Ecological, genetic and metabolic variation in populations of *Tilia cordata*

Carl Barker

This thesis is submitted Edge Hill University in partial fulfilment for the degree of Doctor of Philosophy

November 2017

Contents

Abstract **3**

Chapter 1: Introduction **4**

Chapter 2: Segregation of *Tilia cordata* Mill. and *T. platyphyllos* Scop. (Malvaceae) along environmental gradients **11**

Chapter 3: Reproductive strategy of *Tilia cordata* Mill. across its UK range **24**

Chapter 4: Cryptic spatial genetic structure of a canopy tree (*Tilia cordata*) across a fragmented landscape **57**

Chapter 5: Exploring metabolic variation in natural populations of a temperate canopy tree, *Tilia cordata* **85**

Chapter 6: Conclusions **122**

Acknowledgements **128**

References **129**

Appendices **163**

Abstract

Predicting the responses of tree species to rapid environmental change requires an understanding of their ecology, reproductive strategy, population connectivity and levels of adaptive variation. This project examines these aspects for an understudied UK native tree genus *Tilia* L. Comparison of edaphic and physiographic variables indicated that *Tilia cordata* Mill. is more generalist than *T. platyphyllos* Scop., as well as preferring locations with higher potential incident solar radiation and greater levels of organic carbon content. Examination of fine-scale spatial genetic structure indicates that *T. cordata* has a mixed reproductive system with approximately half of all individuals within sampled populations being of clonal origin. The incidence of clonality was weakly negatively correlated to historic summer temperatures and positively so to the proportion of canopy trees within samples, suggesting both fertility limitations and time since disturbance affect vegetative growth in the species. Clonal reproduction is not expected to impact future outcrossing success due to the fine scale of its effects on spatial genetic structure, being much smaller than typical pollen movements associated with outcrossing. *T. cordata* populations exhibited weak clinal spatial genetic structure at coarser scales (tens of kilometres) across two locations, which likely reflect historic dispersal limitations across a contiguous landscape and effective pollen movement at scales less than two kilometres. Fragmentation has not yet eroded genetic variation except in the demes with the smallest size (not necessarily the most isolated), suggesting that larger fragments may ultimately suffer the same fate. Finally, low metabolic variation between UK populations of *T. cordata* despite contrasting environmental conditions during sampling indicates high levels of phenotypic plasticity, while variation in a functional trait and a group of unidentified metabolite concentrations suggest avenues for examining local adaptation in the future.

Keywords

Ecology; molecular ecology; *Tilia*; niche segregation; clonality; genetic variation; fragmentation; connectivity; metabolic variation; functional traits

1 Introduction

1.1 Forest ecosystems

Globally forest habitats are extremely important reservoirs of biodiversity, provide a multitude of ecosystem services essential to human wellbeing and are a central part of Earth's biogeochemical systems (Shvidenko *et al.*, 2005). Their biodiversity is extremely high, particularly in the tropics, and they provide a habitat for over half of all known terrestrial plant and animal species (Betts *et al.*, 2017). They provide economic benefits in the form of wood and numerous other products, social benefits such as the provision of recreational opportunities and they are important to the conservation of soil and water resources (Hemery *et al.*, 2010). For example a reduction in forest cover reduces water quality, and as most of the freshwater utilized by humans comes from forest catchments the maintenance of these ecosystems is critical; further, loss of tree cover increases the effects of related natural hazards such as soil erosion or the impact of floods and landslides (Kasran and Nik, 1994). They play a significant role in the global carbon cycle, containing approximately 50% of the world's terrestrial organic carbon stocks; forests account for ~80% of terrestrial biomass (Shvidenko *et al.*, 2005). As a result the total amount of carbon contained in forest vegetation alone (i.e. excluding soil) is approximately 359 x 10^9 tonnes (IPCC, 2000), which makes it important as a carbon sink to help mitigate anthropogenic climate change.

Despite their importance, they are at risk due to a variety of factors. They are subject to pressure from habitat fragmentation, degradation and climatic change, which makes their future uncertain. For instance, the amount of tropical forest is declining rapidly due to deforestation for logging and conversion to agricultural land (DeFries *et al.*, 2002; Lambin, Geist and Lepers, 2003). Other areas are faring better in this regard, with coverage of temperate and boreal forest ecosystems stable and in some areas increasing due to natural reforestation and expansion (Shvidenko and Nilsson, 2002). This does not mean that they are not at risk. European forests for example have lost 50-70% of their maximum total area achieved after the last glacial maximum, primarily during the Middle Ages, and as a result many forest ecosystems are highly fragmented. This has implications for population persistence due to a reduction in effective size, gene flow and the concomitant lowering of genetic diversity (Lande, 1988; Aguilar *et al.*, 2008). Habitat degradation is also an issue, with other pressures such as atmospheric pollution and eutrophication impacting European

forests, affecting individual tree health (Nellemann and Thomsen, 2001; Shvidenko *et al.*, 2005; Michel and Seidling, 2016). Finally, the current rapid anthropogenic increases in atmospheric CO_2 are resulting in environmental change at an unprecedented pace (Thomas *et al.*, 2004; IPCC, 2007). This has already had both individual and ecosystem-wide effects on trees and tree communities such as increased mortality and lowered growth rates as a result of climate change impacts related to drought (Jump, Hunt and Peñuelas, 2006; Allen *et al.*, 2010), as well as biome shifts (Peñuelas and Boada, 2003). Climate change impacts will be substantial and have been recognised as one of the largest forthcoming threats to biodiversity across most biomes (Shvidenko *et al.*, 2005; Thuiller *et al.*, 2005, 2008), but with significant uncertainty and complexity in both climatic and species' response projections (Pressey *et al.*, 2007).

1.2 Threats to temperate forests from climate change

As a result of human greenhouse gas emissions the climate is unequivocally warming, as evidenced by observations of global mean atmospheric and oceanic temperatures, widespread loss of ice and snow and a continual rise in sea levels (IPCC, 2007). Evidence from all continents and most oceans shows that regional climatic changes, particularly temperature increases, are already impacting upon natural systems globally (Parmesan and Yohe, 2003). Temperate forest ecosystems are no exception. Projected future impacts will be caused not just by increased temperature and CO_2 concentration but also large-scale stochastic events (Hemery *et al.*, 2010), as both abiotic and biotic disturbance regimes are altered by climate change (Lindner *et al.*, 2010).

The effects of increased temperature and elevated levels of atmospheric CO_2 depend on location, primarily because of the current degree of water-limitation. Broadly speaking forest ecosystems in more northern and western areas of Europe with an oceanic climate could see increases in productivity, and will suffer impacts primarily as a result of alterations to disturbance regimes (Lindner *et al.*, 2010). Continental and Mediterranean Europe forests will also be impacted by changes to disturbance but will also be affected by reduced summer precipitation and increased temperatures, which can decrease photosynthesis and lead to reduced plant growth and altered recruitment (Ogaya *et al.*, 2003; Lloret, Penuelas and Estiarte, 2004). As mentioned, changes to disturbance regimes will affect forest ecosystems. The main abiotic disturbances in Europe are drought, flooding, wind and fire, and all of these may be affected by climate change, potentially occurring with increased frequency and

intensity (Flannigan, Stocks and Wotton, 2000; Bréda and Badeau, 2008). Biotic stresses are also likely to be altered via changes to the population dynamics and distributions of pests and pathogens (Hlásny and Turčáni, 2009). This may result in a higher incidence of pest or disease outbreaks due to temperature changes promoting multivoltinism in insect herbivores (Baier, Pennerstorfer and Schopf, 2007) or climate changes facilitating the spread of fungal pathogens (Desprez-Loustau *et al.*, 2007). Pathogenicity of the latter and susceptibility to insect herbivory can also increase in drought-stressed trees (Desprez-Loustau *et al.*, 2006), which highlights the fact that these impacts do not operate in isolation and interactions between them complicate future projections of responses. As such it is important to have as full an understanding of the species in question as possible (Rabasa *et al.*, 2013).

Historic management practices in European forests have broadly favoured a small set of trees with high economic value such as *Picea abies* (spruce), *Fagus sylvatica* (beech) and *Quercus* spp. (oaks) (Hemery *et al.*, 2010). As a consequence these are well-studied with regard to potential impacts from climate change, and their vulnerability has been noted (e.g. Jump et al., 2006; Ogaya et al., 2003; Peñuelas and Boada, 2003; Rivas-Ubach et al., 2014). More scattered, less commercially important trees in European forests such as *Fraxinus* spp. (ashes), *Ulmus* spp. (elms) and *Tilia* spp. (limes) are relatively understudied and this lack of knowledge introduces considerable uncertainty in determining the impacts of climatic change (Hemery *et al.*, 2010; De Jaegere, Hein and Claessens, 2016). To respond to environmental changes, species in general can either disperse to track shifting climatic envelopes or persist, adapting via phenotypic plasticity and/or evolutionary changes (Hamrick, 2004; Alfaro *et al.*, 2014). There are various constraints on dispersal, as it requires sufficient space (both in terms of suitable physical areas and niche space in newly colonized habitats), and in a fragmented landscape such as that occupied by European forest ecosystems, habitat connectivity dictates the success or otherwise of seed or propagule movement (Sork and Smouse, 2006). Additionally given the unprecedented rate of current changes (Thomas *et al.*, 2004), extremely high migration rates will be required, beyond that exhibited by post-glacial dispersals (Zhu, Woodall and Clark, 2012; Corlett and Westcott, 2013). This makes local persistence the most realistic prospect for tree species. Evolutionary responses might not be possible given long generation times and the extreme rate of change (Christmas, Breed and Lowe, 2015). Natural selection also requires sufficient adaptive genetic variation to operate on, which could potentially be lacking in highly fragmented species such as the scattered examples given above (Hemery *et al.*, 2010). Trees are sessile,

long-lived organism and because of this life strategy they typically possess high levels of phenotypic plasticity in order to weather short-term environmental fluctuations. Whether this is sufficient to enable persistence in areas that will be subject to the most change (i.e. boreal or Mediterranean regions in Europe) is likely to depend on the species on question (e.g. Attorre et al., 2011).

1.3 The genus *Tilia*

One of the widespread but scattered tree genera discussed above is *Tilia* L. It is a genus of 20-30 broadleaved temperate tree species (Pigott, 2012), four of which are native to Europe: *Tilia cordata* Mill., *T. platyphyllos* Scop., *T. dasystyla* Steven and *T. tomentosa* Moench. *T. dasystyla* and *T. tomentosa* have restricted distributions while*T. cordata* and *T. platyphyllos* are more widespread. They possess broad but scattered distributions that are the product of postglacial dispersal followed by anthropogenic influences such as forest clearance and silvicultural practices favouring other taxa (Turner, 1962; Pigott and Huntley, 1980; Pigott, 1991). As a consequence of these human influences both are now relatively rare and occur infrequently, but historically their abundance was much greater, to the extent of being co-dominant in the temperate primeval woodlands of central and eastern Europe (De Jaegere, Hein and Claessens, 2016). *T. cordata* was also widespread as far north as the Lake District in the UK prior to human influence (Pigott and Huntley, 1980). They are more common in areas with a continental rather than oceanic climate (Pigott, 1991). *T. cordata* occurs primarily at low altitudes in the UK (<200 m) but can occur at greater elevations in more southern locations such as Italy and the Caucasus (Pigott, 1991; Pigott and Pigott, 1993). The northern limit of its range correlates broadly with the July isotherm for mean daily maximum temperature of 19 – 20 °C, while the southern and eastern limits are related to the isohyet of 500 – 550 mm mean annual rainfall. Summer drought is likely the determinant of its range limits in southern Europe (Pigott and Pigott, 1993). It occurs on a wide variety of soil types, typically slightly acidic to alkaline, while *T. platyphyllos* occurs primarily on basic soils overlaying limestone (Pigott, 1991). The latter occurs almost exclusively with *T. cordata* despite reports of it colonizing new areas (Pigott, 1981b), but whether the broad ecological preferences described in the literature translate into niche segregation where they occur in sympatry is unknown.

T. cordata exhibits reduced fertility in marginal regions such as the north of England and Finland (Pigott, 1981a; Pigott and Huntley, 1981). At least partly because of this, it has

not recolonized secondary woodland within the UK; however large crops of fertile seed are set in more southern populations (Pigott, 1991). *T. cordata* has low dispersal ability relative to other tree species. Firstly, seed dispersal is wind-assisted but occurs over short distances, typically less than 100 m (Pigott, 1991), while successful pollination events have a maximum recorded distance of ~1666 m (Fromm, 2001). Based on these figures and palynological records of postglacial dispersal (low relative to other European tree species; Birks, 1989), *T. cordata* is likely a poor disperser (De Jaegere, Hein and Claessens, 2016). Whether dispersal limitations or changes in connectivity between fragmented habitats have resulted in reduced gene flow between woodland fragments is unclear.

Information regarding the breeding system in *T. cordata* is contradictory. At least some proportion of individuals are reported to be self-incompatible (Pigott and Huntley, 1981), but a study using parentage analysis showed that all individuals were self-compatible but that selfed progeny have lower survivorship up to the fourth year than do outcrossed individuals (Fromm, 2001). The species is also partially clonal and vegetatively propagates, although the extent and arrangement of clones in natural populations has not been explored. Given that clonality has the potential to reduce future outcrossing success, lower genetic diversity and increase inbreeding (Charpentier, 2001; Vallejo-Marín, Dorken and Barrett, 2010), and inversely to assist in population persistence under conditions unfavourable for sexual reproduction (Silvertown, 2008), this is an important area to address. Finally although many populations of tree species are locally adapted to regional environmental optima (Alfaro *et al.*, 2014), is it unknown whether this is the case for *T. cordata.* Pigott (1991) reports that there is some evidence of variation in physiological attributes, but no provenance trials have been undertaken. How populations vary if at all across the range of the species is therefore unclear.

Existing speculation about the responses of *T. cordata* populations to climate change are based primarily on its characteristics as described broadly in the literature (De Jaegere, Hein and Claessens, 2016), and as such considerable uncertainty remains (Hemery *et al.*, 2010). For instance it seems likely that populations in temperature-limited marginal regions may benefit from increased fertility due to greater warmth during flowering periods or seed development (Pigott, 1981a; Pigott and Huntley, 1981), but conversely these northern areas reportedly persist due to clonality alone (Pigott and Huntley, 1978). As such genotypic diversity may be reduced and therefore its vulnerability increased to any changes in the frequency, severity or character of plant pathogen or pest outbreaks (Lindner *et al.*, 2010;

Barrett, 2015). At its southern range edge *T. cordata* bioclimatic envelope models predict both large reductions in area and an altitudinal shift in suitable habitat due to potential changes in summer precipitation by 2080 (Attorre *et al.*, 2011), but this does not mean that individuals will actually be able to occupy these sites by that time. Many factors unexplored for the species will dictate its response and capacity to colonize new habitat: its ability to adapt to new conditions via plastic or evolutionary responses (Nicotra *et al.*, 2010), reproduction and dispersal ability, competitive capacity, recruitment and growth (Pausas and Lavorel, 2003). Dispersal capacity, reproductive strategy/success and community interactions are especially relevant in a landscape context characterized by a high degree of fragmentation, which has the potential to reduce or prevent migration (Iverson, Schwartz and Prasad, 2004; Christmas, Breed and Lowe, 2015). Therefore, this study will examine aspects of *T. cordata* ecology, reproductive strategy, population connectivity and phenotypic plasticity identified earlier that are yet to be researched.

1.4 Research aims

This thesis will focus primarily on examining ecological, genetic and metabolic variation in UK *T. cordata* individuals and populations in order to address some of the gaps in our knowledge that were identified above. Specifically, the aims of the study are: to more precisely describe the ecological requirements of the genus within the UK and make intra-generic comparisons, in order to facilitate a better understanding of how coexistence occurs; to better understand the life history of the species with regard to vegetative propagation, by quantifying the extent and circumstances of clonal reproduction and how this varies; to determine whether fragmented *T. cordata* populations have suffered a reduction of gene flow between distinct woodland habitats and if so how this relates to the particular landscape context, and if differences in latitude change how demes respond to fragmentation; finally, to explore physiological variation *in situ* across the UK range of the species, exploring plasticity in response to environmental context and generating hypotheses regarding potential routes for examining local adaption.

1.5 Thesis outline

Chapter 2 examines how the UK native *Tilia* species, *T. cordata* and *T. platyphyllos*, segregate along a variety of physiographic and edaphic environmental axes when occurring in sympatry.

Chapter 3 quantifies and describes the extent and character of clonal reproduction in *T. cordata* populations throughout its UK range, and relates this to possible correlates such as climate.

Chapter 4 examines how fragmented demes of *T. cordata* differ with respect to neutral genetic variation, to infer the presence or absence of gene flow between them and any potential landscape constraints on the process. This chapter will also contrast the same between populations at distinct latitudes with subtly differing climates that potentially affect their fertility (and therefore turnover of individuals).

Chapter 5 explores metabolic variation in *T. cordata* individuals collected *in situ* from semi-natural habitats and contrasts this with neutral genetic and adaptive morphological variation.

Chapter 6 will summarise the overall conclusions of the thesis, describe their implications for population persistence and suggest fruitful avenues for future research.

2 Segregation of *Tilia cordata* Mill. and *T. platyphyllos* Scop. (Malvaceae) along environmental gradients

Published as:

Barker, C., Ashton, M., & Ashton, P. A. (2015). Segregation of *Tilia cordata* and *T. platyphyllos* (Malvaceae) along environmental gradients. *New Journal of Botany*, 5 (3), pp. 157–163.

2.1 Abstract

Closely related congeneric species living in sympatry raises the question of how they are separated along abiotic and biotic axes. Within northern Europe, *Tilia cordata* and *Tilia platyphyllos* are two such species. They are large, long-lived trees, often dominant members of their community, and frequently grow sympatrically. This study measured a series of edaphic and topographic variables across a number of sites and examined these via univariate and multivariate approaches to understand how the two species differed. This showed that *T. cordata* occupies soil with significantly greater organic carbon content and is present in areas with greater potential incident solar radiation. There was no significant difference between the two species across soil depths, acidity, moisture content or elevation above sea level. *T. cordata* is the more variable of the two species with regards to soil pH and soil moisture content. This is the first comprehensive account of these characters for the two species. The quantitative values obtained in this study are broadly commensurate with descriptive accounts from the literature. However, contrary to previous descriptions, no difference was observed in pH between the two species and soil nutrient levels were the inverse of those expected. This study presents potentially fruitful areas for examining potential niche separation between the two species alongside potential differences in litter leaf chemistry.

2.2 Introduction

High numbers of plant species can coexist within communities, with species richness reaching highs of 89 per 1 m^2 in temperate grassland, or 942 per 1×10^4 m^2 in tropical rainforest (Wilson *et al.*, 2012). What mechanisms are in operation to maintain these levels of diversity? One suggestion is provided by classical ecological theory: the competitive exclusion principle (Gause's law), based on the Lotka-Volterra competition model (Gause, 1934). That is, competition for finite resources forces coexisting species to occupy discrete niches in order to stably coexist. Defined by Hutchinson (1957) as an n-dimensional hypervolume in environmental space, the dimensions represent ecological variables along which a species can be positioned. Experimental determination of the existence of niche separation sensu Silvertown (2004) requires four tests: a greater impact of intraspecific competition than interspecific competition; segregation along an environmental axis; trade-offs between traits on niche axes and niche shifts following experimental manipulation of competition.

Animal communities can provide examples of the principle at work, often as a clear differentiation between species' trophic niches (e.g. Ashrafi et al., 2011; Dammhahn et al., 2013; Laakmann et al., 2009; Navarro et al., 2013). It has been more difficult to see how the concept of niche separation applies to plants, with the apparent commonality of resource requirements across species: all plants require light/water/CO_2, NPK and a common set of micronutrients. Nevertheless, studies suggest that plants segregate along a variety of environmental niche axes. Niche separation of rooting depth occurs in desert and grassland plant communities (Mamolos, Elisseou and Veresoglou, 1995; Nobel, 1997), and arctic tundra plants adopt a variety of nitrogen uptake strategies, absorbing differing forms of N at different times and depths (Kielland, 1994; McKane *et al.*, 2002). Plants in meadow communities segregate along niche axes of soil drying and soil waterlogging tolerance (Silvertown *et al.*, 1999) while niche differentiation in shade tolerance plays a part in shaping community structure in both temperate and tropical forests (Kobe *et al.*, 1995; Kobe, 1999). Plants may also be organised along micro-topographical gradients (e.g. valley bottom versus slope) in tropical rainforests (Wright, 2002).

If the concept of niche separation has a role in the maintenance of plant community structure, the principle of competitive exclusion may be particularly relevant when examining closely related species whose shared phylogenetic history could indicate

overlapping resource requirements, although the extent and character of niche conservatism across closely related taxa is debated (Wiens and Graham, 2005). Niche characteristics demonstrate a phylogenetic signal in some cases (e.g. Ricklefs and Latham, 1992; cf. Cavender-Bares et al., 2004; Losos et al., 2003), and it has been demonstrated experimentally that species-interactions can increase with phylogenetic closeness (Burns and Strauss, 2011). Therefore for closely-related sympatric species, niche separation may be of paramount importance.

Two such closely related sister species are the broadleaved deciduous trees *Tilia cordata* Miller and *T. platyphyllos* Scop. (McCarthy, 2012). Native to the UK, they occur throughout most of England, often in sympatry, almost exclusively in the case of the latter (Pigott, 1981b). English populations represent the northwest extreme of the species' distributions (**Figure 2.1b**). While a broad overview of the ecological requirements of *T. cordata* can be obtained in the literature (Pigott and Huntley, 1978; Pigott, 1991; Pigott and Pigott, 1993), *T. platyphyllos* has received less attention (Pigott and Huntley, 1981; Sheykholeslami, Namiranian and Sagheb-Talebi, 2011). The bulk of the aforementioned work is primarily descriptive or observational, and no systematic study of ecological parameters exists. Also lacking is a contrast of the requirements of the species where they are growing sympatrically.

This study addresses this gap by undertaking a comparison of environmental variables found within the native range of UK sympatric *Tilia* populations to identify any differences in species' environmental preferences; the second test for identifying niche separation (segregation alone environmental axes) described previously (Silvertown, 2004). To achieve this we have chosen to measure and contrast edaphic and physiographic variables that are analogous to some of the potential axes of niche separation identified earlier, such as soil quality, moisture, and depth. Nutrient status such as nitrate and phosphate levels was considered for inclusion but as nutrient availability is correlated with soil acidity only the latter is measured here. They also include areas identified in the literature as differing between the species; *T. cordata* is described as occupying less basic and typically richer soils than *T. platyphyllos* (Pigott, 1981b, 1991), while the latter is reported as occurring primarily on steep slopes and cliffs (Pigott, 1981b).

2.3 Methods

2.3.1 Field & laboratory work

Sympatric populations of *T. cordata* and *T. platyphyllos* in the UK occur primarily along the Wales/England border, with additional sites in South Yorkshire, Derbyshire and Staffordshire (Pigott, 1981b, 1991). Sampling at 101 trees was conducted within this range across 11 semi-natural ancient woodland sites (**Figure 2.1** & **Table 2.1**). Sites were chosen based on uniformity of substrate, all being present on thin limestone soils (rendzinas and calcareous brown earths) as well as being reported as containing sympatric populations. Roudsea Wood possesses *T. cordata* only and was included as a comparison. All sites sampled are coppiced with standards woodland. Sampling was undertaken between July 2011 and October 2012 to record topographic (slope aspect, gradient, elevation above sea level) and edaphic characters (acidity, depth, moisture and organic carbon content).

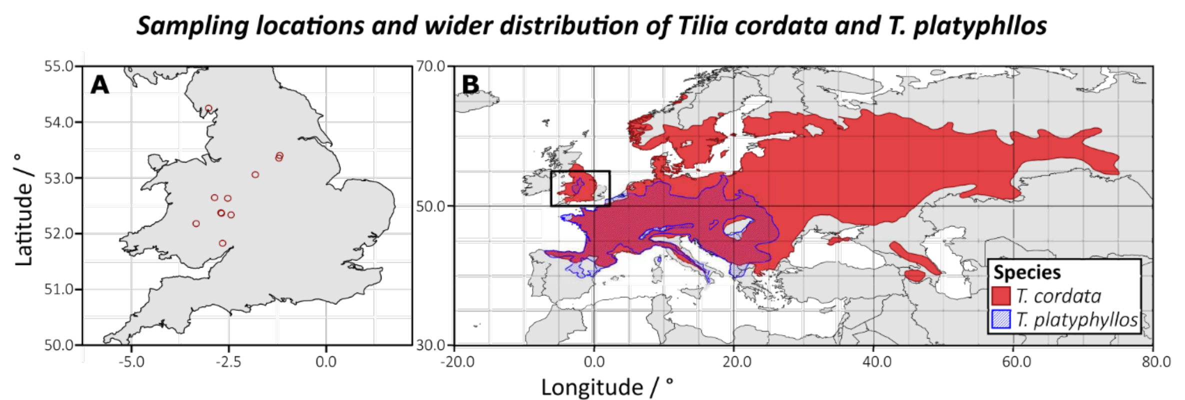


Figure 2.1: A) Map of Britain with open circles indicating mean coordinates for each population at the 11 sites visited. **B)** Map displaying the distribution of *Tilia cordata* and T. platyphyllos across Europe (EUFORGEN, 2009). *T. cordata* is represented by the solid red area while *T. platyphyllos* is represented by the hashed blue area. Inset square: area covered by map A.

Site	Date Sampled	*n*		Latitude	Longitude
		T. cordata	*T. platyphyllos*		
Roudsea Wood, Cumbria	July, 2011	15	0	54.23339	-3.025833
Anston Stones Wood, Yorkshire	August, 2011	2	15	53.34426	-1.207147
King's Wood, Yorkshire	August, 2011	0	18	53.39842	-1.181878
Dumbleton Dingle, Worcestershire	August, 2011	7	2	52.33138	-2.438264
Lady Park Wood, Gloucestershire	August, 2011	14	6	51.82497	-2.659069
Earl's Hill Wood, Shropshire	September, 2011	0	6	52.63727	-2.869548
Crew's Hill Wood,Worcestershire	September, 2011	6	12	52.35649	-2.692415
Highbury Wood, Gloucestershire	September, 2011	15	0	52.17621	-3.341825
Tick Wood, Shropshire	September, 2011	0	11	52.62624	-2.526978
Knapp & Paper Mill, Worcestershire	October, 2011	11	6	52.36835	-2.708353
Hinkley Wood, Staffordshire	October, 2012	12	3	53.05204	-1.80229
Total samples		**82**	**79**		

Table 2.1: Sampling locations. Collection sites and abbreviations for the two *Tilia* species including sampling date, final sample numbers by species per site and mean coordinates of individuals sampled

Leaf samples were collected from each tree and were later identified according to the Anderson morphological hybrid index approach (Anderson, 1949), after Pigott (1969), which scores individuals on a range of 0 - 20. Hybrid indices are a commonplace method for identifying sympatric plant species from populations presenting intermediate and varied morphology due to hybridisation (e.g. Gosler, 1990; Kiaer et al., 2007; Thórsson et al., 2007; Tovar-Sánchez and Oyama, 2004; Wigston, 1975). Leaf colours other than those described by Pigott were observed and therefore this character was ignored, giving a possible HI range of 0 – 18. Only trees which received a hybrid index score lower than six or higher than seven exclusive (*T. cordata* and *T. platyphyllos* respectively) were included in the analysis.

Intermediate individuals were not considered. The distinction between taxa is that made by Pigott (1969), based on the morphology of hybrid individuals of known status and typically corroborated with molecular identification based on microsatellite data (Phuekvilai, 2014).

After collecting leaf samples, individual tree positions, elevations (*ESL*) and slope aspect (if trees were not on level ground) were recorded using a handheld GPS unit (Garmin eTrex Vista). Where appropriate, slope gradient was measured using an analogue clinometer. Given the circular nature of slope aspect (i.e. slopes with facings of 359° and 1° are similar with respect to light levels despite their numeric difference), an estimate of potential annual direct incident solar radiation (*DIR*) was calculated after McCune and Dylan (2002). *DIR* also incorporates latitude and slope gradient so these do not feature as separate factors in this analysis.

Mean soil depth (*SDE*) taken from 4 evenly distributed points between 20 cm and 200 cm from the base of the trunk. 'Soil depth' was taken to be the point at which a metal rod reached an obstruction on insertion into the ground. Beneath each individual, samples of the A-horizon were collected and bulked for subsequent analysis to determine those soil chemistry parameters which have been suggested to differ between the two species (Pigott, 1991, 1981). Standard gravimetric techniques were applied to determine soil moisture content (θ_g) and soil organic content (*SOC*). Recording the mass of samples before and after storage in a drying cabinet at 110 °C for 12 h allowed determination of water mass present, while incineration in a muffle furnace at 450 °C for 4 h allowed for calculation of SOC (dry-weight basis). Soil pH (*pH*) was measured with a pH meter using a 0.1M KCl solution.

2.3.2 Data analysis

All analysis was performed using the R software package, version 3.1.2 (R Core Team, 2014). During the initial data exploration, most data were found to be non-normal with high skew. Typical data manipulation techniques such as *log* transformation did not resolve this issue and thus are not applied in the analysis presented. Therefore a nonparametric hypothesis test (Mann-Whitney *U*) was used to compare the distributions of ecological parameters between species. Dispersion was calculated using Q_n in the R package `robustbase` (Rousseeuw *et al.*, 2015). This is a robust estimator of standard deviation (Rousseeuw and Croux, 1993). Sampling accuracy of the observed medians was determined via a nonparametric bootstrap procedure (resampling with replacement; 10^4 repetitions) in

the R package `boot` (Canty and Ripley, 2015) which generated precise 95% confidence intervals using the bias-corrected and adjusted method.

Multivariate analysis was undertaken using the ROBust method for Principal Components Analysis (ROBPCA) from the R package `rrcov` (Todorov and Filzmoser, 2009) to explore potential differences between the two species due to the interaction of abiotic and edaphic variables. In contrast with standard principal components analysis, this approach reduces the influence of outlying data (Hubert, Rousseeuw and Vanden Branden, 2005). Data was standardised to have zero-means and a standard deviation of 1 to account for the different units of measurement between parameters.

2.4 Results

2.4.1 Comparison of observed ecological parameters

T. cordata occurs on soils with significantly higher organic content (**Figure 2.2D**; *T. cordata* = 208.78 g kg^{-1}; *T. platyphyllos* = 159.52 g kg^{-1}) and in locations with significantly greater potential annual incident solar radiation (**Figure 2.2E**; *T. cordata* = 0.957 MJ cm^{-2} y^{-1}; *T. platyphyllos* = 0.806 MJ cm^{-2} y^{-1}).

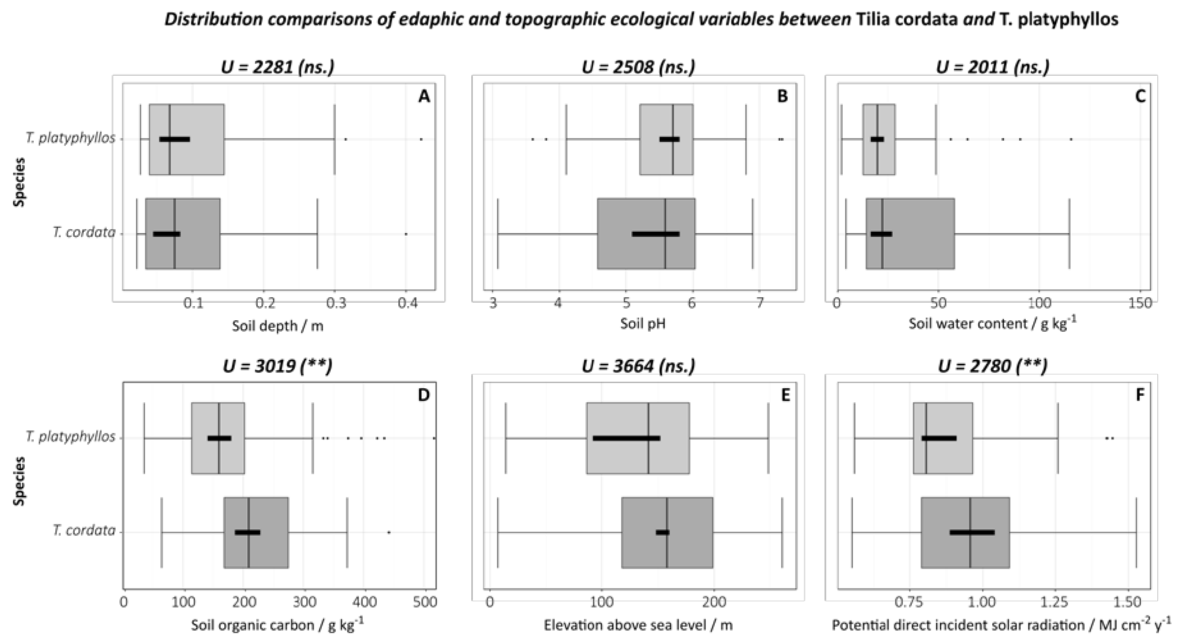


Figure 2.2: Box and whisker plots indicating median values and upper and lower quartiles for each ecological parameter by species. The thick black bar spanning the median indicates the 95% confidence interval for that statistic (see text for details). Extreme values in soil moisture and organic content percentages (>155 g kg^{-1} and >550 g kg^{-1} respectively) are not displayed for legibility's sake (missing $n = 6$ and 8). Shown above each plot is the calculated Mann-Whitney U statistic for between-species comparison. Values in parentheses indicate the significance level (*ns.* = not significant at $\alpha = 0.05$; ** = $p < 0.01$). **A:** distributions of soil depth surrounding individuals of both species. **B:** distributions of A-horizon soil acidity. **C**: distributions of A-horizon soil water content. **D:** distributions of A-horizon soil organic carbon content. **E:** distributions of elevation above sea level occupied by individuals of both species. **F:** potential levels of solar radiation received by locations occupied by individuals of both species, estimated following McCune and Dylan (2002).

No significant difference in distributions was observed between the two species for *SDE*, *pH*, *MPC* or *ESL* (**Figure 2.2A**, **B**, **C**, **E** respectively). Levels of variation (calculated as Q_n) are comparable for the two species across most variables but markedly larger in *T. cordata* for *pH* and θ_g (**Table 2.2**). Summary statistics for the measured ecological parameters are given in **Table 2.2** as well as calculated Mann-Whitney U statistics and significance levels.

	Tilia cordata			*Tilia platyphyllos*				
Variable	**Median**	**Range**	Q_n	**Median**	**Range**	Q_n	**U**	**Significance**
SDE (m)	$0.075^{+0.008}_{-0.031}$	0.02 - 0.40	0.054	$0.068^{+0.028}_{-0.014}$	0.03 - 0.42	0.049	2281	ns.
pH	$5.59^{+0.22}_{-0.50}$	3.08 - 6.90	0.95	$5.70^{+0.10}_{-0.20}$	3.60 - 7.34	0.65	2508	ns.
θ_g (g kg^{-1})	$21.98^{+5.16}_{-5.74}$	4.08 - 290.53	17.53	$19.54^{+3.25}_{-3.49}$	1.85 - 115.76	11.73	2011	ns.
SOC (g kg^{-1})	$208.78^{+19.2}_{-22.55}$	64.74 - 713.60	80.81	$159.52^{+20.48}_{-18.79}$	34.72 - 785.23	77.53	3019	**
ESL (m)	$157.90^{+2.13}_{-9.76}$	7.00 - 260.96	40.30	$141.44^{+10.56}_{-49.44}$	14.00 - 248.66	36.99	3664	ns.
DIR (MJ cm^{-2} y^{-1})	$0.957^{+0.081}_{-0.042}$	0.550 - 1.527	0.230	$0.806^{+0.105}_{-0.016}$	0.558 - 1.446	0.191	2780	**

Table 2.2: Edaphic and physiographic parameters observed across populations of *T. cordata* and *T. platyphyllos*. Includes sample size, observed median (including bootstrapped 95% confidence intervals; see text for details), the Q_n measure of dispersion (Rousseeuw and Croux, 1993), calculated Mann-Whitney *U* scores and significance levels (*ns.* = not significant at $\alpha = 0.05$; ** = $p < 0.01$). Soil parameters: *SDE = soil depth; pH = soil pH;* θ_g *= gravimetric soil moisture content; SOC = gravimetric soil organic content.* Physiographic parameters: *ESL = elevation above sea level; DIR = potential annual direct incident solar radiation.*

2.4.2 Multivariate analysis

Only the first three principal components *PC1*, *PC2* and *PC3* are considered here as they collectively explain over 80% of the variance present in the data (49.5%, 22.3% and 16.4% respectively). See **Table 2.3** for exact variable loadings for all three components. In a scatterplot of PC1 against PC2 individuals from both species form overlapping similarly shaped clusters (**Figure 2.3a**). The remaining pairwise scatterplots (**Figure 2.3b, c**) have similar patterns to each other, with two distinct species clusters which is driven primarily by differences in SOC (**Table 2.3**). Whether a tree is present in a sympatric or allopatric population does not appear to affect the clustering, with both categories being intermingled in all plots (**Figure 2.3**).

Variable	PC1	PC2	PC3
SDE	0.468	-0.097	0.122
pH	0.412	**0.845**	**-0.244**
θ_g	-0.052	0.222	-0.147
SOC	-0.080	-0.250	**-0.944**
DIR	**-0.489**	**0.298**	-0.048
ESL	**-0.603**	0.274	0.105

Table 2.3: Variable loadings for the components of the PCA considered here. The first two parameters with the highest absolute influence are indicated by bold text.

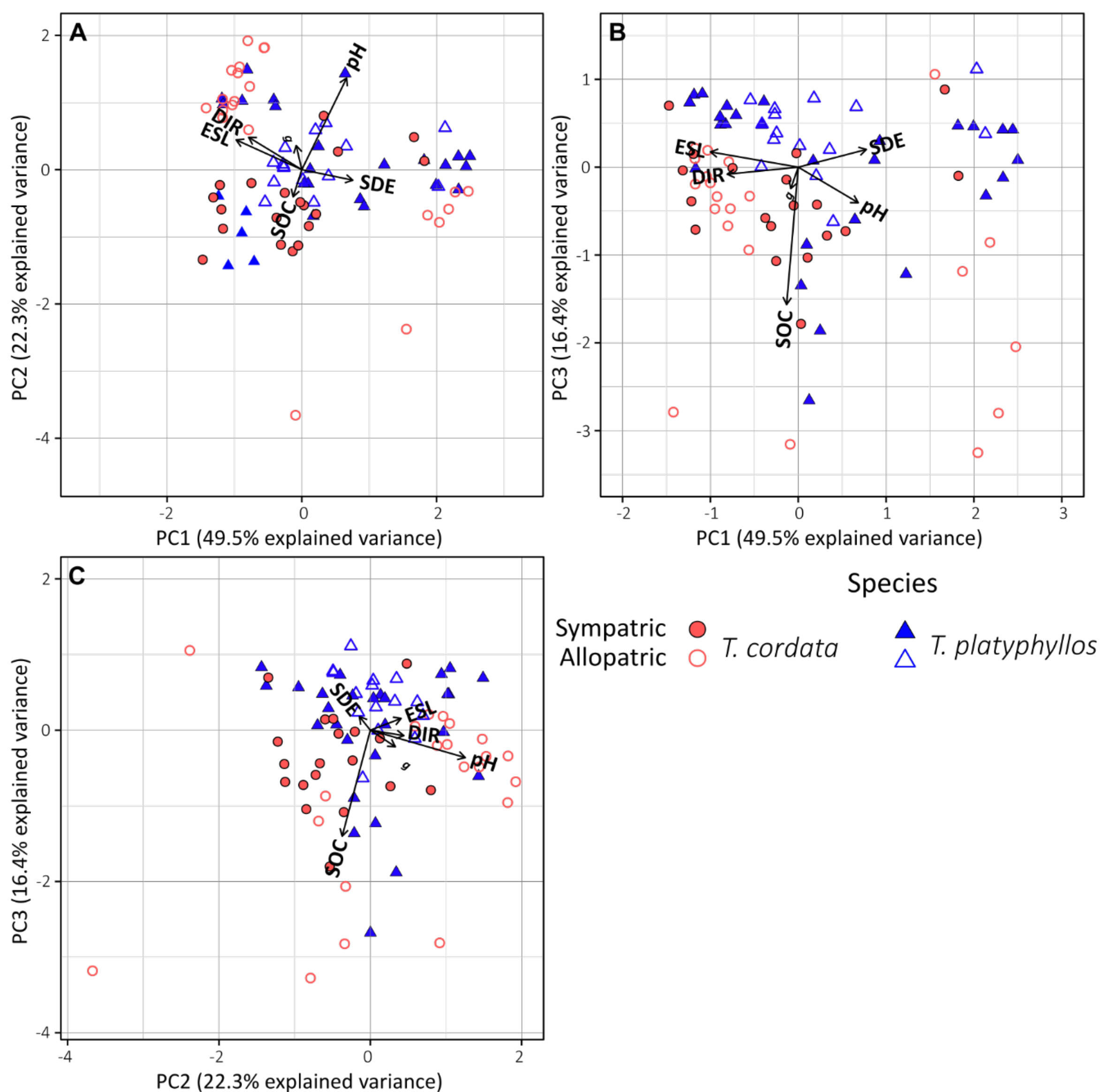


Figure 2.3: Pairwise biplots of the first three principal components (PC1, PC2, PC3) of the RObust Principal Components Analysis (ROBPCA). Circles indicate T. cordata individuals while triangles indicate T. platyphyllos. Open shapes indicate individuals sampled from a sympatric population of both species while closed shapes indicate individuals sampled from sites with only one species present (allopatric).

2.5 Discussion

Evidence for difference in ecological preferences between *T. cordata* and *T. platyphyllos* was found from univariate pairwise comparisons and was also illustrated by ordination. Differences in soil organic carbon content and estimated insolation are statistically significant. Individuals clustering primarily by species rather than by site or presence of the other species on biplots of principal components from the *ROBPCA* supports the former, given the weighting of SOC in *PC3*. For parameters with no significant difference, *T. cordata* occupies soils with greater variability in acidity and moisture content than that of *T. platyphyllos* (**Figure 2.2b**, **c**; **Table 2.2**) and also displays more variation in position within the ordination (**Figure 2.3**).

A qualitative description of the distribution of *T. cordata* with relation to soil classification does indeed suggest that it is present on soils that vary widely over the ecological parameters measured here (Pigott, 1991). Similarly a quantitative assessment of edaphic variables in *T. cordata* populations in northwest England is also in line with the observations presented here (Pigott and Huntley, 1978). There is less comparable data on ecological parameters associated with *T. platyphyllos*, but its distribution is more restricted and apparently tightly linked with soil characteristics and topography (Pigott, 1981b).

Two explanations for the differences observed in estimated insolation between the species present themselves. Firstly, *T. platyphyllos* is most often found on steep slopes and cliffs (Pigott, 1981b; Sheykholeslami, Namiranian and Sagheb-Talebi, 2011), and given that the equation utilised to estimate potential insolation takes slope angle and aspect into account (McCune and Dylan, 2002), the difference between distributions of *DIR* values may be a result of this preference. Alternatively, it may be an indirect indication of a preference for higher light levels in *T. cordata*. It has been suggested that the former species is less shade-tolerant than *T. platyphyllos* across both its UK and wider ranges, according to Ellenberg's indicator values and related work (Ellenberg, 1988; Hill, Preston and Roy, 2004). *DIR* does not consider actual canopy transmission levels, surrounding topography or weather/climatic conditions (e.g. cloud cover), being only an estimate of insolation at a specific location on a particular topography. Hence this potential difference requires further investigation. Soil organic carbon content also differs significantly between the species. This result is expected given that *T. cordata* is noted to be characteristic of typically richer soils than *T. platyphyllos* (Pigott, 1981b, 1991).

Published literature also suggests that *T. cordata* generally occurs on less basic soils than *T. platyphyllos* (Pigott, 1981b, 1991), but no significant difference between soil *pH* was observed here. This is despite the observed range of *pH* for both species (excluding outliers according to box and whisker plots; **Figure 2.2b**) being in line with previous quantitative assessments of the species' preferred edaphic properties (Pigott and Huntley, 1981; Sheykholeslami, Namiranian and Sagheb-Talebi, 2011). Therefore the lack of a statistically significant difference in observed soil acidities may be the result of the preferences of each species being overestimated previously. Alternatively this may be a result of most sampling locations in this study being areas with both species occurring in sympatry (**Table 2.1**). As the geographic distribution of *T. platyphyllos* is more restricted within northern Europe, the greater spread of *T. cordata* may mean it is free to occupy more acidic sites while the inverse would not be true. There is also the possibility that soil characteristics are affected by the species themselves rather than passively utilised. It has been reported that the litter of *T. cordata* generates rich soils (Pigott, 1989), but no basis for comparison with *T. platyphyllos* is available in the literature. Edaphic conditions can be influenced by differences in foliar chemistry (e.g. Finzi et al., 1998; Lovett et al., 2004) which may lead to positive feedback effects (Hobbie, 1992; Gómez-Aparicio and Canham, 2008) that could increase the fitness of conspecific individuals and affect community dynamics (Aponte *et al.*, 2011), 2011).

Observed differences between ecological preferences are not as distinctive as indicated by previous work. It may be the case that the separation between the species lies along unexamined resource or environmental axes. In particular no biotic factors were investigated here. As shade tolerance is reported as differing between the species (Ellenberg, 1988; Hill, Preston and Roy, 2004), they may differ in their responses to competition for available light in forest communities, given that interspecific differences in juvenile mortality as a response to varying light levels are a component of temperate woodland community dynamics and structure (Kobe *et al.*, 1995). There is also for potential for mycorrhizal communities to differ between congeneric species, such as that in *Quercus* (Morris *et al.*, 2008), which in turn influence nutrient uptake. However this potential difference is unexplored in *Tilia* or in many other genera.

Habitat preferences have been found to differ between congeneric tree species along some of the axes examined here, although similar studies are rare. Edaphic separation has been recorded between *Quercus petraea* and *Q. robur* (Bacilieri, Ducousso and Kremer, 1995) and between the sympatric North American oak species, *Q. virginiana* and *Q.*

geminata (Cavender-Bares, Kitajima and Bazzaz, 2004), with the former also exhibiting more variation across edaphic factors (Cavender-Bares and Pahlich, 2009). Light levels separate *Neolistea aciculata* (Lauraceae) and *N. sericea* (Yamasaki, Yamada and Okuda, 2013) and the tropical tree species *Pterospermum diversifolium* (Malvaceae) and *P. javanicum* (Yamada, Ngakan and Suzuki, 2006). Similarly the distributions of *Dryobalanops aromatica* (Dipterocarpaceae) and *D. lanceolata* are highly linked with contrasting preferences in soil physical structure and topographic variables such as slope steepness (Itoh *et al.*, 2003), while *Pourama bicolor* and *P. guianensis* ssp. *guianensis* are found at lower elevations than their sympatric congenerics (Magård, 2002).

The presence of the species on marginally different soils and in areas with differing light levels is an indication of segregation along niche axes, which is the third requirement suggested by Silvertown (2004) to determine the existence of niche separation.. Therefore these findings provide a possible approach to direct future studies of competitive exclusion within the genus.

2.6 Acknowledgements

Funding for the work undertaken was provided by both the award of a Plant Study Grant by the Botanical Society of Britain & Ireland to the lead author, as well as internal grants from Edge Hill University. The authors wish to thank Clare Bugg, Colin Davies, Emily Simmons and Jon Tomkinson for their assistance in the field and with data collection, as well as the anonymous reviewers whose comments and feedback have been used to improve this manuscript.

3 Reproductive strategy of *Tilia cordata* Mill. across its UK range

3.1 Introduction

Although obligate asexual (clonal) reproduction is rare in angiosperms, facultative clonality is extremely common (Harper, 1977; Eckert, 2002; Barrett, 2015), particularly in perennial plants (Holsinger, 2000; Morris *et al.*, 2014). Vegetative propagation (rather than apomictic seed production) is the most common form of clonality with an estimated ~80% of all angiosperm lineages possessing some means of reproducing in this manner (Klimeš *et al.*, 1997). In contrast to exclusively sexual reproduction, a mixed reproductive strategy provides several ecological benefits such as the ability of clones to forage for resources (light, nutrients, water) in patchy environments, lowered chance of death for a particular genotype due to spread mortality risks, the possibility of physiological integration and spatial division of labour (e.g. sharing of root systems) among shoots, etc. (Barrett, 2015). The relative emphasis on sexual versus asexual reproduction differs not just between but also within species (Silvertown, 2008), with different habitats favouring one reproductive mode over the other. For instance, in widely distributed species clonality is often more prevalent in marginal populations which are subject to less favourable environmental conditions (Brzosko *et al.*, 2013). If these lead to a failure or a reduced frequency of sexual reproduction vegetative propagation can contribute to population persistence due to the longevity of clonal individuals (de Witte and Stöcklin, 2010; Barrett, 2015).

Inversely clonality may also have negative consequences on individual fitness and population duration, depending on its frequency and arrangement. It necessarily reduces genotypic diversity, and if species are self-incompatible and the spatial arrangement of vegetative growth is aggregated, outcrossing success can be reduced (Charpentier, 2001; Vallejo-Marín, Dorken and Barrett, 2010). This can affect the viability of small populations due to lower sexual fecundity as a result of intraclone incompatibility (Nuortila, Tuomi and Laine, 2002; Honnay and Bossuyt, 2005). This process can ultimately lead to a complete failure of sexual reproduction and if any particular clonal lineage has a competitive advantage, monoclonal stands (Halkett, Simon and Balloux, 2005; Honnay and Bossuyt, 2005). Reduced genotypic diversity and lack of sexual reproduction can affect future adaptive potential by lowering the efficiency of natural selection or increase the rate at which

deleterious alleles can accumulate, further decreasing reproductive capacity. Either of these factors could affect the likelihood of long-term population persistence (Holsinger, 2000).

Therefore knowledge of the frequency and spatial structure of clonality and the factors that drive reproductive strategy in any particular species is part of a full understanding of their ecology and population genetics, and insight into how and when a species undergoes vegetative reproduction is important in applications which rely on these areas. For instance, the balance of reproductive modes is an important predictor of migration potential (Morris *et al.*, 2014) given the relative effectiveness of seed versus vegetative dispersal (Silvertown, 2008), which makes an understanding of clonality potentially important information in the generation of species distribution models used to forecast responses to climatic change (as these can benefit from the inclusion of life-history traits, e.g. Matthews et al., 2011). Similarly as the impact of some plant pathogens is expected to increase (e.g. *Phytophthora* spp.) as a response to more favourable environmental conditions or an increased susceptibility of host species due to factors such as drought stress (Sturrock *et al.*, 2011), and as biotic stress tolerance generally and pathogen resistance specifically are linked to genetic diversity (or the ability to generate novel genotypes via recombination), an understanding of clonality may be part of mitigating or reducing these impacts (Honnay and Bossuyt, 2005). Finally where genetic variation is at risk and the need for conservation has been realised, knowledge of this aspect of an organism's life history is important in allowing biologists and foresters to choose appropriate management measures that best capture and maintain the species' genetic variation (Namroud *et al.*, 2005).

