UNIVERSITY OF
STIRLING

Landscape genetics of *Alnus glutinosa* across contrasting spatial scales in a natural river system

Gillian F Flint

Biological and Environmental Sciences

School of Natural Sciences

University of Stirling

A thesis submitted for the degree of Doctor of Philosophy

November 2015

Acknowledgements

Thank you to the University of Stirling for providing me with the opportunity to study for a PhD under the Horizon Studentship. In particular, there are many people to thank, past and present, from the Biological and Environmental Sciences Department.

Most especially, I would like to thank my principal supervisor Alistair Jump for enabling me to carry out this research project. His support and understanding was more than I could have asked for, and made all the difference when it mattered most. Thank you.

I would also like to thank Cecile Bacles for recognising my ability, giving me the opportunity, and welcoming me to the department; my second supervisor Nigel Willby for his interest in my work and his river knowledge input; Olivier Lepais for his excellent molecular laboratory training; James Weir for making sure the Controlled Environment Facility was controlled just as I needed it; Ronnie Balfour for always having the equipment ready and the van booked; Lynn MacGregor for seeing me right in processing the paperwork; Scott Jackson for his ability to solve all my IT problems; Jon McArthur for ArcGIS support; Martin Clarke for his logic, patience and talent in resolving my ArcGIS challenges; Patricia González Diáz and Jose Moreno Villena for their immense molecular laboratory help; Antoine Keruzoré for sharing his River Tay knowledge; Liam Cavin and Sarah Greenwood for shedding light on the field of dendrochronology; Mario Vallejo-Marín for his continued interest in my work; and Tim Paine for R inspiration.

For sharing the journey with me, from discussing scientific intricacies to dance-floor joy, I would like to thank Rebecca Barclay, Rachael Cooper-Bohannon, Caroline Griffin, Hannah Hamilton, Elizabeth Herridge, Laura Kubasiewicz, Paul Lintott, Rachel Lintott, Fiona Millar, Rosalind Murray, Zarah Pattison, Caitlin Riddick, Romuald Rouger, Jennifer Sjölund, and Fiona Thompson. You made all the difference.

A special mention goes to Angie MacArthur for her intelligence, kindness, support and friendship. I can't thank you enough.

I give thanks to Helen Taylor of Scottish Natural Heritage for helping obtain the necessary statutory permissions; Emma Rawling of the Scottish Wildlife Trust for providing all the help and information required; and numerous landowners for allowing me to walk, cycle, drive, and park on their land and for pointing me in the direction of alder leaves.

For donning their waterproofs, fending off the ticks, crossing the ricketiest of bridges, climbing trees, taking innumerable measurements, becoming a spotter of alder and parking places, being braver than me when faced with the boggy parts, and continuing to support me, thank you to my parents Armorel and Roy Flint, and my brother Sandy Flint. Thank you for helping me gather my thoughts alongside the leaves. All with the help of a marvellous waterproof pen – thanks Angela McAvoy McClumpha!

Finally, to Stewart Lothian, my partner, fieldwork assistant, excel whizz, equation dissembler, proof-reader, thesis reader, and biggest ally. Thank you for holding my hand throughout.

I remember that evening in May, collecting the D and the N and the A

RM Flint

Abstract

The genetic diversity and genetic structure of populations, and the processes shaping gene flow within and between populations, are influenced by the landscapes they occur within. Within terrestrial landscapes, rivers and their riparian habitat are among the most dynamic, diverse and complex of landscapes and their linear structure appears as an interlinking feature across large landscapes.

This thesis took a landscape genetics approach to examine the influence of river landscape features on *Alnus glutinosa* populations, a widespread keystone tree species of European riparian ecosystems. By accounting for the differing dispersal mechanisms of *A. glutinosa* (wind and water), landscape effects on seed- and pollen-mediated gene flow, genetic diversity, demographic and genetic structure were identified at different spatial scales of a large UK river catchment.

Widespread gene flow within and between *A. glutinosa* populations was identified with no apparent limitation of wind-mediated pollen dispersal. Hydrochorous dispersal of seed between populations was evident, and found to increase genetic connectivity between riparian populations; however an isolation by distance effect was identified between populations located further apart from each other. No pattern of genetic diversity was found, with high levels of genetic diversity identified at all spatial and temporal scales. At the river-catchment scale no genetic clustering was observed, either within or between the six rivers studied. Demographic structuring within *A. glutinosa* populations was evident, and correlated with distance from the main river channel. Interactions between

seed dispersal, hydrological disturbance, colonisation, and historical influences are discussed in relation to fine-scale spatial genetic structure between *A. glutinosa* sapling and adult generations.

Central to the landscape genetics approach taken in this thesis was the incorporation of key *A. glutinosa* life history attributes. By incorporating gene flow analyses, species ecology and landscape features, the research presented here furthers our understanding of riverine landscape influences on their riparian populations at different spatial scales and can be used to inform management principles.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Ecological heterogeneity exists everywhere within nature so that living beings, and the landscapes they occur in, are distributed neither uniformly nor randomly (Legendre 1993). Furthermore, this ecological heterogeneity occurs over a range of spatial and temporal scales. Thus, the field of ecology is fundamentally concerned with issues of pattern and scale (Levin 1992). Crucially, the identification and description of ecological pattern, and the temporal and spatial scales it occurs at, whilst important in and of itself, also enables efforts to discover the processes generating and maintaining the pattern observed. Elucidation of the processes underlying ecological patterns is essential to our understanding of how ecosystems function, interact, and respond to change, and also facilitates the development of management principles (Levin 1992). Over the last 20 years technological developments within ecological genetics, and the wider field of molecular ecology, have seen significant progress in connecting processes with pattern and scale (Chave 2013).

1.2 Ecological genetics

Ecological genetics integrates field ecology with the application of molecular genetic tools (Ford 1975), and seeks to investigate the origin and maintenance of genetic variation within and between populations whilst also accounting for population size and structure (Lowe *et al.* 2008). Molecular markers, fragments of heritable DNA at specific genome loci, can be used to identify differences in the DNA sequences of individuals. Processes that affect individuals ultimately accumulate into effects on populations, thus the application of genetic markers can be used to examine patterns of genetic variation at the individual level, between individuals in a subpopulation, and in subpopulations within

populations (Sunnucks 2000). Technological developments have seen rapid improvement in the power of molecular markers alongside a reduction in processing costs, enabling the use of large numbers of samples and the study of multiple loci, especially for microsatellite markers (Selkoe and Toonen 2006). Codominant in the nuclear genome and selectively neutral, microsatellites reveal length variation in alleles and can be used to identify and measure genetic variation and to subsequently test hypotheses related to processes such as migration, population size, and kinship (Selkoe and Toonen 2006).

To understand the influence of different processes on population genetics, and to ensure the effective management of genetic resources, it is important to describe and quantify genetic diversity, recognised as one of the three forms of biodiversity alongside ecosystem and species diversity (Convention on Biological Diversity 2015). In plants, tree populations typically have higher levels of genetic diversity, and also show less genetic structure, than herbaceous plants and shrubs, in part due to high levels of gene flow (Petit and Hampe 2006). Genetic diversity and its spatial pattern are influenced by gene flow, selection and genetic drift (Loveless and Hamrick 1984).

Spatial genetic structure, i.e. the non-random spatial distribution of genotypes, is often observed in plant populations as a result of restricted seed and pollen dispersal (from several metres to tens of kilometres). This isolation by distance effect (Wright 1943), where there is a decreasing probability of mating as the distance between parents increases, leads to offspring being more likely to occur close to their parent plant(s) as a result of non-random mating. Patterns of isolation by distance, both within and between populations, provide an insight into historical rates of gene flow with ecological factors

affecting gene flow particularly important in determining genetic structure (Loveless and Hamrick 1984).

