Europe. They can hybridize, which results in the common lime, *T. x europaea* (Pigott, 1969). Species relationships within the genus are unknown and complicated by frequent hybridization. Although there are some morphological differences between *T. cordata* and *T. platyphyllos*, it is difficult to identify species in the absence of flowers and from characteristics of leaves at ground level; particularly hybrids can vary in morphology (Pigott, 1991).

There are no nuclear markers suitable for population genetics currently available within this genus. Therefore, microsatellite markers are needed to increase our understanding of population genetics studies in this genus.

The aims of this study are to develop a set of microsatellite markers of the genus. These markers will be tested for polymorphism and their transferability to other species in the genus for further population genetic studies.

3.3 Materials and methods

3.3.1 Plant materials

Polymorphism was studied in two populations of *T. platyphyllos* in France, one from Issole (Alpes de Haute Provence) (44° 2'N 6° 49'E) (N = 20) and the other from Gorges de la Carança (Pyrénées Orientales) (42° 51'N 2° 22'E) (N = 20). For testing of microsatellite primer transferability, leaves were collected from 20 species and two hybrids of this genus (51 individuals) from a living collection in Cartmel, Cumbria UK as described in Table 2.2 (Chapter 2). No voucher specimens have been deposited because only leaves from ground level were available.

3.3.2 DNA extraction

Genomic DNA was extracted from a fresh leaf bud of a mature *T. platyphyllos* tree at Chanstone Wood (52° 0’ 52.36’’N, 2° 56’ 37.98’’W) for constructing the library and genomic DNA of other samples was extracted from dried leaves using a CTAB (cetyltrimethyl ammonium bromide) procedure (Morgan-Richards and Wolff, 1999) as described in Chapter 2. The extracted DNA was dissolved in 100 µl of TE buffer and stored at -20 ^{o}C until use.

3.3.3 Development of microsatellites and genotyping

A microsatellite enrichment library was constructed by Dr. Kirsten Wolff. The protocol was based on Edwards *et al* (1996) and Squirrell and Wolff (2001). In short, 100 ng of purified DNA was digested with *Mbo*I (NEB, Beverly, Massachusetts, USA) and ligated to *Sau*LA and *Sau*LB linkers (Squirrell and Wolff, 2001). The ligated DNA was PCR amplified and the product was hybridized at 47 °C to nylon membranes with dot blots of $(GA)_{15}$ and $(CA)_{15}$ oligos. The eluted enriched DNA was PCR amplified and enriched for a second time, following the same method. The doubly enriched DNA was digested with *Mbo*I and ligated in pUC 19 vector (Qiagen, Hilden, Germany). The plasmids were transformed into competent *E. coli* cells (Bioline, London, UK). Recombinant colonies were selected by blue/white screening and M13 PCR amplification was used to estimate the size of the inserts. Inserts with a size between 300 and 700 bp were sequenced using BigDye terminator V3.1 (Applied Biosystems, Foster City, California, USA).

Clones from the enriched library were sequenced on an ABI PRISM 3130 Genetic Analyser (Applied Biosystems), and inserts with 9-20 dinucleotide repeats were chosen for primer design. Primers were designed using Primer3 software version 0.4.0 using default settings but selecting primers longer than 21 nucleotides (Rozen and Skaletsky, 2000).

The PCR amplification was performed in 10 µl multiplex reactions containing 5 ng of DNA, 1x reaction buffer, 2mM $MgCl_2$, 0.2 mM of each dNTP, 0.1-0.2 mM of each primer and 0.5 U of *Taq* DNA polymerase (Bioline). Forward primers were labeled with FAM, HEX (Integrated DNA technologies, Coralville, Iowa, USA) or NED (Applied Biosystems) (see Table 3.1 for multiplex sets, specific concentration of primers and

fluorescent label). The PCR reaction for all multiplex sets was as follows: an initial pre-denaturation step at 95 °C for 5 min, followed by 15 cycles of 95 °C for 15s, 54 °C for 15s and 72 °C for 15s and 20 cycles of 89 °C for 20s, 52 °C for 20s, 72 °C for 20s, and a final extension step of 72 °C for 30 min. Complete PCR reactions were diluted 1:10 with distilled water and 1 µl of this dilution was mixed with 10 µl of Hi-Di Formamide™ (Applied Biosystems) and 0.1 µl of ROX-500 size standard before analysis on an ABI 3130xl Genetic Analyzer (Applied Biosystems). The data were analyzed by GeneMapper® Software (Applied Biosystems) and GenAlEx6.5 (Peakall and Smouse, 2012). The presence of null alleles was tested using Micro-Checker (Van Oosterhout *et al*, 2004).

Table 3.1 Characterization of 15 microsatellite markers developed in the genus *Tilia*. [a] Size of the original fragment, [b] Fluorescent label on the forward primer.

Locus	Primer sequence (5'-3')	Repeat motif	Size (bp)[a]	Primer conc. (mM)	Fluorescent dye[b]	multiplex sets	GenBank Accession No.
Tc4	F: ATTTTAGAATGCCAACCTGCTAAG	$T_6(GT)_{12}$	224	0.2	HEX	B	JQ289157
	R: TATTGAAGTCCATTTCCAATTGTC						
Tc5	F: TTTTCATACATTTAGAGACTTTTAGCA	$(AG)_{12}$	150	0.2	FAM	D	JQ289158
	R: TGCATGATTTGTATGTTTAGGG						
Tc6	F: CCATATCTTCTGCCAGTTTTCC	$(AG)_{12}$	143	0.2	HEX	A	JQ289159
	R: GGACTAATTTCTTCCTTTTATTAGGC						
Tc7	F: TTTACTTTTGCCAGTTGTGAGG	$(GA)_{13}$	234	0.1	FAM	D	JQ289160
	R: CACCTAGAATGCCTCCTATTCG						
Tc8	F: CGAAGAAACTGTCAAAACAACG	$(GA)_{13}$	160	0.1	HEX	B	JQ289161
	R: AGCTGGGTTTTAGAGGATAGGG						
Tc11	F: AGCTATGAAAGAACTATCAAGAGAAAG	$(AG)_{13}$	146	0.1	NED	C	JQ289162
	R: CCCCAAGACATTGCAGTAGAAC						
Tc31	F: TTTGCAAAGACTACTCCAAGAATC	$(GA)_{12}$	205	0.2	FAM	B	JQ289164
	R: AAATCGATGGTCAAGAACTAAATC						
Tc915	F: ACATCGATTGTATTTCCCTTTAAC	$(CT)_{16}$	165	0.2	HEX	C	JQ289165
	R: GTTGTATTTTGCCCTTAACATTG						
Tc918	F: AACGGCTAATTACTCCTAGTTTCG	$(AC)_9(TC)_2$	240	0.2	HEX	A	JQ289166
	R: TGTTCAGCTCACTACTACCTTTCAC						
Tc920	F: AAATGTCTTCAGAGTGACTAGATGG	$(GA)_2(GT)_{15}(AG)_4$	232	0.1	FAM	A	JQ289167
	R: TGCCTCATTATTCTCCTAATTCTC						
Tc927	F: AGTCCTCCTGTCAAATGCTG	$(AG)_{10}$	157	0.1	FAM	C	JQ289168
	R: ATCACACTCGTTTATGACATCTTG						
Tc937	F: AGCCAACCAACTTTTACAATACAG	$(AG)_{13}$	162	0.1	NED	A	JQ289169
	R: AGATAAAAGCACATAAATCGATGG						
Tc943	F: ATTTCATCTTTCTCTAAAGCCTTG	$(CA)_{10}$	150	0.2	FAM	B	JQ289170
	R: GGGAAAGCCTGTGTTAGTTTC						
Tc951	F: TGTTATGACCTCACTTATAACCAAGT	$(CT)_{12}$	160	0.2	NED	D	JQ289171
	R: GGGTGAGCTGACAATATAGAAGAG						
Tc963	F: CTAACCCCATTCTCTTTAATTCTG	$(CT)_{11}$	238	0.2	HEX	C	JQ289172
	R: GCTTTCATTTCAGTTTTCCTCTAC						

3.4 Results

A total of 104 clones from the enriched library were sequenced and 96 clones contained microsatellite repeats with a minimum of five repeats. In total 31 primer pairs were

designed. These primers were tested on two populations of *T. platyphyllos* from France. Fifteen out of 31 primer pairs provided good patterns with the expected product size and were used for further characterisation. The allele size ranges of each locus in four multiplex sets are shown in Figure 3.1. The other 16 primer pairs failed to amplify targets, amplified non-target sequences, or had nonspecific banding patterns in preliminary tests and were discarded.

All 15 loci were highly polymorphic in the samples analysed. One locus (Tc943) was monomorphic within the Issole population, but was polymorphic in the Gorges de la Carança population. In these two populations, the number of alleles of the 15 loci ranged from 1 to 15, with a mean of 8.96. The observed and expected heterozygosities varied from 0.25 to 1.00 (average = 0.71) and 0.18 to 0.90 (average = 0.70), respectively. Significant departures from Hardy-Weinberg equilibrium were detected at Tc918 and Tc963 in the Issole population ($P<0.05$). For the Gorges de la Carança population, Tc4 and Tc920 deviated from HWE ($P<0.01$), which may indicate the presence of null alleles for those loci (Table 3.2). However, no homozygous null genotypes were detected. In addition, Micro-Checker also did not indicate any null alleles in these populations.

These 15 loci were also tested for their amplification in 20 species and two hybrids in the genus *Tilia*. Most of the loci (12 out of 15) were transferable and polymorphic in most of the related species (Table 3.3). Tc915 failed to amplify in two species (*T. americana* and *T. caroliniana*) and Tc920 failed to amplify in three species (*T.* x *euchlora*, *T. dasystyla* and *T. endochrysea*). Tc918 was successfully amplified in only four species (*T.*x *euchlora*, *T.*x *europaea*, *T. dasystyla* and *T. platyphyllos*).

Table 3.2 Results of 15 microsatelite markers in two populations from France (Issole and Gorges de la Carança). *A*, number of alleles; H_e, expected heterozygosity; H_o, observed heterozygosity; HWE, Hardy–Weinberg equilibrium; *N*, sample size; ns, nonsignificant departure from HWE; * $P < 0.05$; ** $P < 0.01$.

Locus	France: Issole (*N = 20*)				France: Gorges de la Carança (*N = 20*)			
	A	H_o	H_e	HWE	*A*	H_o	H_e	HWE
Tc4	8	0.85	0.82	ns	11	0.70	0.85	**
Tc5	15	1.00	0.86	ns	10	0.90	0.82	ns
Tc6	8	0.70	0.82	ns	7	0.75	0.76	ns
Tc7	9	0.80	0.76	ns	7	0.85	0.79	ns
Tc8	6	0.90	0.77	ns	8	0.80	0.80	ns
Tc11	9	0.95	0.81	ns	8	0.65	0.73	ns
Tc31	9	0.80	0.80	ns	9	0.90	0.80	ns
Tc915	12	0.90	0.88	ns	12	0.75	0.87	ns
Tc918	2	0.10	0.18	*	2	0.25	0.22	ns
Tc920	9	0.65	0.59	ns	5	0.30	0.54	**
Tc927	14	0.85	0.90	ns	11	0.75	0.82	ns
Tc937	10	0.85	0.81	ns	9	0.95	0.80	ns
Tc943	1	0.00	0.00	Monomorphic	2	0.30	0.26	ns
Tc951	8	0.60	0.64	ns	6	0.75	0.71	ns
Tc963	14	0.85	0.87	*	15	0.90	0.88	ns

Table 3.3 Amplification of 15 microsatellite loci across 20 species and two hybrids in the genus *Tilia*. –, failed amplification; +, successful amplification with one allele; ++, successful amplification with more than one allele; [a] Numbers in parentheses show number of samples tested.

Species[a]	Tc4	Tc5	Tc6	Tc7	Tc8	Tc11	Tc31	Tc915	Tc918	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
T. x *euchlora* (1)	++	++	+	++	++	++	++	++	+	-	++	++	++	++	++
T. x *europaea* (4)	++	++	++	++	++	++	+	++	+	++	++	++	++	++	++
T. americana L. (2)	++	++	++	++	+	++	++	-	-	++	++	++	++	++	++
T. amurensis Rupr. (4)	++	++	++	++	+	++	++	++	-	++	++	++	++	++	++
T. callidonta Hung T. Chang (1)	+	+	++	++	+	++	++	++	-	++	++	++	++	+	++
T. caroliniana Mill. (5)	++	++	++	++	++	++	++	-	-	+	++	++	++	++	++
T. chingiana Hu & Cheng (1)	++	++	++	++	+	+	++	++	-	++	+	+	++	++	++
T. concinna Pigott (1)	++	++	+	++	+	++	++	++	-	++	++	++	++	++	++
T. cordata Mill. (6)	++	++	++	++	+	++	++	++	-	++	+	++	++	++	++
T. dasystyla Steven (2)	++	++	++	++	++	++	++	++	++	-	++	++	++	++	++
T. endochrysea Hand.-Mazz. (1)	++	++	++	++	+	++	++	++	-	-	+	++	+	+	+
T. henryana Szyszył. (1)	++	++	++	++	+	++	++	++	-	++	++	++	++	++	++
T. japonica (Miq.) Simonk. (2)	++	++	++	++	+	++	++	++	-	++	++	++	++	++	++
T. kiusiana Makino & Shiras. (1)	++	++	++	++	+	+	++	++	-	+	++	++	++	+	+
T. mandshurica Rupr. & Maxim. (1)	+	++	++	++	+	++	++	+	-	++	++	++	+	+	++
T. maximowicziana Shiras. (2)	++	++	++	++	+	++	++	++	-	++	++	++	+	++	++
T. mongolica Maxim. (1)	++	++	++	++	++	++	++	++	-	++	++	++	++	++	++
T. nobilis Rehder & E. H. Wilson (1)	+	+	++	++	+	+	++	++	-	+	++	++	++	++	++
T. oliveri Szyszył. (1)	++	++	++	++	+	+	+	++	-	+	++	+	++	+	++
T. platyphyllos Scop. (7)	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
T. tomentosa Moench (4)	++	++	++	++	+	++	++	+	-	++	++	++	++	+	++
T. tuan zyszył. (2)	++	++	++	++	+	++	++	++	-	++	++	++	++	++	++

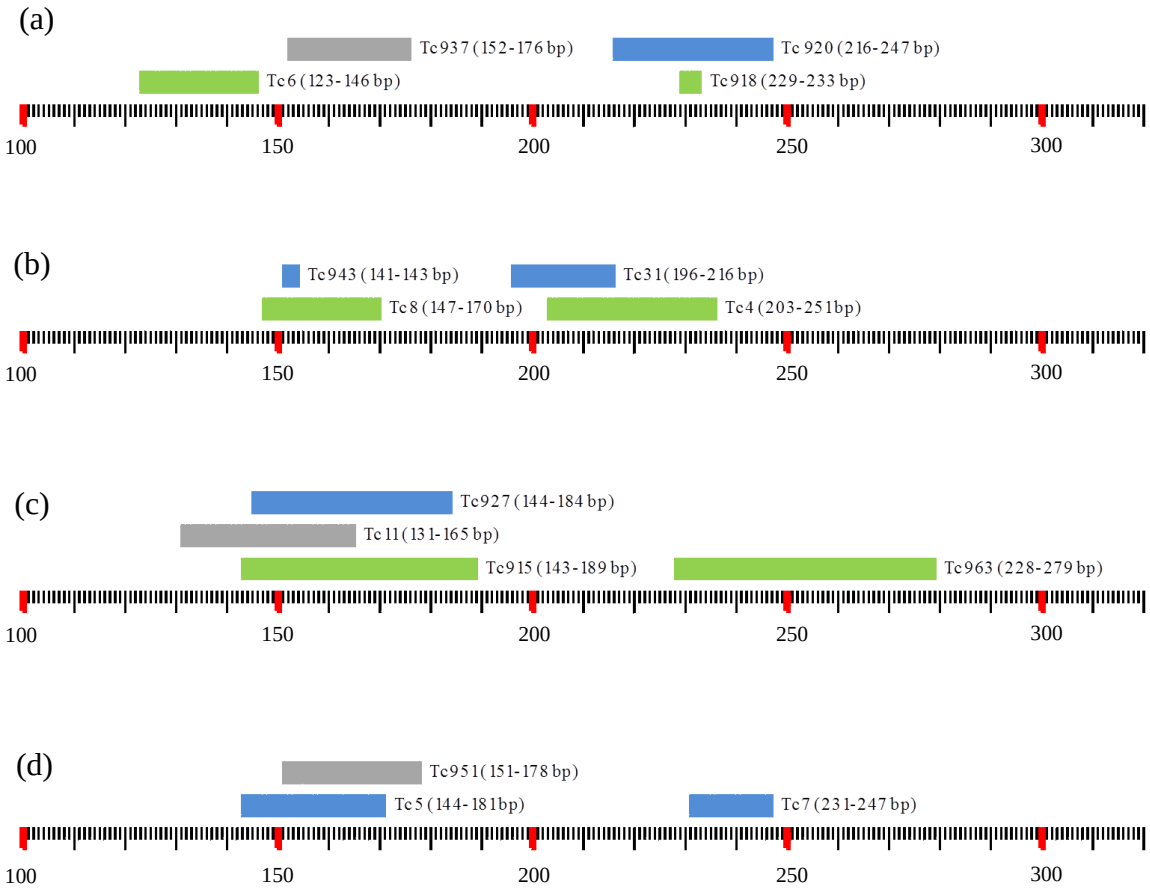


Figure 3.1 The allele size range of each microsatellite primer in four multiplex reactions as described in Table 3.1. (a) = multiplex set A, (b) = multiplex set B, (c) = multiplex set C and (d) = multiplex set D. Colour are assigned to microsatellite primers labelled with FAM ( ), NED ( ) and HEX ( ) fluorescent dyes.

3.5 Discussion

Of the 31 microsatellite primers designed, 15 pairs showed polymorphism. The other 16 primer pairs failed to amplify targets, amplified non-target sequences, or had nonspecific band patterns. These are commonly reported problems of designing microsatellite primers. The absence of PCR products may be the result of null alleles. A mutation in the flanking regions (primer binding sites) can lead to poor or failure of amplification (Jarne and Lagoda, 1996).

The success of amplification of multiple microsatellites in a single PCR reaction (multiplexing) can significantly reduce cost and time for genetic analysis. In general, for

using multiplex PCR, microsatellite primers are combined in such a way that the overlapping allele size ranges are labelled with different fluorescent dyes, while those with non-overlapping size ranges can be labelled with the same fluorescent dye (Guichoux *et al*, 2011). However, in this study, the microsatellite primers were labelled with fluorescent dyes before combining the multiplex sets.

Most of the microsatellite loci developed from *T. platyphyllos* can amplify in many related species. Primer Tc918 was the only primer that amplifies in just four species. Because this primer amplifies in *T. platyphyllos* but not in *T. cordata* this locus may be useful for the identification of the two species and their hybrid. The successful amplification of these loci could facilitate genetic studies in other *Tilia* species as it can save time and cost for developing new primer sets.

3.6 Conclusion

These 15 highly polymorphic microsatellite markers will be useful tools for the study of population structure in *T. platyphyllos* and increase our understanding of their phylogeography and of the hybridization between *Tilia* species. Furthermore, the success of cross-amplification in related species will assist the future study of genetic diversity across the genus.

Chapter 4. Assessing genetic diversity and differentiation between *Tilia cordata* and *T. platyphyllos* across Europe using microsatellite markers

4.1 Abstract

In this study I determined the genetic diversity and differentiation of two *Tilia* species. Twenty-five populations of *T. cordata* and 15 populations of *T. platyphyllos* were collected from natural woods across Europe. Thirteen microsatellite markers developed from *T. platyphyllos* were used for the genetic analysis. These loci show a high level of polymorphism in both *T. cordata* and *T. platyphyllos* and can clearly discriminate these two species. The level of genetic diversity in *T. platyphyllos* is higher than in *T. cordata* for all diversity measures (N_a, A_e, H_o and H_e). However, this could at least partly be the result of ascertainment bias because the markers were isolated from *T. platyphyllos*. Fixation indices were close to zero, indicating an outcrossing mating system. Averaged across all loci, the genetic differentiation between the two species was high and significant (F_{st} = 0.308, P<0.001), which supports the results from Principal Coordinates Analysis (PCO) and Structure analysis. The markers have different levels of differentiation between species and alleles were shared between the two species.

4.2 Introduction

Genetic variation is an important concept for population genetic study. Genetic diversity and genetic differentiation are two components that have been widely studied in various organisms. The term genetic diversity is commonly used to describe the amount of heritable variation within and between populations or species (Brown, 1983; Lowe *et al*, 2004). It is of fundamental importance for populations or communities of the species for adaptation to changing or new environments (Crawford and Whitney, 2010; Dlugosch and Parker, 2008).

The study of genetic variation is important in various fields: evolutionary biology, ecology and conservation biology. It can improve understanding of the process of plant evolution and adaptation to the environment. For example, the study of genetic differentiation between *Quercus petraea* and *Q. robur* indicated low genetic differentiation between these oak species. Although they shared most alleles, the

detection of a low rate of gene flow suggested that the shared alleles resulted from shared ancestral polymorphism (Muir and Schlotterer, 2005).

The analysis of genetic variation also provides an understanding of environmental adaptation and natural selection. By comparing the level of differentiation between species or among populations within a species, it is possible to identify loci important for adaptive divergence or specific traits (Charlesworth *et al*, 1997; Storz, 2005). Molecular markers that are supposed to be selectively neutral could be linked to adaptive genes or genome regions under selection. In addition, the knowledge of genetic variation is essential for establishment of effective conservation programmes. This information can be used as a guideline for the exploitation of genetic resources (Rao and Hodgkin, 2002).

4.2.1 Tools to detect genetic variation

The assessment of population genetic diversity is usually performed at the molecular level. A wide range of molecular markers have been used, from allozymes, which is the first type of genetic marker introduced, to markers which directly measure the variation in the DNA sequences. The markers based on DNA sequence polymorphism, such as restriction fragment length polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphisms (AFLP) and Simple Sequence Repeats (SSR), are commonly used for genetic studies in plant science.

Among the molecular markers, microsatellites or Simple Sequence Repeats (SSR) are potentially the most informative. The popularity of microsatellite markers is because they are widely present and randomly dispersed in the genome. Also, they have co-dominant inheritance, their variation is easy to assess by PCR and they have high polymorphism due to high mutation rate (Tautz, 1989).

4.2.2 Microsatellite ascertainment bias

Microsatellite markers are useful for genetic analysis due to their cross-amplification in closely related species. Although microsatellite repeats have a high rate of mutation, their flanking regions are more stable. Therefore, the primers designed from one species often amplify in closely related species (Coote and Bruford, 1996; Davis *et al*, 2002).

During the isolation of microsatellite loci, the sequences that contain longer repeats are preferred over the shorter and interrupted ones. After specific primers have been designed, they will be optimized and tested with a set of samples. Only markers that show polymorphism will be selected. These processes lead to the average allele length being longer in the original species than that in the other closely related species. This phenomenon is known as ascertainment bias (Vowles and Amos, 2006). The first finding of ascertainment bias in microsatellites was used in the comparison of microsatellite length between human and chimpanzees, using markers developed from humans (Rubinsztein *et al*, 1995). The study revealed that microsatellite loci observed in humans were significantly longer than those in chimpanzees. The researchers pointed out that this could be the result of the different rates of microsatellite mutation in these closely related species. However, Ellegren *et al* (1995) argued that this difference in microsatellite length could be due to ascertainment bias. Therefore, cross-species comparisons of genetic diversity with markers developed from one species need careful consideration.

4.2.3 The study of genetic diversity in *Tilia*

T. cordata and *T. platyphyllos* are the two main *Tilia* species that are distributed in Europe (Pigott, 1991). They are diploid (2n = 82) (Pigott, 2012) and the evidence of mating system is unclear. Pigott (2012) suggested that *T. cordata* in natural population in the UK contains both self-compatible and self-incompatible individuals. Only a few molecular studies have been reported in these species. A genetic study of *T. cordata* and *T. platyphyllos* has been carried out using isozyme markers (Fromm and Hattemer, 2003; Maurer and Tabel, 1995). Chloroplast DNA variation using PCR-RFLP was studied in *T. cordata* populations in 2003 (Fineschi *et al*, 2003). RAPD analysis was used to study genetic diversity in *T. rubra* (Hosseinzadeh Colagar *et al*, 2013) and these markers were also used to clarify some systematic relationships within the genus (Liesebach and Sinko, 2008). In 2012, ITS regions were used to infer the phylogeny of *Tilia* species from Iran (Yousefzadeh *et al*, 2012). Our study is the first use of nuclear microsatellite markers to determine genetic variation in the two *Tilia* species across Europe.

In this study, we aimed to assess the genetic diversity between *T. cordata* and *T. platyphyllos* from natural populations across Europe using 13 nuclear microsatellite markers. The genotypic data were also used to investigate genetic differentiation

between the two species and within species. In addition, the result is also used to confirm that these microsatellite markers can clearly assign individuals to species.

4.3 Materials and methods

4.3.1 Plant materials

Leaves were collected from 25 populations of *T. cordata* (427 samples) and 15 populations of *T. platyphyllos* (222 samples) in natural woods across Europe by Dr. Kirsten Wolff and collaborators. For most locations, the species were already described based on morphology (leaf characters and direction of inflorescence) (refer to Table 6.3, Chapter 6). The two species are mixed in some locations (DEBO, CZST, SKBL, FRCE, AUTH, AUDO and AUSO). In these cases, individuals were genetically analysed before species classification. Samples from Switzerland were collected from several populations in north west Switzerland and maintained in an orchard (CHLO and CHFR). The number of individuals sampled varied from one to 31 per population. Only one *T. cordata* tree was sampled from Spain as this species is very rare there. A leaf of each sample was dried and preserved at 4 °C until use.

Table 4.1 Countries, locations, population codes, coordinates (latitude and longitude) and sampling sizes of *T. cordata* and *T. platyphyllos* after hybrids and further generation samples were omitted.

Species	Country	Location	Label	Latitude (N)	Longitude (E)	No. of samples
T. cordata	Norway	Sogn og Fjordane, Asane nature reserve	NOSO	61.85753	6.19563	16
	Finland	Niinisaari	FINI	61.81667	29.38333	15
		Muukonsaari	FIMU	61.16001	28.48196	14
	Denmark	Åbybjerg	DEAB	57.10241	9.35310	30
		Bolderslev	DEBO	55.0168	9.38901	20
	Germany	Colbitz Wasserwerk parking area	GECO	52.33027	11.55722	20
	United Kingdom	Highbury Wood	UKHB	52.17621	-3.34183	16
	the Netherlands	Margraten	NEMA	50.80000	5.75000	25
		Gulpen	NEGU	50.79595	5.89453	25
	Poland	Lezajsk	PLLE	50.26360	22.38580	10
	Czech Republic	Velky Osek	CZVO	50.10143	15.17778	21
		Šternberk, Vrapač	CZST	49.72490	17.01900	14
	Slovakia	Blatnica, Gaderská dolina	SKBL	48.94860	18.96550	6
	France	Foret Dom de Mouthiers	FRMO	48.91388	4.91441	20
		Cessieres	FRCE	49.55877	3.48875	8
	Ukraine	Rosilna	UARO	48.76970	24.39570	14
	Austria	Thayatal Park near Heardegg	AUTH	48.84747	15.88003	15
		Sommerein	AUSO	48.59052	16.69553	23
		Dobra	AUDO	47.98599	16.69553	21
		Stams	AUST	47.27566	10.97716	22
	Switzerland	Lobsigen Park	CHLO	47.30774	8.26685	25
	Hungary	Oltárc	HUOL	47.00000	19.00077	20
	Italy	Lasen, Dolomites	ITLA	46.07000	11.92000	19
		Veltre, Dolomites	ITVE	46.04000	11.95000	2
	Spain	Huesca	SPHU	42.58778	0.18889	1
T. platyphyllos	Denmark	Bolderslev	DEBO	55.01680	9.38901	10
	United Kingdom	Kings' Wood	UKKI	53.39842	-1.18188	18
	Germany	Lichtenstein	GELC	51.30277	13.01875	31
	Czech Republic	Šternberk, Vrapač	CZST	49.72490	17.01900	2
	France	Cessieres	FRCE	49.55877	3.48875	13
		Gorges de la CaranÇa	FRCC	43.37431	2.45211	20
		Issole	FRIS	43.27211	0.11420	20
	Slovakia	Blatnica, Gaderská dolina	SKBL	48.94860	18.96550	7
	Austria	Thayatal Park near Heardegg	AUTH	48.84747	15.88003	2
		Sommerein	AUSO	48.59052	15.39742	15
		Leopoldsberg	AULE	48.27707	16.35467	25
		Dobra	AUDO	47.98599	16.69553	14
	Switzerland	Frieswil	CHFR	47.07213	7.14241	24
	Spain	Huesca	SPHU	42.58778	0.18889	7
	Greece	Aggistro	GRAG	41.35664	23.48132	10

4.3.2 DNA extraction and microsatellite genotyping

Total genomic DNA was extracted from dried *Tilia* leaves using a CTAB (cetyltrimethyl ammonium bromide) procedure (Morgan-Richards and Wolff, 1999) as described in Chapter 2. The extracted DNA was dissolved in 100 µl of TE buffer and stored at -20 ^{o}C until use. All samples were genotyped using 13 out of 15 microsatellite loci (Chapter 3) (Phuekvilai and Wolff, 2013). Two primer pairs (Tc11 and Tc918) were not included in the analysis because null alleles were detected in these loci and they only amplify in *T. platyphyllos*. PCR reactions were performed as described in Chapter 3. GeneMapper® Software (Applied Biosystems) was used to determine fragment sizes and alleles were manually scored and edited before population genetic analysis. The presence of null alleles was tested using Micro-Checker (Van Oosterhout *et al*, 2004).

4.3.3 Data analysis

After scoring alleles in all 649 samples, initial observations showed that there were nine hybrids and further generation plants (refer to Chapter 6). All individuals in this category (≥ 10% clustering by STRUCTURE with the minority class) were omitted from further population genetic analysis, leaving 422 *T. cordata* and 218 *T. platyphyllos* (Table 4.1). Genetic diversity was analyzed in terms of the number of alleles per locus, allele ranges, average number of alleles (N_a), effective number of alleles (A_e), observed heterozygosity (H_o) and expected heterozygosity (H_e), fixation index (F), and deviation from Hardy-Weinberg by using the computer software GenAlEx version 6.5 (Peakall and Smouse, 2012). The significance of fixation index (F) was tested using FSTAT version 2.9.3.2 (Goudet, 2002). The significance of the difference between the number of alleles per locus of the two *Tilia* species was carried out using pairwise t-tests in Minitab version 16 (Minitab Inc.). GenAlEx was also used to estimate the genetic differentiation between species and between populations within each species in term of F_{st} and to construct a graph of allele frequency of the loci. The genetic differentiation was measured using F_{st} instead of R_{st} because Gaggiotti *et al* (1999) found that in a study using a small number of microsatellite loci (<20 loci) and small sample sizes (<10), F_{st} is more suitable and provides a more conservative approach than R_{st}. Also, in some cases microsatellites may gain or lose a large number of repeats (Ellegren, 2004) and therewith not strictly follow the stepwise mutation model as used in R_{st} estimation.

Genetic differentiation in term of D was also calculated using SMOGD (Crawford, 2010) to compare the results with F_{st}. The genetic distance matrix was constructed to carry out a Principal Coordinates Analysis (PCO) with GenAlEx. Analysis of molecular variance (AMOVA) with 1,000 permutations and linkage disequilibrium was carried out using Arlequin 3.5 (Excoffier and Lischer, 2010).

The identification of clusters (species) and admixed individuals was also analyzed using Bayesian clustering software, STRUCTURE Version 2.3.3 (Pritchard *et al*, 2000). In this software, the number of clusters (K) is estimated and individuals will be assigned to one or more clusters. The number of clusters was set from one to four (total number of expected species plus two). The program was run with 20,000 burn-in period, followed by 20,000 iterations using the Markov Chain Monte Carlo (MCMC) method with the admixture model and correlated allele frequencies. Other parameters were set as default values. To estimate the optimal grouping (K), software STRUCTURE HARVESTER (Earl and vonHoldt, 2012) was used. This program detects the number of K groups that shows the best fit with the data and produces plots of ΔK and K values. The optimal grouping (K) was chosen as the highest ΔK value.

Populations with number of samples less than five (ITVE and SPHU, *T. cordata*, and CZST and AUTH, *T. platyphyllos*) were not included in genetic diversity analysis in order to minimize the variance of gene diversity (Pons and Petit, 1995) but were included in genetic differentiation analysis.

4.4 Results

4.4.1 Genetic diversity estimates

A total of 640 samples of *T. cordata* and *T. platyphyllos* were genotyped using 13 microsatellite markers. Genetic diversity was compared between the two species. The results showed high levels of polymorphism. Marker Tc963 had the highest number of alleles (47 alleles) in both *T. cordata* and *T. platyphyllos.* The number of alleles in *T. cordata* ranged from two (Tc8) to 41 (Tc963), with an average of 5.47 (Table 4.2 and 4.4). The number of alleles in *T. platyphyllos* is between ten (Tc943) and 30 (Tc963), with an average of 8.09 (Table 4.2 and 4.5). A total number of 317 alleles weas identified, with an average of 6.02 alleles per locus (Table 4.3). *T. cordata* had fewer alleles per locus than *T. platyphyllos* (t-value -2.35, $P < 0.05$). *T. cordata* only had more

alleles for Tc943 and Tc963 than *T. platyphyllos*. The largest allele size range was also detected for Tc963, from 171 to 311 bp (Table 4.2). Overall, *T. cordata* usually has shorter fragments (fewer repeats) and a smaller size range than *T. platyphyllos*.

Table 4.2 Number of alleles and allele size ranges of 13 microsatellite loci in 422 *T. cordata* individuals and 218 *T. platyphyllos* individuals

Locus	No. of alleles			Allele size ranges		
	T. cordata	*T. platyphyllos*	Overall	*T. cordata*	*T. platyphyllos*	Overall
Tc4	18	23	25	211-240	203-251	203-251
Tc5	24	25	30	134-189	134-195	134-195
Tc6	17	18	21	122-154	122-171	122-171
Tc7	10	21	23	217-237	226-272	217-270
Tc8	2	16	17	139-141	141-174	139-174
Tc31	11	16	19	192-229	194-227	192-229
Tc915	23	26	31	133-189	143-191	133-191
Tc920	15	20	22	205-242	177-272	177-272
Tc927	4	27	29	142-153	144-197	142-197
Tc937	10	15	16	148-164	150-184	148-184
Tc943	12	10	16	134-158	129-154	129-158
Tc951	9	18	21	146-164	146-186	146-186
Tc963	41	30	47	171-311	222-281	171-311

The highest average number of alleles (N_a), effective number of alleles (A_e) and heterozygosity were detected at locus Tc963 (Table 4.3). The average expected heterozygosity (H_e) per locus over all populations in both species ranged from 0.29 (Tc8) to 0.85 (Tc963) (Table 4.3), with an average of 0.59. The three lowest H_e were observed in Tc8, Tc927 and Tc943 (0.29, 0.40 and 0.45, respectively). The H_e at locus Tc8 was lowest because only two alleles were detected in *T. cordata* populations. This locus was fixed for that allele in 24 out of 25 *T. cordata* populations. This similar pattern, with high degree of fixation in populations of *T. cordata*, was also observed at Tc927. The fixation index (F) over all populations ranged from -0.13 (Tc5) to 0.09 (Tc4), with an average over all loci of -0.04. Tc4 significantly departed from Hardy-Weinberg equilibrium in eight out of 36 populations ($P < 0.001$), exhibiting a deficit of heterozygotes as indicated by positive values of fixation index in those populations (Appendix 5).

Table 4.3 Diversity measures and F value (fixation index) of 13 microsatellites as average over loci in 422 *T. cordata* individuals and 218 *T. platyphyllos* individuals. N_a, Average number of alleles; A_e, Effective number of alleles; H_o, observed; H_e, expected; F, fixation index

Locus	N_a	A_e	H_o	H_e	F
Tc4	7.33	4.50	0.666	0.731	0.085
Tc5	7.73	4.54	0.809	0.729	-0.125
Tc6	7.00	4.31	0.786	0.743	-0.068
Tc7	4.33	2.80	0.560	0.532	-0.046
Tc8	3.40	2.37	0.308	0.289	-0.040
Tc31	4.70	2.74	0.501	0.496	0.008
Tc915	8.55	5.32	0.827	0.779	-0.069
Tc920	6.50	3.94	0.783	0.716	-0.112
Tc927	4.95	3.03	0.408	0.400	-0.024
Tc937	5.05	2.91	0.593	0.538	-0.107
Tc943	2.90	1.94	0.478	0.446	-0.086
Tc951	3.75	2.11	0.463	0.472	0.018
Tc963	12.10	7.70	0.816	0.847	0.021
Average	6.02	3.71	0.615	0.594	-0.043

T. platyphyllos showed a higher level of genetic diversity as average across populations than *T. cordata*. The effective number of alleles (A_e) was smaller in *T. cordata* (average of 3.30) than *in T. platyphyllos* (average of 4.88) (Table 4.4 and 4.5). A_e ranges between 2.48 (FINI) and 3.85 (CHLO) among the populations of *T. cordata*, whereas *T. platyphyllos* had a wider range, from 3.15 (SPHU) to 6.05 (AULE). The average expected heterozygosity (H_e) over 13 loci in *T. cordata* varied between 0.43 and 0.61, with a mean of 0.54. FINI had the lowest H_e, while CZST had the highest H_e. The average H_e in *T. platyphyllos* ranged from 0.64 to 0.79, with a mean of 0.73 (Table 4.5). The population with the lowest H_e was SPHU (7 individuals), while the population with the highest H_e was AULE. A t-test showed significant higher N_a, A_e, H_e in *T. platyphyllos* than in *T. cordata* ($P < 0.001$).