An example of a partially clonal plant species is *Tilia cordata* (small leaved lime), an entomophilous canopy tree with a widespread but patchy distribution (**Figure 3.1**). It is an effective vegetative propagator, producing clonal individuals ('ramets', *sensu* Harper, 1977; cf. 'genets', the clonal group as a whole) in a variety of ways. Shoots at the base of its trunk can develop into multiple stems and the connecting tissue can rot away, leading to separate individuals. These basal shoots can also root if they touch the ground, and fallen trees can produce multiple vertical shoots along their length which ultimately develop into new ramets (Pigott, 1991). Successful sexual reproduction in the species relies on critical temperatures being reached during both pollination (Pigott and Huntley, 1981) and seed development (Pigott, 1981a). The current extent of its range was established during the Holocene climatic optimum when summer temperatures were 1 – 2 °C higher than currently (Birks, 1989; Davis *et al.*, 2003) and its fertility was presumably not an issue. The subsequent

temperature decline means that northern marginal populations rarely set fertile seed except during the warmest summers, due to either a failure of pollination (Pigott and Huntley, 1981) or of fruit development (Pigott, 1981a), in contrast to more southern populations which frequently produce viable fruits (Pigott, 1991). In concert with other potential factors such as herbivory (Pigott, 1991) this ultimately leads to infrequent sexual recruitment, which may promote vegetative reproduction, or at least increase its relative importance (Silvertown, 2008). As a result clonal growth has been suggested as the reason for population persistence in marginal locations such as the north-west of England (Pigott and Huntley, 1978). Fertility is reported to be high in southern marginal populations (Pigott, 1991), where the constraint on its distribution is insufficient water (Pigott and Pigott, 1993; De Jaegere, Hein and Claessens, 2016). Large crops of fertile seed are reported at this southern limit (Pigott, 1991) but whether this results in successful regeneration is unclear. Despite the potential importance of vegetative propagation to the species the extent, character and structure of asexual reproduction for *T. cordata* have never been described, nor has its relationship with climate (or any other factors).

Although not rare its genetic resources have been identified as at risk due to its poor dispersal ability relative to other tree species (De Jaegere, Hein and Claessens, 2016) and its often small and isolated populations (Hemery *et al.*, 2010), with active genetic conservation measures already in place (Turok *et al.*, 1996). In light of the potential effects of clonality described above, knowledge of the degree of clonality and how it varies across a species' range or in response to different ecological constraints is a prerequisite to understanding its genetic structure and ecology and therefore to produce effective genetic conservation strategies (Barsoum, Muller and Skot, 2004; Namroud *et al.*, 2005). These aspects have not yet been studied in *T. cordata*, and so this study is an examination of clonality within 22 populations across the extent of its native UK range, with the aim of assessing both the frequency and arrangement of clonal individuals and their effect on spatial genetic structure (SGS). In addition the potential links between climate, individual density, demography and incidence of clonality are assessed. These factors are known to relate to the frequency of asexual reproduction in other species (e.g. Morris et al., 2014; Weed and Schwarzländer, 2014).

3.2 Methods

3.2.1 Sampling method and location

Twenty-two populations of *Tilia cordata* across its UK range (**Figure 3.1** & **Table 3.1**) were sampled to assess the degree, character and arrangement of clonality present and their fine-scale SGS. All sites are semi-natural ancient woodland (i.e. continuously wooded since 1600 CE) with similar management regimes (coppicing). Sites were chosen to cover a latitudinal gradient from close to the northern UK range edge (**Figure 3.1**) to as close as possible to the mainland Europe), and to avoid populations sympatric with the other UK native *Tilia* species (*T. platyphyllos,* large leaved lime).

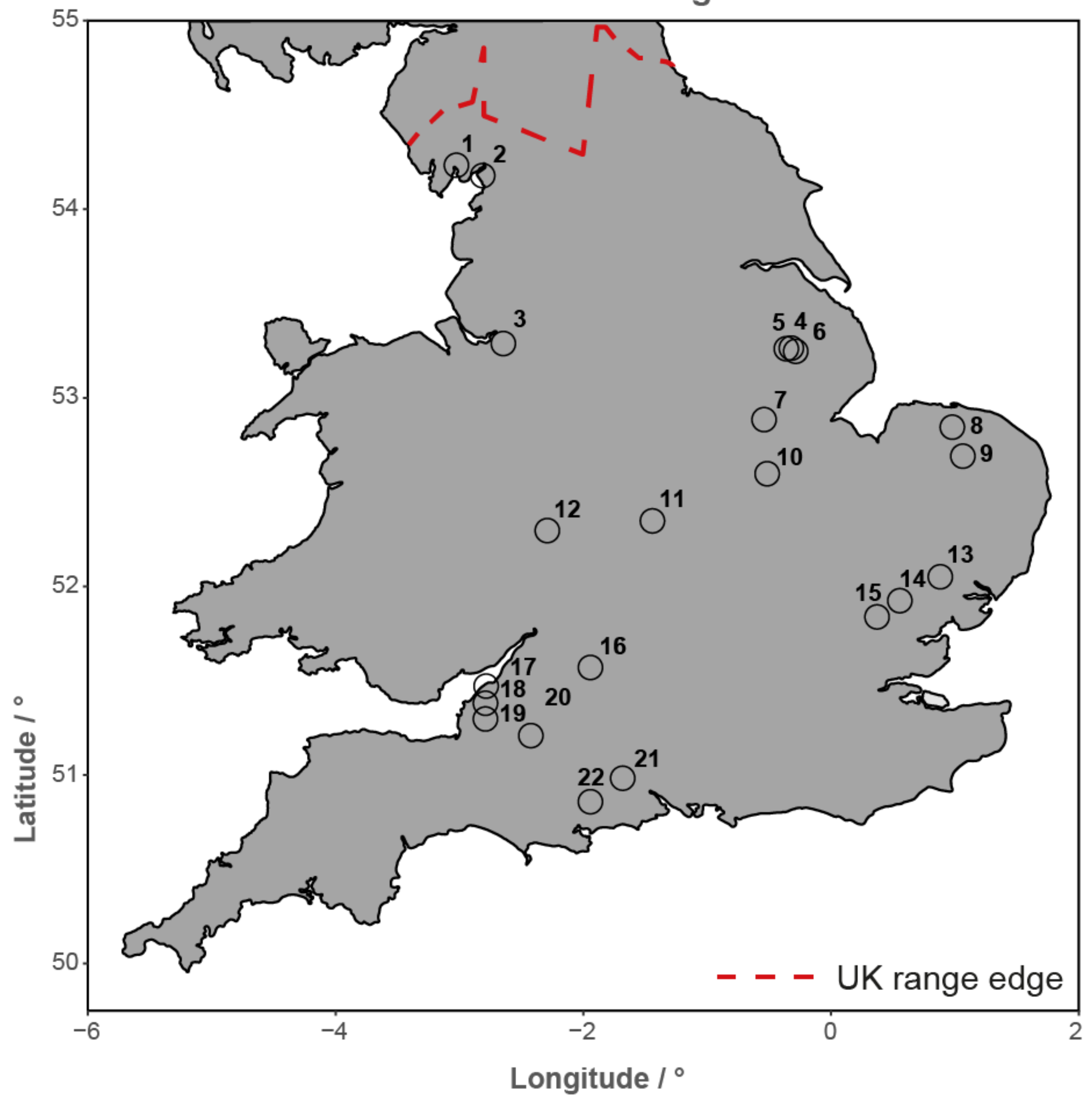


Figure 3.1: geographic position of sampling sites (see **Table 3.1** for key to numbers and exact location); the dashed red line indicates the approximate location of the northern range edge in the UK.

In each population a single 30 x 30 m quadrat (Arnaud-Haond, Duarte, *et al.*, 2007) was randomly placed and all *T. cordata* individuals present were mapped using measuring tapes. Demography data was recorded by taking measurements of diameter at breast height (DBH) as well as position in canopy (understorey, canopy); where trees had multiple stems, the largest was measured. Understorey individuals with a DBH less than 0.1 m were considered saplings (no seedlings were observed). Leaf tissue from each was collected and dried for further laboratory analysis. Where samples were not available due to inaccessible leaves, the position of individuals was recorded regardless (81 *n*). A total of 647 individuals were sampled across all locations (**Table 3.1**).

Number	**Name**	**Code**	***n***	**Ungenotyped *n***	**Latitude**	**Longitude**
1	Roudsea Wood	ROUDS	22	2	54.23	-3.02
2	Eaves Wood	EAVES	34	1	54.18	-2.81
3	Bank Rough	BANKR	13	1	53.29	-2.65
4	College Wood	COLLE	9	0	53.26	-0.32
5	Hardy Gang	HARDY	35	16	53.26	-0.36
6	Ivy Wood	IVYWD	70	5	53.25	-0.29
7	Kirton Wood	KIRTO	17	1	52.88	-0.54
8	Swanton Novers	SWNTN	54	2	52.84	0.98
9	Hockering Wood	HOCKE	19	4	52.69	1.06
10	Collyweston Great Wood	CLWGW	41	4	52.60	-0.52
11	Ryton Wood	RYTON	19	0	52.35	-1.44
12	Shrawley Wood	SHRAW	28	0	52.30	-2.29
13	Groton Wood	GROTN	48	6	52.05	0.88
14	Bovingdon Hall	BOVHL	49	4	51.92	0.56
15	Garnetts Wood	GARNE	40	2	51.84	0.37
16	Webb's Wood	WEBBS	43	19	51.57	-1.94
17	Weston Big Wood	WESTN	18	6	51.47	-2.78
18	King's Wood	KINGW	25	5	51.38	-2.79
19	Cheddar Wood	CHEDD	9	1	51.30	-2.79
20	Asham Wood	ASHAM	6	0	51.21	-2.42
21	Langley Wood	LNGLY	25	2	50.98	-1.68
22	Queens Copse	QUEEN	23	0	50.86	-1.94
		Total:	647	81		

Table 3.1: Sampling site names, abbreviated codes, numbers and locations as well as the number of individuals collected and genotyped at each. Also provided is the number of trees present within quadrats that samples were unavailable for.

3.2.2 DNA extraction & genotyping

Genomic DNA was extracted from 129 individuals using a high-salt CTAB procedure (Tel-zur *et al.*, 1999; Wong, Silvaraj and Phoon, 2014) as standard CTAB methods resulted in extractions of insufficient purity for downstream analysis. 40 mg of dried leaf tissue was homogenised using either a sterile mortar and pestle combined with acid-washed sand, or 1.5 mm Zirconium beads (Sigma-Aldrich, Dorset, UK) in a vortex mixer at 3000 rpm for 3 m. 1 ml of 2% CTAB extraction buffer (2% w/v CTAB, 3 M NaCl, 3% v/v ß-mercaptoethanol, 4% w/v PVP-40, 100 mM Tris-HCl, 20 mM EDTA•Na_2; pH 8.0) at 55 °C was added and then incubated at 55 °C for 1 h. A chloroform:isoamyl alcohol (24:1) wash was performed on the lysate. Further purification of the lysate was achieved by alcohol precipitation using 0.54 volumes isopropanol and 2.5 M ammonium acetate at -20 °C for 1 h. The precipitate was washed using ethanol to remove residual salt, air-dried at 37 °C for 20 m and then rehydrated in TE buffer (10 mM Tris-HCl, 1 mM EDTA•Na_2; pH 8.0).

DNA from the remaining samples was extracted using the 'crude extract' procedure in a KAPA3G Plant PCR Kit (KAPA Biosystems, London, UK). Briefly, a 6.35 mm circular leaf punch was taken from each sample and placed in 125 µl of extraction buffer (50 mM Tris-HCl, 0.1 mM EDTA•Na_2, 2% v/v ß-mercaptoethanol, 1 mM TCEP) before being heated at 95 °C for 5 m.

Samples were genotyped at 10 microsatellite loci (**Table 3.2**) using four multiplex PCR following Phuekvilai and Wolff (2013), with minor modifications. Loci were initially developed for *T. platyphyllos* and the selection here is based on cross-amplification success and polymorphism in *T. cordata* (**Table 3.2**). Amplification was carried out in 10 µl total volume containing either 1X KAPA Plant PCR Buffer (0.2 mM each dNTP, 1.5 mM Mg^{2+}), 0.1-0.2 µM primers, 0.5 mM TCEP, 0.5 U KAPA3G Plant DNA Polymerase or Bioline MyTaq Plant polymerase (Bioline Reagents Ltd., London, UK) and 1 µl of template (either a 1:20 dilution of CTAB-extracted DNA, or undiluted crude extract). Forward primers were tagged with the fluorescent dyes NED, VIC (Applied Biosystems, Warrington, UK), or 6-FAM (Sigma-Aldrich, Dorset, UK). Dye colour and primer concentration can be found in **Table 3.2**. Thermal cycler conditions were as described previously (Phuekvilai and Wolff, 2013).

PCR product size was determined using capillary electrophoresis on an AB3500 Genetic Analyzer (Applied Biosystems, Warrington, UK). Amplicons were diluted 1:10

with nuclease-free water and 1 µl was mixed with 8.9 µl Hi-Di formamide and 0.1 µl LIZ500 size standard before running. Allele peaks were called automatically in GeneMapper 5.0 (Applied Biosystems, Warrington, UK) and then checked manually for errors.

Locus	Multiplex Set	Concentration / µM	Dye
Tc6	A	0.1	VIC
Tc920	A	0.1	6-FAM
Tc937	A	0.1	NED
Tc4	B	0.2	VIC
Tc943	B	0.2	6-FAM
Tc11	C	0.1	NED
Tc915	C	0.2	6-FAM
Tc5	D	0.2	6-FAM
Tc7	D	0.1	6-FAM
Tc951	D	0.2	NED

Table 3.2: Loci used, multiplex groupings, primer concentrations and tagged primer dyes (Phuekvilai and Wolff, 2013).

To reduce the potential for human error in the binning process, the software TANDEM, version 1.07 (Matschiner and Salzburger, 2009) was utilised to produce integer allele sizes from the raw fragment size output produced by GeneMapper.

3.2.3 Data analysis

Unless stated otherwise, all analysis was undertaken in the R software package, version 3.3.3 (R Core Team, 2016).

3.2.4 Marker resolution

To determine whether the markers used have sufficient resolution to successfully discriminate between individuals, a genotype accumulation curve was constructed using the R package `poppr` version 2.1.0 (Kamvar et al., 2015, 2014). This procedure randomly samples from 1 to $n - 1$ (where n = all loci) loci, repeated r times (r = 10 000) and records the number of unique multi-locus genotypes (MLG) observed. If the curve is asymptotic then the addition of more loci to the genotyping is unlikely to reveal many extra MLG and the resolution obtained is sufficient to determine clonal status.

3.2.5 Clonal identification

Following the recommendations of Arnaud-Haond et al. (2007), identification of clonal individuals was divided into two parts: firstly, removal of the possibility that individuals

share identical MLG as a result of chance recombination events, and secondly, confirming that dissimilar MLG actually belong to distinct genets.

To determine whether replicate genotypes are the result of vegetative reproduction or distinct zygotes, the likelihood of re-encountering each MLG multiple times as a result of sexual reproduction based on the allele frequencies present in each population was estimated using `poppr`(p_{sex}). Where p_{sex} was below 0.01 (Arnaud-Haond, Migliaccio, *et al.*, 2007), replicate MLG were considered to be ramets of the same genet and to have originated from vegetative reproduction.

Ascertaining individuality in clonal plants using microsatellite markers is also complicated by somatic mutations (Klekowski, 2003), particularly in the case of long-lived species such as *T. cordata* (Pigott, 1991), and scoring errors (Douhovnikoff and Dodd, 2003; Meirmans and Van Tienderen, 2004). Somatic mutations can cause individuals which are actually the result of vegetative reproduction events from the same parent genet to appear genetically distinct. Similarly a large number of markers, PCR artefacts, the variability of capillary electrophoresis runs and human error can all introduce variation in observed genotypes between ramets. Either scenario can result in underestimating levels of clonality. To avoid this frequency plots of genetic distance (number of distinct alleles) between individuals for each population were examined for the presence of bimodal distributions with a peak at low but non-zero distances. Such a distribution would be characteristic of either scenario (Arnaud-Haond, Duarte, *et al.*, 2007; Rozenfeld *et al.*, 2007). However, if selfing rates are high or inbreeding common in small populations, then this could also produce very similar genotypes. To assess this alternative explanation for bimodal distributions, genetic distance frequency plots from 100 simulated sexual reproductive events (allowing for selfing) were generated with `RClone` and overlaid onto observed distances.

In the presence of bimodal distributions, individuals sharing distinct but very similar genotypes were examined further to determine whether they were actually the result of clonal reproduction. Individuals differing by only one allele temporarily had the distinct locus removed and p_{sex} was recalculated. If the result was below the original threshold for clonal membership (0.01), the individuals were grouped into multi-locus lineages (MLL) and considered to be ramets of the same genet for the remainder of the analysis. This threshold of one allele was chosen as compromise between underestimation of clonal membership and

the reduction in resolution that results from effectively removing marker data (Schnittler and Eusemann, 2010).

3.2.6 Incidence of clonality

Incidence of clonality was assessed using a proportional measure of genotypic richness, *R:*

$$R = \frac{(G-1)}{(N-1)}$$

Where *G* is the number of distinct genotypes (or MLL) and *N* the total number of individuals. This varies from 0 when all sampled individuals are identical to 1 when all observed genotypes are distinct (Dorken and Eckert, 2001). It is preferable to the often used measure of proportion distinguishable (*G/N*), which never reaches 0 even in monoclonal stands (Arnaud-Haond, Duarte, *et al.*, 2007). To determine how equitable clonal reproduction in *T. cordata* is (i.e. whether many lineages reproduce vegetatively or production of clones is dominated by a few), evenness of MLL within sites was described by fitting inverse cumulative frequency data of membership of clonal group size classes to the Pareto distribution. The parameter β (-1 x regression slope) indicates how equitably clonal membership is distributed within the sampled population; a steep slope suggests that all ramets belong to similarly sized groups whereas a shallow slope represents a sample dominated by a small number of large genets (Arnaud-Haond, Duarte, *et al.*, 2007; Ardehed *et al.*, 2015; Jarni, Jakše and Brus, 2015). Other common measures of evenness (e.g. Pielou's *J*; Pielou, 1975) are strongly correlated with *R* and so redundant (Arnaud-Haond, Duarte, *et al.*, 2007).

3.2.7 Spatial analyses

Both the spatial arrangement of clonal lineages (the 'clonal architecture') and how genetic variation is positioned within space (spatial genetic structure) were examined.

Spatial arrangement is described by three statistics: the maximum linear dimension of a MLL (d_{max}), the total area covered by a convex hull containing all ramets of a MLL (A_{MLL}; Jarni et al., 2015) and an aggregation index (A_c) as described in Arnaud-Haond et al. (2007a). The aggregation index produces a value from 0 to 1, indicating that the probability of nearest-neighbour individuals being part of the same MLL does not differ from average or that they always share clonal group membership respectively. The statistical significance of the aggregation index was tested against the null hypothesis of a random distribution in

space by a permutation approach with 999 repetitions, where existing sample positions are randomly assigned to individuals using R package `RClone`, version 1.0 (Arnaud-Haond and Bailleul, 2015). For comparison of D_{max} and A_{MLL} between sites, normality of collected data was first assessed using Shapiro-Wilks tests. Thereafter nonparametric Kruskal-Wallis hypothesis tests were applied. Where these were significant a Bonferroni-corrected Dunn's multiple comparison test was used to identify where the differences lay, due to the uneven sample sizes between groups (Dunn, 1964).

Spatial genetic structure (SGS) was examined for each population using the software SPAGeDi, version 1.5 (Hardy and Vekemans, 2002). In order to assess how clonality affects SGS, two levels of analyses were undertaken: ramet and among-genet. Ramet level analysis proceeded by estimating pairwise relatedness as the kinship coefficient F_{ij} (Loiselle *et al.*, 1995) within the distance classes 0 - 8.64 m, 8.64 - 12.41 m, 12.41 - 15.84 m, 15.84 - 20.12 m, and 20.12-31.40 m. These classes were chosen in order to maintain the most equitable distribution of pairs within groups. The mean kinship coefficient per class was plotted against distance class to produce spatial autocorrelograms. To assess statistical significance of SGS for each, spatial locations were permuted 10 000 times and mean F_{ij} compared with the null hypothesis of no spatial arrangement of genotypes. Among-genet level analysis proceeded identically with the exception that between-ramet kinship coefficient scores were not utilised to calculate mean Fij, therefore reducing the inflationary effects of clonality on SGS. Other approaches such as averaging ramet positions or using the central coordinates of MLL assume isotropic growth and a lack of disturbance, as well as not utilising all spatial information (Alberto *et al.*, 2005). The point at which these two analyses intersect indicates the distance at which clonality ceases to affect SGS, or the 'clonal subrange' (Alberto *et al.*, 2005; Arnaud-Haond, Duarte, *et al.*, 2007).

To allow for simple comparisons of the degree of SGS with other work *Sp* was also calculated (Vekemans and Hardy, 2004), which is a ratio of the regression slope and average F_{ij} at the first distance class to contain all possible pairwise comparisons. Contribution of clonality to this statistic ($\%_{clonal}$) was assessed by calculating the percentage of among-genet to ramet level *Sp* (Schueler, Tusch and Scholz, 2006).

3.2.8 Relationship of levels of clonality with climatic or demographic variables

To explore what if any variables are the best predictors of the level of clonality in a population, environmental and demographic variables were regressed against incidence of

clonality (as R). Although R is continuous it is effectively a proportion and so bounded to the unit interval [0, 1] which makes typical regression models such ordinary least squares (OLS) problematic as they may predict nonsensical values (i.e. >1, <0). The logit transformation is often used when modelling proportional data using OLS (Baum, 2008) - $log_e\left(\frac{y}{[1-y]}\right)$ – but this precludes interpretation of the response parameter on its original scale. Often proportional measures can also fail to meet all of the assumptions of OLS regression, particularly homoscedasticity, as they typically show less variation near to their boundaries. To avoid these issues beta regression can be used (Ferrari and Cribari-Neto, 2004; Cribari-Neto and Zeileis, 2010). This is a derivative of generalized linear models (GLM) explicitly designed for use with proportional response data, where the dependent variable is defined in terms of the beta distribution by two parameters (mean μ, dispersion Φ). As in other GLM, additional flexibility is provided by the use of a link function between the response and independent variables. Fixed variance (Φ) models were applied in the R package `betareg` (Cribari-Neto and Zeileis, 2010), version 3.1. Since the question at hand is the strength of the relationship of particular explanatory variables to clonality, rather the generation of predictions *per se*, each of the following variables was used to generate separate regression models.

3.2.8.1 Climate

It has been suggested that the northernmost populations of *T. cordata* in the UK regenerate from seed extremely infrequently, and persist primarily due to clonal reproduction (Pigott and Huntley, 1978, 1980), possibly due to inadequate temperatures during flowering (Pigott and Huntley, 1981). This is in contrast to populations further from the northern range edge which more often set fertile seed (Pigott, 1991). Clonality should therefore increase alongside latitude and be negatively correlated with typical summer temperatures. To determine whether this was the case, a summary of local climate for each sampling location was generated from the WorldClim2 data set (Fick and Hijmans, 2017). This provided mean maximum July temperatures by day for each site for the period 1970 – 2000. This was regressed against R using a *log* link function.

3.2.8.2 Demography

Clonality has been observed to be density-dependent in other plant species (e.g. Weed and Schwarzländer, 2014) and the balance between reproductive modes has been linked to disturbance (e.g. Cristóbal et al., 2014; Johansson and Lundh, 1988), the effects of which

are likely reflected in demography (i.e. more recent disturbance will reduce the proportion of mature individuals present). To assess whether there was any relationship between density and incidence of clonality in *T. cordata* also, individuals per square meter in each sample was regressed against *R*. As the demography varied between sampled populations, the proportion of mature canopy trees (relative to saplings or understorey individuals) was assessed as a predictor. Both models used a *log-log* link function.

3.3 Results

3.3.1 Marker resolution and clonal identification

The generated genotype accumulation curve plateaued with ~7 markers, indicating that the selected loci have sufficient resolution and the number of distinct genotypes has not been underestimated (**Figure 3.2**). A total of 393 distinct MLG were identified across all 647 sampled individuals. All p_{sex} values for single and multiple re-encounters of replicate genotypes were below the chosen alpha of 0.01.

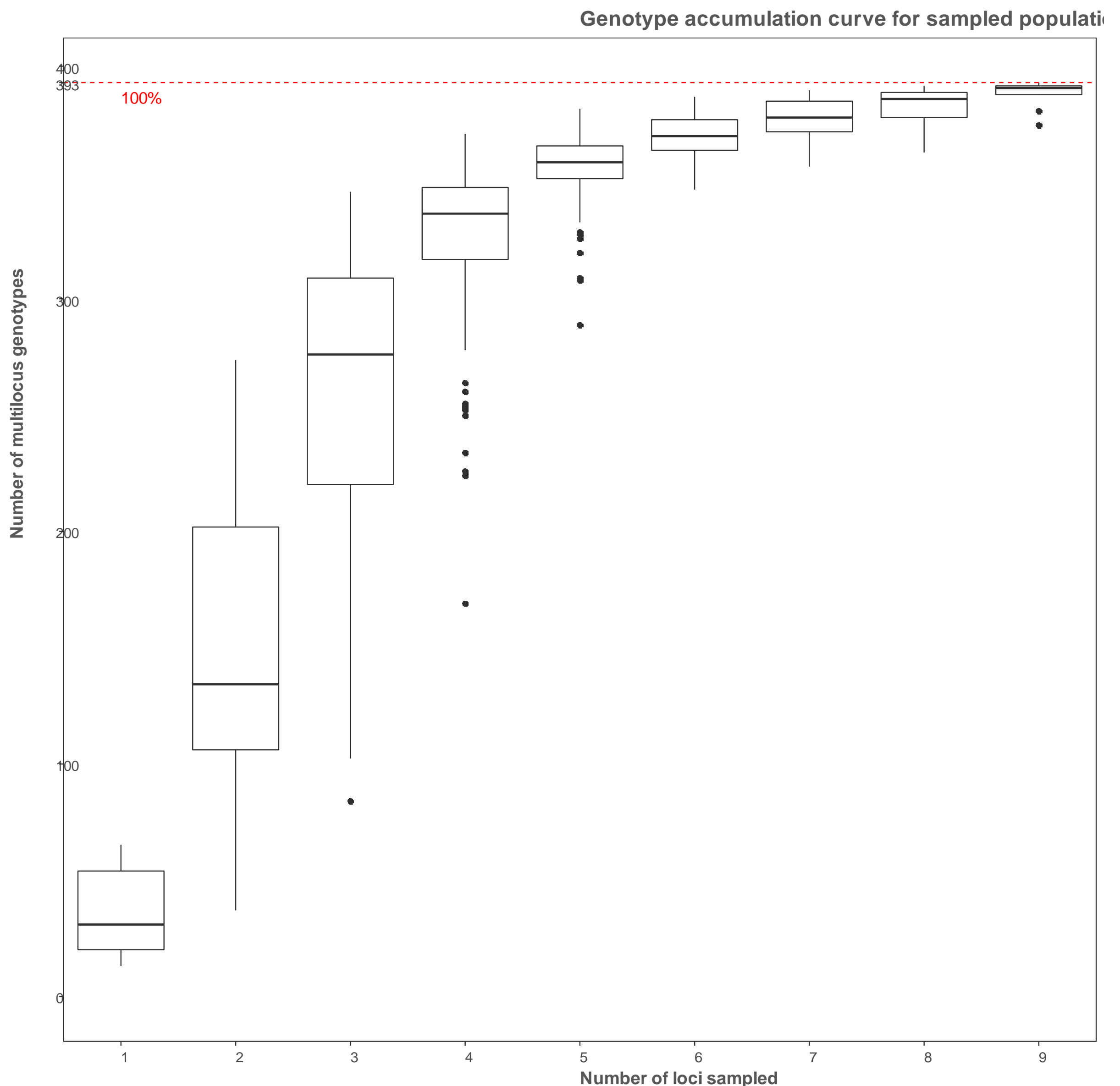


Figure 3.2: Genotype accumulation curve generated by randomly sampling from 1 to n – 1 loci (n = 10) 10 000 times and recording the number of unique genotypes observed.

Seventeen out of 22 frequency distributions of pairwise allelic distance by site were bimodal, with the smaller peak occurring at low but non-null distances (1 – 3 distinct alleles; results not shown). Simulated sexual reproductive events generated frequency distributions that did not overlap these non-null peaks (e.g. **Figure 3.3a, b;** see **Appendix I** for all plots), except where sample sizes were very low (e.g. **Figure 3.3c**). These likely represent artificially deflated levels of genetic variation due to imposed inbreeding as a result of simulating reproductive events with so few individuals. After removing the single distinct locus for each putative MLL and recalculating p_{sex}, all values were below the original threshold of 0.01, lowering the total number of distinct lineages from 393 to 361.

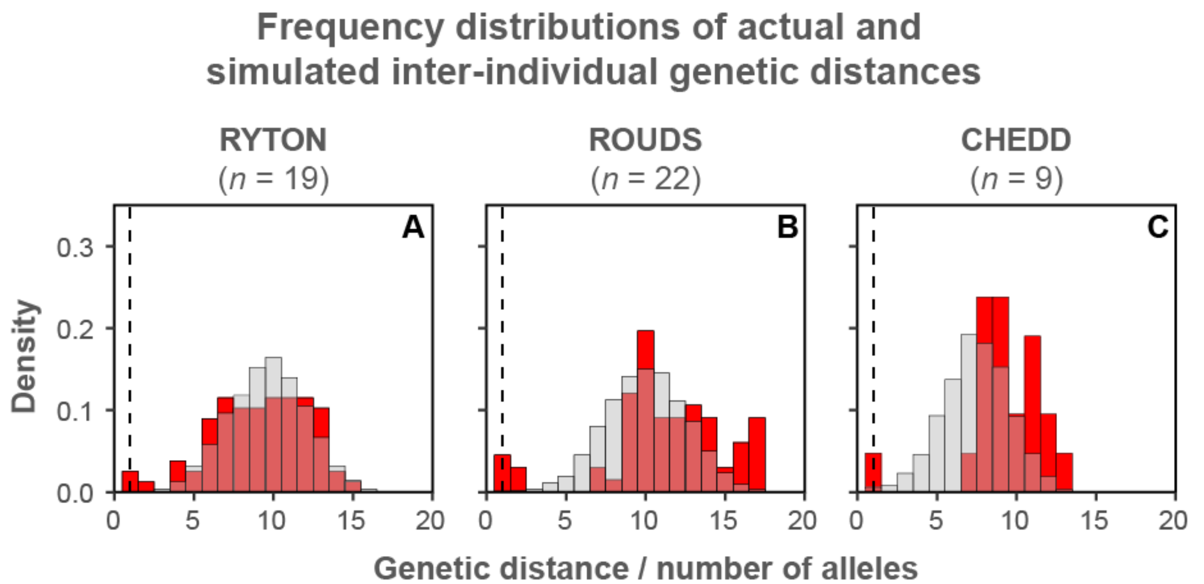


Figure 3.3: Examples of frequency distributions of pairwise genetic distance between individuals (as number of distinct alleles). Red bars are observed distances and translucent grey bars distances from 99 simulated sexual reproductive events (with selfing). The dashed line represents the threshold at which an individual was considered for inclusion in clonal lineage. **A)** and **B)** represent bimodal distributions at site 11 (Ryton Wood: *RYTON*) and site 1 (Roudsea Wood: *ROUDS*) respectively which suggest the occurrence of either somatic mutations in clonal individuals or genotyping error. **C)** is an example of small sample size creating overlap between simulated and actual genetic distance at site 19 (Cheddar Wood: *CHEDD*).

3.3.2 Clonal incidence & diversity

Clonal reproduction occurred at all sites. The number of ramets (N_R) within each genet was typically low with an overall median of two members and a maximum size of no greater than ten (**Figure 3.4**). In contrast to this large groups were found at Ivy Wood (N_R = 3 – 22 *n*).

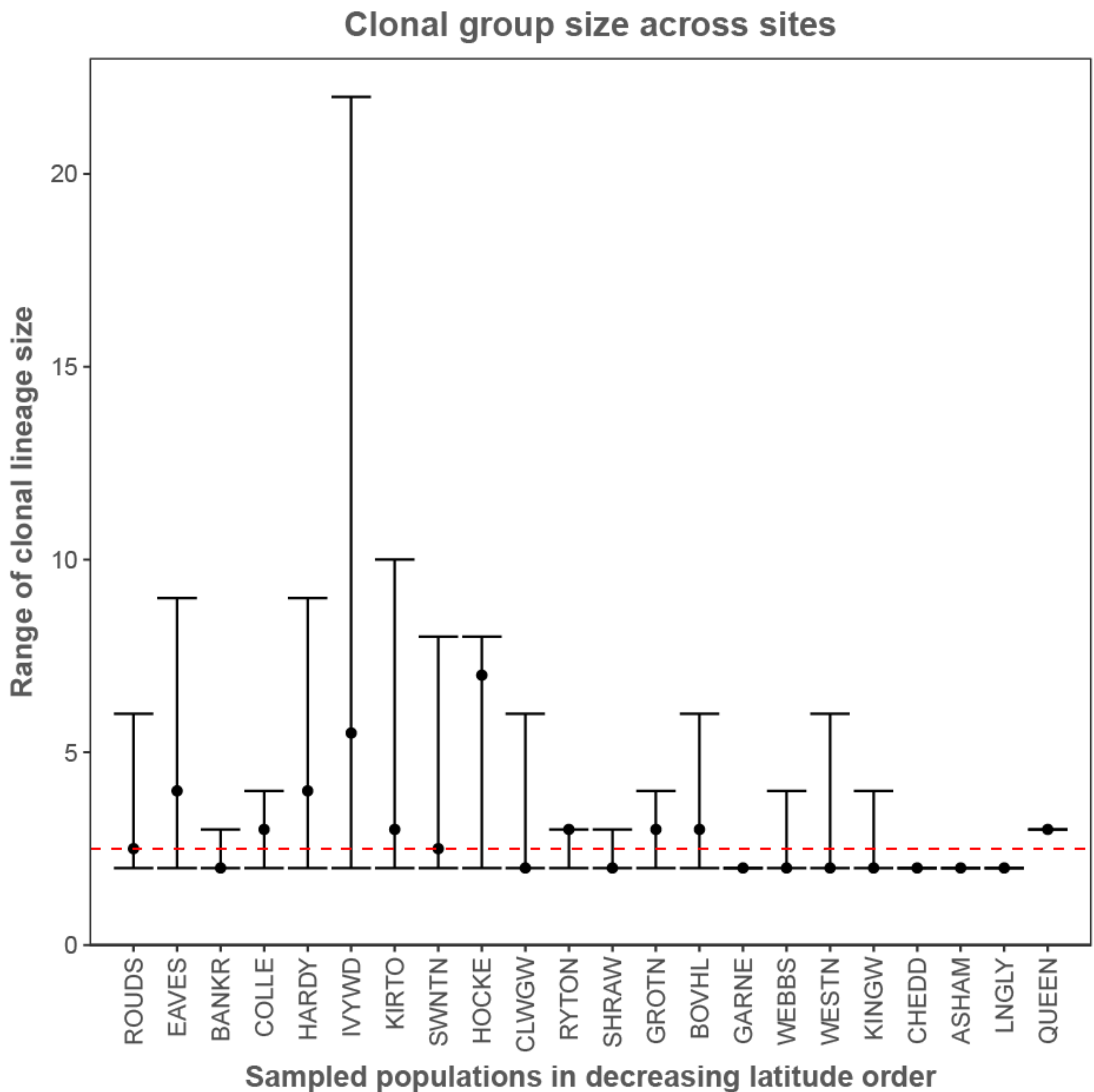


Figure 3.4: The range of sizes of clonal groups (multi-locus lineages; MLL) observed across all sites. Black circles indicate the median size of MLL while error bars indicate minimum and maximum sizes while the red dashed line indicates the overall median.

The proportion of non-clonal individuals (i.e. unique genotypes) *R* was typically around half of all trees sampled (mean = 0.57, standard deviation = 0.22). *R* ranged from 0.195 at *IVYWD* to 0.825 at *GARNE* (**Figure 3.5**).

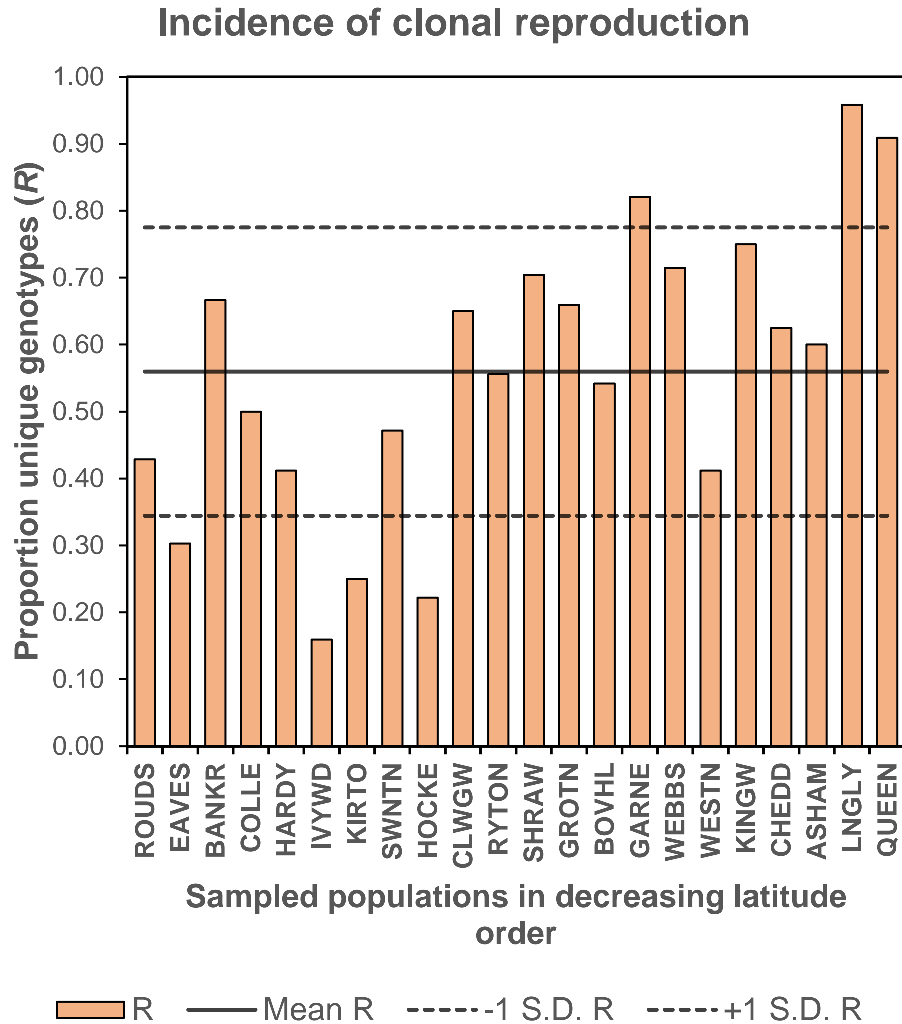
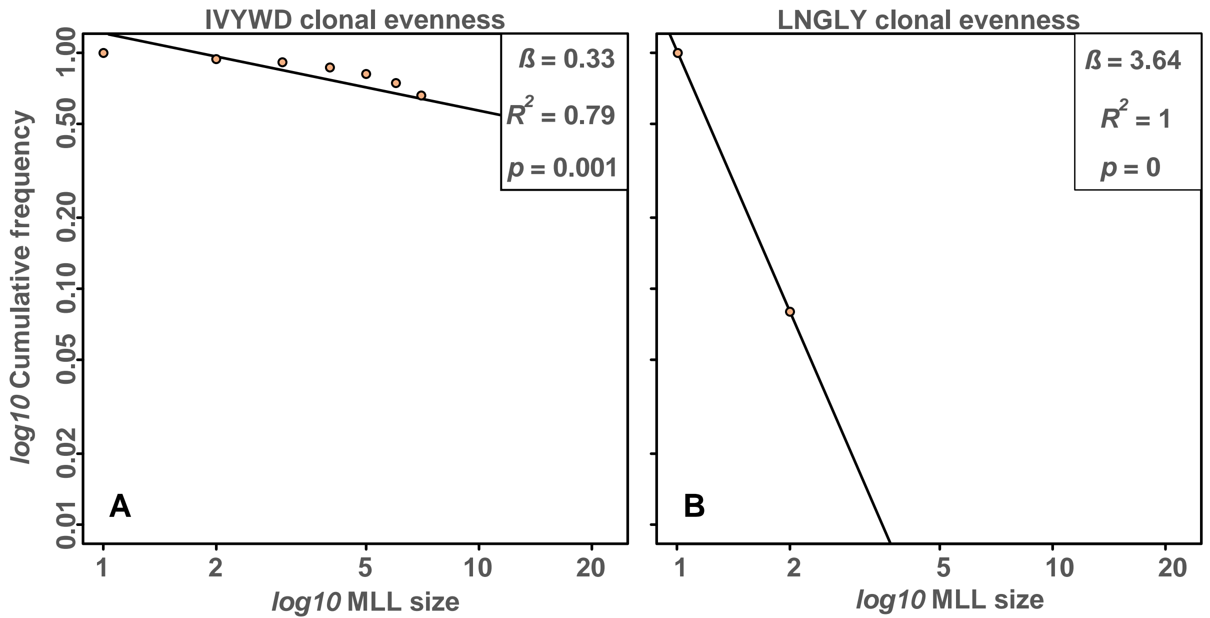


Figure 3.5: Genotypic richness *R* (Dorken and Eckert, 2001) across all sites. A score of one indicates no clonal reproduction (all genotypes are unique) whereas zero would be a monoclonal site (all genotypes identical). Mean *R* is provided by the solid black line while the top and bottom dashed black lines indicate this figure plus or minus one standard deviation respectively.

Inverse cumulative frequency plots (e.g. **Figure 3.6A, B**) of MLL group size membership show that clonal reproduction is not typically dominated by any particular lineage, with mean evenness high across sites ($\beta = 1.09$; **Figure 3.6C**). Evenness does vary between sites (standard deviation = 0.82), ranging from low (e.g. IVYWD $\beta = 0.33$) to extremely high (LNGLY $\beta = 3.64$).



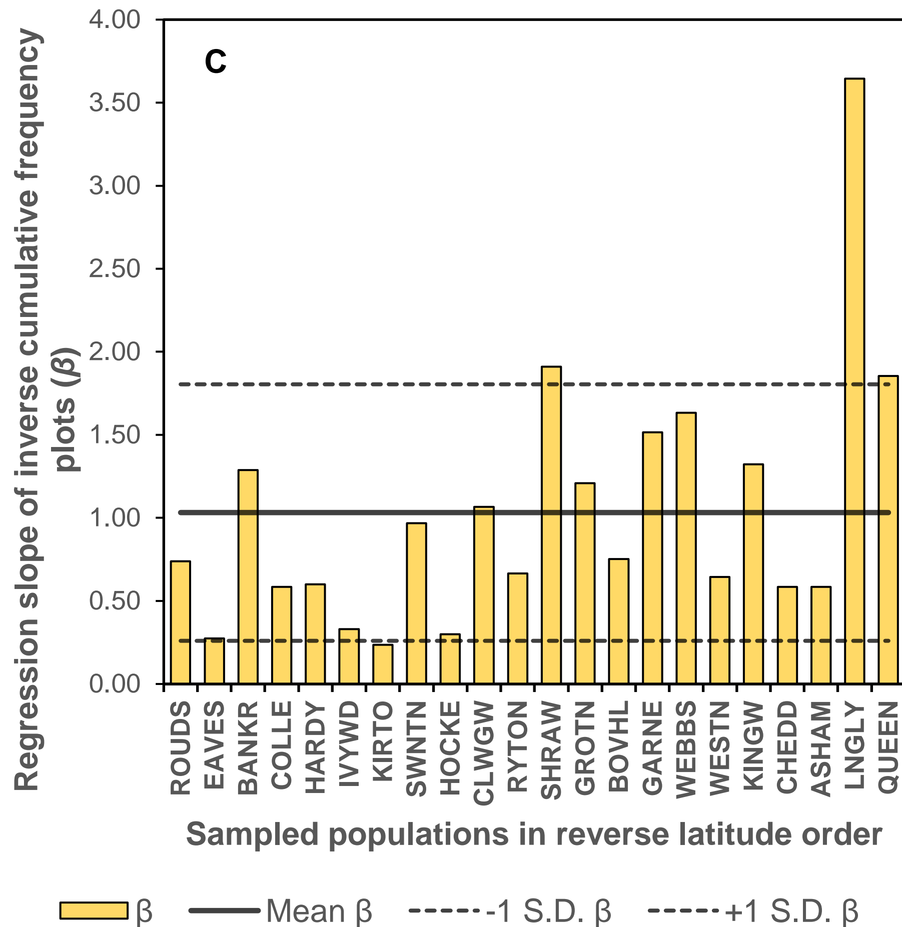


Figure 3.6: Evenness of clonal reproduction across all sampled populations, as represented by the slope of a linear regression model (β) of log inverse cumulative frequency of multi-locus lineage (MLL) group size membership. **A)** An example of a shallow slope at *IVYWD*, indicating that clonal group membership is *not* equitable; asexual reproduction is dominated by a few large lineages. **B)** An example of a steep slope at *LNGLY*, showing that asexual reproduction is not dominated by any particular lineage (all clonal MLL contain only two members). **C)** The parameter β across all sampled populations. The solid line represents mean while the dotted lines represent this +/- standard deviation.

3.3.3 Clonal architecture and spatial genetic structure

Spatial dimensions of clonal groups were typically small but had a wide range in sizes overall (**Figure 3.7**). The maximum linear distance (D_{max}) between ramets of the same genet ranged from 0.33 m (*WEBBS*) to 23.4 m (*KIRTO*) across all sites, with a median of 3.12 m. Due to low numbers of clonal groups differences in D_{max} could not be compared statistically for four sites (*ASHAM, COLLE, LNGLY, QUEEN*). For the remaining sites, D_{max} was not normally distributed at ten out of the 18 according to a Shapiro-Wilks test (results not shown).

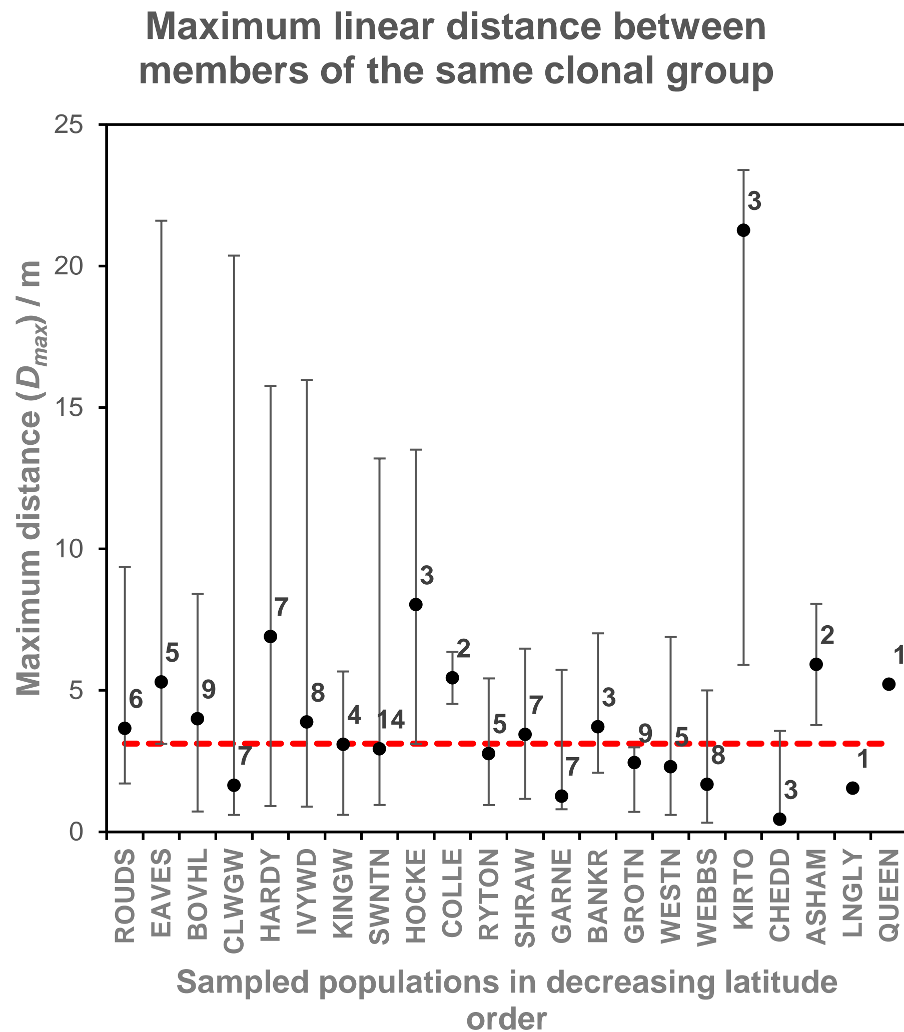


Figure 3.7: Observed values of the maximum linear distance in metres between members of the same clonal multi-locus lineage (MLL), D_{max}. Filled circles indicate median D_{max} while errors bars indicate its range. Numbers indicate how many clonal MLL were observed in that site while the dashed red line indicates overall median D_{max}.

D_{max} differed significantly across the aforementioned 18 sites according to a Kruskal-Wallis test (H = 34.03; df. = 17; $p < 0.01$), but a post-hoc Bonferroni-corrected Dunn's

multiple comparison test did not find any significant difference in pairwise comparisons between sites suggesting that an unexamined factor other than population was responsible (results not shown).

The area covered by each clonal group with more than two individuals (A_{MLL}) was typically low with a median value across all sites of 2.90 m^2, although the dispersion was high with a range of 0.02 m^2 to 137.30 m^2 due to several (spatially) large MLL at two sites (*IVYWD* and *KIRTO*). Only ten populations had at least three MLL with A_{MLL} values (**Figure 3.8**); of these two were not normally distributed according to a Shapiro-Wilks test (results not shown). A Kruskal-Wallis test indicated that clonal group area did not differ significantly between these ten sites ($H = 13.47$; df. = 9; $p = 0.14$).