In plants, gene flow occurs through the dispersal of pollen (which carries the genetic information of the father) and seed (which carries the genetic information of both the father and mother). Dispersal of pollen and seed can be mediated by gravity, wind (anemochory), water (hydrochory), and animals (zoochory), each of which shape the pattern of gene flow in different ways (Ashley 2010). Determining how ecological factors influence contemporary gene flow requires methods that directly estimate gene flow, such as parentage analysis (Sork *et al.* 1999). Parentage analyses, based on the use of statistical analysis models, allow the distance and direction of pollen dispersal to be identified, and how far offspring have travelled from their seed parent. Pollen and seed dispersal patterns and distances can result in different patterns of gene flow. The dispersal pattern and distance of pollen dictates the reproductive neighbourhood size of a plant and the connectivity of plant populations, with seed dispersal influencing not only gene flow, but also patterns of colonisation, recruitment, and demography of plant populations (Ashley 2010). Crucially, understanding intrinsic processes such as dispersal patterns also requires knowledge of how landscape features structure populations.

1.3 Landscape genetics

The term landscape genetics was first described in 2003 (Manel *et al.* 2003). Landscape genetics seeks to explain spatial genetic patterns by using landscape variables, such as forests or open fields, to understand how geographic and environmental landscape heterogeneity influence spatial genetic variation, population structure, and gene flow

(Manel *et al.* 2003). Landscape genetic studies incorporate, alongside the collection of genetic data, the exact geographic location of individuals, enabling tests of landscape heterogeneity on patterns of gene flow and genetic variation within and between populations (Storfer *et al.* 2007). Although a burgeoning field of study, landscape genetic studies of plants is under-explored (Holderegger *et al.* 2010), including the interaction of pollen and seed flow with the landscape (Sork and Smouse 2006). Current priorities for development in this field of research include the consideration of the temporal dimension of landscapes, the incorporation of species life history attributes, and the examination of genetic connectivity based on dispersal, all of which can inform management practice (Bolliger *et al.* 2014).

1.4 Riverine landscapes and their riparian habitat

Within a landscape, river corridors exist as linear features extending across large geographical areas from their headwaters to the sea. Rivers are not homogeneous features of the wider landscape however, rather they form their own landscape characterised by their diverse mosaic of landscape elements and ecological processes that occur across a range of spatial and temporal (seasonal to millennial) scales (Ward *et al.* 2002; Wiens 2002). Landscape elements include the river corridor itself, situated within and upon a network of alluvial channels, areas of lotic (flowing water), semi-lotic (areas only connected to the main river channel at their downstream end e.g. abandoned channel segments) and lentic (standing water e.g. lakes) waterbodies, geomorphic features, and riparian habitat (Ward *et al.* 2002). As distinctive water-based landscapes, their hydrology is a key factor in the dynamic nature of rivers, continuously shaping patch shape and movement, and alternatively shifting floodplains through terrestrial and

aquatic phases (Wiens 2002). Importantly, connectivity provided by water is an important feature of river landscapes, linking the landscape through the exchange of matter, energy, and biota across different spatial scales (Ward *et al.* 2002).

It is these distinctive riverine landscape features that contribute to riparian habitat being among the most diverse and dynamic of terrestrial ecosystems, adding disproportionately to both terrestrial and aquatic ecosystem function and diversity (Gregory *et al.* 1991; Naiman *et al.* 1993). The word riparian refers to the biotic communities on the shores of streams and lakes (Naiman and Décamps 1997) and can include, for example, areas of riparian forest, wetland and floodplain. At the interface between terrestrial and freshwater ecosystems, riparian vegetation forms a complex mosaic of habitats shaped by a range of allogenic (externally imposed environmental influences) and autogenic (vegetation-environment interactions that would not occur without vegetation, and which influence vegetation dynamics) processes (Francis 2006). Key allogenic processes influencing riparian vegetation include the formation of river bars enabling the establishment of pioneer riparian vegetation, the hydrochorous dispersal of plant propagules, and hydrological fluctuations leading to disturbance such as damage to, saturation, and burial of riparian plants (Francis 2006). In turn, autogenic processes can promote bank stability and sedimentation, as well as river island formation, thus influencing both plant and river dynamics from the earliest stages of plant establishment (Francis 2006).

As linear features within a landscape, it is clear that riverine features affect the connectivity and genetic structure of riparian populations at a range of spatial and

temporal scales. Taking an ecological genetics approach enables the identification of gene flow within and between populations, thus providing an insight into connectivity between riparian populations. Examining genetic diversity and genetic structure provides further insight into the influence of riverine features on populations, particularly when undertaken at different spatial scales. By taking a landscape genetics approach to link landscape features to genetic patterns, gene flow, genetic diversity, and genetic structure can be analysed in light of the landscapes they occur in. Moreover, by incorporating species ecological life history traits, our understanding of the interactions between gene flow, ecology, and riverine landscapes, and how they structure plant populations is furthered.

1.5 Thesis outline

The aim of this thesis is to examine patterns of genetic diversity and genetic structure in riparian *A. glutinosa* populations, and to relate underlying processes to the genetic patterns observed at different spatial scales of the River Tay catchment. Three research-based chapters are presented, each in manuscript format.

Chapter 2: Effects of seed- and pollen-mediated dispersal on between-generation genetic diversity and genetic structure within riparian A. glutinosa woodlands

The aim of this chapter is to investigate local-scale gene flow within and between four riparian *A. glutinosa* populations of the River Tummel. In particular, the influence of the different dispersal mechanisms is studied. A paternity analysis of *A. glutinosa* seed is used to identify patterns of wind-mediated pollen dispersal and a maternity analysis of *A. glutinosa* saplings identifies patterns of within-site, wind-mediated and between-site,

water-mediated seed dispersal. Dispersal directions, distances, and the dispersal curves of pollen and seed are compared and the influence of the dispersal processes on between-generation spatial genetic structure is examined.

Chapter 3: Detection of demographic and genetic structure in the riparian A. glutinosa woodlands of a dynamic river system

The aim of this chapter is to identify patterns of demographic and genetic structure in riparian *A. glutinosa* woodland of a downstream reach of the River Tummel, and to relate observed patterns to features of the riverine landscape. A dendrochronology approach is taken to create a size-age standard for *A. glutinosa*, and combined with field measurements, used to examine patterns of spatial and temporal riparian woodland structure. Patterns of demographic structure are compared to patterns of genetic structure and the influence of hydrogeomorphological processes discussed.

Chapter 4: Landscape genetics of a key riparian tree species A. glutinosa at a river catchment scale

The aim of this chapter is to examine landscape-scale effects on genetic connectivity between *A. glutinosa* populations. In particular, the unidirectional nature of flow in rivers is utilised to compare *A. glutinosa* populations connected only by wind-mediated pollen dispersal to populations connected by both wind-mediated pollen dispersal and water-mediated seed dispersal. Patterns of genetic diversity within populations along the course of the river are examined to investigate whether gradients of upstream – downstream genetic diversity occur. Genetic structure of riparian *A. glutinosa*

populations within and between rivers is compared. Long-distance dispersal of pollen and seed is also examined, looking at patterns of isolation by distance.

1.6 Study system

The River Tay catchment is the focus of study for the research presented in this thesis. The River Tay catchment, located in the eastern Highlands of Scotland, is the largest river catchment in Scotland, draining an area of approximately 5,000 km^2. It has six major river tributaries: the Garry, Tummel, Lyon, Isla, Almond, and Earn and a number of large lochs including Lochs Ericht, Rannoch, and Tummel, with Loch Tay being the largest loch at 23 km long. Records from the most downstream river flow gauging station at Ballathie, just south of the confluence of the Rivers Tay and Almond, where the River Tay is 90 m wide, give the mean annual flow to be 165 m^3 s^{-1}, the highest mean flow in the UK (Marsh and Lees 2003). The majority of rivers within the Tay catchment originate in the mountains of the Grampian Highlands and flow south-east across the Highland Boundary Fault to the East Central Lowlands, becoming tidal at Perth and entering the North Sea at Dundee (Forest Research 2013). To the north of the Highland Boundary Fault the upland geology is composed of resistant metamorphic rocks, and to the south it is comprised of softer sedimentary sandstones occurring over gentler gradients (Forest Research 2013). The River Tay catchment has numerous designations including Special Areas of Conservation (SACs), Special Protection Areas (SPAs), Natura 2000, and Sites of Special Scientific Interest (SSSIs). Montane habitat, heath and bog cover the higher areas of the catchment, and rough grazing dominates the uplands with improved grassland and intensive arable land in river valley bottoms (Forest Research 2013). Woodland and plantation forestry account for 15% of the catchments area (Forest Research 2013). The riparian woodland of

the River Tay catchment forms the focus of study in chapter four of this thesis, with chapters two and three focusing on riparian woodland of the River Tummel, one of the major tributaries of the River Tay.