The fixation index (F) of most populations in both species was close to zero with non-significant deviation from zero, which means populations were under random mating and this indicates an outcrossing mating system. A significant positive value of F ($P<0.001$) was detected in CZST (0.22) among *T. cordata* populations and in GRAG

(0.12) among *T. platyphyllos* populations. This indicated a deficit of heterozygotes in these populations.

Pairwise linkage disequilibrium of markers in each population was calculated. The results indicated that 328 out of 2,808 possible pairs (11.7%, $P < 0.05$) showed significant evident of linkage. Each population had different linked loci and the number of linked loci was low; therefore, each microsatellite locus was considered unlinked to others.

Table 4.4 Diversity measures for 13 microsatellite markers of 23 populations of *T. cordata*. N, number of samples analysed; N_a, Average number of alleles; A_e, Effective number of alleles; H_o, observed heterozygosity (mean over loci); H_e, expected heterozygosity (mean over loci); F, fixation index (mean over loci). A significant positive or negative deviation of F from zero is indicated as *** P<0.001.

Population	N	N_a	A_e	H_o	H_e	F
NOSO	16	4.69	2.71	0.553	0.518	-0.078
FINI	15	4.54	2.48	0.456	0.428	-0.049
FIMU	14	4.85	2.78	0.533	0.517	0.022
DEAB	30	6.38	3.45	0.548	0.540	-0.023
DEBO	20	6.31	3.83	0.523	0.559	0.018
GECO	20	5.31	3.10	0.569	0.552	-0.056
UKHB	16	4.38	2.96	0.476	0.517	0.080
NEMA	25	6.38	3.54	0.551	0.566	0.008
NEGU	25	5.23	3.34	0.511	0.543	0.020
PLLE	10	5.15	3.43	0.631	0.570	-0.115
CZVO	21	5.62	3.13	0.480	0.504	0.020
CZST	14	5.85	3.69	0.467	0.606	0.221***
SKBL	6	4.62	3.27	0.526	0.547	0.045
FRMO	20	5.77	3.63	0.500	0.559	0.129
FRCE	8	4.08	2.87	0.510	0.499	0.014
UARO	14	5.46	3.67	0.595	0.603	-0.005
AUTH	15	5.46	3.61	0.544	0.550	0.000
AUSO	23	7.08	3.71	0.575	0.569	-0.013
AUDO	21	6.62	3.51	0.524	0.564	0.045
AUST	22	5.38	3.37	0.549	0.575	0.046
CHLO	25	6.46	3.85	0.551	0.579	0.061
HUOL	20	5.15	3.29	0.508	0.538	0.070
ITLA	19	5.15	2.68	0.513	0.508	0.011
Average	18.20	5.47	3.30	0.530	0.544	0.022

Table 4.5 Diversity measures for 13 microsatellite markers of 13 populations of *T. platyphyllos.* N, number of samples analysed; N_a, Average number of alleles; A_e, Effective number of alleles; H_o, observed heterozygosity (mean over loci); H_e, expected heterozygosity (mean over loci); F, fixation index (mean over loci). A significant deviation of F from zero is indicated as *** P<0.001.

Population	N	N_a	A_e	H_o	H_e	F
DEBO	10	4.85	3.41	0.808	0.662	-0.219
UKKI	18	6.92	4.29	0.696	0.733	0.055
GELC	31	9.92	4.98	0.749	0.733	-0.020
FRCE	13	7.54	4.69	0.746	0.716	-0.023
FRCC	20	9.46	5.27	0.750	0.731	-0.027
FRIS	20	8.62	4.94	0.738	0.745	0.008
SKBL	7	7.31	5.48	0.791	0.717	-0.116
AUSO	15	9.15	5.36	0.753	0.740	-0.028
AULE	25	10.85	6.05	0.785	0.790	0.005
AUDO	14	7.23	4.72	0.775	0.714	-0.088
CHFR	24	10.62	5.55	0.772	0.755	-0.025
SPHU	7	4.54	3.15	0.681	0.636	-0.071
GRAG	10	8.23	5.52	0.638	0.751	0.123***
Average	16.43	8.09	4.88	0.745	0.725	-0.032

4.4.2 Genetic differentiation

The AMOVA revealed high genetic differentiation between the two species explaining 25.21% of the total genetic variance. This variation was almost five times higher than the variation among populations within species (5.62%), while the remainder of the variation (69.17%) was within populations (Table 4.6).

Table 4.6 AMOVA analysis using 13 microsatellite loci of all 36 *Tilia* populations

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among species	1	874.24	1.49	25.21%
Among populations within species	38	554.62	0.33	5.62%
Within populations	1240	5070.98	4.10	69.17%
Total	1279	6499.84	5.92	100%

F_{st} and D indicate genetic differentiation between *T. cordata* and *T. platyphyllos*. Higher F_{st} values (P<0.001) were observed between the two species (average of 0.31) than within each species (average of 0.07 and 0.08 within *T. cordata* and *T. platyphyllos*,

respectively). D had similar results (Table 4.7 and 4.8). The highest average D was observed between the two species at 0.44, while those within *T. cordata* and *T. platyphyllos* were 0.11 and 0.32, respectively.

The three highest levels of interspecific differentiation were observed for Tc8, Tc927 and Tc943 with F_{st} values of 0.68, 0.54 and 0.53, respectively (Table 4.7). This can be caused by specific alleles being present only in one species with a high frequency. For example, allele 141 for Tc8, allele 142 for Tc927 and allele 143 for Tc943 were found only in *T. cordata*, while other alleles with low frequencies were found in *T. platyphyllos* (Figure 4.1 a, b and c).

The three loci with the lowest F_{st} between species were Tc4, Tc963 and Tc915 with F_{st} values of 0.10, 0.08 and 0.13, respectively. The low level of interspecific differentiation at these loci is because alleles of these loci were mostly shared between the two species (Figure 4.1 d, e and f).

The genetic diversity between populations within species expressed as F_{st} and D showed different results at some loci (Table 4.7 and 4.8). The highest F_{st} among populations within *T. cordata* was observed at Tc8 (0.37). In contrast, D at this locus showed almost the lowest. The lowest F_{st} among populations within *T. cordata* was observed in Tc963 (0.04), whereas this alleles had the highest D value (0.38). F_{st} and D within *T. platyphyllos* populations were quite similar in all loci except for Tc943, which had the lowest D, while its F_{st} was in the middle range.

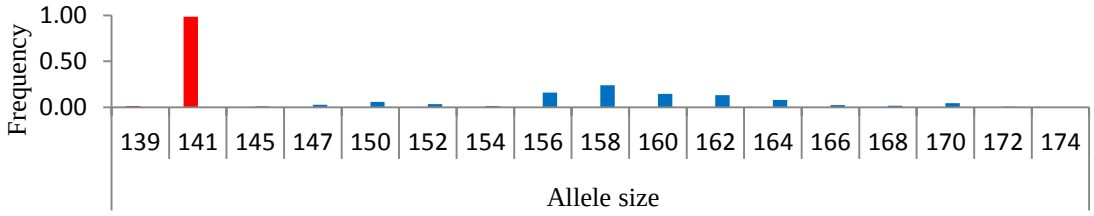
Table 4.7 F_{st} values among populations within *Tilia* species and between species resulting from analysis including all samples and analysis of the two species separately.

Locus	Intraspecific		Interspecific
	T. cordata	*T. platyphyllos*	*T. cordata - T. platyphyllos*
Tc4	0.072	0.070	0.099
Tc5	0.073	0.077	0.185
Tc6	0.064	0.089	0.129
Tc7	0.121	0.080	0.389
Tc8	0.367	0.090	0.675
Tc31	0.036	0.069	0.436
Tc915	0.081	0.055	0.125
Tc920	0.072	0.087	0.158
Tc927	0.045	0.070	0.541
Tc937	0.060	0.066	0.386
Tc943	0.137	0.076	0.533
Tc951	0.052	0.089	0.288
Tc963	0.035	0.072	0.083
Total	0.073	0.076	0.308

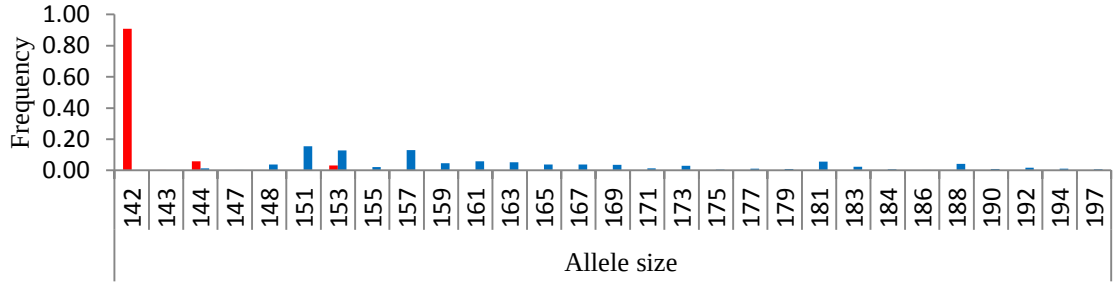
Table 4.8 D values among populations within *Tilia* species and between species

Locus	Intraspecific		Interspecific
	T. cordata	*T. platyphyllos*	*T. cordata - T. platyphyllos*
Tc4	0.272	0.430	0.389
Tc5	0.228	0.536	0.542
Tc6	0.213	0.464	0.390
Tc7	0.134	0.434	0.514
Tc8	0.016	0.419	0.427
Tc31	0.023	0.371	0.430
Tc915	0.376	0.396	0.531
Tc920	0.328	0.269	0.446
Tc927	0.011	0.510	0.413
Tc937	0.048	0.358	0.425
Tc943	0.166	0.031	0.520
Tc951	0.070	0.126	0.294
Tc963	0.383	0.579	0.643
Total	0.108	0.323	0.442

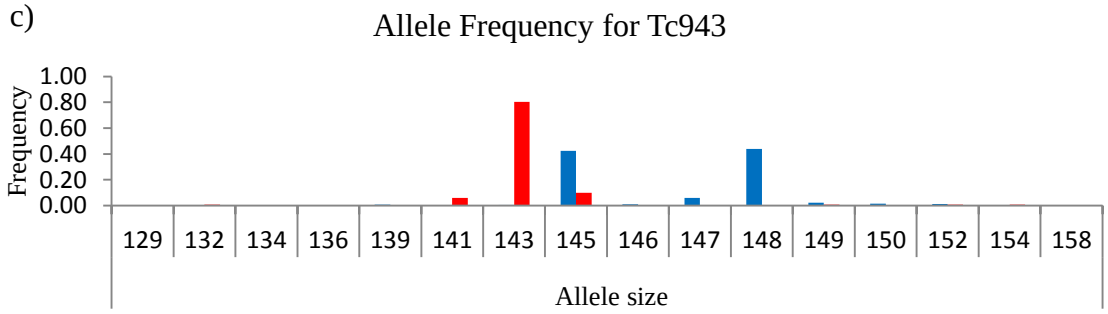
a)


Allele Frequency for Tc8
Frequency
1.00
0.50
0.00
139 141 145 147 150 152 154 156 158 160 162 164 166 168 170 172 174
Allele size

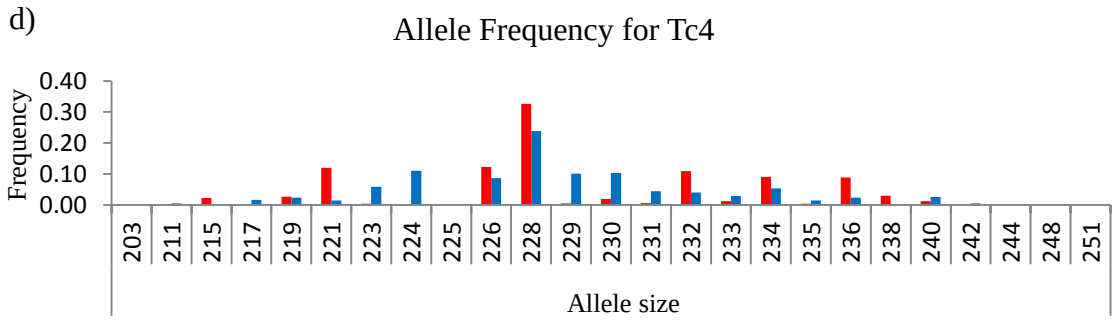
b)


Allele Frequency for Tc927
Frequency
1.00
0.80
0.60
0.40
0.20
0.00
142 143 144 147 148 151 153 155 157 159 161 163 165 167 169 171 173 175 177 179 181 183 184 186 188 190 192 194 197
Allele size

c)


Allele Frequency for Tc943
Frequency
1.00
0.80
0.60
0.40
0.20
0.00
129 132 134 136 139 141 143 145 146 147 148 149 150 152 154 158
Allele size

d)


Allele Frequency for Tc4
Frequency
0.40
0.30
0.20
0.10
0.00
203 211 215 217 219 221 223 224 225 226 228 229 230 231 232 233 234 235 236 238 240 242 244 248 251
Allele size

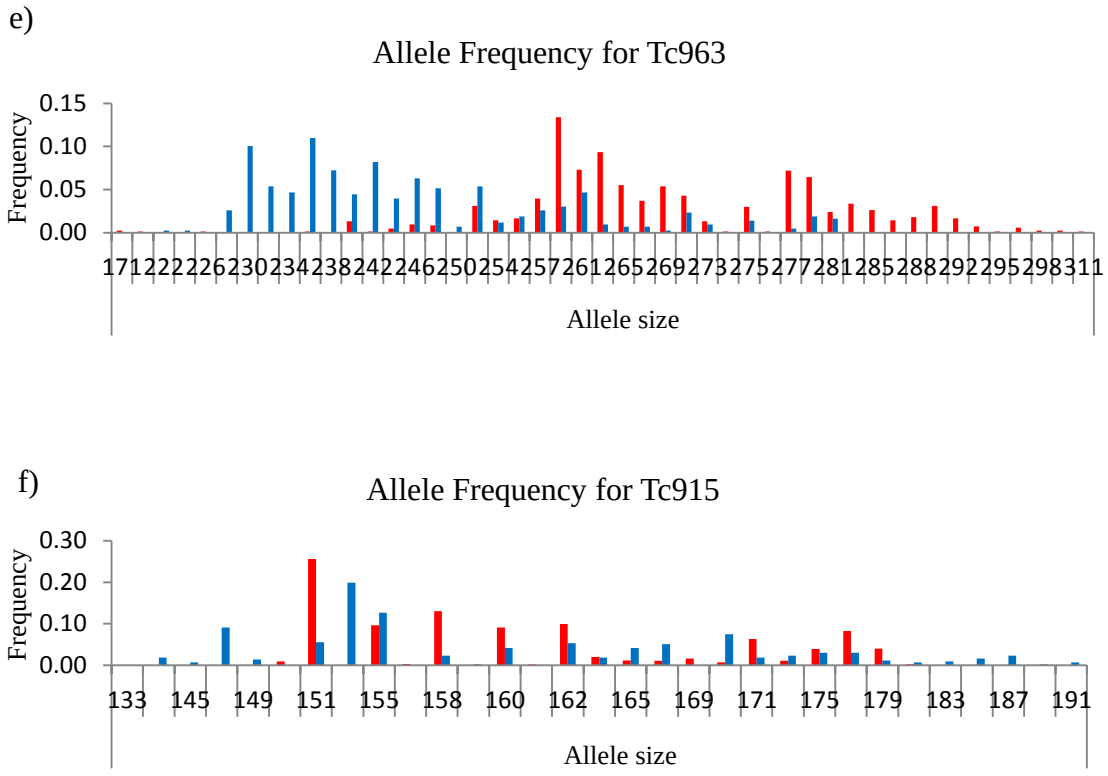


Figure 4.1 Allele frequency at loci Tc8 (a), Tc927 (b), Tc6 (c), Tc4 (d), Tc963 (e) and Tc915 (f). Colours indicate species: ■ *T. cordata* and ■ *T. platyphyllos*

The genetic similarity of the two *Tilia* species (640 samples) was visualized by Principal Coordinates Analysis (PCO) (Figure 4.2). The first two axes of PCO accounted for 75.70% of total variation (67.92% and 7.78% for the first and second axis, respectively) and showed a clear grouping, separating the species.

Bayesian analysis of genetic structure demonstrated that the model with K = 2 (the highest ΔK) (Figure 4.4) gave the optimal grouping. This again revealed that 40 populations of *Tilia* were clearly clustered into two groups according to the species. The 25 populations of *T. cordata* were assigned to cluster I, while the 15 populations of *T. platyphyllos* were assigned to cluster II (Figure 4.3). The result was identical to the species differentiation illustrated by PCO (Figure 4.2).

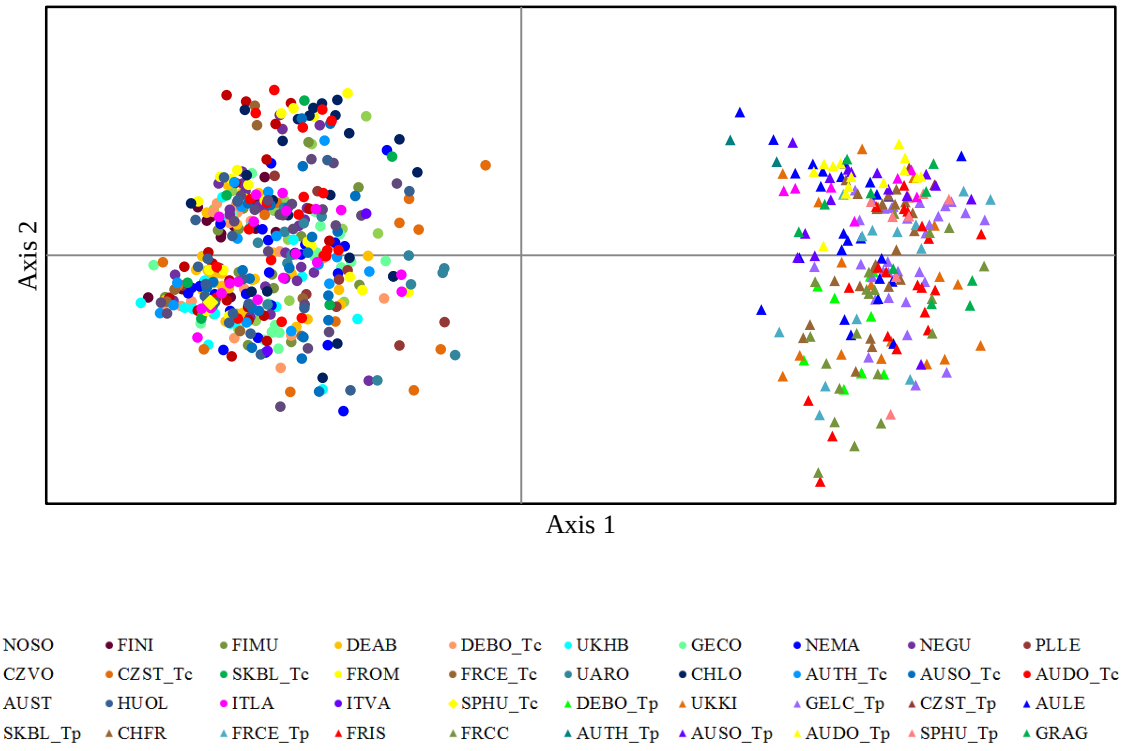


Figure 4.2 Principal Coordinates Analysis (PCO) of genetic distance using 13 microsatellite markers (GenAlEx) of 640 samples in both *T. cordata* and *T. platyphyllos* populations, coding the populations by colour. Dots indicate *T. cordata* samples and triangles indicate *T. platyphyllos* samples.

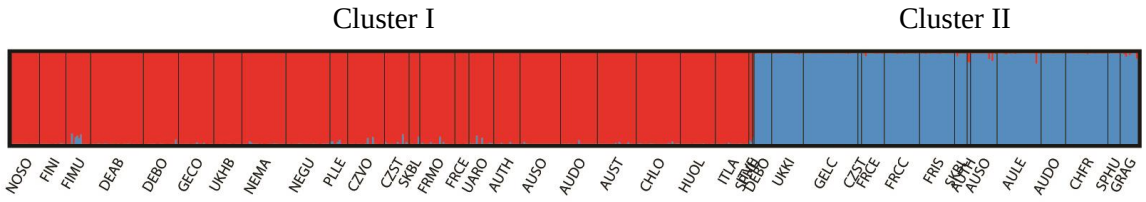


Figure 4.3 Genetic structure analysis between the two *Tilia* species (640 samples) using 13 microsatellite loci, estimated using STRUCTURE. The model with K = 2 showed the highest ΔK value.

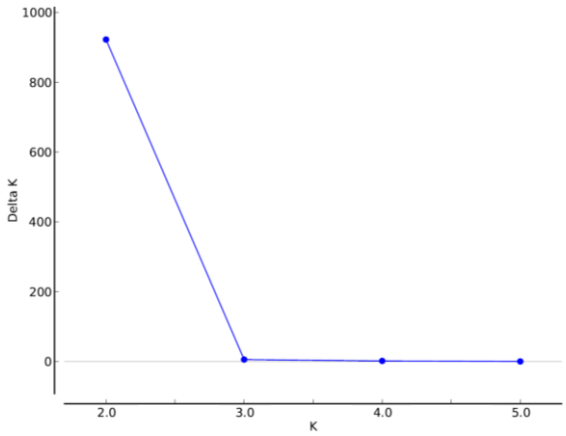


Figure 4.4 Value of Delta K against K of all *Tilia* individuals

4.5 Discussion

4.5.1 Overall genetic diversity

Microsatellite data indicated that both *T. cordata* and *T. platyphyllos* had a high level of genetic diversity. Many microsatellite alleles were detected (16 – 47 alleles per locus) with wide ranges of allele sizes. No samples with more than two alleles per locus were found, which supported that both *T. cordata* and *T. platyphyllos* are diploid (Pigott, 2012). Also, a high observed (H_o) and expected heterozygosity (H_e) over all loci in all populations was observed (0.62 and 0.59, respectively). Most of the fixation indices (F) of the populations were close to zero (Appendix 5 and 6). The results are the first genetic evidence that both *T. cordata* and *T. platyphyllos* have an outcrossing mating system.

Most loci were in Hardy-Weinberg equilibrium in both species, which also indicates that the populations are under random mating and largely outcrossing. However, substantial deviations from Hardy-Weinberg equilibrium ($P<0.001$) with high fixation index ($F>0.5$) found in some populations at a few loci. The high fixation index (heterozygote deficit) at Tc4 of FINI, DEBO, CZVO, CAZT, AUTH, FROM *T. cordata* populations and at Tc951 of FRCE *T. platyphyllos* population could be due to null alleles. Micro-Checker also confirmed that null alleles may be present at these loci. However, from the analysis of 640 samples in both *T. cordata* and *T. platyphyllos* populations using 13 markers, homozygous null alleles were not detected and therefore presence of null alleles in certain populations cannot be excluded, but they were not at a high frequency.

4.5.2 Comparison of genetic diversity between the two species

The level of genetic diversity in *T. platyphyllos* is higher than in *T. cordata* for all measures of diversity (N_a, A_e, H_o and H_e, Table 4.4 and 4.5). In addition, the length and size range of microsatellite alleles observed in *T. platyphyllos* tend to be greater than those in *T. cordata* (Table 4.2). The higher diversity in *T. platyphyllos* than in *T. cordata* could be the result of ascertainment bias: microsatellite repeats tend to be longer in the species from which they have been developed. The longer alleles tend to be more variable than the shorter alleles, thus more polymorphism and higher heterozygosity should be observed in the microsatellite isolated species than in related species (Ellegren *et al*, 1995). In this analysis, all 13 microsatellites used were developed from *T. platyphyllos*. The method to isolate microsatellites from genomic DNA was based on hybridization with repeat oligo probes (as described in Chapter 3). Using this method, longer microsatellite repeats can potentially hybridize better with the probes than shorter repeats. Also, when choosing markers, microsatellites with small numbers of repeats are avoided. This selection of microsatellite loci may create a bias and result in a higher number of repeats than average, compared to that in closely related species.

Another explanation of the different genetic diversity between species is the different time of migration and recolonisation. According to the pollen record, the most abundant pollen found during the Pleistocene period was of *T. cordata*. *T. platyphyllos* is sensitive to low temperature (Pigott, 2012) and this could indicate that *T. platyphyllos* recolonized the north later than *T. cordata*. Therefore, lower genetic diversity would be expected in *T. platyphyllos* because it would have fewer generations compared to those of *T. cordata*. However, we find a higher diversity in *T. platyphyllos* and therefore time of migration does not explain the difference we observed.

Population size can be another explanation for a difference in genetic diversity. In general, small populations have a strong effect of genetic drift, which leads to the loss of variation. *T. platyphyllos* is classified by the IUCN Red List of Threatened Species as Critically Endangered in Norway and Sweden, Endangered in Albania, and Least Concern in many countries, such as Belgium, Denmark, Germany, Switzerland and the United Kingdom (Khela, 2013). The smaller population size of *T. platyphyllos* would lead to lower genetic diversity in *T. platyphyllos* than in *T. cordata*. In contrast, the

results in this study showed higher genetic diversity in *T. platyphyllos* than in *T. cordata*. This can be explained only by bias resulting from using microsatellite markers developed from *T. platyphyllos*. However, to prove this, new microsatellite markers developed from *T. cordata* should be used.

Other studies have also shown that the use of microsatellite markers across species may affect genetic diversity measures, such as heterozygosity, allele size variance or number of segregating alleles (Li and Kimmel, 2013). A number of studies of ascertainment bias have been carried out in various species. For example, Ellegren *et al* (1997) studied the differentiation in length of microsatellite loci between cattle and sheep. Hutter *et al* (1998) studied the variation of microsatellite loci between *Drosophila melanogaster* and *D. simulans.* Thus, it is not surprising that the alleles were longer in *T. platyphyllos*, where markers were isolated from, than those observed in *T. cordata*. The ascertainment bias could lead to the higher genetic diversity detected in *T. platyphyllos* than in *T. cordata*. Therefore, development and use of additional microsatellite markers from *T. cordata* could clarify this ascertainment bias hypothesis.

A significant positive fixation index was observed in CZST and GRAG (*T. cordata* and *T. platyphyllos* populations, respectively). This could be explained by the Wahlund effect, which resulted from population substructure. This phenomenon occurs when more than one population with different allele frequencies are treated as a single population. This result was also support by the deviation from Hardy-Weinberg equilibrium in CZST at several loci (Appendix 5). The high fixation index is not the result of clones because no clone was detected in the population.

4.5.3 Genetic differentiation

Genetic differentiation between and within *T. cordata* and *T. platyphyllos* was investigated using 13 microsatellite loci. The results from PCO and Structure analysis indicated that the microsatellite loci clearly separated the individuals to species categories. This was supported by the high F_{st} of 0.308 over all loci, which indicated a significant differentiation between the two species. The observation of higher interspecific differentiation (F_{st}) than intraspecific differentiation in all loci confirmed that the combination of all 13 microsatellite loci was efficient at species discrimination.

Even with the removal of the loci with the highest F_{st} (Tc8, Tc927 and Tc943) from the analysis, individuals still clustered by species (data not shown).

The analysis of AMOVA revealed that the genetic variation between *T. cordata* and *T. platyphyllos* explained 25% of the total genetic variance ($P < 0.001$) and the interspecific F_{st} value was high compared to those for microsatellite markers in other tree species. For example, the average F_{st} between two oak species, *Quercus petraea* and *Q. robur*, was 0.10 (Neophytou *et al*, 2010). This high and significant F_{st} over all loci in *Tilia* indicates low gene flow between *T. cordata* and *T. platyphyllos* and supports the limited hybridisation and introgression between the two species observed in sympatric UK populations (Chapter 6).

The allele frequencies of individual loci indicated that a high level of differentiation between *T. cordata* and *T. platyphyllos* is observed in some genome regions, while other genome regions are seemingly shared between species. A high frequency of one allele with a high degree of fixation was observed only in *T. cordata* at some loci, namely Tc8, Tc927 and Tc943. Thus these loci can be considered as species specific. This phenomenon is also shown in the high interspecific F_{st} for these loci.

Under directional selection, the frequency of a favoured allele will increase and become more common or eventually fixed. The fixation of an allele can also be the result of genetic drift. However, genetic drift will reduce the diversity of the whole genome rather uniformly. In our case, the three loci with the highest F_{st} (Tc8, Tc927 and Tc943) showed fixation of specific alleles in one species. These loci may reside within the genomic regions associated with genes responsible for morphological traits that strongly discriminate between the two species (Muir and Schlotterer, 2005). For example, genes that control reproductive isolation have been documented in various closely related species. Neophytou *et al* (2010) observed the fixation of specific microsatellite loci between two oak species, *Q. petraea* and *Q. robur*. The results revealed that the species discriminant locus QrZAG96 locates within the genome associated with a morphological quantitative trait locus (QTL) for petiole length, which is different between the two species. Therefore, in our study the three loci with a high degree of fixation in one species could represent the genome regions linked to selected loci maintaining species integrity or reproductive isolation, such as gene control of flowering time. Further studies on mapping the markers and quantitative trait loci (QTL) could be carried out to support these results.

For some loci, allele ranges were shared between the two species, which resulted in a low F_{st} value. The three lowest F_{st} were observed in Tc4, Tc963 and Tc915. The shared alleles between the two species could be explained by shared ancestral variation or gene flow between the two species (Muir and Schlotterer, 2005). However, the gene flow between the two species seems to be limited, with nine hybrids and further generations detected among 649 individuals across Europe and low numbers of hybrids detected in sympatric populations in the UK (refer Chapter 6). Therefore, shared ancestral variation could be the reason for the shared alleles.

The measurement of population differentiation using F_{st} and D has been debated recently. F_{st} is calculated based on within population diversity (expected heterozygosity), which can lead to difficulties in comparison of the index among species or among loci. On the other hand, D uses a multiplication partitioning of genetic diversity, based on effective number of alleles, which scales linearly with the genetic diversity, while expected heterozygosity does not (Meirmans and Hedrick, 2011). In this study, the two indices were calculated. The genetic differentiation within species showed differences between F_{st} and D at some loci. The highest F_{st} within *T. cordata* was observed at Tc8 (0.37) compared to other loci. In general, the more alleles are detected, the lower the maximum F_{st}. With more than two alleles, the F_{st} cannot reach one. Thus, the high F_{st} in Tc8 could be due to the small number of alleles (2 alleles). On the other hand, almost the lowest D (0.016) was observed at this locus, which reflected the similarity among populations. From the observation of allele frequency at this locus, most populations were almost fixed at one allele (141 bp) (Figure 4.1 a). This phenomenon indicated that D better reflected the similarity among populations than F_{st}. Low values of D were also observed in Tc927 in *T. cordata* and Tc943 in *T. platyphyllos* with a high frequency of one distinct allele. Meirmans and Hedrick (2011) suggested that F_{st} reflects the fixation and not the differentiation in allele frequencies among populations in the case where different alleles are fixed in different populations. However, F_{st} performs better than D for demographic inference (Whitlock, 2011). For example, if there are three alleles in three populations and the same alleles go to fixation in three subpopulations, in this case, F_{st} value will be high (one), while D will be zero. On the other hand, if different alleles go to fixation in different subpopulations both F_{st} and D in this case will be one. This suggests that F_{st} would be better for describing the distribution of genetic variation, while D would be better for measuring the differentiation of allelic frequencies among

populations. Meirmans and Hedrick (2011) also suggested that in highly variable markers alternative indices such as D should be used in addition to F_{st}.

4.6 Conclusion

Genetic diversity and differentiation between *T. cordata* and *T. platyphyllos* were studied for the first time using microsatellite markers. This study revealed that the two *Tilia* species have an outcrossing mating system. They were clearly distinct and high levels of genetic diversity were detected by using 13 microsatellite loci. Higher genetic diversity was observed in *T. platyphyllos* than in *T. cordata*. However, this could be an artefact of developing markers from *T. platyphyllos*. Ideally, *T. cordata* microsatellite markers would be used to support the hypothesis that the lower genetic diversity in *T. cordata* is because of ascertainment bias. Although microsatellite markers revealed clearly differentiation between the two *Tilia* species, some shared alleles were also observed. The allele sharing could be explained by ancestral polymorphism or these loci are potentially linked to genes that have adaptive advantages. Some species specific alleles were detected in some loci. This could suggest directional selection of some specific genome regions, which maintain species integrity. Further study on QTL mapping or genome wide association analysis could be carried out to support these results.

Chapter 5. Population structure and phylogeography of *Tilia cordata* and *T. platyphyllos* across Europe

5.1 Abstract

In this study we investigate patterns of population genetic structure and gain insight into postglacial recolonization in *Tilia* across Europe. Thirteen microsatellite markers were used to analyse 24 *T. cordata* and 15 *T. platyphyllos* populations. In addition, eight chloroplast regions were analysed with two individuals per country from each species. We identified low genetic diversity in the peripheral populations, which supports the rapid expansion of tree species from southern Europe during postglacial periods. Some chloroplast haplotypes were shared between the two species. The haplotype network suggested that these shared haplotypes could be the result of incomplete lineage sorting rather than recent hybridization. Bayesian analysis revealed strong genetic structure in *T. platyphyllos* but weaker in *T. cordata*. This could be because migration and colonization in the northern areas by *T. cordata* occurred before those by *T. platyphyllos*. Microsatellite analysis suggested different possible colonization routes between two *Tilia* species. However, *T. cordata* and *T. platyphyllos* seem to share the three main refugia in southern Europe (Iberia, Italy and Balkans). In addition, *T. cordata* seems to have additional putative refugia in eastern areas (Caucasus).

5.2 Introduction

Population genetic structure is the distribution of genotypes across space through time (Hewitt and Butlin, 1997). The patterns of population structure are shaped by various factors from natural processes to human impact. Phylogeography is the study of the historical and contemporary processes responsible for the geographic distribution of genealogical lineages (Avise *et al*, 1987). Thus, phylogeographic study is a powerful tool to understand population structure and evolutionary processes in living species.

The climatic fluctuation during the Pleistocene glaciation is believed to have an important impact on distribution ranges and differentiation in both plants and animals species. The multiple oscillations between warm and cold glacial conditions during these periods caused the extinctions of some European tree flora, while some dispersed

to new locations and some survived in refugia and expanded again during the warm periods. These phenomena had a strong effect on population structure of the species that are living today (Hewitt, 2000). Phylogeography has been studied under the hypothesis that considers this period as a major factor for species distribution.

5.2.1 Ice ages and glacial refugia

The Quaternary cold periods in Europe (2.4 Myr to present), particularly in the last glacial cycle (115,000 to 15,000 years ago) (Birks, 1986), appear to have had a dramatic influence on the distribution of plants and animals. During the last ice age, the polar ice sheets spread across northern Europe, including the whole of Scandinavia and northern Britain. Mountain blocks, such as the Alps, Pyrenees and Carpathians, were also covered with ice caps (Hewitt, 2000). In that period, European temperate forests were considerably more restricted than today. They occupied mainly the southern peninsulas, such as Iberia, Italy, and the Balkans (the southern refugia hypothesis), where the climate was more suitable to survive (Bennett *et al*, 1991). This long-term isolation in the different southern refugia led to genetic differentiation among populations. The expansion and colonisation of the species from southern refugia toward the north occurred during the warm interglacial periods and post glaciation (began 15,000 to 10,000 years ago).

During the post-glacial periods, populations at the northern limit range of the refugia expanded into the northern areas and colonized new suitable territory. This expansion occurred remarkably rapidly as the climate warmed suddenly. The leading edge populations would have dispersed and colonized the areas far from the main refugia. The pioneers would rapidly expand to fill the new territory before others arrived, thus their genotypes would dominate in the new areas. These newly colonized populations would be favoured by natural selection and might act as barriers, which prevent the later migrants from occupying the colonized areas. The repeat of successive founder events over a long colonisation route of the expansion led to the loss of genetic variation of the leading populations. This rapid expansion has been considered to reduce genetic diversity in large areas of northern Europe (Hewitt, 1999). The end of the last ice age allowed trees to move northwards about 8000-13000 years ago. This evidence was observed and documented from the pollen records (Huntley and Birks, 1983).