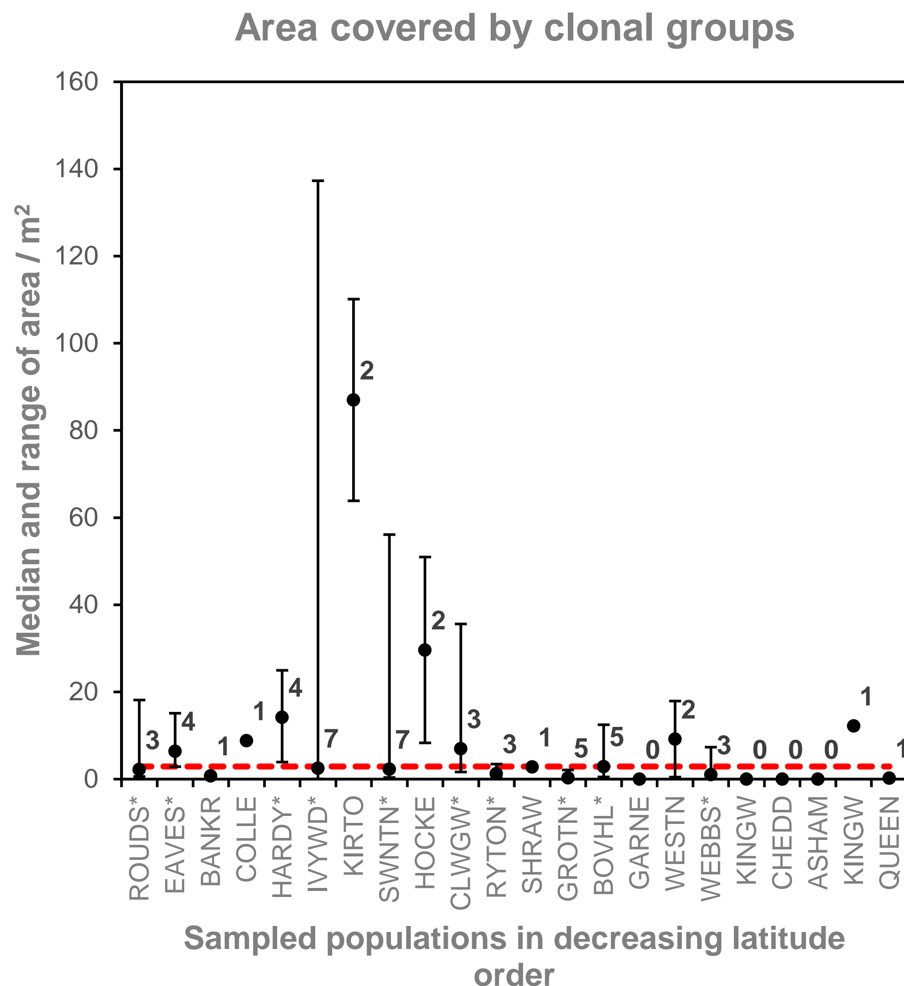


Figure 3.8: The area occupied by clonal groups (A_{MLL}) with more than two members.. Filled circles represent median A_{MLL} for that site while error bars indicate the range of A_{MLL} observed. The red dashed lines is the median A_{MLL} for all sites combined while the numbers indicate the how many clonal multi-locus lineages (MLL) contained at least three members. A_{MLL} at sites marked with an asterisk was not significantly different based on a Kruskal-Wallis test ($H = 13.47$; df. = 9; $p = 0.14$).

Aggregation of ramets within MLL was significant at sixteen out of 22 sites (**Figure 3.9**) according to a permutation test and ranged from low (*GARNE:* $A_c = 0.17$; $p < 0.05$; **Figure 3.9, Figure 3.10B**) to high (*IVYWD*: $A_c = 0.83$; $p < 0.05$; **Figure 3.9, Figure 3.10C**). Where clonal MLL contained larger numbers of ramets they often formed exclusive groups not containing other genets (e.g. **Figure 3.10C**). More typically they were arranged in pairs, often close to non-clonal individuals (mean $A_c = 0.42$; e.g. **Figure 3.10D**). Another cause of more intermingled genets was ramets separated by unusually long distances (e.g. **Figure 3.10A**). See **Appendix II** for the spatial arrangement of all sites.

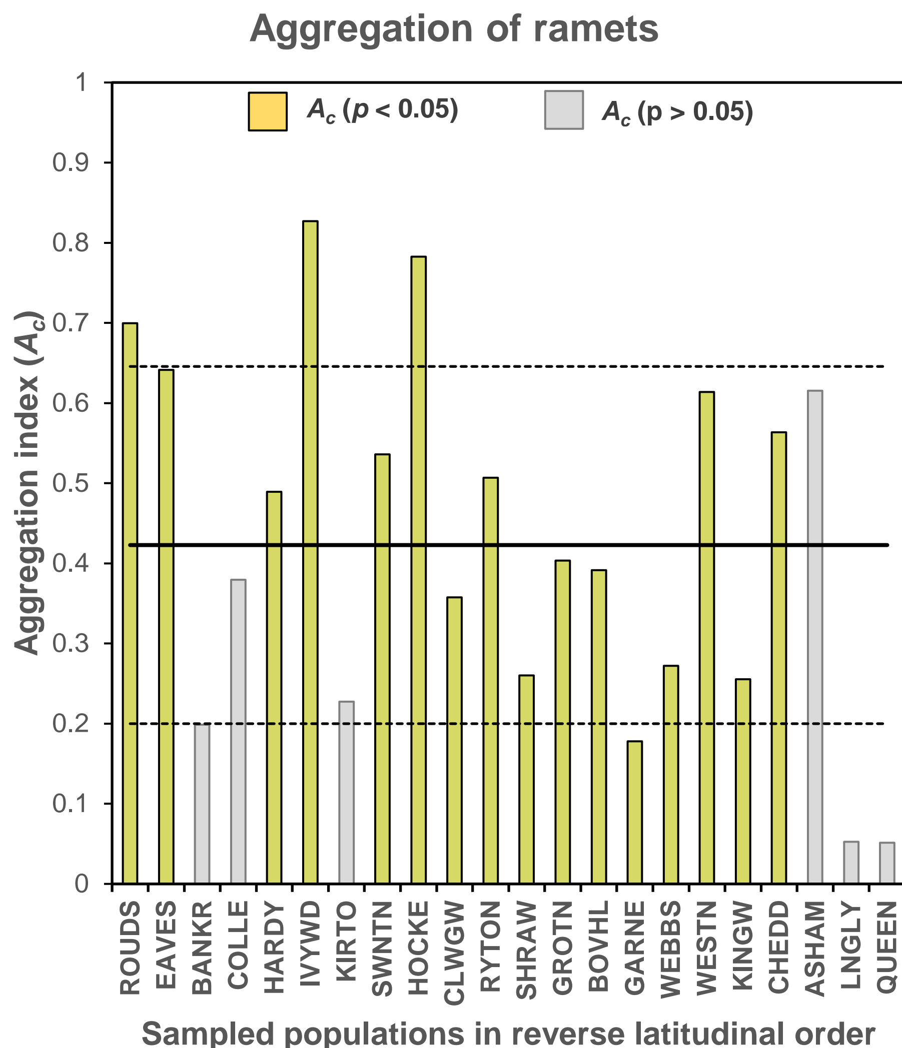


Figure 3.9: Aggregation of ramets (A_c) across all sites, which indicates how likely a ramet's nearest neighbour is to be another ramet of the same genet. Yellow bars represent significant aggregation according to a permutation test ($n = 999$; $p < 0.05$). Grey bars were not significant ($p > 0.05$). The solid black line is mean A_c across all sites while the dashed black lines represent mean A_c +/- one standard deviation.

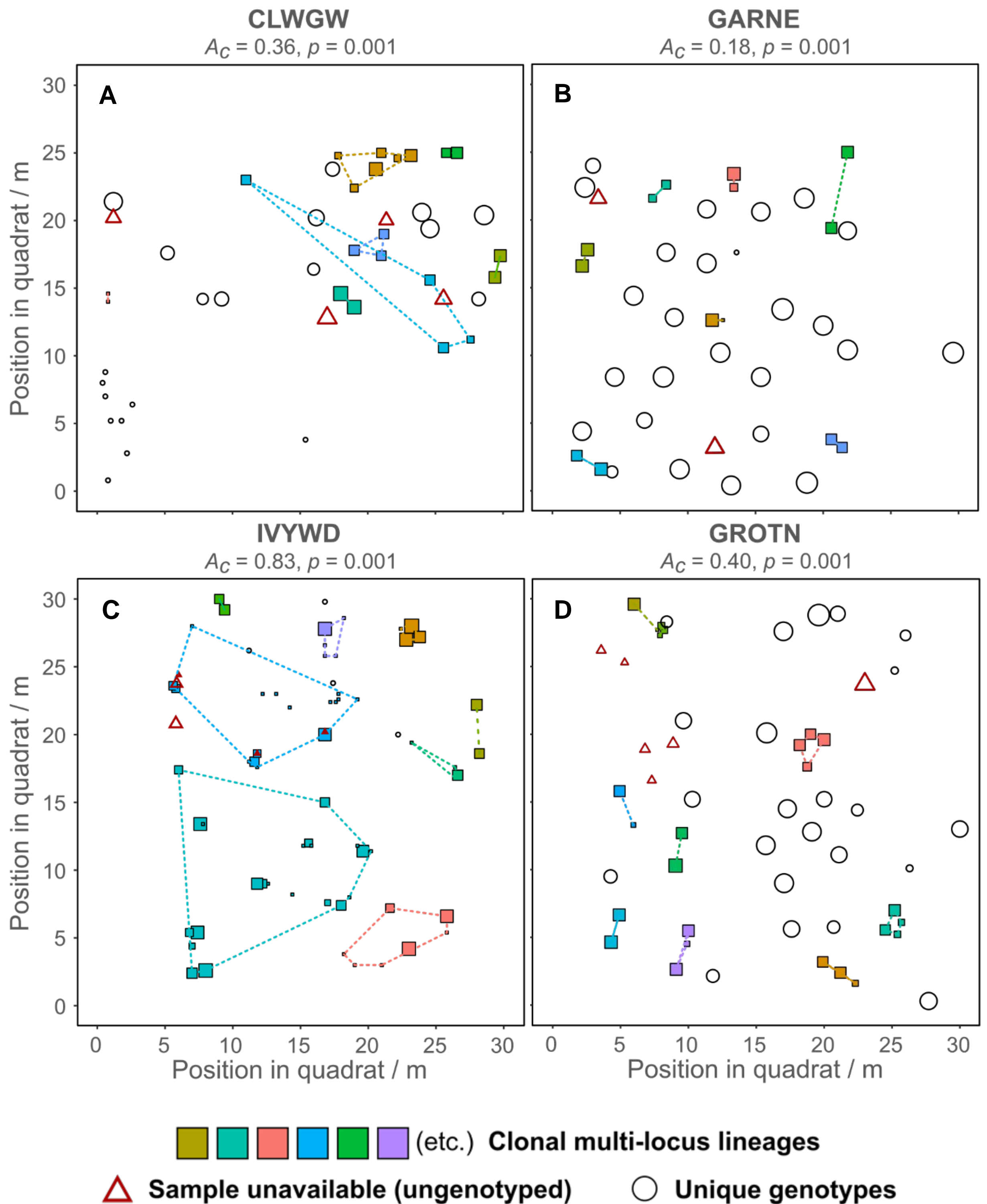


Figure 3.10: examples of the spatial arrangement of clonality at sampling sites. Unique genets are represented by hollow circles while clonal MLL groups are represented as filled squares of the same colour. Clonal multi-locus lineages (MLL) are contained within convex hulls described by dashed lines. Red hollow triangles represent individuals whose presence was recorded but not genotyped. The extent of aggregation A_c and its significance after 1 000 permutations are provided. DBH of each individual is indicated by the size of the shapes. **A)** *CLWGW*: this site has some of the largest distances between ramets of the same genet. **B)** *GARNE:* the site with lowest significant aggregation, A_c. **C)** *IVYWD*: this site has the largest MLL in both numbers and area covered as well as the highest A_c. **D)** *GROTN*: a typical site with close to average A_c and small MLL, in terms of both size and numbers.

Significant but weak SGS at both ramet and among-genet level (mean *Sp* = 0.12, 0.06 respectively) was found in fifteen of 22 sites (**Figure 3.11A**), although two sites (ROUDS and WESTN) stood out as having markedly stronger structure than the others (**Figure 3.11B, C**). Three of the remaining populations had insufficient pairs within each distance class (<10) for analysis, and are not considered here. One population, *EAVES*, showed an overall positive relationship between pairwise kinship (as F_{ij}) and distance (**Figure 3.7**, **Figure 3.11D**), possibly as a result of several clones separated by large distances; calculating *Sp* is therefore inappropriate (Vekemans and Hardy, 2004).

Clonality was typically the primary driver of SGS, explaining the majority of *Sp* (mean $\%_{clonal}$ = 51.34 %, standard deviation = 17.61 %), although this ranged from 20.61 % at *LNGLY* to 87.41 % at *IVYWD* (**Figure 3.11A**). The effect of ramets on SGS was removed beyond a very fine scale, with mean kinship for both analyses becoming identical (the clonal subrange) at the 12.41 m distance class in most instances (**Figure 3.11A**).

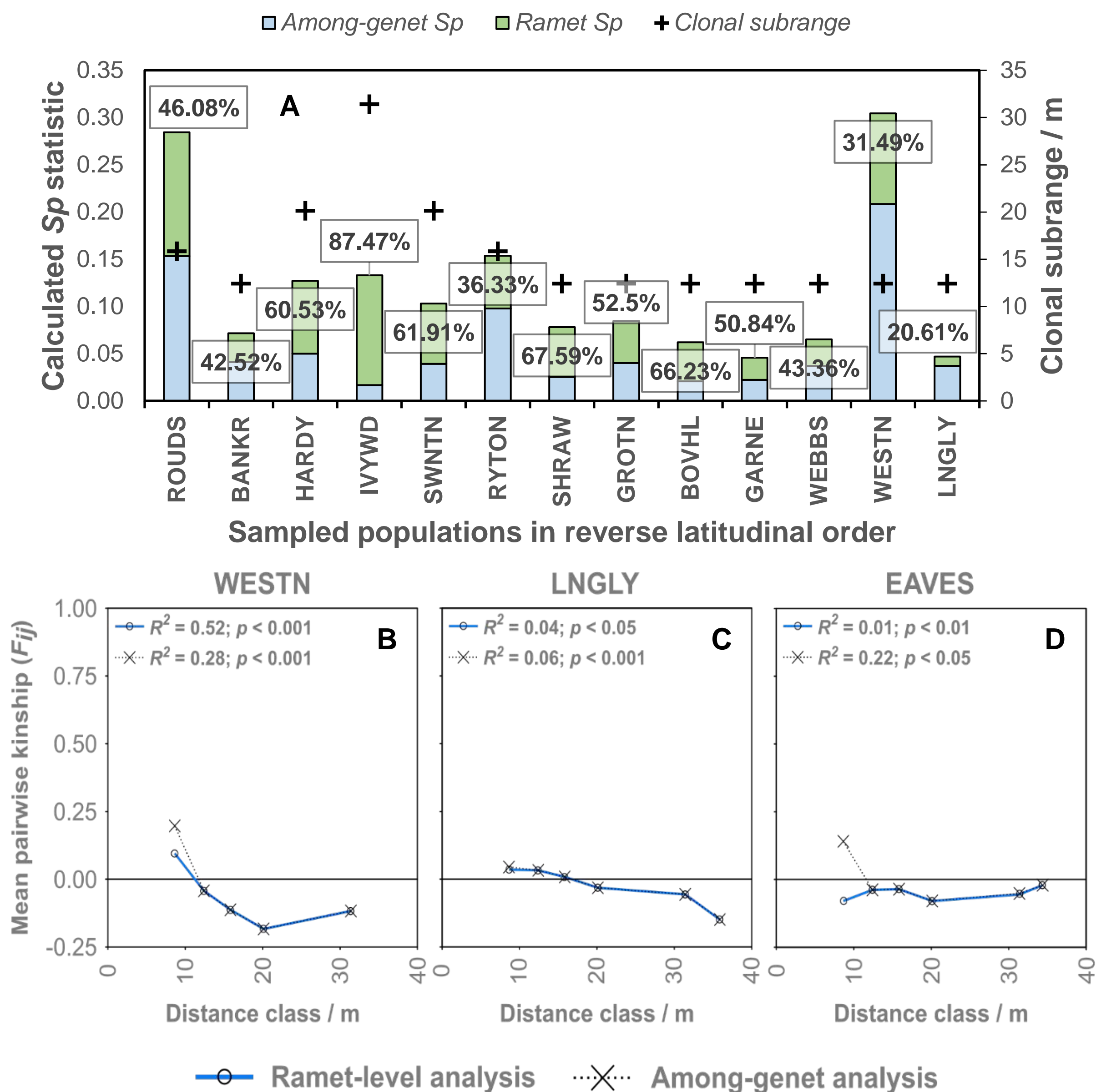


Figure 3.11: Significant spatial genetic structure (SGS) in sampled populations and the contribution of clonality thereof. Pairwise relatedness (as kinship coefficient F_{ij}; Loiselle et al., 1995) within distance classes was regressed against distance and the slope of this regression, b, is used to calculate the statistic Sp ($-b/\left[1-\widehat{F_{(1)}}\right]$, where $\widehat{F_{(1)}}$ is the mean F_{ij} in the first distance class to include all pairs (Vekemans and Hardy, 2004). High Sp indicates a sharp decline in relatedness between individuals (and therefore strong SGS). This analysis was performed between all ramets and also among genets only (i.e. discarding any kinship coefficients calculated between ramets of the same clone), in order to determine how much clonality contributes to SGS (Schueler, Tusch and Scholz, 2006). The scale of this contribution was assessed by determining the clonal subrange, the point at which these two analyses intersect (Alberto *et al.*, 2005; Arnaud-Haond, Duarte, *et al.*, 2007). **A)** Calculated Sp from regressions significant in both analyses ($p < 0.05$). Ramet-level Sp is indicated by the total height of the bars while among-genet level Sp is represented by the height of the lower blue segments. Contribution of clonality to Sp (as $\%_{clonal}$) is indicated by the floating numbers while the clonal subrange is marked on the second vertical axis by the black crosses. **B)** An example of the spatial autocorrelograms used to quantify SGS; this site, *WESTN*, has the strongest observed structure. **C)** Another example, this time illustrating the weakest SGS at *LNGLY*. **D)** Finally, at one site (*EAVES*), kinship increased with distance due to large linear distances between ramets and therefore it was not appropriate to calculate Sp.

3.3.4 Relationship of clonality with climate and demography

Both climate and demography were significant but weak predictors of the incidence of clonality (as *R*); there was no relationship between density of individuals in sample and clonality (**Table 3.3**).

Link function	Predictor	Estimate	SE	*p*
log(μ)	*Intercept*	-9.95	2.61	***
log(μ)	***Mean maximum July temperature***	0.48	0.13	***
Φ	*Intercept*	6.57	1.89	***
-log(-log[μ])	*Intercept*	0.90	0.29	**
-log(-log[μ])	***Density of individuals***	-10.26	7.23	ns.
Φ	*Intercept*	4.78	1.32	***
-log(-log[μ])	*Intercept*	-0.37	0.32	ns.
-log(-log[μ])	***Proportion of mature individuals***	1.28	0.44	*
Φ	*Intercept*	5.97	1.68	***

Table 3.3: Link functions (left-hand column) and coefficients of beta regression models as well as their standard errors (SE) and statistical significance (*** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$; ns. = $p > 0.05$).

Of the former two, typical July temperatures explained slightly more variance in *R* (pseudo-R^2: 0.17; **Figure 3.12A**) than the proportion of mature individuals (pseudo-R^2: 0.10; **Figure 3.12C**).

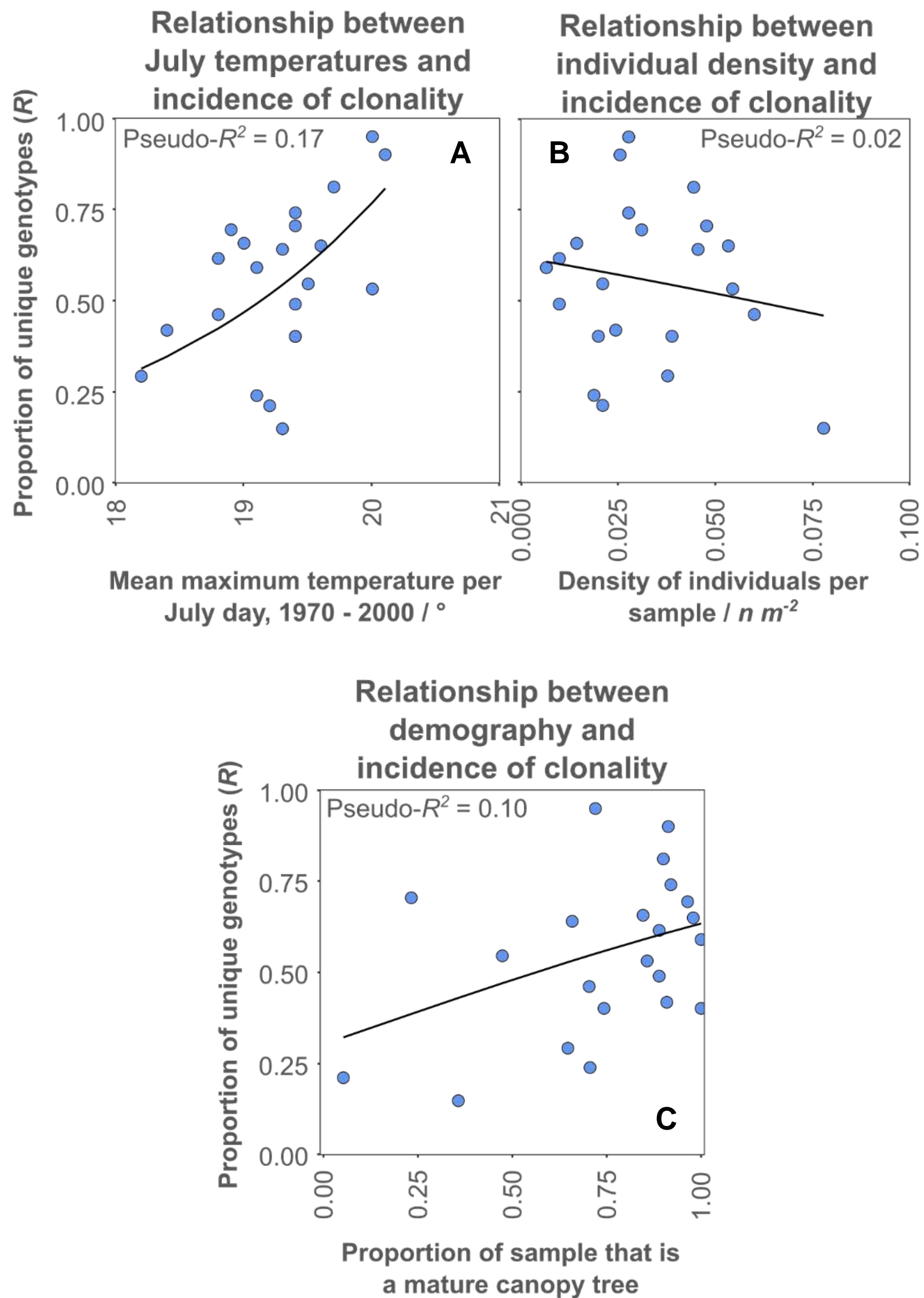


Figure 3.12: Scatter plots of predictors against incidence of clonality, R, and an estimate of how variance in the latter they explain. **A)** The relationship of summer temperatures over a period of 30 years (1970 – 2000) and clonality. Line of best fit is predicted values from the *log* link function. **B)** The relationship of individual density within each sampled quadrat and clonality. Line of best fit is predicted values from the *log*-*log* link function, $-log(-log[\mu])$. **C)** The relationship of demography (as the proportion of sampled individuals that are mature canopy trees) and clonality. Line of best fit is predicted values from the *log*-*log* link function.

3.4 Discussion

In this study it was observed that *T. cordata* is indeed partially clonal throughout its UK range, although the relative contribution of sexual versus asexual varies. Typically clonality was not dominant over recruitment from seed; on average of over half of all individuals were very likely the result of sexual reproduction. These levels of clonality are higher than the only previously reported figure. Logan et al. (2015) observed a lower incidence of clonality overall (R = 0.75), but this was a pooled figure from two species (both *T. cordata* and *T. platyphyllos*). Their study was also not designed to examine clonality, and as a consequence they used a sampling scheme which likely underestimated the incidence of vegetative reproduction, being both unstructured and not comprehensive, so this discrepancy is not unexpected. Clonality is also more common in *T. cordata* than in its North American congeneric, *T. americana* var. *caroliniana* (American basswood; Evans and Morris, 2016), which has higher genotypic richness (mean R = 0.86), although the lower bound of clonal incidence (R = 0.50) was very similar to the mean observed here. In comparison to other genera the amount of clonality was not unusual. The proportion of unique genotypes is very similar to those reported for a variety of tree species such as *Populus euphratica* (Euphrates poplar), *Ulmus minor* (field elm), and *Quercus pyrenaica* (Pyrenean oak; Buiteveld et al., 2016; Schnittler and Eusemann, 2010; Valbuena-Carabaña and Gil, 2017).

Even the samples with the highest levels of clonality (e.g. *EAVES*, *KIRTO*, *HOCKE*, and *IVYWD*) were not exclusively clonal, with at least ~16 % of the sample composed of individuals that are the result of sexual reproduction. Similarly there were no samples with a total lack of clonal individuals, although some of the most southern sites came close (e.g. *LNGLY, QUEEN*). These extremes are distinctive by their maintenance of both modes of reproduction, as many partially clonal tree species have populations with either a complete absence of clonality or monoclonal stands (e.g. Buiteveld et al., 2016; Evans and Morris, 2016; Fuentes-Utrilla et al., 2014; Morris et al., 2014), even when overall incidence is similar to that observed here.

The size of clonal groups was small with a median number of two ramets per genet. As a result clonal reproduction was characterised by high evenness, with many lineages reproducing vegetatively. There were exceptions to this, such as the sites *EAVES*, *KIRTO*, *HOCKE*, and *IVYWD,* which had a greater than average proportion of clonal individuals. These possessed several lineages with disproportionately large groups, producing lower

measures of evenness. Taken together this means that levels of genotypic diversity can remain relatively high despite ubiquitous clonality, and even where incidence of asexual reproduction is more common, some genotypic diversity remains. Similar evenness or clonal diversity is often observed in other species, with most tree genets having low numbers of ramets and correspondingly high measures of equitability or diversity (e.g. Alfonso-Corrado et al., 2005; Fuentes-Utrilla et al., 2014; Suvanto and Latva-Karjanmaa, 2005), although monoclonal stands do occur (see above).

In spatial terms clonal groups were also typically small. As expected from the typical mode of vegetative propagation (shoots from the stem base; Pigott, 1991), the distances separating ramets were typically less than 10 m, with the median maximum distance (D_{max}) between members of the same genet being just 3.2 m. D_{max} was not totally homogenous however, with significant differences across all groups (although it is unclear where these differences lie); occasionally D_{max} was greater than 20 m. As a result of the small group size and mostly low distances between ramets, genets did not occupy large areas (median A_{MLL} = 2.90 m^2) and there was no significant difference between sites. This is in contrast with other species, except where clonal groups contained many members (e.g. *IVYWD*). For example, *Populus alba* (white poplar), *Prunus avium* (wild cherry) or *Q. pyreniaca* possess ramets that are separated by greater distances and occupy larger areas than those observed here (Dering, Chybicki and Rączka, 2015; Jarni, Jakše and Brus, 2015; Valbuena-Carabaña and Gil, 2017). This is not surprising: as is common in partially clonal tree species, these examples are capable of root suckering, which is a more effective method of vegetative dispersal given the larger distances between ramets reported in these examples. Other trees have ramets separated by distances that differ by orders of magnitude (hundreds or even thousands of metres), but these are typically riparian species due to their potential for long-distance dispersal via water (e.g. Barsoum et al., 2004; Buiteveld et al., 2016).

The low distances between ramets and the small size of clonal groups meant that the aggregation index (A_c) was significant at most sites, but not particularly high (mean A_c = 0.42), meaning that the likelihood of a clone's nearest neighbour being of the same genotype was less than 50 % on average. The exception to this was again the sites with a high incidence of clonality, where the large genets formed exclusive groups not usually containing unique individuals, increasing A_c. Despite the smaller size of clonal groups in lime this aggregation was similar to that of species with comparable levels of clonality such as *U. minor* or *Athrotaxis cupressoides* (pencil pine; Buiteveld et al., 2016; Worth et al., 2016).

A significant decline in relatedness with increasing Euclidean distance was observed at thirteen of the sampled sites. A fourteenth site, *EAVES*, was observed to have the opposite relationship, probably as a result of the large linear distances between certain ramets of the same genet described above (elevating mean kinship within a particular higher distance class). This SGS was primarily influenced by clonality, with $\%_{clonal}$ greater than 50 % in a majority of sites. This did not go lower than 20.61 % at *LNGLY*, where clonality was at its lowest incidence, indicating that vegetative propagation is important in generating very fine-scale SGS in *T. cordata*. Despite this the small spatial scale of clonal groups means the influence of clonality on SGS disappears after a short distance, with the modal clonal subrange being the 12.43 m class. Most sites with significant SGS have similar *Sp* to other insect-pollinated tree species (Dering, Chybicki and Rączka, 2015), but those with strong structure (e.g. *ROUDS*, *WESTN*) have high values relative to other plant species in general (Vekemans and Hardy, 2004).

None of the examined variables (climate, density or demography) were particularly good predictors of the incidence of clonality observed. Asexual reproduction was not density-dependent as indicated by the nonsignificant regression model generated. There was a subtle trend for higher *R* values at decreasing latitudes, with mean maximum July temperatures negatively correlated with clonality as expected, but the relationship was weak (pseudo-R^2 = 0.17). The proportion of mature canopy trees in the sample was also significantly related to *R* but this relationship was even weaker than summer temperatures (pseudo-R^2 = 0.10). These results suggest that other, unexamined factors are also responsible for generating the observed patterns in incidence of clonality. That climate and demography are related to the proportion of clones in a population is not unexpected, given both the link between *T. cordata* fertility and temperature and the factors that affect clonality explored in other research. Trends in asexual reproduction related to climate have been observed in other species. For example, *Fagus grandifolia* (American beech) is more clonal at higher elevations and at more northern latitudes (Kitamura and Kawano, 2001; Morris *et al.*, 2014), *Populus tremula* (European aspen) populations have greater genotypic diversity in southern versus northern Europe (Cristóbal *et al.*, 2014), and the *Banksia ionthocarpa* subspecies which occupies more unfavourable habitats than its sister taxa is more clonal (Millar, Byrne and Coates, 2010). At the community level the proportion of species which participate in asexual reproduction is greater at higher latitudes (Ye *et al.*, 2014, 2016). Similarly the incidence of clonality differs with the maturity of *P. nigra* (black poplar) groups, with ramets

more common in middle-aged stands (Barsoum, Muller and Skot, 2004), and mature stands of *F. grandifolia* are more clonal than immature stands (Morris *et al.*, 2014).

These results show that vegetative reproduction is indeed common in *T. cordata*, although its relative importance varies. The incidence of clonality shows a latitudinal trend with southern populations having proportionally fewer clones than more northern ones. This is as predicted by the link between temperature and pollination success/seed development, and the resulting difference in fertility between northern and southern populations. It does suggest that previous work (Pigott and Huntley, 1978) may have overestimated the importance of clonality for marginal populations by reporting that they persist entirely as a result of vegetative propagation, as there is still an appreciable fraction of individuals present that are likely derived from sexual reproduction (no totally clonal populations were observed). The weak relationship between clonal incidence and summer temperatures observed here reflects this. Whether the increased clonality in more northern populations represents a by-product of sexual failure or an actual shift in reproductive strategy by *T. cordata* as a response to unfavourable conditions (i.e. expending more energy in vegetative propagation rather than sexual reproduction) is unclear (Silvertown, 2008). Just as the most clonal populations never completely lose sexually derived individuals, the least clonal southern populations were always observed to retain some clonal trees (e.g. *LNGLY, QUEEN*). Whether this holds and clonality is ubiquitous across the entire range of *T. cordata*, or clonality disappears completely, is unknown. More central populations such as those in France have been reported to regenerate freely from seed, even colonizing new habitats such as abandoned pasture, something that is very uncommon in the UK (Pigott, 1991). Similarly if asexual reproduction is a response to unfavourable conditions then the extent of clonality may also increase in marginal southern populations in the Mediterranean, although these are constrained by inadequate moisture rather than temperature. Fertility in these areas is reported to be high, but there no information on how successfully seedlings establish; vegetative propagation might again be a more successful strategy.

The principle mechanism of vegetative propagation for *T. cordata* described previously explains the clonal architecture observed here, where ramets were typically separated by small distances (median D_{max} = 3.12 m), usually less than 10 m, genets covered a small area, and as a result ramets are fairly aggregated. This is to be expected if most ramets are produced via basal shoots (rather than root suckers), and the longer distances sometimes observed are likely to be the result of layering from these shoots. The occasional relatively

large distances between clonal individuals that were observed are probably the result of treefall producing new ramets. This was observed during sampling, with a sprouting fallen tree being present at *CLWGW*.

Where SGS was observed, clonality was usually responsible for the majority of it. This could be expected to have negative effects on population persistence (Vallejo-Marín, Dorken and Barrett, 2010), by producing neighbourhoods of closely related trees. This has the potential to reduce sexual reproductive success as selfed *T. cordata* progeny exhibit inbreeding depression, having higher mortality rates in their first four years of life than outcrossed seedlings (Fromm, 2001). However, even the highest clonal subrange observed (31.4 m at *IVYWD*) is much less than the average pollen transport distance reported (~79 m; Fromm, 2001). It seems unlikely then the effects of clonality on SGS will impact future reproductive success very much in *T. cordata*, especially since just under half of the populations sampled exhibited no significant SGS at all.

Although climate explained the largest proportion of variance in incidence of clonality, the proportion of mature trees was quite similar. The relationship of demography with asexual reproduction suggests that at least part of the role of vegetative propagation in *T. cordata* is rapidly responding to disturbance. As all sampled populations are or have been subject to periodic coppicing, and since more recently coppiced locations will be characterised by proportionally younger demography (i.e. fewer matures trees), the proportion of canopy trees here probably reflects time since cutting. Coppicing is likely the highest intensity disturbance stands of *T. cordata* are subject to, with all trees within an area cut down to the base of the stem. In other species sprouting is often a response to disturbance and the intensity or frequency of this regime strongly affects the incidence of asexual reproduction (Diemer and Schmid, 2001; Morris, Small and Cruzan, 2004). For example, the frequency of and time since disturbance affects clone size in *Q. chrysolepis* (canyon oak; Montalvo et al., 1997) and the warmth and increased light levels provided by disruptions to the canopy promote root suckering in *P. tremula* (Johansson and Lundh, 1988; Cristóbal *et al.*, 2014). More specifically coppicing promotes clonality in *Q. pyrenaica* (Valbuena-Carabaña and Gil, 2017). As clonal groups of trees age, intra-clonal competition can result in self-thinning and a reduction in the size of the genets (Krasny and Johnson, 1992; Hartnett, 1993; Rautiainen, Koivula and Hyvärinen, 2004), which would decrease clonal incidence and produce the pattern observed here. This could explain why *posthoc* testing failed to

locate the difference in clone dimensions (*D_{max}*) between sites; it may lie between groups of sites that have been coppiced at similar times.

Considering this and the spatially aggregated clonal architecture (relative to other tree species) described above, it is unlikely that clonality allows *T. cordata* ramets to maximise resource use in heterogeneous environments. It probably provides additional benefits however, such as lowering mortality risk for genets. The capacity for vegetative propagation in *T. cordata* stems is reported to not diminish with age (Pigott, 1991), which can be considerable (~300 y). This likely confers exceptional longevity on individual genotypes. Clonality may also facilitate resource sharing between ramets (Barrett, 2015), although it is unknown to what extent they stay connected after above-ground material has disappeared. Additionally as *T. cordata* is insect-pollinated the production of additional stems increases floral display size, which can enhance outcrossing success and therefore fitness (Vallejo-Marín, Dorken and Barrett, 2010; Barrett, 2015). If this is the case here then a small amount of clonality should be advantageous even in areas with favourable conditions for sexual reproduction, which would explain why many lineages within populations reproduce clonally (high clonal evenness), although it should be noted that if coppicing promotes vegetative propagation in most individuals then that could also produce high equitability.

In summary, here the incidence, character and architecture of clonality in *T. cordata* across its UK range is described for the first time, as well as the effects of clonal individuals on SGS. As expected it is partially clonal throughout its UK range, although the incidence of clonality declines with latitude, potentially due to increased summer temperatures during flowering. The importance of vegetative growth in northern marginal populations may have been overestimated, although asexual reproduction is certainly more common there. These results suggest several avenues for further examining the role of asexual reproduction in *T. cordata.* Most obviously, it should be determined if the declining latitudinal trend in incidence of clonality continues into continental Europe, or if there is some level of clonality across the entire range. If southern marginal populations are included, this would indicate whether vegetative propagation is a more successful strategy when moisture regime is the limiting factor in *T. cordata*'s distribution. The role of asexual reproduction in responding to disturbance should be examined more closely by comparing incidence of clonality between populations with known differences in time since coppicing, controlling for the confounding influence of climate. Finally, the benefits (or disadvantages) of ramet

production should be explored. If clonality is a resource-sharing strategy in *T. cordata* then clonal individuals should share root systems, which can be determined via excavation work; if additional stems increase outcrossing success then clonal groups should have higher fecundity than unique individuals, which can be easily assessed by examining seeds (Pigott and Huntley, 1981).

4 Cryptic spatial genetic structure of a canopy tree (*Tilia cordata*) across a fragmented landscape

4.1 Introduction

Global terrestrial biodiversity is currently under threat due to anthropogenic disturbances such as habitat loss, degradation and fragmentation as well as rapid climatic change (Sala *et al.*, 2000; Parmesan and Yohe, 2003). Small patch sizes as a result of fragmentation mean smaller populations and higher extinction risk due to environmental or demographic stochasticity (Shaffer, 1981; Lande, 1988), by eroding genetic diversity due to increased drift and an increase in isolation (Andrén, 1994; Aguilar *et al.*, 2008). As well as increasing the effects of genetic drift lowered population size can lead to an increase in inbreeding and selfing (Schaal and Leverich, 1996; Young, Boyle and Brown, 1996). In the short term this leads to inbreeding depression via decreased heterozygosity and the expression of deleterious alleles, reducing individual fitness (Keller and Waller, 2002; Reed and Frankham, 2003). If fragments are sufficiently isolated by either distance or the landscape context then this loss of genetic variation cannot be alleviated by gene flow from other demes (Aguilar *et al.*, 2008). In the long term, lowered genetic variation can limit a species' potential for evolutionary responses to environmental change (Booy *et al.*, 2000) which may be key to persistence in a changing climate (Christmas, Breed and Lowe, 2015). It is important therefore to conserve the genetic diversity of species at risk to loss of variation from anthropogenic influences (Reed and Frankham, 2003).

Although most research examining the genetic structure of fragmented populations focuses on rare or endangered species (Honnay and Jacquemyn, 2007), common species are still susceptible to genetic erosion (e.g. Galeuchet et al., 2005; Hooftman et al., 2004; Lienert et al., 2002). However, the specific genetic response of any given species to habitat fragmentation depends on a variety of factors (Aguilar *et al.*, 2008): the characteristics of the species such as life history traits, dispersal ability, mating system, longevity and its capacity for clonal growth, etc. as well as the particular landscape context. This makes responses to habitat fragmentation idiosyncratic and difficult to generalize about, although certain groups possess traits that should make their response more or less severe. For example, woody plant species have characteristics that can make them slow to respond (Honnay and Jacquemyn, 2007; Kramer *et al.*, 2008; Vranckx *et al.*, 2012).

Individual trees and shrubs can be extremely long-lived, especially those which are able to propagate vegetatively (Honnay and Bossuyt, 2005), which allows for population persistence even in the absence of sexual reproduction (Hamrick, 2004; Lowe *et al.*, 2005). This means that for any particular landscape context, everything else being equal, woody plants will experience less drift than shorter lived species due longer generation times and lower turnover (Young, Boyle and Brown, 1996). They also typically possess high levels of phenotypic plasticity, allowing them to respond to adverse changes on a shorter timescale than an individual lifespan which may allow persistence even in very small or disturbed habitat fragments (Jump and Peñuelas, 2005). Finally their large size and correspondingly extensive pollen and seed production creates a high potential for gene flow which may counteract this loss of variation (Hamrick, 2004).

Despite this there is increasing evidence that trees and shrubs are ultimately affected by fragmentation in the same way as other species (Vranckx *et al.*, 2012). This has been demonstrated even in effective long-distance dispersers such as wind-pollinated species which might not be expected to suffer a reduction in gene flow, like beech (*Fagus sylvatica*; Jump and Peñuelas, 2006) and oaks (*Quercus* spp.; Knapp et al., 2001; Sork et al., 2002). Insect-pollinated species are particularly vulnerable as fragmentation also impacts pollinator behaviour, with pollinators typically staying within fragments (e.g. Didham et al., 1996; Goverde et al., 2002), as well reducing their population size which may limit pollen availability (Ashworth *et al.*, 2004). How a species is responding or will respond to habitat loss and fragmentation is important to understand in order to effectively undertake conservation efforts that prevent genetic erosion.

This need for genetic conservation has been recognised in *Tilia cordata*, a tree species which occupies particularly fragmented habitats. Although widespread across Europe, it has a patchy distribution. Formerly far more abundant and even a dominant species in central and Eastern European primeval woodland, it now often occupies small and isolated forest ecosystems (Pigott, 2012; De Jaegere, Hein and Claessens, 2016). This decline has been attributed to human clearance of woodland and a failure to recolonise secondary woodlands (Turner, 1962; Pigott, 1991). The lack of recolonization may be at least partly a result of low fertility, which is temperature-linked (Pigott, 1981a; Pigott and Huntley, 1981), or perhaps its low dispersal ability relative to other temperate canopy tree species (De Jaegere, Hein and Claessens, 2016). The current extent of its range was established during the Holocene climatic optimum when summer temperatures during the

critical periods of flowering and seed development were 1 – 2 °C higher than currently and its fertility was presumably not an issue (Birks, 1989; Davis *et al.*, 2003). As fertility and therefore sexual recruitment may be linked to climate, populations at different latitudes may be responding to habitat fragmentation at different rates; more sexual reproduction means more recombination and a greater opportunity for drift to remove genetic diversity.

As a result of this reduction in its distribution and some of its particular characteristics, it is predicted to be subject to the erosion of variation described above and genetic conservation schemes have already been established (Turok *et al.*, 1996). In order to do this effectively it is important to understand which areas are most at risk (Logan, Phuekvilai and Wolff, 2015). Despite this certain life history aspects of *T. cordata* suggest it will respond only slowly. It is consistently partially clonal throughout its UK range (see **Chapter 3**), meaning that individual genotypes can possess extreme longevity, preserving genetic variation over long timespans (Balloux, Lehmann and de Meeûs, 2003). On the other hand it is entomophilous and excessive or spatially aggregated clonal reproduction may promote selfing via competition with outcrossed pollen (Charpentier, 2001; Vallejo-Marín, Dorken and Barrett, 2010), adding to genetic diversity loss. This would make gene flow even more important to offset erosion of variation.

It is unclear whether a reduction in gene flow between fragments has occurred. If it has then given sufficient time drift should differentiate these via loss of rare alleles, generating spatial genetic structure (SGS) correlated with the arrangement of woodland populations inhabit. Some aspects of the genetic structure of its fragmented demes have been examined, as well as pollen movement within a contiguous habitat (Fromm, 2001; Logan, Phuekvilai and Wolff, 2015). Genetic erosion has not *yet* occurred within *T. cordata*: Logan (2015) observed high diversity across the UK range of the species. Populations present SGS on a scale of hundreds of kilometres, but this likely represents historic dispersal limitations across a formerly far more contiguous habitat (isolation by distance), and is to be expected as maximum estimated pollen movement derived from parentage analysis is only ~1.7 km (Fromm, 2001). It is unclear if habitat loss and fragmentation has limited gene flow and caused populations to become differentiated at a finer scale more relevant to realistic pollen or seed dispersal distances.

To assess this a different approach is required. Logan et al. (2015) considered each fragment to be distinct populations *a priori* and used these demes as the unit of study, but it

can be difficult to delineate what actually constitutes a distinct population (Manel *et al.*, 2003). By averaging genetic information over individuals, metrics of population differentiation based on allelic frequencies also suffer from a loss of resolution and are thus slow to respond to a reduction in gene flow (Kelly *et al.*, 2010). In a species where the genetic response to fragmentation is likely to be delayed or slow for reasons outlined above, an individual-based sampling scheme and an analysis based on inter-individual genetic distances will have increased sensitivity, which is important in order to detect the likely subtle SGS generated by either a reduction in gene flow or isolation by distance (IBD; Prunier et al., 2013). If there has been a reduction in gene flow, it is also important to understand what, if any, factors promote or resist pollen or seed dispersal between fragments (Manel and Holderegger, 2013). Knowledge of this helps to ensure genetic conservation methods are effective by highlighting which populations are most at risk to genetic erosion.

This study then has three objectives. Firstly, to examine the SGS of *T. cordata* demes using more sensitive methods and at a finer scale than previously, in order to assess its arrangement and infer how it relates to fragmentation (i.e. has there been a reduction in gene flow between distinct woodlands?). Secondly, to examine whether these patterns differ at differing latitudes. Finally, to assess whether certain habitats or land uses promote or reduce gene flow between fragments.

4.2 Methods

4.2.1 Sampling strategy and locations

Two complexes of ancient semi-natural woodland (ASNW) at contrasting latitudes within the United Kingdom were chosen for sampling: the remnants of Rockingham Forest (**RF**; **Figure 4.1A**) and Arnside & Silverdale AONB (**AS**; **Figure 4.1B**). ASNW has been consistently wooded since 1600 CE, i.e. before large-scale tree planting began, and therefore represent natural – albeit typically managed – populations. **AS** is close to the northern range edge of *Tilia cordata* in the UK and fragments here have increased levels of clonality relative to **RF** (see **Chapter 3**). Both areas possess woodland habitats of varying size and degree of isolation. An individual-based sampling scheme was used to maximise power to detect genetic structure due to the likely slow response of the species to fragmentation (Prunier *et al.*, 2013). Fragments known to contain *T. cordata* were randomly sampled with leaf material collected from maximum of 40 individuals (Hale, Burg and Steeves, 2012), with the position of each plant recorded using a handheld GPS unit (Garmin, Southamptom, UK; various models). This material was taken and stored in silica gel for later laboratory analysis. A minimum distance of 10 m between samples was maintained to reduce the collection of clonal replicates. The aggregated distribution and differing abundance of individuals within fragments is reflected by the differences in sample size and density (**Table 4.1**).

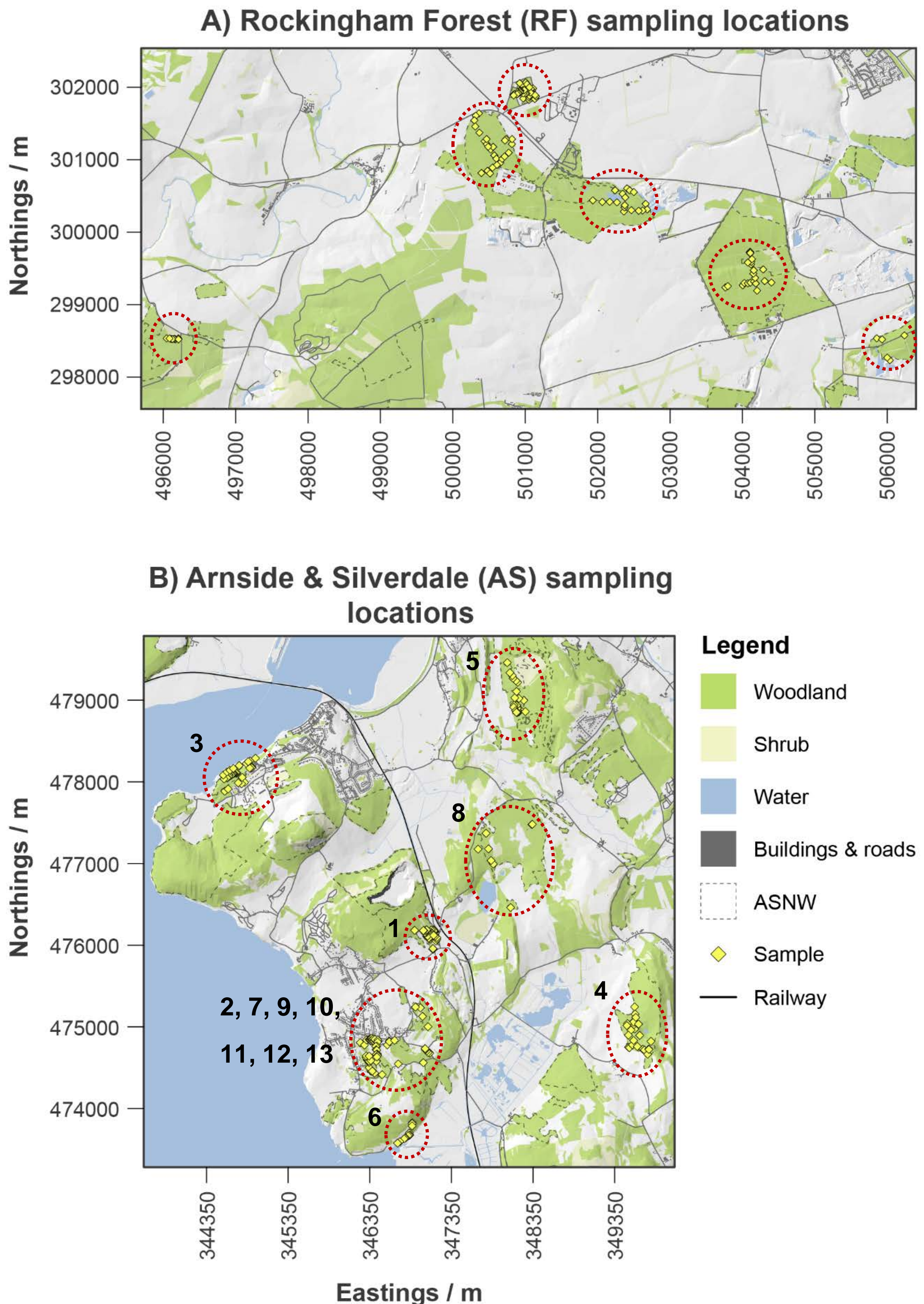


Figure 4.1: Position of sampled individuals within each landscape: **A)** the remnants of Rockingham Forest (RF), Northamptonshire, UK; **B)** Arnside and Silverdale AONB, Cumbria/Lancashire, UK. Within each the location of woodland fragments, roads and urban areas are marked. Dashed red ellipses provide a reference for the name of each sampled wood, where possible (black numbers; see **Table 4.1** for key). Coordinates are given as Ordnance Survey National Grid for ease of interpretation due to use of metres rather than degrees. "ASNW" is ancient semi-natural woodland, i.e. areas that have been continuously wooded since at least 1600 CE (before large-scale tree planting began).