The River Tummel is a large and active wandering gravel-bed river. It has a catchment area of 1670 km^2 and originates at Loch Rannoch in the west, flowing for 93 km before joining the River Tay at Logierait. Records from the river flow gauging station at Pitlochry, where the river is approximately 40 – 50 m wide (Parsons and Gilvear 2002) give the mean annual flow to be 73 m^3 s^{-1} (Marsh and Lees 2003). The River Tummel is dynamic in nature with lateral movement evident from the analysis of old maps, aerial photography, and recent events (Gilvear and Winterbottom 1992; Winterbottom 2000; Parsons and Gilvear 2002). Early maps show the lower 10 km of the river, downstream of its confluence with the River Garry, to be multi-channelled and unstable, however embankments constructed during the 18th and 19th centuries led to channel change, principally the confinement of the channel to a single course (Gilvear and Winterbottom 1992; Winterbottom 2000). Following a large flood event in 1903 the embankments were allowed to fall into disrepair and the river has since returned to a more natural state with further channel change and movement occurring, characteristic of a 'mobile' gravel-bed river (Figure 1.1) (Gilvear and Winterbottom 1992; Winterbottom 2000; Parsons and Gilvear 2002). Since abandonment of the flood embankments, fluvial landforms and mosaic patches of vegetation have evolved principally as a consequence of flood-induced planform change and fluvial disturbance (Parsons and Gilvear 2002). In particular, this has included the movement of river gravels, particles larger than sand (>2 mm diameter) but smaller than boulders (<256 mm in diameter), for which the Rivers Tay and Tummel have

the largest volumes of in the UK (Scottish Natural Heritage 2008). Accumulations of exposed gravel are important for establishing riparian vegetation and key gravel areas of the River Tummel are designated as the Shingle Islands Site of Special Scientific Interest and the Shingle Islands Special Area of Conservation for Annex I feature 'alluvial forests with *Alnus glutinosa* and *Fraxinus excelsior*' (JNCC 2014).

Figure 1.1 Features of the River Tummel. Top left: Abandoned former river channel, lined with mature *Alnus glutinosa* trees. Top right: upstream-end of backwater of River Tummel; Middle left: downstream-end backwater of River Tummel (visible in background); Middle right: gravel deposition at Moulinearn on the River Tummel, with alluvial forest visible in the background; Bottom left: gravel and woody-debris deposition at Moulinearn: Bottom right: gravel bar formation, with the main River Tummel channel visible to the left and a narrow channel just visible to the right. All photographs taken along the River Tummel, within the designated Shingle Islands area. [Photographs by G Flint]

1.7 Study species – *Alnus glutinosa*

Alnus glutinosa (L.) Gaertn., commonly known as black alder, is one of 29 – 35 *Alnus* species that form part of the Betulaceae family, with four – five *Alnus* species native to Europe, nine species native to the Americas, and 18 – 23 species native to Asia (Chen and Li 2004). *Alnus glutinosa* is distributed throughout most of Europe, from mid-Scandinavia to the Mediterranean, with rear edge populations occurring as far as the mountains of Turkey and North Africa (Figure 1.2). Although there is uncertainty over the arrival time of *A. glutinosa* in Britain, evidence shows a major expansion of *A. glutinosa* occurring across Britain from 7500 BP (Chambers 1989). *Alnus glutinosa* is considered the only native *Alnus* species in the UK although *A. incana* and *A. cordata* also occur, having been introduced for cultivation in 1780 and 1820 respectively, and subsequently recorded in the wild in 1922 and 1935 respectively (Preston *et al.* 2002).

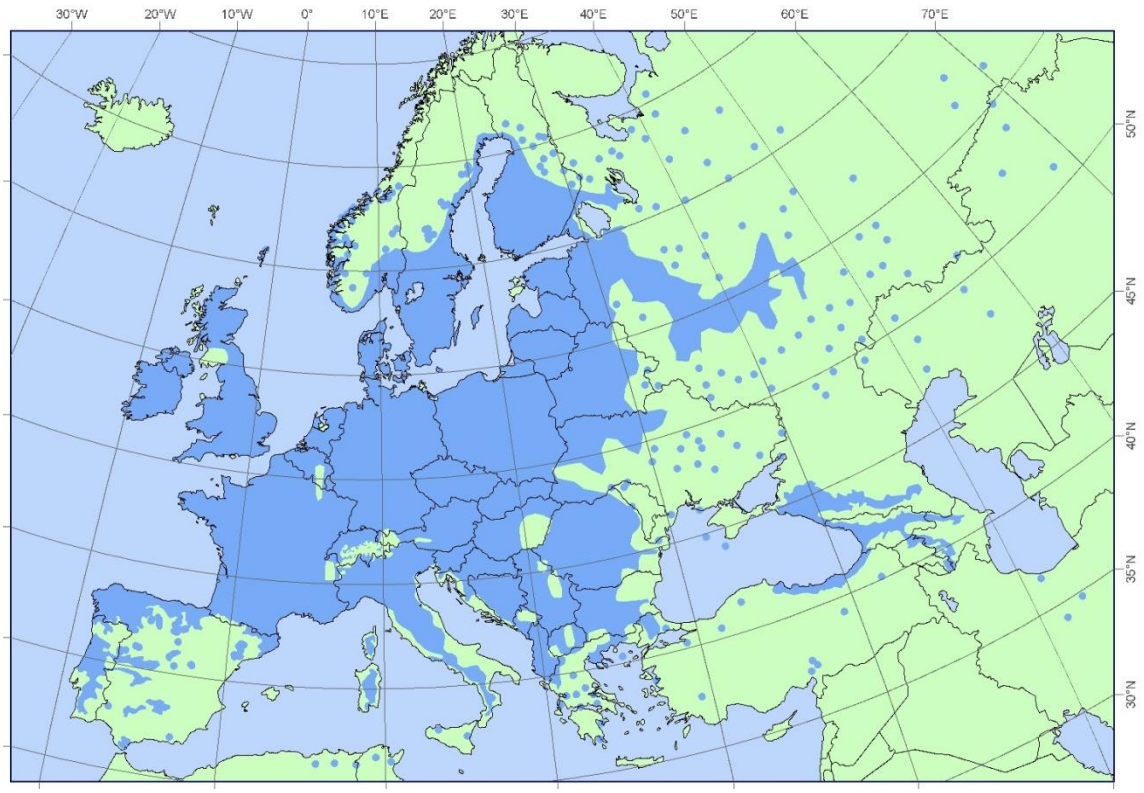

Figure 1.2 Distribution map of *Alnus glutinosa.* Blue shaded areas show natural distribution of *A. glutinosa*. (EUFORGEN 2008).

Alnus glutinosa is an important tree species of European riparian ecosystems. It contributes to biodiversity by supporting a range of flora and fauna both on the tree itself and in the flooded root system. In waterlogged soils it contributes to water filtration and purification. The root system helps to control floods and reduce river bank erosion. In addition, it influences water temperature and nutrient cycling, and is able to fix atmospheric nitrogen in symbiotic root nodules with *Frankia* bacteria (Claessens *et al*. 2010). It is an important species in alluvial and marshy ecosystem restoration projects, contributing to both nature conservation and watershed management (Claessens *et al.* 2010). It is listed as a key species for conservation under the European Habitats Directive under the Annex I priority habitat 91E0* 'alluvial forests with *Alnus glutinosa* and *Fraxinus excelsior*' (Council of the European Communities 1992). It is also an important forestry species, and is grown for timber, paper, joinery purposes, and for use underwater where it is very long-lasting (Claessens *et al*. 2010).