5.2.2 Molecular markers suitable for phylogeographic studies

Phylogeographic patterns in plants have been studied using various molecular approaches. Chloroplast DNA is one of the useful tools for phylogeographic studies and identification of postglacial colonization routes in plants (Petit *et al*, 2002; Taberlet *et al*, 1998). Since chloroplast DNA is non-recombinant and maternally inherited in most angiosperms, it is transmitted through the maternal line only. Therefore, clearer population structure and colonization patterns of chloroplast DNA can be observed than from nuclear genes. Considering these characteristics, chloroplast DNA seems to be an ideal tool for the study of evolutionary processes in the historical events that shaped the genetic structure of plant species. However, due to the absence of recombination of chloroplast DNA, it can only represent a single gene genealogy. Thus, using this genome will hardly capture all historical events that occurred within the species (Heuertz *et al*, 2006). In addition, the low mutation rates of chloroplast genomes in some species seems to be one of the obstacles to obtaining sufficient variation to explain the historical events of the species (Provan *et al*, 2001).

Nuclear markers have been suggested for phylogeographic studies since they show recombination and are biparentally inherited and, therefore, integrate several genealogical processes. The availability of highly polymorphic microsatellite markers and the development of Bayesian approaches provide opportunities to investigate the population genetic processes during recolonization that affected the current patterns of population genetic structure (Heuertz *et al*, 2006). The combination of both chloroplast and microsatellite markers could reveal more information about the historical events of the species than using only a single marker (Bai *et al*, 2010).

5.2.3 Postglacial colonization patterns of plants in Europe

The influence of the Quaternary period on the geographic distribution of plants across Europe has received a lot of attention in the past decade. The pollen record is the traditional data source that has effectively explained the expansion of tree species. Huntley and Birks (1983) mapped the distribution of pollen of trees across Europe at 500 year intervals for each tree species, for the last 13,000 years. The patterns indicated most forest trees expanded from the southern areas and they suggested the possible refugia for each tree species. The combination of molecular data and the evidence of the

pollen record revealed the postglacial migration history and recolonization of many plant species. Also, these data can suggest locations of putative refugia in southern Europe.

The studies of phylogeographic inferences of forest tree species across Europe revealed that most of the species shared three distinct major refugia in the southern peninsulas (Iberia, Italy and Balkan) during the last ice age. Also, the tree species shared some postglacial migration lineages during recolonization towards the north (Taberlet *et al*, 1998). For example, the homogeneity of chloroplast DNA observed in common beech (*Fagus sylvatica*) from northern Spain, France, German, Poland and Balkans together with the pollen evidence (Huntley and Birks, 1983) indicated a postglacial colonization route from the Balkan refugia, while the migration from the Iberian and Italian refugia was inhibited due to the rapid expansion from the Balkan population that filled the area first (Demesure *et al*, 1996). The colonization route from the main refugia in the Balkans is similar to that of the black alder (*Alnus glutinosa*) (King and Ferris, 1998). The study of chloroplast DNA variation in oaks (*Quercus* spp.) suggested postglacial colonization routes from at least three southern refugia, the Iberian peninsula, Italy and the Balkans (Dumolin-Lapegue *et al*, 1997a). The expansion to northern and western Europe came from the Iberian peninsula. Other species have a similar lineage, such as hazel (*Corylus avellana*) (Palme and Vendramin, 2002) and English holly (*Ilex aquifolium* L.) (Rendell and Ennos, 2003). The recolonization of oaks into eastern Europe from the Balkan refugia was also observed in ash (*Fraxinus excelsior*) (Heuertz *et al*, 2004) and silver fir (*Abies alba*) (Konnert and Bergmann, 1995).

In *Tilia*, pollen evidence shows that during the end of the last glaciation, about 18,000 BP, *Tilia* was present only in Greece and southern Italy. The rapid expansion of *Tilia* from the Balkan peninsula across Bulgaria to the Black Sea coast and the expansion from Italy to the southern and southeastern borders of the Alps occurred at the end of the Late-glacial (11,500 BP). Then *Tilia* spread to the north and northwest of Europe and reached southern Scandinavia and England by 8,000 BP (Figure 5.1). *Tilia* reached the edge of the species range in Finland and Scandinavia then stopped the migration at around 6,000 BP (Huntley and Birks, 1983). The migration rate of *Tilia* during the first half of the Holocene, based on pollen material, is between 350 and 820 metres per year (Lang, 2003). As *Tilia* is an insect pollinated plant, the abundance of pollen in deposits is more limited than for wind pollinated plants. However, its location is more reliable as

pollen is not prone to be wind dispersed. In addition, apart from ice age refugia, hybridisation may have impacted on the distribution of variation in *Tilia* species.

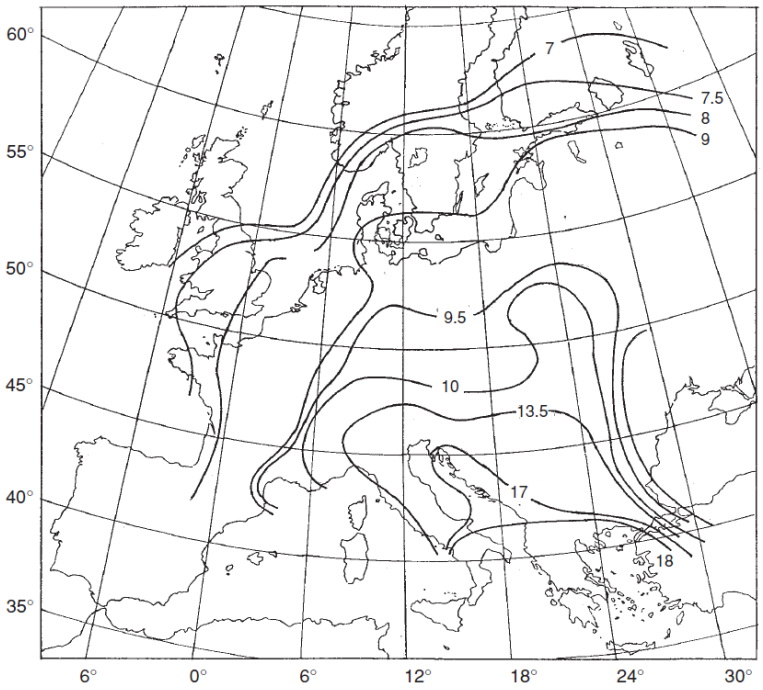


Figure 5.1 Distribution of *Tilia* during the Late and Postglacial periods based on pollen data. Lines indicate millennia before present (Huntley and Birks, 1983).

Molecular markers are one of the reliable tools for clarification of the distribution of variation of a species. The only molecular study on *Tilia* phylogeography is based on chloroplast variation in *T. cordata* using PCR RFLP. The results indicated that the distribution of 14 chloroplast haplotypes revealed low geographic structure and low genetic differentiation among 17 populations (Fineschi *et al*, 2003).

The objectives of our study are to detect the patterns of population genetic structures of *T. cordata* and *T. platyphyllos* across Europe by using microsatellites and chloroplast DNA. It is also to interpret the observed geographical patterns in terms of recolonisation after postglacial periods and identify possible refugia of the two *Tilia* species.

5.3 Materials and methods

5.3.1 Plant materials: microsatellite analysis

For population structure and phylogeographic analysis using microsatellite markers, the genotypic data of 13 microsatellite loci in 24 populations of *T. cordata* (421 individuals) and 15 populations of *T. platyphyllos* (218 individuals) obtained in Chapter 4 (Table 4.1) were analysed. One population of *T. cordata* from Spain, which contains only one individual, is not included in the microsatellite analysis.

5.3.2 Plant materials: chloroplast analysis

Two individuals per country from each *Tilia* species were used, except for *T. cordata* in Spain where only one individual was available. In total 51 individuals from 16 populations (31 individuals) of *T. cordata* and 10 populations (20 individuals) of *T. platyphyllos* were analysed (Table 5.1). DNA was extracted from leaves using CTAB based procedure (see Chapter 2).

Table 5.1 Countries, locations, population codes and coordinates (latitude and longitude) of *T. cordata* and *T. platyphyllos* used for chloroplast analysis

Species	Country	Location	Label	Latitude (N)	Longitude (E)
T. cordata	Norway	Sogn og Fjordane, Asane nature reserve	NOSO	61.85753	6.19563
	Finland	Niinisaari	FINI	61.81667	29.38333
	Denmark	Åbybjerg	DEAB	57.10241	9.35310
	Germany	Colbitz	GECO	52.33027	11.55722
	United Kingdom	Anston Stones Wood	UKANS	53.34426	-1.20715
	the Netherlands	Gulpen	NEGU	50.79595	5.89453
	Poland	Lezajsk	PLLE	50.26360	22.38580
	Czech Republic	Velky Osek	CZVO	50.10143	15.17778
	Slovakia	Blatnica, Gaderská dolina	SKBL	48.94860	18.96550
	France	Foret Dom de Mouthiers	FRMO	48.91388	4.91441
	Ukraine	Rosilna	UARO	48.76970	24.39570
	Austria	Stams	AUST	47.27566	10.97716
	Switzerland	Lobsigen Park	CHLO	47.30774	8.26685
	Hungary	Oltárc	HUOL	47.00000	19.00077
	Italy	Lasen, Dolomites	ITLA	46.07000	11.92000
	Spain	Huesca	SPHU	42.58778	0.18889
T. platyphyllos	Denmark	Bolderslev	DEBO	55.01680	9.38901
	United Kingdom	Kings' Wood	UKKI	53.39842	-1.18188
	Germany	Lichtenstein	GELC	51.30277	13.01875
	Czech Republic	Šternberk, Vrapač	CZST	49.72490	17.01900
	France	Cessieres	FRCE	49.55877	3.48875
	Slovakia	Blatnica, Gaderská dolina	SKBL	48.94860	18.96550
	Austria	Thayatal Park near Hardegg	AUTH	48.84747	15.88003
	Switzerland	Frieswil	CHFR	47.07213	7.14241
	Spain	Huesca	SPHU	42.58778	0.18889
	Greece	Aggistro	GRAG	41.35664	23.48132

5.3.3 Chloroplast sequencing

Eight chloroplast regions were amplified. Primer sequences are listed in Table 5.2. The PCR amplifications were carried out in 15 µl reactions containing 5 ng of template DNA, 1x reaction buffer, 2mM $MgCl_2$, 0.2 mM of each dNTP, 0.4 mM of each primer and 0.5 U of *Taq* DNA polymerase (Bioline). The PCR reactions were amplified as follows: an initial pre-denaturation step at 94 °C for 3 min, followed by 30-35 cycles of denaturation step at 94 °C for 15s, annealing temperature and time are as described in Table 5.2, and extension step at 72 °C for 60s, and a final extension step of 72 °C for 4 min. PCR products were purified with Exo-SAP and sequenced using the same reaction as described in Chapter 2 before visualisation on an ABI 3100 automated sequencer (Applied Biosystem). The level of chloroplast variation is typically low, thus the samples were only sequenced in one direction.

Table 5.2 Primer sequences and PCR details of eight chloroplast regions

Primer name	Sequence	length (bp)	Annealing Temp (°C)	Annealing time (second)	Number of PCR cycles	Reference
atpI	TATTTACAAGYGGTATTCAAGCT	1157	53	30	35	Shaw *et al* (2007)
atpH	CCAAYCCAGCAGCAATAA C					
ndhF	GAAAGGTATKATCCAYGMATATT	768	53	30	35	Shaw *et al* (2007)
*rpl32*R	CCAATATCCCTTYYTTTTCCAA					
petL	AGTAGAAAACCGAAATAACTAGTTA	1161	50-65	ramp 0.3 °C/s	30	Shaw *et al* (2007)
psbE	TATCGAATACTGGTAATAATATCAGC					
psbA	CGAAGCTCCATCTACAAATGG	495	53	30	35	Hamilton (1999)
trnH	ACTGCCTTGATCCACTTGGC					
psbJ	ATAGGTACTGTARCYGGTATT	1269	53	30	35	Shaw *et al* (2007)
petA	AACARTTYGARAAGGTTCAATT					
*rpl32*F	CAGTTCCAAAAAAACGTACTTC	649	53	30	35	Shaw *et al* (2007)
trnL	CTGCTTCCTAAGAGCAGCGT					
trnG	GAACGAATCACACTTTTACCAC	844	53	30	35	Hamilton (1999)
trnS	GCCGCTTTAGTCCACTCAGC					
trnQ	GCGTGGCCAAGYGGTAAGGC	1204	50-65	ramp 0.3 °C/s	30	Shaw *et al* (2007)
*rps16*x1	GTTGCTTTYTACCACATCGTTT					

5.3.4 Microsatellite data analysis

In order to understand whether genetic diversity is correlated with geographical gradients, the correlation between expected heterozygosity (H_e) of each population and geographic-coordinates (latitude) was analysed for each species using Minitab16 software. The genetic distance matrices of individuals within each species were constructed to carry out a Principal Coordinates Analysis (PCO) using GenAlEx version 6.5 (Peakall and Smouse, 2012). In addition, genetic differentiations between populations (Pairwise F_{st}) of each species were determined by GenAlEx. To test the patterns of Isolation by Distance (IBD) of the populations in each species, geographic distance matrices were constructed by calculating the pairwise distances between each population in kilometres using the Mantel test in GenAlEx. Both pairwise F_{st} and pairwise genetic distance matrices were uploaded onto Genepop on the Web version 4.2 (Raymond and Rousset, 1995), option 6. The pairwise F_{st} were converted to $F_{st}/(1-F_{st})$, and the correlation with the log transformed (ln) geographic distance was calculated (Mantel test).

The patterns of population structure in both *T. cordata* and *T. platyphyllos* were analyzed using Bayesian clustering software, STRUCTURE Version 2.3.3 (Pritchard *et al*, 2000). A number of clusters (K) are estimated and individuals will be assigned to one or more clusters. The number of clusters was set from 1 to 26 for *T. cordata*

analysis and from 1 to 17 for *T. platyphyllos* analysis (total number of populations plus two). The program was run using the same parameters as described in Chapter 4. The results from STRUCTURE were then downloaded onto software STRUCTURE HARVESTER (Earl and vonHoldt, 2012) to estimate the optimal grouping (K). To visualize population clusters on a geographical scale, PhyloGeoViz (Tsai, 2011) was used to convert the proportion assignment to each cluster to pie charts and overlaid on the map.

5.3.5 Chloroplast data analysis

Chloroplast sequences were double-checked using the chromatograms, edited and aligned using Geneious version 6.1.4 (Drummond *et al*, 2011). All eight chloroplast regions of each sample were concatenated and regarded as a single locus using Mesquite Version 2.75 (Maddison and Maddison, 2001). A chloroplast network was constructed based on statistical parsimony using median-joining network using Network version 4.6.1.1 (Bandelt *et al*, 1999). Insertions and deletions (indels) were manually edited as a single base substitution and set as double weight for the analysis. The distribution maps were constructed based on the frequencies of the haplotypes using PhyloGeoViz (Tsai, 2011).

5.4 Results

5.4.1 Geographical patterns of genetic diversity

Geographical patterns of genetic diversity in *T. cordata* and *T. platyphyllos* based on microsatellite markers showed a negative correlation between expected heterozygosity (H_e) and degrees north (latitude) (Figure 5.2 a and b). A significant decline ($P<0.05$) in genetic diversity toward the north was detected in *T. cordata* ($R^2 = 0.28$), while no significant correlation ($R^2 = 0.01$, $P>0.05$) was detected in *T. platyphyllos*. It must be noted that *T. platyphyllos* has a smaller range of latitudes (41.35664-55.01680 ^{o}N) compared to *T. cordata* (46.07000-61.85753 ^{o}N). (As the distribution range of *T. platyphyllos* is limited in northern Europe, the highest latitude of population we can get is from Denmark.)

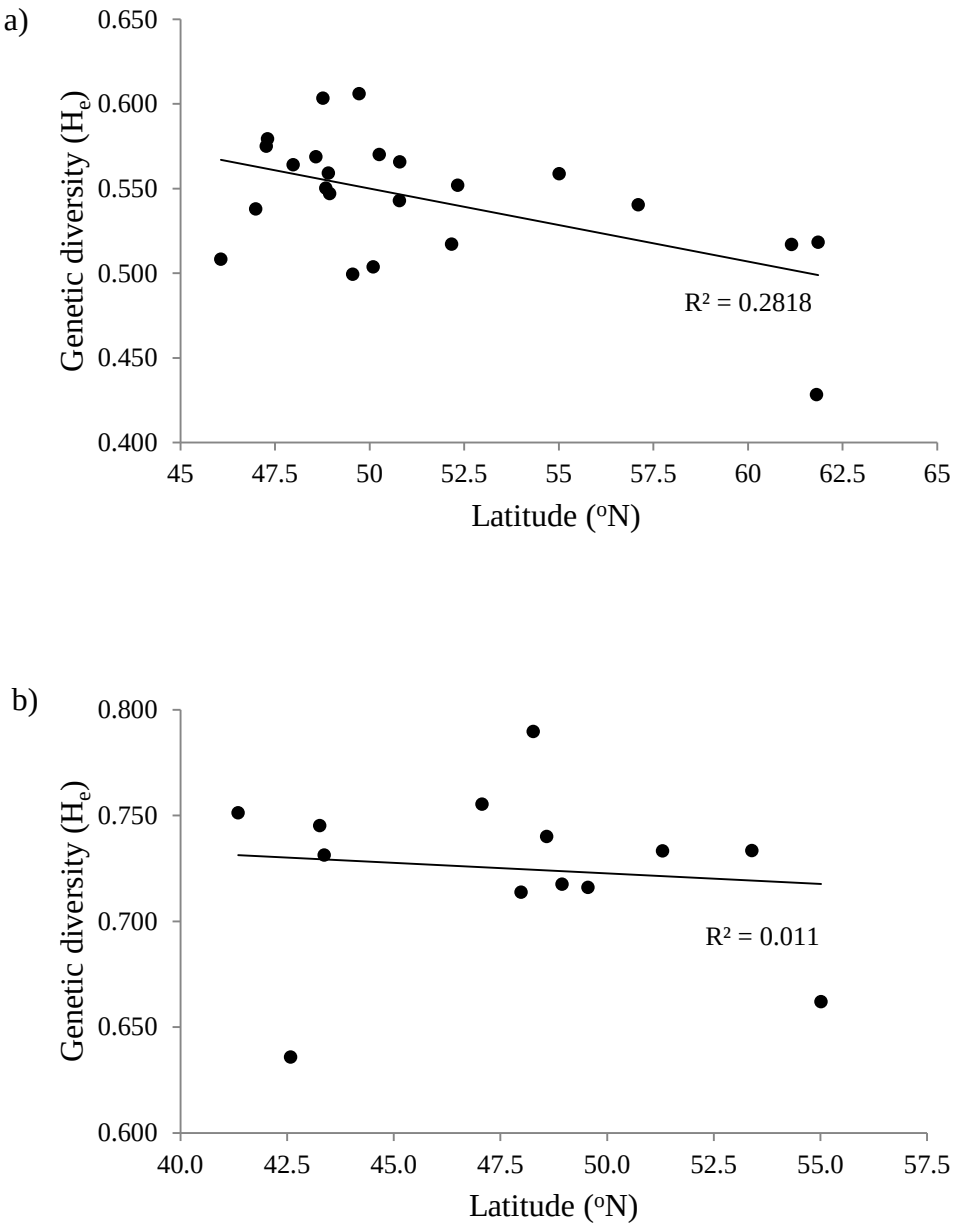


Figure 5.2 Correlation between the genetic diversity (H_e) and latitude (^oN) in *T. cordata* populations (a) and in *T. platyphyllos* populations (b)

5.4.2 Isolation by Distance (IBD)

The correlation between genetic and geographic distance (Mantel test) indicated significant patterns of isolation by distance in both *T. cordata* and *T. platyphyllos* (R^2 = 0.22 and 0.12, respectively, $P<0.05$) (Figure 5.3 a and b). The positive correlations reveal that populations are more similar when they are geographically close than those at longer distances. The matrices of pairwise F_{st} and geographic distance of *T. cordata*

are shown in appendix 7 and 9, respectively and those of *T. platyphyllos* are shown in appendix 8 and 10, respectively.

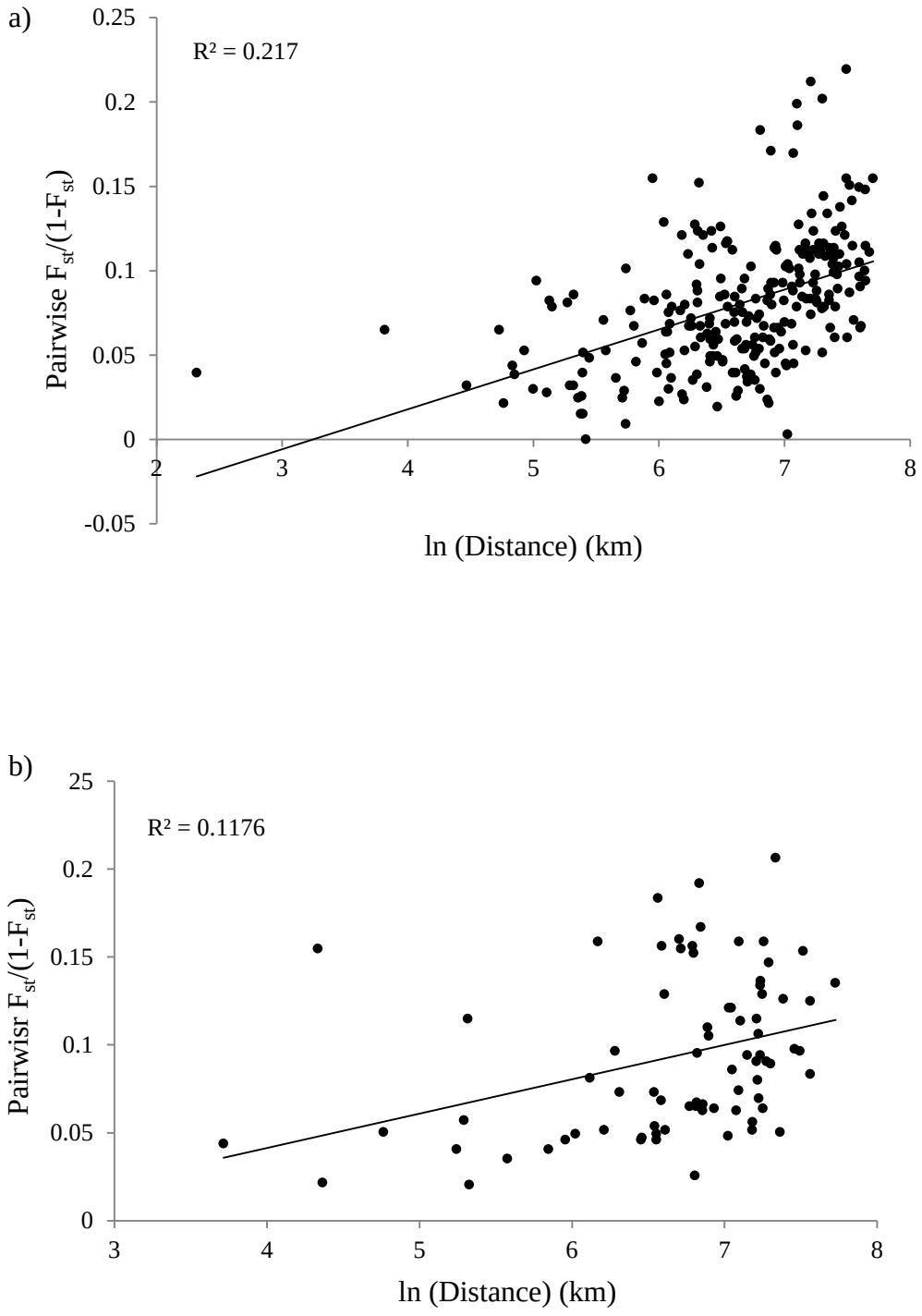


Figure 5.3 Correlation between genetic distance (based on converted pairwise F_{st}) and log (ln) geographic distance (based on pairwise distance in km) of *T. cordata* populations (a) and of *T. platyphyllos* populations (b).

5.4.3 Patterns of population genetic structure based on microsatellite analysis: *T. cordata*

The population structures of 24 populations (421 individuals) of *T. cordata* were analysed using PCO and STRUCTURE. The PCO approach based on microsatellite markers did not show clear grouping of individuals by populations. Figure 5.4 shows a scatter plot of genetic distance within *T. cordata* populations with 41.91% of total variation (22.82% and 19.09% for the first two axes). Although the PCO illustrated low genetic differentiation between populations, some groupings were observed. For example, FINI population (light green dots) was presented only on the left in the graph.

The model-based clustering method implemented in STRUCTURE, which assesses the most likely number of clusters based on allele frequency of microsatellite markers, cannot determine the optimal grouping in *T. cordata* individuals. Although the highest ΔK was detected at K = 2, the natural logarithm of estimated probability of the data increased with K value. Thus, it is difficult to determine the best value of K (Figure 5.5 a and b). Therefore, the structure analyses at different K values were inspected (K = 2, 3, 4 and 6). K = 5 is not presented due to only minor changes from K= 4. K = 2 did not reveal a clear structure of the populations in central Europe, while populations from the north (FINI, FIMU in Finland) were largely clustered together (Figure 5.6 and 5.7 a, at K = 2). K = 3 and 6 also did not give clear information based on geographical pattern, whereas K = 4 reveals the best clustering based on geographical regions (Figure 5.6 and 5.7 b, at K = 4).

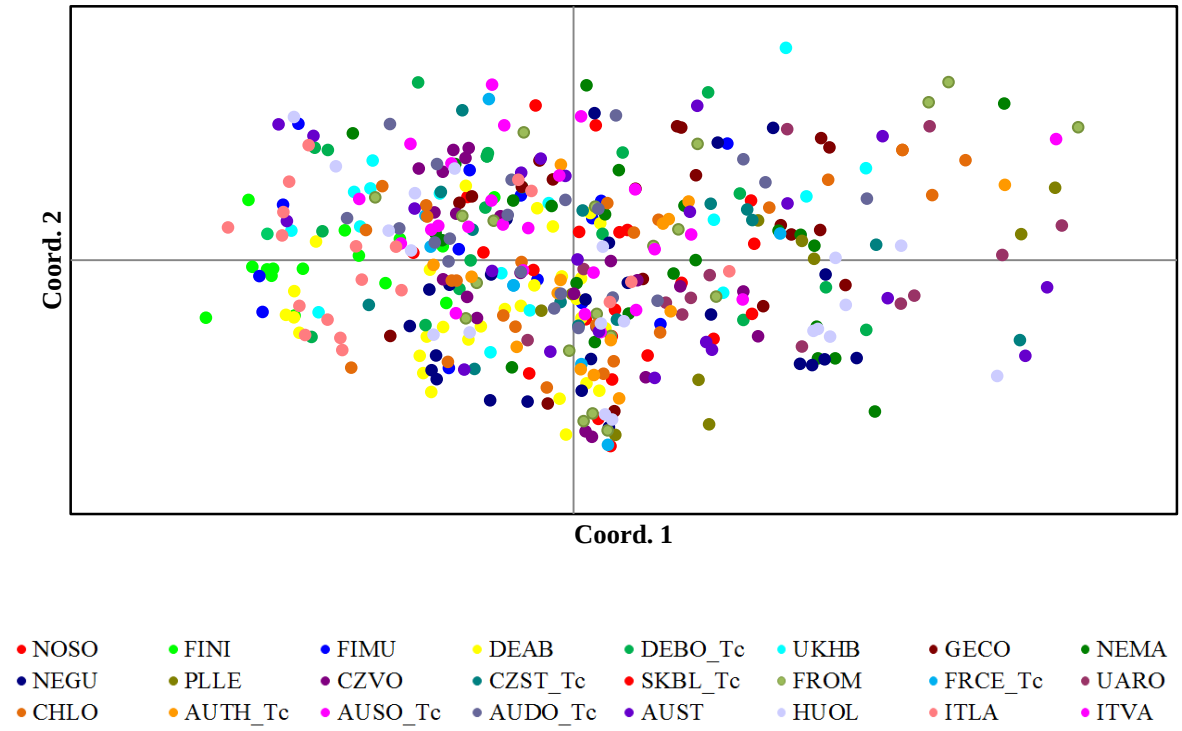


Figure 5.4 Principal Coordinates Analysis (PCO) of genetic distance using 13 microsatellite markers (GenAlEx) of 421 individuals in *T. cordata* populations, coding the populations by colour.

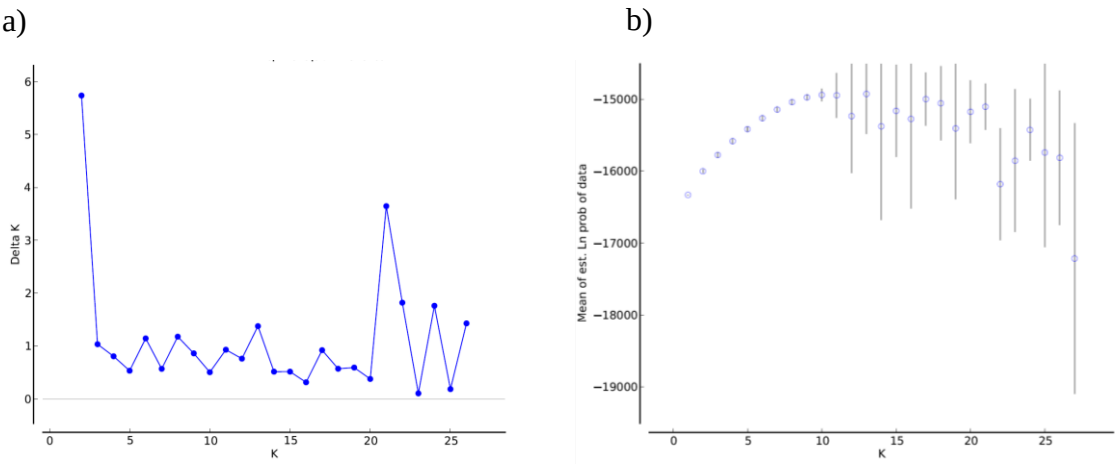


Figure 5.5 Value of ΔK against K of *T. cordata* individuals (a) and means of the estimated natural logarithm probability of the data against K (b).

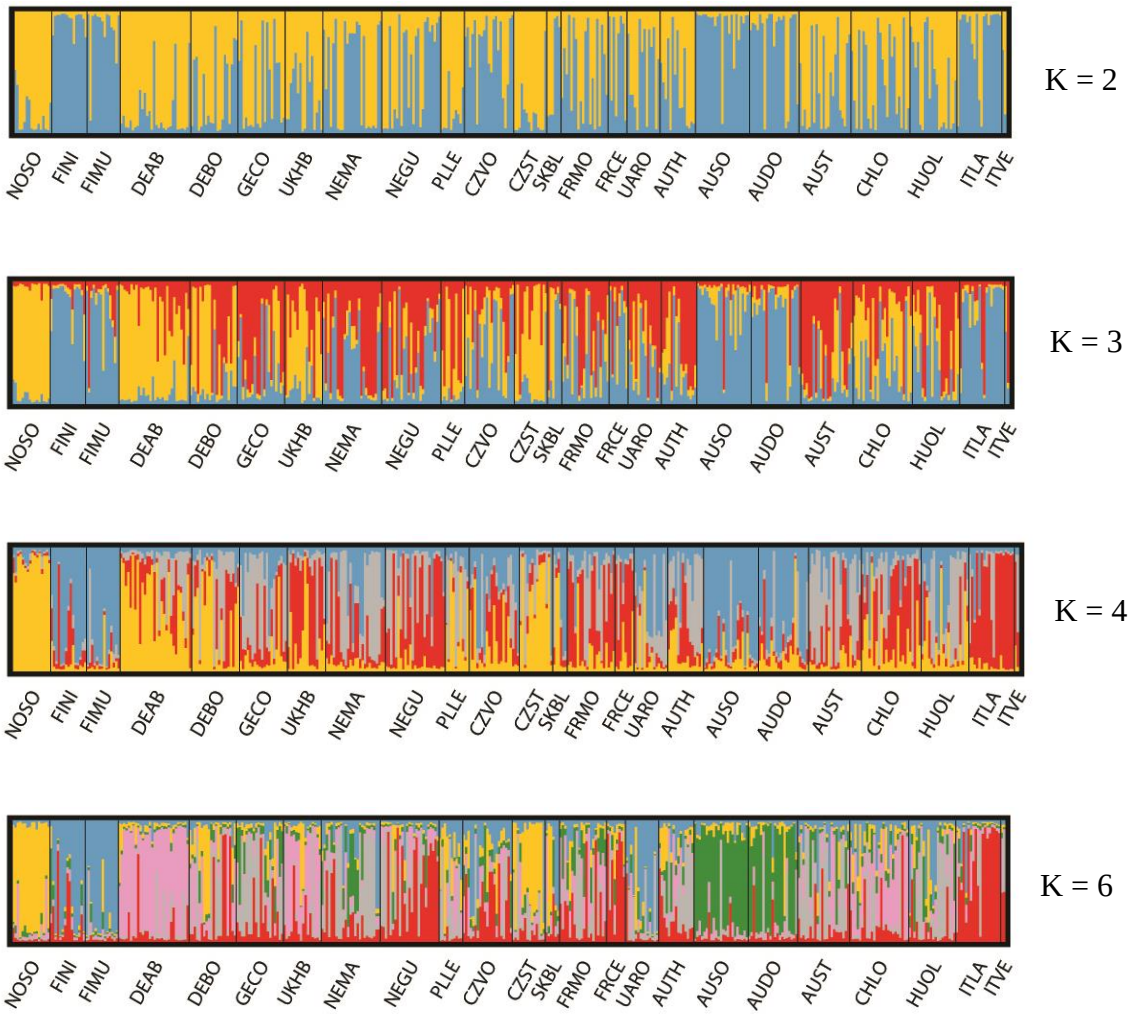


Figure 5.6 Genetic structure of 421 *T. cordata* individuals (24 populations) based on microsatellite markers according to the Bayesian analysis using the program STRUCTURE at different K. Each vertical bar represents individuals. Colours indicate population assignment.

a)

b)

Figure 5.7 Geographical distribution of 24 populations of *T. cordata* based on microsatellite markers at K = 2 (a) and K = 4 (b). Colours on pies indicate frequency of population assignment in each location.

5.4.4 Patterns of population genetic structure based on microsatellite analysis: *T. platyphyllos*

The microsatellites within 15 populations (218 samples) of *T. platyphyllos* were visualised by the PCO. The results show that a total of 42.01% of the variation is accounted by the first two axes (22.27% and 19.75%, Figure 5.8). Some individuals were clustered by populations. For example, individuals from the AUDO population are very near each other and plotted in the upper left, individuals from UKKI are on the lower left, while those from DEBO are on the lower right of the graph. In addition, the grouping based on geographical regions was also observed. For example, individuals from Austria (4 populations: AULE, AUTH, AUSO and AUDO) were largely plotted in the upper left on the graph, while individuals from southern France (FRIS and FRCC) and northern Spain were largely plotted on the right.

In STRUCTURE analysis of *T. platyphyllos*, the highest ΔK showed that K = 4 is the best clustering (Figure 5.9 and 5.10). Population assignment visualized on a geographical map separates the populations by geographical regions. Individuals from eastern populations (7 populations) were assigned largely to the same cluster (Figure 5.11). Individuals from southern France and northern Spain were largely clustered separately from those in northern France and central Europe, while individuals from the UK were largely assigned to a separate cluster. K = 9, which had the second highest ΔK, did not depict a clear distribution pattern. Therefore, only genetic structure at K = 4 is shown.

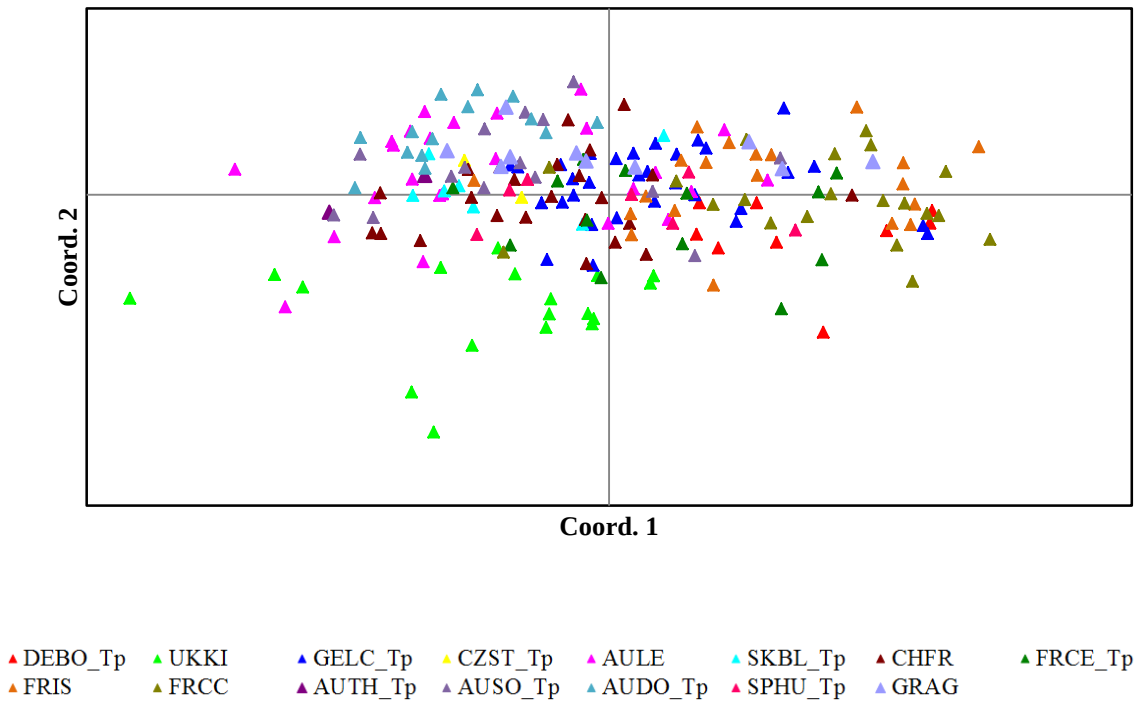


Figure 5.8 Principal Coordinates Analysis (PCO) of genetic distance using 13 microsatellite markers (GenAlEx) of 218 individuals in *T. platyphyllos* populations, coding the populations by colour.