		Fragment name	*n*	Eastings	Northings
AS	**1**	Eaves Wood	40	347096.0	476128.5
	2	Bottoms Wood	35	346363.7	474716.1
	3	Grubbins Wood	35	344716.0	478102.9
	4	Cringlebarrow & Deepdale	25	349614.4	474914.3
	5	Underlaid Wood	23	348142.8	479020.7
	6	Heald Brow	12	346805.9	473688.6
	7	Scout Wood	10	346440.0	474666.9
	8	Gaitbarrows	8	347961.1	477148.4
	9	Clark's Lot	6	347033.9	474721.1
	10	Burton Well Wood	4	346983.5	475153.8
	11	Sharp's Lot	4	346597.8	474817.1
	12	Misc.*	1	347004.0	474564.5
	13	Misc.*	1	346697.1	474548.9
		Total:	**204**	-	-
RF	**1**	Vigo Wood	40	501002.5	301933.1
	2	Collyweston Great Wood	27	500544.8	301169.9
	3	Easton Hornstocks	23	502756.5	300269.0
	4	Bedford Purlieus	20	504117.9	299394.8
	5	Wakerley Great Wood	9	496139.6	298526.4
	6	Old Sulehay	5	506012.1	298425.1
		Total:	**124**	-	-

Table 4.1: Sampled fragment name, sample size *n* (after removal of suspected clonal individuals), and location. Rows marked with an asterisk represent two individuals in close proximity to other fragments that were opportunistically sampled.

4.2.2 Laboratory analysis

In order to facilitate rapid genotyping, DNA from all samples was extracted using a 'crude extract' procedure described in a KAPA3G Plant PCR Kit (KAPA Biosystems, London, UK). Briefly, ~6 mm of leaf tissue was placed in 125 µl of an extraction buffer (50 mM Tris-HCl, pH 8.0; 0.1 mM EDTA•Na_2; 2% v/v ß-mercaptoethanol; 1 mM TCEP) and heated for 5 m at 95 °C. This leaf extract was then used as a template in all following PCR.

Samples were genotyped across 10 microsatellite loci following Phuekvilai and Wolff (2013) with minor modifications, using four multiplex reactions. The loci were initially identified in *T. platyphyllos* and the selection here is based on cross-amplification success and polymorphism in *T. cordata* (**Table 4.2**). Amplification was carried out in 10 µl volume using either a KAPA3G Plant PCR Kit or a Bioline MyTaq Plant-PCR Kit (Bioline Reagents Ltd., London, UK) as directed by the manufacturer, plus 0.5 mM TCEP. This

included 0.1-0.2 µM of each primer (**Table 4.2**) and 1 µl of template extract. Forward primers were fluorescently tagged (**Table 4.2**) with the dyes NED and VIC (Applied Biosystems, Warrington, UK) or 6-FAM (Sigma-Aldrich, Dorset, UK).

Locus	Multiplex Set	Concentration µM	Dye
Tc6	A	0.1	VIC
Tc920	A	0.1	6-FAM
Tc937	A	0.1	NED
Tc4	B	0.2	VIC
Tc943	B	0.2	6-FAM
Tc11	C	0.1	NED
Tc915	C	0.2	6-FAM
Tc5	D	0.2	6-FAM
Tc7	D	0.1	6-FAM
Tc951	D	0.2	NED

Table 4.2: Loci used, their multiplex groupings, concentration in reactions and the fluorescent dye used as a tag on the forward primer.

PCR product size was determined using capillary electrophoresis on an AB3500 Genetic Analyzer (Applied Biosystems, Warrington, UK). Reactions were diluted 1:10 with nuclease-free water and 1 µl of this dilution mixed with 8.9 µl Hi-Di formamide and 0.1 µl LIZ500 size standard (Life Technologies, Warrington, UK) before running. Allele peaks were called automatically in GeneMapper 5.0 (Applied Biosystems, Warrington, UK) and checked manually for errors. Raw fragment data was binned to integer allele sizes using the software TANDEM (version 1.07) in order to maintain consistency by reducing the potential for human error (Matschiner and Salzburger, 2009).

4.2.3 Statistical analysis

Unless stated otherwise all analysis was undertaken using the software R, version 3.3.3 (R Core Team, 2016).

4.2.3.1 Detection and removal of clonal individuals

The R package `poppr` version 2.3.0 was used to assess whether any individuals sampled were clonal replicates (Kamvar, Tabima and Grünwald, 2014; Kamvar, Brooks and Grünwald, 2015). Although the simplest method of identifying clonality is to check for the presence of repeated multilocus genotypes (MLG), this may over- or underestimate the number of clones sampled (Arnaud-Haond, Duarte, *et al.*, 2007). If genetic diversity is low, levels of inbreeding high or selfing common, identical MLG may be produced via sexual reproduction. Inversely, somatic mutation in long-lived clonal individuals or genotyping error may create subtly different MLG across individuals that are the product of vegetative reproduction from the same parent genotype (Douhovnikoff and Dodd, 2003; Klekowski, 2003; Meirmans and Van Tienderen, 2004).

To determine whether chance recombination events have produced identical MLG, the likelihood of re-encountering the genotypes as a result of sexual reproduction was calculated (p_{sex}; Parks and Werth, 1993). Where p_{sex} was less than 0.01 (Arnaud-Haond, Migliaccio, *et al.*, 2007), repeated MLG were considered to be clones and all but one removed from the analysis randomly (**AS**: 10 *n* removed; **RF**: 1 *n* removed).

If either somatic mutations or genotyping error have occurred, frequency distributions of genetic distance between individuals should be bimodal, with an additional small peak at low but non-null distances generated by these processes (Arnaud-Haond, Duarte, *et al.*, 2007; Rozenfeld *et al.*, 2007). This was observed and so individuals differing by only one allele (Schnittler and Eusemann, 2010) temporarily had the distinct locus removed and their p_{sex} recalculated. Where $p_{sex} < 0.01$ individuals were also considered clones and all but one randomly removed (**AS**: 5 *n* removed).

4.2.3.2 Genetic diversity and population structure

Departures from Hardy-Weinburg equilibrium at all loci were assessed across each landscape using a χ^2-test with a Monte-Carlo permutation procedure (999 replicates) to

determine statistical significance in the R package `pegas` (Paradis, 2010), version 0.9. Linkage disequilibrium (LD) was assessed using a multilocus estimate of the relationship between loci, $\overline{r}_d$ (Agapow and Burt, 2001). This is based on the index of association (I_A; Brown et al., 1980) but standardized to take into account the number of markers sampled, allowing for comparisons to be made with other studies. Departure from linkage equilibrium was tested with a permutation approach in `poppr` by comparing observed $\overline{r}_d$ with 999 permuted datasets exhibiting the $\overline{r}_D$ expected under a lack of association between loci. The presence of LD has been used as an indicator of clonality due to the linked fate of alleles during asexual reproduction (e.g. Goss et al., 2014; Guillemin et al., 2008). To assess whether the landscapes have responded differently to fragmentation, genetic variation across each was contrasted. Allelic richness for all loci (calculated in the R package `PopGenReport`; Adamack and Gruber, 2014) was compared using a Mann-Whitney *U* test (data were non-normal according to a Shapiro-Wilks test; results not shown). This measure standardizes the number of alleles to be independent of sample size, allowing for direct comparisons between different populations or studies (El Mousadik and Petit, 1996). Similarly observed heterozygosity (H_O) was compared using a *t*-test.

To assess whether individuals exhibited isolation by distance (IBD), the relationship of geographic location with genetic differentiation was assessed using distance-based redundancy analysis (dbRDA; Meirmans, 2015), implemented in the R package `vegan` (Oksanen *et al.*, 2017), version 2.4-2. Unlike the Mantel test more often utilised this approach properly decomposes genetic variance (in other words, it provides a measure of how much variation is explained by explanatory variables, i.e. geographic location) as well as possessing increased statistical power (Legendre and Fortin, 2010). This is ideal where the structure may be weak as a result of slow response times. Distanced-based RDA requires the calculation of a pairwise genetic distance metric between individuals. Percentage dissimilarity of shared alleles (1 - D_{PS}) was chosen due to its simplicity and acceptable accuracy (Bowcock *et al.*, 1994; Shirk, Landguth and Cushman, 2017).

Although dbRDA or Mantel tests both identify whether IBD is present, neither is able to describe more complex, non-clinal structures that may result from a reduction in gene flow or other processes, nor give an indication as to the scale at which this structure is present. Mantel correlograms are one solution to the latter problem (Borcard and Legendre, 2012) but they are difficult to interpret and their principle assumption (second-order stationarity, i.e. variance is same across the landscape – the structure can be described by

one function) is not guaranteed to be met in natural populations (Legendre and Legendre, 1998). Bayesian clustering methods are also often used for the purpose of describing the arrangement of SGS (e.g. STRUCTURE, Geneland, etc.; Guillot, 2008; Pritchard et al., 2000) but as they operate by assigning individuals to one of an arbitrary number of population clusters, they may not be suitable if population structure is clinal (as in IBD). Further they are model-based and their assumptions (typically that populations will be under Hardy-Weinburg and linkage equilibrium) will not necessarily be met in a partially clonal species such as *T. cordata,* given the linked fate of alleles when reproducing genotypes identically as touched on above. A method of extracting and describing spatially-structured genetic variation that is not subject to the same assumptions is spatial principal component analysis (sPCA; Jombart et al., 2008). This effectively summarises spatially autocorrelated allelic differences between individuals. Therefore in order to describe the arrangement of SGS across both landscapes sPCA was implemented using the R package `adegenet` (Jombart, 2008; Jombart and Ahmed, 2011), version 2.0.1. As maximum pollen transport distances have been estimated empirically using data from parentage analysis (Fromm, 2001), a neighbourhood-by-distance connection network was used so that only individuals within a distance of 1666.4 m were considered neighbours. In **AS**, three positive axes were retained, while in **RF** only two positive axes were retained. These decisions were based on both variance explained and spatial autocorrelation of each relative to the others (Jombart *et al.*, 2008), i.e. the composite eigenvalues were clearly distinct for these components.

4.2.3.3 Relationship of landscape factors with population structure

4.2.3.3.1 Assessment of relationship

In addition to IBD, a wide variety of landscape factors can work in concert to generate spatially-structured genetic variation by resisting gene flow between fragmented habitats (e.g. Cushman et al., 2006). This relationship is often assessed using partial Mantel tests or similar to examine the correlation of pairwise individual or population genetic distance with pairwise landscape resistance values. This approach requires the generation of properly parameterised resistance surfaces that describe how various landscape variables restrict gene flow (and their weighting relative to each other), and errors in their creation can lead to misleading conclusions (Spear *et al.*, 2010). To remove the need for this step, partial dbRDA was used assess the relationship of genetic distance to landscape variables across both areas. This is a constrained ordination technique and the multivariate analogue of

multiple regression (Legendre and Legendre, 1998), allowing for the use of raw data describing landscape variables hypothesised to affect gene flow. It allows for specification of conditioning variables whose effects are ‘partialled out’, i.e. the variance that is explained by these variables is specified separately in the model. In this instance, geographic location (as eastings and northings) were used as conditioning variables to control for the effects of spatial autocorrelation in genetic data generated by IBD, allowing the relationship of the landscape context and genetic distance alone to be considered.

4.2.3.3.2 Initial validation

Partial RDA can have elevated type I error rates (false positives) in individual-based studies due to spatial autocorrelation in landscape variables (Kierepka and Latch, 2015). To assess how likely erroneous conclusions were to be drawn, the error rate for this particular study was estimated using simulations in a similar manner to Kierepka and Latch (2016). Firstly, the software EASYPOP (Balloux, 2001) was used to generate 100 independent replicates of spatially-structured population genetics data that exhibited IBD but whose genetic structure was not influenced by landscape variables. These simulations included thirteen (**AS**) or six (**RF**) populations with a location provided by the mean location of actual sampled fragments (**Table 4.3**).

	Population	X	Y
AS	*1*	1.68	1.14
	2	2.30	1.51
	3	2.35	1.06
	4	4.91	1.23
	5	2.43	2.48
	6	2.32	0.89
	7	3.17	3.51
	8	0.00	4.43
	9	2.02	0.88
	10	2.15	0.00
	11	1.76	1.00
	12	1.91	1.15
	13	3.47	5.26
RF	*1*	7.98	0.97
	2	4.41	2.74
	3	6.62	1.84
	4	9.87	0.00
	5	4.86	3.51
	6	0.00	0.10

Table 4.3: Location in kilometres of each population within all replicates of simulated data. Relative distance between simulated and actual fragments is the same.

Each population consisted of 100 diploid hermaphroditic individuals with a clonal reproduction rate of 30% (see **Chapter 3**), a selfing rate of 30% (Fromm, 2001) and the same spatially explicit migration scheme throughout the simulation (mean dispersal distance: 79.8 m; *ibid.*). To ensure comparable levels of genetic diversity were generated, 10 loci were specified with an initial 10 alleles each. The simulation was run for either 10 (**AS**) or 20 (**RF**) generations, with the difference in length due to presumed lower frequency of sexual reproduction within northern *T. cordata* populations (Pigott and Huntley, 1981; Pigott, 1991). Full parameters are given in **Table 4.4**.

Description	Value
Ploidy level	Diploid
Two sexes	N
Random mating	N
Proportion of clonal reproduction	0.3*
Selfing rate	0.3†
Number of populations	13 (**AS**) / 6 (**RF**)
Same number of individuals per population	Y
Number of individuals	100
Migration scheme does not change	Y
Migration model	Spatial
Proportion of migration	0.05
Number of dimensions that define space	2
Mean dispersal distance	0.0798 km†
Number of loci	10
Free recombination between loci	Y
All loci possess same mutation scheme	Y
Mutation rate	0.00005‡
Mutation model	SMM
Number of possible allelic states	10
Initial variability	Maximal
Number of generations	10 (**AS**) / 20 (**RF**)
Number of individuals in each sample	75

Table 4.4: All parameters used to generate simulated data. The clonal proportion (*) is derived from (see **Chapter 3**). Selfing-rate and mean dispersal distance (†) are both taken from empirical estimates of pollen movement derived from parentage analysis (Fromm, 2001). Mutation rate of *T. cordata* microsatellites (‡) is unknown and so a conservative value observed in another tree species was used (Provan *et al.*, 1999).

Once complete, a total of 75 n per population were retained, clonal individuals removed, and further randomly subsampled to match their respective fragments (**Table 4.1**). Individuals were randomly assigned a position from the actual samples within said fragments. Presence of IBD across all 100 simulated replicates in each landscape was verified using dbRDA (**AS**: median $R^2 = 0.02$, all $p < 0.01$; **RF**: median $R^2 = 0.03$, all $p < 0.01$).

The relationship of landscape variables to each of the simulated datasets was then examined using the final partial dbRDA models described above. The proportion of false positives (i.e. detection of a relationship between landscape factors and genetic distance where none could exist) was recorded to determine type I error rate. A high rate (greater than 5 %) suggested that spatial autocorrelation remained despite conditioning out the effects of geographic location (Kierepka and Latch, 2016). To avoid reaching erroneous conclusions, the sampling density was reduced before assessing the relationship of landscape variables to genetic distance. Individuals were randomly removed until no overlap between sampling radii occurred. The produced subsets at **AS** and **RF** containing 31 and 25 *n* respectively. Only these subsets were used in the final model.

4.2.3.3.3 Choice of landscape variables

As an entomophilous plant species, gene flow in *T. cordata* could be affected twice by landscape factors: once by influences on seed dispersal and then again by changes to pollinator abundance and behaviour, and these factors may be different (Holderegger *et al.*, 2010). Both processes were therefore considered when assessing which variables might influence genetic structure. A total of four topographical and three land use/cover variables considered likely to affect gene flow were derived from a variety of sources (**Table 4.5**) using the software QGIS, version 2.18.11 (QGIS Development Team, 2016). These data were sampled in a 79 m radius around each individual (average estimated pollen movement distance; Fromm, 2001) using the R package `raster` (Hijmans, 2016). This produced a data table with seven total explanatory variables describing the topography or land cover around each individual which could be regressed against inter-individual genetic distance using geographic location (X/Y) as conditioning variables to control for the effects of IBD.

Variable	Description	Summary statistic	Data source(s)	Stage of effect	Potential effect
Altitude	Terrain altitude in meters	Mean	OS Terrain 5 DTM*	Seed	Changes in prevailing winds
Slope	Terrain slope in degrees	Mean	OS Terrain 5 DTM*	Pollen	Increasing slope may increase *Bombus* nest abundance
cos(Aspect)	Cosine of aspect	Mean	OS Terrain 5 DTM*	Seed/Pollen	Changes in prevailing winds; certain UK *Bombus* spp. prefer north-facing banks as nest sites
sin(Aspect)	Sine of aspect	Mean	OS Terrain 5 DTM*	Seed	Changes in prevailing winds
Urban	Percent of sampling radius occupied by roads or buildings	Percent cover	OS Open Map Local†	Pollen	Urban gardens promote *Bombus* nest abundance; paved areas may discourage *Bombus* movement
Woodland	Percent of sampling radius occupied by woodland	Percent cover	OS Open Map Local†, OS VectorMap Local‡, Natural England Priority Habitat Inventory§	Pollen	*Bombus* abundance has been observed to be negatively correlated with forest cover
Open	Percent of sampling radius occupied by open habitats with high nesting or floristic value	Percent cover	OS VectorMap Local‡, Natural England Priority Habitat Inventory§	Pollen	*Bombus* abundance is correlated with either floristic abundance or richness; some spp. prefer to nest in tussocky habitats

Table 4.5: *The landscape variables chosen and their sources, as well as stage at which they are hypothesised to affect gene flow and how. *: Ordnance Survey (GB) (2016). †: Ordnance Survey (GB) (2016b). ‡: Ordnance Survey (GB) (2017). §: Natural England (2016).*

T. cordata seeds are wind-dispersed and typically fall within 100 m of the parent tree (Pigott, 1991). Topography affects prevailing winds and even low hills can affect airflow within and above the canopy (Belcher, Finnigan and Harman, 2008; Lapworth and McGregor, 2008). It is therefore important to consider terrain in the analysis (Trakhtenbrot, Katul and Nathan, 2014). Three of four topographic variables were chosen for this reason: average altitude in metres and two aspect variables. Given the circular nature of aspect (i.e. 359° is likely to be ecologically almost identical to 0° and yet is numerically distinct), it was decomposed into two variables: sin(aspect) and cos(aspect) (Legendre and Legendre, 1998).

The remaining variables (one topographic and five land cover/use) were all chosen due to presumed effects on gene flow via influences on pollinator abundance or movement. *T. cordata* flowers are visited by a diverse array of insects both diurnal and nocturnal, including but not limited to hoverflies (Syrphidae) and bumblebees (*Bombus* spp.) which are abundant during the day and a variety of moth species during the night (Pigott, 1991). This diversity of potential pollinators is undoubtedly reflected in a diversity of potential landscape influences on gene flow, making it necessary for practical reasons to focus on only the most important taxa to *T. cordata*. *Bombus* spp. are major pollinators of a variety of plant families, including the Malvaceae so it is likely that bumblebees are influential in this specific case as well (Corbet, Williams and Osborne, 1991). A variety of *Bombus* species (*B. lucorum*, *B. pratorum*, *B. pascuorum*, etc.) are reported as common visitors to *T. cordata* and its flowers are a major pollen source for *B. lucorum* (Free, 1970; Corbet, Unwin and Prŷr-Jones, 1979; Pigott, 1991). The genus is also relatively well-studied and therefore only landscape factors likely to affect bumblebee abundance or behaviour are considered here.

Land use or habitat which promotes or discourages *Bombus* nesting and therefore abundance were included as explanatory variables. *B. lucorum* and *B. lapidarius* queens are associated with habitats containing banks due to a preference for subterranean nest sites (Kells and Goulson, 2003) and thus average terrain slope in degrees within each buffer was included. Similarly, UK *Bombus* species prefer north-facing banks with loose soil for nests (Alford, 1975). Other *Bombus* species that visit *T. cordata* (*B. pascuorum* and *B. hortorum)* prefer nesting in habitats containing tussock-forming grasses (Kells and Goulson, 2003) and so a composite variable describing the coverage of open, semi-natural habitats such as calcareous grassland was created (**Table 4.5**); percent cover of these within each buffer was included. Paved areas such as roads negatively impact nesting density in *B. vosnesenskii* (Jha and Kremen, 2013) but conversely urban gardens have higher *Bombus* nest density than

woodland or grassland in the UK (Osborne *et al.*, 2008), and therefore percent cover of urban areas (either roads or buildings) was included in the model.

Finally, habitats which affect the abundance of *Bombus* individuals were included as explanatory variables. Both floral abundance (Bennett et al., 2014; Grundel et al., 2010) and diversity (Ghazoul, 2006; Jha and Kremen, 2013) have been suggested as important to bee abundance. The open semi-natural areas described above are expected to have a high availability of flowers during the flowering period of *T. cordata.* Canopy cover has been negatively correlated with bee abundance (Grundel *et al.*, 2010) and so the percent cover of wooded areas in each buffer was also included.

To avoid collinearity in the explanatory variables, two steps were taken. Firstly a pairwise correlation matrix was calculated for each landscape; where the Pearson's correlation coefficient had an absolute value greater than 0.7, one of the variables was removed from the analysis based on ease of interpretation (**AS**: only altitude and slope, latter retained; **RF**: none). Secondly, the remaining variables were used in an initial partial dbRDA model (conditioned by geographic location to remove effects of IBD) and from this, a variance inflation factor (VIF) for each was calculated in `vegan`. All VIF were less than five in both instances, indicating low levels of collinearity. These initial models were tested for significance using a permutational approach (999 permutations; Legendre et al., 2011).

4.3 Results

4.3.1 Genetic diversity and structure

Overall genetic diversity was quite high in both landscapes. Sampled loci had high polymorphism at both locations with an average of 10.3 alleles per locus (**Figure 4.2**) and a Mann-Whitney U test found no significant difference in allelic richness (R_S) between them ($U = 52$, $n = 10$, $p = 0.92$). Heterozygosity was reasonably high overall (**Figure 4.3**), with over 50% of individuals heterozygous at most loci in both **AS** (mean $H_O = 0.61$) and **RF** (mean $H_O = 0.63$). There was no significant difference in H_O between sites ($t = 0.31$, df = 18, $p = 0.88$). There is a deficiency of heterozygotes at **AS** (**Figure 4.3A**), with a paired t-test finding a significant difference between observed and expected heterozygosity ($t = 4.89$, df = 9, $p < 0.001$), but not at **RF** ($t = 2.05$, df = 9, $p = 0.07$; **Figure 4.3B**).

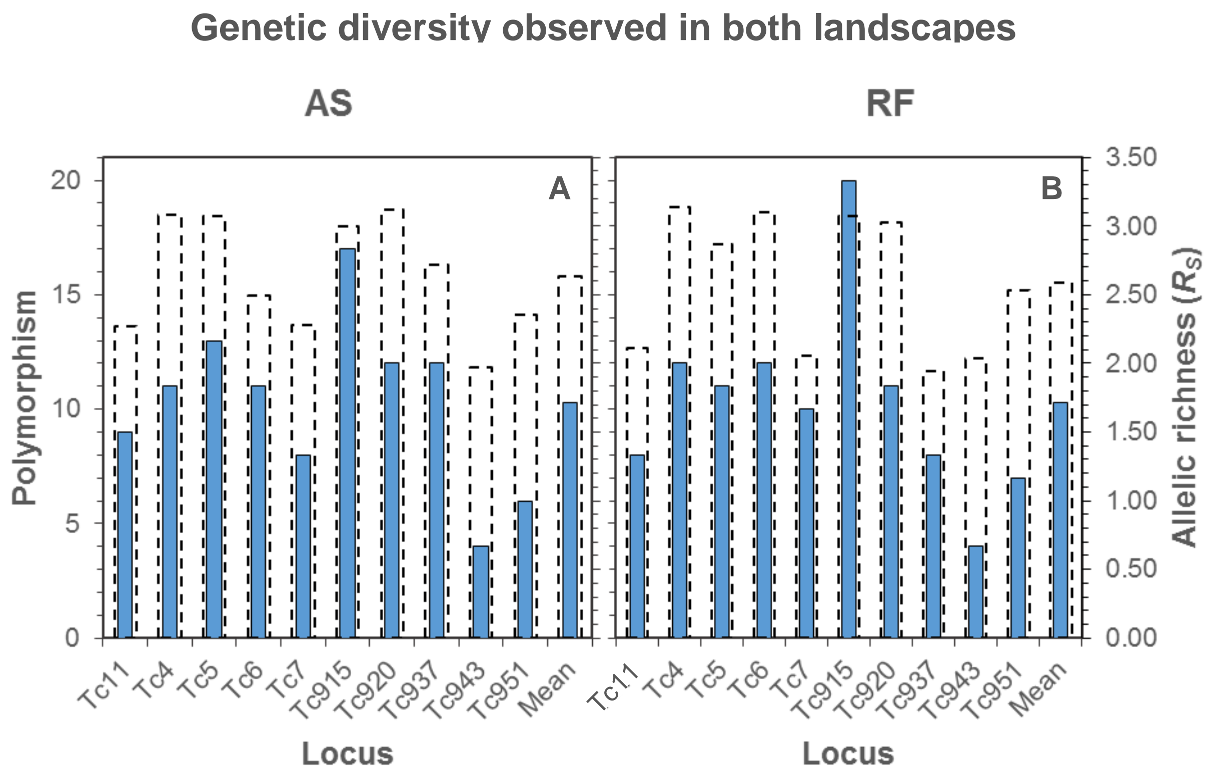


Figure 4.2: Polymorphism and allelic richness for each locus in both landscapes. Polymorphism (solid blue bars, primary axis) is the number of alleles at each locus, while allelic richness (R_S) is a rarefied measure of the same (El Mousadik and Petit, 1996) that is independent of sample size (dashed bars, secondary axis), allowing landscapes to be compared.

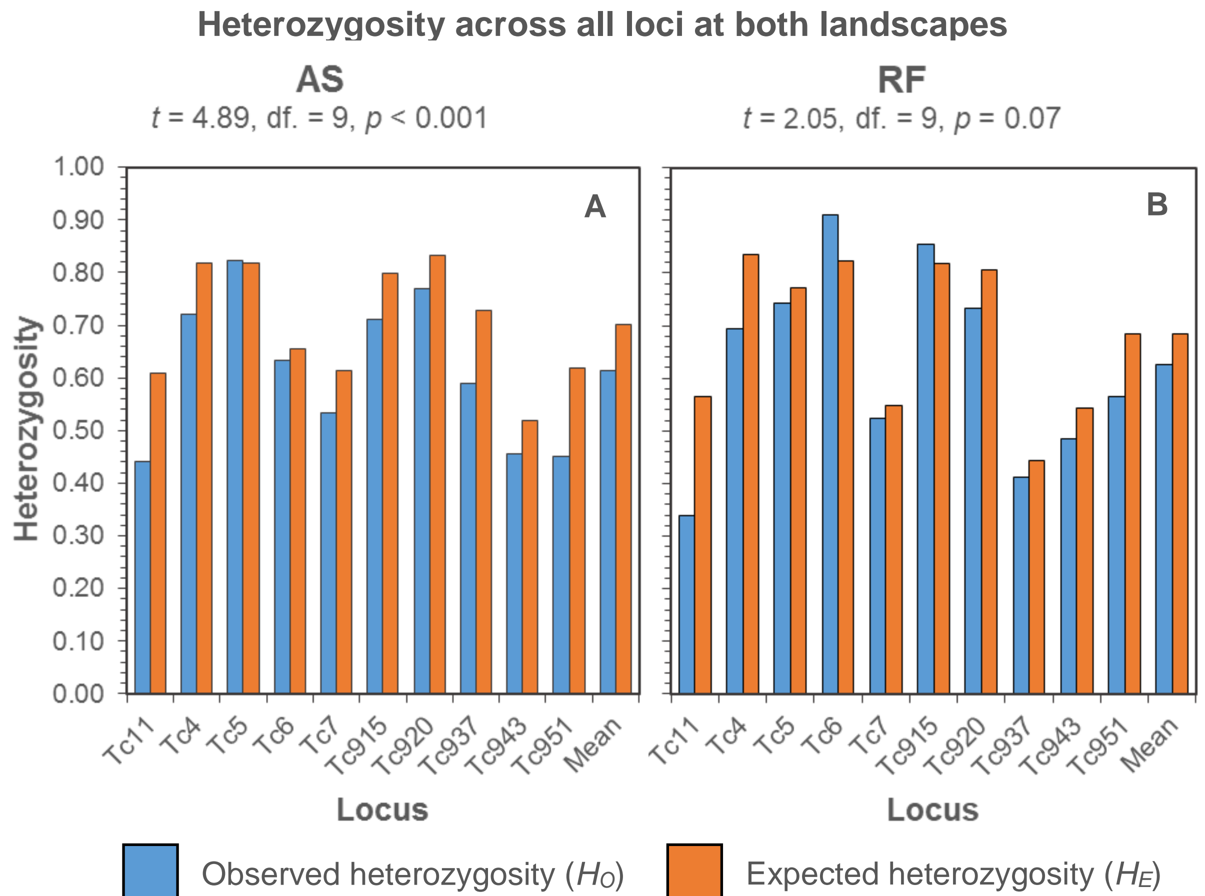


Figure 4.3: heterozygosity for each locus in both landscapes. The observed heterozygosity (H_O) of loci is indicated by the solid blue bar and the expected amount under Hardy-Weinberg equilibrium (H_E) by the solid orange bar. Results of a paired *t*-test for differences between H_O and H_E are given for each location.

Similarly when considering each landscape as a single population most loci were not in Hardy-Weinburg equilibrium according to a χ^2-test (**Table 4.6**), except for *Tc5* at **RF** ($\chi^2 = 35.87$, df = 55, $p = 0.87$) and *Tc915* ($\chi^2 = 287.66$, df = 190, $p = 0.14$). Both locations had significant linkage disequilibrium (**AS**: $\overline{r}_d = 0.04$, $n = 204$, $p < 0.01$; **RF**: $\overline{r}_d = 0.05$, $n = 124$, $p < 0.01$).

	AS			RF		
	χ^2	df	*p*	χ^2	df	*p*
Tc11	938.41	36	***	685.35	28	***
Tc4	115.86	55	***	195.96	66	***
Tc5	253.93	78	***	35.87	55	ns.
Tc6	270.48	55	*	76.36	66	***
Tc7	270.40	28	***	256.03	45	***
Tc915	544.94	136	***	287.66	190	ns.
Tc920	274.21	66	***	203.49	55	**
Tc937	773.48	66	***	158.82	28	**
Tc943	25.35	6	***	40.90	6	**
Tc951	239.38	15	***	145.92	21	**

Table 4.6: Results of a χ^2 test to determine departures from Hardy-Weinberg equilibria for all loci when considering each landscape as a distinct population. Arnside and Silverdale (**AS**) is in the left column and Rockingham Forest (**RF**) the right. Significance is indicated by asterisks or letters (* = p < 0.05; ** = p < 0.01; *** = p < 0.001; ns. = p > 0.05).

Geographic location was significantly related to genetic distance in both **AS** (R^2_{adj} = 0.01; F = 2.24, df = 2, $p < 0.001$) and **RF** (R^2_{adj} = 0.04; F = 3.44, df = 2, $p < 0.01$) according to dbRDA, indicating the presence of weak IBD given the low R^2_{adj} values (i.e. only a low proportion of variation in genetic distance was explained by position).

Spatial PCA of both **AS** and **RF** samples supported the detection of IBD above. The SGS was broadly arranged as southwest-northeast or northwest-southeast clines of differentiation respectively (**Figure 4.4**), with the exception of the two most distant fragments in both instances (**AS**: Grubbins Wood, Deepdale & Cringlebarrow; **RF**: Wakerley Great Wood, Old Sulehay), which appeared distinct. The latter two locations also represent the smallest demes in **RF**, and while Old Sulehay is the second most isolated fragment it is only marginally more distant (2.13 km) from its nearest neighbour than other examples (e.g. Bedford Purlieus to Easton Hornstocks, 1.62 km).

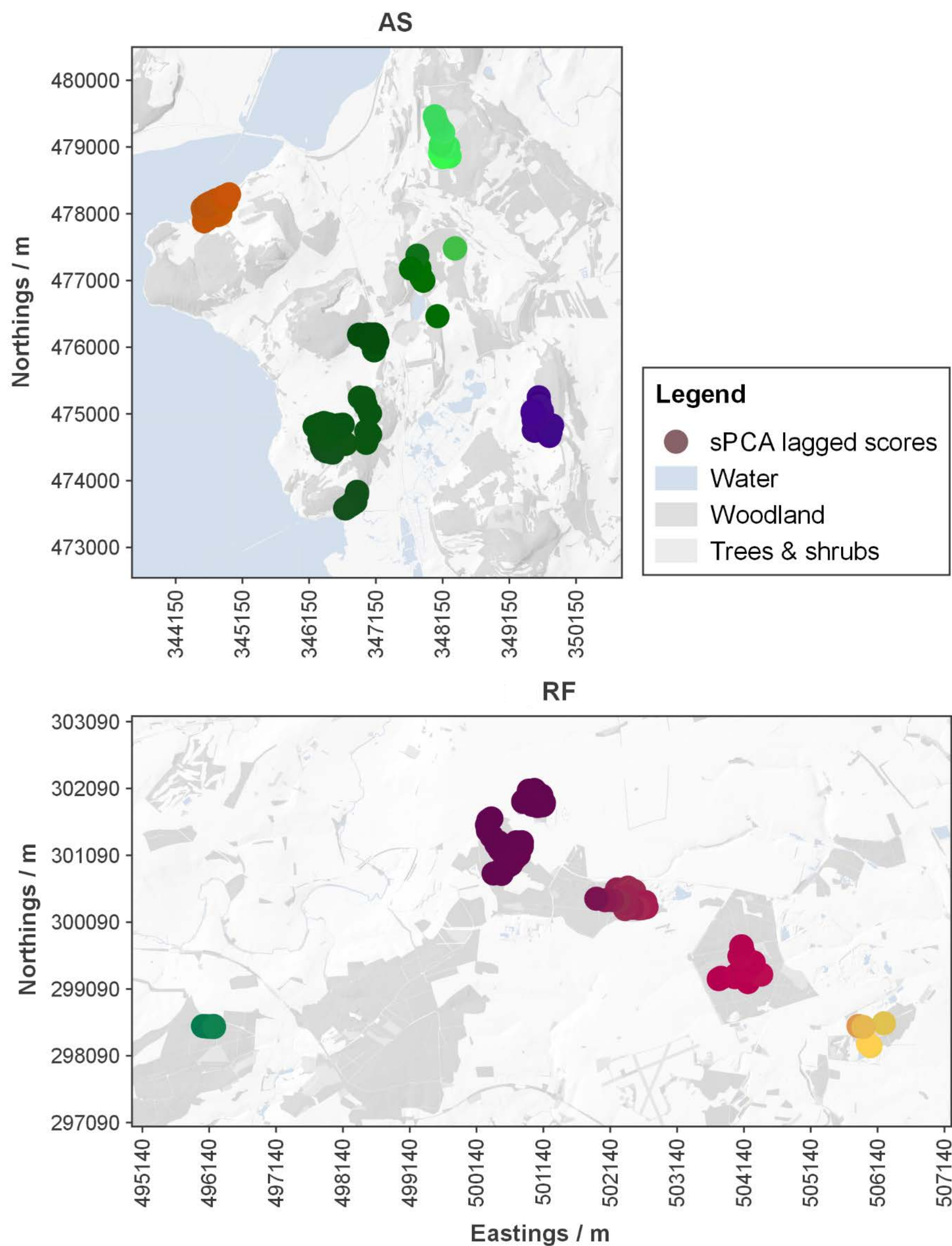


Figure 4.4: 'Colour plots' of the lagged scores (smoothed versions of the principal components scores created by replacing the original values with means of nearest neighbour scores; Jombart et al., 2008) from spatial principal components analysis (sPCA) of the allele frequency data across each landscape (**AS**, top; **RF**, bottom). To illustrate the spatially autocorrelated genetic structure, the shading of each individual point is generated by assigning the red/green/blue bands that define colours in computer graphics according to the lagged score of the first three principal components. Where only two components were retained (**RF**), the final band is set to a constant value. In both instances the SGS is clinal, suggesting it is generated by dispersal limitations (IBD).

4.3.2 Relationship of genetic distance with landscape variables

Genetic distance between individuals was unrelated to the examined landscape context, with no significant relationship between selected variables and genetic distance found at either **AS** (R^2_{adj} = 0.01, F = 1.04, df. = 7, p = 0.384) or **RF** (R^2_{adj} = 0.05, F = 1.24, df. = 5, p = 0.134). The largest proportion of the variance in genetic distance was unexplained by the model (**AS** = 0.669; **RF** = 0.623). Percent cover of urban areas in **AS** and average cos (Aspect) in **RF** constituted the largest proportion of constrained variance respectively. The remaining variables differed in their contributions between landscapes. In all instances the conditioned-out effects of geographic location explained a larger proportion of variance than any given landscape variable.

		Constrained	**Conditional**
AS	Urban	0.048	0.282
	Slope	0.040	0.290
	sin(Aspect)	0.036	0.295
	Woodland	0.035	0.296
	Altitude	0.032	0.298
	cos(Aspect)	0.023	0.308
	Open	0.021	0.309
	Full model	0.232	0.099
RF	cos(Aspect)	0.069	0.308
	Urban	0.069	0.309
	Slope	0.060	0.317
	sin(Aspect)	0.042	0.336
	Woodland	0.039	0.338
	Full model	0.248	0.129

Table 4.7: Proportion of variance explained (constrained) by partial dbRDA model, and the proportion of this explained by specific landscape variables as well as that 'partialled' (conditioned) out by geographic location. Unconstrained variance (**AS** = 0.669; **RF** = 0.623)

4.4 Discussion

Habitat fragmentation and the resulting small population sizes and isolation can erode genetic variation, which has implications for population persistence (Booy *et al.*, 2000; Christmas, Breed and Lowe, 2015). It has been recognised that *T. cordata* possesses characteristics which put it at risk of this (Turok *et al.*, 1996; De Jaegere, Hein and Claessens, 2016). This process is likely to affect populations at different latitudes to a greater or lesser degree, given that the amount of sexual recruitment varies with climate (see **Chapter 3**). Previous work has found no evidence of genetic erosion but did observe significant but weak population differentiation between demes (Logan, Phuekvilai and Wolff, 2015). How this relates to fragmentation is unclear given the coarse scale of sampling (i.e. much larger than expected seed or pollen dispersal distances). This study examines how genetic variation is arranged in two landscapes at contrasting latitudes with different levels of sexual recruitment and how this relates to landscape factors.

Genetic diversity (both polymorphism and heterozygosity) was reasonably high and very similar to previously reported results for *T. cordata* (Logan, Phuekvilai and Wolff, 2015), as expected, given that high levels of genetic variation are characteristic of tree species in general (Hamrick, 2004; Petit and Hampe, 2006). Specifically, heterozygosity was similar to other self-compatible woody plants (e.g. Noreen and Webb, 2013; Thakur et al., 2016; Zong et al., 2015), but lower than in primarily outcrossing wind-pollinated species (e.g. Cottrell et al., 2003; Jump and Peñuelas, 2006; Rasmussen et al., 2010). Polymorphism is more difficult to compare given its dependence on both mutation rate and sample size, both of which are specific to loci or particular studies respectively (Anmarkrud *et al.*, 2008), but where standardized allelic richness has been reported for other tree species the values observed here are similar (Dubreuil *et al.*, 2010).

Although genetic diversity was relatively high, there was a deficiency of heterozygotes at **AS**, but not in the most southern landscape, **RF**. There are several plausible explanations for this, none of which are mutually exclusive: technical difficulties, population substructure and increased inbreeding. A common problem when using microsatellite markers is the presence of null alleles (mutations in primer binding sites that prevent PCR amplification), but if this has occurred then homozygous null individuals should occur (Rico *et al.*, 2017), which was not observed. More likely the difference is the result of either increased inbreeding/selfing or population substructure. The former may be occurring at elevated levels in at least two fragments at **AS** due to a higher incidence of clonality (see

Chapter 3, and also note that proportionally more clones were removed during analysis here). Clonal reproduction is important in generating very fine-scale SGS (see **Chapter 3**), and the production of neighbourhoods of very closely related or identical individuals can increase inbreeding and selfing (Vallejo-Marín, Dorken and Barrett, 2010). This seems less likely than population substructure due to the potential for effective pollen transport within sites, with the mean distance for pollen movement being several times the maximum scale at which clonality affects SGS (Fromm, 2001).

Taken as a whole most markers at both landscapes are not in Hardy-Weinberg equilibrium (HWE). As with heterozygosity the most likely explanation for this is a departure from panmixia resulting in population substructure together with clonal reproduction. Larger populations are more likely to exhibit population substructure and when considering these subdivisions as one unit, the Wahlund effect leads to reduced heterozygosity and departures from HWE (e.g. Rasmussen et al., 2010). Clonality can also directly lead to departures from HWE due to the linked fate of alleles when generating replicate genotypes, which is illustrated by significant linkage disequilibrium in both landscapes, a pattern typical of clonally reproducing organisms (Guillemin *et al.*, 2008; Goss *et al.*, 2014). The fact that aside from a slight heterozygote deficiency at **AS** both landscapes had similar levels of genetic variation is unexpected given differences in the incidence of clonality (see **Chapter 3**), which imply differences in fertility levels. The increased amount of clonal reproduction at **AS** may not distinct enough to result in an appreciable difference in recombination and thus drift experienced at **RF**. Alternatively, since populations in **RF** still possess notable amounts of clonality, this may have reduced the impact of drift, preserving rare alleles (de Meeûs, Prugnolle and Agnew, 2007).

Spatial genetic structure was found in both landscapes. Genetic distance was significantly related to geographic location, indicating the presence of dispersal limitations (IBD). This contradicts a previous finding of no IBD detectable in the species (Logan, Phuekvilai and Wolff, 2015). The latter is a surprising result given the much coarser scale (hundreds of kilometres) in comparison with previously reported maximum pollen dispersal distances (Fromm, 2001). This is likely a result of the increased sensitivity of individual-based analyses utilized here (Prunier *et al.*, 2013). An exploratory ordination technique (sPCA) also illustrated the presence of clinal SGS, with differentiation of individuals increasing with geographic distance, supporting the conclusion of IBD. There was no apparent relationship between the observed SGS and the arrangement of woodland

fragments across both landscapes, except that the smallest and most isolated locations were the most distinct (although the southeastern-most deme at **RF**, Old Sulehay, is not particularly isolated). Genetic distance was also not related to topography or land cover/use. These patterns of SGS seem a more likely cause for the departures from HWE (via generation of population substructure) than clonality, given that previous work examining vegetative propagation specifically or population genetics generally within smaller, partially clonal populations do not find HWE departures (Morris *et al.*, 2014; Logan, Phuekvilai and Wolff, 2015). Similar results – clinal 'core' populations, distinct outliers – have been observed in other taxa using sPCA (Favre-Bac *et al.*, 2015).

Taken together these results suggest two possible explanations: that there is no response to fragmentation in *T. cordata* due to the lag time likely inherent in such a process (insufficient time has passed for genetic erosion to occur), or that there is no response because gene flow between fragments is sufficient to offset genetic drift. The clinal structure observed should occur in both of these instances, but in the former scenario it reflects historic dispersal limitations over a contiguous landscape while in the latter it is a result of current dispersal limits. There is no way to definitively separate the two here as the maximum pollen transport distance observed from successful pollination events is in principal sufficient to connect most of the fragments within this study (Fromm, 2001). In either instance however the nature of the SGS – with genetic distance being only weakly correlated with geographic location – suggests effective dispersal on a scale of one to two kilometres, with distinct individuals occurring where separated from other woodland fragments by larger distances (e.g. Grubbins Wood, Cringlebarrow & Deepdale, Underlaid Wood, **Figure 4.1a**; Wakerley Great Wood, **Figure 4.1b**). Although location explains the distinctiveness of these individuals, one of the most clearly differentiated fragments at **RF** (Old Sulehay) is not very much further from its nearest neighbour than other, less obviously different fragments. What makes this wood distinct is the small number of *T. cordata* individuals present. A smaller population is subject to an increased susceptibility to drift and elevated levels of inbreeding (Vranckx *et al.*, 2014), either of which would result in a more differentiated fragment. Therefore even if we consider the SGS to mostly reflect historic dispersal limitations, the smallest or most isolated fragments at least have responded to fragmentation. This justifies the continuing preservation of genetic resources via conservation measures, and additionally suggests populations that might be most at risk via their isolation or size.

Whether the most closely linked areas will respond similarly in the long-term is unclear given the difficulty of separating historic from current patterns. However entomophilous tree species respond more strongly to fragmentation than anemophilous ones (Vranckx *et al.*, 2012; Breed *et al.*, 2015). Considering that widespread wind-pollinated species which are far more effective dispersers than *T. cordata* (De Jaegere, Hein and Claessens, 2016) have already responded to habitat fragmentation (e.g. Dubreuil et al., 2010; Jump and Peñuelas, 2006), it seems unlikely that they will escape genetic erosion in the long-term, but how long that will be real terms is unclear. Based on historical map data the woodlands at **RF** have been separate for at least 129 y (Ordnance Survey (GB), 1888), and even longer at **AS** (Ordnance Survey (GB), 1848, 1862). Given this and the reproductive and life history traits of *T. cordata,* the lag time from fragmentation to response is clearly considerable, for several reasons. An appreciable fraction of the populations will be clonal replicates (see **Chapter 3**), which has the potential to bestow extreme longevity on vegetatively propagating genotypes (Honnay and Bossuyt, 2005; de Witte and Stöcklin, 2010). Individual stems are also long-lived (300 y) and possess low or at least inconsistent fertility at this point in its range (Pigott, 1991), meaning that generation times are long and there is low overall turnover. Genetic drift is therefore likely to be slow to act, even without any gene flow between fragments. As the connectivity between fragmented habitats is often modified by the characteristics of the intervening landscape (e.g. Hirao and Kudo, 2004; Trénel et al., 2008), the fact that there was no observed relationship between genetic distance and any of the examined landscape variables also supports an interpretation of the SGS as reflecting historical circumstance.

This study highlights the utility of individual-based population genetics studies to detect weak SGS via multiple methods where previously none was reported (Prunier *et al.*, 2013; Logan, Phuekvilai and Wolff, 2015). Although the results likely represent historical processes, meaning that the sampled *T. cordata* populations have not for the most part responded to habitat fragmentation, the distinctiveness of small or isolated locations shows that this is still likely to occur in more populations given sufficient time. This supports the continuation of genetic resource conservation efforts and illustrates which areas are most in danger of genetic erosion. It also suggests additional fruitful avenues of research relating to responses to fragmentation in *T. cordata.* Firstly if infrequent recruitment as a result of low or inconsistent fertility has increased the lag time in response to fragmentation then more central populations which regenerate freely should examined for signs of genetic erosion

using a similar approach (such as those in France; Pigott, 1991). Secondly, if the SGS of larger or more connected populations reflects historic patterns but they will ultimately respond to fragmentation, then more recent recruitment may be less genetically diverse than older cohorts as has been observed elsewhere (Vranckx *et al.*, 2012). Therefore a comparative study exploring differences in genetic diversity between juvenile and mature individuals would help to illustrate whether this interpretation is correct. Finally, plants can respond to habitat fragmentation twice, via restriction of both pollen movement and seed dispersal, and this should be explored for *T. cordata* in order to better understand the main factors affecting connectivity between fragments. The relative influence of pollen versus seed-mediated gene flow can be inferred via comparison of differentiation exhibited by bi- versus uni-parentally molecular markers (Oddou-Muratorio *et al.*, 2001), such as nuclear and chloroplast microsatellites (Provan, Powell and Hollingsworth, 2001).

5 Exploring metabolic variation in natural populations of a temperate canopy tree, *Tilia cordata*

5.1 Introduction

The current rapid increase in atmospheric CO_2 concentration driven primarily human activity is driving climatic changes at an unprecedented rate (IPCC, 2007), causing climatic zones to shift polewards or upwards altitudinally. Population persistence for organisms will therefore require either dispersal to track suitable areas, evolutionary adaptation to altered conditions or sufficient phenotypic plasticity to withstand them (Aitken *et al.*, 2008; Christmas, Breed and Lowe, 2015), although these responses are not mutually exclusive (Nicotra *et al.*, 2010). The extremely rapid speed of these changes likely outstrips the ability of some organisms to respond to them (Root *et al.*, 2003).

In the past range shifts have been a common response to climatic change in plant species, as indicated by paleo-ecological studies (Pardi and Smith, 2012), but this was not universal and some taxa failed to keep up with shifting climatic envelopes after the last glacial maximum (Normand *et al.*, 2011). Similarly although many species have moved recently in response to current changes (Parmesan and Yohe, 2003; Chen *et al.*, 2011), it is likely that the rate of change underway outstrips the dispersal ability of many others (Loarie *et al.*, 2009; Corlett and Westcott, 2013). For example, some contemporary plant populations have tracked recent warming only partly or not at all (Bertrand *et al.*, 2011; Chen *et al.*, 2011; Zhu, Woodall and Clark, 2012; Gray and Hamann, 2013). Even given sufficient potential for dispersal, there needs be both available space as well as suitable niche space within new locations and this depends on the communities present at or co-migrating to any destination (Christmas, Breed and Lowe, 2015). The degree to which the landscapes inhabited by species resist dispersal will also affect the success of range shifts, an important consideration given the degree of habitat fragmentation experienced by many species (Pearson and Dawson, 2005; Kremer *et al.*, 2012). Finally, sufficient time is also required, and this is dictated by the speed of environmental change and the species' life history and ecology, i.e. generation time, time to maturity and individual or propagule dispersal distance (Corlett and Westcott, 2013).