Alnus glutinosa is a monoecious, self-incompatible broad-leaved tree (McVean 1953; Steiner and Gregorius 1999). They are a relatively short-lived tree with a maximum lifespan between 100 -160 years old (Claessens *et al*. 2010). *Alnus glutinosa* mature between the ages of three to 30, with the male and female flowers forming as catkins during February / March (McVean 1955), producing seed every three to four years (Claessens *et al.* 2010). Following wind-pollination and fertilisation the female flower forms as a woody, cone-like fruit containing approximately 60 seeds (strictly achenes) that are dispersed in autumn (McVean 1953). A mature tree can produce approximately 4,000 cones in one year (Claessens *et al.* 2010). The viability of *A. glutinosa* seed is generally low and highly variable (0 – 80%), whether between years, between

populations, or between trees, and appears to be a function of a high level of unfertilised seed (McVean 1955; Claessens *et al.* 2010). *Alnus glutinosa* seed is small (1 – 2 mm in diameter) and light and is dispersed by both wind and water. Water dispersal is the principal dispersal mechanism and the seeds possess adaptations particularly suitable for water dispersal including lateral cork-like float chambers, aiding buoyancy, and an oily outer coat (McVean 1955). The seed is able to float in still water for over 12 months and wind and wave action concentrates seed along shore lines (McVean 1955). Wind dispersal of the seed occurs over a distance of 30 m – 60 m, and most wind-dispersed saplings are 20 m – 30 m from the parent tree (McVean 1955). Germination of the seed occurs on the surface of the vegetation or soil, due to its buoyancy (McVean 1953), from late September onwards (McVean 1955). Seed dispersal is influenced by abiotic events such as high winds and floods. Wind dispersed seed forms clusters of saplings adjacent to parent plants, or linear populations along the edge of woods where concentric lines of even-aged trees may develop. Water dispersed seed becomes concentrated in river meanders, and following floods forms lines of seed along high-water marks, establishing on river alluvium and mud (McVean 1956). The hydrophyte seedling requires a high water-table, or high rainfall, as well as high light intensity (McVean 1953; Claessens *et al.* 2010).

Growth of *A. glutinosa* is rapid when young, with most diameter growth occurring in the first 15 years of life (Claessens *et al.* 2010). Growth habit is described as variable by McVean (1953), from low multiple-stemmed bush form to tall single-bole trees. Although most *A. glutinosa* reproduction is by seed, various types of vegetative reproduction also occur (McVean 1953). Root suckers, or sprouting, although not rare, are more frequent in grazed forests and dynamic systems; ageing trees produce trunk suckers that develop into

individual trees when the parent dies; and adventitious roots may form on uprooted trees or fallen branches (Koop 1987)(Figure 1.3).

Figure 1.3 *Alnus glutinosa* growth. Top left: Single-bole *A. glutinosa* tree. Top middle: Multi-stemmed *A. glutinosa* with two stems; Top right: ageing *A. glutinosa* tree with suckers; Bottom left: *A. glutinosa* regeneration on shingle; Bottom right: uprooted and deposited *A. glutinosa* tree establishing on area of bare shingle. All photographs taken along the River Tummel, within the designated Shingle Islands area. [Photographs by G Flint]

Alnus glutinosa occurs throughout the River Tay catchment and was sampled from a range of habitats including river bank trees, riparian woodland, and slopes with wet flushes next to rivers, lochs, and abandoned river channels. The history of most *A. glutinosa* sampling locations within the River Tay catchment is unknown, however all populations are considered natural although it is likely that some populations may have been managed in the past, as coppice woodlands for example (personal observation) (Figure 1.4).

Figure 1.4 Photographs of *Alnus glutinosa* woodlands sampled within the River Tay catchment. Clockwise, from top left: Edinchip wet-flush woodland, River Earn; Old *A. glutinosa* at Coille Criche, Loch Earn; riparian woodland at Drumlochlan, River Earn; *A. glutinosa* woodland adjacent to the Loch of Butterstone, part of the Lunan; wet *A. glutinosa* woodland located in disconnected River Tay channel at Bloody Inches; lochside *A. glutinosa* growth at Dalerb, Loch Tay; *A. glutinosa* coppice stool at Fiddlers Bay, Loch Tay; wet-flush slopes of Loch Tummel; *A. glutinosa* lined River Almond at Newton Bridge; meadow-like growth of *A. glutinosa* at Drumlochlan, River Earn. [Photographs by G Flint]

Chapter 2

Effects of seed- and pollen-mediated dispersal on between-generation genetic diversity and genetic structure within riparian *Alnus glutinosa* woodlands

2.1 Abstract

In plants, gene flow within and between populations is maintained through a combination of pollen and seed dispersal. The identification and characterisation of the distance and direction of gene flow is therefore key to informing many areas of plant science including population connectivity, recruitment, and demography of plant populations. In this study microsatellite markers were used to assess gene flow and spatial genetic structure of *Alnus glutinosa* located in four riparian populations of the River Tummel, Scotland. A maternity analysis of 167 saplings was undertaken, assessing levels of within-population wind dispersal of seed and between-population hydrochorous seed dispersal. A complementary paternity analysis of 398 seeds was undertaken, and spatial genetic structure, within the pollen donor, sapling, and adult (n = 653) generations was investigated. Within-population seed dispersal occurred over distances up to 98 m although between-population hydrochorous seed dispersal, covering distances up to 2.6 km, accounted for most seed dispersal. Pollen dispersal was observed up to distances of 4.3 km although most pollen dispersal was within-population. The extensive gene flow revealed was consistent with the lack of spatial genetic structure identified in the sapling generation. However, although weak, significant spatial genetic structure was observed in the adult generation. By revealing the extent of both seed and pollen dispersal in riparian populations of *A. glutinosa* this study highlights the importance of incorporating ecological processes in the measurement of gene flow. Differences in spatial genetic structure between generations indicates the response to ecological and evolutionary influences varies in the different life stages of *A. glutinosa* and the detection of spatial genetic structure in the adult generation is discussed.

2.2 Introduction

The identification and characterisation of gene flow is key to understanding the processes underpinning gene flow, genetic diversity and spatial genetic structure within populations (Slatkin 1987). In plant populations gene flow is maintained by the dispersal of pollen and seed. If pollen and seed dispersal is unrestricted, little genetic differentiation between populations is expected. However, differences in the pattern and distance of pollen and seed dispersal lead to varying levels of within population genetic diversity and between population genetic differentiation (Ennos 1994; Petit *et al.* 2005).

Although seed dispersal is often assumed to be more spatially restricted than pollen dispersal, a growing body of empirical evidence reports that seed-mediated dispersal may be as widespread as pollen dispersal (Ashley 2010). Although most seed dispersal occurs over short distances, long-distance seed dispersal events, although rare, are found to be disproportionately important (Nathan 2006). Species that occupy habitats prone to extreme flooding events, such as river banks, are more likely to experience long-distance seed dispersal (Nathan *et al.* 2008). Further to this, the seeds of many riparian species have adaptations enabling hydrochory, the passive dispersal of organisms by water (Nilsson *et al.* 2010).

In plant populations, direct measures of inter-population gene flow can be based on bi-parentally inherited nuclear and / or uni-parentally inherited organelle genomes (Ennos 1994). DNA microsatellites can provide accurate markers to assign parentage and parentage analysis provides a practical method of estimating dispersal (Ashley 2010).

When using nuclear markers, parentage analysis, based on sampling all possible parents, or a proportion of possible parents within a prescribed area, enables the identification of both the maternal and paternal parent. Paternal analysis, undertaken to identify the pollen parent, is typically based on offspring from a known, genotyped seed parent, thus the maternal allele of the seed can be identified and excluded from the offspring genotype. Pollen parents are thus identified based on finding a matching paternal allele from the sampled pool of possible parents (Ashley 2010). Maternal analysis, to identify the maternal parent, poses further challenges, particularly when the origin and destination of seed has not been tracked (Ashley 2010). In this case, although parentage analysis can assign two potential parents it is not possible to confirm whether an assigned parent is the seed or pollen parent (Ashley 2010) although, where knowledge of the species ecology is available, assumptions can be made as to which assigned parent is the maternal parent (e.g. Bacles *et al.* 2006; Nakanishi *et al.* 2009; Vranckx *et al.* 2014). For example, in wind-pollinated, co-sexual plants, the geographically closest parent is assumed to be the maternal parent (Bacles *et al.* 2006; Nakanishi *et al.* 2009; Vranckx *et al.* 2014). Inevitably, any assumption made introduces bias into any insight gained although other methodological approaches, such as examination of spatial genetic structure, can be used to ascertain whether the parentage analysis results obtained are consistent with the results and insight gained from other analyses.