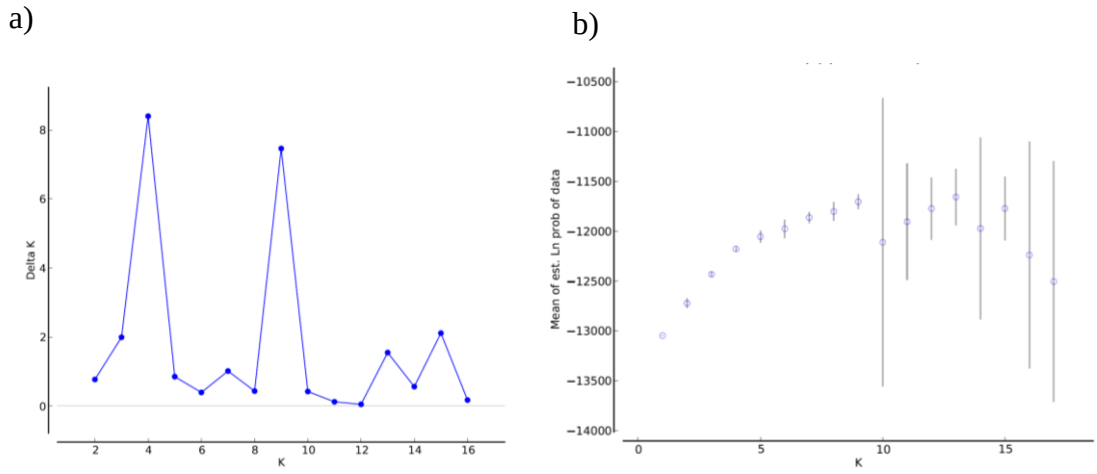


Figure 5.9 Value of ΔK against K of *T. platyphyllos* individuals (a) and Means of the estimate natural logarithm probability of the data against K (b).

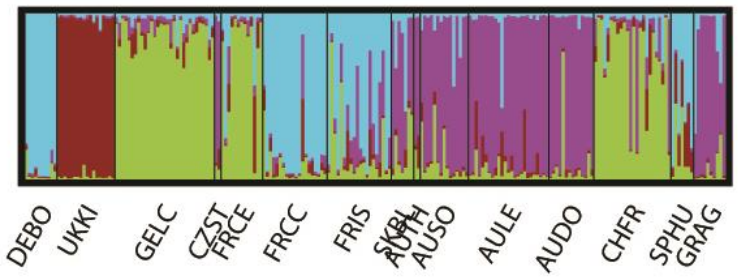


Figure 5.10 Genetic structure of *T. platyphyllos* individuals based on microsatellite markers according to the Bayesian analysis using the program STRUCTURE at K = 4. Each vertical bar represents individuals. Colours indicate population assignment.

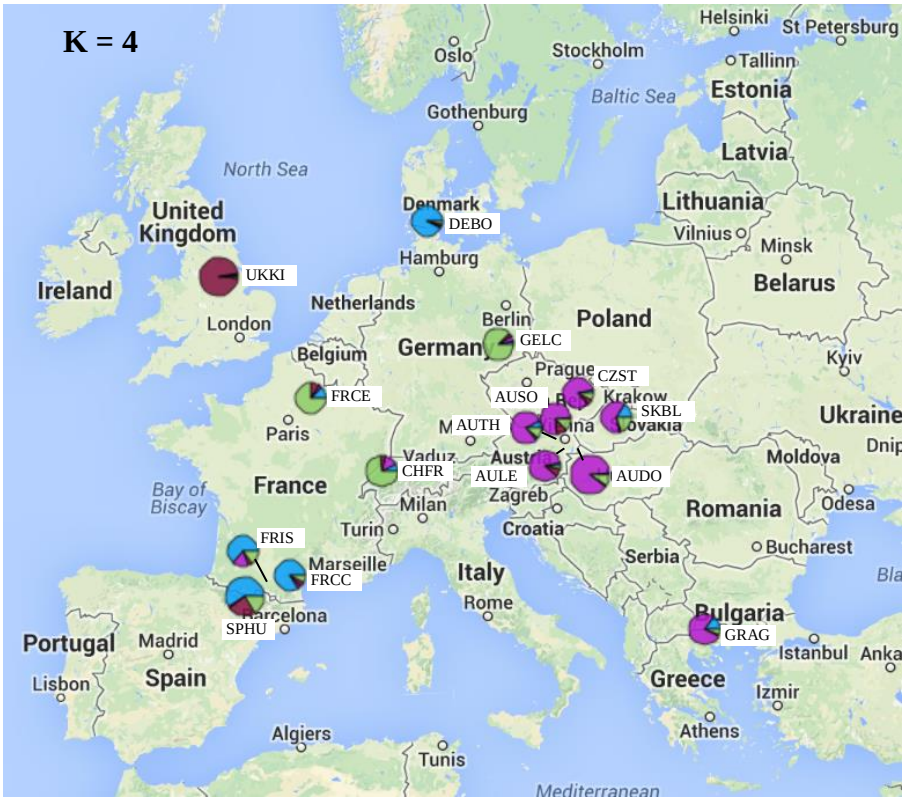


Figure 5.11 Geographical distribution of 15 populations of *T. platyphyllos* based on microsatellite markers at K = 4. Colours on pies indicate frequency of population assignment in each location.

5.4.5 Chloroplast variation and geographic distribution

The chloroplast variation of eight regions in both *T. cordata* and *T. platyphyllos* was limited (Table 5.3). The total length of the concatenated eight chloroplast regions was 2179 bp. Nine substitutions and eleven indels (from five bp in the *psbJ* to 72 bp in the *trnQ*) were detected. A total of eleven haplotypes were identified. Among these haplotypes, four haplotypes (Hap2, Hap4, Hap5 and Hap7) were found only in *T. cordata*, while five haplotypes (Hap6, Hap8, Hap9, Hap10 and Hap11) were specific to *T. platyphyllos*. Two haplotypes (Hap1 and Hap3) were shared by the two *Tilia* species (Table 5.4).

The network of chloroplast haplotypes analysed revealed close relationships of some haplotypes of the two *Tilia* species (Figure 5.12). Hap1 is the most frequent haplotype, found in 27.45% of the individuals in both species (32.26% in *T. cordata* and 20% in *T. plartyphyllos* individuals). Hap10, which is specific to *T. platyphyllos*, has the largest number of differences from other haplotypes, with seven sites different from *T. cordata* Hap7. Two haplotypes of *T. platyphyllos* (Hap8 and Hap9), which were found in Greek and German populations, respectively, linked by just one step mutation to Hap7, which is found only in a *T. cordata* population from Finland. Also, *T. cordata* Hap5 was closely linked to *T. platyphyllos* Hap6 with three different mutations.

There was some geographic structure to the distribution of haplotypes in *T. cordata* (Figure 5.13). Hap1 exhibited the most widespread distribution across mainland Europe with some structure to the distribution from Northern Spain to Denmark and Norway. Hap2, which is linked to Hap1 by one base substitution, had a scattered distribution in central Europe. Hap5 showed a mosaic distribution, being found in two populations in the east (Czech Republic and Hungary) and also in the geographically remote UK population, whereas Hap7 occurred only in one relatively distant population, Finland. Within *T. platyphyllos*, a little stronger structure of haplotype distribution was observed (Figure 5.14). Hap10, which is different from other haplotypes, was distributed in the western populations (Spain, France and UK), while some haplotypes were restricted to only one population of *T. platyphyllos*, such as Hap6, Hap8 and Hap11.

Table 5.3 Polymorphic sites in the eight regions of chloroplast DNA in *T. cordata* and *T. platyphyllos*. Indels are indicated with a number followed by letter (see Notes)

Cp regions	*atp*I			*ndh*F		*pet*L	*psb*A				*psb*J				*rpl*32F		*trn*G		*trn*Q	
Positions	313	319-336	426	101	179	295	215-223	328-337	357-363	364-373	385	400-404	503	614-646	343-359	421	268-290	511	272-288	392-463
Polymorphic site	1	2a	3	4	5	6	7b	8c	9d	10e	11	12f	13	14g	15h	16	17i	18	19j	20k
Hap1	A	-	C	G	C	G	-	-	-	-	A	+	A	+	-	T	-	A	-	-
Hap2	A	-	C	G	C	G	-	-	-	-	C	+	A	+	-	T	-	A	-	-
Hap3	A	-	C	G	C	G	-	-	-	-	A	+	A	+	-	G	-	A	-	-
Hap4	A	-	C	G	C	G	-	-	-	-	A	-	A	+	-	G	-	A	-	-
Hap5	A	-	C	G	T	A	-	-	+	-	A	+	A	-	-	G	-	G	-	+
Hap6	A	-	C	G	T	A	-	-	-	-	A	+	A	+	-	G	-	G	-	-
Hap7	A	-	C	T	C	G	-	-	-	-	A	+	C	+	-	G	-	G	-	-
Hap8	A	-	C	T	C	G	+	-	-	-	A	+	C	+	-	G	-	G	-	-
Hap9	A	+	C	T	C	G	-	-	-	-	A	+	C	+	-	G	-	G	-	-
Hap10	A	-	T	T	C	G	-	+	+	+	A	-	C	+	-	G	+	G	+	-
Hap11	C	-	C	G	T	A	-	-	-	-	A	+	A	+	+	G	-	G	-	-

Notes: Eleven indel sites found in all chloroplast haplotypes are as follows:

a: TTTTTATTATTTATATTT
b: TTATTACCA
c: TTTATATTAT
d: TATAAATAA
e: TATTATAA
f: CTTTT
g: TTATACTTATTATATTATCCTTATTACTTATT
h: AATTAATAATAAATAAT
i: TTATTATTTATTCTAGTTATTCT
j: TCTATCGAGTTATTTGA
k: GAAAAGTTTAATTTTAATATTTAATTTTAATAAATAGAAAAGTATATTTATAATATATAATTTTAATAAATA

Table 5.4 Number of chloroplast haplotypes found in individuals of *T. cordata* and *T. platyphyllos*.

Haplotypes	Number of individuals found	
	T. cordata	*T. platyphyllos*
Hap1	10	4
Hap2	8	0
Hap3	4	3
Hap4	1	0
Hap5	6	0
Hap6	0	2
Hap7	2	0
Hap8	0	2
Hap9	0	1
Hap10	0	6
Hap11	0	2

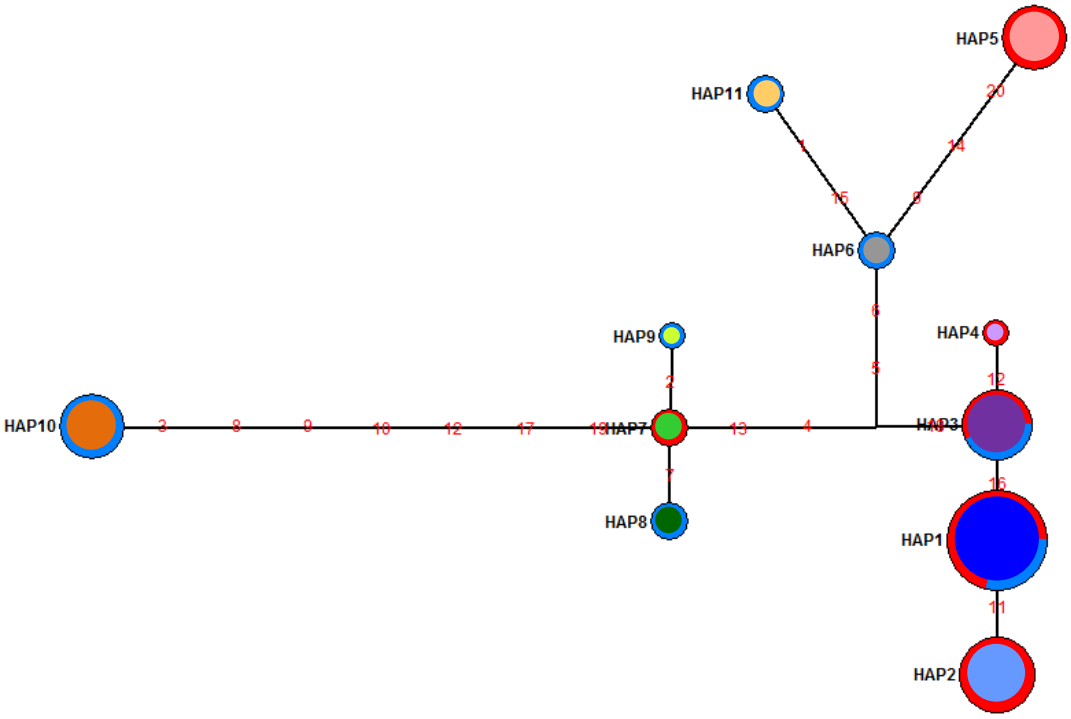
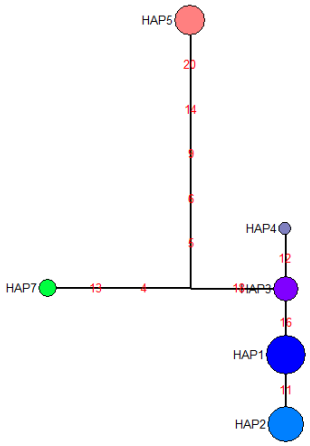


Figure 5.12 Haplotype network of chloroplast haplotypes in *T. cordata* and *T. platyphyllos* using statistical parsimony. Sizes of circles are proportional to the haplotype frequency. Colours of the rim indicate the *Tilia* species found with that haplotype (red = *T. cordata* and blue = *T. platyphyllos*). Numbers on the line indicate the polymorphic sites (Table 5.3).

(a)



(b)

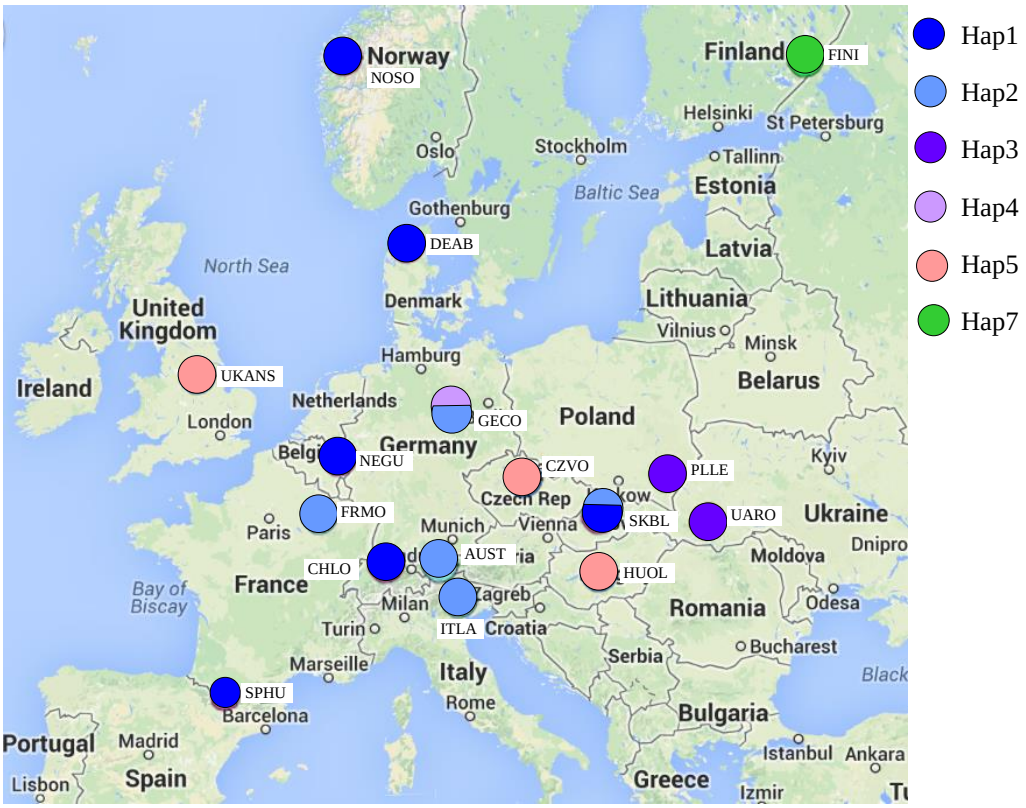
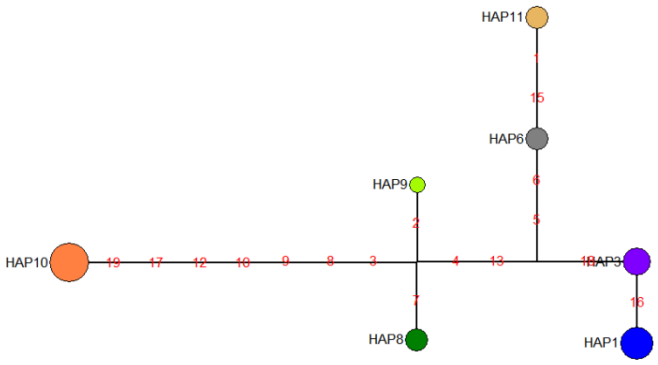


Figure 5.13 Haplotype network (a) and distribution of chloroplast haplotypes of *T. cordata* populations (b).

(a)



(b)

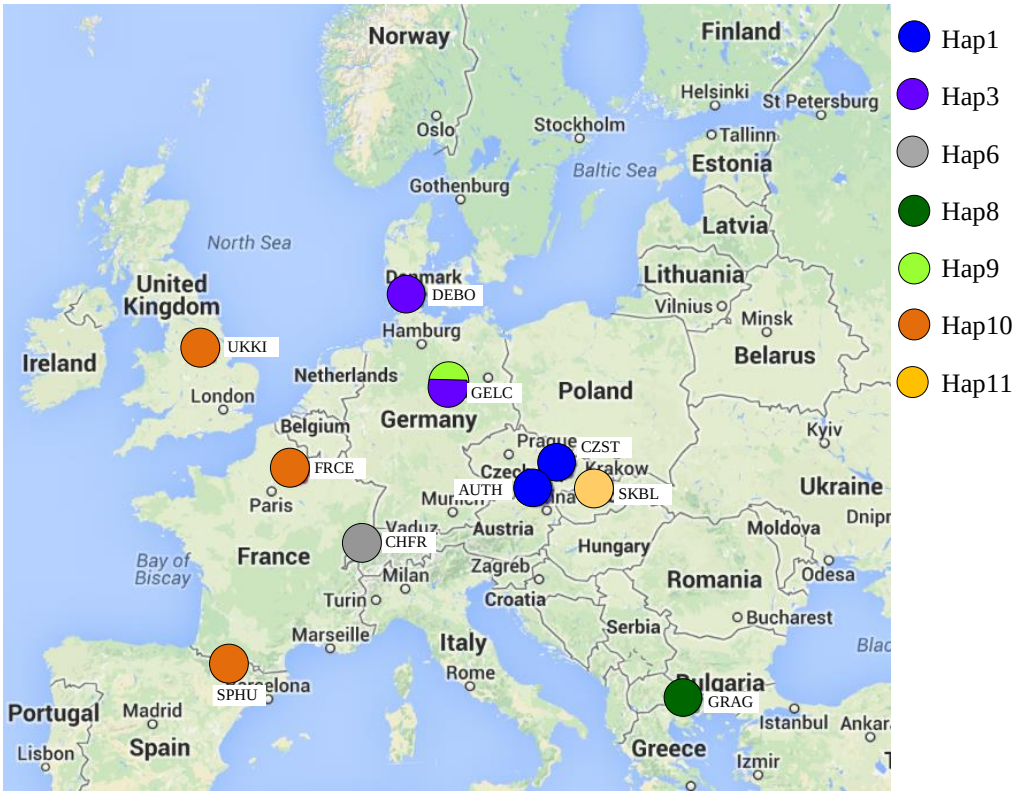


Figure 5.14 Haplotype network (a) and distribution of chloroplast haplotypes of *T. platyphyllos* populations (b).

5.5 Discussion

5.5.1 Genetic diversity and associated geographical patterns

The genetic diversity in term of H_e decreased as latitude increased (Figure 5.2), with a significant correlation in *T. cordata*. The higher genetic diversity observed in southern populations than in the north supports the hypothesis of migration of plants from the

southern refugia after the postglacial period. The high genetic diversity in southern populations could be due to the favourable climate and stable environments in southern refugia during glacial periods. These conditions have provided suitable habitats for living species, including plants, to survive, whereas other parts in northern Europe were largely covered with ice sheets. Thus, genetic diversity tends to be more accumulated in southern populations than in the northern populations. The decline of genetic diversity toward the north observed in both *Tilia* species could be explained by the leading edge model (Hewitt, 1999). This model implies that during warm interglacial periods, populations from the southern refugia expanded into new ice-free territories toward north. This expansion occurred rapidly to colonize and dominate the new habitats before others arrived. The phenomenon would occur several times during setting up new colonisation. As a consequence, the leading edge populations would experience repeated bottlenecks or genetic drift, which could lead to the loss of genetic diversity (Hewitt, 1999). In addition, edge populations can have less gene flow compared to the central population because they can only receive migration from one direction (GarciaRamos and Kirkpatrick, 1997).

The genetic diversity, which is usually higher in southern populations and tends to reduce along latitudinal gradients, has been reported in several plant species. Similar patterns were detected in white oak *Quercus* spp. (Dumolin-Lapegue *et al*, 1997b), in black alder *Alnus glutinosa* (King and Ferris, 2000), in heather *Calluna vulgaris* (Rendell and Ennos, 2002) and in hornbeam *Carpinus betulus* (Grivet and Petit, 2003).

5.5.2 Population genetic structure of *T. cordata* and *T. platyphyllos*

The population genetic structure analysis using a Bayesian method based on microsatellite data revealed the geographical patterns in both *T. cordata* and *T. platyphyllos*. Although Principal Coordinates Analysis (PCO) detected no structure within each species, some grouping of populations was observed. Some structure based on geographical regions in both species was detected by STRUCTURE analysis. Even though the precise K value was difficult to infer from our dataset, the overall patterns can tentatively be interpreted.

The patterns of genetic structure illustrated by STRUCTURE analysis in both species of *Tilia* revealed high genetic differentiation of the range edge populations, which are

more recently colonized areas. In *T. cordata*, populations in the north (Norway and Finland), which is the highest distribution range in degrees north of *T. cordata*, almost formed clusters of their own. Isolated clusters were also observed in the *T. platyphyllos* populations in Denmark and the UK. These results could be because during the rapid northward and westward expansion from southern refugia after the post-glacial periods, *Tilia* may have experienced genetic drift or bottleneck during the long distances of migrations. A loss of alleles seems to have occurred in the range edge populations during colonisation. The number of alleles (N_a) and effective number of alleles (A_e) was lower on average in northern populations than in southern populations in both species (Table 4.4 and 4.5, in Chapter 4). In addition, the largely isolated populations can be due to the barriers to gene flow along the migration routes. The water barriers (North and Baltic Sea) seem to limit the amount of gene flow between populations and lead to population isolation, particularly separation between *T. platyphyllos* in the UK and the mainland populations.

The significant Isolation by Distance (IBD) patterns in both *T. cordata* and *T. platyphyllos* (Figure 5.3) also supported the divergence of the range edge populations due to the longer distance of migration from the south compare to populations in central Europe. Also, these significant trends of IBD may reflect population colonisation from different glacial refugia in southern Europe or even further east from unknown refugia. It is clearly seen in eastern populations of *T. platyphyllos* that are near the Greek population (Balkan refugia), and the results from STRUCTURE also assigned these populations largely in a single cluster.

A stronger genetic structure based on geographical region was observed in *T. platyphyllos* compared to *T. cordata*. This could be because migration and recolonisation toward north after postglacial periods in *T. cordata* occurred before that in *T. platyphyllos*. Thus, *T. cordata* would have more generations in the newly colonized area, which could lead to more opportunities for gene flow between populations in *T. cordata* than in *T. platyphyllos*. This phenomenon may have produced a reduction of genetic differentiation among populations and less clear population structure of *T. cordata*. The fewer generations of *T. platyphyllos* due to the later recolonisation could aid in the maintenance of genetic differentiation among the original sources, namely southern refugia. This result supported the modelling of potential range limit distributions of tree species in Europe during the Last Glacial Maximum (LGM) (Svenning *et al*, 2008). This modelling, predicted by simulation of LGM climates and

tree distribution using the data of 22 current tree species distribution, revealed that *T. cordata* would have distributed relatively far north from the southern refugia compared to *T. platyphyllos*. In addition, the pollen record indicated that the most abundant pollen found during the Pleistocene period was of *T. cordata* and the considerable tolerance to low temperature of *T. cordata* (Pigott, 2012) could also support the possibility of the earlier migration and recolonisation toward the north of *T. cordata* than of *T. platyphyllos*.

5.5.3 Post-glacial colonisation routes and possible refugia of *T. cordata* and *T. platyphyllos*

The analysis of microsatellite data suggested different possible colonisation routes of *T. cordata* and *T. platyphyllos* from southern putative refugia. At least two putative glacial refugia could exist in *T. cordata*, whereas three putative refugia could be defined in *T. platyphyllos*. Chloroplast haplotypes observed in *T. cordata* suggest different colonisation routes compared to microsatellite analysis, particularly in the range edge populations, such as the British and Norwegian populations. In contrast, chloroplast haplotypes observed in *T. platyphyllos* agree with the potential colonisation route to the UK from the Iberian peninsula defined by microsatellite analysis.

Considering the distribution of *T. cordata* based on population structure analysis from microsatellites, K = 4 reveals the best clustering. This clustering indicates some signals of post-glacial colonisation and migration routes of *T. cordata* from southern to northern Europe (Figure 5.6 and 5.7 b, at K = 4). The pattern of population structure could identify at least two putative refugia in the Balkan and Italian peninsulas. The different clusters found in Finnish and Norwegian populations were present in the eastern European populations. This could imply that the genotypes in the two northern populations could derive from Balkan refugia after the ice age. The distribution pattern observed also suggests the expansion from the southern refugia to fill the north of Europe was divided into two possible different migration routes. The first possible routes could expand from the Balkans via Russia to Finland, while the second route could expand through central Europe to Denmark and Norway. The main cluster present in the UK could suggest the contribution of refugia from the Italian peninsula (Figure 5.15). These expansion routes from the two main refugial areas support the spread of *Tilia* analysed using pollen evidence during Late Glacial and postglacial periods (Lang,

2003). The study revealed that during the last glaciation (18,000 BP), *Tilia* was present only in the Balkan Peninsula and southern Italy. Then *Tilia* spread to the north and northwest of Europe and reached southern Scandinavia and England by 8,000 BP (Figure 5.1). However, the main genotype in Finland and Eastern Europe could have distributed from refugia in the Caucasus and the main genotype in the UK could derive from Iberia. Due to the lack of samples from these areas, the putative Caucasus and Iberia refugia cannot be confirmed.

Considering chloroplast haplotype distribution of *T. cordata*, it is possible that three refugia existed in Iberia, Italy and the Balkan Peninsula. The haplotypes suggest some different colonisation routes compared to microsatellite analysis (Figure 5.16). Hap1, which is the most widespread and predominant haplotype, could suggest the expansion route was from Iberia to Denmark and Norway. However, due to the limited sampling of *T. cordata* from Spain (one individual), it is difficult to determine the reliable colonisation routes and confirm the existence of this refugium in *T. cordata*. If the chloroplast haplotypes and microsatellite analysis are combined, the colonisation route to Norway may have come from two potential refugia (Italy and Iberia). The distribution of Hap2 could suggest the existence of putative glacial refugia in the Italian peninsula, which supports the results of microsatellite distribution. Cross-over of haplotype distribution was observed in Hap5, which indicates the UK haplotype probably came from the Balkan Peninsula or eastern areas during the postglacial expansion (Figure 5.16). This colonisation route is different from the pattern implied by microsatellite analysis, which suggests an origin from refugia in Italy. From the results of both chloroplast haplotypes and microsatellite data, it is possible that the colonisation of *T. cordata* in the UK could originate from two putative refugia, in Italy and the Balkans. However, to confirm the colonisation route from the southern refugia, more samples need to be analysed, particularly in the UK. The mixing of different lineages in the north has also been observed in white oaks (*Quercus* sp.), which indicated that the colonisation in southern Scandinavia could derive from two routes from refugia in Iberia and Italy (Dumolin-Lapegue *et al*, 1997a; Petit *et al*, 2002).

The distribution pattern based on microsatellite data in *T. platyphyllos* indicated three putative refugia in Iberia, Italy and the Balkan Peninsula (Figure 5.17). The main cluster found in six populations in Eastern Europe, which were grouped into the same cluster as the Greek population, could suggest a distribution route from the Balkans during the postglacial period. Despite the lack of *T. platyphyllos* sampling from Italy, the main

cluster observed in central Europe (German, France and Swiss populations) could imply a postglacial colonisation route from Italian refugia. The main clusters in the UK population and the Danish population are both found in the populations from southern France and Spain, which could suggest that they may have shared the potential refugia in Spain (Iberia) and there was colonisation by two different routes toward the north. Although there is no pollen evidence of *Tilia* in Spain and Portugal during the last glaciation (18,000 BP), the coherent results of both microsatellite analysis and chloroplast haplotype (Haplotype 10) (Figure 5.18) in the expansion of *T. platyphyllos* from northern Spain to the UK could support the existence of this potential refugium. Also, the result could be explained by the presence of a land bridge during the post-glacial period, due to the low level of the sea (Rohling *et al*, 1998).

In the chloroplast haplotype analysis, the results indicated that some haplotypes were restricted to single populations. Thus, sequencing more samples in other locations would allow more reliable determination of the patterns of chloroplast distribution. In addition, adding samples from the potential refugial areas would build a complete picture of post-glacial colonisation routes and more accurately define the possible refugia in the two species of *Tilia.*

Figure 5.15 The possible migration routes of *T. cordata* based on microsatellite data.

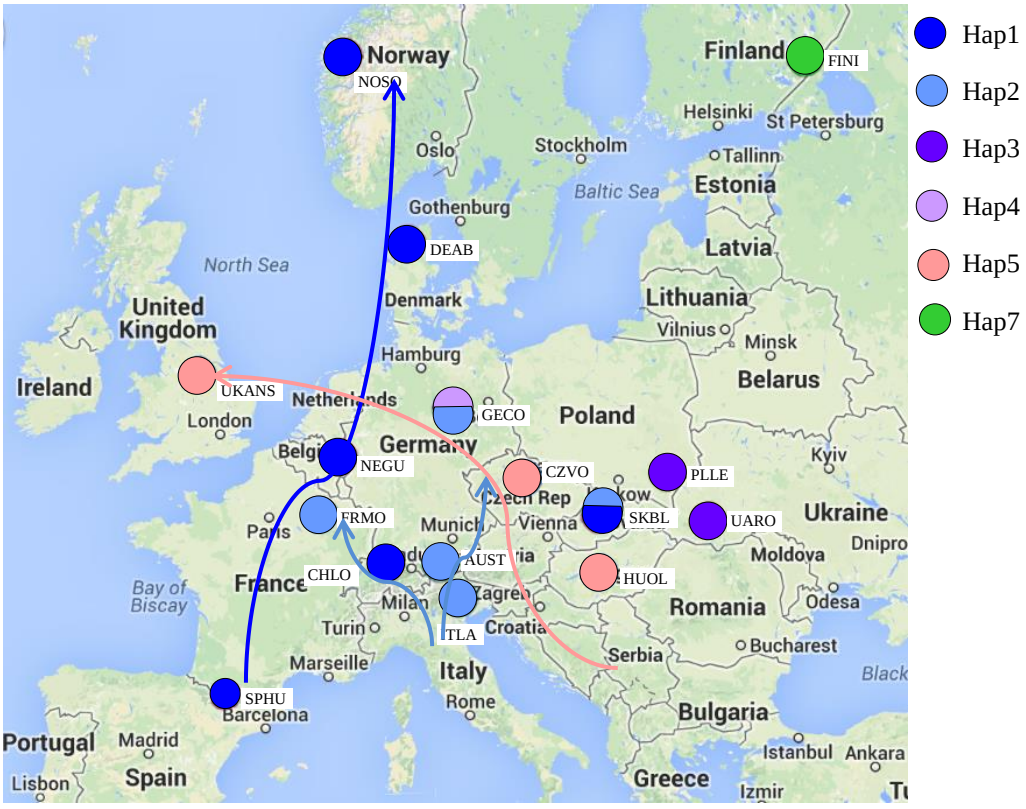


Figure 5.16 The possible migration routes of *T. cordata* based on chloroplast haplotypes.

Figure 5.17 The possible migration routes of *T. platyphyllos* based on microsatellite data.

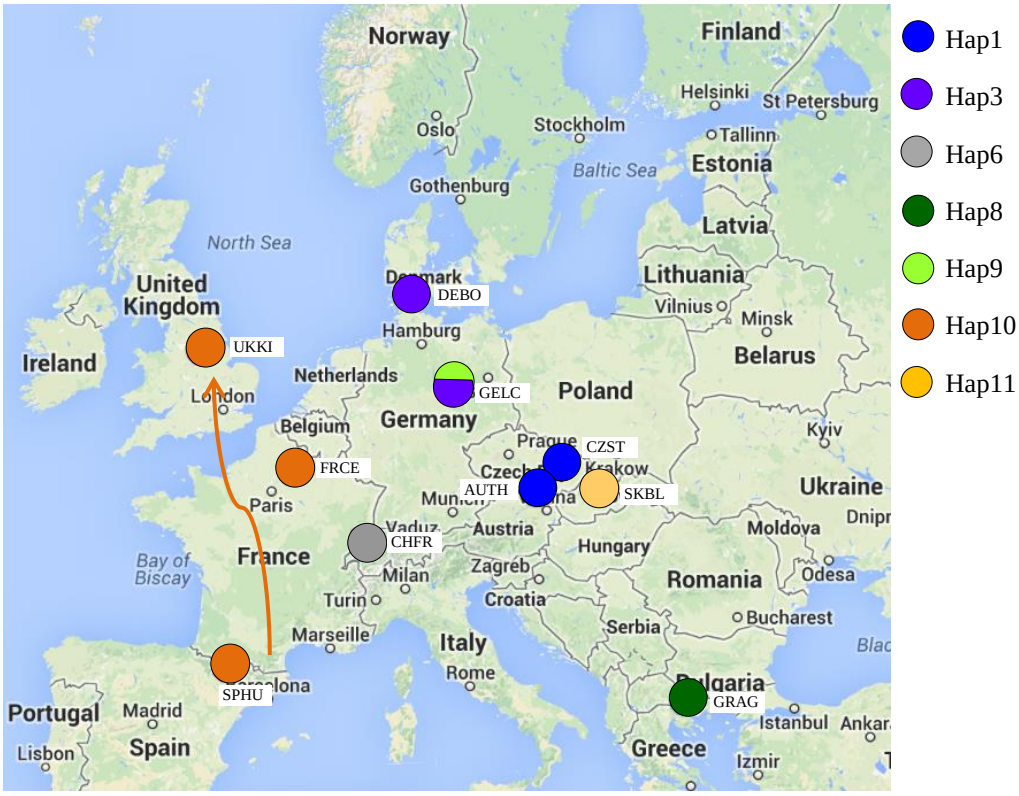


Figure 5.18 The possible migration routes of *T. platyphyllos* based on chloroplast haplotypes.

5.5.4 The sharing of chloroplast haplotypes between *T. cordata* and *T. platyphyllos*

The haplotype network indicated some sharing of haplotypes between *T. cordata* and *T. platyphyllos*. In general, sharing haplotypes in closely related species may be caused by incomplete lineage sorting of ancient polymorphisms and by hybridization. In our study, the haplotype network suggests that incomplete lineage sorting is a more likely cause of sharing haplotypes between *T. cordata* and *T. plattyphyllos* than recent hybridisation.

The long generation time of *Tilia* could be a reason for the incomplete lineage sorting. The age at which *Tilia* trees begin to flower and produce seed ranges from six to 40 years old, depending on the environmental conditions, particularly related to light exposure (Pigott, 2012). In addition, in natural woodlands this may be even later if there are no gaps. The long period of time for producing the next generation may be insufficient for clearly separating the two species, particularly by using DNA regions with a low rate of evolution, such as chloroplast DNA.