Taken together these constraints make range shifts an unlikely or at least insufficient response for many taxa (Aitken *et al.*, 2008). If dispersal is not an option then populations

must possess sufficient phenotypic plasticity to weather changes in the short term and enough adaptive variation to allow for evolutionary responses in the long term (Bradshaw, 2006; Jump, Marchant and Peñuelas, 2009). These processes are not separate, as the ability of individuals to sense environmental cues and generate phenotypic changes is ultimately under genetic control. Plasticity can therefore buffer against rapid change but may also assist in rapid evolutionary adaptation via selection for stronger plastic responses (Lande, 2009; Chevin, Lande and Mace, 2010; Nicotra *et al.*, 2010). Given this some tree species might be well-placed to respond rapidly to environmental change. In many instances they are locally adapted already (Alberto *et al.*, 2013), meaning they have high among-population adaptive variation. Given their large size they also possess typically high fecundity and gene flow which may allow for the movement of suitably adapted alleles from elsewhere in their range to the necessary populations (Petit and Hampe, 2006; Kremer *et al.*, 2012). As they are sessile, long-lived organisms, they also possess high levels of phenotypic plasticity to act as buffer against change (or even as the material for change as described above; Alberto et al., 2013; Nicotra et al., 2010).

Evolutionary responses to environmental change require the presence of sufficient adaptive variation within a species. Local adaptation is less likely in species with small or fragmented ranges, as selection is less efficient due to the increased effects of drift on any given locus (Leimu and Fischer, 2008). Although the presence of local adaptation is well-established in many economically important species given their long history of study in this regard (Alberto *et al.*, 2013), for less commercially relevant taxa the status of adaptive variation within populations is unknown. For example, *Tilia cordata* is a widespread temperate canopy tree that possesses a fragmented distribution, often occurring in small and isolated populations. Although it possesses fairly high levels of neutral genetic variation, and low levels of genetic differentiation (see **Chapter 4**; Logan et al., 2015), these cannot be used as evidence of either the presence of adaptive variation or a lack of local adaptation respectively, as neutral and adaptive variation are only weakly related (Reed and Frankham, 2001; Mittell, Nakagawa and Hadfield, 2015). Therefore the level of adaptive variation within the species is unknown and it is unclear whether it will be able to make evolutionary responses to changing environmental conditions. It is a poor disperser relative to other species based on postglacial migration rates (De Jaegere, Hein and Claessens, 2016), which may have been overestimated regardless (McLachlan, Clark and Manos, 2005). It also possesses low or at least intermittent fertility in many parts of its range, rarely colonizing

secondary woodlands (Pigott, 1991), and when combined with frequent clonal reproduction this likely results in a low turnover of individuals (see **Chapter 3**). Movement of populations, colonization of new areas or evolutionary responses can therefore all be expected to be slow, which suggests that phenotypic plasticity will be important in buffering against rapid environmental changes.

Rapid responses may be particularly important at lower latitude range margins (Sánchez-Salguero *et al.*, 2017). Being limited by colder temperatures (Pigott, 1981a; Pigott and Huntley, 1981) northern populations could benefit from climatic change (although altered disturbance regimes could still have a negative impact; De Jaegere et al., 2016; Lindner et al., 2010), while southern marginal populations face increased risk. Inadequate moisture regimes limit its distribution at this range edge, and in areas with annual precipitation of around 500 mm or less, *T. cordata* only survives in shaded or north-facing areas with a greater availability of water (Pigott and Pigott, 1993). Given the significant reductions in precipitation that are projected for the Mediterranean (Jacob *et al.*, 2014), declines in abundance and local extinctions have been predicted in Italian populations (Attorre *et al.*, 2011). Furthermore although northern range-edge populations have the potential to benefit from adaptive gene flow from lower latitude locations that are locally adapted to warmer conditions (assisted or otherwise), the inverse is not true (Hampe and Petit, 2005).

As increased tree mortality due to drought or heat stress has already been observed in forest ecosystems globally (Allen *et al.*, 2010) it is important to understand both how *T. cordata* will respond to environmental changes in the short-term (via phenotypic plasticity), and to assess the likelihood of evolutionary change in threatened populations. Knowledge of these processes should suggest the necessity of conservation measures such as facilitated migration (Aitken *et al.*, 2008). Reciprocal transplant or common garden experiments that examine the genetic basis of phenotypic traits such as specific leaf area (e.g. Ramírez-Valiente et al., 2010; Scheepens et al., 2010) are a common method of establishing the presence of local adaptation in plant species ('provenance trials'), and therefore the existence of adaptive variation among populations (Alberto *et al.*, 2013). Gene sequencing or population genomics techniques can also indicate the presence of local selection (Oleksyk, Smith and O'Brien, 2010). Either approach, while effective, requires significant investments of time and money and are therefore typically restricted to commercially important groups such as conifers (e.g. *Picea*, *Pinus*; Kujala and Savolainen, 2012; Mimura and Aitken, 2010; Palmé et al., 2009). Gene sequencing requires foreknowledge of the traits of interest and

genomics does not necessarily lend itself to understanding the actual functional distinction between populations, especially in non-model organisms, as gene homology is not necessarily a good guide for gene function (Fiehn, 2002).

Metabolomics or metabolic phenotyping is a promising and complimentary approach. It is the qualitative or quantitative study of all low molecular weight (less than one kDa) compounds within a sample, such as sugars or amino acids (Kosmides *et al.*, 2013). As some of these metabolites represent the end-products of cellular regulatory processes, their levels can represent the ultimate response of biological systems to genetic or environmental changes (Fiehn, 2002). Consequently it has been used to understand plant-environment interactions (e.g. Rivas-Ubach et al., 2014; Sardans et al., 2014), as well as to detect signatures of local adaptation in natural populations (Davey *et al.*, 2008; Kunin *et al.*, 2009; Field and Lake, 2011). It may be useful tool to examine both of these aspects in lime, because it allows for an exploratory approach which has the scope to uncover unexpected relationships or metabolite responses (Bundy, Davey and Viant, 2009). When the particular trait(s) of interest are not known, as is the case here, an examination of a wide array of metabolite concentrations can be used, an approach called metabolic fingerprinting (Halket *et al.*, 2005). In order to assess the presence of local adaptation both field-based provenance trials and metabolomics approaches require the removal of confounding environmental variables that affect phenotypes via plastic responses, by equalising it during development or acclimatizing samples to homogeneous conditions (Kim and Verpoorte, 2010; Field and Lake, 2011).

Unfortunately *T. cordata* is not a particularly experimentally tractable species, given its intermittent fertility and involved germination (Pigott and Huntley, 1981; Pigott, 1991; Gosling, 2007), which makes common garden, reciprocal transplant or glasshouse approaches difficult. Despite this an exploratory approach characterizing metabolic variation *in situ* within and between populations using metabolomic techniques may have utility in a hypothesis-generating role (Bundy, Davey and Viant, 2009), and would also be totally novel. As the scale at which metabolic variation was detectable in previous work is large (hundreds of kilometres), an appreciable geographic range will need to be sampled (Kunin *et al.*, 2009). To provide additional information, neutral genetic and adaptive morphological variation can also be assessed. Stochastic processes such as drift affect both neutral and functional genetic variation (Linhart and Grant, 1996). Therefore if differences in metabolite fingerprints are observed and they have a genetic basis, but differentiation across selectively neutral

molecular markers is not found then this would suggest deterministic processes are responsible in generating the former pattern (i.e. local selection). Additionally the metabolome is the product of many factors, including daily fluctuating metabolic processes and the current environmental context (Gibon *et al.*, 2006; Jones *et al.*, 2013). As a result it is a 'snapshot' of processes occurring at a variety of temporal scales (Brunetti *et al.*, 2013). In contrast with this a functional trait such as specific leaf area (SLA), although exhibiting diurnal variation, is largely influenced by climatic conditions (Poorter *et al.*, 2009). As a result it potentially integrates information on environmental processes operating over longer timescales. SLA is an indicator of plant strategy in distinct environments, varying with factors such as resource availability (e.g. water, nutrients) or climate (Lavorel and Garnier, 2002; Cornelissen *et al.*, 2003), and has been shown to be adaptive and exhibit inter-population variation (Ramírez-Valiente *et al.*, 2010). It will therefore also be contrasted across populations of *T. cordata*.

In summary this study has two objectives: to explore if samples collected *in situ* from natural populations and metabolically fingerprinted vary within and between populations, and in what way, which has not yet been attempted, and in doing so, explore how metabolic variation across different populations and conditions relates to relevant 'metadata' (both environmental context and methodological information), to allow for both speculation on the plastic capacity of the species and potentially to identify avenues for further research into local adaptation in the species.

5.2 Methods

5.2.1 Sampling strategy and locations

As previous metabolomics studies examining variation between populations found local adaptation only at large, regional scales, six populations of *T. cordata* spread across its UK range were sampled in order to explore metabolic differences between them (**Figure 5.1**). To minimize variation in environmental conditions, the context within sites was standardized as far as possible. East-facing trees that receive similar insolation were located, and the specific individuals to be sampled chosen randomly. To reduce differences in ontogeny sampling dates (as day length) were kept as similar as practicably possible. From each tree four branches spread across the canopy that received full sunlight were collected. Branches with obvious signs of pathogen damage or herbivory such as galls were avoided. The terminal leaves of these were cored, pooled and placed in dry-shippers charged with liquid nitrogen in order to quench metabolism and eliminate enzymatic activity. Cores were stored in the laboratory at -80 °C. In order to examine whether SLA differed between sites, the fourth and fifth leaves of each sampled branch were removed. These were placed in a plastic bag containing moistened tissue to avoid moisture loss and stored at 4 °C until measurement.

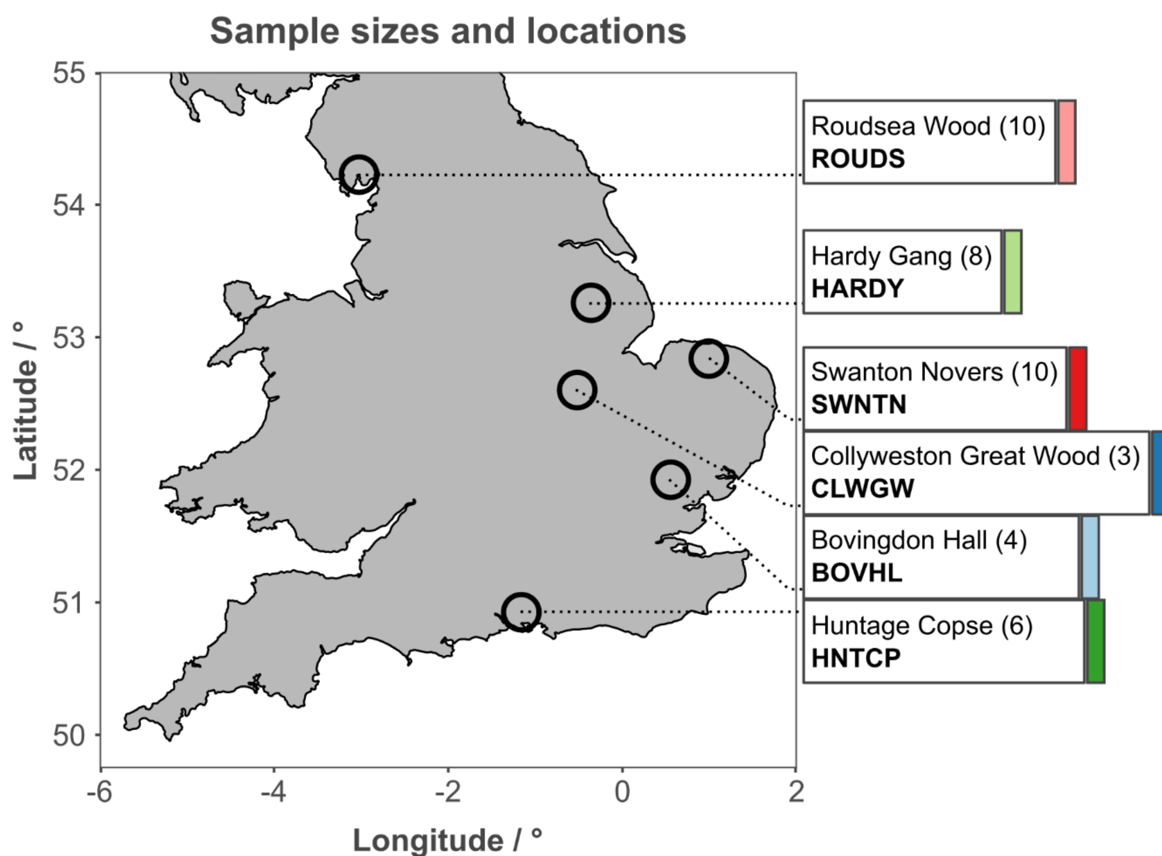


Figure 5.1: The geographic location of all sampling sites along with the number of samples per population. Open circles indicate the mean position of samples, while boxes connected by a dotted line to these indicate the respective name, sample size (in brackets), and abbreviation used for each. Coloured bars indicate the colour used to refer to each site in all further figures.

To explore the contribution of sampling methods and laboratory procedures to the metabolic variation, relevant 'metadata' during both collection and analysis of each sample was recorded (Goodacre *et al.*, 2007). As metabolite concentrations can show temporal variability (Gibon *et al.*, 2006), time of day was noted and converted to hours since dawn based on latitude and date using R software `maptools`, version 0.8-41 (Bivand and Lewin-Koh, 2017). Other factors likely to affect primary metabolism via changes to photosynthetic rate or transpiration were also measured: wind speed and ambient temperature were recorded with a handheld anemometer; relative humidity was noted with a handheld hygrometer; finally, the current amount of photosynthetically active radiation (PAR) reaching each sample was recorded using a handheld photometer.

5.2.2 Metabolite extraction

The four leaf cores from each individual were split into pairs and extracted, run and analysed separately as biological replicates, to allow for an assessment of the variability

within individuals. Metabolites were extracted using a similar procedure to Davey et al. (2008). Each leaf pair was weighed, ground in a mortar and pestle with sterile sand and then transferred to a 2 ml microcentrifuge tube. 40 µl mg^{-1} leaf tissue of a solvent mixture containing methanol, chloroform and water ($MeOH/CHCl_3/H_2O$, 2.5:1:1) was added, vortexed to mix and then left on ice for 30 min. Tubes were then centrifuged at 16 000 x *g* for 2 m and the supernatant transferred to a fresh 15 ml centrifuge tube and stored on ice. A solvent mixture ($MeOH/CHCl_3$, 1:1) was added to the pellet (20 µl mg^{-1}), vortexed to mix and the tube stored on ice for a further 30 min. This was then centrifuged as before and the supernatant transferred to the 15 ml tube. This extract was separated into nonpolar, $CHCl_3$ and aqueous $MeOH:H_2O$ phases by the addition of distilled H_2O at 4 °C (10 µl mg^{-1}) and centrifugation at 16 000 x g for 1 m. 300 µl of the polar phase was transferred to glass vials and stored at -80 °C until analysis. Both solvent mixtures were chilled to -20 °C before use, and the extraction procedure was undertaken in small, randomly chosen groups (two to three pairs) in a randomized order. These sequences were also retained as metadata to assess if the handling of samples induced any of the observed variation between samples. 'Extraction blanks' were generated by repeating this procedure three times without any leaf material.

5.2.3 Metabolite fingerprinting

Despite the *–omics* suffix, no current metabolomics data collection techniques provide information on all metabolites present within a sample due to the wide array of compounds produced by the plant kingdom and their physical differences (i.e. polarity). Estimates put the total number at 90 000 - 200 000 (Fiehn, Kloska and Altmann, 2001; Peñuelas and Sardans, 2009) and typically only a subset of these compounds are analysed (Brunetti *et al.*, 2013). Therefore, for reasons of time and expense only the polar metabolites present in the aqueous phase were examined. Metabolic fingerprints were generated using a Synapt G2 (Waters Ltd., Elstree, UK) matrix-assisted laser desorption/ionization time-of-flight mass-spectrometer (MALDI-TOF MS) in both positive and negative ion modes. Samples were combined 1:1 with a 5 mg ml^{-1} α-Cyano-4-hydroxycinnamic acid (CHCA) matrix and ionized with a solid-state Nd:YAG UV laser (355 nm). Spectra were collected in the mass range of 100 – 800 Da at a rate of one spectrum s^{-1} (1 s scan time, 0.02 s inter-scan delay). In positive mode, the mass spectrometer operated with an accelerating voltage of 10 V, an 11 V hexapole bias and an aperture voltage of 7 V. Negative mode settings were identical except for lower hexapole bias and aperture voltage (10 V and 5 V respectively).

5.2.4 Metabolite data pre-processing

Raw spectra of mass/charge ratios were processed as described in Overy et al. (2005). Putative identification of metabolites was made based on comparison of detected masses with the KEGG database (Kanehisa Laboratories, 2017). Spectra were rounded into 0.2 Da bins and the relative abundance of each used to produce profiles of percent total positive and negative ion content for each replicate (%TIC).

5.2.5 Assessment of neutral genetic variation

For comparison of potentially adaptive variation and neutral genetic variation all individuals were genotyped across 10 microsatellite loci using the 'crude extract' procedure described previously (see **Chapters 3** & **4**). Allele frequencies at these loci are in Hardy-Weinberg equilibrium in a variety of populations of *T. cordata* (Logan, Phuekvilai and Wolff, 2015). These markers are therefore not expected to be under selective pressure.

5.2.6 Collection of functional trait data

Specific leaf area (SLA) was recorded following Cornelissen et al. (2003). One-sided leaf area was measured using an AM350 Area Meter (ADC BioScientific Ltd., Hoddesdon, UK), and then leaves were oven-dried at 80 °C for 48 h and weighed individually. SLA was calculated for each individual as the mean of each of its leaves' area divided by their mass (10 *n* for each tree).

5.2.7 Data exploration and analysis

All data analysis was undertaken in the R software package, version 3.3.3 (R Core Team, 2016).

5.2.7.1 Data pre-treatment

Metabolite concentrations can differ by orders of magnitude across a dataset and yet these differences do not necessarily correspond to their biological importance. For example, the average concentration of a signalling molecule is likely to be much lower than a ubiquitous compound involved in primary metabolism such as ATP (van den Berg *et al.*, 2006). Scaling the data can help focus on biologically relevant variation by adjusting for these fold changes between different metabolites, by converting it into differences relative to a scaling factor. Pareto-scaling (division of each variable by the square root of its standard deviation) was applied to both positive and negative %TIC datasets here, as it reduces the relative importance of large values but keeps the structure of the data relatively intact (van den Berg

et al., 2006; Yang *et al.*, 2015). To avoid generation of nonsensical data, several invariant bins with a standard deviation of zero had to be removed prior to scaling (**Table 5.1**). Finally, before analysis, the bins representing the CHCA matrix used in the ionization process were removed from both datasets (**Table 5.1**). These signals are also detected by the instrument and can obscure biological information (Worley and Powers, 2013).

	Data removed		**Rationale**
	Positive %TIC	**Negative %TIC**	
Mass bin	172.0	172.0	Matrix bins
	379.0	379.0	
	100.0	800.2	Invariant
	752.8		
	753.0		
	753.2		
	753.4		
	753.6		
	765.8		
	800.4		
	800.6		

Table 5.1: Data bins removed from each metabolite concentration (percent total ion content, %TIC) dataset prior to analysis and the rationale behind removal.

5.2.7.2 Multivariate data exploration

Due to its sensitivity and large mass-range, TOF mass spectrometry generates large datasets with many collinear variables, making it necessary to use multivariate analyses that can handle this collinearity (Worley and Powers, 2013). The use of ordination methods such as principal components analysis (PCA) or partial least squares-discriminant analysis (PLS-DA) can mitigate the problem of high dimensionality by effectively summarising the dataset. PCA produces orthogonal variables (components), linear combinations of the original variables that describe the axes of greatest variation through the original data, providing an informative summary of structures present. This approach is useful in exploratory studies where differences between groups are unknown or unpredictable as they are here (Worley and Powers, 2013). Therefore PCA was applied to both positive and negative %TIC datasets using the R package `vegan`, version 2.4-2 (Oksanen *et al.*, 2017). This was performed both with and without extraction blanks in order to assess whether the procedure to assess the level of variation induced by technical issues such as instrumental noise. If these 'samples' are not distinct from actual metabolite profiles, then a large amount of the variation has been

generated by the procedures used (van den Berg *et al.*, 2006). To explore whether environmental context or laboratory methods were related to observed variation in metabolite concentrations, all metadata were used to colour-code plots displaying the first two principal components of each.

As PCA summarises the axes of greatest variation through the data, it only reveals group structure when within-group variation is less than between-group variation (Kosmides *et al.*, 2013). In contrast PLS-DA reduces the dimensionality of the data by producing latent variables that are linear combinations of the original variables that maximise covariance between response and class membership data (here site of origin). In other words, the lower dimensional space created is primarily formed by the predictive components of the data (Barker and Rayens, 2003; Worley and Powers, 2013). Due to this approach it can also fortuitously find separation between groups where none exists, and so should be used in concert with PCA to guide interpretation (Worley and Powers, 2013). Therefore PLS-DA was also applied to both positive and negative %TIC datasets using the R package `mixOmics`, version 6.2.0 (Le Cao *et al.*, 2017). An assessment of how well these models discriminate between groups was performed using leave-one-out cross-validation (LOOCV), which generates new PLS-DA models excluding one observation across all observations, and uses these to predict class membership for each excluded observation (Hastie, Tibshirani and Friedman, 2001). This allows for the calculation of error rates, i.e. what proportion of individuals are misclassified.

Finally, the genetic data was explored in a comparable way. As population genetics studies typically have higher sample sizes than used here, the relationship of genetic distance (1 – percentage shared alleles [DPS]; Bowcock et al., 1994) with geographic location was assessed using distance-based redundancy analysis (dbRDA) as in **Chapter 4** (Legendre and Fortin, 2010; Meirmans, 2015). If sufficient numbers of individuals have been genotyped then the populations can be expected to exhibit isolation-by-distance (IBD) as observed previously, given the spatial scale of the study. PCA was then applied to allele frequency data for comparison with the ordinations of metabolite concentration.

5.2.7.3 Functional trait variation

To compare SLA between sites, normality and homoscedasticity of collected data was first assessed using Shapiro-Wilks and Bartlett tests respectively (Bartlett, 1937). While this indicated that data were normal, groups were significantly heteroscedastic. Therefore a

nonparametric Kruskal-Wallis hypothesis test was applied. If this is significant a Bonferroni-corrected Dunn’s multiple comparison test will be used to identify where the differences lie, due to the uneven sample sizes between groups (Dunn, 1964).

5.3 Results

Both positive and negative mode TOF mass spectrometry revealed variation in metabolite concentrations across sites, as illustrated by averaged %TIC within each molecular mass bin (**Figure 5.2**).

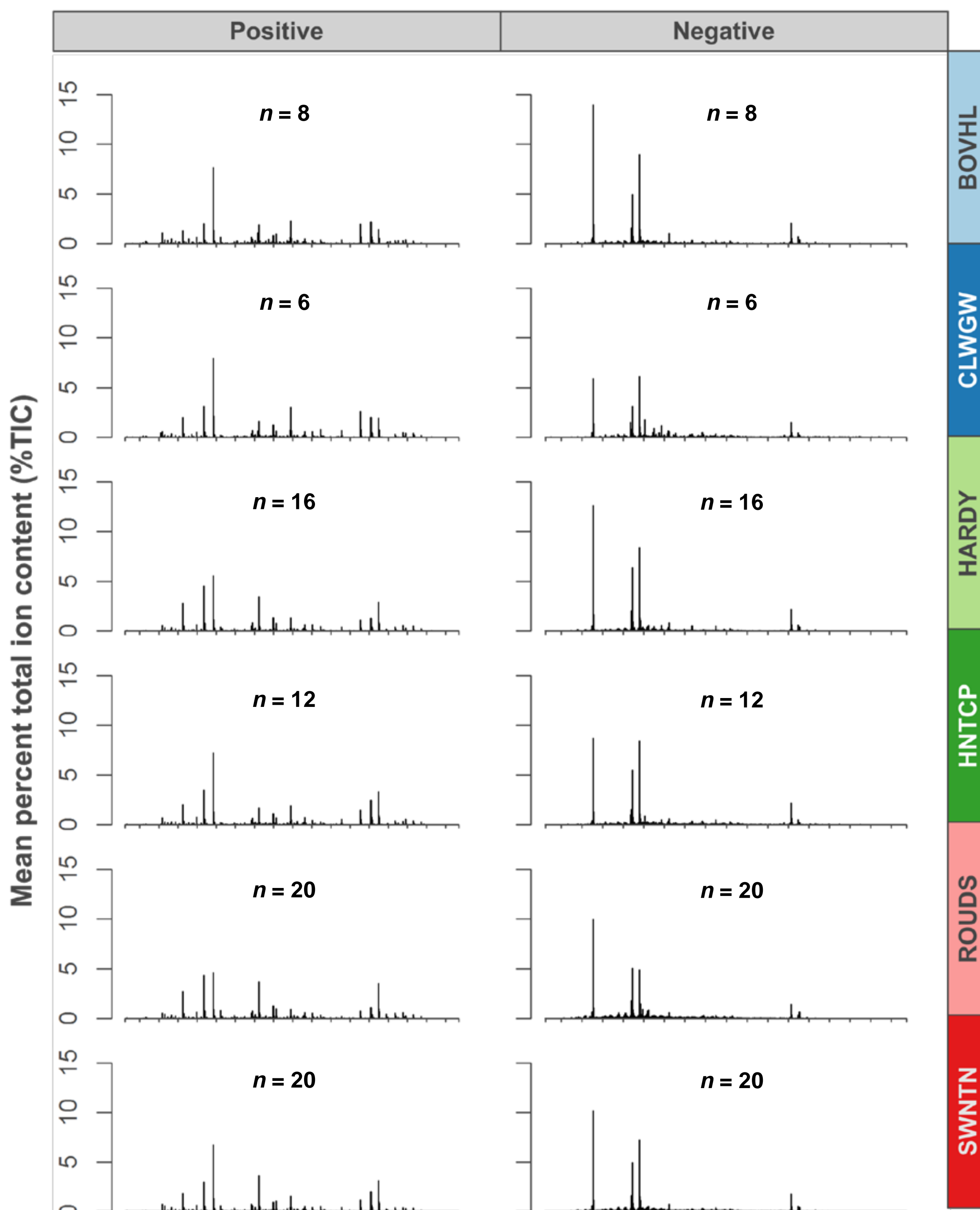
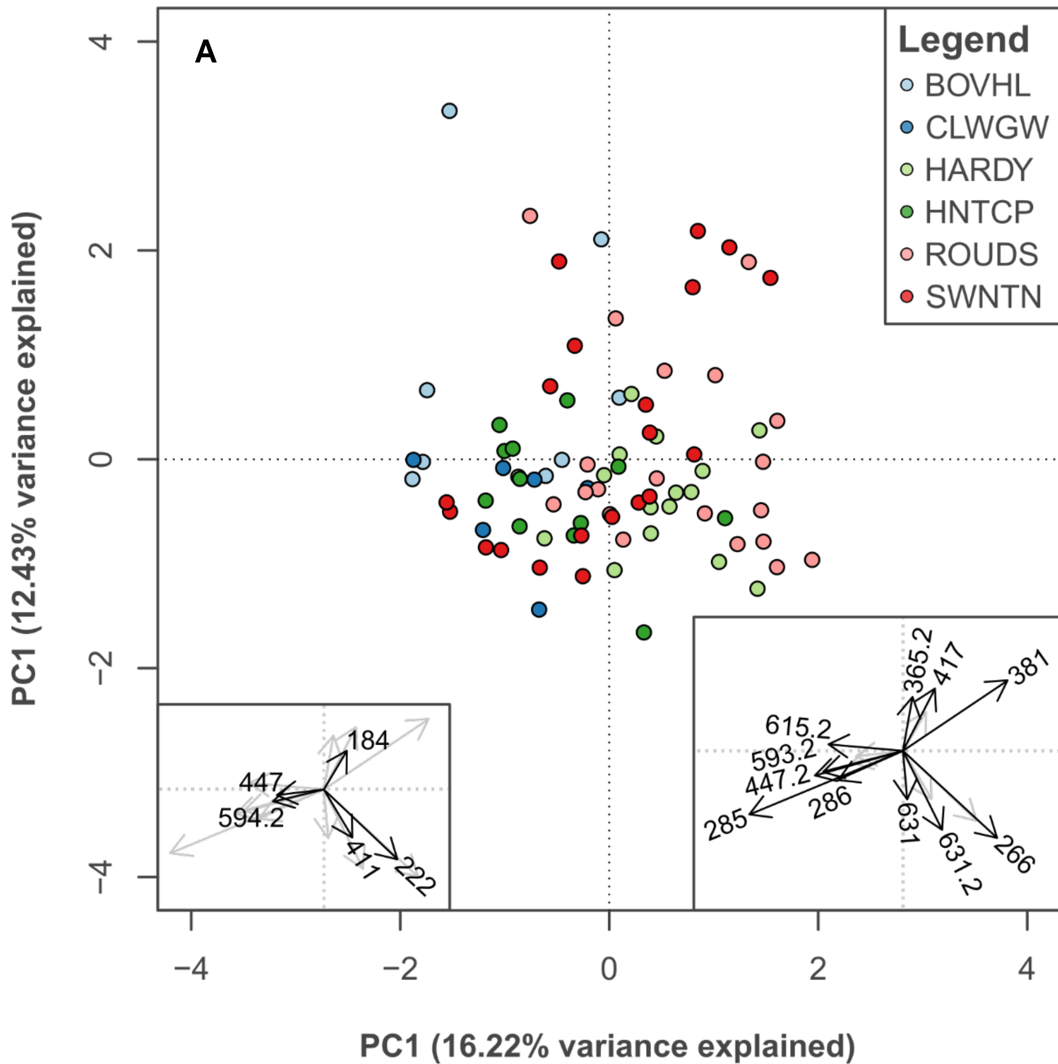


Figure 5.2: Site-wide means of percent total ion content (%TIC) for each mass bin across all individuals and biological replicates within each site. Left-hand charts contains %TIC for metabolites detected in positive mode while charts on the right contain %TIC for those detected in negative mode. Site codes are given on the right:. **BOVHL** (light blue): Bovingdon Hall; **CLWGW** (dark blue): Collyweston Great Wood; **HARDY** (light green): Hardy Gang; **HNTCP** (dark green): Huntage Copse; **ROUDS** (pink): Roudsea Wood; **SWNTN** (red): Swanton Novers.

Multivariate analyses show that actual variation between sites is low, however, except for that of the extraction blanks, which appear distinct from all other samples (results not shown). The first two components of PCA of both datasets summarise an acceptable proportion of the variation in %TIC between samples (positive: 28.65%; negative: 38.39%), given the extremely high dimensionality of the data (3493 and 3499 variables respectively). These two axes of greatest variation through the positive dataset reveal little variation between sites, as illustrated by the large amount of overlap in an ordination of the first two principal components (PC1/PC2), and although many individuals are similar to others from the same location (**Figure 5.3A**).



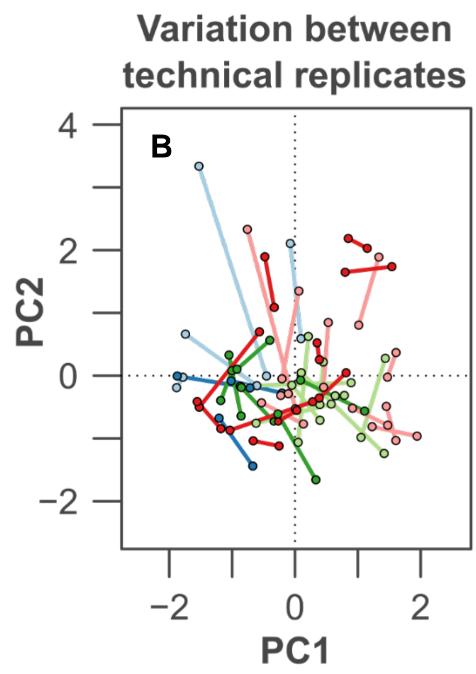


Figure 5.3: A) Principal components analysis (PCA) of percent total ion content (%TIC) in all samples. Only the first two components (PC1/PC2) are shown. The filled circles indicate the scores received by each sample; colour indicates site of origin. Inset bottom left and right: variable loadings for this ordination, which illustrate the most influential mass bins in the analysis. For the sake of clarity only the ten highest absolute loadings for each component are labelled with their respective mass bin size.

B) To illustrate intra-individual variation, the same ordination as above is provided, but with biological replicates connected by a coloured line.

The same technique reveals two obvious structures in the negative dataset however: all sites bar ROUDS contain individuals that receive similar scores along PC1 and appear together in a separate cluster, while the remaining individuals primarily differ by scores received on PC2 (**Figure 5.4A**). In particular the sites ROUDS and SWNTN show the greatest variation along this axis, and several individuals from each are distinct from the other groups as a result. Scores received by biological replicates in both ordinations reveal that there is heterogeneity in intra-individual variation. Metabolite profiles within some samples are very similar, causing them to cluster tightly with the other from the same individual. Equally some biological replicates are quite distinct as indicated by large distances between them in PCA of both positive and negative %TIC (**Figure 5.3B**, **Figure 5.4B**).

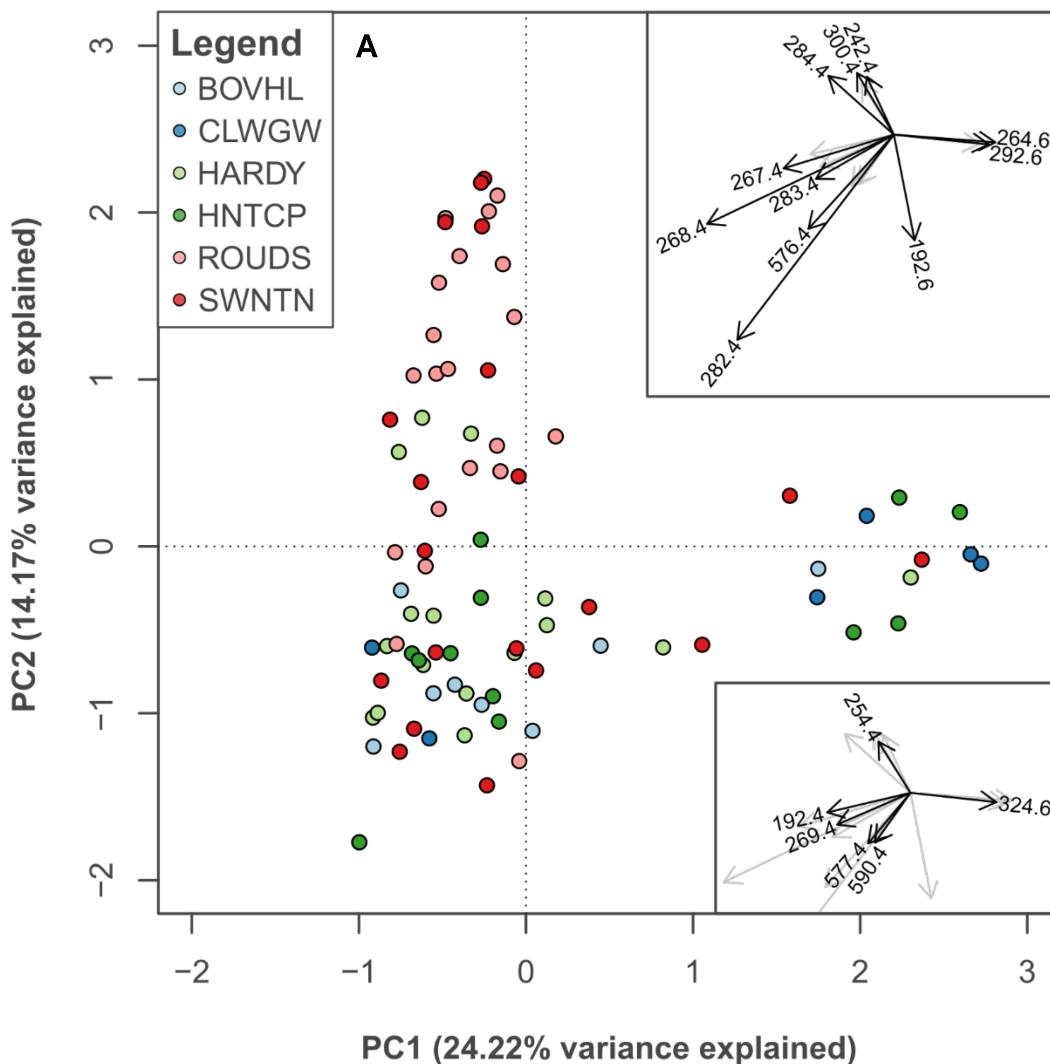
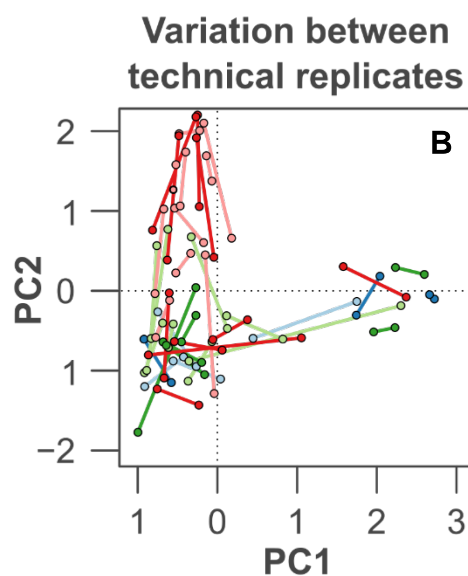


Figure 5.4: A) Principal components analysis (PCA) of negative percent total ion content (%TIC) in all samples. Only the first two components (PC1/PC2) are shown. The filled circles indicate the scores received by each sample; colour indicates site of origin. Inset top and bottom right: variable loadings for this ordination, which illustrate the most influential mass bins in the analysis. For the sake of clarity only the ten highest absolute loadings for each component are labelled with their respective mass bin size.

B) To illustrate intra-individual variation, the same ordination as above is provided, but with biological replicates connected by a coloured line.



Each principal component is composed of contributions from all mass bins, but the relative importance of each varies. Due to the extreme number of variables within the data only the ten highest loadings for each component of both ordinations are provided. The mass bins with the highest contributions to principal components derived from positive %TIC profiles range from 184 to 632.2 Da (**Table 5.2**), and many are strongly correlated with each other as indicated by the narrow angle between loading arrows (**Figure 5.3A**, inset bottom left and right). The ten highest ranked loadings from the negative ordination covered a similar but slightly narrower range (192.4 – 590.4 Da), and the majority of important bins are between 254.4 and 324.6 Da (**Table 5.3**). Comparisons of detected metabolite masses with the KEGG database suggest many potential candidates for the identity of compounds in the positive %TIC profiles (see **Appendix III** for an unbridged list of the ten highest ranking loadings). In most bins there was no clear pattern to the putative identifications, with metabolites involved in a wide array of metabolic pathways (**Table 5.2**), but size matches in bins 381 (PC1: 2nd, PC2: 3rd) and 365.2 (PC2: 7th) were primarily oligosaccharides. There were far fewer matches for the negative profiles, but two mass bins did result in putative identifications (**Table 5.3**).

Mode	Component	Mass Bin	Rank	Loading	Putative identification	Amino acid metabolism	Synthesis of secondary metabolites	Cell signalling	Cell membrane metabolism	Carbohydrate metabolism	Citric acid cycle	Purine metabolism	Pyrimidine metabolism
Positive	PC1	285	1	-0.42	Various (10)	✓	✓	✓	✓	✓	✓		
		381	2	0.29	Oligosaccharides; *galactinol; pantheteine 4'-PO_4*	✓	✓	✓	✓	✓	✓		
		266	3	0.26	Amino acids (3), nucleosides (2)	✓	✓	✓	✓	✓	✓		✓
		447.2	4	-0.24	Various (10)	✓	✓	✓	✓	✓	✓		
		593.2	5	-0.22	-								
		615.2	6	-0.20	-								
		222	7	0.20	Amino acids (5), amino sugars (3)	✓	✓	✓	✓	✓	✓		
		286	8	-0.18	*Pyridoxal-PO_4; pyrodoxine-PO_4; linamarin*	✓	✓	✓	✓	✓	✓		
		594.2	9	-0.14	-								
		447	10	-0.13	*CDP-ethanolamine; Khellol glucoside*				✓				
	PC2	266	1	-0.27	Amino acids (3), nucleosides (2)	✓	✓	✓	✓	✓	✓		✓
		631.2	2	-0.25	*Diosmin; reserpine*								
		381	3	0.22	Oligosaccharides; *galactinol; pantheteine 4'-PO_4*	✓	✓	✓	✓	✓	✓		
		222	4	-0.22	Amino acids (5), amino sugars (3)	✓	✓	✓	✓	✓	✓		
		285	5	-0.20	Various (10)	✓	✓	✓	✓	✓	✓		
		417	6	0.19	-								
		365.2	7	0.16	Oligosaccharides; *XMP; GA44; galactinol; ajmaline*	✓	✓	✓	✓	✓	✓	✓	
		631	8	-0.15	-								
		411	9	-0.15	*2'-Deoxyuridine 5'-diphosphate*								✓
		184	10	0.12	Amino acids (4); *selenophosphate*	✓	✓	✓	✓	✓	✓		

Table 5.2: Ten highest absolute loadings for each principal component derived from positive percent total ion content (%TIC). Where possible putative identifications from the KEGG database (Kanehisa Laboratories, 2017) along with their respective metabolic pathways.

Mode	Component	Mass Bin	Rank	Loading	Putative identification	Amino acid metabolism	Synthesis of secondary metabolites	Cell signalling	Cell membrane metabolism	Carbohydrate metabolism	Citric acid cycle	Purine metabolism	Pyrimidine metabolism
Negative	PC1	268.4	1	-0.32	-								
		282.4	2	-0.27	-								
		267.4	3	-0.19	-								
		264.6	4	0.17	-								
		292.6	5	0.16	-								
		324.6	6	0.15	-								
		576.4	7	-0.15	*Phosphoribosyl-formimino-AICAR-phosphate*	✓	✓						
		192.4	8	-0.14	-								
		283.4	9	-0.13	-								
		269.4	10	-0.13	*Estrone*			✓					
	PC2	282.4	1	-0.46	-								
		192.6	2	-0.24	-								
		576.4	3	-0.21	*Phosphoribosyl-formimino-AICAR-phosphate*	✓	✓						
		268.4	4	-0.20	-								
		300.4	5	0.14	-								
		284.4	6	0.13	-								
		242.4	7	0.13	-								
		254.4	8	0.11	-								
		577.4	9	-0.11	-								
		590.4	10	-0.11	-								

Table 5.3: Ten highest absolute loadings for each principal component derived from negative percent total ion content (%TIC). Where possible putative identifications from the KEGG database (Kanehisa Laboratories, 2017) along with their respective metabolic pathways.

Supervised ordinations of positive and negative %TIC profiles had similar patterns to the unsupervised PCA. The two latent variables (LV1/LV2) of PLS-DA of the positive dataset illustrate the lack of variation between sites, with distinctly overlapping groups (**Figure 5.5A**). A similar structure is present in the PLS-DA of negative %TIC to the PCA of the same. Most individuals vary along one direction, with the sites ROUDS and SWNTN exhibiting the most variation and the former group the most distinct (but still overlapping). As before there is an additional cluster in an orthogonal direction containing outliers from all sites bar ROUDS. However, both patterns here are rotated approximately 45° relative to

the PCA of the same data, i.e. individuals are differentiated along both latent variables rather than each component being responsible for one axis of separation (**Figure 5.6A**). Despite these obvious patterns neither dataset effectively distinguished between sites. Based on leave-on-out cross-validation, overall error rates for both PLS-DA models were high (>60%; **Figure 5.7**).

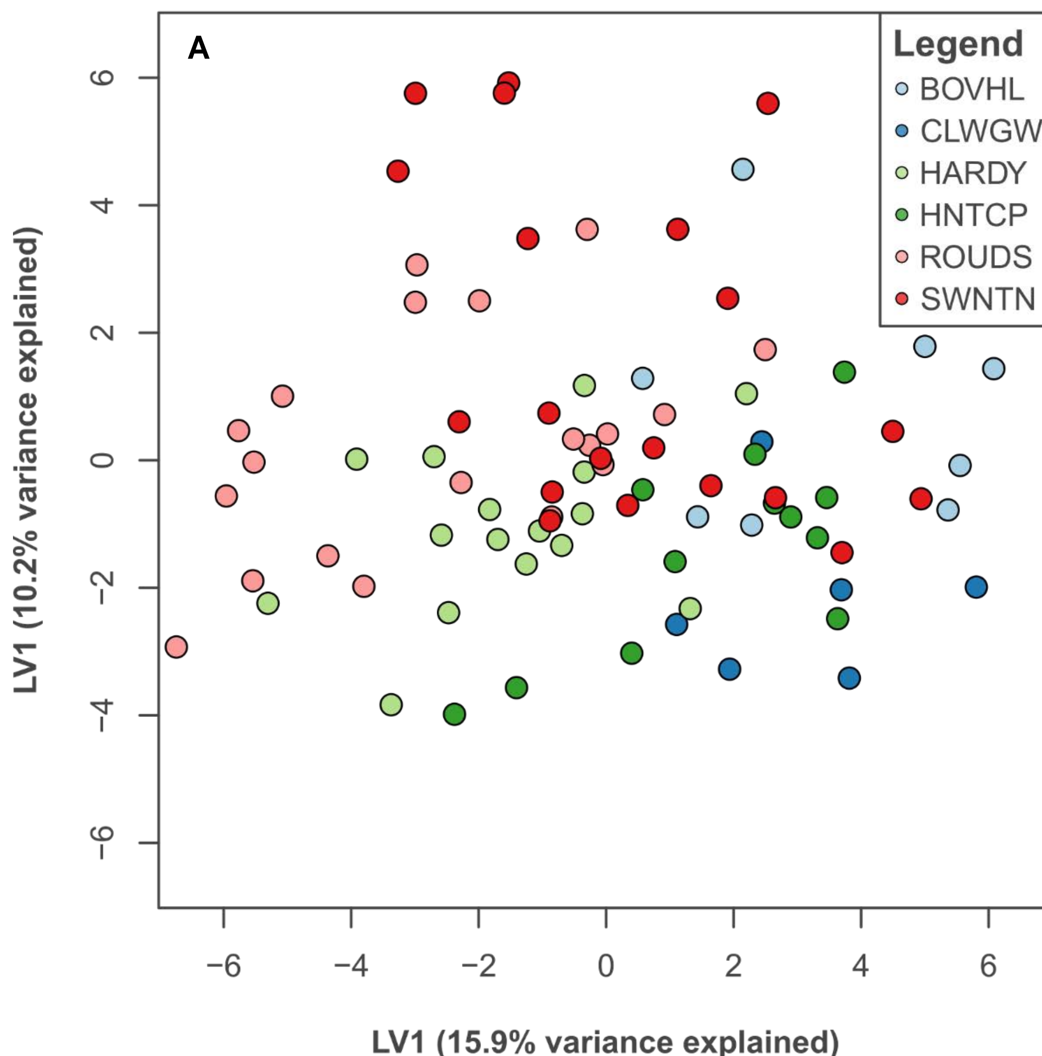
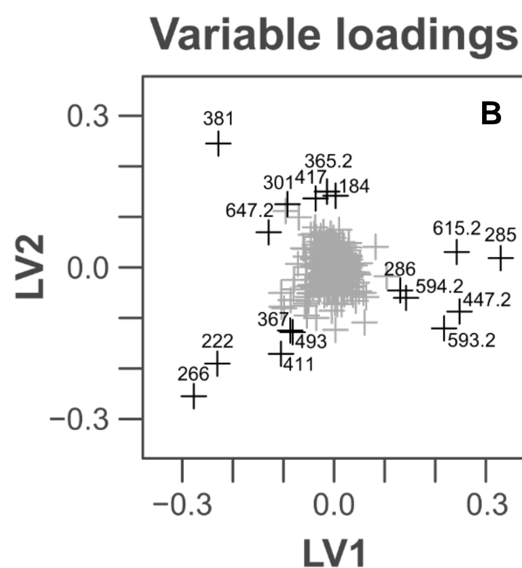
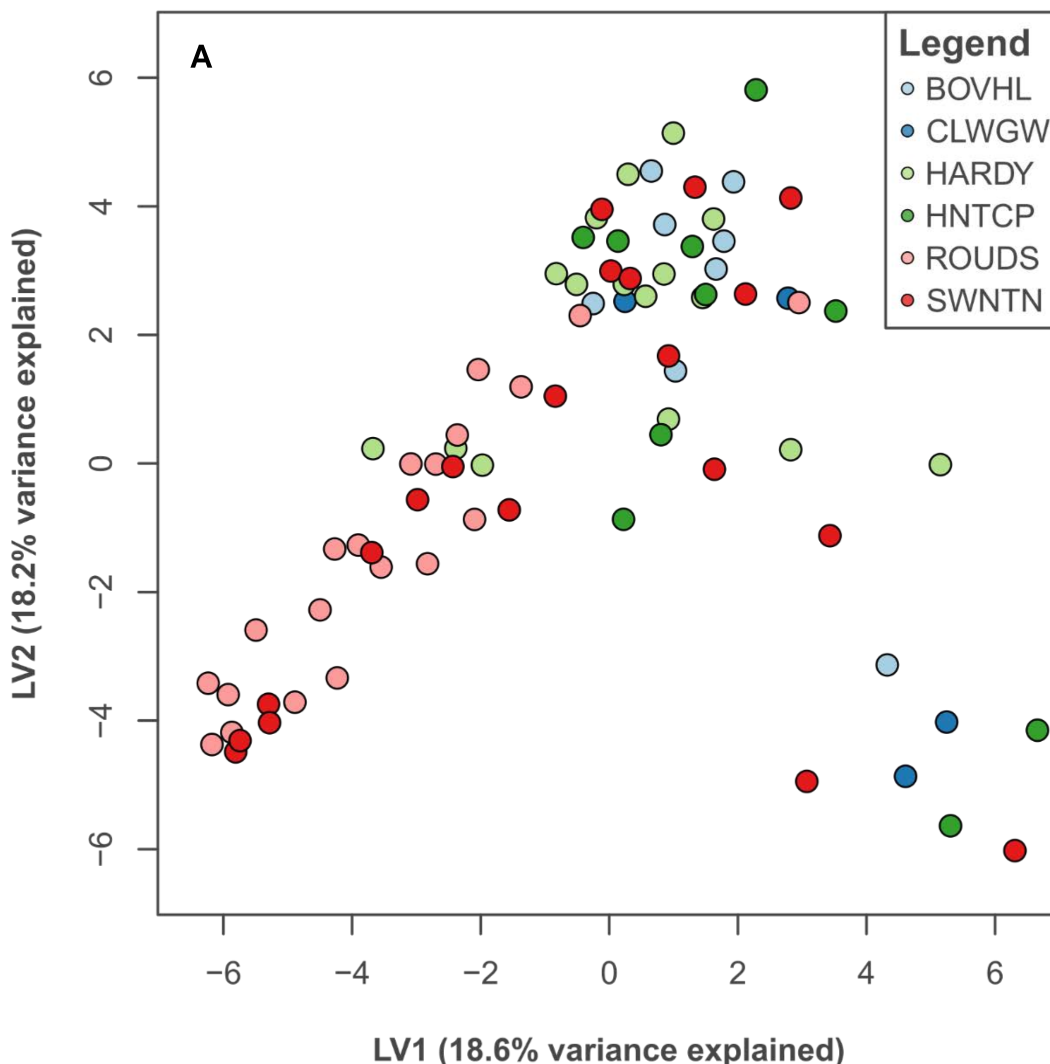


Figure 5.5: A) Partial least squares-discriminant analysis (PLS-DA) of Pareto-scaled positive total ion content (%TIC) of samples. Only two latent variables were generated for this model.