As patterns in the dispersal distances of pollen and seed often occur, spatial genetic structure (SGS), the non-random spatial distribution of genotypes, is expected to occur frequently in plant populations (Vekemans and Hardy 2004). Typically, most gene flow between plant populations is expected to occur via the movement of pollen because

pollen may disperse over longer distances and in greater numbers than seed dispersal, especially where pollen is wind dispersed (Ennos 1994; Petit *et al.* 2005). In comparison, seed dispersal is often spatially limited therefore offspring tend to be more spatially aggregated. If seed dispersal is spatially restricted, as expected in plant populations, the spatial distribution of maternal half-sibling offspring will be more constrained than the spatial distribution of paternal half-sibling offspring (Vekemans and Hardy 2004; Nakanishi *et al.* 2009). Thus, the formation of kinship structures is considered the most prevalent cause of fine-scale spatial genetic structure in plant populations (Vekemans and Hardy 2004).

Studies of SGS benefit by incorporating comparison between different age classes, enabling the detection of changes in SGS across life stages (Kalisz *et al.* 2001; Fuchs and Hamrick 2010). This information can provide some insight into the various ecological and evolutionary processes shaping SGS at different life stages. For example, the SGS of newly dispersed seeds may reflect the SGS of the parent cohort and / or the dispersal distances of pollen and seed (Kalisz *et al.* 2001). A subsequent loss of kinship structure in successive demographic stages is expected to reflect recruitment patterns, density-dependent mortality, or post-dispersal selection (Kalisz *et al.* 2001; Fuchs and Hamrick 2010).

Although the level of, and change in, SGS gives some insight into the demography and evolution of populations it is clear that quantification of pollen and seed dispersal patterns would greatly improve our understanding of demographic processes. To better understand the processes shaping genetic diversity across generations, this study aimed to simultaneously assess contemporary patterns of gene flow alongside examining kinship

structure in four populations of *Alnus glutinosa*, a keystone tree species of riparian ecosystems. Parentage analyses were conducted to assess wind mediated pollen dispersal and hydrochorous seed dispersal within and between populations, and genetic diversity and kinship structure across the adult and sapling generations were examined. Specifically, the following hypotheses were tested:

I. Greater genetic diversity will be observed in the sapling generation than the adult generation as a consequence of widespread pollen dispersal increasing initial genetic diversity, and recruitment and density-dependent mortality reducing genetic diversity in the adult generation.

II. Extensive between-population gene flow will occur as a consequence of widespread, wind-dispersed pollen.

III. Seed dispersal will be less extensive than pollen dispersal, as typically reported in tree species where widespread, wind-dispersed pollen occurs.

IV. Greater spatial genetic structure will be observed in the sapling generation, compared to the adult generation, as a consequence of the spatially aggregated growth of *A. glutinosa* saplings

2.3 Materials and methods

2.3.1 Study species

Alnus glutinosa (L.) Gaertn, commonly known as black alder, is an important temperate riparian tree species distributed across Europe. Growing alongside the edge of rivers and standing water, *A. glutinosa* can act as a pioneer species and forms a key element of

dynamic river systems (McVean 1953). It can grow on a wide range of soils and is able to fix atmospheric nitrogen in symbiotic root nodules with *Frankia* bacteria (Claessens *et al.* 2010). A monoecious tree, the male and female flowers form as catkins during February / March and are wind-pollinated (McVean 1955). Following fertilisation the female flower forms as a woody cone-like fruit containing approximately 60 seeds (strictly achenes) that are dispersed in autumn (McVean 1953). The seeds, which have lateral cork-like float chambers and an oily outer coat, are principally dispersed by water, although seed dispersal by wind can occur up to a distance of 60m (McVean 1955). Trees mature between the ages of three and 30 years and produce seed every three to four years (Claessens *et al.* 2010). *Alnus glutinosa* is considered self-incompatible (Steiner and Gregorius 1999). Growth of *A. glutinosa* is rapid when young, with most diameter growth occurring in the first 15 years of life (Claessens *et al.* 2010). Growth habit is described as variable by McVean (1953), from low multiple-stemmed bush form to tall single-bole trees. At the locations studied here some trees appeared as multi-stemmed trees, characterised by trunks growing individually but so close to one another at the base that is was not possible to distinguish whether they were the same tree or not. Future reference to multi-stemmed trees refers to this growth characteristic.

2.3.2 Study area

Sampling of *A. glutinosa* trees took place within four riparian populations adjacent to the River Tummel in the eastern Highlands of Scotland (Figure 2.1).

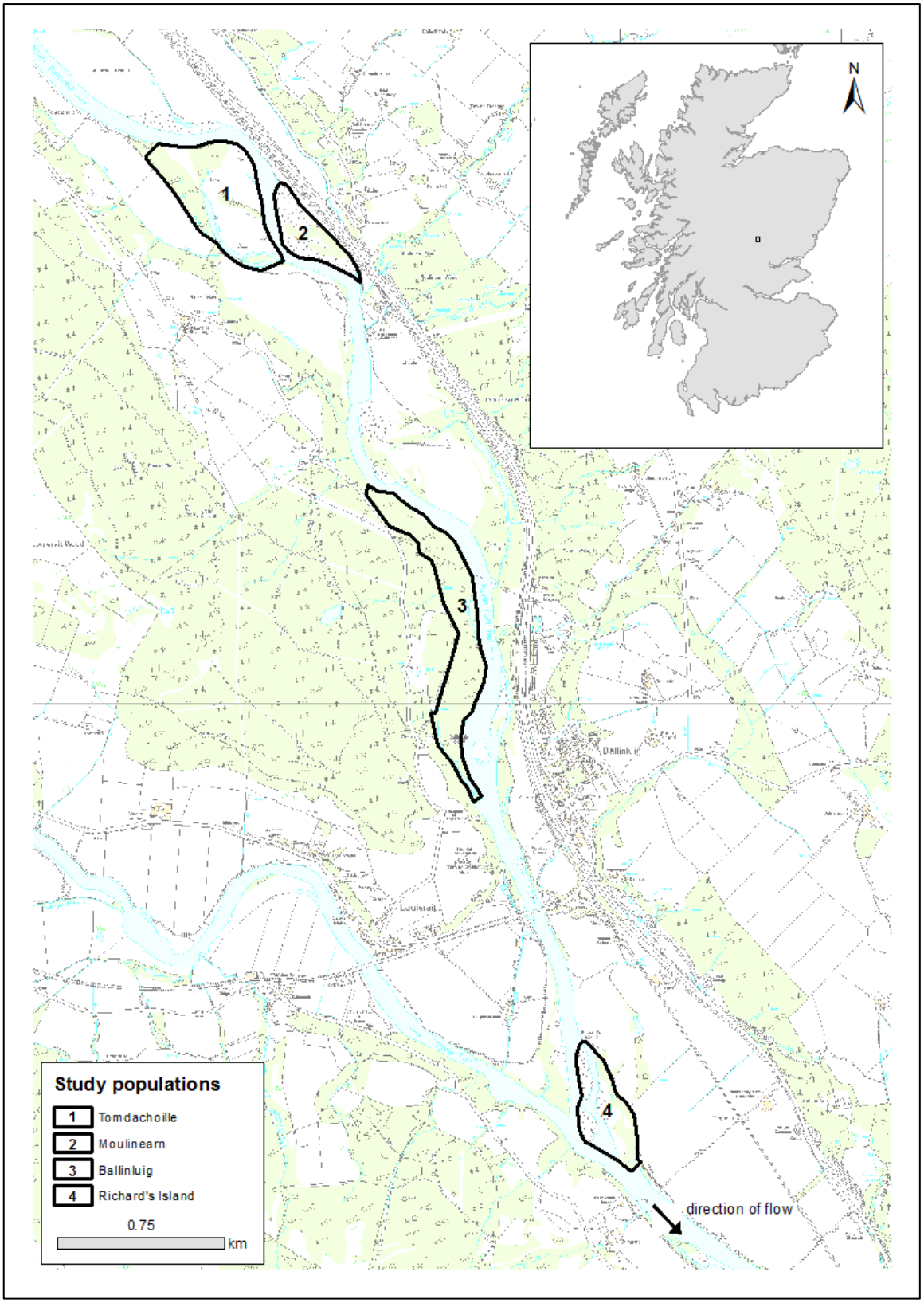

Figure 2.1 Map showing the location of the four study sites on the River Tummel. Inset map shows the location of the River Tummel in Scotland.