The haplotypes Hap1 and Hap3 were shared by both species and are at an intermediate position in the network. This could indicate that these haplotypes are ancient and present in the common ancestor, and now present in *T. cordata* and *T.platyphyllos* after the two species split. This hypothesis is also supported by the fact that the terminal haplotypes in the network are all species specific. In addition, several lines showed a mixture of haplotypes belonging to the two species, which also suggests incomplete sorting. For example, Hap8 and Hap9, which were found only in *T. platyphyllos*, are separated by one mutation step (indel) from the neighbouring haplotype (Hap7) found only in *T. cordata*. Moreover, the shared haplotypes occurred in non-overlapping sampling areas of the two species. Also populations in mixed stands had different haplotypes. *T. cordata* in Spain possessed Hap1, while *T. platyphyllos* had Hap10. This could indicate that recent hybridisation between *T. cordata* and *T. platyphyllos* is infrequent. These results are also supported by the low hybridization and introgression detected in the UK (Chapter 6). However, pollen evidence of the hybrid found during the middle of the Pleistocene (Hoxanian interglacial) could indicate that the hybridisation was more frequent at that ancient time (Pigott, 2012).

Some intermediate haplotypes not observed in this study could indicate a high extinction rate of chloroplast lineages, which may result from high rates of lineage sorting. However, as the sampling numbers and sites were limited, this hypothesis

cannot be tested. Adding samples would complete the chloroplast network and may aid in explanation of the lineage sorting of chloroplast haplotypes in the species.

In plants, sharing of chloroplast haplotypes between closely related species has often been observed, such as in two sympatric beeches (*Facus lucida* and *F. longipetiolatta*) (Zhang *et al*, 2013), in red maple and silver maple (*Acer rubrum* L. and *A. saccharinum* L.) (Saeki *et al*, 2011), in birches (*Betula pendula*, *B. pubescens* and *B. nana*) (Palme *et al*, 2004) and in ash (Fraxinus excelsior and F. angustifilia) (Heuertz *et al*, 2006).

5.6 Conclusion

Our study confirmed the expectation that populations at the range edge have low genetic diversity and are genetically more isolated than the populations near the southern refugia. This supports the rapid recolonisation of tree species during the postglacial periods. In addition, the stronger genetic structure observed in *T. platyphyllos* than in *T. cordata* suggests that *T. cordata* could migrate and colonize the northern areas before *T. platyphyllos*. According to the results of both microsatellite and chloroplast haplotypes, the colonisation routes of *T. cordata* are different from *T. platyphyllos*. However, the two *Tilia* species seem to share the three main refugia in southern Europe (Iberia, Italy and Balkans). *T. cordata* seems to have additional putative refugia in Caucasus (eastern areas).To clarify the colonisation routes and confirm the existence of ice age refugia, more samples, particularly from the putative refugia, need to be analysed. Considering the chloroplast haplotype network, the shared haplotypes of the two *Tilia* species could be the result of incomplete lineage sorting rather than recent hybridization.

Chapter 6. Contrasting molecular and morphological methods for the identification of interspecific gene flow between *Tilia cordata* and *T. platyphyllos* in the UK

6.1 Abstract

Hybridisation is a significant process in plant evolution. The accurate identification of hybrids is a necessary process required for understanding evolutionary consequences of hybridisation. *Tilia cordata* and *T. platyphyllos* are common in the UK and they can form a hybrid (*T.* x *europaea*). There is little genetic information available on natural hybridisation and introgression of these species. In this study, nuclear microsatellite markers and morphological characters were used to determine natural hybridisation and introgression between *T. cordata* and *T. platyphyllos* in ten sympatric UK populations and compare the use of these two methods for detection of the hybrid. The results revealed that hybridisation has occurred between *T. cordata* and *T. platyphyllos*. Microsatellite results indicated clear separation of the two species. The hybrids detected are mainly first generation hybrids, but introgression was also detected. The results from molecular and morphological analysis were partially congruent. The hybrids had overlapping ranges of hybrid index (HI) with the pure parental species. The results suggest that molecular markers are more reliable than HI analysis, based on leaf characters, for detecting hybridization. The hybridization and introgression observed could indicate gene flow between the two species, of importance for adaptation and other evolutionary processes.

6.2 Introduction

Hybridisation is a common phenomenon in plants (Grant, 1981; Harrison, 1993). It occurs when the reproductive isolation (RI) between closely related species is incomplete (Stebbins, 1959). It has been suggested that approximately 25% of plant species hybridise with at least one other species (Mallet, 2005). Hybridisation is a significant process in plant evolution as it produces genetic combinations between different species (Arnold, 1997; Arnold *et al*, 2004; Rieseberg and Carney, 1998). This process can lead to the origin of new species, subspecies, races or ecotypes through

homoploid (parental species and hybrid have the same ploidy level) and polyploid (parental species and hybrid have different ploidy level) hybridisation (Rieseberg and Wendel, 1993).

The consequences of hybridisation and introgression are also a potential source of evolutionary novelty. These processes increase genetic diversity and also provide a possibility of transferring genes involved in plant adaptation (Rieseberg and Wendel, 1993). Anderson (1949) was the first to describe that introgression may have an advantage if genes are exchanged between two species. This could facilitate adaptation and contribute to colonisation of novel habitats as exemplified by the adaptive introgression of *Helianthus paradoxus* (Rieseberg, 1991).

Often hybrids are fully fit or even if they have a reduced fitness compared to their parental species, they are often able to contribute to the next generation and may introgress with their parental species, which results in hybrid swarms. The phenotype and genotype in hybrid swarms can vary considerably depending upon the type of crosses, such as back-crossing or selfing of first generation hybrids (Abbott, 1992). Therefore, the correct identification of hybrids is important and necessary for the understanding of evolutionary processes.

6.2.1 Methods used for studying hybrids in plants

Hybridisation and hybrid detection in plants have been studied in many groups. Both morphological and molecular methods have been commonly used to distinguish species and to detect hybrids (Duminil and Di Michele, 2009). The morphological approach is the traditional way for species and hybrid identification, relying on the fact that hybrids typically show a morphology intermediate between the parental species. Historically hybrids were identified using a hybrid index approach, introduced by Anderson (1949). This approach identifies hybrids using a number of different traits of the two parental species. Traits from one species are scored as zero, while those from the second species are scored as two. Each sample is scored for all morphological characters and the total scores are calculated. While parental types score extreme values, hybrids have a total score intermediate between the two parental species. This has been a successful approach and continues to be used (e.g. Kiaer *et al* (2007), Thorsson *et al* (2007), Tovar-Sanchez and Oyama (2004)).

Variation of morphological characters can be due to both genetic and environmental effects (Rieseberg, 1995). Even different samples of the same genotype may have variation of morphology because of phenotypic plasticity (Garzon *et al*, 2011). In addition, the morphological characters of hybrids, or further generations, may not be intermediate between their parents but closer toward one parental species. Therefore, the detection of hybrids using only morphological traits may lead to detection error. In the last few decades, molecular methods have been widely used for species determination and hybrid detection (Duminil and Di Michele, 2009).

Molecular markers are powerful tools to identify species and to detect hybrids. Study at the DNA level provides an advantage over the morphological approach because genetic material is stable in all plant tissue and does not vary due to the environment or the stage of plant development. Often morphological and molecular approaches are used in combination to identify species and hybrids. For example, hybridisation between species in the genus *Quercus* (the oaks) was studied using various types of genetic markers. By using nine RAPD markers and morphological characteristics of leaves, the hybridisation between two closely related Mexican red oaks, *Quercus affinis* and *Q. laurina,* was analyzed. Sixteen populations including 12 hybrid zones showed partially congruent results between morphological characters and genetic markers (Gonzalez-Rodriguez *et al*, 2004), with morphological analyses indicating a smaller number of intermediate samples than molecular analyses.

Evidence for natural hybridisation and introgression was reported in four naturally co-occurring oak species (*Quercus robur, Q. petrea, Q. pubescens* and *Q. frainetto*) (Curtu *et al*, 2007). In this study, 13 morphological characters (mainly leaf morphology), chloroplast DNA, seven isozymes and six microsatellite markers were used to define species and assess hybridisation among the four oak species (269 samples). Genetic clustering showed a high correlation with morphological analyses. Some incongruence between morphological and genetic discrimination was found. Curtu *et al* (2007) suggested that those with intermediate morphology are not necessarily hybrids. Therefore, hybrid identification based only on morphological characters can lead to wrong conclusions.

In 2010, seven nuclear microsatellite markers and leaf morphology were used to analyse interspecific gene flow in hybrid zones of *Q. hypoleucoides, Q. scytophylla* and *Q.*

sideroxyla in north-western Mexico (Penaloza-Ramirez *et al*, 2010). Hybrids between each pair of species, back-crosses and a probable triple hybrid of three species were detected. Other studies of hybridisation and hybrid detection between oak species have been reported. Some had incongruent or partially congruent results between morphological and genetic characters (Bacilieri *et al*, 1996; Craft and Ashley, 2006; Craft *et al*, 2002; Kremer *et al*, 2002), while some had correspondence between these two approaches (Fortini *et al*, 2009; Gailing *et al*, 2012; Howard *et al*, 1997).

6.2.2 Hybridisation between *T. cordata* and *T. platyphyllos*

Tilia cordata and *T. platyphyllos* are both regarded as native in England and Wales (Pigott, 1969) and the former was dominant in ancient woodland 5000 yr BP (Huntley and Birks, 1983). In the UK, *Tilia* has not been widely planted in woodland and regenerates poorly from seed. Therefore, the presence of *Tilia* indicates ancient woodlands (Pigott, 1969). *T. cordata* is naturally distributed in many parts of the UK, being found naturally as far north as the Lake District (latitude 54^{o} 30'N) and North Yorkshire (Pigott and Huntley, 1978). *T. platyphyllos* mostly occurs at the same latitudes as *T. cordata*, although it does not extend naturally as far north. If these two species are sympatric, as is typically the case, hybridisation may occur as they have overlapping flowering times (Pigott, 2012). The hybrid between *T. cordata* and *T. platyphyllos*, *T.* x *europaea* L. (common lime), shows intermediate morphology between its parental species. Elwes and Henry (1906) considered that *T.* x *europaea* occurred rarely naturally. Little concrete evidence of introgression has been reported. At Harton, Cambridge, seedlings grown from a *T. platyphyllos* tree growing close to *T.* x *europaea* showed a hybrid phenotype (Pigott, 1969). Introgression in *Tilia* species is thought to be limited as hybrids produce few viable seeds (Fromm, 2001). However, this evidence is limited and introgression of *Tilia* species remains to be confirmed. Therefore, the occurrence of natural hybridisation and introgression needs to be clarified.

6.2.3 Detection of *Tilia* species and hybrids using morphological and molecular approaches

T. cordata and *T. platyphyllos* are frequently found together in woods, which could promote hybridisation between them. The identification of the pure species and hybrids using morphological traits is difficult as the hybrids show intermediate traits with

overlapping ranges of characteristics between the species. The identification is even more difficult if only a few traits are used. The main characteristics for species identification are the morphology of leaf, flower and fruit (Pigott, 2012). The two species differ in many traits, such as the shape of leaf buds, the colour of leaves, the presence/ absence of hairs on mature leaves from fully exposed parts of the crowns, the sharpness of the marginal teeth of leaves, the number of flowers in cymes, the number of stamens, the thickness of fruit walls (Pigott, 2012), the position of inflorescences and bracts, the scent of flowers (Pigott, 1991) and the size and shape of pollen grains (Andrew, 1971; Chambers and Godwin, 1971; Wicksell and Christensen, 1999). Although these morphological characters are reliable and clearly separate the two species, it is difficult to obtain some characteristics, such as the number of flowers, inflorescence position and pollen type because flowers are not always present.

In addition, the variability of traits within species is extremely high. Morphological traits are highly affected by environments. Samples differ in leaf shape, size, form of marginal teeth and number of flowers in cymes. Leaf characters are the main feature used in many analyses and these are particularly sensitive to the environment. The amount of light and water are important factors that influence leaf traits, such as size, dimension, margin and structure (Traiser *et al*, 2005). Even in the same tree, leaf size and shape from different parts can be different. The leaves from basal sprouts are normally larger than those from the crown and also have different shapes. In addition, full light exposed leaves have reddish-brown tufted hairs on the lower surface, while leaves from the shade have pale brown or colourless tufted hairs (Pigott, 1991). Therefore, using a morphological approach the location of the leaves on the tree is very important for hybrid detection. Also, hybrids have both intermediate and a mixture of characters of each of the two parental species (Pigott, 1991). In contrast, molecular methods detect species differentiation and hybrids at the DNA level with limited environmental variation.

A small number of studies on hybridisation and species identification in *Tilia* have been published. The characters used in the HI approach are based on the morphology of leaves, petioles and young twigs. In 1969, Pigott used Anderson's method to examine the variation of the two *Tilia* species and the hybrids in Derbyshire Limestone, England. The results reveal that many trees in this location have intermediate morphology between the two parental species, which was proposed to be the result of hybridisation.

However, the occurrence of introgression could not be established, possibly due to the method of detection (morphology).

Wicksell and Christensen (1999) studied the hybridisation between *T. cordata* and *T. platyphyllos* in Denmark, using leaf morphology with seven quantitative and eight qualitative characters (Table 6.1). From the analysis of 111 samples in 12 *Tilia* stands, collected from both natural population and planted avenues, 50 samples were identified as pure *T. cordata*, while 46 and 15 were defined as pure *T. platyphyllos* and hybrids, respectively. Among these 15 hybrids, 10 were from natural populations (Wicksell and Christensen, 1999).

Table 6.1 Morphological characteristics of *T. cordata* and *T. platyphyllos* described by Wicksell and Christensen's method (Wicksell and Christensen, 1999)

Characters	*T. cordata*	*T. platyphyllos*
Length of leaf, including basal lobe	45-106 mm	53-144 mm
Length of lamina, excluding basal lobe	40-91 mm	48-135 mm
Width of lamina	40-82 mm	40-113 mm
Width of apex	2-9 mm	3-15 mm
Length of petiole	25-62 mm	22-62 mm
Number of teeth per cm on the broadest part of leaf	3-7	3-5
Number of lateral veins of first order	4-6.5	6-10
Presence of hairs on upper and lower surface	glabrous	pubescent
Type of hairs	stellate or forked	simple
Colour of hair	reddish-brown	white
Colour of abaxial surface	glaucous	green
Lateral veins of second and third order on abaxial surface	not raised	raised
Presence of hairs on petiole	glabrous	pubescent
Presence of hairs on twig	glabrous	pubescent
Direction of inflorescence	obliquely erect	pendulous

Pollen is also used for identification of *T. cordata, T. platyphyllos* and the putative hybrid. Chambers and Godwin (1971) used scanning electron microscope to investigate whole and fractured pollen grains of *Tilia* taken from the herbarium in Cambridge, UK. The results indicate differences in size and pattern of pollen grains and in structure of pollen walls between the two species and their hybrid (Chambers and Godwin, 1971). Andrew (1971) published his study of pollen in *Tilia* in the same year. 23 samples of

British *Tilia* showed that the average size of pollen was smaller in *T. cordata* than in *T. platyphyllos*. In *T. cordata*, pollen had finer reticulation and a rounder outline than in *T. platyphyllos*. In the hybrid, *T.* x *europaea*, the pollen showed a mixed pattern of the two parental species.

Molecular markers, which are not affected by environmental variation, have also been used for species and hybrid detection. The isozyme method was used for identification of clones of *T. cordata* in a seed orchard and detection of clonal samples along a country road in Germany (Maurer and Tabel, 1995). Fromm and Hattemer (2003) used allozymes to study species differentiation between *T. cordata*, *T. platyphyllos* and their hybrid, using buds and seeds of 140 *T. cordata*, 67 buds of *T. platyphyllos* and seven hybrids in Germany. The study suggested that hybridisation and introgression between *T. cordata* and *T. platyphyllos* is limited as there are several species-specific enzyme variants (Fromm and Hattemer, 2003).

This study aims to utilise and compare molecular markers and a morphological approach to examine the extent of hybridisation between *T. cordata* and *T. platyphyllos* across the range of their sympatric populations in the UK. I have used highly polymorphic nuclear microsatellite markers developed from *T. platyphyllos* (Phuekvilai and Wolff, 2013), which also amplify well in *T. cordata*. This is combined with a Hybrid Index (HI) approach (Pigott, 1969). Individuals from ten sympatric locations in the UK have been investigated to address the question whether hybrids are formed between *T. cordata* and *T. platyphyllos* in natural populations and to evaluate whether morphological and molecular data give consistent results.

6.3 Materials and methods

6.3.1 Plant materials

Leaf and young twig samples from 144 *Tilia* trees were collected from nine locations, considered to hold sympatric populations of *T. cordata* and *T. platyphyllos* based upon SSSI citations (Natural England website). An additional population (Roudsea) was beyond the northern range of *T. platyphyllos* in the UK and was considered to have only *T. cordata* (Table 6.2). Branches fully exposed to sunlight were utilised to provide samples from each tree. The number of trees per location ranged from five to 19, depending on population sizes (mean number of trees sampled per population = 14.4). A

minimum of twelve leaf samples per tree were taken, ten for morphometric analysis and two leaves for DNA extraction. The location of each tree was also recorded using GPS (Garmin eTrex). Among these trees sampled, individuals from two locations were collected by the author (ANS and KIN), the others were collected by colleagues at Edge Hill University.

Table 6.2 Locations, population codes, coordinates (latitude and longitude) and number of trees sampled of the ten *Tilia* study sites

Location	Pop. code	Latitude and longitude	Sample size
Anston Stones Wood	ANS	53.34426, -1.207147	18
Crews Hill Wood	CHW	52.35649, -2.692415	16
Dumbleton Dingle	DUD	52.33138, -2.438264	12
Earl's Hill	EAR	52.63727, -2.869548	5
Highbury Wood	HBW	52.17621, -3.341825	16
Kings' Wood	KIN	53.39842, -1.181878	18
Knapp and Papermill	KPM	52.36835, -2.708353	13
Lady Park Wood	LPW	51.82497, -2.659069	17
Roudsea	ROU	54.23339, -3.025833	19
Tick Wood	TIC	52.62624, -2.526978	10

6.3.2 Molecular analysis

DNA extraction and microsatellite genotyping

144 samples were genotyped using microsatellites. A leaf of each sample was dried and kept at 4 °C. Genomic DNA was extracted using a cetyltrimethylammonium bromide (CTAB) procedure (Morgan-Richards and Wolff, 1999) as described in Chapter 2. The extracted DNA was dissolved in 100 µl of TE buffer and stored at -20 ^{o}C until use. All samples were genotyped using 13 microsatellite loci (Phuekvilai and Wolff, 2013). PCR reactions were performed as described in Chapter 3. GeneMapper® Software (Applied Biosystems) was used to determine fragment sizes and alleles were manually scored and edited before analysing in GenAlEx version 6.5 (Peakall and Smouse, 2012).

Hybrid detection using microsatellite data

Principal Coordinates Analysis (PCO) was performed using GenAlEx6.5 (Peakall and Smouse, 2012). To allocate samples to genealogical classes, Bayesian clustering analysis was performed using the NewHybrids program Version 1.1 (Anderson and Thompson, 2002). By using the default setting, the program was run with 10^5 iterations of the Markov Chain Monte Carlo (MCMC) method and did not use prior allele frequency information. The samples were assigned to one of the six possible genealogical classes (parental species, F_1, F_2 or backcross to each parent), which is visualized with different coloured bars. If multiple samples with the same genotype were detected, only one was included in NewHybrids.

6.3.3 Morphological analysis

A minimum of ten leaves per tree were collected and measured. The morphology was analysed based on the methods of Pigott (1969). Nine characters of leaves, petioles and young twigs were scored as 0 for *T. cordata* character states, 1 for intermediate character between *T. cordata* and *T. platyphyllos* and 2 for *T. platyphyllos* character states (Table 6.3). The total Hybrid Index (HI) score of each tree was calculated. A HI of six or seven was used as the reference HI of a putative first generation hybrid (Pigott, 1969). The HI scores were obtained from our colleagues at Edge Hill University. The frequency distributions of HI were plotted for each population.

Table 6.3 Characters of *T. cordata*, *T. platyphyllos* and their hybrids used for assessing hybridization based on Pigott (1969) method

Score	*T. cordata* (0)	Hybrid (1)	*T. platyphyllos* (2)
Leaves	- Largest leaves on second order shoots <8 cm long	Largest leaves on second order shoots 8-10 cm long	Largest leaves on second order shoots >10 cm long
	- Adaxial surface flat; terriary veins not raised on abaxial surface	Intermediate	Adaxial surface rugose; tertiary veins prominent on abaxial surface
	- Abaxial surface glaucous	Intermediate	Abaxial surface green
	- Veins hairsless on abaxial surface	Veins with scattered hairs	Veins very hairy
	- No hairs between veins on abaxial surface	Scattered hairs between veins on abaxial surface	Hairy between veins on abaxial surface
	- No hairs on adaxial surface	Scattered hairs on adaxial surface	Many hairs on adaxial surface
Petiole	- < 1.2 mm diameter	1.2-1.5 mm diameter	>1.5 mm diameter
	- No hairs	Few hairs	Many hairs
Young twigs	- No hairs	Few hairs	Many hairs

6.3.4 Correlation between morphological characters and molecular analysis

The consistency of morphological characters and molecular analysis was calculated as the correlation between HI and the value loading on the first axis of PCO of all samples, (dot plots with trend lines) using Microsoft Excel 2010.

6.4 Results

6.4.1 Molecular analysis

Microsatellite analysis of 144 samples from ten UK populations showed high genetic diversity. However, five groups of individuals in four different populations were found to share the same genotype across all 13 loci, presumably as a result of clonal reproduction. Two clones were detected in Dumbleton Dingle (DUD). Seven samples (DUD03, 04, 05, 06, 07, 10 and 11) showed the same genotype at all 13 loci, and two other samples in this population (DUD01 and 02) were also a clone. The other three clones were detected in three populations with two samples per clone, namely Highbury

Wood (HBW08 and 09), Knapp and Papermill (KPM05 and 06) and Tick Wood (TIC02 and 05).

The PCO (Principal Coordinate Analysis, Figure 6.1) shows that the UK samples were clustered into two groups. Considering the morphology, samples from one cluster belonged to the species *T. cordata*, and the other to *T. platyphyllos*. The first two axes of PCO account for 59.88% and 11.44% of the molecular variance (71.32% of the total variation). The *T. platyphyllos* cluster (right) was more compact than the *T. cordata* cluster (left). Samples from Kings' Wood (KIN) and Earl's Hill (EAR) clearly fell into the *T. platyphyllos* cluster and samples from Highbury Wood (HBW) and Roudsea (ROU) clearly fell into the *T. cordata* cluster. Other populations were a mix of *T. cordata* and *T. platyphyllos*, with some hybrids. Twelve samples that were plotted in between the two main clusters (Figure 6.1, with data label) were considered to be hybrids (NewHybrids). These were one sample from Crews Hill Wood (CHW08), seven samples from Dumbleton Dingle (DUD03, 04, 05, 06, 07, 10 and 11), two samples from Knapp and Papermill (KPM05 and 06) and two samples from Tick Wood (TIC02 and 05). However, eleven of these samples were identified as clones, as mentioned earlier. Therefore, four genetically different hybrids were detected from ten UK populations using genetic markers illustrated by PCO. In addition, three samples (CHW05, LPW11 and 12) were plotted closer towards the *T. platyphyllos* cluster. These samples were defined later by NewHybrids as F_2 (CHW05) and *T. platyphyllos* (LPW11 and 12).

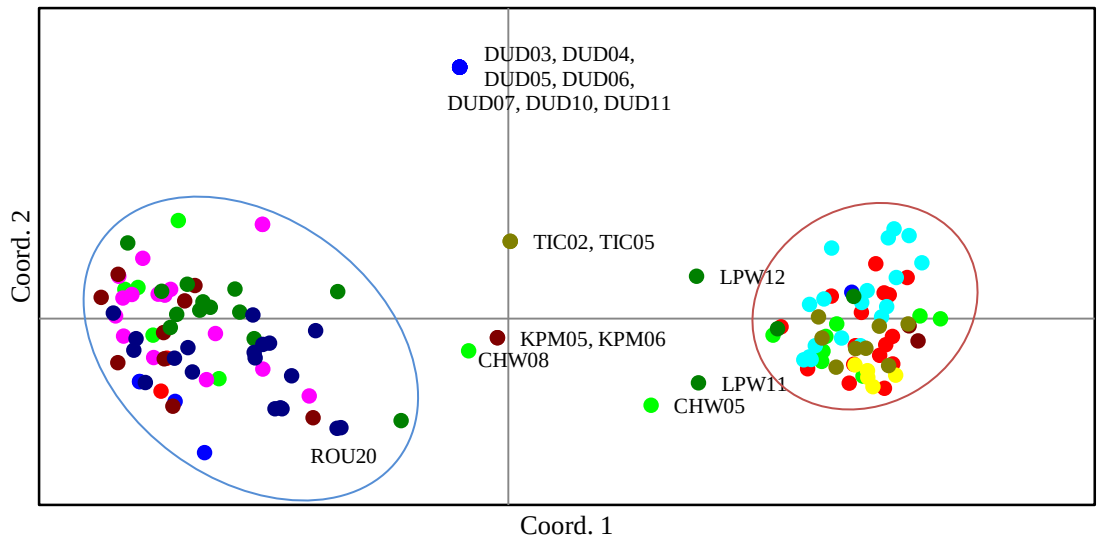


Figure 6.1 Principal Coordinate Analysis of ten UK populations using 13 microsatellite markers. Blue circle indicates samples which coincided largely with *T. cordata* morphology, red circle indicates samples which coincided largely with *T. platyphyllos* morphology. The populations are coded by colour (GenAlEx).

● = ANS, ● = CHW, ● = DUD, ● = EAR, ● = HBW, ● = KIN, ● = KPM, ● = LPW, ● = ROU and ● = TIC

For Baysian clustering only unique genotypes were analysed (a single sample per clone, 134 samples). NewHybrids analysis classified samples into six genealogical classes: *T. cordata*, *T. platyphyllos*, F_1, F_2 and backcross to each parental species (Figure 6.2). The results are consistent with the result from the PCO (CHW, DUD, KPM and TIC). One sample of each population from Crews Hill Wood (CHW8), Dumbleton Dingle (DUD), Knapp and Papermill (KPM) and Tick Wood (TIC) were F_1, whereas one sample from Crews Hill Wood (CHW5) was identified as F_2. Two samples from Lady Park Wood (LPW11 and 12), placed in between parental species in PCO, but closer towards *T. platyphyllos*, were assigned as *T. platyphyllos* in NewHybrids. There was one sample (ROU20) that is undetermined, equally likely F_2 hybrid, backcross to *T. cordata* and pure *T. cordata*, and was plotted near the edge of the *T. cordata* cluster by PCO (Figure 6.1, with data label). The other samples were assigned to be pure parental species (either *T. cordata* or *T. platyphyllos*), which was consistent with the results from the PCO analysis.

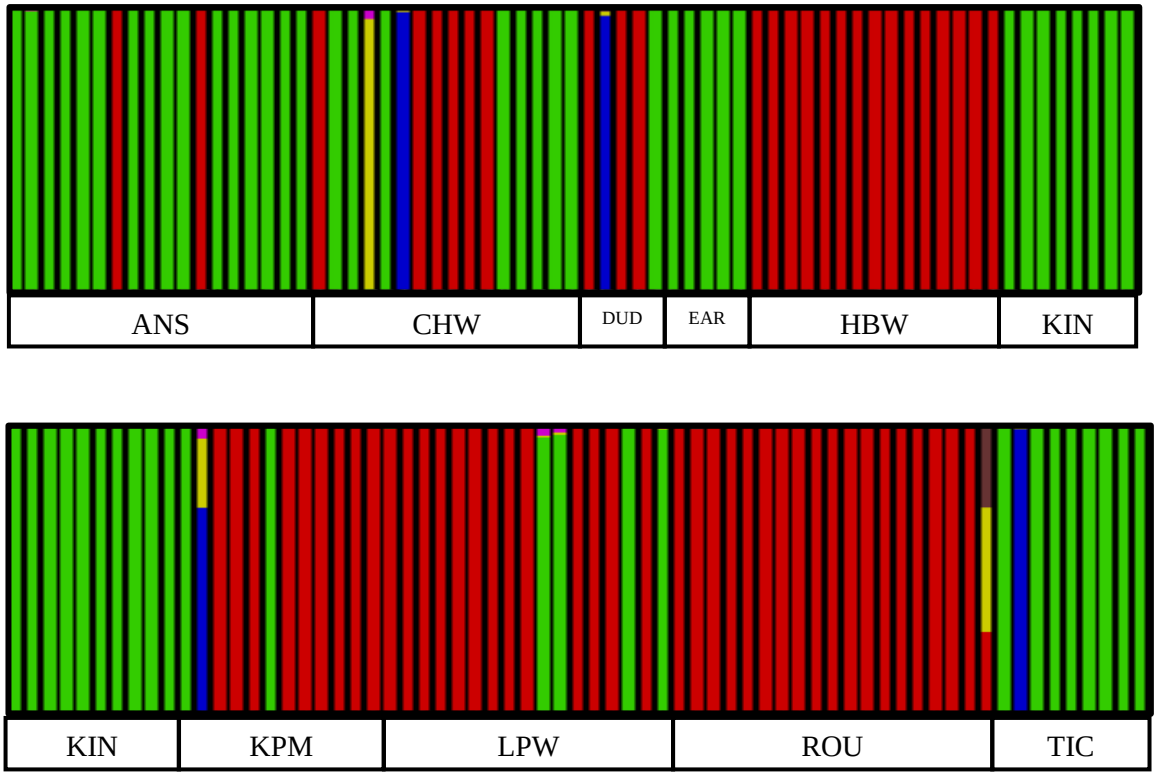


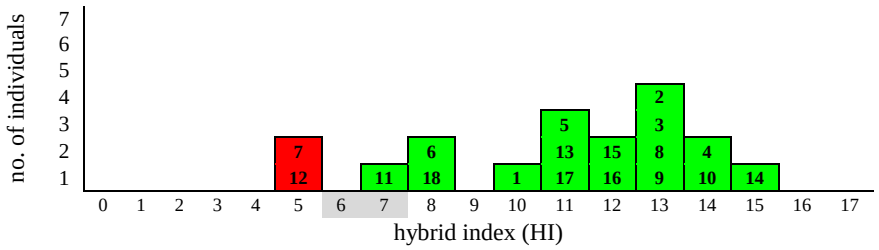
Figure 6.2 Genealogical classes analysis of 136 samples from 10 UK populations. The colours in each column indicate the probabilities of the corresponding classes for each sample (NewHybrids). = *T. cordata*, = *T. platyphyllos*, = F_1 hybrid, = F_2 hybrid, = backcross to *T. cordata*, = backcross to *T. platyphyllos*

6.4.2 Morphological analysis

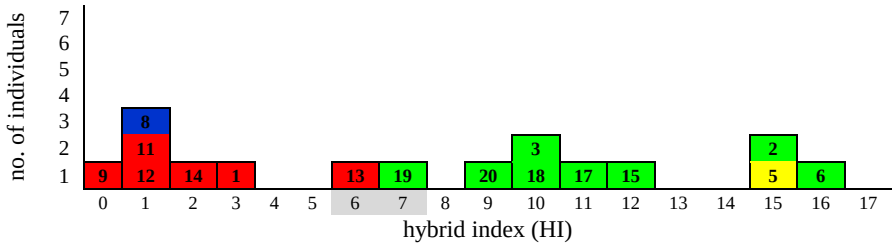
The hybrid index (HI) of our samples ranged from 0 to 17. Based only on morphology, samples with low total scores (0-5) are indicated as *T. cordata*, while samples with a high HI (8-17) are *T. platyphyllos* and those with HI of 6 or 7 are hybrid. The broadest range of HI was in Crews Hill Wood (CHW), Dumbleton Dingle (DUD) and Lady Park Wood (LPW), while Knapp and Papermill (KPM) had morphologically identified mixed species with a distribution in the direction of *T. cordata* and Anston Stones Wood (ANS) in the direction of *T. platyphyllos*. Four populations appeared not to have both pure species in sympatry. Highbury Wood (HBW) and Roudsea (ROU) only had trees identified as *T. cordata* and hybrids, while Tick Wood (TIC) and Kings' Wood (KIN) only had *T. platyphyllos* and hybrids.

Sixteen out of 144 samples (11.11%) were identified as hybrids. Hybrids (with HI of 6 or 7) were detected in almost all of the populations, except in Lady Park Wood (LPW), where only pure species of *T. cordata* and *T. platyphyllos* were identified and in Earl's Hill (EAR), where only *T. platyphyllos* was detected. The frequency distribution of the HI in each population did not reveal a consistent pattern across the sites surveyed (Figure 6.3).

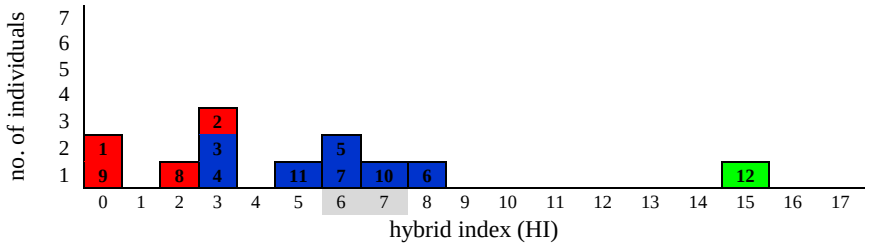
(a) Anston Stones Wood (ANS)



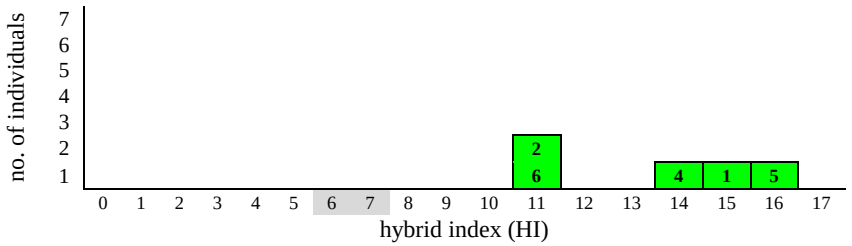
(b) Crews Hill Wood (CHW)



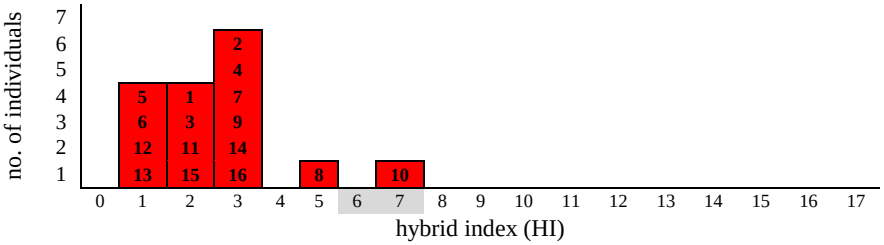
(c) Dumbleton Dingle (DUD)



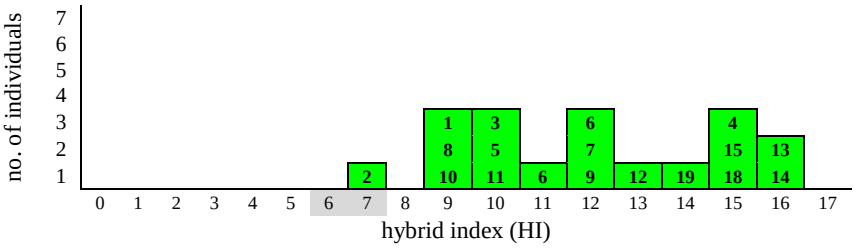
(d) Earl's Hill (EAR)



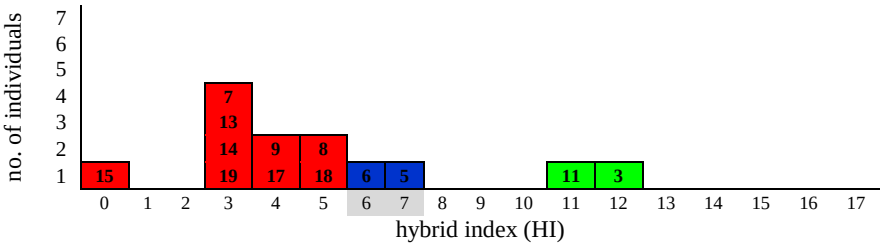
(e) Highbury Wood (HBW)



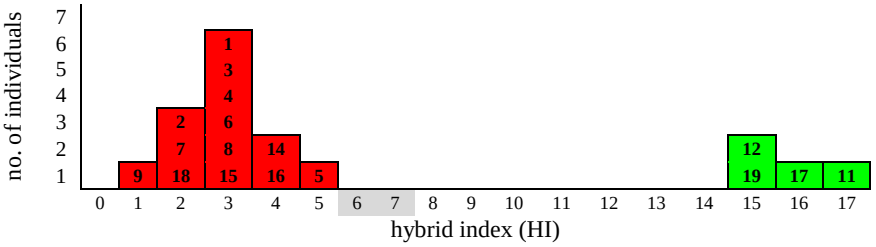
(f) Kings' Wood (KIN)



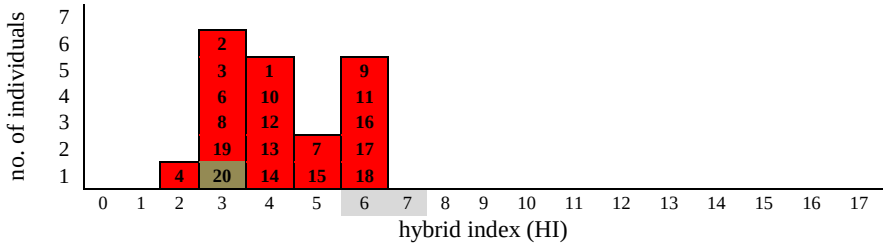
(g) Knapp and Papermill (KPM)



(h) Lady Park Wood (LPW)



(i) Roudsea (ROU)



(j) Tick Wood (TIC)

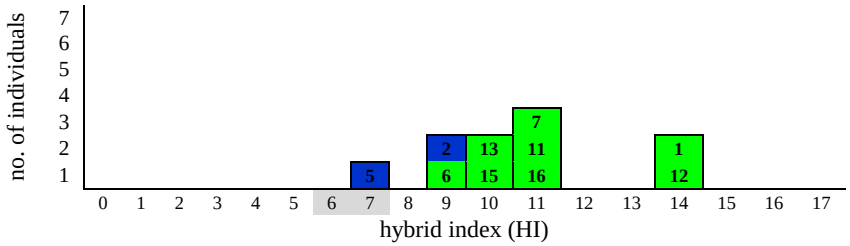


Figure 6.3 The frequency distribution of hybrid index (HI) in 10 populations with sympatric *T. cordata* and *T. platyphyllos*. Numbers on bars are the names of each individual. Colours indicate species and hybrid classes as defined by NewHybrids.