B) The most influential mass bins within the analysis are illustrated by the variable loadings for the PLS-DA model. For the sake of clarity only the ten highest absolute loadings for each latent variable are labelled.





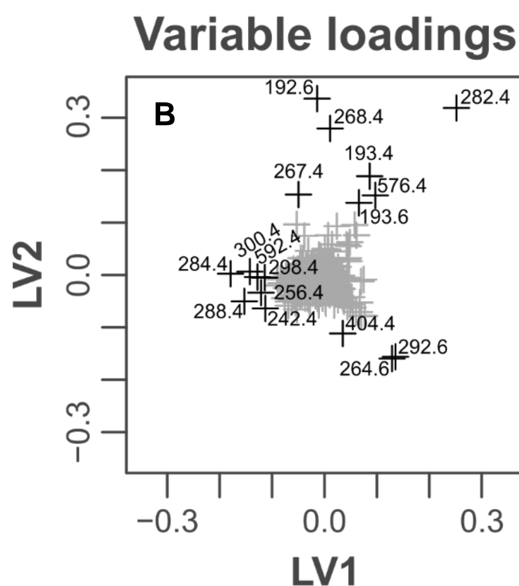


Figure 5.6: A) Partial least squares-discriminant analysis (PLS-DA) of Pareto-scaled negative total ion content (%TIC) of samples. Only two latent variables were generated for this model.

B) The most influential mass bins within the analysis are illustrated by the variable loadings for the PLS-DA model. For the sake of clarity only the ten highest absolute loadings for each latent variable are labelled.

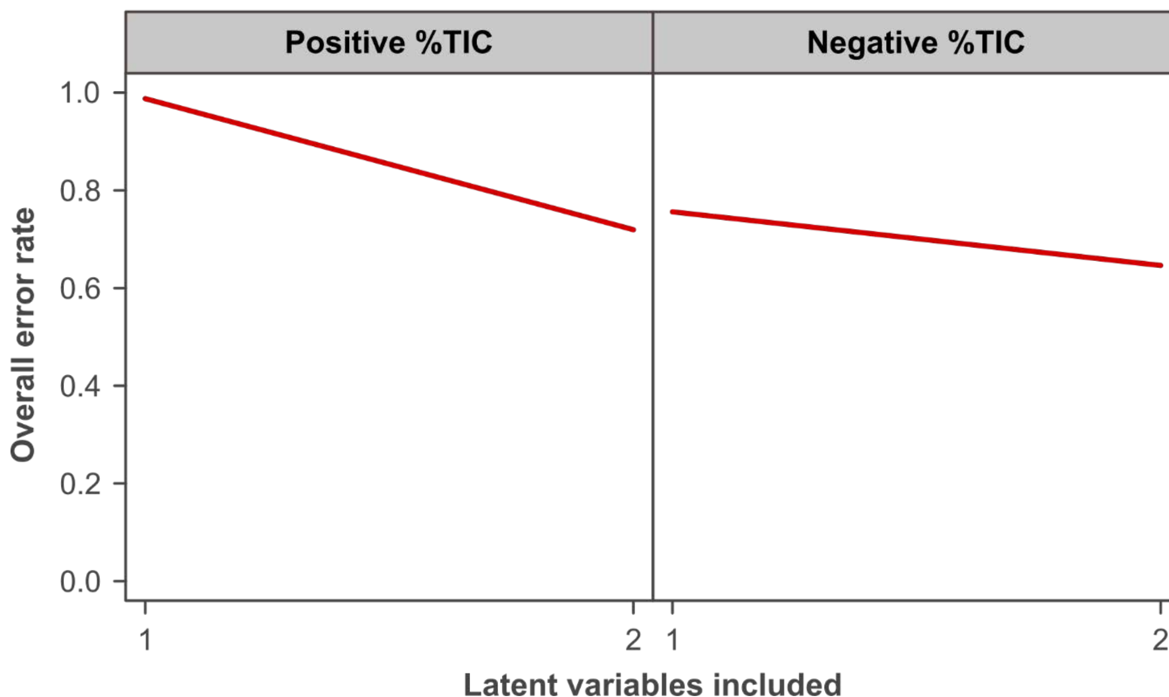


Figure 5.7: Overall error rate (i.e. what proportion of individuals are misclassified across all groups) of the partial least square-discriminant analysis (PLS-DA) models for both positive and negative %TIC.

As with PCA, the latent variables of PLS-DA are composed of contributions from all mass bins, and the contribution of the latter to each varies. Due to the extreme number of variables within each dataset only the ten highest loadings for each axis are provided. The mass bins with the highest contributions to the latent variables derived from positive %TIC profiles range from 184 to 647.2 Da, a very similar range to high ranking loadings in the equivalent PCA (184 to 632.2 Da). Here as there many are strongly correlated with each other, as indicated by their proximity (**Figure 5.5B**). The most important mass bins to PLS-DA of negative %TIC profiles range from 192.6 – 592.4 Da, also a very similar to result to that of PCA on the same data (192.4 – 590.4 Da). There was a comparable pattern in the masses of important variables, with the majority of high ranking bins being between 242.4 and 300.4 Da. Overall, the supervised ordinations do not provide any more information than the unsupervised examples. Given the considerable overlap in high ranking variables and the similar patterns observed, they appear to be summarising the same patterns of variation between individuals. Taken together with the lack of or weakness of group structure and the resulting high classification error rates, no putative identifications of metabolites are given

for PLS-DA for the sake of brevity. Similarly no comparison between supervised ordinations and recorded metadata will be made.

Comparison of recorded metadata with PCA ordinations derived from both positive and negative %TIC revealed few apparent relationships. Wind speed (**Figure 5.8**), hours since dawn, time of year (**Figure 5.9**), order of extraction on the bench and the time spent in storage before analysis (**Figure 5.10**) varied independently of all axes examined. There is a potential relationship between relative humidity and temperature and the scores received in both ordinations (**Figure 5.8**, **Figure 5.11**). However by comparison with group membership it is apparent that these patterns in fact represent site of origin (i.e. individuals sampled at the same time in the same environmental context share similar metadata), rather than the variance summarised by the axis in question (cf. **Figure 5.3A**, **Figure 5.4A**, and BOVHL/ROUDS individuals respectively). Only the amount of photosynthetically active radiation (PAR) recorded appears related to the variance summarised by the first component of the ordination derived from positive %TIC across groups.

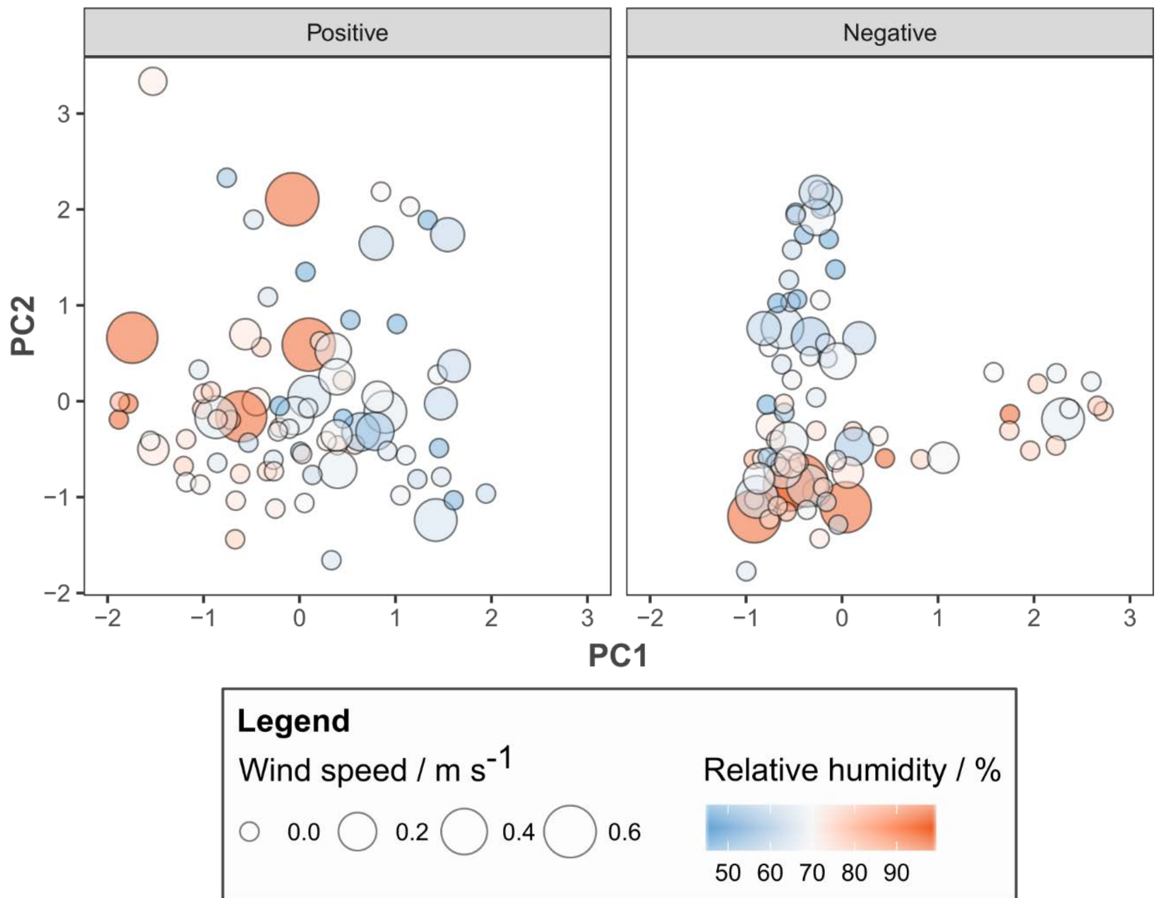


Figure 5.8: Principal components analysis of Pareto-scaled positive and negative percent total ion content (%TIC), coded by metadata indicating the environmental context at time of sampling. Individuals are represented by filled circles whose colour indicates the relative humidity while the size of these circles reflects the recorded wind speed.

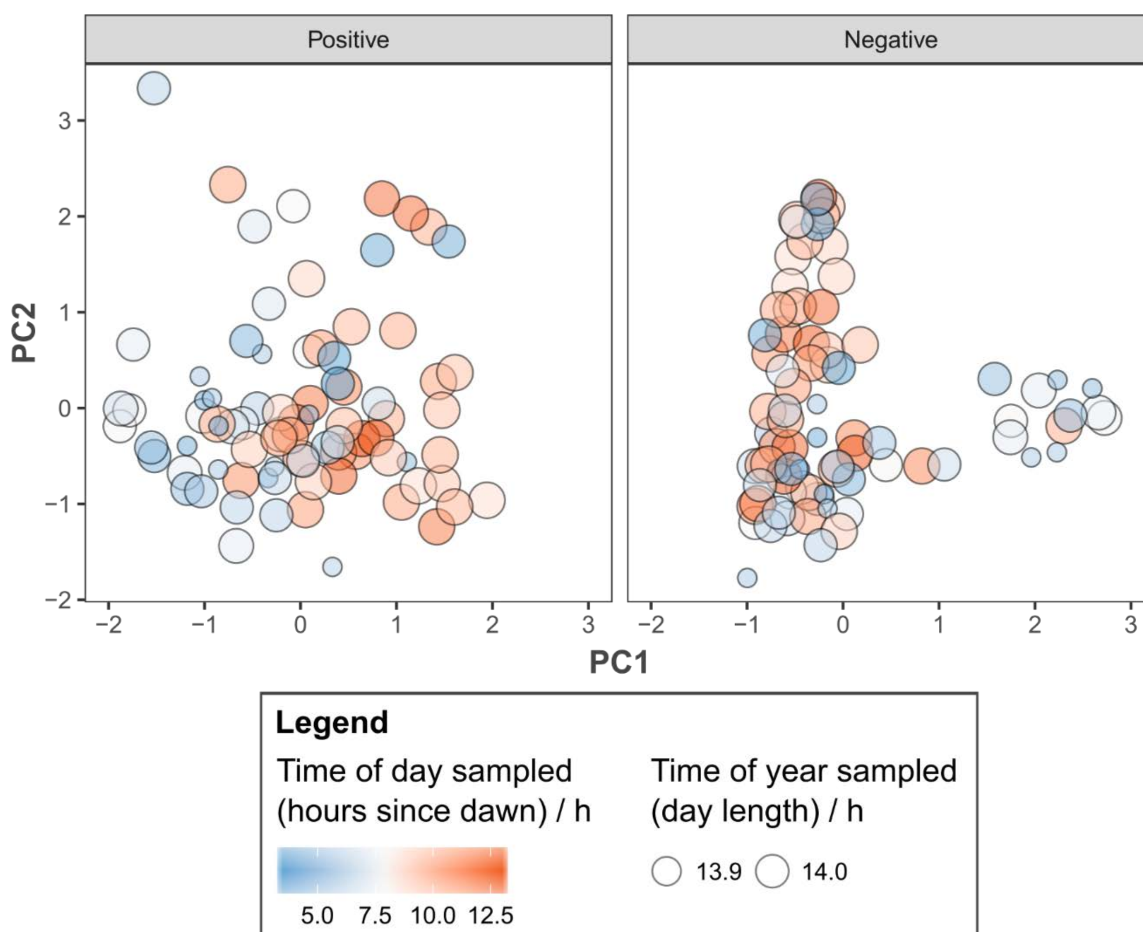


Figure 5.9: Principal components analysis of Pareto-scaled positive and negative percent total ion content (%TIC), coded by temporal metadata. Individuals are represented by filled circles whose colour indicates the time of the day (as hours of daylight since dawn) they were sampled. The size of these circles reflects the sampling date (as day length).

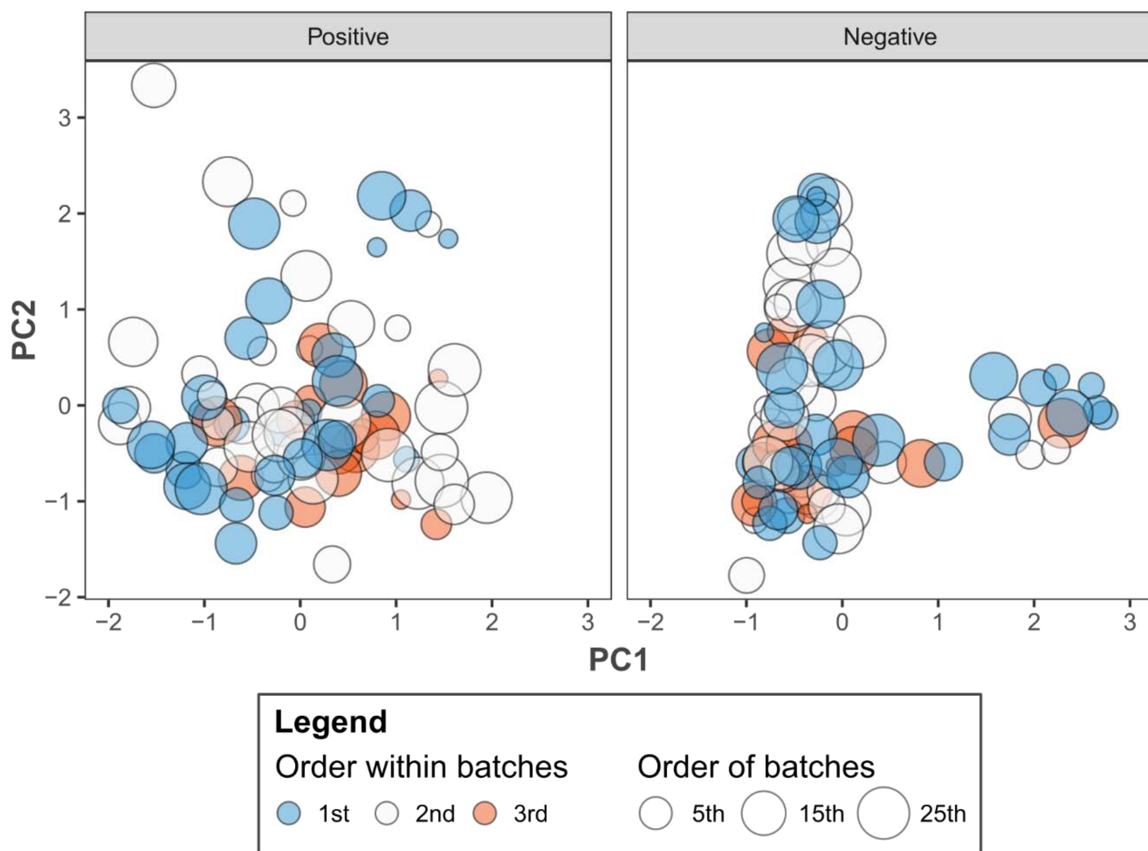


Figure 5.10: Principal components analysis of Pareto-scaled positive and negative percent total ion content (%TIC), coded by metadata describing laboratory procedure. Individuals are represented by filled circles whose colour indicates the order in which samples were processed on the bench, while the size of these circles reflects the overall order of metabolite extraction.

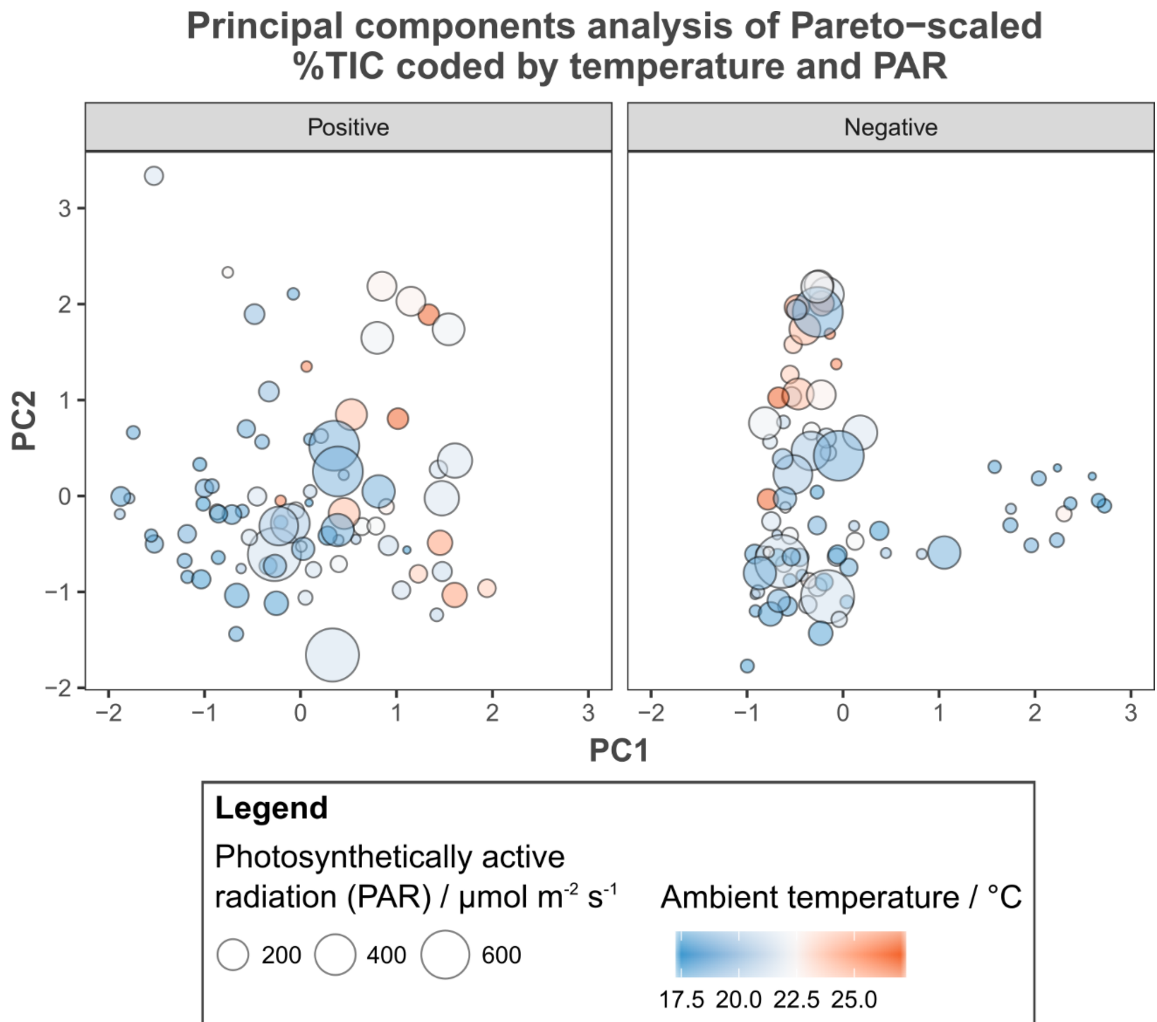


Figure 5.11: Principal components analysis of Pareto-scaled positive and negative percent total ion content (%TIC), coded by metadata indicating the environmental context at time of sampling. Individuals are represented by filled circles whose colour indicates the ambient air temperature while the size of these circles reflects the amount of photosynthetically active radiation potentially received.

An examination of neutral genetic variation produced broadly similar results to that of metabolic variation, with differentiation between groups being low. Genetic distance (proportion of shared alleles) was however significantly but weakly related to geographic distance between individuals (R^2 = 0.054; F = 2.25, df. = 1, p = 0.013), indicating the occurrence of isolation by distance. Low differentiation is reflected in a principal components analysis of allele frequencies which exhibits no group structure (**Figure 5.12**).

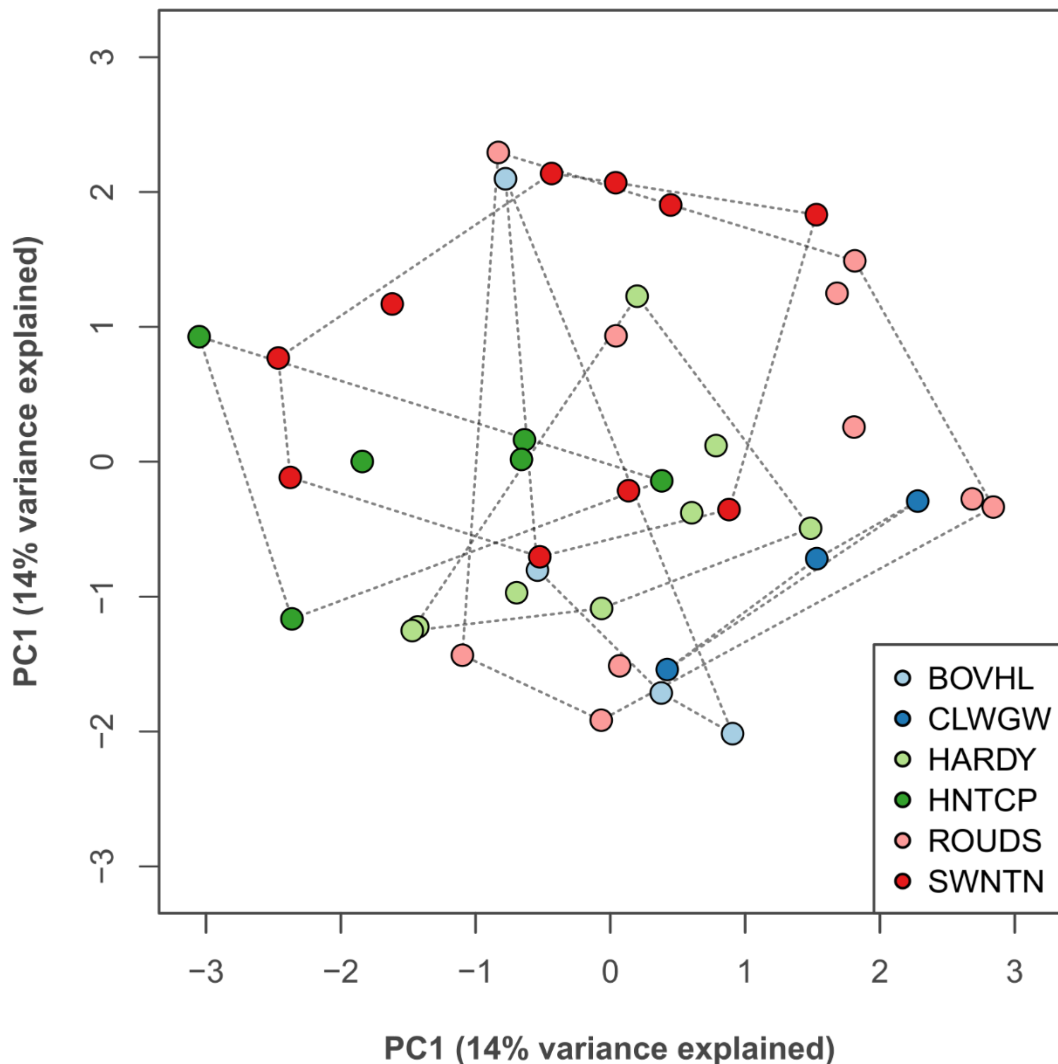


Figure 5.12: Principal components analysis (PCA) of microsatellite allele frequency data for all sampled individuals. Only the first two components (PC1/PC2) are shown. Filled coloured circles represent the scores received by each individual for each of these axes. Site groups are surrounded by a convex hull (dashed line).

In contrast to positive %TIC profiles and neutral genetic variation the measured functional trait was distinct between certain locations. SLA was significantly different between sites according to a Kruskal-Wallis test (H = 19.384, df. = 5, $p < 0.01$). A Bonferroni-corrected Dunn's test for multiple comparisons indicated that this result was generated by significantly different median SLA at ROUDS versus all other sites bar HARDY (

Table 5.4). Variation in SLA was also lower in ROUDS than at other locations (**Figure 5.13**).

	BOVHL	CLWGW	HARDY	HNTCP	ROUDS
CLWGW	-0.66	-	-	-	-
	1.00				
HARDY	0.90	1.56	-	-	-
	1.00	*0.90*			
HNTCP	-0.13	0.59	-1.18	-	-
	1.00	*1.00*	*1.00*		
ROUDS	**2.79**	**3.27**	2.32	**3.36**	-
	0.04	***0.01***	*0.15*	***0.01***	
SWNTN	1.43	2.04	0.61	1.79	-1.81
	1.00	*0.31*	*1.00*	*0.55*	*0.53*

Table 5.4: Results of a Bonferroni-corrected Dunn's multiple comparison test of specific leaf area between all pairwise combinations of sites. White cells contain the test statistic for that comparison, z, while the grey values are calculated p values. Statistically significant comparisons ($p < 0.05$) are indicated by green cells.

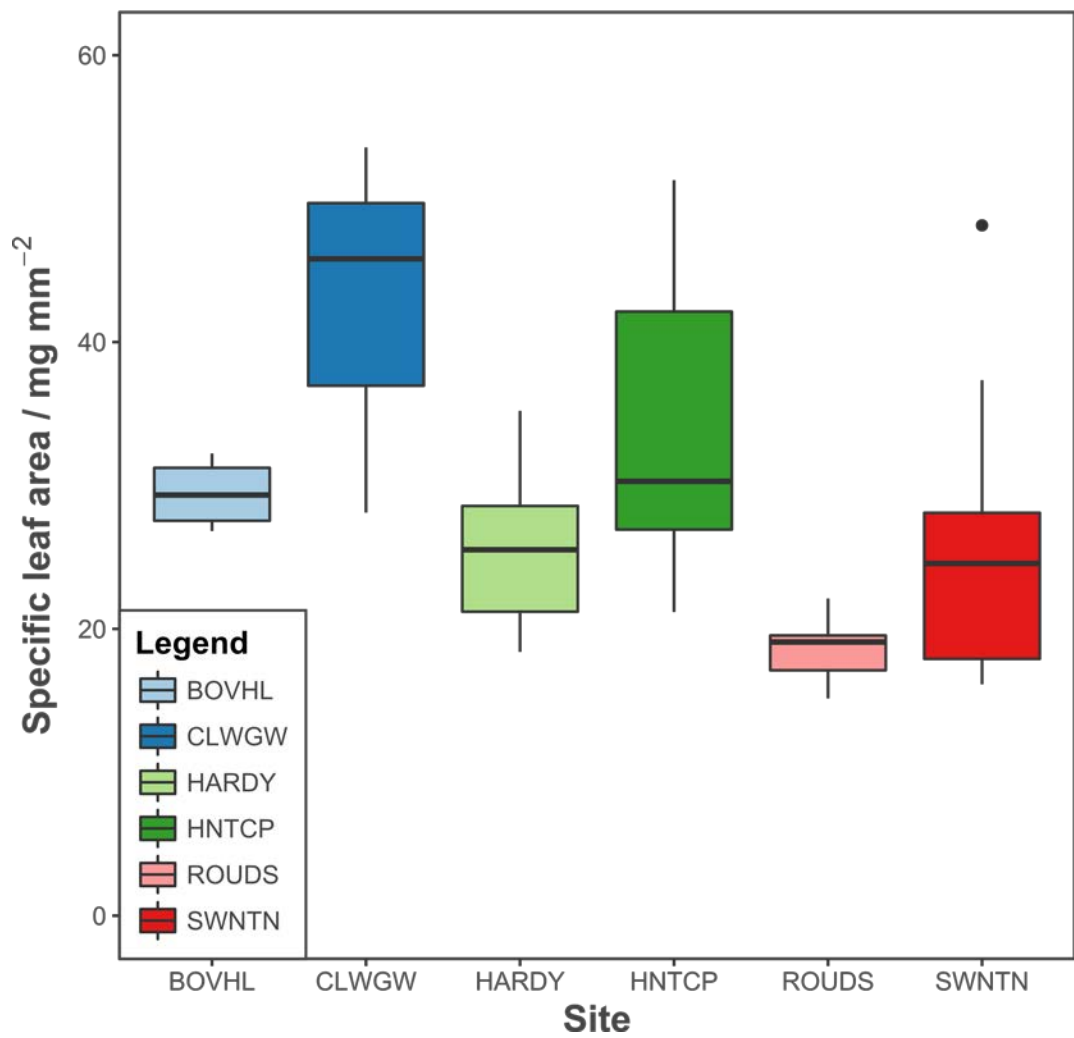


Figure 5.13: Box-and-whisker plot of specific leaf area (SLA: one-sided leaf area / leaf dry mass) for all sampled individuals across all sites.

5.4 Discussion

Metabolomics approaches have been used to detect the signatures of local adaption to distinct selective pressures as well as elucidate plant-environment interactions (Davey *et al.*, 2008; Bundy, Davey and Viant, 2009; Brunetti *et al.*, 2013), both of which have implications for population persistence in response to rapid environmental change (Jump, Marchant and Peñuelas, 2009; Alfaro *et al.*, 2014). Here for the first time an exploratory approach to characterizing *in situ* metabolic and morphological variation within and between populations was used. This variation was also contrasted with neutral genetic variation derived from microsatellite markers.

Metabolic fingerprints derived from leaf material collected and fixed *in situ* from individuals in six populations of *T. cordata* across its UK range were broadly similar despite contrasting environmental conditions during sampling. Both unsupervised (PCA) and supervised (PLS-DA) ordinations revealed little group structure, illustrating that most of the variation in metabolite concentrations summarised by these techniques occurred across rather than between populations. These analyses reasonably summarised datasets with extremely high dimensionality (~3500 variables), reducing them down to two variables which described an appreciable amount of the variance in the overall datasets. Populations exhibited considerable overlap and this was reflected in the high overall error rates in the classification performed by PLS-DA (>60%).

Despite this two patterns were observed in ordinations of negatively charged metabolite concentrations. Firstly a small group of individuals were distinct from the main cluster, but this outlying group was mostly unrelated to site of origin, containing members of all populations bar ROUDS. The most important metabolites involved in the generation of this structure were in the 242.4 to 300.4 Da mass range. Secondly individuals from both SWNTN and ROUDS exhibited a greater amount of variability in a subset of all metabolites, which was observed in both supervised and unsupervised ordinations (**Figure 5.4A**, **Figure 5.6A**). The same methods applied to positively charged metabolite concentrations did not reveal any patterns at all, either in the arrangement of individuals within ordinations or in the variable loadings. Unsupervised ordinations did reveal heterogeneity in intra-individual variation however. Many biological replicates were clustered tightly, but others appeared quite distinct as indicated by the length of the connecting lines in **Figure 5.3B** and **Figure 5.4B**.

Putative identifications of metabolites were made based on their detected masses. Searches of the KEGG database provided many candidates for the most important positive %TIC mass bins, but only two for those of negative %TIC. Broadly speaking there was no pattern to the former identifications, with a variety of compounds involved in many different metabolic processes being returned, although the majority of proposed substances within bin 381 (ranked 2nd and 3rd respectively in PC1 and PC2 from PCA of positive %TIC) were potentially oligosaccharides such as sucrose. These loadings are positive, and so individuals with higher scores along these axes possess higher concentrations of simple carbohydrates. All of the potential identifications made for positive %TIC bins should be interpreted with caution due to the nature of the ionization process used. More specifically the charged metabolites detected during MALDI-TOF spectrometry are composed of the compound itself (M) and the addition or removal other ions (H+, Na+, K+). Positively charged ions can take the form of $[M+H]^+$, $[M+Na]^+$ or $[M+K]^+$, while negatively charged ions are only formed via the removal of a proton ($[M-H]^-$). Therefore the mass of positive metabolites generated via MALDI is more variable than that of negatively charged (i.e. the mass of M +1, +23, or +39 Da versus the mass of M – 1 Da), which generates a wider range of putative identifications when matching detected masses with databases of biological compounds.

In contrast there was a distinct pattern to most loadings of negative %TIC ordinations, as indicated by the fairly narrow mass range of the highest ranking bins. Unfortunately given the greater precision of detected masses for these metabolites few candidate compounds were returned from the KEGG database. Only the bins 576.4 and 269.4 resulted in matches, with phosophoribosyl-formimino-AICAR-phosphate (PRFAR) and estrone respectively being suggested. PRFAR is in intermediate in histidine biosynthesis, and histidine, being a proteinogenic amino acid, is involved in many aspects of plant metabolism (Stepansky and Leustek, 2006). Estrone is an exclusively mammalian hormone and so this identification is likely in error (Kuhl, 2005), but many plant compounds in the flavonoid group have similar molecular weights and structures (e.g. isoflavones; Coward et al., 1993). It is possible that the patterns present in both PCA and PLS-DA of negative %TIC are created primarily by differences in concentration of this class of secondary metabolites, several of which have already been observed in the species previously (Negri, Santi and Tabach, 2013). These compounds also often occur in leaf material as glycosides (i.e. with the addition of a sugar molecule), bringing their molecular weight into the 500 – 600 Da range which is in line with the more massive high ranking bins. If concentrations of secondary metabolites are indeed

generating these structures then the lack of putative identifications is not surprising, since although primary metabolites are often conserved between taxa and are therefore easily identified, the majority of secondary products are genus or even species specific (Macel, van dam and Keurentjes, 2010). Whatever their identity, most of the highest ranking loadings are negative, and so individuals scoring highly along both axes have lower concentrations of these compounds.

Both PCA and PLS-DA approaches produced equivalent results with similar variable loadings, and therefore metadata was not compared with the supervised ordinations for the sake of brevity. Comparison of recorded metadata with the output of both PCA suggested that relative humidity and temperature at the time of sampling may have been related to metabolite concentrations, but by cross-referencing group identity with these apparent patterns it becomes clear that they reflect similarity due to spatial and temporal proximity. In other words, individuals from the same site possess similar environmental metadata, but not other individuals with similar metabolic fingerprints, indicating that neither of these two factors were responsible. Similarly neither wind speed nor time of sampling (either time of day or year) appeared related to the variation summarised by PCA. More importantly none of the laboratory metadata was related either, which alongside distinct extraction blanks illustrates that the results reflect actual variation in metabolic status at the time of sampling rather than induced variation as a result of technical issues (van den Berg *et al.*, 2006). However, incident PAR appeared related to the first principal component derived from positive %TIC, which is not unexpected given that this is a key determinant of photosynthetic rate (Hopkins, 1995). It could be argued that the lack of a link between temperature and PC1 (and therefore oligosaccharide concentration) does not support this interpretation, since carbon fixation is an enzymatic reaction and therefore its rate (and the generation of carbohydrates) is partially dependent on temperature. However the rate of CO_2 fixation reaches a plateau before dropping off and so all sampled individuals may be within the optimum temperature range for the species (Hew, Krotkov and Canvin, 1969; Sage and Kubien, 2007). Both the lack of an obvious relationship between metadata and metabolite concentrations and the diversity of metabolic roles played by flavonoids (Hernández *et al.*, 2009) make it difficult to comment on what other external environmental factors might be driving the variation summarised by ordinations of negative %TIC. For example variation in flavonoid concentration can be caused by an array of processes including biotic and abiotic stressors such as temperature, herbivory, plant-pathogen interactions, drought, nutrient

deprivation, etc. (Dixon and Paiva, 1995; Winkel-Shirley, 2001, 2002; Sampaio, Edrada-Ebel and Da Costa, 2016). Generally speaking metabolic fingerprints are the result of a dense network of biosynthesis routes (Benning and Stitt, 2004) and so the plastic portion of the variance examined here is likely generated by a variety of environmental factors.

Non-metabolic data returned mixed results. Allele frequencies produced similar patterns to metabolite concentrations, i.e. no group structure, indicating low differentiation between populations (little between-site variation). Inter-individual distance was however weakly related to geographic location, indicating the presence of weak IBD. This is in line with previously observed results (see **Chapter 4**, Logan et al., 2015), which shows that the sample size was sufficient to detect genetic structure. Conversely, leaf morphology was distinct between sites, with SLA lower at ROUDS than all other more southerly sites bar HARDY. Along with BOVHL it is also possesses low variability relative to the other groups. ROUDS experiences the lowest average temperature while BOVHL experiences both the highest average temperature and the lowest level of precipitation of all sampled sites (Met Office, 2017). Since both temperature and moisture regime affect SLA, these results are consistent with observations that abiotic filtering is stronger in certain environments. For instance lower leaf area per unit mass as well as reduced variability in the same has been observed along both latitudinal and altitudinal gradients of temperature (Li, Suzuki and Hara, 1998; Scheepens, Frei and Stöcklin, 2010; Hulshof *et al.*, 2013), as well as in response to reduced precipitation (Poorter *et al.*, 2009; Robson *et al.*, 2012). In adverse growing conditions or a less productive environment, lower SLA can provide a fitness advantage (Poorter *et al.*, 2009).

Whether differences in SLA and variation in (potential) secondary metabolite concentrations represent responses to local selection or phenotypic plasticity cannot be determined here. Both however have been observed to be components of local adaptation in other tree species. For instance more drought-tolerant *Fagus sylvatica* individuals in Spanish populations had higher flavonoid concentrations than more northern populations when grown in a common garden setting (Aranda *et al.*, 2017), and similarly differentiation in SLA was linked to drought-tolerance in populations of *Quercus suber* (Ramírez-Valiente *et al.*, 2010). Temperature has also been observed to be a cause of local selective differences that generate metabolic differentiation between populations, e.g. changes in nitrogen metabolism in *Arabidopsis lyrata* along a latitudinal gradient (Davey *et al.*, 2008). Stochastic processes (e.g. genetic drift, founder effect) can also cause chemical differentiation (Linhart

and Grant, 1996), but these should affect neutral genetic variation too. Given the low genetic differentiation observed, any of the observed differences between populations (variation in flavonoid concentration, SLA changes) can be expected to be the result of local selective pressures provided they have a genetic basis.

This study shows that metabolite profiles derived from samples collected *in situ* from natural populations have potential in a hypothesis generating role, as does non-targeted metabolite profiling generally (Macel, van dam and Keurentjes, 2010). Although not evidence of local adaptation the results here suggest fruitful avenues for exploring local adaptation in the future. For instance, now that a functional trait difference between populations has been established, it should be examined further, to determine if it has a genetic basis, by either common garden or glasshouse experiments or more realistically, given the potential logistical difficulties of such an approach in *T. cordata,* a targeted genetic study (e.g. sequencing of homologous genes linked to SLA in other species; ter Steege et al., 2005; Yin et al., 1999). Similarly further chemical analysis of the variable unidentified metabolites generating the structure in negative %TIC, with a view to elucidating the cause of such variation and ultimately assessing the presence of a genetic basis if necessary. The partially clonal nature of the species may also represent a way to explore the genetic basis of both SLA and metabolic differences by comparing intra- and inter-clonal variation in homogenous environments. These results also illustrate the high level plasticity of *T. cordata,* as it was able to maintain very similar metabolic profiles within distinct environmental settings. Even if the distinctions made between populations here do not reflect the presence of adaptive variation, this phenotypic plasticity may assist population persistence in the future (Aitken *et al.*, 2008; Lande, 2009; Nicotra *et al.*, 2010).

6 Conclusions

The aims of this study were to examine ecological, genetic and metabolic variation in UK *T. cordata* individuals or populations in order to answer questions or generate hypotheses regarding intra-generic competition, reproductive strategy, responses to habitat fragmentation and phenotypic plasticity, and provide suggestions as to how these results relate to population persistence in response to rapid environmental change. The key findings support an ecological distinction between *T. cordata* and its sympatric congeneric *T. platyphyllos*, illustrate that clonal reproduction is ubiquitous and correlated weakly with climate and demography, show that populations have likely not responded to habitat fragmentation and finally suggest that the species is highly plastic due to individuals possessing consistent metabolic fingerprints despite environmental variation.

6.1 Segregation of *Tilia cordata* Mill. and *T. platyphyllos* Scop. (Malvaceae) along environmental gradients

Based on broad descriptions within the literature, *T. cordata* and *T. platyphyllos* were expected to possess distinct ecological requirements, both edaphic and topographical. The former should also exhibit greater variation in examined parameters as is reflected in its greater abundance and wider distribution. The results obtained were in general agreement with these expectations, as *T. cordata* was found to occupy more fertile soils and areas with higher potential insolation, and also to exhibit greater variation in most characteristics. In contrast to previous descriptions no difference in soil acidity was observed. Other examined variables such as soil depth or moisture were not distinct between the species. Due to the method used to calculate potential insolation (McCune, 2007), differences in preferred light levels may actually reflect distinct topographical preferences, as *T. platyphyllos* is noted to typically occupy steep slopes or cliffs. Overall the segregation along the niche axes examined here was not as great as expected given previous descriptions of ecological preferences, but the species may differ along unexamined axes or at other points in their life cycle (e.g. juvenile mortality differences in response to light levels). The results suggest approaches to direct future studies of competitive exclusion within the genus.

6.2 Reproductive strategy of *Tilia cordata* Mill. across its UK range

Data show that *T. cordata* has mixed reproductive modes across its UK range but that the relative contribution of sexual versus asexual recruitment varies. Neither monoclonal stands nor a total lack of clones was ever observed, which contrasts with studies of other partially clonal tree species for unknown reasons. Despite this distinction levels of clonality were fairly typical for a canopy tree. Based on previous descriptions of clonality within the species, the most northern populations had proportionally fewer clones than expected; in these locations vegetative propagation has been suggested as the reason for population persistence, and yet young individuals with distinct genotypes were observed. This suggests that sexual recruitment in marginal locations is more common than previously thought.

Clonality was also not typically dominated by any one lineage. Clonal groups were also mostly small in terms of both number of members and spatial size which is expected given that basal shoots are the principle method of vegetative propagation, although occasional relatively long-distance dispersal events were inferred based on the distances separating genetically identical individuals. As a result of the low group size clonality did not generate excessive amounts of spatial genetic structure and is not expected to affect outcrossing success via competition with foreign pollen. Similarly, even though asexual reproduction was ubiquitous genotypic richness was still appreciable even in the most marginal populations.

There was a weak relationship between both climate and demography and the incidence of clonality observed in samples, meaning that the proportion of individuals that are likely to be the result of sexual reproduction was negatively correlated with latitude and the proportion of mature trees. The link with climate was not as strong as expected given the role of summer temperatures in limiting *T. cordata* fertility, and so the failure of sexual recruitment alone is not the only factor that promotes vegetative propagation in the species. Specifically the link with demography suggests it has a role as a rapid response to disturbance (here coppicing), as has been observed in other species. This aspect should be explored further by directly comparing time since coppicing with incidence of clonality. More southern locations including marginal Mediterranean populations should also be examined to determine if the identified trend in declining clonal incidence continues.

6.3 Cryptic spatial genetic structure of a canopy tree (*Tilia cordata*) across a fragmented landscape

The results of this study confirm that despite significant fragmentation, populations of *T. cordata* have not suffered from genetic erosion and maintain high levels of genetic diversity, as has been observed previously. Populations closer to the northern range edge of the species did not have reduced variation compared to those in a more southern location, despite elevated levels of clonality in the former. At a landscape scale (5 – 10 km) population substructure in *T. cordata* demes as a result of dispersal limitations was evident. The pattern of isolation by distance (IBD) was also reflected in a weak association between genetic distance and geographic location, as well as in clinal structures produced by exploratory ordination methods. This finding contradicts a previous observation of no IBD in UK *T. cordata* populations (Logan, Phuekvilai and Wolff, 2015), a fact which illustrates the increased power of the individual-based approach used here compared to population-based schemes.

Broadly speaking these structures were not related to the pattern of habitat fragmentation observed, i.e. there was low differentiation between woodland fragments for the most part. Similarly there was no relationship between the landscape context (e.g. surrounding land use) and genetic distance. Several of the smallest or most spatially isolated fragments however were more genetically distinct than other demes. It was inferred that the former pattern likely reflects historic processes operating over a formerly contiguous landscape rather than contemporary dispersal limits. Although separating these two explanations definitively is not possible here, *T. cordata* genotypes should possess extreme longevity due to long individual lifespans and the common occurrence of vegetative propagation. Combined with intermittent fertility the turnover within populations is likely to be extremely low and as a result, even in the absence of any gene flow, drift should be slow to act. Further the weak relationship between structure and location and the lack of a relationship between landscape context and individual differentiation also suggest that it has not responded to a reduction in population size and gene flow, except in the locations with the lowest numbers of individuals.

This latter point suggests that this genetic erosion may still occur in more populations given sufficient time. This supports the continuation of genetic resource conservation efforts and highlights which locations are most in danger of genetic erosion. Further research in

areas with higher rates of sexual recruitment or comparisons of genetic diversity between distinct cohorts could both help elucidate the response of the species to fragmentation. The relative importance of seed versus pollen mediated gene flow should also be examined to identify the routes by which connectivity occurs which may inform landscape conservation measures.

6.4 Exploring metabolic variation in natural populations of a temperate canopy tree, *Tilia cordata*

Metabolic fingerprints obtained from *T. cordata* leaves collected and fixed in a natural context exhibited greater variability within rather than between populations, producing results that showed little group structure. The most variable metabolites were diverse and involved in a variety of biochemical pathways based on the putative identifications made. There was little apparent correlation between the environmental context at the time of sampling and the aforementioned variation with the exception of light level, perhaps as a result of differences in photosynthetic activity, an inference based on differences in oligosaccharide concentrations. Importantly the same applied to laboratory metadata indicating that recorded fingerprints reflected actual metabolic status at the time of sampling. The similarity of data across sites despite differences in environmental factors suggests that *T. cordata* possesses high levels of phenotypic plasticity.

Although sites were metabolically similar overall two locations exhibited greater variability in concentrations of a group of unidentified compounds, potentially flavonoids based on their molecular masses. One of these locations was the most northern population sampled, Roudsea Wood, and individuals here also had significantly different specific leaf area to all other more southern sites, bar one (the second most northern site, Hardy Gang). Whether these morphological and metabolic differences reflect genetic differentiation or phenotypic plasticity cannot be determined here. If the former is responsible then local adaptation to more marginal conditions is likely responsible due to the similarity of neutral genetic variation observed between sites. More specifically, stochastic processes such as drift can also generate biochemical differences between populations, but these will also affect selectively neutral variation, producing separation between populations which was not observed in the microsatellite marker data. These results suggest that a future direction is to examine in a targeted fashion the presence of local adaptation in the species, and they also

demonstrate that metabolite profiles derived from samples collected *in situ* in natural populations have potential in a hypothesis generating role.

6.5 Recommendations for future research

- Niche separation between *Tilia* congenerics is not confirmed; if competitive exclusion promotes coexistence between *T. cordata* and *T. platyphyllos* then further tests can be made, such a detecting niche shifts when species occur in sympatry.
- A fuller understanding of the balance between reproductive modes across the entire range of *T. cordata* should be obtained; does the latitudinal trend in declining incidence of clonality continue into continental Europe to the point at which there are totally sexual populations, or is there some level of clonality across the entire range? Similarly, does clonality become a more optimal strategy in populations which are marginal for different reasons to those in the UK, such as Italian locations which are moisture limited?
- Further research into the role of asexual reproduction in *T. cordata*; it may be a response to disturbance and this could be examined more closely by comparing incidence of clonality between populations with distinct disturbance regimes.
- The benefits (or disadvantages) of clonal reproduction should be explored, such as increases or decreases in outcrossing success via enlarged floral displays.
- The possibility of significant lag time from fragmentation to genetic erosion should be examined by contrasting genetic diversity and its relationship with landscape context in populations with likely higher turnover such as those in more continental regions. If the process of genetic erosion is occurring this could also be detected by contrasting diversity between mature and juvenile cohorts of individuals.
- If any response to fragmentation is detected in larger populations then knowledge of the constraints on connectivity is important, but this is complicated by the fact that plants can respond twice via restriction of both pollen movement and seed dispersal. The relative influence of each may be inferred via comparison of differentiation exhibited by bi- versus uni-parentally molecular markers.
- The genetic basis of the morphological functional trait difference observed in *T. cordata* populations should be determined, most practically by a targeted genetic study (e.g. sequencing of homologous genes linked to SLA in other species). Other approaches such as common garden or reciprocal transplants are likely to be impractical given intermittent fertility of the species in the examined locations.

- Identification of the most variable unidentified metabolites should be made so that ultimately the cause of said variation and the presence of a genetic basis if necessary can be determined.
- The genetic basis of both SLA and metabolic differences may also be explored by exploiting the partially clonal nature of the species to examine and compare the intra- and inter-clonal variation of these aspects.