Each site encompasses areas of dynamic river shingle, where river action deposits and re-arranges shingle, and a wide range of successional vegetation communities occur. *Alnus glutinosa* woodland occurs on shingle and other alluvial soils, alongside areas of bare shingle, neutral grassland, and open water, including abandoned river channels and backwaters. Here, the term 'backwater' refers to a former river channel that, through the deposition of alluvial or woody debris, has lost its upstream connection with the main river channel but maintains a downstream connection with the river channel. The four populations, referred to here, in upstream to downstream order, as Tomdachoille, Moulinearn, Ballinluig, and Richard's Island, occur over an approximate 6 km stretch of the River Tummel with Richard's Island located on the confluence of the River Tummel and the River Tay. Each population varies in size, ranging from 6.5 Ha to 19.3 Ha (Table 2.1). These four populations are of national importance, designated as the Shingle Sands Site of Special Scientific Interest (SSSI), forming a series of extensive and dynamic river shingle areas in various stages of colonisation (Scottish Natural Heritage 2013). The study populations are also of European conservation importance forming part of the Shingle Islands Special Area of Conservation (SAC) for Annex I priority feature 'alluvial forests with *Alnus glutinosa* and *Fraxinus excelsior*' (JNCC 2014). All necessary permissions were gained from the landowners, land manager, and the statutory authority prior to fieldwork commencing.

2.3.3 Sample collection

In 2011, leaf material was collected from adult and sapling trees from each of the study populations. Within each population up to 277 trees were randomly sampled at a range of distances apart across each site (Table 2.1). Trees were identified as adults where ≥20

cm circumference at 30 cm height. Trees were identified as saplings based on their height (>30 cm high) and size (<20 cm circumference at 30 cm height). Occasionally, in the case of multi-stemmed trees, leaves were collected from more than one stem. In this situation each stem was recorded as an individual sample, but shared the same geographical location as the other sampled stems. The geographical location of each sampled tree was recorded using a Garmin GPSMAP 62s handheld navigator. Leaf samples were immediately placed in silica gel (Chase and Hill 1991) and subsequently stored at room temperature.

Seed samples were collected from the Tomdachoille population only. To ensure certainty of maternal origin of progeny, seed samples were collected directly from the canopy of all fruiting trees sampled for leaf material, a subset of 42 individual trees (38 adults and four saplings). Up to ten seed cones were collected from each maternal tree, placed in individual paper bags and subsequently stored at room temperature.

2.3.4 Seed germination

Seeds were germinated on moist filter paper (Whatman 90 mm) following the protocol outlined in Gosling *et al.* (2009). Principally, seed was soaked in water for 24 hours at 10 ^{o}C with no light, followed by a pre-chill period of 21 days at 4 ^{o}C in the dark, followed by a growing period of up to 28 days, with each day consisting of 16 hours in darkness at 20 ^{o}C and eight hours in UV fluorescent light at 30 ^{o}C. Germinated seeds were harvested when the cotyledons emerged and stored in individual eppendorfs at -80 ^{o}C until DNA extraction.

2.3.5 DNA extraction and microsatellite analysis

Genomic DNA from the germinated seed samples was obtained using the Isolate Plant DNA Mini Kit (Bioline) according to manufacturer instructions and eluted into a final volume of 30 µL. Genomic DNA from the leaf samples was obtained using the DNeasy 96 Plant Kit (Qiagen) following manufacturer instructions. All extracted DNA was stored at -20 ^{o}C until further use. DNA concentration was quantified using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific) and samples adjusted to 10 ng/ µL for PCR amplification. DNA was amplified using the multiplex of 12 nuclear microsatellite markers (Ag01, Ag05, Ag09, Ag10, Ag13, Ag14, Ag20, Ag23, Ag25, Ag27, Ag30, Ag35) of Lepais and Bacles (2011). Multiplex reactions were carried out in a total volume of 5 µL, using 1X Type-it Microsatellite PCR Kit (Qiagen) and 0.5 µL of template DNA, and performed in a Veriti thermocycler (Applied Biosystems). Each PCR amplification included negative controls to monitor contamination and reference samples were included to standardise scoring, as recommended by Bonin *et al.* (2004). PCR conditions followed those described in Lepais and Bacles (2011): 5 min denaturation at 95 ^{o}C followed by 30 cycles of 95 ^{o}C for 30 s, 58 ^{o}C for 180 s, and 72 ^{o}C for 30 s, finishing with a final elongation step of 60 ^{o}C for 30 min. Following test amplicon success on 2% agarose gel 1 x TBE electrophoresis, samples were sent to DNA Sequencing and Services (Dundee, UK) for fragment analysis on a Biosystems 3730 capillary sequencer at a 1:50 dilution using GeneScan 500 LIZ size standard (Applied Biosystems). The resulting electropherograms were analysed using GeneMarker v.2.4.0 software (Softgenetics) and the correct assignment of allele size class checked in FlexiBin (Amos *et al.* 2007).

Genotypic linkage equilibrium between loci pairs within each population, across all four populations, and within the seed cohort was checked using FSTAT v.2.9.3.2 (Goudet 1995). Significant associations between loci were identified by randomly associating genotypes at pairs of loci over 1100 and 8800 permutations respectively, based on the 5% nominal level after Bonferonni correction.

Table 2.1 Details of the four study sites and the number of *A. glutinosa* adult and sapling trees sampled at each site.

Site name	Site code	Latitude	Longitude	Size of site (Ha)	No. adults sampled	No. saplings sampled
Tomdachoille	Tom	3° 41' 43''	56° 40' 38''	19.3	190	38
Moulinearn	Mou	3° 41' 20''	56° 40' 35''	6.5	128	70
Ballinluig	Bal	3° 40' 36''	56° 39' 37''	18.5	232	45
Richard's Island	RIs	3° 39' 52''	56° 38' 26''	9.6	180	19

2.3.6 Genotyping error

Based on Bonin *et al.* (2004), repeat amplification (including some blind samples) and fragment analysis was undertaken to estimate allele and genotype mismatch errors within each population and across all populations, as well as within the seed cohort. Across all four populations a total of 160 leaf samples (20% of the total genotyped sample size) were repeated, and within the seed cohort, 74 seed samples (18% of the total) were repeated. Genotyping errors and null alleles were quantified using two methods, as advocated by Dąbrowski *et al.* (2014), implemented in Micro-Checker (Van Oosterhout *et al.* 2004) and Cervus v.3.0.7 (Kalinowski *et al.* 2007). In addition, mismatches between seed genotypes and maternal parent genotypes were identified, and genotyping error rates quantified by direct comparison of offspring-mother genotypes in Cervus v.3.0.7 (Kalinowski *et al.* 2007).

2.3.7 Genetic diversity of adult and sapling generations

Initial analysis sought to clarify whether leaves collected from more than one stem of the same multi-stemmed tree shared the same genotype. Implemented in GenClone v.2.0 (Arnaud-Haond and Belkhir 2007), genotypes from all sampled trees were compared and, where identical genotypes were revealed, field records were consulted to determine whether they were sampled from the same multi-stemmed tree (*i.e.* shared the same geographical coordinates). Where clones were detected, only one individual was retained for subsequent analysis. Prior to assessing genetic diversity between the adult and sapling generations the potential for pseudo-replication within the sapling generation was excluded by undertaking parentage analysis. As described in Section 2.3.8, the parentage analysis sought to identify whether any saplings shared the same maternal allele (*i.e*. half sibs) so that saplings sharing the same maternal parent could be excluded from further analyses comparing the genetic diversity between generations. Similarly, to avoid pseudo-replication, genetic diversity statistics of the seed cohort were not calculated as individual seeds shared the same maternal parent. Genetic diversity statistics for the adult and sapling generations in each population and across all four populations were then calculated. The mean number of alleles per locus (N_A), allelic richness (A_R) (Petit *et al.* 1998), gene diversity (H_E) corrected for sample size (Nei 1978) and the inbreeding coefficient (F_{IS}) (Weir and Cockerham 1984) were calculated using SPAGeDi 1.4c (Hardy and Vekemans 2002). Rarefaction analysis of A_R was based on 36 gene copies and significance of F_{IS} values were obtained following 10,000 permutations of gene copies within individuals relative to each population. Significant differences between the adult and sapling generations were tested using the Wilcoxon signed rank test, implemented

using R (R Core Team 2014). Finally, the percentage of rare alleles in each generation within each population was calculated by dividing the number of alleles occurring with a <0.05 frequency by the number of alleles present within the relevant generation of each population. Significant differences between the adult and sapling generations were tested using the Wilcoxon signed rank test implemented using R (R Core Team 2014).