■ = *T. cordata*, ■ = *T. platyphyllos*, ■ = F_1 hybrid, ■ = F_2 hybrid, ■ = undetermined class

6.4.3 Comparison of morphological and molecular analysis

Molecular analysis indicated four first generation hybrids, one second generation hybrid and one undetermined class (ROU20). The four F_1 hybrids consist of three clones in DUD, KPM and TIC and one sample in CHW (12 samples in total) while morphological analysis defined more hybrids (16 samples) with a HI of 6 or 7. Of the twelve hybrids detected in the molecular analysis, only six samples (DUD05, DUD07 DUD10, KPM05, KPM06 and TIC05) had a HI of 6 -7. In some populations, hybrids (HI of 6 or 7) were detected, e.g. in Highbury Wood (HBW) and King's Wood (KIN), but molecular analysis defined these populations as pure *T. cordata* and *T. platyphyllos*, respectively (Figure 6.3).

There was a strong positive correlation between HI and the loading of the first axis of PCO in the 144 samples (R^2 =0.762, Figure 6.4), indicating overall strong coherence of molecular and morphological identification of species and hybrids. The trend line crosses the HI axis (PCO value = 0) at approximately HI 7. However, the HI of the samples identified as the same clone varied from 3 to 8 in DUD (seven samples, DUD03, 04, 05, 06, 07, 10, 11), 6 and 7 in KPM (two samples, KPM05, 06) and 7 and 9 in TIC (two samples, TIC02, 05).

Samples deviating from the trend line indicated conflicting results in hybrid detection between molecular analysis (PCO and NewHybrids) and morphological analysis, if relying on HI classification of 1 – 5 as *T. cordata*, 6 and 7 as hybrid and 8 – 18 as *T. platyphyllos* (Fig 6.3). In the CHW population, CHW08 was assigned as F_1 in NewHybrids, but the HI was low (1, indicating *T. cordata*) and another (CHW05) was defined as F_2 hybrid with a high HI (15, indicating *T. platyphyllos*). DUD03 and DUD04 are assigned as F_1 hybrid, but have a HI indicating *T. cordata*. Although one sample in Roudsea (ROU20) was undetermined class by NewHybrids, but most likely F_2 hybrid, the value of the first HI (3) indicated it as *T. cordata*.

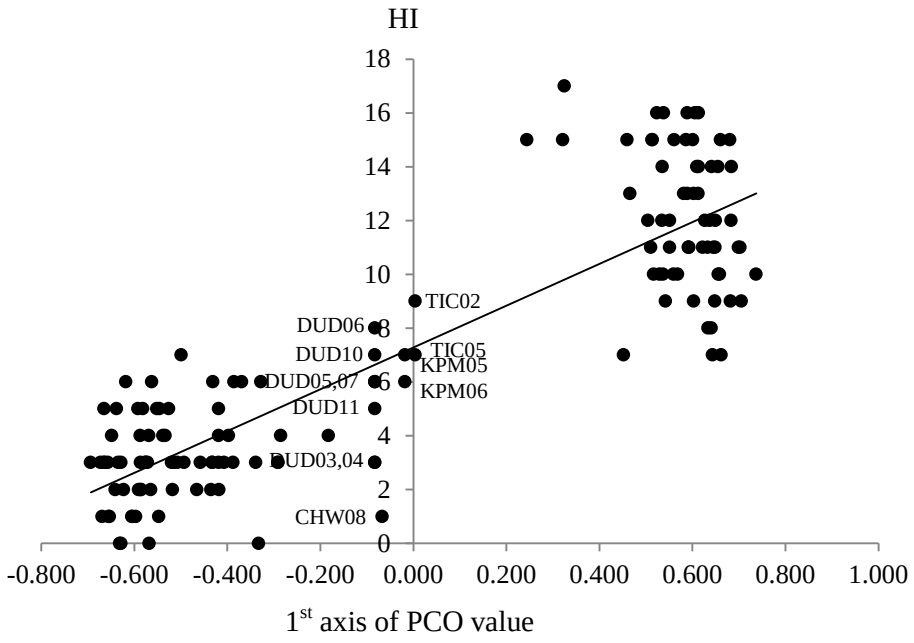


Figure 6.4 Correlation between hybrid index (HI) and the first axis of PCO of all individuals. Labels indicates the hybrids detected by molecular approach.

Details of the HI, the value of the first axis of PCO and species detection results from NewHybrids analysis of each sample are shown in Appendix 11.

6.5 Discussion

6.5.1 Hybridisation and introgression

Here we used molecular and morphological characters to study the biology of *T. cordata* and *T. platyphyllos* in UK populations. The study showed that recently developed nuclear microsatellites have the power to genetically identify the two *Tilia* species native to the UK, as well as their hybrid. We have unequivocally shown that hybridisation and introgression occurs. Although the identification of trees as pure species or hybrid with both methods is not always identical, there is a strong coherence between both methods. The HI ranges of hybrids and parental species overlap and this suggests that a HI of six and seven is not a definite identification of a hybrid. Therefore, molecular analysis is likely to be more reliable than morphological analysis to identify hybrids and subsequent generations.

The detection of hybrids in four of the ten populations supports the previous evidence that hybridisation between these species occurs in nature (Elwes and Henry, 1906; Pigott, 1969; Wicksell and Christensen, 1999). *T. cordata* flowers 14 days later than *T. platyphyllos* and they are both protandrous (Pigott, 2012). Hybrids flower in the intermediate period of the parental species (Weryszko-Chmielewska and Sadowska, 2010). As the flowering times of the species and their hybrids partially overlap, the occurrence of hybridisation, introgression and back-crossing can be expected. The chance of hybridisation can increase if the flowering of *T. platyphyllos* is delayed, for example, when *T. platyphyllos* grows in a colder location (higher up on a mountain) than *T. cordata*.

We detected a small proportion (3%, 4 out of 134) of first generation hybrids and less than 2% (two out of 134) of further generation hybrids. This is a small percentage compared to some other tree species. For example, in the genus *Quercus* 15-55% hybrids were detected (Jensen *et al*, 2009; Moran *et al*, 2012). This low proportion in *Tilia* supports the restricted hybridisation and introgression between *T. cordata* and *T. platyphyllos* reported in a German population using allozymes (Fromm and Hattemer, 2003). It also fit with the fact that the two species have remained two separate

evolutionary genetic entities. However, considering the fact that introgression is evident (two samples), there is scope for gene flow between the species, resulting in possible further evolution and adaptation. Even in populations that currently have only one of the two species present (ROU and TIC) a hybrid was detected. Inevitably, the flowers from these hybrid trees will be pollinated with pollen from the species present in majority, resulting in introgression.

Five clones in four populations (DUD, HBW, KPM and TIC) were identified by molecular markers as three hybrids and two *T. cordata*. It is well-known that *Tilia* has the ability for vegetative reproduction. The tree can sprout from cut or fallen stems and branches that touch the ground can generate new shoots and become a new tree. They can also produce new stems from vertical shoots or from dormant buds in the root collar (Pigott, 1991).

6.5.2 Accord between methods

The results of hybrid detection from the PCO and NewHybrids analyses are largely congruent. Four samples that were plotted between the two parental species (PCO) were defined as first generation hybrids (F_1) after NewHybrids analysis. In addition, there are three samples plotted between the two parental clusters, but closer to *T. platyphyllos* in PCO, and NewHybrids analysis identified one (CHW05) as a second generation hybrid (F_2) and the other two (LPW11 and LPW12) were defined as *T. platyphyllos*. Therefore, NewHybrids analysis not only confirmed the results from PCO that there are hybrids in the populations but also classified the genealogical classes of samples that were plotted in the area where it is hard to define their class. The class of one sample (ROU20) cannot be defined by NewHybrids, being F_2 hybrid (most likely), pure *T. cordata* or backcross to *T. cordata*. It is worth considering that Roudsea is considered to be beyond the range of *T. platyphyllos*.

One second generation hybrid (F_2) was detected in the CHW population, where also a first generation hybrid (F_1) was detected. This indicates that the hybrid between *T. cordata* and *T. platyphyllos* is fertile. Pigott and Huntley (1981) claimed that *T. cordata* is probably self-sterile and therefore out-crossing. Alleles detected in the F_2 (CHW5) were not present in the F_1 hybrid (CHW8), which indicated that this F_1 is not one of the parents of the F_2. Therefore, there must have been other F_1 hybrids in the population. In

the TIC population, the F_1 was detected among pure *T. platyphyllos* samples. This indicates that the other parental species (*T. cordata*) must have been present in the past or *T. cordata* is present in the location but was not sampled. Hybridisation and introgression between these two species was clearly established.

Correlation between morphological characters (HI) and the first axis of PCO showed a strong positive correlation; *T. cordata* samples have a low value of HI and a negative value of the first axis PCO, while *T. platyphyllos* samples have a high HI value and positive value of the first axis PCO. The trend line crosses the hybrid index axis at approximately 7, which is close to the value (6-7) used as a reference for hybrid identification (Pigott, 1969).

The molecular and morphological analyses largely lead to the same conclusion regarding classification of species and hybrids with some conflicting results. Out of 144 samples analysed, 121 samples had consistent results, while 13 samples did not give consistent results. Molecular analysis identified 12 hybrids (F_1) (3 clones) and one F_2 hybrid, while only six of them had HI value of 6 or 7 and the rest were defined as pure species based on morphological analysis (HI). The results showed that it is impossible to identify F_2 trees by morphological analysis. One F_2 detected by molecular analysis (NewHybrids) had HI of 15, which was much higher than any F_1. The HI of *T. cordata* ranged from 0-7, *T. platyphyllos* ranged from 7-17 and the hybrids ranged from 1-9. Studies in other species in some cases show congruent results between these two approaches (Fortini *et al*, 2009; Gailing *et al*, 2012; Howard *et al*, 1997), while others had incongruent or partial congruent results between morphological and genetic characters (Bacilieri *et al*, 1996; Craft and Ashley, 2006; Craft *et al*, 2002; Kremer *et al*, 2002).

This variation and overlap of HI between species and hybrids is probably caused by plasticity and the effect of the environment. The amount of light and position on the tree affects leaf morphology in *Tilia* (Pigott, 1991; Traiser *et al*, 2005). However, we cannot rule out that genes that affect leaf morphology could have a different evolutionary history than 'selectively neutral' microsatellite loci. It is possible that natural selection has affected genes associated with morphological characters and has resulted in weaker distinction of species and hybrids with HI than molecular markers.

This implies that it is difficult to use HI of leaf characters on its own to identify the species or hybrid status of a tree. Similarly, samples that have been shown to be a clone

indicate that the HI of genetically identical samples can have a wide range: HI values of of 0 and 3 (DUD), 3 and 5 (HBW) for *T. cordata* and 6 and 7 (KPM), 7 and 9 (TIC), 3 to 8 (DUD) for F_1. This confirms the strong effect of plasticity and the environment (Abrams, 1994; Garzon *et al*, 2011). A tree with a HI of 6 can be *T.cordata* or hybrid, a tree with HI of 7 can be *T. cordata, T. platyphyllos* or hybrid. The use of standard samples may alleviate this problem to some extent, in particular for amount of hairs on adaxial surface of leaves or young twigs. Therefore, using these nine morphological characters to identify species and hybrids without standard samples carries a risk of an error of detection. Adding characteristics of inflorescences and fruits, which are clearly different between the two parental species, would help an accurate identification. However, it is not universally possible because of the short duration of flowering, flowering higher up in the crown of trees or the tree may not be flowering at all.

Microsatellite analysis seems to be more reliable to detect species and hybrids as this method is not affected by environmental factors. However, the complementarity of the data also highlights opportunities. For example, the variable morphology of trees that were identified as clones may help us understand which characters are the most reliable for identification and which characters seem to have the greatest plasticity. Studying genes associated with those characters could reveal the evolutionary potential for between species gene flow.

Molecular data are certainly more helpful for determining the extent of introgression (generations beyond F_1) than morphological data (Rieseberg and Ellstrand, 1993) and may indicate the direction of introgression, showing the evolutionary potential of the two species. However, in some situations genotyping is not practical and the morphological methods could suffice for the detection of first generation hybrids, in particular if characters from flowers and fruits are included.

6.6 Conclusion

The present study indicates that natural hybridisation between *T. cordata* and *T. platyphyllos* has occurred. Molecular analysis using microsatellite markers clearly separated the two *Tilia* species and accurately detected first and second generation hybrids. Morphological analysis showed some conflict with molecular analysis. The hybrids had overlapping ranges of HI with the pure parental species as morphological

characters are highly affected by environment. This suggested that a HI of six and seven is not a definite identification of a hybrid. In addition, using morphological analysis could identify only a possible first generation hybrid, but identifying a second generation hybrid is clearly not possible. Therefore, molecular analysis should be more reliable for species determination and hybrid detection than morphological analysis. However, in some situations it is not practical to use genotyping. Morphological methods could suffice for the detection of first generation hybrids, if characters from flowers and fruits are included. Accurate species detection using morphological characters could also depend on the experience of the person doing it.

Chapter 7. General discussion

This thesis reports on the genetic diversity and the evolutionary history in an understudied genus, *Tilia*. Microsatellite markers were developed and potentially have a broad application in science and forestry. In Chapter 6, species specific alleles and species specific non-amplification were observed, clearly identifying the two UK species of the genus, *T. cordata* and *T. platyphyllos*, which are difficult to identify using only morphological characters. Many species in the genus are morphologically very similar (Pigott, 2012). In particular hybrids are morphologically not discrete. The cross-amplification of the microsatellite markers in other species in the genus gives the potential to use them for accurate species identification, similar to their use in *T. cordata* and *T. platyphyllos*. In addition, these markers could facilitate other genetic studies in other species, saving time and cost for developing new primer sets.

A combination of microsatellite markers can be used for generating genetic fingerprints of individuals, thus they are important for cultivar identification, which is useful for landscape planners and botanists. The microsatellite markers from this study have already been used to identify *Tilia* trees (*T.* x *europaea*, common lime) to find potential sources of propagating *Tilia* trees instead of the trees that were potentially to be felled in Walpole Park, London (Phuekvilai, unpublished data). In this study, these microsatellite markers were used to investigate the relationships of those individual trees, whether they are genetically identical or closely related. Also, the basal sprout and epicormic buds on the trunk were genotyped to confirm that they are genetically identical to their canopy buds, which indicated that the trees were not grafted on rootstocks. Microsatellite markers have been used to identify cultivars in various plant species, such as olive (*Olea europaea* L.) (Rekik *et al*, 2008), fig (*Ficus carica* L.) (Achtak *et al*, 2009), almond (*Prunus dulcis* Mill.) (Dangl *et al*, 2009) and apple (*Malus* × *domestica* Borkh.) (Patzak *et al*, 2012). In addition, microsatellites developed in one species may allow analysis in taxonomically related species. For example, 52 microsatellite markers developed from two oak species, *Quercus robur* and *Q. petraea*, were used to identify twelve European chestnut (*Castanea sativa*) cultivars grown in Italy (Boccacci *et al*, 2004). Therefore, *Tilia* microsatellite markers are not only useful for genetic fingerprinting within the same genus but may also be useful for other closely related genera, such as cacao (*Theobroma cacao*) and cotton (*Gossypium hirsutum*).

Microsatellite markers can be used to understand the population dynamics within the landscape and broad geographical areas. The investigation of *T. cordata* and *T. platyphyllos* populations across Europe indicated that there is some genetic structure of *Tilia*, based on geographical regions. This information can be important for restoration projects, as it allows practitioners to select source populations for new establishment. These genetic diversities may be associated with local adaptation to the habitat (Linhart and Grant, 1996). At a small scale, assessing the dynamics of genetic diversity and differentiation of a population in a wooded landscape can increase understanding of the effects of population fragmentation. Habitat fragmentation disrupts the processes of population connectivity and reduces effective population size. This leads to the reduction of genetic diversity within populations and increases genetic differentiation between populations (Young *et al*, 1996). In many habitats, where genetic diversity is degraded or species have gone locally distinct, reintroducing and restocking programs need to be considered. For such plans, microsatellites can aid in selection of sources of reproductive materials, such as those from local provenance, to maintain or increase genetic diversity and improve habitat connectivity. Therefore, assessing genetic diversity is essential not only for landscape planners, but also for conservationists to plan the best strategies and select materials to maintain genetic diversity in wooded landscapes.

Microsatellites offer insight into historical processes of colonisation and population evolutionary history. This information is essential for making predictions of contemporary and future colonisation and distribution range of the species. The potential for the survival of a species depends on its adaptation to environmental alterations, such as climate change. Climate is one of the important factors that drive plant distribution. The effects of global climate warming on living species are becoming apparent during the last decades (Yang and Rudolf, 2010). There is much evidence of climate change impacting upon tree species. For example, using microsatellite markers, Truong *et al* (2007) reported that the recent shift in the range of mountain birch, *Betula pubescens* ssp. *tortuosa*, above the treeline in northern Sweden is attributed to climate warming. *Tilia* is a warmth-demanding forest tree. With the warmer climate, they are likely to become more prolific throughout their distribution range. It is predicted that their distribution range may shift northwards (Hemery *et al*, 2010). Microsatellites we developed can be used to increase understanding of the adaptation of *Tilia* to climate changes. An example is use for exploring genetic data along the range edge of the

species to investigate new founding individuals in the newly established populations outside its current range. The investigation of migration rate could indicate whether the range shift of the species is the result of warming climate.

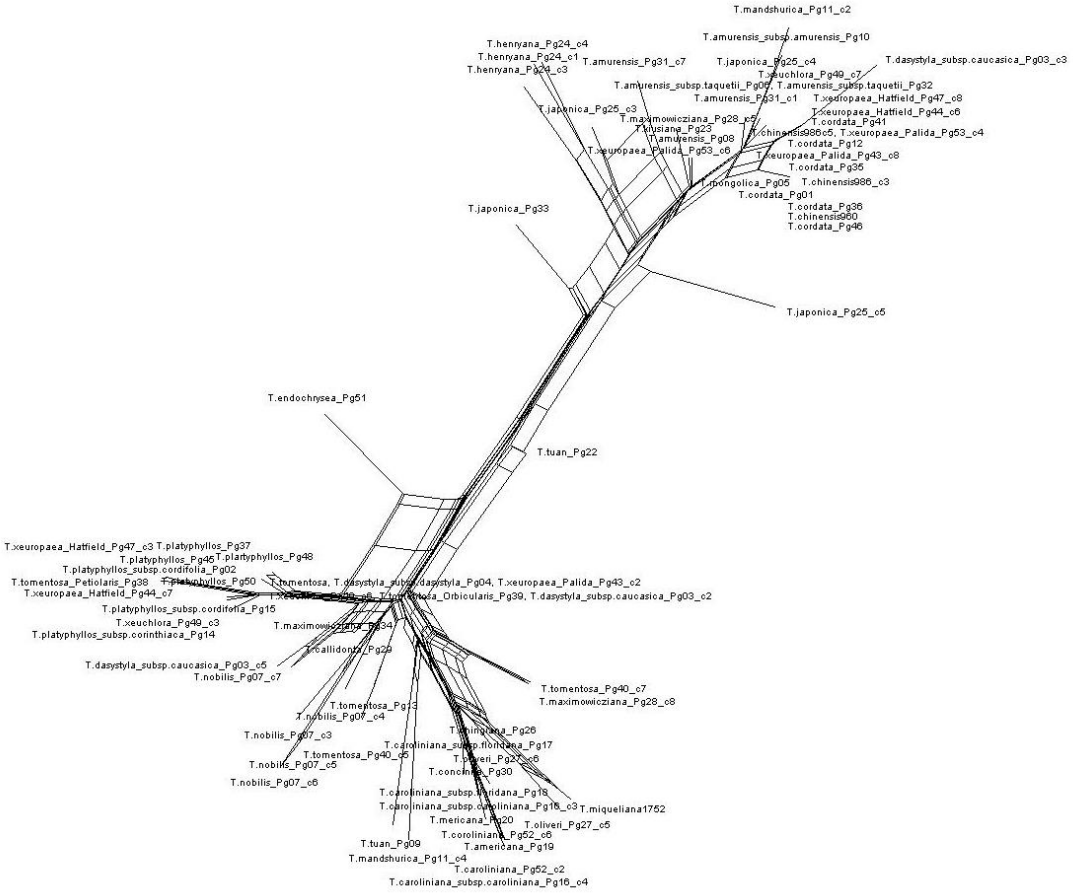
Phylogenetic analysis reveals that the genus *Tilia* is evolving slowly. The species within this genus seems to be genetically fairly close to each other. Thus, a low resolution of phylogenetic relationships was observed (Chapter 2). This is also supported by the cross-amplification of microsatellite markers in most species in the genus (Chapter 3) and the fact that large numbers of species can hybridise. The relationships observed in the phylogeny are largely congruent with their morphology. From this study, it is interesting that the two main species in Europe, *T. cordata* and *T. platyphyllos* seem to be evolutionarily distant, although these two species hybridise naturally (Pigott, 2012). In the trees generated (Chapter 2), the hybrids, *T.* x *europaea*, are clustered into both parental species clades. However, phylogenetic reconstruction in this study had limited numbers of individuals of each species. Therefore, adding more samples of each species could strengthen the species relationships of the genus *Tilia* and could increase understanding of extensive hybridisation among the species.

As shown in Chapter 6, the analysis of hybridisation between *T. cordata* and *T. platyphyllos* in sympatric UK populations indicated that the hybrids are mainly first generation hybrids with limited introgression. Fewer hybrids were detected using microsatellite markers than using morphological approaches, as morphological characters are highly affected by environment. In addition, a few hybrids and introgression were also observed in sympatric populations in Austria and Spain. These individuals were first thought to be pure species from their leaf morphology (Phuekvilai, unpublished data). These results illustrate that molecular analysis using microsatellites should be more reliable for hybrid detection than morphological analysis. In particular, using morphological analysis it is not possible to identify a second generation hybrid. In addition, introgression seems not to be apparent in sympatric populations. This may be because hybridisation is actually infrequent. Thus, more extensive sampling is needed. However, the low level of introgression detected may also be because hybrids are not fertile. Therefore, both pollen and seed fertility of the hybrids need to be investigated and quantified.

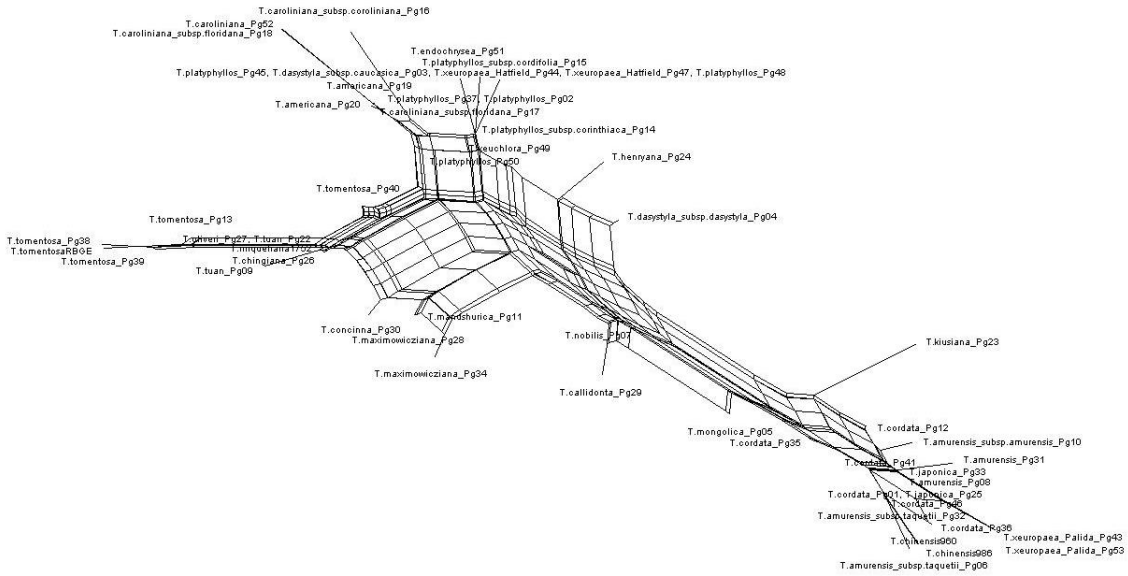
Appendix

Appendix 1. Nuclear primers tested for phylogenetic analysis. * indicated the primers that were successfully amplified in all *Tilia* species.

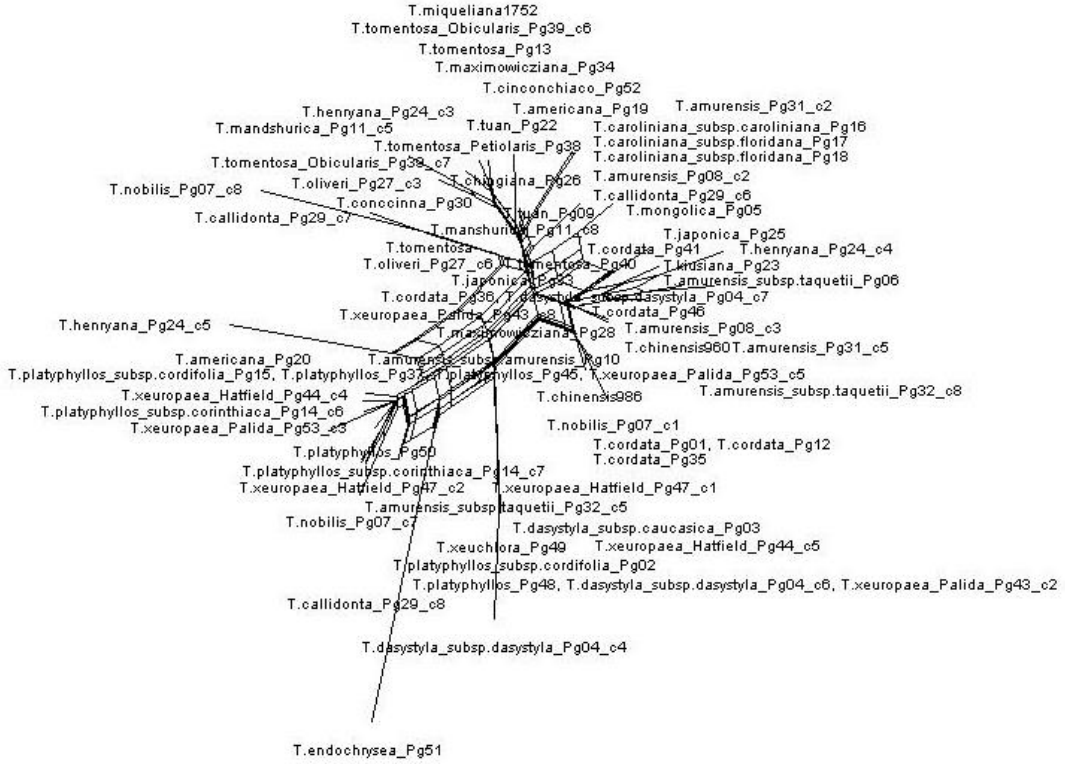
Primer name	Sequence	species of origin	Reference
CelA-F	GATGGAATCTGGGGTTCCTGTTTGC	*Gossypium*	Cronn *et al* (2002)
CelA-R	GGGAACTGATCCAACACCCAGGA		
A1341-F	GCATGCTGAATTGACAGAACCAGCY	*Gossypium*	Cronn *et al* (2002)
A1341-R	CACTCACAAAGTTATGCCGGATGY		
*ADH*x4-3	GGGCAGACTAGGTTTTCCAAAG	*Gossypium*	Cronn *et al* (2002)
ADH-P2	GCACAGCCACACCCCAACCCTG		
*hsp90_15*F	ACGGACAAGAGCAAGCTGGATG	Rosids	Steele *et al* (2008)
*hsp90_15*F	TTGTAGTCTTCCTTGTTCTCAG		
*atpB-51*F*	CCTAGCTTGATGACACCAC	Rosids	Steele *et al* (2008)
*atpB-51*R*	CTTGGACGTATCCTGAAT		
*actin_61*F	ATGGGACAAAAAGATGCTTA	Rosids	Steele *et al* (2008)
*actin_61*R	TAGAAGCACTTCCTGTGGA		
*TcWRKY-11*F	GGTAGTGAATATCCAAGAAGC	*Theobroma cacao*	Borrone *et al* (2007)
*TcWRKY-11*R	ACAGGACATCCAGGAGTTG		
*TcWRKY-12*F	ACGCATCCTAATTGTGAAGTG	*Theobroma cacao*	Borrone *et al* (2007)
*TcWRKY-12*R	TTTTCTAACAGGGCAACCG		
*TcWRKY-13*F*	AAGCAAGTGAAAGGAAGTGAG	*Theobroma cacao*	Borrone *et al* (2007)
*TcWRKY-13*R*	TGAAAGCTCTTGGATCATCCGATGC		
*Grx*F*	TTTCAGCAAGTCGTACTGTCC	*T. phatyphyllos*	Newly designed
*Grx*R*	AATCAGCTTCCCATTCTTGTG		
Nr-F	AACCGCTGATAACTGGATCG	*T. phatyphyllos*	Newly designed
Nr-R	TCGAACCGATATCTCCTTGG		
*ITS4**	TCCTCCGCTTATTGATATGC		(White *et al* (1990))
*ITS5**	GGAAGTAGAAGTCGTAACAAGG		



Appendix 2. Phylogenetic network based on *Grx* region from 22 *Tilia* species (Split Tree4 version 4.1.3.1)



Appendix 3. Phylogenetic network based on *TcWRKY-13* region from 22 *Tilia* species (Split Tree4 version 4.1.3.1)



Appendix 4. Phylogenetic network based on *Atp*B region from 22 *Tilia* species (Split Tree4 version 4.1.3.1)

Appendix 5. Summary data of diversity measures by locus for each population of *Tilia cordata*. N, number of samples analysed; N_a, Average number of alleles; A_e, effective number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; F, fixation index. The significant from Hardy Weinberg (Signif HW) is indicated as ***P<0.001; **P<0.01; *P<0.05; NS, no significant; Mono, monomorphic loci. Null, Null allele detected by Micro-Checker

Pop		Tc4	Tc5	Tc6	Tc7	Tc8	Tc31	Tc915	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
NOSO	N	16	16	16	16	16	16	16	16	16	16	16	16	16
	N_a	7	5	7	3	1	3	4	5	3	3	4	5	11
	A_e	5.020	2.415	4.531	1.459	1.000	1.373	2.522	2.844	1.547	1.373	2.994	2.667	5.505
	H_o	0.875	0.750	0.813	0.375	0.000	0.313	0.688	0.813	0.313	0.313	0.438	0.688	0.813
	H_e	0.801	0.586	0.779	0.314	0.000	0.271	0.604	0.648	0.354	0.271	0.666	0.625	0.818
	F	-0.093	-0.280	-0.043	-0.193	#N/A	-0.151	-0.139	-0.253	0.116	-0.151	0.343	-0.100	0.007
	Signif HW	NS	NS	NS	NS	Mono	NS	NS	NS	NS	NS	**	NS	NS
FINI	N	15	15	15	15	15	15	15	15	15	15	15	15	15
	N_a	4	5	8	2	1	1	8	7	2	1	4	5	11
	A_e	1.915	2.356	3.629	1.471	1.000	1.000	2.368	4.737	1.069	1.000	2.632	2.296	6.818
	H_o	0.200	0.467	0.800	0.400	0.000	0.000	0.533	1.000	0.067	0.000	0.800	0.733	0.933
	H_e	0.478	0.576	0.724	0.320	0.000	0.000	0.578	0.789	0.064	0.000	0.620	0.564	0.853
	F	0.581	0.189	-0.104	-0.250	#N/A	#N/A	0.077	-0.268	-0.034	#N/A	-0.290	-0.299	-0.094
	Signif HW	$***^{Null}$	NS	NS	NS	Mono	Mono	*	NS	NS	Mono	NS	NS	NS
FIMU	N	14	14	14	14	14	14	14	14	14	14	14	14	14
	N_a	3	4	7	3	1	4	6	7	2	5	3	4	14
	A_e	2.390	3.379	2.481	2.074	1.000	1.581	2.010	3.806	1.237	1.876	2.279	2.481	9.561
	H_o	0.500	0.786	0.714	0.429	0.000	0.286	0.500	0.786	0.071	0.500	0.643	0.714	1.000
	H_e	0.582	0.704	0.597	0.518	0.000	0.367	0.503	0.737	0.191	0.467	0.561	0.597	0.895
	F	0.140	-0.116	-0.197	0.172	#N/A	0.222	0.005	-0.066	0.627	-0.071	-0.145	-0.197	-0.117
	Signif HW	NS	NS	NS	NS	Mono	*	NS	NS	*	NS	NS	NS	NS

Appendix 5. continued.

Pop		Tc4	Tc5	Tc6	Tc7	Tc8	Tc31	Tc915	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
DEAB	N	29	30	30	30	30	30	30	30	30	30	30	30	30
	N_a	9	10	6	2	1	3	11	7	3	3	3	6	19
	A_e	4.751	3.673	3.147	1.991	1.000	1.185	7.826	5.678	1.439	1.185	1.909	2.951	8.072
	H_o	0.793	0.833	0.767	0.400	0.000	0.167	1.000	0.833	0.367	0.167	0.400	0.700	0.700
	H_e	0.790	0.728	0.682	0.498	0.000	0.156	0.872	0.824	0.305	0.156	0.476	0.661	0.876
	F	-0.005	-0.145	-0.124	0.196	#N/A	-0.068	-0.146	-0.011	-0.202	-0.068	0.160	-0.059	0.201
	Signif HW	NS	*	NS	NS	Mono	NS	NS	***	NS	NS	NS	NS	NS^{Null}
DEBO	N	20	20	20	20	20	20	20	20	20	20	20	20	20
	N_a	9	8	8	2	2	4	11	10	3	4	4	2	15
	A_e	7.339	3.279	4.545	1.956	1.051	1.597	6.957	5.229	1.107	1.597	2.319	1.663	11.111
	H_o	0.300	0.700	0.800	0.550	0.050	0.400	0.900	0.750	0.100	0.400	0.400	0.550	0.900
	H_e	0.864	0.695	0.780	0.489	0.049	0.374	0.856	0.809	0.096	0.374	0.569	0.399	0.910
	F	0.653	-0.007	-0.026	-0.125	-0.026	-0.070	-0.051	0.073	-0.039	-0.070	0.297	-0.379	0.011
	Signif HW	$***^{Null}$	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
UKHB	N	16	16	16	16	16	16	16	16	16	16	16	16	16
	N_a	6	4	5	3	1	3	6	7	1	4	3	3	11
	A_e	3.459	2.427	1.969	2.216	1.000	1.575	5.020	4.741	1.000	2.124	2.599	1.662	8.678
	H_o	0.688	0.750	0.438	0.375	0.000	0.188	0.750	0.750	0.000	0.438	0.688	0.500	0.625
	H_e	0.711	0.588	0.492	0.549	0.000	0.365	0.801	0.789	0.000	0.529	0.615	0.398	0.885
	F	0.033	-0.276	0.111	0.317	#N/A	0.487	0.063	0.050	#N/A	0.173	-0.117	-0.255	0.294
	Signif HW	NS	NS	NS	NS	Mono	*	NS	NS	Mono	NS	NS	NS	NS^{Null}
GECO	N	20	20	20	20	20	20	20	20	20	20	20	20	20
	N_a	6	11	6	2	1	3	8	7	2	4	3	3	13
	A_e	2.867	5.634	3.478	1.600	1.000	1.766	4.545	4.348	1.220	2.116	1.995	2.247	7.547
	H_o	0.600	0.750	0.650	0.400	0.000	0.600	0.900	0.800	0.200	0.750	0.400	0.600	0.750
	H_e	0.651	0.823	0.713	0.375	0.000	0.434	0.780	0.770	0.180	0.528	0.499	0.555	0.868
	F	0.079	0.088	0.088	-0.067	#N/A	-0.383	-0.154	-0.039	-0.111	-0.422	0.198	-0.081	0.135
	Signif HW	***	NS	*	NS	Mono	NS	NS	NS	NS	NS	NS	NS	NS

Appendix 5. continued.