6.6 Concluding remarks

This study expands our knowledge of *T. cordata* ecology, reproductive strategy, population connectivity and variation. The findings have implications for population persistence and the application of effective genetic conservation methods. Firstly although clonality is ubiquitous, even the most marginal UK populations still possess an appreciable level of genetic diversity, and the typically small size and scale of clonal groups should not reduce outcrossing success or increase levels of inbreeding. Therefore asexual reproduction is not expected to have a negative impact on population persistence in the future, and may even enhance it if sexual reproduction fails due to unfavourable climatic changes. Secondly this confirms that small populations are particularly vulnerable to genetic erosion even in marginal locations with low turnover and supports the continued conservation of genetic resources. Secondly the phenotypic plasticity observed will likely be important in population persistence although whether it is sufficient to provide the capacity to weather the potential changes in marginal southern populations is unclear. Finally, the variation in a functional trait and certain metabolite concentrations hint at the possibility of local adaptation which may help in longer term.

Acknowledgements

Thank you to all stakeholders for permission to sample and fielding constant queries regarding lime trees, in particular the Natural England, National Trust and Cumbria Wildlife Trust staff that provided a great deal of useful information and guidance. I should thank Craig McCoy and the rest of the team at Arnside & Silverdale Natural Trust specifically for taking the time to walk around and point out particularly elusive trees. Many people also provided an extra pair of hands during fieldwork for which I'm very grateful, so thank you to Mark Ashton, Robert Ashton, Jordan Bilsborrow, Clare Bugg, Pedro de Carvalho, Thom Dallimore, Colin Davies, Emma Dumbill, Bruna Garbes, Marie Heaps, Emily Simmons and Jon Tomkinson your assistance. Thank you to my supervisory team, Professor Paul Ashton, Dr. Fay Voller and Dr. Clare Strode, for advice and support and some excellent form signing. Similarly thank you also to Dr Matt Davey for his collaboration and assistance with the final data chapter. Thank you to my family and friends for general support and enduring any thesis-related grumpiness from me above and beyond my baseline grumpiness. Specifically I should thank the Bank of Mum & Dad for their favourable interest rates, which helped a lot with the whole not being homeless thing when the studentship money ran out. On that note Chapter 2 was part funded by a Plant Study Grant from the Botanical Society of Britain & Ireland so thanks to them too.

References

Adamack, A. T. and Gruber, B. (2014) 'PopGenReport: Simplifying basic population genetic analyses in R', *Methods in Ecology and Evolution*. Edited by S. Dray, 5(4), pp. 384–387.

Agapow, P.-M. and Burt, A. (2001) 'Indices of multilocus linkage disequilibrium', *Molecular Ecology Notes*. Blackwell Science Ltd, 1(1–2), pp. 101–102. doi: 10.1046/j.1471-8278.2000.00014.x.

Aguilar, R. *et al.* (2008) 'Genetic consequences of habitat fragmentation in plant populations: Susceptible signals in plant traits and methodological approaches', *Molecular Ecology*, 17(24), pp. 5177–5188. doi: 10.1111/j.1365-294X.2008.03971.x.

Aitken, S. N. *et al.* (2008) 'Adaptation, migration or extirpation: climate change outcomes for tree populations', *Evolutionary Applications*. WILEY-BLACKWELL PUBLISHING, INC, COMMERCE PLACE, 350 MAIN ST, MALDEN 02148, MA USA, 1(1), pp. 95–111. doi: 10.1111/j.1752-4571.2007.00013.x.

Alberto, F. *et al.* (2005) 'Within-population spatial genetic structure, neighbourhood size and clonal subrange in the seagrass Cymodocea nodosa.', *Molecular ecology*, 14(9), pp. 2669–81. doi: 10.1111/j.1365-294X.2005.02640.x.

Alberto, F. J. *et al.* (2013) 'Potential for evolutionary responses to climate change - evidence from tree populations', *Global Change Biology*, 19(6), pp. 1645–1661. doi: 10.1111/gcb.12181.

Alfaro, R. I. *et al.* (2014) 'The role of forest genetic resources in responding to biotic and abiotic factors in the context of anthropogenic climate change', *Forest Ecology and Management*, 333, pp. 76–87. doi: 10.1016/j.foreco.2014.04.006.

Alfonso-Corrado, C. *et al.* (2005) 'Clonal and genetic structure of two Mexican oaks: Quercus eduardii and Quercus potosina (Fagaceae)', *Evolutionary Ecology*, 18(5–6), pp. 585–599. doi: 10.1007/s10682-004-5145-5.

Alford, D. V. (1975) *Bumblebees*. 1st edn. London: Davis-Poynter.

Allen, C. D. *et al.* (2010) 'A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests', *Forest Ecology and Management*. Elsevier, 259(4), pp. 660–684. doi: 10.1016/j.foreco.2009.09.001.

Anderson, E. (1949) *Introgressive Hybridisation*. London: Chapman & Hall Limited.

Andrén, H. (1994) ‘Effects of Habitat Fragmentation on Birds and Mammals in Landscapes with Different Proportions of Suitable Habitat: A Review’, *Oikos*, 71(3), p. 355. doi: 10.2307/3545823.

Anmarkrud, J. A. *et al.* (2008) ‘Microsatellite evolution: mutations, sequence variation and homoplasy in the hypervariable avian microsatellite locus HrU10’, *BMC Evolutionary Biology*, 8(138).

Aponte, C. *et al.* (2011) ‘Oak trees and soil interactions in Mediterranean forests: a positive feedback model’, *Journal of Vegetation Science*, 22, pp. 856–867.

Aranda, I. *et al.* (2017) ‘Fagus sylvatica L. provenances maintain different leaf metabolic profiles and functional response’, *Acta Oecologica*. Elsevier Masson, 82, pp. 1–9. doi: 10.1016/j.actao.2017.05.003.

Ardehed, A. *et al.* (2015) ‘Complex spatial clonal structure in the macroalgae Fucus radicans with both sexual and asexual recruitment.’, *Ecology and evolution*, 5(19), pp. 4233–45. doi: 10.1002/ece3.1629.

Arnaud-Haond, S., Duarte, C. M., *et al.* (2007) ‘Standardizing methods to address clonality in population studies.’, *Molecular Ecology*. WILEY-BLACKWELL, 16(24), pp. 5115–39. doi: 10.1111/j.1365-294X.2007.03535.x.

Arnaud-Haond, S., Migliaccio, M., *et al.* (2007) ‘Vicariance patterns in the Mediterranean Sea: east?west cleavage and low dispersal in the endemic seagrass Posidonia oceanica’, *Journal of Biogeography*, 34(6), pp. 963–976. doi: 10.1111/j.1365-2699.2006.01671.x.

Arnaud-Haond, S. and Bailleul, D. (2015) ‘RClone: Partially Clonal Populations Analysis’.

Ashrafi, S. *et al.* (2011) ‘Trophic niche partitioning of cryptic species of long-eared bats in Switzerland: implications for conservation’, *European Journal of Wildlife Research*. SPRINGER, 57(4), pp. 843–849. doi: 10.1007/s10344-011-0496-z.

Ashworth, L. *et al.* (2004) ‘Why do pollination generalist and specialist plant species show similar reproductive susceptibility to habitat fragmentation?’, *Journal of Ecology*, 92(4), pp. 717–719. doi: 10.1111/j.0022-0477.2004.00910.x.

Attorre, F. *et al.* (2011) ‘Evaluating the effects of climate change on tree species abundance

and distribution in the Italian peninsula', *Applied Vegetation Science*. Blackwell Publishing Ltd, 14(2), pp. 242–255. doi: 10.1111/j.1654-109X.2010.01114.x.

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

Baier, P., Pennerstorfer, J. and Schopf, A. (2007) 'PHENIPS-A comprehensive phenology model of Ips typographus (L.) (Col., Scolytinae) as a tool for hazard rating of bark beetle infestation', *Forest Ecology and Management*. Elsevier, 249(3), pp. 171–186. doi: 10.1016/j.foreco.2007.05.020.

Balloux, F. (2001) 'EASYPOP (Version 1.7): A Computer Program for Population Genetics Simulations', *Journal of Heredity*. OXFORD UNIV PRESS INC, 92(3), pp. 301–302. doi: 10.1093/jhered/92.3.301.

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

Barker, M. and Rayens, W. (2003) 'Partial least squares for discrimination', *Journal of Chemometrics*. John Wiley & Sons, Ltd., 17(3), pp. 166–173. doi: 10.1002/cem.785.

Barrett, S. C. H. (2015) 'Influences of clonality on plant sexual reproduction', *Proceedings of the National Academy of Sciences*, 112(29), pp. 8859–8866. doi: 10.1073/pnas.1501712112.

Barsoum, N., Muller, E. and Skot, L. (2004) 'Variations in levels of clonality among Populus nigra L. stands of different ages', in *Evolutionary Ecology*. Kluwer Academic Publishers, pp. 601–624. doi: 10.1007/s10682-004-5146-4.

Bartlett, M. S. (1937) 'Properties of Sufficiency and Statistical Tests', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. The Royal Society, 160(901), pp. 268–282. doi: 10.1098/rspa.1937.0109.

Baum, C. F. (2008) 'Modeling proportions', *The Stata Journal*, 8(2), pp. 299–303.

Belcher, S. E., Finnigan, J. J. and Harman, I. N. (2008) 'Flows through forest canopies in complex terrain', *Ecological Applications*, 18(6), pp. 1436–1453. doi: 10.1890/06-1894.1.

Benning, C. and Stitt, M. (2004) 'Physiology and metabolism: Reacting to the full

complexity of metabolic pathways in a postgenomic era', *Current Opinion in Plant Biology*. Elsevier Current Trends, 7(3), pp. 231–234. doi: 10.1016/j.pbi.2004.03.010.

van den Berg, R. A. *et al.* (2006) 'Centering, scaling, and transformations: improving the biological information content of metabolomics data.', *BMC genomics*. BioMed Central, 7, p. 142. doi: 10.1186/1471-2164-7-142.

Bertrand, R. *et al.* (2011) 'Changes in plant community composition lag behind climate warming in lowland forests', *Nature*. Nature Research, 479(7374), pp. 517–520. doi: 10.1038/nature10548.

Betts, M. G. *et al.* (2017) 'Global forest loss disproportionately erodes biodiversity in intact landscapes', *Nature*. Nature Research, 547(7664), pp. 441–444. doi: 10.1038/nature23285.

Birks, H. J. B. (1989) 'Holocene Isochrone Maps and Patterns of Tree-Spreading in the British Isles', *Journal of Biogeography*. BLACKWELL SCIENCE LTD, 16(6), pp. 503–540. doi: 10.2307/2845208.

Bivand, R. and Lewin-Koh, N. (2017) 'maptools: Tools for Reading and Handling Spatial Objects'.

Booy, G. *et al.* (2000) 'Genetic diversity and the survival of populations', *Plant Biology*, pp. 379–395. doi: 10.1055/s-2000-5958.

Borcard, D. and Legendre, P. (2012) 'Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study.', *Ecology*, 93(6), pp. 1473–81.

Bowcock, A. M. *et al.* (1994) 'High resolution of human evolutionary trees with polymorphic microsatellites', *Nature*, 368(6470), pp. 455–457. doi: 10.1038/368455a0.

Bradshaw, A. D. (2006) 'Unravelling phenotypic plasticity - Why should we bother?', *New Phytologist*. Blackwell Publishing Ltd, pp. 644–648. doi: 10.1111/j.1469-8137.2006.01761.x.

Bréda, N. and Badeau, V. (2008) 'Forest tree responses to extreme drought and some biotic events: Towards a selection according to hazard tolerance?', *Comptes Rendus - Geoscience*. Elsevier Masson, pp. 651–662. doi: 10.1016/j.crte.2008.08.003.

Breed, M. F. *et al.* (2015) 'Mating patterns and pollinator mobility are critical traits in forest fragmentation genetics', *Heredity*, 115(2), pp. 108–114. doi: 10.1038/hdy.2013.48.

Brown, A. H. D., Feldman, M. W. and Nevo, E. (1980) 'Multilocus structure of natural populations of Hordeum spontaneum', *Genetics*. Genetics Society of America, 96, pp. 523–536.

Brunetti, C. *et al.* (2013) 'Metabolomics in plant environmental physiology', *Journal of Experimental Botany*, 64(13), pp. 4011–4020.

Brzosko, E. *et al.* (2013) 'High level of genetic variation within clonal orchid Goodyera repens', *Plant Systematics and Evolution*. SPRINGER WIEN. doi: 10.1007/s00606-013-0817-7.

Buiteveld, J. *et al.* (2016) 'Human impact on the genetic diversity of Dutch field elm (Ulmus minor) populations in the Netherlands: Implications for conservation', *Plant Ecology and Evolution*, 149(2), pp. 165–176. doi: 10.5091/plecevo.2016.977.

Bundy, J. G., Davey, M. P. and Viant, M. R. (2009) 'Environmental metabolomics: a critical review and future perspectives', *Metabolomics*, 5, pp. 3–21.

Burns, J. H. and Strauss, S. Y. (2011) 'More closely related species are more ecologically similar in an experimental test.', *Proceedings of the National Academy of Sciences*. NATL ACAD SCIENCES, 108(13), pp. 5302–5307. doi: 10.1073/pnas.1013003108.

Canty, A. and Ripley, B. (2015) 'boot: Bootstrap R (S-Plus) Functions. R package version 1.3-14'.

Le Cao, K.-A. *et al.* (2017) 'mixOmics: Omics Data Integration Project'.

Cavender-Bares, J., Kitajima, K. and Bazzaz, F. A. (2004) 'Multiple trait associations in relation to habitat differentiation among 17 Floridian oak species', *Ecological Monographs*. ECOLOGICAL SOC AMER, 74(4), pp. 635–662. doi: 10.1890/03-4007.

Cavender-Bares, J. and Pahlich, A. (2009) 'Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, Quercus virginiana and Q. geminata (Fagaceae).', *American journal of botany*. BOTANICAL SOC AMER INC, 96(9), pp. 1690–702. doi: 10.3732/ajb.0800315.

Charpentier, A. (2001) 'Consequences of clonal growth for plant mating', *Evolutionary Ecology*. Kluwer Academic Publishers, 15(4–6), pp. 521–530. doi: 10.1023/A:1016057503722.

Chen, I.-C. *et al.* (2011) ‘Rapid Range Shifts of Species Associated with High Levels of Climate Warming’, *Science*, 333(6045), pp. 1024–1026. doi: 10.1126/science.1206432.

Chevin, L.-M., Lande, R. and Mace, G. M. (2010) ‘Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory’, *PLoS Biology*. Public Library of Science, 8(4), p. e1000357. doi: 10.1371/journal.pbio.1000357.

Christmas, M. J., Breed, M. F. and Lowe, A. J. (2015) ‘Constraints to and conservation implications for climate change adaptation in plants’, *Conservation Genetics*, 17(2), pp. 305–320. doi: 10.1007/s10592-015-0782-5.

Corbet, S. A., Unwin, D. M. and Prŷr-Jones, O. E. (1979) ‘Humidity, nectar and insect visits to flowers, with special reference to Crataegus, Tilia and Echium’, *Ecological Entomology*. Blackwell Publishing Ltd, 4(1), pp. 9–22. doi: 10.1111/j.1365-2311.1979.tb00557.x.

Corbet, S. A., Williams, I. H. and Osborne, J. L. (1991) ‘Bees and the pollination of crops and wild flowers in the European Community’, *Bee World*. Taylor & Francis, 72(2), pp. 47–59. doi: 10.1080/0005772X.1991.11099079.

Corlett, R. T. and Westcott, D. A. (2013) ‘Will plant movements keep up with climate change?’, *Trends in ecology & evolution*, 28(8), pp. 482–8. doi: 10.1016/j.tree.2013.04.003.

Cornelissen, J. H. C. *et al.* (2003) ‘A handbook of protocols for standardised and easy measurement of plant functional traits worldwide’, *Australian Journal of Botany*. CSIRO PUBLISHING, pp. 335–380. doi: 10.1071/BT02124.

Cottrell, J. E. *et al.* (2003) ‘Comparison of fine-scale genetic structure using nuclear microsatellites within two British oakwoods differing in population history’, *Forest Ecology and Management*, 176(1–3), pp. 287–303. doi: 10.1016/S0378-1127(02)00289-X.

Coward, L. *et al.* (1993) ‘Genistein, Daidzein, and Their β-Glycoside Conjugates: Antitumor Isoflavones in Soybean Foods from American and Asian Diets’, *Journal of Agricultural and Food Chemistry*. American Chemical Society, 41(11), pp. 1961–1967. doi: 10.1021/jf00035a027.

Cribari-Neto, F. and Zeileis, A. (2010) ‘Beta Regression in R’, *Journal of Statistical Software*, 34(2), pp. 1–24.

Cristóbal, D. *et al.* (2014) ‘Clonal structure and dynamics of peripheral Populus tremula L. populations’, *IForest*. SISEF - Italian Society of Silviculture and Forest Ecology, 7(3), pp.

140–149. doi: 10.3832/ifor0971-007.

Cushman, S. A. *et al.* (2006) ‘Gene flow in complex landscapes: testing multiple hypotheses with causal modeling’, *Am Nat*, 168(4), pp. 486–499. doi: 10.1086/506976.

Dammhahn, M., Soarimalala, V. and Goodman, S. M. (2013) ‘Trophic Niche Differentiation and Microhabitat Utilization in a Species-rich Montane Forest Small Mammal Community of Eastern Madagascar’, *Biotropica*. WILEY-BLACKWELL, 45(1), pp. 111–118. doi: 10.1111/j.1744-7429.2012.00893.x.

Davey, M. P. *et al.* (2008) ‘Population-specific metabolic phenotypes of Arabidopsis lyrata ssp. petraea.’, *The New Phytologist*. BLACKWELL PUBLISHING, 177(2), pp. 380–8. doi: 10.1111/j.1469-8137.2007.02282.x.

Davis, B. A. S. *et al.* (2003) ‘The temperature of Europe during the Holocene reconstructed from pollen data’, *Quaternary Science Reviews*, 22(15–17), pp. 1701–1716. doi: 10.1016/S0277-3791(03)00173-2.

DeFries, R. S. *et al.* (2002) ‘Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 99(22), pp. 14256–14261. doi: 10.1073/pnas.182560099.

Dering, M., Chybicki, I. J. and Rączka, G. (2015) ‘Clonality as a driver of spatial genetic structure in populations of clonal tree species’, *Journal of Plant Research*. Springer Japan, 128(5), pp. 731–745. doi: 10.1007/s10265-015-0742-7.

Desprez-Loustau, M.-L. *et al.* (2006) ‘Interactive effects of drought and pathogens in forest trees’, *Annals of Forest Science*. EDP Sciences, 63(6), pp. 597–612. doi: 10.1051/forest:2006040.

Desprez-Loustau, M.-L. *et al.* (2007) ‘Simulating the effects of a climate-change scenario on the geographical range and activity of forest-pathogenic fungi’, *Canadian Journal of Plant Pathology*. Taylor & Francis Group, 29(2), pp. 101–120. doi: 10.1080/07060660709507447.

Didham, R. K. *et al.* (1996) ‘Insects in fragmented forests: A functional approach’, *Trends in Ecology and Evolution*, pp. 255–260. doi: 10.1016/0169-5347(96)20047-3.

Diemer, M. and Schmid, B. (2001) ‘Effects of biodiversity loss and disturbance on the

survival and performance of two Ranunculus species with differing clonal architectures', *Ecography*. MUNKSGAARD INT PUBL LTD, 24(1), pp. 59–67. doi: 10.1034/j.1600-0587.2001.240108.x.

Dixon, R. A. and Paiva, N. L. (1995) 'Stress-Induced Phenylpropanoid Metabolism.', *The Plant cell*. American Society of Plant Biologists, 7(7), pp. 1085–1097. doi: 10.1105/tpc.7.7.1085.

Dorken, M. E. and Eckert, C. G. (2001) 'Severely reduced sexual reproduction in northern populations of a clonal plant, Decodon verticillatus (Lythraceae)', *Journal of Ecology*. BLACKWELL SCIENCE LTD, 89(3), pp. 339–350. doi: 10.1046/j.1365-2745.2001.00558.x.

Douhovnikoff, V. and Dodd, R. S. (2003) 'Intra-clonal variation and a similarity threshold for identification of clones: application to Salix exigua using AFLP molecular markers.', *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*. SPRINGER-VERLAG, 106(7), pp. 1307–15. doi: 10.1007/s00122-003-1200-9.

Dubreuil, M. *et al.* (2010) 'Genetic effects of chronic habitat fragmentation revisited: Strong genetic structure in a temperate tree, Taxus baccata (Taxaceae), with great dispersal capability.', *American journal of botany*, 97(2), pp. 303–10. doi: 10.3732/ajb.0900148.

Dunn, O. J. (1964) 'Multiple Comparisons Using Rank Sums', *Technometrics*. Taylor & Francis, Ltd.American Statistical AssociationAmerican Society for Quality, 6(3), p. 241. doi: 10.2307/1266041.

Eckert, C. G. (2002) 'The loss of sex in clonal plants', *Evolutionary Ecology*, 15, pp. 501–520. doi: 10.1023/a:1016005519651.

Ellenberg, H. (1988) *Vegetation Ecology of Central Europe*. 4th edn. Cambridge: Cambridge University Press.

EUFORGEN (2009) *Distribution map of Lime (Tilia cordata)*. Available at: www.euforgen.org.

Evans, J. P. and Morris, A. B. (2016) 'Isolated coastal populations of Tilia americana var. Caroliniana persist long-term through vegetative growth', *American Journal of Botany*. Botanical Society of America, 103(9), pp. 1687–1693. doi: 10.3732/ajb.1600233.

Favre-Bac, L. *et al.* (2015) 'Ditch network sustains functional connectivity and influences

patterns of gene flow in an intensive agricultural landscape', *Heredity*. Nature Publishing Group, 116(2), pp. 1–13. doi: 10.1038/hdy.2015.90.

Ferrari, S. and Cribari-Neto, F. (2004) 'Beta Regression for Modelling Rates and Proportions', *Journal of Applied Statistics*, 31(7), pp. 799–815. doi: 10.1080/0266476042000214501.

Fick, S. E. and Hijmans, R. J. (2017) 'WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas', *International Journal of Climatology*, 15 May. doi: 10.1002/joc.5086.

Fiehn, O. (2002) 'Metabolomics - The link between genotypes and phenotypes', *Plant Molecular Biology*. Kluwer Academic Publishers, 48(1–2), pp. 155–171. doi: 10.1023/A:1013713905833.

Fiehn, O., Kloska, S. and Altmann, T. (2001) 'Integrated studies on plant biology using multiparallel techniques', *Current Opinion in Biotechnology*. Elsevier Current Trends, pp. 82–86. doi: 10.1016/S0958-1669(00)00165-8.

Field, K. J. and Lake, J. A. (2011) 'Environmental metabolomics links genotype to phenotype and predicts genotype abundance in wild plant populations.', *Physiologia Plantarum*. WILEY-BLACKWELL, 142(4), pp. 352–60. doi: 10.1111/j.1399-3054.2011.01480.x.

Finzi, A. C., Van Breemen, N. and Canham, C. D. (1998a) 'Canopy tree-soil interactions within temperate forests: Species effects on soil carbon and nitrogen', *Ecological Applications*, pp. 440–446. doi: 10.2307/2641084.

Finzi, A. C., Van Breemen, N. and Canham, C. D. (1998b) 'Canopy tree soil interactions within temperate forests: Species effects on soil carbon and nitrogen', *Ecological Applications*, 8(2), pp. 440–446. doi: 10.2307/2641083.

Flannigan, M. D., Stocks, B. J. and Wotton, B. M. (2000) 'Climate change and forest fires', *Science of the Total Environment*. Elsevier, 262(3), pp. 221–229. doi: 10.1016/S0048-9697(00)00524-6.

Free, J. B. (1970) 'The flower constancy of bumblebees', *Journal of Animal Ecology*, 39(2), pp. 395–402.

Fromm, M. (2001) 'Reproduktion einer entomophilen Baumart in geringer

Populationsdichte – Das Beispiel der Winterlinde (Tilia cordata Mill .)'.

Fuentes-Utrilla, P. *et al.* (2014) 'Population clustering and clonal structure evidence the relict state of Ulmus minor Mill. in the Balearic Islands.', *Heredity*. Nature Publishing Group, 113(1), pp. 21–31. doi: 10.1038/hdy.2014.21.

Galeuchet, D. J., Perret, C. and Fischer, M. (2005) 'Microsatellite variation and structure of 28 populations of the common wetland plant, Lychnis flos-cuculi L., in a fragmented landscape', *Molecular Ecology*, 14(4), pp. 991–1000. doi: 10.1111/j.1365-294X.2005.02485.x.

Gause, G. F. (1934) *The Struggle for Existence*. Baltimore: The Williams & Wilkins company.

Gibon, Y. *et al.* (2006) 'Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes.', *Genome biology*. BioMed Central, 7(8), p. R76. doi: 10.1186/gb-2006-7-8-r76.

Gómez-Aparicio, L. and Canham, C. D. (2008) 'NEIGHBORHOOD MODELS OF THE EFFECTS OF INVASIVE TREE SPECIES ON ECOSYSTEM PROCESSES', *Ecological Monographs*. ECOLOGICAL SOC AMER, 1707 H ST NW, STE 400, WASHINGTON, DC 20006-3915 USA, 78(1), pp. 69–86. doi: 10.1890/06-2036.1.

Goodacre, R. *et al.* (2007) 'Proposed minimum reporting standards for data analysis in metabolomics', *Metabolomics*. Springer US, 3(3), pp. 231–241. doi: 10.1007/s11306-007-0081-3.

Gosler, A. G. (1990) 'Introgressive hybridization between Crataegus monogyna Jacq. and Crataegus laevigata Poiret DC. in the Upper Thames Valley', *Watsonia*, 18(1), pp. 49–62.

Gosling, P. (2007) *Raising trees and shrubs from seed. Foresty Commision Practice Guide*. Edinburgh: Forestry Commission.

Goss, E. M. *et al.* (2014) 'The Irish potato famine pathogen Phytophthora infestans originated in central Mexico rather than the Andes', *PNAS USA*. National Academy of Sciences, 111(24), pp. 8791–8796. doi: 10.1073/pnas.1401884111.

Goverde, M. *et al.* (2002) 'Small-scale habitat fragmentation effects on pollinator behaviour: Experimental evidence from the bumblebee Bombus veteranus on calcareous grasslands', *Biological Conservation*, pp. 293–299. doi: 10.1016/S0006-3207(01)00194-X.

Gray, L. K. and Hamann, A. (2013) 'Tracking suitable habitat for tree populations under climate change in western North America', *Climatic Change*. Springer Netherlands, 117(1–2), pp. 289–303. doi: 10.1007/s10584-012-0548-8.

Grundel, R. *et al.* (2010) 'Floral and nesting resources, habitat structure, and fire influence bee distribution across an open-forest gradient', *Ecological Applications*. Ecological Society of America, 20(6), pp. 1678–1692. doi: 10.1890/08-1792.1.

Guillemin, M. L. *et al.* (2008) 'Genetic variation in wild and cultivated populations of the haploid- diploid red alga Gracilaria chilensis: How farming practices favor asexual reproduction and heterozygosity', *Evolution*. Blackwell Publishing Inc, 62(6), pp. 1500–1519. doi: 10.1111/j.1558-5646.2008.00373.x.

Guillot, G. (2008) 'Inference of structure in subdivided populations at low levels of genetic differentiation - The correlated allele frequencies model revisited', *Bioinformatics*. Springer-Verlag, New York, 24(19), pp. 2222–2228. doi: 10.1093/bioinformatics/btn419.

Hale, M. L., Burg, T. M. and Steeves, T. E. (2012) 'Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies.', *PloS one*, 7(9), p. e45170. doi: 10.1371/journal.pone.0045170.

Halket, J. M. *et al.* (2005) 'Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS', in *Journal of Experimental Botany*. Oxford University Press, pp. 219–243. doi: 10.1093/jxb/eri069.

Halkett, F., Simon, J.-C. and Balloux, F. (2005) 'Tackling the population genetics of clonal and partially clonal organisms.', *Trends in ecology & evolution*. ELSEVIER SCIENCE LONDON, 20(4), pp. 194–201. doi: 10.1016/j.tree.2005.01.001.

Hampe, A. and Petit, R. J. (2005) 'Conserving biodiversity under climate change: The rear edge matters', *Ecology Letters*. Blackwell Science Ltd, pp. 461–467. doi: 10.1111/j.1461-0248.2005.00739.x.

Hamrick, J. L. (2004) 'Response of forest trees to global environmental changes', *Forest Ecology and Management*, 197(1–3), pp. 323–335. doi: 10.1016/j.foreco.2004.05.023.

Hardy, O. J. and Vekemans, X. (2002) 'spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels', *Molecular Ecology Notes*, 2(4), pp. 618–620. doi: 10.1046/j.1471-8286.2002.00305.x.

Harper, J. L. (1977) *Population Biology of Plants*. Academic Press.

Hartnett, D. C. (1993) ‘Regulation of clonal growth and dynamics of Panicum virgatum (Poaceae) in tallgrass prairie: Effects of neighbor removal and nutrient addition’, *American Journal of Botany*. Botanical Society of America, Inc., 80(10), pp. 1114–1120. doi: 10.2307/2445538.

Hastie, T., Tibshirani, R. and Friedman, J. (2001) *The Elements of statistical learning: Data mining, inference and prediction*. 2nd editio. Springer Series in Statistics.

Hemery, G. E. *et al.* (2010) ‘Growing scattered broadleaved tree species in Europe in a changing climate: a review of risks and opportunities’, *Forestry*. Edward Arnold, London, 83(1), pp. 65–81. doi: 10.1093/forestry/cpp034.

Hernández, I. *et al.* (2009) ‘How relevant are flavonoids as antioxidants in plants?’, *Trends in Plant Science*. Elsevier Current Trends, 14(3), pp. 125–132. doi: 10.1016/j.tplants.2008.12.003.

Hew, C. S., Krotkov, G. and Canvin, D. T. (1969) ‘Effects of Temperature on Photosynthesis and CO(2) Evolution in Light and Darkness by Green Leaves.’, *Plant physiology*. American Society of Plant Biologists, 44(5), pp. 671–7. doi: 10.1104/pp.44.5.671.

Hijmans, R. J. (2016) ‘raster: Geographic Data Analysis and Modeling’.

Hill, M. O., Preston, C. D. and Roy, D. B. (2004) *PLANTATT - attributes of British and Irish plants: status, size, life history, geography and habitats*. Abbotts Ripton: Centre forEcology & Hydrology.

Hirao, A. S. and Kudo, G. (2004) ‘Landscape genetics of alpine-snowbed plants: comparisons along geographic and snowmelt gradients.’, *Heredity*, 93(3), pp. 290–298. doi: 10.1038/sj.hdy.6800503.

Hlásny, T. and Turčáni, M. (2009) ‘Insect pests as climate change driven disturbances in forest ecosystems’, in *Bioclimatology and Natural Hazards*. Dordrecht: Springer Netherlands, pp. 165–177. doi: 10.1007/978-1-4020-8876-6_15.

Hobbie, S. E. (1992) ‘Effects of plant species on nutrient cycling.’, *Trends in ecology & evolution*. ELSEVIER SCI LTD, THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD, OXON, ENGLAND OX5 1GB, 7(10), pp. 336–9. doi: 10.1016/0169-5347(92)90126-V.

Holderegger, R. *et al.* (2010) ‘Landscape genetics of plants’, *Trends in Plant Science*, pp. 675–683. doi: 10.1016/j.tplants.2010.09.002.

Holsinger, K. E. (2000) ‘Reproductive systems and evolution in vascular plants.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 97(13), pp. 7037–42.

Honnay, O. and Bossuyt, B. (2005) ‘Prolonged clonal growth: Escape route or route to extinction?’, *Oikos*, pp. 427–432. doi: 10.1111/j.0030-1299.2005.13569.x.

Honnay, O. and Jacquemyn, H. (2007) ‘Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation’, *Conservation Biology*, 21(3), pp. 823–831. doi: 10.1111/j.1523-1739.2006.00646.x.

Hooftman, D. A. P. *et al.* (2004) ‘Genetic effects of habitat fragmentation on common species of Swiss fen meadows’, *Conservation Biology*, 18(4), pp. 1043–1051. doi: 10.1111/j.1523-1739.2004.00223.x.

Hopkins, W. G. (1995) *Introduction to Plant Physiology*. 1st ed. New York, United States: John Wiley & Sons, Inc.

Hubert, M., Rousseeuw, P. J. and Vanden Branden, K. (2005) ‘ROBPCA: A New Approach to Robust Principal Component Analysis’, *Technometrics*, 47(1), pp. 64–79.

Hulshof, C. M. *et al.* (2013) ‘Intra-specific and inter-specific variation in specific leaf area reveal the importance of abiotic and biotic drivers of species diversity across elevation and latitude’, *Journal of Vegetation Science*. Edited by F. de Bello, 24(5), pp. 921–931. doi: 10.1111/jvs.12041.

Hutchinson, G. E. (1957) ‘Concluding remarks’, *Cold Spring Harbor Symposium on Quantitative Biology*, 22(2), pp. 415–427. doi: 10.1159/000345197.

IPCC (2000) ‘Land Use, Land-Use Change and Forestry’, in *Special Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

IPCC (2007) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by S. Solomon et al. Cambridge: Cambridge University Press.

Itoh, A. *et al.* (2003) ‘Importance of topography and soil texture in the spatial distribution

of two sympatric dipterocarp trees in a Bornean rainforest', *Ecological Research*. SPRINGER JAPAN KK, 18(3), pp. 307–320. doi: 10.1046/j.1440-1703.2003.00556.x.

Iverson, L. R., Schwartz, M. W. and Prasad, A. M. (2004) 'How fast and far might tree species migrate in the eastern United States due to climate change?', *Global Ecology and Biogeography*, 13(3), pp. 209–219. doi: 10.1111/j.1466-822X.2004.00093.x.

Jacob, D. *et al.* (2014) 'EURO-CORDEX: New high-resolution climate change projections for European impact research', *Regional Environmental Change*. Springer Berlin Heidelberg, 14(2), pp. 563–578. doi: 10.1007/s10113-013-0499-2.

De Jaegere, T., Hein, S. and Claessens, H. (2016) 'A review of the characteristics of small-leaved lime (Tilia cordata Mill.) and their implications for silviculture in a changing climate', *Forests*. Multidisciplinary Digital Publishing Institute, 7(3), pp. 1–21. doi: 10.3390/f7030056.

Jarni, K., Jakše, J. and Brus, R. (2015) 'Vegetative propagation: linear barriers and somatic mutation affect the genetic structure of a Prunus avium L. stand', *Forestry*, 88(5), pp. 612–621. doi: 10.1093/forestry/cpv029.

Jha, S. and Kremen, C. (2013) 'Resource diversity and landscape-level homogeneity drive native bee foraging', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 110(2), pp. 555–558. doi: 10.1073/pnas.1208682110.

Johansson, T. and Lundh, J.-E. (1988) 'Sucker production from root cuttings of Populus tremula in relation to growing conditions', *Scandinavian Journal of Forest Research*, 3(1), pp. 75–82. doi: 10.1080/02827588809382497.

Jombart, T. (2008) 'adegenet: a R package for the multivariate analysis of genetic markers', *Bioinformatics*, 24, pp. 1403–1405. doi: 10.1093/bioinformatics/btn129.

Jombart, T. *et al.* (2008) 'Revealing cryptic spatial patterns in genetic variability by a new multivariate method', *Heredity*. Nature Publishing Group, 101(1), pp. 92–103. doi: 10.1038/hdy.2008.34.

Jombart, T. and Ahmed, I. (2011) 'adegenet 1.3-1: new tools for the analysis of genome-wide SNP data', *Bioinformatics*, 27(21), pp. 3070–3071. doi: 10.1093/bioinformatics/btr521.

Jones, O. A. H. *et al.* (2013) 'Metabolomics and its use in ecology', *Austral Ecology*.

WILEY-BLACKWELL, 38(6), pp. 713–720. doi: 10.1111/aec.12019.

Jump, A. S., Hunt, J. M. and Peñuelas, J. (2006) ‘Rapid climate change-related growth decline at the southern range edge of Fagus sylvatica’, *Global Change Biology*, 12(11), pp. 2163–2174. doi: 10.1111/j.1365-2486.2006.01250.x.

Jump, A. S., Marchant, R. and Peñuelas, J. (2009) ‘Environmental change and the option value of genetic diversity’, *Trends in Plant Science*. Elsevier Current Trends, pp. 51–58. doi: 10.1016/j.tplants.2008.10.002.

Jump, A. S. and Peñuelas, J. (2005) ‘Running to stand still: adaptation and the response of plants to rapid climate change’, *Ecology Letters*, 8(9), pp. 1010–1020. doi: 10.1111/j.1461-0248.2005.00796.x.

Jump, A. S. and Peñuelas, J. (2006) ‘Genetic effects of chronic habitat fragmentation in a wind-pollinated tree’, *Proceedings of the National Academy of Sciences USA*, 103(21), pp. 8096–8100. doi: 10.1073/pnas.0510127103.

Kamvar, Z. N., Brooks, J. C. and Grünwald, N. J. (2015) ‘Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality.’, *Frontiers in genetics*, 6, p. 208. doi: 10.3389/fgene.2015.00208.

Kamvar, Z. N., Tabima, J. F. and Grünwald, N. J. (2014) ‘Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction.’, *PeerJ*, 2, p. e281. doi: 10.7717/peerj.281.

Kanehisa Laboratories (2017) *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Available at: http://www.genome.jp/kegg/.

Kasran, B. and Nik, A. R. (1994) ‘Suspended sediment yield resulting from selective logging practices in a small watershed in peninsular Malaysia’, *Forest Research*. Forest Research Institute Malaysia, 7(2), pp. 286–295. doi: 10.2307/43581814.

Keller, L. F. and Waller, D. M. (2002) ‘Inbreeding effects in wild populations’, *Trends in Ecology and Evolution*, pp. 230–241. doi: 10.1016/S0169-5347(02)02489-8.

Kells, A. R. and Goulson, D. (2003) ‘Preferred nesting sites of bumblebee queens (Hymenoptera: Apidae) in agroecosystems in the UK’, *Biological Conservation*, 109(2), pp. 165–174. doi: 10.1016/S0006-3207(02)00131-3.

Kelly, R. P. *et al.* (2010) ‘A method for detecting population genetic structure in diverse, high gene-flow species’, *Journal of Heredity*. Island Press, Washington (DC), 101(4), pp. 423–436. doi: 10.1093/jhered/esq022.

Kiaer, L. P. *et al.* (2007) ‘Genealogy, morphology and fitness of spontaneous hybrids between wild and cultivated chicory (Cichorium intybus).’, *Heredity*. NATURE PUBLISHING GROUP, MACMILLAN BUILDING, 4 CRINAN ST, LONDON N1 9XW, ENGLAND, 99(1), pp. 112–20. doi: 10.1038/sj.hdy.6800973.

Kielland, K. (1994) ‘Amino acid absorption by Arctic plants: implications for plant nutrition and nitrogen cycling’, *Ecology*, 75(8), pp. 2373–2383.

Kierepka, E. M. and Latch, E. K. (2015) ‘Performance of partial statistics in individual-based landscape genetics’, *Molecular Ecology Resources*, 15(3), pp. 512–525. doi: 10.1111/1755-0998.12332.

Kierepka, E. M. and Latch, E. K. (2016) ‘Fine-scale landscape genetics of the American badger (Taxidea taxus): disentangling landscape effects and sampling artifacts in a poorly understood species.’, *Heredity*. Nature Publishing Group, 116(1), pp. 33–43. doi: 10.1038/hdy.2015.67.

Kim, H. K. and Verpoorte, R. (2010) ‘Sample preparation for plant metabolomics’, *Phytochemical Analysis*, pp. 4–13. doi: 10.1002/pca.1188.

Kitamura, K. and Kawano, S. (2001) ‘Regional Differentiation in Genetic Components for the American Beech, Fagus grandifolia Ehrh., in Relation to Geological History and Mode of Reproduction’, *Journal of Plant Research*. The Botanical Society of Japan, 114(3), pp. 353–368. doi: 10.1007/PL00013997.

Klekowski, E. J. (2003) ‘Plant clonality, mutation, diplontic selection and mutational meltdown’, *Biological Journal of the Linnean Society*, 79(1), pp. 61–67. doi: 10.1046/j.1095-8312.2003.00183.x.

Klimeš, L. *et al.* (1997) *Clonal plant architecture: a comparative analysis of form and function*. Edited by H. de Kroon and V. Groenendael. Leiden: Backhuys.

Knapp, E. E., Goedde, M. A. and Rice, K. J. (2001) ‘Pollen-limited reproduction in blue oak: Implications for wind pollination in fragmented populations’, *Oecologia*, 128(1), pp. 48–55. doi: 10.1007/s004420000623.

Kobe, R. K. *et al.* (1995) ‘Juvenile tree survivorship as a component of shade tolerance’, *Ecological Applications*, 5(2), pp. 517–532.

Kobe, R. K. (1999) ‘Light gradient partitioning among tropical tree species through differential seedling mortality and growth’, *Ecology*. ECOLOGICAL SOC AMER, 80(1), pp. 187–201. doi: 10.1890/0012-9658(1999)080[0187:LGPATT]2.0.CO;2.

Kosmides, A. K. *et al.* (2013) ‘Metabolomic fingerprinting: challenges and opportunities.’, *Critical reviews in biomedical engineering*. NIH Public Access, 41(3), pp. 205–21. doi: 10.1615/CritRevBiomedEng.2013007736.

Kramer, A. T. *et al.* (2008) ‘The paradox of forest fragmentation genetics.’, *Conservation biology : the journal of the Society for Conservation Biology*. WILEY-BLACKWELL, 22(4), pp. 878–85. doi: 10.1111/j.1523-1739.2008.00944.x.

Krasny, M. E. and Johnson, E. A. (1992) ‘Stand development in aspen clones’, *Canadian Journal of Forest Research*. NRC Research Press Ottawa, Canada, 22(9), pp. 1424–1429. doi: 10.1139/x92-190.

Kremer, A. *et al.* (2012) ‘Long-distance gene flow and adaptation of forest trees to rapid climate change’, *Ecology Letters*. Blackwell Publishing Ltd, pp. 378–392. doi: 10.1111/j.1461-0248.2012.01746.x.

Kuhl, H. (2005) ‘Pharmacology of estrogens and progestogens: Influence of different routes of administration’, *Climacteric*. Taylor & Francis, 8(SUPPL. 1), pp. 3–63. doi: 10.1080/13697130500148875.

Kujala, S. T. and Savolainen, O. (2012) ‘Sequence variation patterns along a latitudinal cline in Scots pine (Pinus sylvestris): Signs of clinal adaptation?’, *Tree Genetics and Genomes*. Springer-Verlag, 8(6), pp. 1451–1467. doi: 10.1007/s11295-012-0532-5.

Kunin, W. E. *et al.* (2009) ‘Variation at range margins across multiple spatial scales: environmental temperature, population genetics and metabolomic phenotype.’, *Proceedings. Biological sciences / The Royal Society*, 276(1661), pp. 1495–1506. doi: 10.1098/rspb.2008.1767.

Laakmann, S., Kochzius, M. and Auel, H. (2009) ‘Ecological niches of Arctic deep-sea copepods: Vertical partitioning, dietary preferences and different trophic levels minimize inter-specific competition’, *Deep Sea Research Part I: Oceanographic Research Papers*.

PERGAMON-ELSEVIER SCIENCE LTD, 56(5), pp. 741–756. doi: 10.1016/j.dsr.2008.12.017.

Lambin, E. F., Geist, H. J. and Lepers, E. (2003) 'Dynamics of land use and land cover change in tropical regions', *Annual Review of Environment and Resources*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 28(1), pp. 205–241. doi: 10.1146/annurev.energy.28.050302.105459.

Lande, R. (1988) 'Genetics and demography in biological conservation', *Science*, 241(4872), pp. 1455–1460. doi: 10.1126/science.3420403.

Lande, R. (2009) 'Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation', *Journal of Evolutionary Biology*. Blackwell Publishing Ltd, 22(7), pp. 1435–1446. doi: 10.1111/j.1420-9101.2009.01754.x.

Lapworth, A. and McGregor, J. (2008) 'Seasonal variation of the prevailing wind direction in Britain', *Weather*. John Wiley & Sons, Ltd., 63(12), pp. 361–364. doi: 10.1002/wea.242.

Lavorel, S. and Garnier, E. (2002) 'Predicting changes in community composition and ecosystem functioning from plant traits: Revisiting the Holy Grail', *Functional Ecology*. Blackwell Science Ltd, 16(5), pp. 545–556. doi: 10.1046/j.1365-2435.2002.00664.x.

Legendre, P. and Fortin, M.-J. (2010) 'Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data', *Molecular Ecology Resources*. Blackwell Publishing Ltd, 10(5), pp. 831–844. doi: 10.1111/j.1755-0998.2010.02866.x.

Legendre, P. and Legendre, L. (1998) *Numerical Ecology*. 3rd edn. Amsterdam: Elsevier.

Legendre, P., Oksanen, J. and ter Braak, C. J. F. (2011) 'Testing the significance of canonical axes in redundancy analysis', *Methods in Ecology and Evolution*. Blackwell Publishing Ltd, 2(3), pp. 269–277. doi: 10.1111/j.2041-210X.2010.00078.x.

Leimu, R. and Fischer, M. (2008) 'A meta-analysis of local adaptation in plants.', *PloS one*. Public Library of Science, 3(12), p. e4010. doi: 10.1371/journal.pone.0004010.

Li, B., Suzuki, J.-I. and Hara, T. (1998) 'Latitudinal variation in plant size and relative growth rate in Arabidopsis thaliana', *Oecologia*, 115(3), pp. 293–301. doi: 10.1007/s004420050519.

Lienert, J. *et al.* (2002) 'Isozyme variability of the wetland specialist Swertia perennis (Gentianaceae) in relation to habitat size, isolation, and plant fitness', *American Journal of Botany*, 89(5), pp. 801–811. doi: 10.3732/ajb.89.5.801.

Lindner, M. *et al.* (2010) 'Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems', *Forest Ecology and Management*. Elsevier, 259(4), pp. 698–709. doi: 10.1016/j.foreco.2009.09.023.

Linhart, Y. B. and Grant, M. C. (1996) 'Evolutionary significance of local genetic differentiation in plants', *Annual Review of Ecology and Systematics*, 27(1), pp. 237–277. doi: 10.1146/annurev.ecolsys.27.1.237.

Lloret, F., Penuelas, J. and Estiarte, M. (2004) 'Experimental evidence of reduced diversity of seedlings due to climate modification in a Mediterranean-type community', *Global Change Biology*. Blackwell Publishing, 10(2), pp. 248–258. doi: 10.1111/j.1365-2486.2004.00725.x.

Loarie, S. R. *et al.* (2009) 'The velocity of climate change', *Nature*. Nature Publishing Group, 462(7276), pp. 1052–1055. doi: 10.1038/nature08649.

Logan, S. A., Phuekvilai, P. and Wolff, K. (2015) 'Ancient woodlands in the limelight: delineation and genetic structure of ancient woodland species Tilia cordata and Tilia platyphyllos (Tiliaceae) in the UK', *Tree Genetics & Genomes*, 11(3), p. 52. doi: 10.1007/s11295-015-0872-z.

Loiselle, B. A. *et al.* (1995) 'Spatial Genetic Structure of a Tropical Understory Shrub, Psychotria officinalis (Rubiaceae)', *American journal of botany*, 81(11), pp. 1420–1425.

Losos, J. B. *et al.* (2003) 'Niche lability in the evolution of a Caribbean lizard community.', *Nature*. NATURE PUBLISHING GROUP, 424(6948), pp. 542–5. doi: 10.1038/nature01814.

Lovett, G. M. *et al.* (2004) 'Nitrogen cycling in a northern hardwood forest: Do species matter?', *Biogeochemistry*. KLUWER ACADEMIC PUBL, VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS, 67(3), pp. 289–308. doi: 10.1023/B:BIOG.0000015786.65466.f5.

Lowe, A. J. *et al.* (2005) 'Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees.', *Heredity*, 95(4),

pp. 255–73. doi: 10.1038/sj.hdy.6800725.

Macel, M., van dam, N. M. and Keurentjes, J. J. B. (2010) ‘Metabolomics: The chemistry between ecology and genetics’, *Molecular Ecology Resources*. Blackwell Publishing Ltd, 10(4), pp. 583–593. doi: 10.1111/j.1755-0998.2010.02854.x.

Magård, E. (2002) ‘Possible mechanisms for the coexistence of congeneric (Pourouma, Cecropiaceae) Amazonian tree species’, *Ecotropica*, 8, pp. 15–22.

Mamolos, A. P., Elisseou, G. K. and Veresoglou, D. S. (1995) ‘Depth of Root Activity of Coexisting Grassland Species in Relation to N and P Additions, Measured Using Nonradioactive Tracers’, *The Journal of Ecology*. BLACKWELL SCIENCE LTD, 83(4), p. 643. doi: 10.2307/2261632.

Manel, S. *et al.* (2003) ‘Landscape genetics: combining landscape ecology and population genetics’, *Trends in Ecology & Evolution*, 18(4), pp. 189–197. doi: 10.1016/S0169-5347(03)00008-9.

Manel, S. and Holderegger, R. (2013) ‘Ten years of landscape genetics.’, *Trends in Ecology & Evolution*, 28(10), pp. 614–21.

Matschiner, M. and Salzburger, W. (2009) ‘TANDEM: integrating automated allele binning into genetics and genomics workflows.’, *Bioinformatics (Oxford, England)*. OXFORD UNIV PRESS, GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND, 25(15), pp. 1982–3. doi: 10.1093/bioinformatics/btp303.

Matthews, S. N. *et al.* (2011) ‘Modifying climate change habitat models using tree species-specific assessments of model uncertainty and life history-factors’, *Forest Ecology and Management*, 262(8), pp. 1460–1472. doi: 10.1016/j.foreco.2011.06.047.

McCarthy, D. (2012) *Systematics and Phylogeography of the Genus Tilia in North America*. University of Illinois at Chicago.

McCune, B. (2007) ‘Improved estimates of incident radiation and heat load using non-parametric regression against topographic variables’, *Journal of Vegetation Science*, 18, pp. 751–754.

McCune, B. and Dylan, K. (2002) ‘Equations for potential annual direct incident radiation and heat load’, *Journal of Vegetation Science*, 13(4), pp. 603–606.

McKane, R. B. *et al.* (2002) ‘Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra’, *Nature*, 415, pp. 68–71.

McLachlan, J. S., Clark, J. S. and Manos, P. S. (2005) ‘Molecular indicators of tree migration capacity under rapid climate change’, *Ecology*. Ecological Society of America, 86(8), pp. 2088–2098. doi: 10.1890/04-1036.

de Meeûs, T., Prugnolle, F. and Agnew, P. (2007) ‘Asexual reproduction: Genetics and evolutionary aspects’, *Cellular and Molecular Life Sciences*. Birkhäuser-Verlag, pp. 1355–1372. doi: 10.1007/s00018-007-6515-2.

Meirmans, P. G. (2015) ‘Seven common mistakes in population genetics and how to avoid them’, *Molecular Ecology*, pp. 3223–3231. doi: 10.1111/mec.13243.