2.3.8 Parentage analysis

Two types of parentage analyses were undertaken. A maternity analysis was implemented to identify the presence of siblings within the sapling generation, and to enable investigation into seed dispersal. A paternity analysis was implemented to identify the pollen parents of the seed cohort, enabling assessment of contemporary pollen-mediated gene flow.

Maximum likelihood (ML) parentage analysis was implemented in Cervus v.3.0.7 (Kalinowski *et al.* 2007). For each putative candidate parent, Cervus calculates an overall log-likelihood (LOD) score, obtained by taking the natural log of the ratio between the likelihood that the candidate parent is the true parent and the likelihood that the candidate parent is not the true parent. Confidence in the LOD score is determined by running simulations, based on parent allele frequencies, to estimate critical LOD values above which candidate parents can be considered a true parent at a given confidence level.

Here, 10,000 simulations were run to simulate parent allele frequencies, parent pair genotypes, offspring genotypes based on Mendelian sampling of the alleles from the two

parental individuals, and a series of random genotypes representing unrelated candidate parents. For each simulated offspring Cervus identified the most likely candidate parent which may or may not be the true parent. By comparing the distribution of the LOD scores for offspring where the most likely candidate parent was the true parent, with the distribution of the LOD scores for offspring where the most likely candidate was an unrelated individual, critical LOD scores were identified at the 80% and 95% confidence level. Parentage was assigned where the LOD score obtained from the parentage analysis exceeded the (simulated) critical LOD score.

2.3.8.1 Maternity analysis

Maternity analysis was undertaken to identify the maternal parent of the 167 saplings sampled across all four populations. Parent trees were considered to be any of the 653 genotyped adult trees sampled across all four populations. The most-likely parent and parent pair analysis was implemented using the default error rate of 1%, higher than the allele mismatch error (reported in Section 2.4.2) thus providing some allowance for undetected error. Allele frequencies calculated across all four populations were used, and the proportion of loci typed set to 0.9966 (based on Cervus allele frequency output). To enable maximum assignment of maternal parents it was assumed that all possible maternal parents had been sampled. Subsequent analysis based on maternal parent assignment assumed that all maternal parents were located upstream of the sired sapling and / or within 60 m of the sired sapling. These assumptions are based on the findings that *A. glutinosa* seeds are predominantly dispersed by water, presumably in a downstream direction, but may also be dispersed by wind up to a distance of 60 m (McVean 1955; Chambers and Elliot 1989). Consequently, if a single parent was identified

it was excluded from further analysis as it was unknown whether it was the maternal or paternal parent. If a parent pair was identified the parent located upstream and / or within 60 m of the sapling was assigned as the maternal parent. If both parents were located upstream, or within 60 m, of the sapling the tree nearest to the sapling was assigned as the maternal parent. If both parents were located downstream and >60m from the sapling they were excluded from further analysis. Importantly, because the maternity analysis is based on established saplings, subsequent estimates of seed dispersal reflect both the movement (dispersal *per se*) and establishment of saplings. Hence, future reference to seed dispersal estimates obtained here refer to 'effective seed dispersal', incorporating both dispersal and establishment processes (Cain *et al.* 2000).

Following assignment of maternal parents, the number of local (*i.e.* maternal parent and sapling located in the same population) seed dispersal events were counted and compared to the number of seed dispersal events from neighbouring populations (*i.e.* maternal parent and sapling located in different populations). Seed dispersal distances were calculated as the distance between each sapling and its maternal parent. The resultant distribution of seed dispersal distances was then compared to that expected under random mating, *i.e.* the distance between each sapling and each possible maternal parent sampled in the field.

2.3.8.2 Paternity analysis

Paternity analyses were undertaken to identify the pollen parent of the 398 seeds that shared a compatible multilocus genotype with their mother (excluding 18 seeds presenting at least one mismatching allele with their mother). Pollen parents were

considered to be any of the 820 genotyped adult and sapling trees sampled across all four populations, excluding the possibility for the mother tree to self. Allele frequencies calculated across all four populations were used, and the proportion of loci typed set to 0.9978 (based on Cervus allele frequency output).

Confidence in parentage assignment is influenced by two user-supplied variables, the genotyping error rate and the proportion of candidate parents sampled (Jones *et al*. 2010). In Cervus, both of these variables are assumed to be known *a priori* and are user-supplied in the simulation analysis stage. Here, an error rate of 1% was used, higher than the mismatch error rate between the known parent and their offspring (0.5%) and also higher than the repeat genotyping error rate per allele (0.1%), providing some allowance for undetected error.

To accommodate uncertainty around the proportion of candidate parents sampled both ML and exclusion paternity analyses were undertaken. As the proportion of candidate parents sampled declines the critical LOD score increases, reducing the success rate of parentage analysis. In this study the proportion of candidate parents sampled was unknown, particularly given *A. glutinosa* is a wind-pollinated tree occurring more or less continuously in the wider landscape. As it was known that not all candidate parents were sampled, either within the seed sample population (Tomdachoille) or within neighbouring populations, paternity analysis was undertaken for varying proportions of parents sampled (N = 820 (100%), 1,093 (75%), 1,640 (50%), 3,280 (25%), 8,200 (10%), 16,400 (5%)), enabling, as advocated by Koch *et al*. (2008), a sensitivity analysis to be carried out for this parameter.

In a strong data set, with few errors, an exclusion-based parentage analysis will be relatively insensitive to the proportion of parents sampled (Jones *et al*. 2010). Therefore, by applying an error rate of zero, exclusion-based paternity analyses where no mismatch between parent and offspring is allowed, was also implemented in Cervus. The results of the ML and the exclusion-based paternity analyses were compared to examine the effect the proportion of parents sampled had on confidence in paternity assignment.

2.3.9 Contemporary pollen-mediated gene flow

Pollen dispersal distances were calculated as the distance between each seed (based on the location of the maternal parent) and its pollen parent. The resultant distribution of pollen dispersal distances was then compared to that expected under random mating, *i.e.* the distance between each seed and each possible pollen parent sampled in the field.

2.3.10 Spatial genetic structure

Assessment of differences in spatial genetic structure (SGS) between the adult and sapling generations was undertaken in SPAGeDI v.1.4c (Hardy and Vekemans 2002). The extent of SGS was assessed in the adult generation occurring within each population and across all populations; in the sapling generation occurring across all populations as the low sample size within each population did not allow for analysis for each population; and between the adult and sapling generations across all four populations. Kinship coefficients (F_{ij}) between individuals *i* and *j* were estimated using Nason's kinship coefficient (Loiselle *et al*. 1995), as it is found to be statistically robust (Vekemans and Hardy 2004). Nason's kinship coefficient is based on the probability that a random gene from *i* is identical to a

random gene from *j*, and defined as F_{ij} = (*Qij* – *Qm*) / (1 - *Qm*), where *Qij* is the average probability of identity by state for random gene copies from individuals *i* and *j*, and *Qm* is the average probability of identity by state for gene copies coming from random individuals from the reference population (Vekemans and Hardy 2004). In each dataset, the association between all pairs of F_{ij} and spatial distances (*r*) was characterised by averaging the pairwise statistics to a set of predefined distance intervals. Preliminary tests were undertaken to establish suitable distance classes that would enable comparison between the adult and sapling generations, and to ensure as close to >100 pairwise comparisons within each distance class, as advised by Hardy and Vekemans (2013). Subsequently, eighteen distance classes were defined, at 20m, 40m, 60m, 80m, 100m, 200m, 300m, 400m, 500m, 600, 700m, 800m, 900m, 1,000m, 2,000m, 3,000m, 4,000m, and 5,100m. For further investigation, and due to the larger dataset, additional analysis for the adult generation was undertaken using smaller distance classes (smallest distance class 5 m). Over 100 pairwise comparisons occurred within each distance class except within the sapling generation where only 40 pairwise comparisons occurred in the 900m – 1000m distance class. Analyses of the adult generation within each population were restricted to a maximum distance of 400m to ensure a minimum of 100 pairwise comparisons with distance class. Averaged F_{ij} were then regressed to the natural logarithm of the distance ln(r_{ij}) to provide the regression slope (*b*). To test for SGS, and to obtain 95% confidence intervals, the regression slope was compared to that obtained following 10,000 random permutations of the spatial positions of individuals under the null hypothesis that F_{ij} and r_{ij} are uncorrelated. Standard errors and mean multilocus F_{ij} estimates within each distance class were obtained through jackknifing over loci following Sokal and Rohlf (1995).