Pop		Tc4	Tc5	Tc6	Tc7	Tc8	Tc31	Tc915	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
NEMA	N	25	25	25	25	25	25	25	25	25	25	25	25	25
	N_a	8	9	8	2	1	4	9	8	3	4	4	5	18
	A_e	5.656	4.386	3.541	1.625	1.000	1.696	3.492	4.941	1.274	2.039	2.244	2.166	12.019
	H_o	0.720	0.760	0.720	0.360	0.000	0.440	0.760	0.720	0.240	0.560	0.520	0.560	0.800
	H_e	0.823	0.772	0.718	0.385	0.000	0.410	0.714	0.798	0.215	0.510	0.554	0.538	0.917
	F	0.125	0.016	-0.003	0.064	#N/A	-0.072	-0.065	0.097	-0.115	-0.099	0.062	-0.040	0.127
	Signif HW	***	NS	NS	NS	Mono	NS	NS	NS	NS	NS	NS	NS	NS
NEGU	N	25	25	25	25	25	25	25	25	25	25	25	25	25
	N_a	6	10	9	2	1	3	9	6	1	4	2	3	12
	A_e	4.237	7.310	5.208	1.855	1.000	1.486	5.435	4.417	1.000	1.778	1.676	2.399	5.556
	H_o	0.480	0.800	0.840	0.640	0.000	0.320	0.920	0.760	0.000	0.520	0.400	0.720	0.240
	H_e	0.764	0.863	0.808	0.461	0.000	0.327	0.816	0.774	0.000	0.438	0.403	0.583	0.820
	F	0.372	0.073	-0.040	-0.389	#N/A	0.022	-0.127	0.018	#N/A	-0.188	0.008	-0.235	0.707
	Signif HW	$***^{Null}$	NS	NS	NS	Mono	NS	***	NS	Mono	NS	NS	NS	$***^{Null}$
PLLE	N	10	10	10	10	10	10	10	10	10	10	10	10	10
	N_a	4	7	9	1	1	3	8	5	4	5	4	3	13
	A_e	2.597	4.167	5.556	1.000	1.000	2.020	4.878	3.846	1.869	2.740	2.326	2.532	10.000
	H_o	0.800	0.800	0.900	0.000	0.000	0.400	0.800	1.000	0.600	0.700	0.700	0.700	0.800
	H_e	0.615	0.760	0.820	0.000	0.000	0.505	0.795	0.740	0.465	0.635	0.570	0.605	0.900
	F	-0.301	-0.053	-0.098	#N/A	#N/A	0.208	-0.006	-0.351	-0.290	-0.102	-0.228	-0.157	0.111
	Signif HW	NS	NS	NS	Mono	Mono	NS	NS	NS	NS	NS	NS	NS	NS
CZVO	N	21	21	21	21	21	21	21	21	21	21	21	21	21
	N_a	6	7	8	2	1	3	9	8	3	4	2	4	16
	A_e	2.358	2.061	4.846	1.747	1.000	1.213	6.041	4.955	1.156	1.418	1.982	2.557	9.383
	H_o	0.238	0.571	0.905	0.524	0.000	0.190	0.905	0.762	0.143	0.333	0.524	0.429	0.714
	H_e	0.576	0.515	0.794	0.427	0.000	0.176	0.834	0.798	0.135	0.295	0.495	0.609	0.893
	F	0.587	-0.110	-0.140	-0.225	#N/A	-0.084	-0.084	0.045	-0.059	-0.131	-0.057	0.296	0.201
	Signif HW	$***^{Null}$	***	NS	NS	Mono	NS	NS	*	NS	NS	NS	NS	$**^{Null}$

Appendix 5. continued.

Pop		Tc4	Tc5	Tc6	Tc7	Tc8	Tc31	Tc915	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
CZST	N	14	14	14	14	14	14	14	14	14	14	14	14	14
	N_a	9	6	8	4	2	3	10	8	3	3	4	3	13
	A_e	5.158	2.253	4.000	1.903	1.960	1.782	6.759	5.370	1.338	1.537	3.769	1.840	10.316
	H_o	0.286	0.500	0.786	0.429	0.000	0.286	0.786	0.643	0.286	0.357	0.500	0.357	0.857
	H_e	0.806	0.556	0.750	0.474	0.490	0.439	0.852	0.814	0.253	0.349	0.735	0.457	0.903
	F	0.646	0.101	-0.048	0.097	1.000	0.349	0.078	0.210	-0.131	-0.022	0.319	0.218	0.051
	Signif HW	^{**Null}	NS	NS	NS	^{***Null}	**	NS	NS	NS	NS	*	NS	NS
SKBL	N	6	6	6	6	6	6	6	6	6	6	6	6	6
	N_a	7	6	6	3	1	3	8	7	2	4	2	3	8
	A_e	4.235	4.500	4.500	1.412	1.000	1.412	6.545	5.143	1.385	1.714	1.800	2.323	6.545
	H_o	0.500	0.833	1.000	0.167	0.000	0.333	1.000	0.833	0.333	0.500	0.333	0.333	0.667
	H_e	0.764	0.778	0.778	0.292	0.000	0.292	0.847	0.806	0.278	0.417	0.444	0.569	0.847
	F	0.345	-0.071	-0.286	0.429	#N/A	-0.143	-0.180	-0.034	-0.200	-0.200	0.250	0.415	0.213
	Signif HW	NS	NS	NS	**	Mono	NS	NS	NS	NS	NS	NS	NS	NS
FROM	N	20	20	20	20	20	20	20	20	20	20	20	20	20
	N_a	11	9	6	3	1	2	7	8	3	3	2	3	17
	A_e	6.897	4.255	3.846	1.766	1.000	1.471	4.469	4.082	1.288	1.865	1.835	2.292	12.121
	H_o	0.350	0.800	0.800	0.500	0.000	0.100	0.800	0.800	0.150	0.350	0.400	0.700	0.750
	H_e	0.855	0.765	0.740	0.434	0.000	0.320	0.776	0.755	0.224	0.464	0.455	0.564	0.918
	F	0.591	-0.046	-0.081	-0.153	#N/A	0.688	-0.031	-0.060	0.330	0.245	0.121	-0.242	0.183
	Signif HW	^{***Null}	NS	NS	NS	Mono	^{**Null}	NS	*	NS	NS	NS	NS	NS^{Null}
FRCE	N	8	8	8	8	8	8	8	8	8	8	8	8	8
	N_a	5	6	4	2	1	3	6	5	1	3	3	3	11
	A_e	2.977	3.282	2.844	1.753	1.000	1.471	4.571	3.657	1.000	1.471	2.246	1.855	9.143
	H_o	0.500	0.750	0.750	0.125	0.000	0.375	1.000	0.875	0.000	0.375	0.750	0.125	1.000
	H_e	0.664	0.695	0.648	0.430	0.000	0.320	0.781	0.727	0.000	0.320	0.555	0.461	0.891
	F	0.247	-0.079	-0.157	0.709	#N/A	-0.171	-0.280	-0.204	#N/A	-0.171	-0.352	0.729	-0.123
	Signif HW	NS	NS	NS	*	Mono	NS	NS	NS	Mono	NS	NS	^{*Null}	NS

Appendix 5. continued.

Pop		Tc4	Tc5	Tc6	Tc7	Tc8	Tc31	Tc915	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
UARO	N	14	14	14	14	14	12	14	14	14	14	14	14	14
	N_a	5	6	10	2	1	4	6	8	3	5	5	4	12
	A_e	3.500	3.322	7.000	1.324	1.000	2.304	4.000	5.158	1.537	2.925	2.596	3.187	9.800
	H_o	0.500	0.786	0.500	0.286	0.000	0.167	0.786	0.857	0.429	0.786	0.786	0.929	0.929
	H_e	0.714	0.699	0.857	0.245	0.000	0.566	0.750	0.806	0.349	0.658	0.615	0.686	0.898
	F	0.300	-0.124	0.417	-0.167	#N/A	0.706	-0.048	-0.063	-0.226	-0.194	-0.278	-0.353	-0.034
	Signif HW	*	NS	$*^{Null}$	NS	Mono	$**^{Null}$	NS	NS	NS	NS	NS	NS	NS
CHLO	N	25	25	25	25	25	25	25	25	25	25	25	25	25
	N_a	9	9	6	3	1	6	11	7	2	6	3	3	18
	A_e	5.507	4.921	3.551	2.319	1.000	2.013	5.580	5.020	1.041	2.046	2.197	1.947	12.887
	H_o	0.880	0.800	0.680	0.560	0.000	0.360	0.880	0.840	0.040	0.440	0.440	0.400	0.840
	H_e	0.818	0.797	0.718	0.569	0.000	0.503	0.821	0.801	0.039	0.511	0.545	0.486	0.922
	F	-0.075	-0.004	0.053	0.015	#N/A	0.285	-0.072	-0.049	-0.020	0.139	0.192	0.178	0.089
	Signif HW	NS	NS	NS	NS	Mono	NS^{Null}	NS	NS	NS	*	*	NS	**
AUTH	N	15	15	15	15	15	15	15	15	15	15	15	15	15
	N_a	8	7	5	4	1	2	9	8	2	3	3	3	16
	A_e	5.000	3.913	3.061	1.982	1.000	1.471	6.000	5.556	1.069	1.737	2.018	2.261	11.842
	H_o	0.333	0.933	0.600	0.600	0.000	0.267	0.933	0.800	0.067	0.467	0.600	0.533	0.933
	H_e	0.800	0.744	0.673	0.496	0.000	0.320	0.833	0.820	0.064	0.424	0.504	0.558	0.916
	F	0.583	-0.254	0.109	-0.211	#N/A	0.167	-0.120	0.024	-0.034	-0.099	-0.189	0.044	-0.019
	Signif HW	$**^{Null}$	NS	NS	*	Mono	NS	NS	NS	NS	NS	NS	NS	NS
AUSO	N	23	23	23	23	23	23	23	23	23	23	23	23	23
	N_a	10	7	9	5	1	3	14	9	2	3	2	5	22
	A_e	4.661	3.792	4.541	2.543	1.000	1.615	7.007	5.264	1.139	1.615	1.784	2.416	10.907
	H_o	0.739	0.826	0.870	0.609	0.000	0.391	0.783	0.913	0.130	0.391	0.391	0.565	0.870
	H_e	0.785	0.736	0.780	0.607	0.000	0.381	0.857	0.810	0.122	0.381	0.440	0.586	0.908
	F	0.059	-0.122	-0.115	-0.003	#N/A	-0.027	0.087	-0.127	-0.070	-0.027	0.110	0.035	0.043
	Signif HW	NS	NS	NS	NS	Mono	***	NS	NS	NS	***	NS	NS	NS

Appendix 5. continued.

Pop		Tc4	Tc5	Tc6	Tc7	Tc8	Tc31	Tc915	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
AUDO	N	21	21	21	21	21	21	21	21	21	21	21	21	21
	N_a	9	8	7	4	1	5	13	9	2	5	2	4	17
	A_e	5.128	6.485	4.240	1.960	1.000	1.282	5.690	4.846	1.385	1.500	1.982	2.609	7.475
	H_o	0.429	0.905	0.857	0.524	0.000	0.238	0.810	0.714	0.333	0.381	0.238	0.619	0.762
	H_e	0.805	0.846	0.764	0.490	0.000	0.220	0.824	0.794	0.278	0.333	0.495	0.617	0.866
	F	0.468	-0.070	-0.122	-0.069	#N/A	-0.082	0.018	0.100	-0.200	-0.143	0.519	-0.004	0.120
	Signif HW	***^{Null}	NS	***	NS	Mono	NS	NS	**	NS	NS	*^{Null}	NS	NS
AUST	N	22	22	22	22	22	22	22	22	22	22	22	22	22
	N_a	10	7	8	2	1	2	10	8	2	3	3	3	11
	A_e	4.676	3.216	5.232	1.996	1.000	1.658	5.095	4.033	1.146	2.495	2.091	2.547	8.566
	H_o	0.864	0.682	0.682	0.500	0.000	0.273	0.727	0.773	0.136	0.636	0.409	0.636	0.818
	H_e	0.786	0.689	0.809	0.499	0.000	0.397	0.804	0.752	0.127	0.599	0.522	0.607	0.883
	F	-0.099	0.010	0.157	-0.002	#N/A	0.313	0.095	-0.027	-0.073	-0.062	0.216	-0.048	0.074
	Signif HW	NS	NS	NS	NS	Mono	NS	NS	*	NS	NS	NS	NS	NS
HUOL	N	20	20	20	20	20	20	20	20	20	20	20	20	20
	N_a	8	6	6	3	1	3	10	7	1	3	3	3	13
	A_e	5.229	3.137	4.145	1.766	1.000	1.629	6.897	5.634	1.000	1.559	2.005	2.168	6.557
	H_o	0.600	0.800	0.800	0.350	0.000	0.350	0.900	0.900	0.000	0.300	0.450	0.450	0.700
	H_e	0.809	0.681	0.759	0.434	0.000	0.386	0.855	0.823	0.000	0.359	0.501	0.539	0.848
	F	0.258	-0.174	-0.054	0.193	#N/A	0.094	-0.053	-0.094	#N/A	0.164	0.102	0.165	0.174
	Signif HW	NS^{Null}	NS	NS	NS	Mono	NS	NS	NS	Mono	NS	NS	NS	**^{Null}
ITLA	N	19	19	19	19	19	19	18	19	18	19	19	19	18
	N_a	8	8	7	4	1	4	7	5	2	4	2	4	11
	A_e	3.592	4.056	2.490	2.117	1.000	1.391	3.857	3.989	1.057	1.566	1.819	2.195	5.684
	H_o	0.579	0.842	0.632	0.526	0.000	0.158	0.667	0.737	0.056	0.316	0.474	0.737	0.944
	H_e	0.722	0.753	0.598	0.528	0.000	0.281	0.741	0.749	0.054	0.361	0.450	0.544	0.824
	F	0.198	-0.118	-0.056	0.003	#N/A	0.438	0.100	0.017	-0.029	0.126	-0.052	-0.354	-0.146
	Signif HW	NS	NS	NS	NS	Mono	***^{Null}	NS	NS	NS	NS	NS	NS	NS

Appendix 6. Summary data of diversity measures by locus for each population of *T. platyphyllos*. N, number of samples analysed; N_a, Average number of alleles; A_e, effective number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; F, fixation index. The significant from Hardy Weinberg (Signif HW) is indicated as ***P<0.001; **P<0.01; *P<0.05; NS, no significant; Mono, monomorphic loci. Null, Null allele detected by Micro-Checker

Pop		Tc4	Tc5	Tc6	Tc7	Tc8	Tc31	Tc915	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
DEBO	N	10	10	10	10	10	10	10	10	10	10	10	10	10
	N_a	6	3	4	4	6	6	7	3	6	6	2	5	5
	A_e	4.545	2.020	2.857	2.985	2.857	4.762	5.128	2.299	4.348	4.762	1.342	3.226	3.226
	H_o	1.000	0.500	1.000	0.600	0.800	1.000	1.000	0.900	0.800	1.000	0.300	0.600	1.000
	H_e	0.780	0.505	0.650	0.665	0.650	0.790	0.805	0.565	0.770	0.790	0.255	0.690	0.690
	F	-0.282	0.010	-0.538	0.098	-0.231	-0.266	-0.242	-0.593	-0.039	-0.266	-0.176	0.130	-0.449
	Signif HW	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS
UKKI	N	18	18	18	18	18	18	18	18	18	18	18	16	18
	N_a	8	9	6	4	5	7	8	7	9	7	4	5	11
	A_e	5.635	5.635	4.235	3.192	1.940	3.927	4.909	4.985	6.894	4.263	2.274	2.639	5.268
	H_o	0.889	0.833	0.889	0.722	0.444	0.611	0.611	0.889	0.833	0.667	0.611	0.375	0.667
	H_e	0.823	0.823	0.764	0.687	0.485	0.745	0.796	0.799	0.855	0.765	0.560	0.621	0.810
	F	-0.081	-0.013	-0.164	-0.052	0.083	0.180	0.233	-0.112	0.025	0.129	-0.091	0.396	0.177
	Signif HW	NS	NS	NS	NS	NS	NS^{Null}	NS	NS	NS	NS	NS	$***^{Null}$	NS
GELC_Tp	N	31	31	31	31	31	31	31	31	31	31	31	31	31
	N_a	9	13	9	12	11	8	12	8	13	8	4	6	16
	A_e	4.866	7.421	5.309	7.226	5.653	4.916	5.523	2.490	5.125	4.916	1.220	2.010	8.110
	H_o	0.871	0.839	0.774	0.935	0.903	0.839	0.839	0.613	0.839	0.839	0.194	0.419	0.839
	H_e	0.794	0.865	0.812	0.862	0.823	0.797	0.819	0.598	0.805	0.797	0.181	0.503	0.877
	F	-0.096	0.031	0.046	-0.086	-0.097	-0.053	-0.024	-0.024	-0.042	-0.053	-0.072	0.166	0.043
	Signif HW	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS

Appendix 6. continued.

Pop		Tc4	Tc5	Tc6	Tc7	Tc8	Tc31	Tc915	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
AULE	N	25	25	25	25	25	25	25	25	25	25	25	25	25
	N_a	11	14	12	14	9	12	12	9	12	12	4	6	14
	A_e	4.562	7.764	6.378	8.446	5.459	7.485	7.440	3.666	5.708	7.576	2.137	1.947	10.081
	H_o	0.800	0.760	0.880	0.880	0.880	0.760	0.960	0.680	0.920	0.800	0.640	0.400	0.840
	H_e	0.781	0.871	0.843	0.882	0.817	0.866	0.866	0.727	0.825	0.868	0.532	0.486	0.901
	F	-0.025	0.128	-0.044	0.002	-0.077	0.123	-0.109	0.065	-0.115	0.078	-0.203	0.178	0.067
	Signif HW	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*
SKBL	N	7	7	7	7	7	7	7	7	7	7	7	7	7
	N_a	8	11	8	9	7	6	11	5	10	7	3	1	9
	A_e	6.533	8.909	6.125	7.000	5.444	3.161	9.800	3.063	8.167	3.920	1.556	1.000	6.533
	H_o	1.000	0.857	0.714	0.857	0.857	0.857	1.000	0.857	0.857	1.000	0.429	0.000	1.000
	H_e	0.847	0.888	0.837	0.857	0.816	0.684	0.898	0.673	0.878	0.745	0.357	0.000	0.847
	F	-0.181	0.034	0.146	0.000	-0.050	-0.254	-0.114	-0.273	0.023	-0.342	-0.200	#N/A	-0.181
	Signif HW	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	Mono	NS
CHFR	N	24	24	24	24	24	24	24	24	24	24	24	24	24
	N_a	11	13	11	8	10	10	14	7	15	10	4	6	19
	A_e	4.683	6.436	5.878	5.166	6.508	4.861	9.000	3.740	8.727	4.721	1.293	1.993	9.143
	H_o	0.917	0.917	0.792	0.958	0.833	0.917	0.958	0.667	0.875	0.792	0.250	0.458	0.708
	H_e	0.786	0.845	0.830	0.806	0.846	0.794	0.889	0.733	0.885	0.788	0.227	0.498	0.891
	F	-0.166	-0.085	0.046	-0.188	0.015	-0.154	-0.078	0.090	0.012	-0.004	-0.103	0.080	0.205
	Signif HW	NS	NS	NS	**	NS	*	NS	NS	NS	*	NS	NS	***[Null]
FRCE	N	13	13	13	13	13	13	13	13	13	13	13	13	13
	N_a	8	14	6	5	7	7	9	6	10	7	4	4	11
	A_e	5.729	10.903	4.278	2.380	4.390	5.121	6.377	1.965	3.976	5.121	1.617	2.099	7.042
	H_o	0.769	1.000	0.923	0.615	0.769	1.000	0.923	0.615	0.769	1.000	0.385	0.077	0.846
	H_e	0.825	0.908	0.766	0.580	0.772	0.805	0.843	0.491	0.749	0.805	0.382	0.524	0.858
	F	0.068	-0.101	-0.205	-0.061	0.004	-0.243	-0.095	-0.253	-0.028	-0.243	-0.008	0.853	0.014
	Signif HW	NS	NS	**	NS	NS	NS	NS	NS	NS	NS	NS	***[Null]	NS

Appendix 6. continued.

Pop		Tc4	Tc5	Tc6	Tc7	Tc8	Tc31	Tc915	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
FRIS	N	20	20	20	20	20	20	20	20	20	20	20	20	20
	N_a	8	15	8	9	6	9	12	9	14	10	1	8	14
	A_e	5.556	7.080	5.556	4.167	4.420	5.063	8.081	2.432	9.639	5.128	1.000	2.740	7.692
	H_o	0.850	1.000	0.700	0.800	0.900	0.800	0.900	0.650	0.850	0.850	0.000	0.600	0.850
	H_e	0.820	0.859	0.820	0.760	0.774	0.803	0.876	0.589	0.896	0.805	0.000	0.635	0.870
	F	-0.037	-0.164	0.146	-0.053	-0.163	0.003	-0.027	-0.104	0.052	-0.056	#N/A	0.055	0.023
	Signif HW	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	Mono	NS	*
FRCC	N	20	20	20	20	20	20	20	20	20	20	20	20	20
	N_a	11	10	7	7	8	9	12	5	11	9	2	6	15
	A_e	6.452	5.674	4.188	4.790	4.908	4.969	7.547	2.180	5.674	4.908	1.342	3.448	8.081
	H_o	0.700	0.900	0.750	0.850	0.800	0.900	0.750	0.300	0.750	0.950	0.300	0.750	0.900
	H_e	0.845	0.824	0.761	0.791	0.796	0.799	0.868	0.541	0.824	0.796	0.255	0.710	0.876
	F	0.172	-0.093	0.015	-0.074	-0.005	-0.127	0.135	0.446	0.090	-0.193	-0.176	-0.056	-0.027
	Signif HW	**	NS	NS	NS	NS	NS	NS	**	NS	NS	NS	NS	NS
AUSO	N	15	14	15	14	15	15	15	15	15	15	15	14	15
	N_a	10	12	9	8	9	10	13	9	10	10	4	4	11
	A_e	6.081	8.167	5.233	2.545	6.000	6.716	6.522	4.327	6.716	6.716	1.724	1.248	7.627
	H_o	0.933	1.000	0.800	0.571	0.800	0.867	0.933	0.733	0.733	0.867	0.533	0.214	0.800
	H_e	0.836	0.878	0.809	0.607	0.833	0.851	0.847	0.769	0.851	0.851	0.420	0.199	0.869
	F	-0.117	-0.140	0.011	0.059	0.040	-0.018	-0.102	0.046	0.138	-0.018	-0.270	-0.077	0.079
	Signif HW	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
AUDO	N	14	14	14	14	14	14	14	14	14	14	14	14	14
	N_a	9	9	8	8	9	8	8	6	10	8	2	2	7
	A_e	6.426	6.426	5.091	5.521	5.765	4.723	5.158	3.698	6.877	4.723	1.508	1.074	4.308
	H_o	1.000	0.929	0.929	0.857	1.000	0.857	0.929	0.786	1.000	0.857	0.429	0.071	0.429
	H_e	0.844	0.844	0.804	0.819	0.827	0.788	0.806	0.730	0.855	0.788	0.337	0.069	0.768
	F	-0.184	-0.100	-0.156	-0.047	-0.210	-0.087	-0.152	-0.077	-0.170	-0.087	-0.273	-0.037	0.442
	Signif HW	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*[Null]

Appendix 6. continued.

Pop		Tc4	Tc5	Tc6	Tc7	Tc8	Tc31	Tc915	Tc920	Tc927	Tc937	Tc943	Tc951	Tc963
SPHU	N	7	7	7	7	7	7	7	7	7	7	7	7	7
	N_a	5	3	6	4	4	5	7	4	5	5	2	3	6
	A_e	3.063	2.000	3.920	2.970	1.849	3.769	4.900	2.970	4.083	3.769	1.690	1.556	4.455
	H_o	0.857	0.714	0.571	0.714	0.286	0.714	0.857	0.857	0.714	0.714	0.571	0.286	1.000
	H_e	0.673	0.500	0.745	0.663	0.459	0.735	0.796	0.663	0.755	0.735	0.408	0.357	0.776
	F	-0.273	-0.429	0.233	-0.077	0.378	0.028	-0.077	-0.292	0.054	0.028	-0.400	0.200	-0.289
	Signif HW	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS
GRAG	N	10	10	10	10	10	10	10	10	10	10	10	10	10
	N_a	12	6	10	9	12	8	6	6	12	8	3	5	10
	A_e	7.407	3.571	8.333	5.882	8.333	5.882	4.878	3.571	8.333	6.250	1.361	1.724	6.250
	H_o	0.800	0.700	0.900	0.900	1.000	0.300	0.700	0.600	0.800	0.400	0.300	0.500	0.400
	H_e	0.865	0.720	0.880	0.830	0.880	0.830	0.795	0.720	0.880	0.840	0.265	0.420	0.840
	F	0.075	0.028	-0.023	-0.084	-0.136	0.639	0.119	0.167	0.091	0.524	-0.132	-0.190	0.524
	Signif HW	NS	NS	NS	*	NS	$**^{Null}$	NS	*	NS	$*^{Null}$	NS	NS	$**^{Null}$

Appendix 7. Pairwise F_{st} between 23 *T. cordata* populations

	NOSO	FINI	FIMU	DEAB	DEBO	GECO	UKHB	NEMA	NEGU	PLLE	CZVO	CZST	SKBL	FROM	FRCE	UARO	AUTH	AUSO	AUDO	AUST	CHLO	HUOL	ITLA
NOSO	0.000																						
FINI	0.166	0.000																					
FIMU	0.145	0.031	0.000																				
DEAB	0.094	0.089	0.081	0.000																			
DEBO	0.082	0.085	0.077	0.046	0.000																		
GECO	0.093	0.126	0.076	0.081	0.063	0.000																	
UKHB	0.157	0.134	0.100	0.079	0.051	0.102	0.000																
NEMA	0.092	0.103	0.057	0.055	0.034	0.029	0.056	0.000															
NEGU	0.113	0.124	0.094	0.065	0.063	0.060	0.078	0.038	0.000														
PLLE	0.102	0.175	0.099	0.092	0.103	0.070	0.134	0.053	0.083	0.000													
CZVO	0.081	0.076	0.072	0.057	0.044	0.054	0.101	0.045	0.087	0.063	0.000												
CZST	0.073	0.118	0.101	0.082	0.051	0.071	0.104	0.053	0.087	0.076	0.050	0.000											
SKBL	0.057	0.102	0.049	0.043	0.043	0.019	0.094	0.021	0.023	0.035	0.024	0.027	0.000										
FROM	0.099	0.129	0.083	0.076	0.038	0.064	0.064	0.025	0.049	0.078	0.056	0.067	0.038	0.000									
FRCE	0.101	0.086	0.063	0.057	0.038	0.047	0.063	0.024	0.038	0.077	0.037	0.056	0.003	0.042	0.000								
UARO	0.131	0.168	0.089	0.118	0.101	0.085	0.130	0.069	0.097	0.015	0.079	0.084	0.038	0.075	0.098	0.000							
AUTH	0.079	0.089	0.062	0.049	0.033	0.023	0.077	0.028	0.025	0.050	0.029	0.037	0.000	0.040	0.029	0.053	0.000						
AUSO	0.100	0.121	0.090	0.101	0.068	0.074	0.110	0.070	0.101	0.113	0.076	0.073	0.050	0.074	0.077	0.112	0.061	0.000					
AUDO	0.082	0.112	0.073	0.094	0.063	0.044	0.104	0.053	0.093	0.108	0.066	0.075	0.031	0.049	0.055	0.108	0.061	0.021	0.000				
AUST	0.110	0.095	0.066	0.065	0.034	0.057	0.064	0.037	0.052	0.069	0.049	0.067	0.047	0.026	0.056	0.062	0.022	0.077	0.070	0.000			
CHLO	0.097	0.103	0.062	0.076	0.047	0.067	0.074	0.043	0.048	0.085	0.059	0.073	0.036	0.028	0.060	0.073	0.030	0.075	0.060	0.031	0.000		
HUOL	0.080	0.108	0.093	0.050	0.042	0.051	0.099	0.051	0.062	0.064	0.035	0.044	0.015	0.060	0.043	0.073	0.009	0.071	0.079	0.047	0.065	0.000	
ITLA	0.180	0.091	0.088	0.085	0.085	0.105	0.104	0.105	0.104	0.155	0.099	0.132	0.102	0.110	0.078	0.146	0.079	0.134	0.114	0.086	0.092	0.110	0.000

Appendix 8. Pairwise F_{st} between 13 *T. platyphyllos* populations

	DEBO	UKKI	GELC	FRCE	FRCC	FRIS	SKBL	AUSO	AULE	AUDO	CHFR	SPHU	GRAG
DEBO	0.000												
UKKI	0.155	0.000											
GELC	0.137	0.095	0.000										
FRCE	0.135	0.088	0.044	0.000									
FRCC	0.118	0.108	0.059	0.068	0.000								
FRIS	0.128	0.108	0.049	0.049	0.039	0.000							
SKBL	0.143	0.082	0.049	0.046	0.060	0.048	0.000						
AUSO	0.134	0.086	0.039	0.061	0.079	0.053	0.034	0.000					
AULE	0.135	0.083	0.047	0.059	0.069	0.065	0.020	0.021	0.000				
AUDO	0.161	0.120	0.075	0.099	0.102	0.086	0.054	0.048	0.042	0.000			
CHFR	0.132	0.087	0.044	0.044	0.068	0.051	0.025	0.045	0.047	0.064	0.000		
SPHU	0.171	0.137	0.096	0.138	0.103	0.134	0.112	0.103	0.114	0.137	0.114	0.000	
GRAG	0.133	0.119	0.074	0.088	0.089	0.077	0.063	0.060	0.062	0.061	0.083	0.111	0.000

Appendix 9. Pairwise distances (km) between 23 *T. cordata* populations

	NOSO	FINI	FIMU	DEAB	DEBO	GECO	UKHB	NEMA	NEGU	PLLE	CZVO	CZST	SKBL	FROM	FRCE	UARO	AUTH	AUSO	AUDO	AUST	CHLO	HUOL	ITLA
NOSO	0.00																						
FINI	1210.47	0.00																					
FIMU	1178.83	87.29	0.00																				
DEAB	557.84	1240.61	1175.48	0.00																			
DEBO	782.79	1379.63	1306.59	231.92	0.00																		
GECO	1106.87	1498.45	1416.61	549.12	331.07	0.00																	
UKHB	1218.69	2219.91	2156.36	981.20	895.96	1012.54	0.00																
NEMA	1229.84	1885.27	1807.69	739.13	528.42	435.82	647.36	0.00															
NEGU	1230.13	1879.02	1801.27	736.62	524.44	426.84	657.20	10.17	0.00														
PLLE	1625.22	1354.32	1269.05	1143.29	1021.57	786.31	1794.77	1174.86	1164.75	0.00													
CZVO	1418.56	1566.98	1480.31	867.44	671.72	353.49	1308.75	671.55	661.38	513.35	0.00												
CZST	1504.69	1545.21	1457.95	963.61	783.05	479.19	1447.07	809.10	798.93	388.20	138.32	0.00											
SKBL	1639.66	1570.45	1483.45	1109.19	939.38	643.17	1609.12	967.75	957.61	286.53	301.90	165.35	0.00										
FROM	1441.47	2086.03	2006.72	957.22	744.27	602.84	686.36	218.11	220.75	1265.36	752.10	881.01	1024.99	0.00									
FRCE	1377.73	2094.54	2017.65	924.00	727.03	643.32	560.44	212.06	219.70	1351.78	839.72	973.11	1123.33	125.92	0.00								
UARO	1842.50	1483.65	1401.53	1363.78	1238.17	987.96	1986.35	1353.79	1343.62	220.54	682.32	545.65	397.67	1421.86	1517.76	0.00							
AUTH	1566.92	1668.58	1581.39	1015.55	816.94	492.87	1404.02	757.73	747.72	494.80	148.38	127.84	225.81	801.20	902.69	623.34	0.00						
AUSO	1582.83	1709.56	1622.43	1028.88	824.71	496.65	1382.71	735.46	725.57	538.32	168.76	172.66	264.48	768.77	873.16	660.55	45.50	0.00					
AUDO	1677.16	1731.68	1644.39	1127.26	930.01	605.76	1498.06	850.70	840.87	485.18	259.93	194.80	198.65	874.10	982.17	575.09	113.13	117.23	0.00				
AUST	1649.38	1990.79	1905.03	1098.20	867.81	563.58	1161.80	546.33	538.74	898.54	440.10	521.62	621.21	485.54	607.81	1010.31	404.00	360.24	435.62	0.00			
CHLO	1623.15	2091.35	2007.42	1091.59	860.74	606.16	992.64	429.39	424.60	1083.42	594.36	697.90	813.99	306.30	432.20	1207.79	590.65	549.65	635.57	204.43	0.00		
HUOL	1840.40	1774.18	1687.59	1299.91	1114.50	798.11	1702.07	1054.81	1045.04	439.87	444.98	336.48	216.69	1068.67	1180.28	447.71	310.24	322.02	204.97	607.40	811.74	0.00	
ITLA	1793.31	2076.20	1989.57	1239.27	1010.61	696.61	1297.86	695.08	687.84	904.26	509.30	556.23	618.08	613.68	738.84	984.13	428.86	383.63	419.89	152.14	310.74	551.20	0.00

Appendix 10. Pairwise distances (km) between 13 *T. platyphyllos* populations

	DEBO	UKKI	GELC	FRCE	FRCC	FRIS	SKBL	AUSO	AULE	AUDO	CHFR	SPHU	GRAG
DEBO	0.00												
UKKI	709.84	0.00											
GELC	478.51	990.45	0.00										
FRCE	727.03	535.40	701.76	0.00									
FRCC	1387.60	1146.09	1185.60	692.23	0.00								
FRIS	1466.99	1130.01	1317.42	745.18	189.46	0.00							
SKBL	939.38	1483.15	498.01	1123.33	1410.82	1578.92	0.00						
AUSO	824.71	1273.60	346.26	873.16	1154.03	1318.45	264.48	0.00					
AULE	889.33	1352.06	412.87	949.65	1204.89	1373.51	205.94	78.75	0.00				
AUDO	930.01	1390.63	453.86	982.17	1217.13	1388.28	198.65	117.23	41.08	0.00			
CHFR	897.14	918.21	634.96	386.50	551.16	693.86	902.90	638.56	702.18	723.93	0.00		
SPHU	1534.52	1206.35	1370.19	815.65	203.82	76.34	1613.25	1355.20	1407.37	1420.22	740.63	0.00	
GRAG	1836.01	2270.45	1364.03	1798.26	1737.51	1926.76	915.03	1024.16	952.06	910.99	1445.18	1924.35	0.00

Appendix 11. Details of hybrid index, value of the first axis of PCO and species detection results from NewHybrids analysis of each sample in 10 locations in the UK

No.	Sample	Hybrid index	1st axis of PCO value	NewHybrids identification
1	ANS01	10	0.656	*T. platyphyllos*
2	ANS02	13	0.466	*T. platyphyllos*
3	ANS03	13	0.582	*T. platyphyllos*
4	ANS04	14	0.642	*T. platyphyllos*
5	ANS05	11	0.511	*T. platyphyllos*
6	ANS06	8	0.634	*T. platyphyllos*
7	ANS07	5	-0.592	*T. cordata*
8	ANS08	13	0.603	*T. platyphyllos*
9	ANS09	13	0.589	*T. platyphyllos*
10	ANS10	14	0.655	*T. platyphyllos*
11	ANS11	7	0.644	*T. platyphyllos*
12	ANS12	5	-0.582	*T. cordata*
13	ANS13	11	0.551	*T. platyphyllos*
14	ANS14	15	0.680	*T. platyphyllos*
15	ANS15	12	0.627	*T. platyphyllos*
16	ANS16	12	0.650	*T. platyphyllos*
17	ANS17	11	0.592	*T. platyphyllos*
18	ANS18	8	0.641	*T. platyphyllos*
19	CHW01	3	-0.493	*T. cordata*
20	CHW02	15	0.561	*T. platyphyllos*
21	CHW03	10	0.537	*T. platyphyllos*
22	CHW05	15	0.244	F_2
23	CHW06	16	0.605	*T. platyphyllos*
24	CHW08	1	-0.067	F_1
25	CHW09	0	-0.631	*T. cordata*
26	CHW11	1	-0.605	*T. cordata*
27	CHW12	1	-0.653	*T. cordata*
28	CHW13	6	-0.562	*T. cordata*
29	CHW14	2	-0.587	*T. cordata*
30	CHW15	12	0.535	*T. platyphyllos*
31	CHW17	11	0.702	*T. platyphyllos*
32	CHW18	10	0.737	*T. platyphyllos*
33	CHW19	7	0.452	*T. platyphyllos*
34	CHW20	9	0.542	*T. platyphyllos*
35	DUD01[a]	0	-0.628	*T. cordata*
36	DUD02[a]	3	-0.628	*T. cordata*
37	DUD03[b]	3	-0.083	F_1
38	DUD04[b]	3	-0.083	F_1
39	DUD05[b]	6	-0.083	F_1

Appendix 11. Continued.