Meirmans, P. G. and Van Tienderen, P. H. (2004) ‘genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms’, *Molecular Ecology Notes*, 4(4), pp. 792–794. doi: 10.1111/j.1471-8286.2004.00770.x.

Met Office (2017) *Climate summaries*. Available at: https://www.metoffice.gov.uk/climate/uk/summaries (Accessed: 30 October 2017).

Michel, A. and Seidling, W. (2016) *Forest Condition in Europe: 2016 Technical Report of ICP Forests. Report under the UNECE Convention on Long-Range Transboundary Air Pollution (CLRTAP)*. Vienna, Austria.

Millar, M. A., Byrne, M. and Coates, D. J. (2010) ‘The maintenance of disparate levels of clonality, genetic diversity and genetic differentiation in disjunct subspecies of the rare Banksia ionthocarpa.’, *Molecular Ecology*. WILEY-BLACKWELL, 19, pp. 4217–4227. doi: 10.1111/j.1365-294X.2010.04817.x.

Mimura, M. and Aitken, S. N. (2010) ‘Local adaptation at the range peripheries of Sitka spruce’, *Journal of Evolutionary Biology*. Blackwell Publishing Ltd, 23(2), pp. 249–258. doi: 10.1111/j.1420-9101.2009.01910.x.

Mittell, E. A., Nakagawa, S. and Hadfield, J. D. (2015) ‘Are molecular markers useful predictors of adaptive potential?’, *Ecology letters*, 18(8), pp. 772–8. doi: 10.1111/ele.12454.

Montalvo, A. M. *et al.* (1997) ‘Population Structure, Genetic Diversity, and Clone Formation in Quercus chrysolepis (Fagaceae)’, *American Journal of Botany*. Botanical Society of America, Inc., 84(11), pp. 1553–1564. doi: 10.1017/S0007485311000721.

Morris, A. B. *et al.* (2014) ‘Mature beech trees (Fagus grandifolia; Fagaceae) are persistently clonal in coves and beech gaps in the Great Smoky Mountains.’, *American journal of botany*, 101(2), pp. 381–8. doi: 10.3732/ajb.1300161.

Morris, A. B., Small, R. L. and Cruzan, M. B. (2004) ‘Variation in Frequency of Clonal Reproduction Among Populations of Fagus grandifolia Ehrh. in Response to Disturbance’, *Castanea*, 69(1), pp. 38–51. doi: 10.2179/0008-7475(2004)069<0038:VIFOCR>2.0.CO;2.

Morris, M. H. *et al.* (2008) ‘Contrasting ectomycorrhizal fungal communities on the roots of co-occurring oaks (Quercus spp.) in a California woodland.’, *The New Phytologist*. BLACKWELL PUBLISHING, 9600 GARSINGTON RD, OXFORD OX4 2DQ, OXON, ENGLAND, 178(1), pp. 167–76. doi: 10.1111/j.1469-8137.2007.02348.x.

El Mousadik, A. and Petit, R. J. (1996) ‘High level of genetic differentiation for allelic richness among populations of the argan tree [Argania spinosa (L.) Skeels] endemic to Morocco.’, *Theoretical and Applied Genetics*. Springer, 92(7), pp. 832–839. doi: 10.1007/BF00221895.

Namroud, M.-C. *et al.* (2005) ‘Clonal and spatial genetic structures of aspen (Populus tremuloides Michx.).’, *Molecular Ecology*. BLACKWELL PUBLISHING, 14(10), pp. 2969–80. doi: 10.1111/j.1365-294X.2005.02653.x.

Natural England (2016) *Priority Habitat Inventory (England) [SHAPE Geospatial Data]*. Available at: https://naturalengland-defra.opendata.arcgis.com.

Navarro, J. *et al.* (2013) ‘Ecological segregation in space, time and trophic niche of sympatric planktivorous petrels.’, *PloS one*, 8(4), p. e62897.

Negri, G., Santi, D. and Tabach, R. (2013) ‘Flavonol glycosides found in hydroethanolic extracts from Tilia cordata, a species utilized as anxiolytics’, *Brazilian Journal of Medicinal Plants*. Sociedade Brasileira de Plantas Medicinais, 15(2), pp. 217–224. doi: 10.1590/S1516-05722013000200008.

Nellemann, C. and Thomsen, M. G. (2001) ‘Long-term changes in forest growth: Potential effects of nitrogen deposition and acidification’, *Water, Air, and Soil Pollution*. Kluwer Academic Publishers, 128(3–4), pp. 197–205. doi: 10.1023/A:1010318800180.

Nicotra, A. B. *et al.* (2010) ‘Plant phenotypic plasticity in a changing climate’, *Trends in Plant Science*, pp. 684–692. doi: 10.1016/j.tplants.2010.09.008.

Nobel, P. S. (1997) 'Root Distribution and Seasonal Production in the Northwestern Sonoran Desert for a C 3 Subshrub, a C 4 Bunchgrass, and a CAM Leaf Succulent', *American Journal of Botany*. BOTANICAL SOC AMER INC, 84(7), p. 949. doi: 10.2307/2446285.

Noreen, A. M. E. and Webb, E. L. (2013) 'High genetic diversity in a potentially vulnerable tropical tree species despite extreme habitat loss', *PLoS ONE*. Edited by T. Brown. Public Library of Science, 8(12), p. e82632. doi: 10.1371/journal.pone.0082632.

Normand, S. *et al.* (2011) 'Postglacial migration supplements climate in determining plant species ranges in Europe.', *Proceedings. Biological sciences / The Royal Society*. ROYAL SOC, 278(1725), pp. 3644–53. doi: 10.1098/rspb.2010.2769.

Nuortila, C., Tuomi, J. and Laine, K. (2002) 'Inter-parent distance affects reproductive success in two clonal dwarf shrubs, Vaccinium myrtillus and Vaccinium vitis-idaea (Ericaceae)', *Canadian Journal of Botany*, 884(8), pp. 875–884. doi: 10.1139/B02-079.

Oddou-Muratorio, S. *et al.* (2001) 'Pollen-versus seed-mediated gene flow in a scattered forest tree species.', *Evolution; international journal of organic evolution*, 55(6), pp. 1123–35.

Ogaya, R. *et al.* (2003) 'Effect of drought on diameter increment of Quercus ilex, Phillyrea latifolia, and Arbutus unedo in a holm oak forest of NE Spain', *Forest Ecology and Management*. Elsevier, 180(1–3), pp. 175–184. doi: 10.1016/S0378-1127(02)00598-4.

Oksanen, J. *et al.* (2017) 'vegan: Community Ecology Package'.

Oleksyk, T. K., Smith, M. W. and O'Brien, S. J. (2010) 'Genome-wide scans for footprints of natural selection.', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1537), pp. 185–205. doi: 10.1098/rstb.2009.0219.

Ordnance Survey (GB) (1848) 'County Series 1:10560 (Lancashire, 1st edition)'.

Ordnance Survey (GB) (1862) 'County Series 1:10560 (Westmorland, 1st edition)'.

Ordnance Survey (GB) (1888) 'County Series 1:10560 (Northamptonshire, 1st edition)'.

Ordnance Survey (GB) (2016a) *OS Open Map - Local [SHAPE Geospatial Data]*, *EDINA Digimap*. EDINA Digimap. Available at: http://digimap.edina.ac.uk.

Ordnance Survey (GB) (2016b) *OS Terrain 5 DTM [TIFF Geospatial Data]*, *EDINA Digimap*. Available at: http://digimap.edina.ac.uk.

Ordnance Survey (GB) (2017) *OS VectorMap® Local [SHAPE Geospatial Data]*, *EDINA Digimap*. Available at: http://digimap.edina.ac.uk.

Osborne, J. L. *et al.* (2008) 'Quantifying and comparing bumblebee nest densities in gardens and countryside habitats', *Journal of Applied Ecology*. Blackwell Publishing Ltd, 45(3), pp. 784–792. doi: 10.1111/j.1365-2664.2007.01359.x.

Overy, S. A. *et al.* (2005) 'Application of metabolite profiling to the identification of traits in a population of tomato introgression lines', in *Journal of Experimental Botany*, pp. 287–296. doi: 10.1093/jxb/eri070.

Palmé, A. E. *et al.* (2009) 'Selection on nuclear genes in a Pinus phylogeny', *Molecular Biology and Evolution*, 26(4), pp. 893–905. doi: 10.1093/molbev/msp010.

Paradis, E. (2010) 'pegas: an {R} package for population genetics with an integrated--modular approach', *Bioinformatics*, 26, pp. 419–420.

Pardi, M. I. and Smith, F. A. (2012) 'Paleoecology in an era of climate change: How the past can provide insights into the future', in *Paleontology in Ecology and Conservation*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 93–116. doi: 10.1007/978-3-642-25038-5_6.

Parks, J. C. and Werth, C. R. (1993) 'A Study of Spatial Features of Clones in a Population of Bracken Fern, Pteridium aquilinum (Dennstaedtiaceae)', *American Journal of Botany*. BOTANICAL SOC AMER INC, 80(5), p. 537. doi: 10.2307/2445369.

Parmesan, C. and Yohe, G. (2003) 'A globally coherent fingerprint of climate change impacts across natural systems.', *Nature*. NATURE PUBLISHING GROUP, 421(6918), pp. 37–42. doi: 10.1038/nature01286.

Pausas, J. G. and Lavorel, S. (2003) 'A hierarchical deductive approach for functional types in disturbed ecosystems', *Journal of Vegetation Science*. Blackwell Publishing Ltd, 14(3), pp. 409–416. doi: 10.1658/1100-9233(2003)014[0409:AHDAFF]2.0.CO;2.

Pearson, R. G. and Dawson, T. P. (2005) 'Long-distance plant dispersal and habitat fragmentation: identifying conservation targets for spatial landscape planning under climate change', *Biological Conservation*, 123(3), pp. 389–401. doi: 10.1016/j.biocon.2004.12.006.

Peñuelas, J. and Boada, M. (2003) 'A global change-induced biome shift in the Montseny mountains (NE Spain)', *Global Change Biology*. Blackwell Science, Ltd, 9(2), pp. 131–140. doi: 10.1046/j.1365-2486.2003.00566.x.

Peñuelas, J. and Sardans, J. (2009) ‘Ecological metabolomics’, *Chemistry and Ecology*, 25(4), pp. 305–309. doi: 10.1080/02757540903062517.

Petit, R. J. and Hampe, A. (2006) ‘Some Evolutionary Consequences of Being a Tree’, *Annual Review of Ecology, Evolution, and Systematics*. Annual Reviews, 37(1), pp. 187–214. doi: 10.1146/annurev.ecolsys.37.091305.110215.

Phuekvilai, P. (2014) *Relicts , Refugia and Reticulation : A Study of Population History , Hybrids and Phylogeny in the Long-Lived Flowering Tree Genus Tilia*. Newcastle University.

Phuekvilai, P. and Wolff, K. (2013) ‘Characterization of Microsatellite Loci in Tilia platyphyllos (Malvaceae) and Cross-Amplification in Related Species’, *Applications in Plant Sciences*, 1(4), p. 1200386. doi: 10.3732/apps.1200386.

Pielou, E. C. (1975) *Ecological Diversity*. New York: Wiley-Interscience.

Pigott, C. D. (1969) ‘The status of Tilia cordata and T. platyphyllos on the Derbyshire limestone’, *The Journal of Ecology*. JSTOR, 57(2), pp. 491–504.

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

Pigott, C. D. (1981b) ‘The status, ecology and conservation of Tilia platyphyllos in Britain’, in Synge, H. (ed.) *The Biological Aspects of Rare Plant Conservation*. Chichester: John Wiley & Sons Ltd., pp. 305–317.

Pigott, C. D. (1989) ‘The Growth of Lime Tilia Cordata in an Experimental Plantation and its Influence on Soil Development and Vegetation’, *Quarterly Journal of Forestry*, 83(1).

Pigott, C. D. (1991) ‘Tilia cordata Miller’, *Journal of Ecology*. JSTOR, 79(4), pp. 1147–1207.

Pigott, C. D. and Huntley, J. P. (1978) ‘Factors controlling the distribution of Tilia cordata at the northern limits of its geographical range. I. Distribution in North-West England’, *New Phytologist*. Wiley Online Library, 81(2), pp. 429–441. doi: 10.1111/j.1469-8137.1978.tb02648.x.

Pigott, C. D. and Huntley, J. P. (1980) ‘Factors controlling the distribution of Tilia cordata at the northern limits of its geographical range. II. History in north-west England’, *New*

Phytologist. JSTOR, 84(1), pp. 145–164.

Pigott, C. D. and Huntley, J. P. (1981) 'Factors controlling the distribution of Tilia cordata at the northern limits of its geographical range. III. Nature and causes of seed sterility', *New Phytologist*. JSTOR, 8(4), pp. 817–839.

Pigott, C. D. and Pigott, S. (1993) 'Water as a determinant of the distribution of trees at the boundary of the Mediterranean zone', *Journal of Ecology*. JSTOR, pp. 557–566.

Pigott, D. (2012) *Lime-trees and Basswoods: A Biological Monograph of the Genus Tilia*. 1st edn. New York: Cambridge University Press.

Poorter, H. *et al.* (2009) 'Causes and consequences of variation in leaf mass per area (LMA): A meta-analysis', *New Phytologist*. Blackwell Publishing Ltd, pp. 565–588. doi: 10.1111/j.1469-8137.2009.02830.x.

Pressey, R. L. *et al.* (2007) 'Conservation planning in a changing world', *Trends in Ecology and Evolution*. Elsevier Current Trends, 22(11), pp. 583–592. doi: 10.1016/j.tree.2007.10.001.

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

Provan, J. *et al.* (1999) 'A low mutation rate for chloroplast microsatellites', *Genetics*, 153(2), pp. 943–947. doi: 10.1073/PNAS.84.24.9054.

Provan, J., Powell, W. and Hollingsworth, P. M. (2001) 'Chloroplast microsatellites: New tools for studies in plant ecology and evolution', *Trends in Ecology and Evolution*, pp. 142–147. doi: 10.1016/S0169-5347(00)02097-8.

Prunier, J. G. *et al.* (2013) 'Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: Towards a better assessment of functional connectivity using an individual-based sampling scheme', *Molecular Ecology*, 22(22), pp. 5516–5530. doi: 10.1111/mec.12499.

QGIS Development Team (2016) 'QGIS Geographic Information System'.

Rabasa, S. *et al.* (2013) 'Disparity in elevational shifts of European trees in response to recent climate warming.', *Global Change Biology*, 19(8), pp. 2490–9.

R Core Team (2016) 'R: A Language and Environment for Statistical Computing'. Vienna,

Austria.

Ramírez-Valiente, J. A. *et al.* (2010) ‘Phenotypic plasticity and local adaptation in leaf ecophysiological traits of 13 contrasting cork oak populations under different water availabilities’, *Tree Physiology*. Oxford University Press, 30(5), pp. 618–627. doi: 10.1093/treephys/tpq013.

Rasmussen, K. K., Strandby, U. and Kollmann, J. (2010) ‘High genetic diversity within but limited differentiation among populations of the vulnerable Guatemalan fir’, *Journal of Tropical Forest Science*. Forest Research Institute Malaysia, 22(3), pp. 247–259.

Rautiainen, P., Koivula, K. and Hyvärinen, M. (2004) ‘The effect of within-genet and between-genet competition on sexual reproduction and vegetative spread in Potentilla anserina ssp. egedii’, *Journal of Ecology*. Blackwell Science Ltd, 92(3), pp. 505–511. doi: 10.1111/j.0022-0477.2004.00878.x.

Reed, D. H. and Frankham, R. (2001) ‘How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis’, *Evolution*. Blackwell Publishing Ltd, 55(6), pp. 1095–1103. doi: 10.1111/j.0014-3820.2001.tb00629.x.

Reed, D. H. and Frankham, R. (2003) ‘Correlation between fitness and genetic diversity’, *Conservation Biology*. Blackwell Science, Inc., 17(1), pp. 230–237. doi: 10.1046/j.1523-1739.2003.01236.x.

Ricklefs, R. E. and Latham, R. E. (1992) ‘Intercontinental Correlation of Geographical Ranges Suggests Stasis in Ecological Traits of Relict Genera of Temperate Perennial Herbs’, *The American Naturalist*. UNIV CHICAGO PRESS, 139(6), p. 1305. doi: 10.1086/285388.

Rico, C. *et al.* (2017) ‘Null alleles are ubiquitous at microsatellite loci in the Wedge Clam (Donax trunculus)’, *PeerJ*. PeerJ Inc., 5, p. e3188. doi: 10.7717/peerj.3188.

Rivas-Ubach, A. *et al.* (2014) ‘Drought enhances folivory by shifting foliar metabolomes in Quercus ilex trees’, *New Phytologist*, 202(3), pp. 874–885. doi: 10.1111/nph.12687.

Robson, T. M. *et al.* (2012) ‘Variation in functional leaf traits among beech provenances during a Spanish summer reflects the differences in their origin’, *Tree Genetics and Genomes*. Springer-Verlag, 8(5), pp. 1111–1121. doi: 10.1007/s11295-012-0496-5.

Root, T. L. *et al.* (2003) ‘Fingerprints of global warming on wild animals and plants’, *Nature*. Nature Publishing Group, 421(6918), pp. 57–60. doi: 10.1038/nature01333.

Rousseeuw, P. J. *et al.* (2015) ‘robustbase: Basic Robust Statistics. R package version 0.92-3’.

Rousseeuw, P. J. and Croux, C. (1993) ‘Alternatives to the Median Absolute Deviation’, *Journal of the American Statistical Association*, 88(424), pp. 1273–1283.

Rozenfeld, A. F. *et al.* (2007) ‘Spectrum of genetic diversity and networks of clonal organisms.’, *Journal of the Royal Society, Interface / the Royal Society*, 4(17), pp. 1093–1102. doi: 10.1098/rsif.2007.0230.

Sage, R. F. and Kubien, D. S. (2007) ‘The temperature response of C3 and C4 photosynthesis’, *Plant, Cell and Environment*, 30(9), pp. 1086–1106. doi: 10.1111/j.1365-3040.2007.01682.x.

Sala, O. E. *et al.* (2000) ‘Global Biodiversity Scenarios for the Year 2100’, *Science*, 287(5459), pp. 1770–1774. doi: 10.1126/science.287.5459.1770.

Sampaio, B. L., Edrada-Ebel, R. A. and Da Costa, F. B. (2016) ‘Effect of the environment on the secondary metabolic profile of Tithonia diversifolia: a model for environmental metabolomics of plants’, *Scientific Reports*. Nature Publishing Group, 6(1), p. 29265. doi: 10.1038/srep29265.

Sánchez-Salguero, R. *et al.* (2017) ‘Assessing forest vulnerability to climate warming using a process-based model of tree growth: bad prospects for rear-edges’, *Global Change Biology*, 23(7), pp. 2705–2719. doi: 10.1111/gcb.13541.

Sardans, J. *et al.* (2014) ‘Metabolic responses of Quercus ilex seedlings to wounding analysed with nuclear magnetic resonance profiling’, *Plant Biology*. Edited by G. Noctor, 16(2), pp. 395–403. doi: 10.1111/plb.12032.

Schaal, B. A. and Leverich, W. J. (1996) ‘Molecular Variation in Isolated Plant Populations.’, *Plant Species Biology*. Blackwell Publishing Ltd, 11(1), pp. 33–40. doi: 10.1111/j.1442-1984.1996.tb00106.x.

Scheepens, J. F., Frei, E. S. and Stöcklin, J. (2010) ‘Genotypic and environmental variation in specific leaf area in a widespread Alpine plant after transplantation to different altitudes’, *Oecologia*. Springer Berlin Heidelberg, 164(1), pp. 141–150. doi: 10.1007/s00442-010-1650-0.

Schnittler, M. and Eusemann, P. (2010) ‘Consequences of genotyping errors for estimation

of clonality: a case study on Populus euphratica Oliv. (Salicaceae)', *Evolutionary Ecology*. SPRINGER, VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS, 24(6), pp. 1417–1432. doi: 10.1007/s10682-010-9389-y.

Schueler, S., Tusch, A. and Scholz, F. (2006) 'Comparative analysis of the within-population genetic structure in wild cherry (Prunus avium L.) at the self-incompatibility locus and nuclear microsatellites.', *Molecular ecology*, 15(11), pp. 3231–43. doi: 10.1111/j.1365-294X.2006.03029.x.

Shaffer, M. L. (1981) 'Minimum Population Sizes for Species Conservation', *BioScience*. Oxford University Press, 31(2), pp. 131–134. doi: 10.2307/1308256.

Sheykholeslami, A., Namiranian, M. and Sagheb-Talebi, K. (2011) 'A study of large-leaved lime (Tilia platyphyllos Scop.) in forests of Western Mazandaran', *Journal of Biodiversity and Ecological Sciences*, 1(1), pp. 65–75.

Shirk, A. J., Landguth, E. L. and Cushman, S. A. (2017) 'A comparison of individual-based genetic distance metrics for landscape genetics', *Molecular Ecology Resources*. doi: 10.1111/1755-0998.12684.

Shvidenko, A. *et al.* (2005) 'Forest and Woodland Systems', in *Ecosystems and Human Well-being: Synthesis (Millenium Ecosystem Assessment).* Washington DC, USA: Island Press, pp. 585–621.

Shvidenko, A. and Nilsson, S. (2002) 'Dynamics of Russian forests and the carbon budget in 1961-1998: An assessment based on long-term forest inventory data', *Climatic Change*. Kluwer Academic Publishers, 55(1–2), pp. 5–37. doi: 10.1023/A:1020243304744.

Silvertown, J. *et al.* (1999) 'Hydrologically defined niches reveal a basis for species richness in plant communities', *Nature*, 400(July), pp. 61–64.

Silvertown, J. (2004) 'Plant coexistence and the niche', *Trends in Ecology & Evolution*, 19(11), pp. 607–611.

Silvertown, J. (2008) 'The Evolutionary Maintenance of Sexual Reproduction: Evidence from the Ecological Distribution of Asexual Reproduction in Clonal Plants', *International Journal of Plant Sciences*. The University of Chicago Press, 169(1), pp. 157–168. doi: 10.1086/523357.

Sork, V. L. *et al.* (2002) 'Pollen movement in declining populations of California Valley

oak, Quercus lobata: Where have all the fathers gone?', *Molecular Ecology*, 11(9), pp. 1657–1668. doi: 10.1046/j.1365-294X.2002.01574.x.

Sork, V. L. and Smouse, P. E. (2006) 'Genetic analysis of landscape connectivity in tree populations', *Landscape Ecology*, 21(6), pp. 821–836. doi: 10.1007/s10980-005-5415-9.

Spear, S. F. *et al.* (2010) 'Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis.', *Molecular Ecology*. WILEY-BLACKWELL, 19(17), pp. 3576–91. doi: 10.1111/j.1365-294X.2010.04657.x.

ter Steege, M. W. *et al.* (2005) 'Genetic and physiological architecture of early vigor in Aegilops tauschii, the D-genome donor of hexaploid wheat. A quantitative trait loci analysis.', *Plant physiology*, 139(2), pp. 1078–94. doi: 10.1104/pp.105.063263.

Stepansky, A. and Leustek, T. (2006) 'Histidine biosynthesis in plants', *Amino Acids*, 30(2), pp. 127–142. doi: 10.1007/s00726-005-0247-0.

Sturrock, R. N. *et al.* (2011) 'Climate change and forest diseases', *Plant Pathology*. Blackwell Publishing Ltd, pp. 133–149. doi: 10.1111/j.1365-3059.2010.02406.x.

Suvanto, L. I. and Latva-Karjanmaa, T. B. (2005) 'Clone identification and clonal structure of the European aspen (Populus tremula).', *Molecular Ecology*. BLACKWELL PUBLISHING, 14(9), pp. 2851–60. doi: 10.1111/j.1365-294X.2005.02634.x.

Tel-zur, N. *et al.* (1999) 'Modified CTAB Procedure for DNA Isolation from Epiphytic Cacti of the Genera Hylocereus and Selenicereus (Cactaceae)', *Plant Molecular Biology Reporter*. Kluwer Academic Publishers, 17(3), pp. 249–254. doi: 10.1023/A:1007656315275.

Thakur, S. *et al.* (2016) 'Genetic diversity and population structure of Melia azedarach in North-Western Plains of India', *Trees - Structure and Function*, 30(5), pp. 1483–1494. doi: 10.1007/s00468-016-1381-x.

Thomas, C. D. *et al.* (2004) 'Extinction risk from climate change', *Nature*, 427(6970), pp. 145–8. doi: 10.1038/nature02121.

Thórsson, A. T. *et al.* (2007) 'Morphological variation among Betula nana (diploid), B. pubescens (tetraploid) and their triploid hybrids in Iceland.', *Annals of Botany*. OXFORD UNIV PRESS, GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND, 99(6), pp. 1183–93. doi: 10.1093/aob/mcm060.

Thuiller, W. *et al.* (2005) ‘Climate change threats to plant diversity in Europe.’, *Proceedings of the National Academy of Sciences of the United States of America*. NATL ACAD SCIENCES, 102(23), pp. 8245–50. doi: 10.1073/pnas.0409902102.

Thuiller, W. *et al.* (2008) ‘Predicting global change impacts on plant species’ distributions: Future challenges’, *Perspectives in Plant Ecology, Evolution and Systematics*. Urban & Fischer, 9(3–4), pp. 137–152. doi: 10.1016/j.ppees.2007.09.004.

Todorov, V. and Filzmoser, P. (2009) ‘An Object-Oriented Framework for Robust Multivariate Analysis’, *Journal of Statistical Software*, 32(3), pp. 1–47.

Tovar-Sánchez, E. and 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*. BOTANICAL SOC AMER INC, 91(9), pp. 1352–63. doi: 10.3732/ajb.91.9.1352.

Trakhtenbrot, A., Katul, G. G. and Nathan, R. (2014) ‘Mechanistic modeling of seed dispersal by wind over hilly terrain’, *Ecological Modelling*, 274, pp. 29–40. doi: 10.1016/j.ecolmodel.2013.11.029.

Trénel, P. *et al.* (2008) ‘Landscape genetics, historical isolation and cross-Andean gene flow in the wax palm, Ceroxylon echinulatum (Arecaceae)’, *Molecular Ecology*, 17(15), pp. 3528–3540. doi: 10.1111/j.1365-294X.2008.03853.x.

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

Turok, J. *et al.* (1996) ‘Report of the First Meeting’, *International Plant Genetic Resources Institute: Rome, Italy*, p. 172.

Valbuena-Carabaña, M. and Gil, L. (2017) ‘Centenary coppicing maintains high levels of genetic diversity in a root resprouting oak (Quercus pyrenaica Willd.)’, *Tree Genetics and Genomes*. Springer Berlin Heidelberg, 13(1), p. 28. doi: 10.1007/s11295-017-1105-4.

Vallejo-Marín, M., Dorken, M. E. and Barrett, S. C. H. (2010) ‘The Ecological and Evolutionary Consequences of Clonality for Plant Mating’, *Annual Review of Ecology, Evolution, and Systematics*. ANNUAL REVIEWS, 41(1), pp. 193–213. doi: 10.1146/annurev.ecolsys.110308.120258.

Vekemans, X. and Hardy, O. J. (2004) ‘New insights from fine-scale spatial genetic structure

analyses in plant populations', *Molecular Ecology*, 13(4), pp. 921–935. doi: 10.1046/j.1365-294X.2004.02076.x.

Vranckx, G. *et al.* (2012) 'Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation.', *Conservation biology : the journal of the Society for Conservation Biology*, 26(2), pp. 228–37. doi: 10.1111/j.1523-1739.2011.01778.x.

Vranckx, G. *et al.* (2014) 'Tree density and population size affect pollen flow and mating patterns in small fragmented forest stands of pedunculate oak (Quercus robur L.)', *Forest Ecology and Management*, 328, pp. 254–261. doi: 10.1016/j.foreco.2014.05.044.

Weed, A. S. and Schwarzländer, M. (2014) 'Density dependence, precipitation and biological control agent herbivory influence landscape-scale dynamics of the invasive Eurasian plant Linaria dalmatica', *Journal of Applied Ecology*. Edited by Q. Paynter, 51(3), pp. 825–834. doi: 10.1111/1365-2664.12226.

Wiens, J. J. and Graham, C. H. (2005) 'Niche conservatism: Integrating Evolution, Ecology, and Conservation Biology', *Annual Review of Ecology, Evolution, and Systematics*. ANNUAL REVIEWS, 36(1), pp. 519–539. doi: 10.1146/annurev.ecolsys.36.102803.095431.

Wigston, D. L. (1975) 'The distribution of Quercus robur L., Q. petraea (Matt.) Leibl. and their hybrids in south-western England: 1. The assessment of the taxonomic status of populations from leaf characters', *Watsonia*, 10, pp. 345–349.

Wilson, J. B. *et al.* (2012) 'Plant species richness: the world records', *Journal of Vegetation Science*. Edited by M. Palmer. WILEY-BLACKWELL, 23(4), pp. 796–802. doi: 10.1111/j.1654-1103.2012.01400.x.

Winkel-Shirley, B. (2001) 'It takes a garden. How work on diverse plant species has contributed to an understanding of flavonoid metabolism.', *Plant physiology*, 127(4), pp. 1399–1404. doi: 10.1104/pp.010675.

Winkel-Shirley, B. (2002) 'Biosynthesis of flavonoids and effects of stress', *Current Opinion in Plant Biology*. Elsevier Current Trends, 5(3), pp. 218–223. doi: 10.1016/S1369-5266(02)00256-X.

de Witte, L. C. and Stöcklin, J. (2010) 'Longevity of clonal plants: why it matters and how to measure it.', *Annals of botany*. OXFORD UNIV PRESS, 106(6), pp. 859–70. doi:

10.1093/aob/mcq191.

Wong, L.-M., Silvaraj, S. and Phoon, L.-Q. (2014) 'An Optimised High-Salt CTAB Protocol for Both DNA and RNA Isolation from Succulent Stems of Hylocereus sp.', *Journal of Medical and Bioengineering*, 3(4), pp. 236–240.

Worley, B. and Powers, R. (2013) 'Multivariate Analysis in Metabolomics', *Current Metabolomics*, 1(1), pp. 92–107. doi: 10.2174/2213235X11301010092.

Worth, J. R. P. *et al.* (2016) 'Gondwanan conifer clones imperilled by bushfire', *Scientific Reports*, 6(1), p. 33930. doi: 10.1038/srep33930.

Wright, S. J. (2002) 'Plant diversity in tropical forests : a review of mechanisms of species coexistence', *Oecologica*, 130(1), pp. 1–14. doi: 10.1007/s004420100809.

Yamada, T., Ngakan, O. P. and Suzuki, E. (2006) 'Differences in growth and light requirement of two sympatric congeneric tree species in an Indonesian floodplain forest', *Journal of Tropical Ecology*. CAMBRIDGE UNIV PRESS, 22(3), p. 349. doi: 10.1017/S026646740500310X.

Yamasaki, N., Yamada, T. and Okuda, T. (2013) 'Coexistence of two congeneric tree species of Lauraceae in a secondary warm-temperate forest on Miyajima Island, south-western Japan', *Plant Species Biology*. WILEY-BLACKWELL, 28(1), pp. 41–50. doi: 10.1111/j.1442-1984.2011.00356.x.

Yang, J. *et al.* (2015) 'A data preprocessing strategy for metabolomics to reduce the mask effect in data analysis.', *Frontiers in molecular biosciences*. Frontiers, 2, p. 4. doi: 10.3389/fmolb.2015.00004.

Ye, D. *et al.* (2014) 'Clonality-climate relationships along latitudinal gradient across china: Adaptation of clonality to environments', *PLoS ONE*. Edited by F. de Bello. Public Library of Science, 9(4), p. e94009. doi: 10.1371/journal.pone.0094009.

Ye, D. *et al.* (2016) 'Strong but diverging clonality - climate relationships of different plant clades explain weak overall pattern across China', *Scientific Reports*, 6(June), p. 26850. doi: 10.1038/srep26850.

Yin, X., Kropff, M. J. and Stam, P. (1999) 'The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley', *Heredity*, 82(4), pp. 415–421. doi: 10.1038/sj.hdy.6885030.

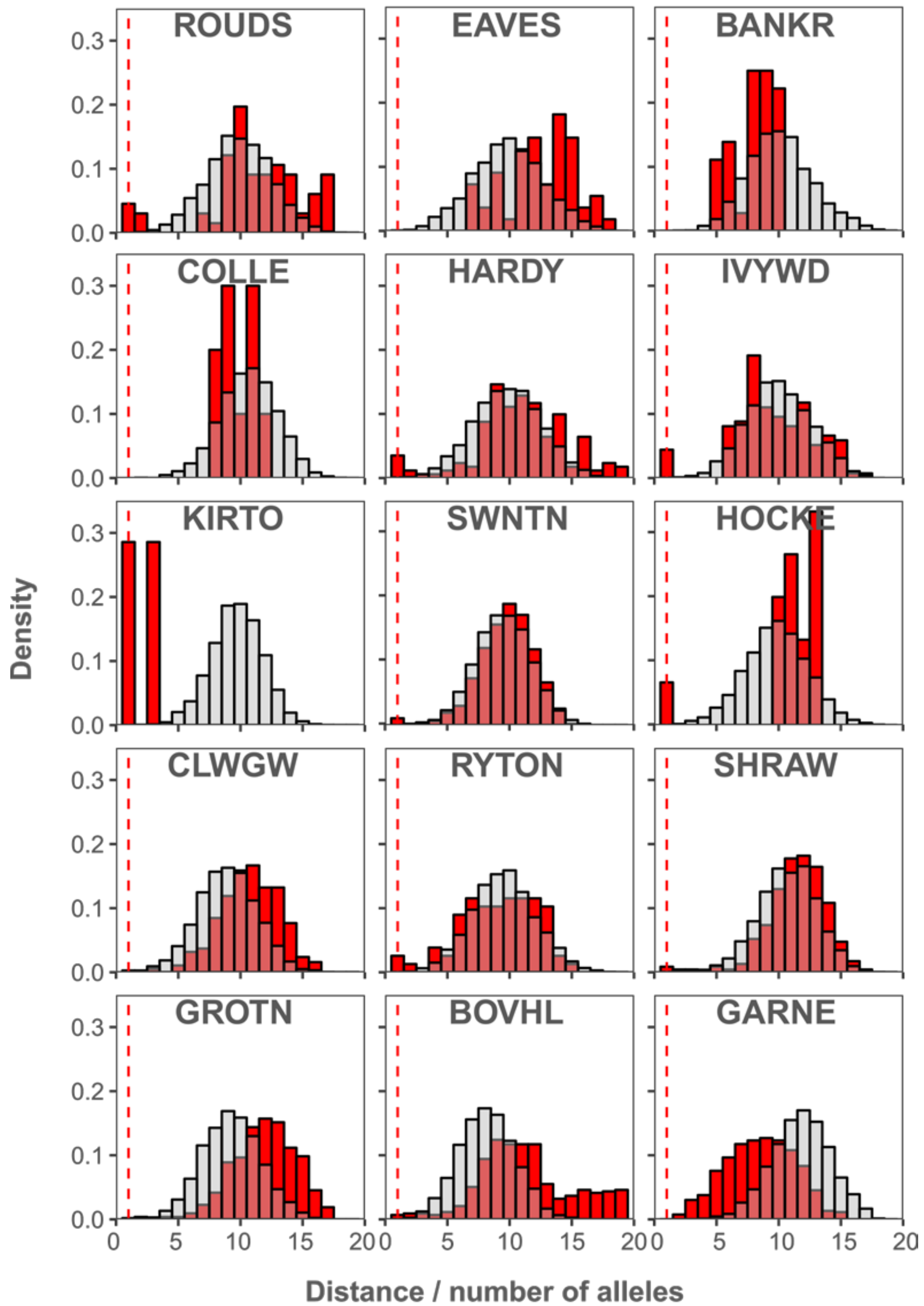
Young, A., Boyle, T. and Brown, T. (1996) ‘The population genetic consequences of habitat fragmentation for plants’, *Trends in Ecology & Evolution*, 11(10), pp. 413–418.

Zhu, K., Woodall, C. W. and Clark, J. S. (2012) ‘Failure to migrate: Lack of tree range expansion in response to climate change’, *Global Change Biology*, 18(3), pp. 1042–1052. doi: 10.1111/j.1365-2486.2011.02571.x.

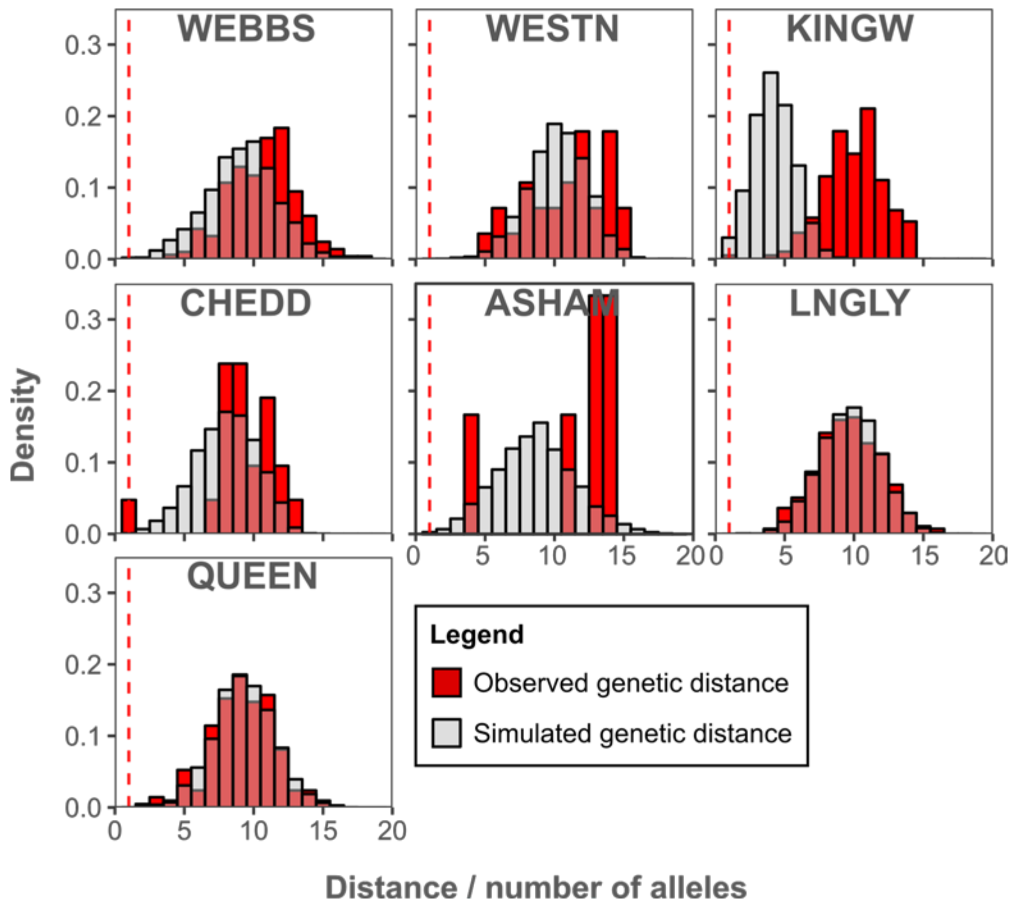
Zong, J.-W. *et al.* (2015) ‘Assessment of Genetic Diversity and Population Genetic Structure of Corylus mandshurica in China Using SSR Markers’, *PLoS One*. Edited by X. Pang, 10(9), p. e0137528. doi: 10.1371/journal.pone.0137528.

Appendices

Appendix I: distributions of inter-individual genetic distance

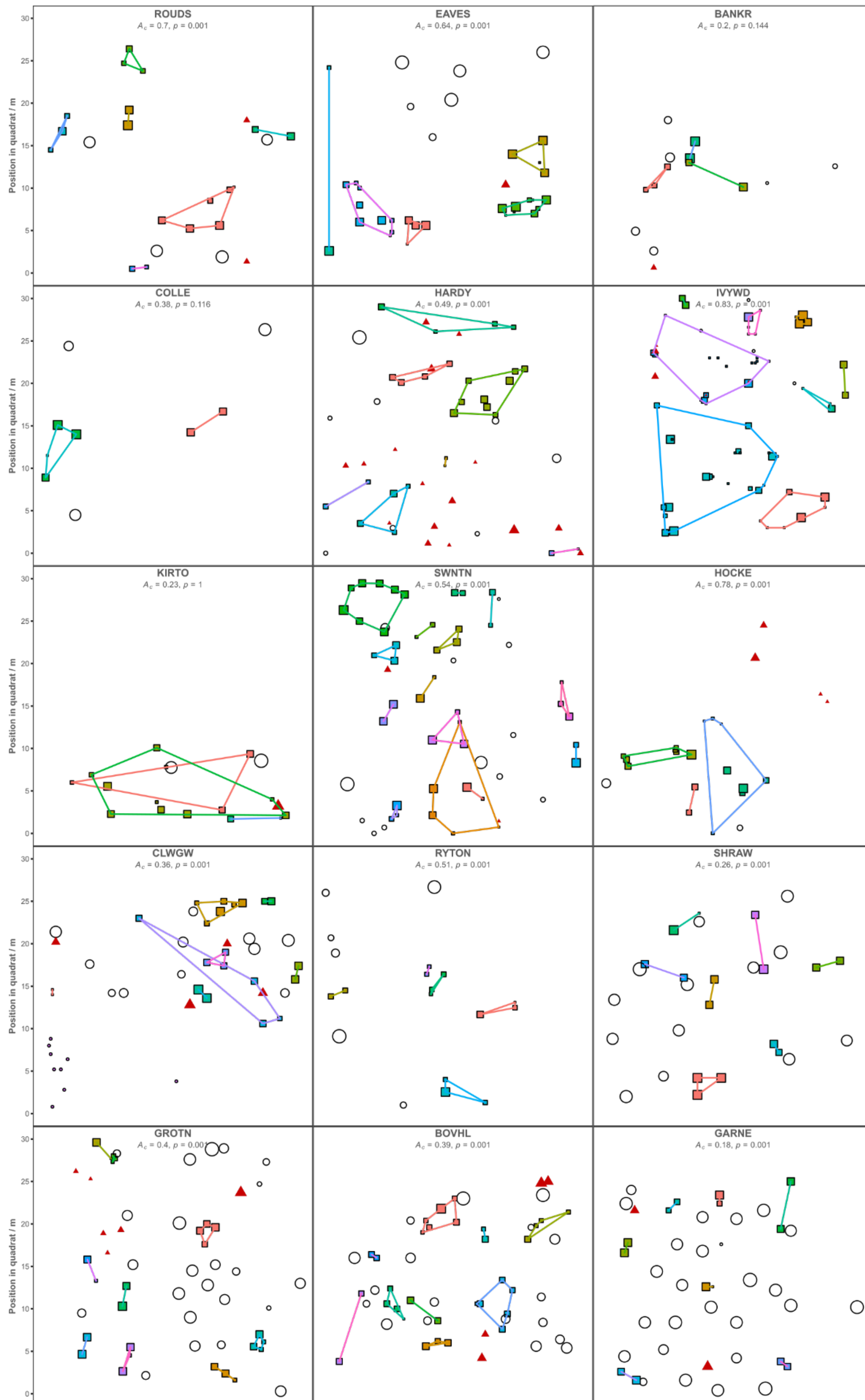


(continued on next page)



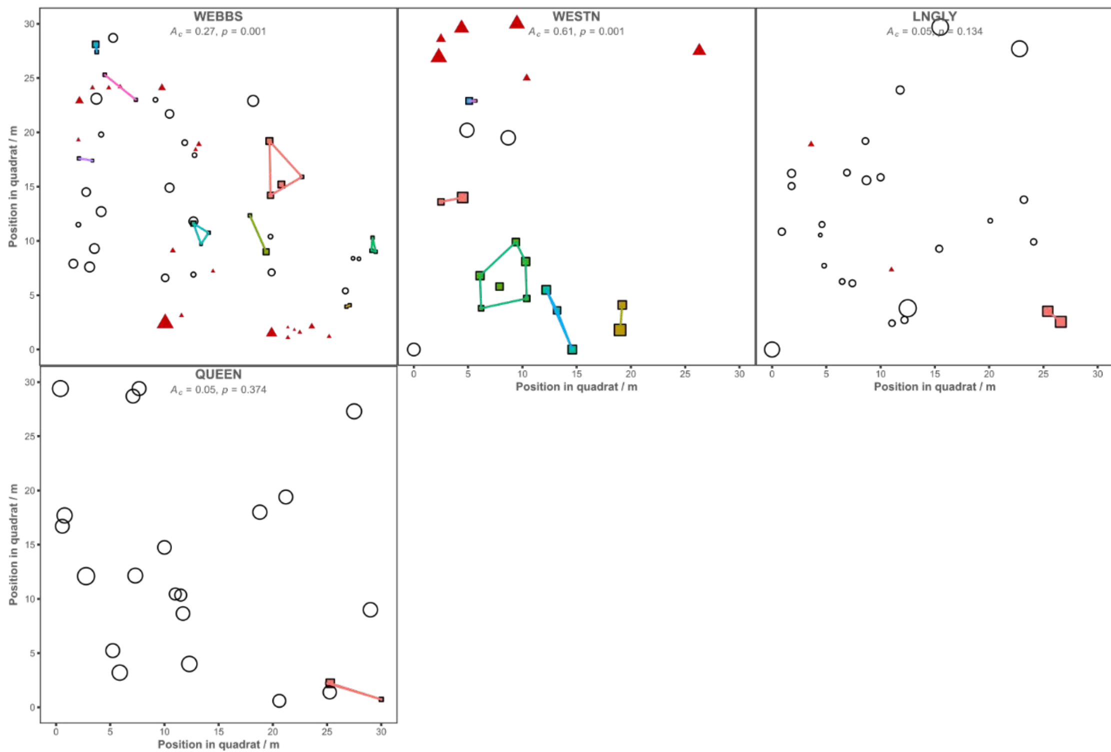
Appendix I: Frequency distributions of pairwise genetic distance between individuals (as number of distinct alleles). Red bars are observed distances and translucent grey bars distances from 99 simulated sexual reproductive events (with selfing). The dashed line represents the threshold at which an individual was considered for inclusion in clonal lineage.

Appendix II: spatial arrangement of clonality



(continued on next page)

Spatial arrangement of clonality



(etc.) **Clonal multi-locus lineages**

Sample unavailable (ungenotyped) **Unique genotypes**

Appendix II: the spatial arrangement of clonality at sampling sites. Unique genets are represented by hollow circles while clonal MLL groups are represented as filled squares of the same colour. Clonal MLL are contained within convex hulls described by dashed lines. Red triangles represent individuals whose presence was recorded but not genotyped. The extent of aggregation A_c and its significance after 1 000 permutations are provided. DBH of each individual is indicated by the size of the shapes.

Appendix III: putative identifications for the ten highest absolute ranking mass bins in principal components analysis

Detection mode	Component	Mass bin	Ranking	Loading	Metabolite
Negative	PC1	268.4	1	-0.32	
		282.4	2	-0.27	
		267.4	3	-0.19	
		264.6	4	0.17	
		292.6	5	0.16	
		324.6	6	0.15	
		576.4	7	-0.15	Phosphoribosyl-formimino-AICAR-phosphate
		192.4	8	-0.14	
		283.4	9	-0.13	
		269.4	10	-0.13	Estrone
	PC2	282.4	1	-0.46	
		192.6	2	-0.24	
		576.4	3	-0.21	Phosphoribosyl-formimino-AICAR-phosphate
		268.4	4	-0.20	
		300.4	5	0.14	
		284.4	6	0.13	
		242.4	7	0.13	
		254.4	8	0.11	
		577.4	9	-0.11	
		590.4	10	-0.11	
Positive	PC1	285	1	-0.42	Delta3-Isopentenyl diphosphate
					Delta2-Isopentenyl diphosphate
					Isopimpinellin
					Pimpinellin
					D-Mannitol 1-phosphate
					Decursinol
					Marmesin
					Acacetin
					Xanthosine
					Octopine
		381	2	0.29	Trehalose
					alpha-D-glucopyranose
					Nigerose
					Sucrose
					Laminaribiose
					Rutinose
					Maltose
					Melebiose
					Sophorose
					Pantetheine 4'-phosphate
					Cellobiose
					Epimelibiose
					Galactinol
					Gentiobiose
					Isomaltose
		266	3	0.26	alpha-D-Glutamyl phosphate

Detection mode	Component	Mass bin	Ranking	Loading	Metabolite
Positive	PC1	266	3	0.26	L-Glutamate 5-phosphate
					L-Arogenate
					Deoxycytidine (dC)
					Cytidine (C)
		447.2	4	-0.24	Ginkgolide A
					Nodakenin
					Adifoline
					Rutarin
					Bleekerine
					Vernoflexuoside
					6,8-Diprenylnarigenin
					Abyssinone V
					Cascarillin
					Dimethamine
		593.2	5	-0.22	
		615.2	6	-0.20	
		222	7	0.20	O-Phospho-L-homoserine
					Isowillardiine
					Willardiine
					L-Normetanephrine
					L-Histidinol phosphate
					N-acetyl-D-galactosamine
					N-acetyl-D-glucosamine
					N-acetyl-D-mannosamine
		286	8	-0.18	Pyridoxal phosphate
					Pyridoxine phosphate
					Linamarin
		594.2	9	-0.14	
		447	10	-0.13	CDP-ethanolamine
					Khellol glucoside
	PC2	266	1	-0.27	*See above (repeated)*
		631.2	2	-0.25	Diosmin
					Reserpine
		381	3	0.22	*See above (repeated)*
		222	4	-0.22	*See above (repeated)*
		285	5	-0.20	*See above (repeated)*
		417	6	0.19	
		365.2	7	0.16	Trehalose
					alpha-D-glucopyranose
					Nigerose
					Sucrose
					Laminaribiose
					Rutinose
					Maltose
					Melebiose
					Sophorose
					Cellobiose
					Epimelibiose
					Galactinol
					Gentiobiose

Detection mode	Component	Mass bin	Ranking	Loading	Metabolite
Positive	PC2	365.2	7	0.16	Isomaltose
					Ajmaline
					Gibberellin A44
					Xanthosine 5'-phosphate (XMP)
		631	8	-0.15	
		411	9	-0.15	2'-Deoxyuridine 5'-diphosphate (dUDP)
		184	10	0.12	Selenophosphate
					L-2-aminoadipic acid
					O-Acetyl-L-homoserine
					2-Ketoglutaramate
					L-Normetanephrine

Appendix III: Unabridged putative identifications for the ten highest ranking loadings of both components of both principal components analyses (using both positive and negative %TIC). The identity of the compounds given are derived from matching observed masses with the KEGG database (Kanehisa Laboratories, 2017). Blank cells indicate that no match was found.