No.	Sample	Hybrid index	1st axis of PCO value	NewHybrids identification
40	DUD06[b]	8	-0.083	F_1
41	DUD07[b]	6	-0.083	F_1
42	DUD08	2	-0.517	*T. cordata*
43	DUD09	0	-0.568	*T. cordata*
44	DUD10[b]	7	-0.083	F_1
45	DUD11[b]	5	-0.083	F_1
46	DUD12	15	0.586	*T. platyphyllos*
47	EAR1	15	0.661	*T. platyphyllos*
48	EAR2	11	0.592	*T. platyphyllos*
49	EAR4	14	0.613	*T. platyphyllos*
50	EAR5	16	0.613	*T. platyphyllos*
51	EAR6	11	0.622	*T. platyphyllos*
52	HBW01	2	-0.418	*T. cordata*
53	HBW02	3	-0.664	*T. cordata*
54	HBW03	2	-0.641	*T. cordata*
55	HBW04	3	-0.657	*T. cordata*
56	HBW05	1	-0.604	*T. cordata*
57	HBW06	1	-0.669	*T. cordata*
58	HBW07	3	-0.575	*T. cordata*
59	HBW08[c]	5	-0.419	*T. cordata*
60	HBW09[c]	3	-0.419	*T. cordata*
61	HBW10	7	-0.499	*T. cordata*
62	HBW11	2	-0.585	*T. cordata*
63	HBW12	1	-0.597	*T. cordata*
64	HBW13	1	-0.656	*T. cordata*
65	HBW14	3	-0.339	*T. cordata*
66	HBW15	2	-0.623	*T. cordata*
67	HBW16	3	-0.514	*T. cordata*
68	KIN01	9	0.603	*T. platyphyllos*
69	KIN02	7	0.662	*T. platyphyllos*
70	KIN03	10	0.569	*T. platyphyllos*
71	KIN04	15	0.513	*T. platyphyllos*
72	KIN05	10	0.517	*T. platyphyllos*
73	KIN06	11	0.646	*T. platyphyllos*
74	KIN07	12	0.552	*T. platyphyllos*
75	KIN08	9	0.706	*T. platyphyllos*
76	KIN09	12	0.637	*T. platyphyllos*
77	KIN10	9	0.648	*T. platyphyllos*
78	KIN11	10	0.658	*T. platyphyllos*

Appendix 11. Continued.

No.	Sample	Hybrid index	1st axis of PCO value	NewHybrids identification
79	KIN12	13	0.612	*T. platyphyllos*
80	KIN13	16	0.524	*T. platyphyllos*
81	KIN14	16	0.538	*T. platyphyllos*
82	KIN15	15	0.601	*T. platyphyllos*
83	KIN16	12	0.504	*T. platyphyllos*
84	KIN18	15	0.514	*T. platyphyllos*
85	KIN19	14	0.684	*T. platyphyllos*
86	KPM03	12	0.684	*T. platyphyllos*
87	KPM05[d]	7	-0.018	F_1
88	KPM06[d]	6	-0.018	F_1
89	KPM07	3	-0.586	*T. cordata*
90	KPM08	5	-0.665	*T. cordata*
91	KPM09	4	-0.587	*T. cordata*
92	KPM11	11	0.699	*T. platyphyllos*
93	KPM13	3	-0.694	*T. cordata*
94	KPM14	3	-0.665	*T. cordata*
95	KPM15	0	-0.333	*T. cordata*
96	KPM17	4	-0.534	*T. cordata*
97	KPM18	5	-0.551	*T. cordata*
98	KPM19	3	-0.571	*T. cordata*
99	LPW01	3	-0.290	*T. cordata*
100	LPW02	2	-0.590	*T. cordata*
101	LPW03	3	-0.433	*T. cordata*
102	LPW04	3	-0.458	*T. cordata*
103	LPW05	5	-0.526	*T. cordata*
104	LPW06	3	-0.520	*T. cordata*
105	LPW07	2	-0.565	*T. cordata*
106	LPW08	3	-0.507	*T. cordata*
107	LPW09	1	-0.547	*T. cordata*
108	LPW11	17	0.325	*T. platyphyllos*
109	LPW12	15	0.321	*T. platyphyllos*
110	LPW14	4	-0.182	*T. cordata*
111	LPW15	3	-0.576	*T. cordata*
112	LPW16	4	-0.648	*T. cordata*
113	LPW17	16	0.589	*T. platyphyllos*
114	LPW18	2	-0.466	*T. cordata*
115	LPW19	15	0.459	*T. platyphyllos*
116	ROU01	4	-0.538	*T. cordata*
117	ROU02	3	-0.388	*T. cordata*

Appendix 11. Continued.

No.	Sample	Hybrid index	1st axis of PCO value	NewHybrids identification
118	ROU03	3	-0.433	*T. cordata*
119	ROU04	2	-0.435	*T. cordata*
120	ROU06	3	-0.673	*T. cordata*
121	ROU07	5	-0.638	*T. cordata*
122	ROU08	3	-0.407	*T. cordata*
123	ROU09	6	-0.431	*T. cordata*
124	ROU10	4	-0.397	*T. cordata*
125	ROU11	6	-0.369	*T. cordata*
126	ROU12	4	-0.569	*T. cordata*
127	ROU13	4	-0.418	*T. cordata*
128	ROU14	4	-0.285	*T. cordata*
129	ROU15	5	-0.546	*T. cordata*
130	ROU16	6	-0.618	*T. cordata*
131	ROU17	6	-0.328	*T. cordata*
132	ROU18	6	-0.385	*T. cordata*
133	ROU19	3	-0.634	*T. cordata*
134	ROU20	3	-0.291	undetermined
135	TIC01	14	0.610	*T. platyphyllos*
136	TIC02[e]	9	0.004	F_1
137	TIC05[e]	7	0.004	F_1
138	TIC06	9	0.682	*T. platyphyllos*
139	TIC07	11	0.593	*T. platyphyllos*
140	TIC11	11	0.633	*T. platyphyllos*
141	TIC12	14	0.535	*T. platyphyllos*
142	TIC13	10	0.559	*T. platyphyllos*
143	TIC15	10	0.529	*T. platyphyllos*
144	TIC16	11	0.649	*T. platyphyllos*

Note: [a-e] indicated five clones detected in four populations.

References

Abbott RJ (1992). Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends in Ecology & Evolution* **7**(12)**:** 401-405.

Abrams MD (1994). Genotypic and phenotypic variation as stress adaptations in temperate tree species: a review of several case studies. *Tree Physiology* **14**(7-9)**:** 833-842.

Achtak H, Oukabli A, Ater M, Santoni S, Kjellberg F, Khadari B (2009). Microsatellite Markers as Reliable Tools for Fig Cultivar Identification. *Journal of the American Society for Horticultural Science* **134**(6)**:** 624-631.

Alvarez I, Wendel JF (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**(3)**:** 417-434.

Alverson WS, Karol KG, Baum DA, Chase MW, Swensen SM, McCourt R *et al* (1998). Circumscription of the Malvales and relationships to other Rosidae: Evidence from *rbcL* sequence data. *American Journal of Botany* **85**(6)**:** 876-887.

Anderson E (1949). *Introgressive hybridization*. John Wiley and Sons, Inc. New York, USA.

Anderson EC, Thompson EA (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**(3)**:** 1217-1229.

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

Arnold ML (1997). *Natural hybridization and evolution*. Oxford University Press: New York, USA.

Arnold ML, Bouck AC, Cornman RS (2004). Verne Grant and Louisiana Irises: Is there anything new under the sun? *New Phytologist* **161**(1)**:** 143-149.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE *et al* (1987). Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics* **18:** 489-522.

Bacilieri R, Ducousso A, Petit RJ, Kremer A (1996). Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution* **50**(2)**:** 900-908.

Bai WN, Liao WJ, Zhang DY (2010). Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. *New Phytologist* **188**(3)**:** 892-901.

Baldwin BG (1992). Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the compositae. *Molecular Phylogenetics and Evolution* **1**(1)**:** 3-16.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995). The ITS region of nuclear ribosomal DNA - a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* **82**(2)**:** 247-277.

Bandelt HJ, Forster P, Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**(1)**:** 37-48.

Barbara T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C (2007). Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology* **16**(18)**:** 3759-3767.

Bennett KD, Tzedakis PC, Willis KJ (1991). Quaternary refugia of the North Europeans trees. *Journal of Biogeography* **18**(1)**:** 103-115.

Birks HJB (1986). Late Quaternary biotic changes in terrestrial and lacustrine environments with particular reference to north-west Europe. In: Berglund BE (ed) *Handbook of Holocene Palaeoecology and Palaeohydrology*. J.W. Wiley & Sons Ltd.: London, pp 3-65.

Boccacci P, Akkak A, Marinoni DT, Bounous G, Botta R (2004). Typing European chestnut (*Castanea sativa* Mill.) cultivars using oak simple sequence repeat markers. *Hortscience* **39**(6)**:** 1212-1216.

Borrone JW, Meerow AW, Kuhn DN, Whitlock BA, Schnell RJ (2007). The potential of the *WRKY* gene family for phylogenetic reconstruction: An example from the Malvaceae. *Molecular Phylogenetics and Evolution* **44**(3)**:** 1141-1154.

Bradley P (1992). *British Herbal Compendium*, Vol 1: British Herbal Medicine Association, Dorset, Great Britain.

Bremer B, Bremer K, Chase MW, Reveal JL, Soltis DE, Soltis PS *et al* (2003). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* **141**(4)**:** 399-436.

Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B (1998). Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat. *American Journal of Human Genetics* **62**(6)**:** 1408-1415.

Broome A, Quine C, Trout R, Poulsom E, Mayle B. (2004). Forestry commision: Forestry commision, UK.

Brown WL (1983). Genetic diversity and genetic vulnerability - an appraisal *Economic Botany* **37**(1)**:** 4-12.

Buckler ES, Ippolito A, Holtsford TP (1997). The evolution of ribosomal DNA: Divergent paralogues and phylogenetic implications. *Genetics* **145**(3)**:** 821-832.

Calonje M, Martin-Bravo S, Dobes C, Gong W, Jordon-Thaden I, Kiefer C *et al* (2009). Non-coding nuclear DNA markers in phylogenetic reconstruction. *Plant Systematics and Evolution* **282**(3-4)**:** 257-280.

Chambers TC, Godwin H (1971). Scanning electron microscopy of *Tilia* pollen. *New Phytologist* **70**(4)**:** 687-692.

Charlesworth B, Nordborg M, Charlesworth D (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research* **70**(2)**:** 155-174.

Chase M, Kesseli R, Bawa K (1996). Microsatellite markers for population and conservation genetics of tropical trees. *American Journal of Botany* **83**(1)**:** 51-57.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD *et al* (1993). Phylogenetics of seed plants - An analysis of nucleotide-sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* **80**(3)**:** 528-580.

Coote T, Bruford MW (1996). Human microsatellites applicable for analysis of genetic variation in apes and old world monkeys. *Journal of Heredity* **87**(5)**:** 406-410.

Craft KJ, Ashley MV (2006). Population differentiation among three species of white oak in northeastern Illinois. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **36**(1)**:** 206-215.

Craft KJ, Ashley MV, Koenig WD (2002). Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. *American Journal of Botany* **89**(11)**:** 1792-1798.

Crawford KM, Whitney KD (2010). Population genetic diversity influences colonization success. *Molecular Ecology* **19**(6)**:** 1253-1263.

Crawford NG (2010). smogd: software for the measurement of genetic diversity. *Molecular Ecology Resources* **10**(3)**:** 556-557.

Cronn RC, Small RL, Haselkorn T, Wendel JF (2002). Rapid diversification of the cotton genus (*Gossypium* : Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany* **89**(4)**:** 707-725.

Cuénoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW (2002). Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *American Journal of Botany* **89**(1)**:** 132-144.

Curtu AL, Gailing O, Finkeldey R (2007). Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evolutionary Biology* **7**: 218.

Dangl GS, Yang J, Golino DA, Gradziel T (2009). A practical method for almond cultivar identification and parental analysis using simple sequence repeat markers. *Euphytica* **168**(1)**:** 41-48.

Davis CS, Gelatt TS, Siniff D, Strobeck A (2002). Dinucleotide microsatellite markers from the Antarctic seals and their use in other Pinnipeds. *Molecular Ecology Notes* **2**(3)**:** 203-208.

Demesure B, Comps B, Petit RJ (1996). Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L) in Europe. *Evolution* **50**(6)**:** 2515-2520.

Dlugosch KM, Parker IM (2008). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* **17**(1)**:** 431-449.

Drummond AJ, Ashton B, Buxton S, Cheung M, Coope A, Duran C *et al*. (2011). *Geneious* (Version 6.1.4) [Computer program]. Available at: http://www.geneious.com/.

Duminil J, Di Michele M (2009). Plant species delimitation: A comparison of morphological and molecular markers. *Plant Biosystems* **143**(3)**:** 528-542.

Dumolin-Lapegue S, Demesure B, Fineschi S, LeCorre V, Petit RJ (1997a). Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**(4)**:** 1475-1487.

Dumolin-Lapegue S, Pemonge MH, Petit RJ (1997b). An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology* **6**(4)**:** 393-397.

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

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996). Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques* **20**(5)**:** 758-760.

Elder JF, Turner BJ (1995). Concerted evolution of repetitive DNA-sequences in eukaryotes. *Quarterly Review of Biology* **70**(3)**:** 297-320.

Ellegren H (1991). DNA typing of museum birds. *Nature* **354**(6349)**:** 113-113.

Ellegren H (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics* **5**(6)**:** 435-445.

Ellegren H, Moore S, Robinson N, Byrne K, Ward W, Sheldon BC (1997). Microsatellite evolution - A reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Molecular Biology and Evolution* **14**(8)**:** 854-860.

Ellegren H, Primmer CR, Sheldon BC (1995). Microsatellite evolution - directionality or bias. *Nature Genetics* **11**(4)**:** 360-362.

Elwes HJ, Henry A (1906). *The trees of Great Britain and Ireland*. Index: Edinburgh.

Excoffier L, Lischer HEL (2010). Arlequin suite version 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**(3)**:** 564-567.

FAO (2010). Global Forest Resources Assessment. *FAO forestry paper***:** 163.

Fay MF, Swensen SM, Chase MW (1997). Taxonomic affinities of *Medusagyne oppositifolia* (*Medusagynaceae*). *Kew Bulletin* **52**(1)**:** 111-120.

Fehrer J, Gemeinholzer B, Chrtek J, Brautigam S (2007). Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (*Hieracium*, Cichorieae, Asteraceae). *Molecular Phylogenetics and Evolution* **42**(2)**:** 347-361.

Felsenstein J (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**(4)**:** 783-791.

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

Fortini P, Viscosi V, Maiuro L, Fineschi S, Vendramin GG (2009). Comparative leaf surface morphology and molecular data of five oaks of the subgenus *Quercus* Oerst (Fagaceae). *Plant Biosystems* **143**(3)**:** 543-554.

Fromm M (2001). Reproduktion einer entomophilen Baumart in geringer Populationdichte -Das Beispiel der Winter Linde (*T. cordata* Mill.). Ph.D. thesis, Universität Göttingen.

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

Gaggiotti OE, Lange O, Rassmann K, Gliddon C (1999). A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology* **8**(9)**:** 1513-1520.

Gailing O, Lind J, Lilleskov E (2012). Leaf morphological and genetic differentiation between *Quercus rubra* L. and *Q. ellipsoidalis* E.J. Hill populations in contrasting environments. *Plant Systematics and Evolution* **298**(8)**:** 1533-1545.

GarciaRamos G, Kirkpatrick M (1997). Genetic models of adaptation and gene flow in peripheral populations. *Evolution* **51**(1)**:** 21-28.

Garzon MB, Alia R, Robson TM, Zavala MA (2011). Intra-specific variability and plasticity influence potential tree species distributions under climate change. *Global Ecology and Biogeography* **20**(5)**:** 766-778.

Gaut BS (1998). Molecular clocks and nucleotide substitution rates in higher plants. In: Hecht MK, Macintyre RJ and Clegg MT (eds) *Evolutionary Biology*. **30**: 93-120.

Gonzalez-Rodriguez A, Arias DM, Valencia S, Oyama A (2004). Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. *American Journal of Botany* **91**(3)**:** 401-409.

Goudet J (2002). *FSTAT, a program to estimate and test gene diversities and fixation indices* (Version 2.9.3.2) [Computer program]. Available at: http://www2.unil.ch/popgen/softwares/fstat.htm.

Grant V (1981). *Plant speciation*. Columbia University Press: New York, USA.

Grivet D, Petit RJ (2003). Chloroplast DNA phylogeography of the hornbeam in Europe: Evidence for a bottleneck at the outset of postglacial colonization. *Conservation Genetics* **4**(1)**:** 47-56.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Leger P, Lepais O *et al* (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**(4)**:** 591-611.

Hamilton MB (1999). Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* **8**(3)**:** 521-523.

Harrison RG (1993). *Hybrid Zones and the Evolutionary Process*. Oxford University Press: New York, USA.

Hemery GE, Clark JR, Aldinger E, Claessens H, Malvolti ME, O'Connor E *et al* (2010). Growing scattered broadleaved tree species in Europe in a changing climate: a review of risks and opportunities. *Forestry* **83**(1)**:** 65-81.

Heuertz M, Carnevale S, Fineschi S, Sebastiani F, Hausman JF, Paule L *et al* (2006). Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (Oleaceae): roles of hybridization and life history traits. *Molecular Ecology* **15**(8)**:** 2131-2140.

Heuertz M, Fineschi S, Anzidei M, Pastorelli R, Salvini D, Paule L *et al* (2004). Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in Europe. *Molecular Ecology* **13**(11)**:** 3437-3452.

Hewitt GM (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**(1-2)**:** 87-112.

Hewitt GM (2000). The genetic legacy of the Quaternary ice ages. *Nature* **405**(6789)**:** 907-913.

Hewitt GM, Butlin RK (1997). Causes and consequences of population structure. In: Krebs JR and Davies NB (eds) *Behavioural Ecology: An Evolutionary Approach*. Blackwell publishing company: Oxford, pp 350-372.

Hilu KW, Liang HP (1997). The matK gene: Sequence variation and application in plant systematics. *American Journal of Botany* **84**(6)**:** 830-839.

Hogan CM (2012). *Encyclopedia of Earth*. National Council for Science and the Environment: Washington DC.

Hosseinzadeh Colagar A, Yusefi M, Zarei M, 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)**:** 341-348.

Howard DJ, Preszler RW, Williams J, Fenchel S, Boecklen WJ (1997). How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution* **51**(3)**:** 747-755.

Huntley B, Birks HJB (1983). *An atlas of past and present pollen maps for Europe: 0–13,000 B.P*. Cambridge University Press: Cambridge, United Kingdom.

Hutter CM, Schug MD, Aquadro CF (1998). Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: A reciprocal test of the ascertainment bias hypothesis. *Molecular Biology and Evolution* **15**(12)**:** 1620-1636.

Ishida Y, Roca AL, Fratpietro S, Greenwood AD (2012). Successful Genotyping of Microsatellites in the Woolly Mammoth. *Journal of Heredity* **103**(3)**:** 459-464.

Jarne P, Lagoda PJL (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* **11**(10)**:** 424-429.

Jensen J, Larsen A, Nielsen LR, Cottrell J (2009). Hybridization between *Quercus robur* and *Q. petraea* in a mixed oak stand in Denmark. *Annals of Forest Science* **66**(7): 706.

Jones AG, Small CM, Paczolt KA, Ratterman NL (2010). A practical guide to methods of parentage analysis. *Molecular Ecology Resources* **10**(1)**:** 6-30.

Kandpal RP, Kandpal G, Weissman SM (1994). Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. *Proceedings of the National Academy of Sciences of the United States of America* **91**(1)**:** 88-92.

Kang JH, Park JY, Jo HS (2012). Rapid Development of Microsatellite Markers with 454 Pyrosequencing in a Vulnerable Fish, the Mottled Skate, *Raja pulchra*. *International Journal of Molecular Sciences* **13**(6)**:** 7199-7211.

Khela S (2013) *Tilia platyphyllos. In: IUCN 2013. IUCN Red List of Threatened Species. Version 2013.2.* Available at: www.iucnredlist.org (Accessed: 1 May 2014).

Kiaer LP, Philipp M, Jorgensen RB, Hauser TP (2007). Genealogy, morphology and fitness of spontaneous hybrids between wild and cultivated chicory (Cichorium intybus). *Heredity* **99**(1)**:** 112-120.

King RA, Ferris C (1998). Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology* **7**(9)**:** 1151-1161.

King RA, Ferris C (2000). Chloroplast DNA and nuclear DNA variation in the sympatric alder species, *Alnus cordata* (Lois.) Duby and *A. glutinosa* (L.) Gaertn. *Biological Journal of the Linnean Society* **70**(1)**:** 147-160.

Konnert M, Bergmann F (1995). The geographical distribution of genetic variation of silver fir (*Abies alba, Pinaceae*) in relation to its migration history. *Plant Systematics and Evolution* **196**(1-2)**:** 19-30.

Kremer A, Dupouey JL, Deans JD, Cottrell J, Csaikl U, Finkeldey R *et al* (2002). Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science* **59**(7)**:** 777-787.

Kress WJ, Erickson DL (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcL* Gene Complements the Non-Coding *trnH-psbA* Spacer Region. *PloS ONE* **2**(6):e501.

Lang G (2003). *Immigration and expansion of Tilia in Europe since the last glacial*. Pensoft Publisher: Moscow.

Levin RA, Blanton J, Miller JS (2009). Phylogenetic utility of nuclear nitrate reductase: A multi-locus comparison of nuclear and chloroplast sequence data for inference of relationships among American Lycieae (Solanaceae). *Molecular Phylogenetics and Evolution* **50**(3)**:** 608-617.

Li B, Kimmel M (2013). Factors influencing ascertainment bias of microsatellite allele sizes: Impact on estimates of mutation rates. *Genetics* **195**(2)**:** 563-572.

Li J, Chen Y, Pigott CD (2002). *Phylogenetic relationships of diploid species of Tilia inferred from sequences of nuclear gene NIA*. Unpublished poster.

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

Linhart YB, Grant MC (1996). Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics* **27:** 237-277.

Loria MC (1967). Sistematicheskii obzor lip Kavkaza. *Botanicheskii Zhurnal SSSR* **52:** 1789-1791.

Loveless MD, Hamrick JL (1984). Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics* **15:** 65-95.

Lowe A, Harris S, Ashton P (2004). *Ecological Genetics: Design, Analysis, and Application*. Blackwell Publishing.

Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A (2005). *MINISEED3* (*MINI3*), a *WRKY f*amily gene, and *HAIKU2* (*IKU2*), a leucine-rich repeat (*LRR*) *KINASE* gene, are regulators of seed size in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **102**(48)**:** 17531-17536.

Maddison WP, Maddison DR (2001). *Mesquite: a modular system for evolutionary analysis* (Version 2.75) [Computer program]. Available at: http://mesquiteproject.org.

Mallet J (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**(5)**:** 229-237.

Marriage TN, Hudman S, Mort ME, Orive ME, Shaw RG, Kelly JK (2009). Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). *Heredity* **103**(4)**:** 310-317.

Maurer WD, Tabel U (1995). A methodical study to improve the isozyme analysis for identification of clones of *Tilia* (Linden syn lime tree). *Silvae Genetica* **44**(5-6)**:** 351-356.

McCarty D (2012). Systematic and phylogeography of the genus Tilia in north America. Ph.D. thesis, University of Illinois at Chicago, University of Illinois at Chicago, Chicago, Illinois, USA.

McEwen JR, Vamosi JC, Rogers SM (2011). Rapid isolation and cross-amplification of microsatellite markers in *Plectritis congesta* (Valerianaceae) with 454 sequencing. *American Journal of Botany* **98**(12)**:** E369-E371.

Meirmans PG, Hedrick PW (2011). Assessing population structure: F_{st} and related measures. *Molecular Ecology Resources* **11**(1)**:** 5-18.

Moran EV, Willis J, Clark JS (2012). Genetic evidence for hybridization in red oaks (*Quercus* sect. *Lobatae*, Fagaceae). *American Journal of Botany* **99**(1)**:** 92-100.

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

Morgante M, Rafalski A, Biddle P, Tingey S, Olivieri AM (1994). Genetic mapping and variability of seven soybean simple sequence repeat loci. *Genome* **37**(5)**:** 763-769.

Mort ME, Archibald JK, Randle CP, Levsen ND, O'Leary TR, Topalov K *et al* (2007). Inferring phylogeny at low taxonomic levels: Utility of rapidly evolving cpDNA and nuclear ITS loci. *American Journal of Botany* **94**(2)**:** 173-183.

Muir G, Schlotterer C (2005). Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology* **14**(2)**:** 549-561.

Neophytou C, Aravanopoulos FA, Fink S, Dounavi A (2010). Detecting interspecific and geographic differentiation patterns in two interfertile oak species (*Quercus petraea* (Matt.) Liebl. and *Q. robur* L.) using small sets of microsatellite markers. *Forest Ecology and Management* **259**(10)**:** 2026-2035.

Oldfield S, Lusty C, MacKinven A (1998). *The World List of Threatened Trees*. World Conservation Press: Cambridge, UK.

Palme AE, Su Q, Palsson S, Lascoux M (2004). Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. *Molecular Ecology* **13**(1)**:** 167-178.

Palme AE, Vendramin GG (2002). Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, *Corylus avellana*. *Molecular Ecology* **11**(9)**:** 1769-1780.

Park YJ, Lee JK, Kim NS (2009). Simple Sequence Repeat Polymorphisms (SSRPs) for Evaluation of Molecular Diversity and Germplasm Classification of Minor Crops. *Molecules* **14**(11)**:** 4546-4569.

Patzak J, Paprstein F, Henychova A, Sedlak J (2012). Genetic diversity of Czech apple cultivars inferred from microsatellite markers analysis. *Horticultural Science* **39**(4)**:** 149-157.

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

Penaloza-Ramirez JM, Gonzalez-Rodriguez A, Mendoza-Cuenca L, Caron H, Kremer A, Oyama K (2010). Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. *Annals of Botany* **105**(3)**:** 389-399.

Petit RJ, Brewer S, Bordacs S, Burg K, Cheddadi R, Coart E *et al* (2002). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* **156**(1-3)**:** 49-74.

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

Pigott CD (1969). The status of *Tilia cordata* and *T. platyphyllos* on Derbyshire limestones. *Journal of Ecology* **57**(2)**:** 491-504.

Pigott CD (1991). Biological Flora of the British Isles No 174 *Tilia cordata* Miller. *Journal of Ecology* **79:** 1147-1207.

Pigott CD (2002). A review of chromosome numbers in the genus *Tilia* (Tiliaceae). *Edinburgh Journal of Botany* **59**(2)**:** 239–246.

Pigott CD (2006). *Tilia endochrysea*. *Curtis's Botanical Magazine* **23**: 56-61.

Pigott CD (2012). *Lime-trees and Basswoods : a biological monograph of the genus Tilia*. Cambridge University Press: New York. USA.

Pigott CD, Huntley JP (1978). Factors comtrolling the distribution of *Tlia cordata* at the northern limits of its geographical range. *New Phytologist* **81:** 429-441.

Pons O, Petit RJ (1995). Estimation, variance and optimal sampling of gene diversity. *Theoretical and Applied Genetics* **90**(3-4)**:** 462-470.

Primmer CR, Merila J (2002). A low rate of cross-species microsatellite amplification success in Ranid frogs. *Conservation Genetics* **3**(4)**:** 445-449.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**(2)**:** 945-959.

Provan J, Powell W, Hollingsworth PM (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* **16**(3)**:** 142-147.

Rao VR, Hodgkin T (2002). Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell, Tissue and Organ Culture* **68:** 1-19.

Raymond M, Rousset F (1995). GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**(3)**:** 248-249.

Rekik I, Salimonti A, Kamoun NG, Muzzalupo I, Lepais O, Gerber S *et al* (2008). Characterization and identification of Tunisian olive tree varieties by microsatellite markers. *Hortscience* **43**(5)**:** 1371-1376.

Rendell S, Ennos RA (2002). Chloroplast DNA diversity in *Calluna vulgaris* (heather) populations in Europe. *Molecular Ecology* **11**(1)**:** 69-78.

Rendell S, Ennos RA (2003). Chloroplast DNA diversity of the dioecious European tree *Ilex aquifolium* L. (English holly). *Molecular Ecology* **12**(10)**:** 2681-2688.

Rieseberg LH (1991). Homoploid Reticulate Evolution in Helianthus (Asteraceae): Evidence from Ribosomal Genes. *American Journal of Botany* **78**(9)**:** 1218-1237.

Rieseberg LH (1995). The Role of Hybridization in Evolution: Old Wine in New Skins. *American Journal of Botany* **82**(7)**:** 944-953.

Rieseberg LH, Carney SE (1998). Plant hybridization. *New Phytologist* **140**(4)**:** 599-624.

Rieseberg LH, Ellstrand NC (1993). What can molecular and morphological markers tell us about plant hybridization? *Critical Reviews in Plant Sciences* **12**(3)**:** 213-241.

Rieseberg LH, Wendel JF (1993). *Introgression and its consequences in plants*. Oxford University Press: New York.

Rodriguez F, Spooner DM (2009). Nitrate reductase phylogeny of potato (*Solanum* sect. *Petota*) genomes with emphasis on the origins of the polyploid species. *Systematic Botany* **34**(1)**:** 207-219.

Rohling EJ, Fenton M, Jorissen FJ, Bertrand P, Ganssen G, Caulet JP (1998). Magnitudes of sea-level lowstands of the past 500,000 years. *Nature* **394**(6689)**:** 162-165.

Rouhier N, Couturier J, Jacquot JP (2006). Genome-wide analysis of plant glutaredoxin systems. *Journal of Experimental Botany* **57**(8)**:** 1685-1696.

Royal Forestry Society (2014). *Limes*. Available at: http://www.rfs.org.uk/learning/limes (Accessed: 1 May 2014).

Rozen S, Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology (Clifton, NJ)* **132:** 365-386.

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH *et al* (1995). Microsatellite evolution-evidence for directionality and variation in rate between species. *Nature Genetics* **10**(3)**:** 337-343.

Saeki I, Dick CW, Barnes BV, Murakami N (2011). Comparative phylogeography of red maple (*Acer rubrum* L.) and silver maple (*Acer saccharinum* L.): impacts of habitat specialization, hybridization and glacial history. *Journal of Biogeography* **38**(5)**:** 992-1005.

Salanoubat M, Ha DBD (1993). Analysis of the petunia nitrate reductase apoenzyme-encoding gene - a first step for sequence modification analysis. *Gene* **128**(2)**:** 147-154.

Sang T (2002). Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* **37**(3)**:** 121-147.

Schlotterer C (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**(6)**:** 365-371.

Schlotterer C, Amos B, Tautz D (1991). Conservation of polymorphic simple sequence loci in cetacean species. *Nature* **354**(6348)**:** 63-65.

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF (1998). The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution* **15**(12)**:** 1751-1760.

Selkoe KA, Toonen RJ (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**(5)**:** 615-629.

Setsuko S, Uchiyama K, Sugai K, Yoshimaru H (2012). Rapid development of microsatellite markers for *Pandanus boninensis* (Pandanaceae) by pyrosequencing technology. *American Journal of Botany* **99**(1)**:** E33-E37.

Shaw J, Lickey EB, Schilling EE, Small RL (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *American Journal of Botany* **94**(3)**:** 275-288.

Shaw J, Small RL (2004). Addressing the "hardest puzzle in American pomology": Phylogeny of *Prunus* sect. *Prunocerasus* (Rosaceae) based on seven noncoding chloroplast DNA regions. *American Journal of Botany* **91**(6)**:** 985-996.

Small RL, Cronn RC, Wendel JF (2004). Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* **17**(2)**:** 145-170.

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998). The tortoise and the hare: Choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* **85**(9)**:** 1301-1315.

Smith SA, Donoghue MJ (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science* **322**(5898)**:** 86-89.

Soltis PS, Soltis DE, Chase MW (1999). Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**(6760)**:** 402-404.

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M *et al* (2003). How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology* **12**(6)**:** 1339-1348.

Squirrell J, Wolff K (2001). Isolation of polymorphic microsatellite loci in *Plantago major* and *P. intermedia*. *Molecular Ecology Notes* **1**(3)**:** 179-181.

Stebbins GL (1959). The role of hybridization in evolution. *Proceedings of the American Philosophical Society* **103:** 231-251.

Steele PR, Guisinger-Bellian M, Linder CR, Jansen RK (2008). Phylogenetic utility of 141 low-copy nuclear regions in taxa at different taxonomic levels in two distantly related families of rosids. *Molecular Phylogenetics and Evolution* **48**(3)**:** 1013-1026.

Storz JF (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**(3)**:** 671-688.

Svejgaard Jensen J. (2003). *Vol. 2014*: International Plant Genetic Resources Institute, Rome, Italy.

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

Swofford DL (2003). *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) Version 4.0b10*. Sinauer: Sunderland, Massachusetts, USA.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**(4)**:** 453-464.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**(5)**:** 1105-1109.

Tang DQ, Lu JJ, Fang W, Zhang S, Zhou MB (2010). Development, characterization and utilization of GenBank microsatellite markers in *Phyllostachys pubescens* and related species. *Molecular Breeding* **25**(2)**:** 299-311.

Tautz D (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**(16)**:** 6463-6471.

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

Thorsson A, Palsson SP, Sigurgeirsson A, Anamthawat-Jonsson K (2007). Morphological variation among Betula nana (diploid), B-pubescens (tetraploid) and their triploid hybrids in Iceland. *Annals of Botany* **99**(6)**:** 1183-1193.

Tovar-Sanchez E, Oyama K (2004). Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: Morphological and molecular evidence. *American Journal of Botany* **91**(9)**:** 1352-1363.

Traiser C, Klotz S, Uhl D, Mosbrugger V (2005). Environmental signals from leaves - a physiognomic analysis of European vegetation. *New Phytologist* **166**(2)**:** 465-484.

Truong C, Palme AE, Felber F (2007). Recent invasion of the mountain birch *Betula pubescens* ssp. *tortuosa* above the treeline due to climate change: genetic and ecological study in northern Sweden. *Journal of Evolutionary Biology* **20**(1)**:** 369-380.

Tsai Y-HE (2011). PhyloGeoViz: a web-based program that visualizes genetic data on maps. *Molecular Ecology Resources* **11**(3)**:** 557-561.

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

Vijayan K, Tsou CH (2010). DNA barcoding in plants: taxonomy in a new perspective. *Current Science* **99**(11)**:** 1530-1541.

Vowles EJ, Amos W (2006). Quantifying ascertainment bias and species-specific length differences in human and chimpanzee microsatellites using genome sequences. *Molecular Biology and Evolution* **23**(3)**:** 598-607.

Wang HX, Walla JA, Zhong SB, Huang DQ, Dai WH (2012). Development and cross-species/genera transferability of microsatellite markers discovered using 454 genome sequencing in chokecherry (*Prunus virginiana* L.). *Plant Cell Reports* **31**(11)**:** 2047-2055.

Weising K, Nybom H, Wolff K, Kahl G (2005). *DNAFingerprinting in Plants: Principles, Methods and Applications*. CRC Press.

Weryszko-Chmielewska E, Sadowska DA (2010). The phenology of flowering and pollen release in four species of linden (*Tilia* L.). *Journal of Apicultural Science* **54**(2)**:** 99-108.

White TJ, Bruns T, Lee S, Taylor J (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ and White TJ (eds) *PCR Protocols: A Guide to Methods and Applications*. Academic Press, Inc.: New York., pp 315-322.

Whitlock MC (2011). G $'_{st}$ and D do not replace F_{st}. *Molecular Ecology* **20**(6)**:** 1083-1091.

Whittaker RH (1975). *Communities and ecosystems*. Macmillian: London, UK.

Wicksell U, Christensen KI (1999). Hybridization among *Tilia cordata* and *T. platyphyllos* (Tiliaceae) in Denmark. *Nordic Journal of Botany* **19**(6)**:** 673-684.

William R, Catton J (2009). *Bottleneck: Humanity's Impending Impasse*. Xlibris Corporation.

Yang LH, Rudolf VHW (2010). Phenology, ontogeny and the effects of climate change on the timing of species interactions. *Ecology Letters* **13**(1)**:** 1-10.

Yi T, Wen J, Golan-Goldhirsh A, Parfitt DE (2008). Phylogenetics and reticulate evolution in *Pistacia* (Anacardiaceae). *American Journal of Botany* **95**(2)**:** 241-251.

Young A, Boyle T, Brown T (1996). The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution* **11**(10)**:** 413-418.

Yousefzadeh H, Colagar AH, Tabari M, Sattarian A, 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)**:** 947-961.

Zhang N, Zeng LP, Shan HY, Ma H (2012). Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytologist* **195**(4)**:** 923-937.

Zhang ZY, Wu R, Wang Q, Zhang ZR, Lopez-Pujol J, Fan DM *et al* (2013). Comparative phylogeography of two sympatric beeches in subtropical China: Species-specific geographic mosaic of lineages. *Ecology and Evolution* **3**(13)**:** 4461-4472.

Zhou JZ, Kleinhofs A (1996). Molecular evolution of nitrate reductase genes. *Journal of Molecular Evolution* **42**(4)**:** 432-442.

Zwickl DJ (2006). Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. thesis, The University of Texas at Austin.