To investigate the effect of pollen flow on the kinship structure of the seed cohort, SGS analysis was undertaken on the paternal allele of each seed, following Nakanishi *et al*. (2009) and Hampe *et al*. (2010). Briefly, the known maternal alleles were subtracted from the offspring genotypes and the paternal haplotype converted into a diploid homozygous genotype. Where both the parent and offspring were heterozygotes with the same alleles, the maternal and paternal haplotypes were converted into the corresponding heterozygote genotype. The SGS of the resultant paternal seed genotypes was analysed using the same parameters as described above, up to a maximum distance of 600m.

To compare the extent of SGS among the adult and sapling generations the *Sp* statistic was used (Vekemans and Hardy 2004), as it accounts for differences in SGS due to variation in sampling schemes. The *Sp* statistic was calculated as $-b / (1 - F_{(1)})$, where *b* is the regression slope of F_{ij} on the natural logarithm of the distance classes, and $F_{(1)}$ is the mean F_{ij} between individuals belonging to the first distance interval. Thus *Sp* considers average kinship across individuals relative to the extent of the decrease in *F* across distance intervals. The standard error of *b* is given as an estimate of the variability of *Sp*, calculated by jackknifing over loci (Hardy *et al*. 2006).

2.4 Results

2.4.1 DNA extraction and microsatellite analysis

A total of 714 adults and 173 sapling leaf samples were successfully genotyped at 12 microsatellite loci (Table 2.2). Across all four populations missing data occurred in seven of the 12 loci, with most missing data occurring in locus Ag14 (7.5%) and loci Ag25 and Ag27 (1.2%) with all other loci showing <1% missing data. A total of 416 seeds were

genotyped, with an average of 10 seeds from each parent tree (range 2 – 14). As with the tree samples, most missing data in the seed cohort occurred at loci Ag14 (5.3%) and loci Ag27 (3.8%); all other loci contained zero or <1% missing data. All loci pairs, within each population, across all four populations, and within the seed cohort, were found to be in linkage equilibrium.

Table 2.2 Number of adults and saplings within each study population successfully genotyped, and the number of individual adults and saplings retained for subsequent analysis following exclusion of repeat individuals identified as clones.

Population	# Genotyped adults	# Genotyped saplings	# Individual adults	# Individual saplings
Tom	187	39	165	37
Mou	127	70	108	66
Bal	226	45	222	45
Ris	174	19	158	19
All 4 sites	714	173	653	167

2.4.2 Genotyping error

Repeat genotyping of samples revealed low levels of allele and genotype mismatch errors. The mean mismatch error rate per allele varied within each population but was consistently <1% within each population and within the seed cohort (Table 2.3). The mean mismatch genotype error rate was higher and variable between populations, but low across all four populations at 2.50% and 4.05% in the seed cohort (Table 2.3). Null alleles were consistently revealed at locus Ag14 in every population as well as the seed cohort in both Micro-Checker and Cervus analyses. Other null alleles were identified in each population but with no consistency between populations or between analysis methods. Comparison of the parent and seed cohort data in Cervus identified 47 seeds that presented a mismatch with their mother at one or more loci (23 mismatches at locus

Ag14, 11 at locus Ag01, six at locus Ag27, five at locus Ag05, and one at loci Ag09 and Ag35). Comparing the parent and seed cohort data also revealed null alleles in loci Ag01, Ag05, Ag14 and Ag27and high error rates for loci Ag14 (7.1%) and Ag27 (12.0%), with all other loci presenting <5% error (range 0.0% - 3.2%).

All subsequent analysis is based on accepting the error checking results which were consistently positive between different analysis methods, as recommended by Dąbrowski *et al.* (2014). Therefore, locus Ag14 was excluded from subsequent analysis (as previously reported in Lepais and Bacles 2011). Further to this, locus Ag27 was excluded from the parentage analysis following Cervus guidance regarding loci revealing >5% genotyping error (Kalinowski *et al*. 2007). Removal of locus Ag14 and locus Ag27, for parentage analysis, resulted in lower allele and genotype mismatch errors within each population data although not within the seed cohort (Table 2.3).

Table 2.3 Mean mismatch error rates per allele and per genotype based on repeat amplification and genotyping of individuals within each population. Results are shown for all 12 amplified loci, and for 11 loci and 10 loci following exclusion of one locus due to the presence of null alleles (loci Ag14) and another locus >5% error rate (Ag27).

		12 loci		11 loci (excluding Ag14)		10 loci (excluding Ag14 & Ag27)	
Population	**Repeat samples (%)**	**Allele error rate (%)**	**Genotype error rate (%)**	**Allele error rate (%)**	**Genotype error rate (%)**	**Allele error rate (%)**	**Genotype error rate (%)**
Tom	38	0.18	2.63	0.20	2.63	0.00	0.00
Mou	11	0.00	0.00	0.00	0.00	0.00	0.00
Bal	18	0.00	0.00	0.00	0.00	0.00	0.00
RIs	10	0.96	11.11	0.00	0.00	0.00	0.00
All 4 sites	20	0.19	2.50	0.09	1.25	0.00	0.00
Seed cohort	18	0.18	4.05	0.18	4.05	0.18	4.05

2.4.3 Genetic diversity of adult and sapling generations

Across all four populations, 65 genotypes were found to be shared by two or more individual trees. Shared genotypes only occurred within study populations *i.e.* no genotypes were shared between populations. All subsequent analysis was therefore based on a total of 653 adults and 167 saplings (Table 2.2). Parentage analysis of the saplings confirmed that no sapling shared the same maternal parent.

Across all four populations a total of 98 different alleles were revealed across the 11 loci, ranging from 3 to 15 alleles per locus. Genetic diversity measures were very similar between the sapling and adult generations with expected heterozygosity and the inbreeding coefficient being almost identical both between generation and between populations (Table 2.4). Although allelic richness varied between generations, and between sites, no significant differences between generations were detected (Table 2.4). The percentage of rare alleles, within and among each site, was consistently lower in the sapling generation compared to the adult generation. This difference was significant in the Richard's Island population (V = 4, *p* value = 0.002) (Figure 2.2).

Four seeds were found to have two alleles not sampled within the parent cohort: one allele at loci Ag10 occurred in one seed and one allele at loci Ag35 occurred in three seeds.

Table 2.4 Multilocus genetic diversity between the adult and sapling generations with each population, and across all four populations: N, number of samples; N_A, mean number alleles per locus; A_R, allelic richness; H_E, gene diversity, corrected for sample size; and F_{IS}, inbreeding coefficient with significance based on 10,000 permutations as implemented in SGAGeDi.

	Tomdachoille		Moulinearn		Ballinluig		Richard's Is.		All 4 sites	
	Adult	**Sapling**	**Adult**	**Sapling**	**Adult**	**Sapling**	**Adult**	**Sapling**	**Adult**	**Sapling**
N	165	37	108	66	222	45	158	19	653	167
N_A	8.27	6.91	8.18	7.64	8.55	7.09	8.09	5.73	8.82	8.27
A_R	5.85	5.96	6.07	5.87	5.87	5.81	5.88	5.65	8.24	8.21
H_E	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.62	0.64	0.64
F_{IS}	0.01	0.02	-0.01	-0.01	0.02	0.03	0.01	-0.01	0.01	0.01

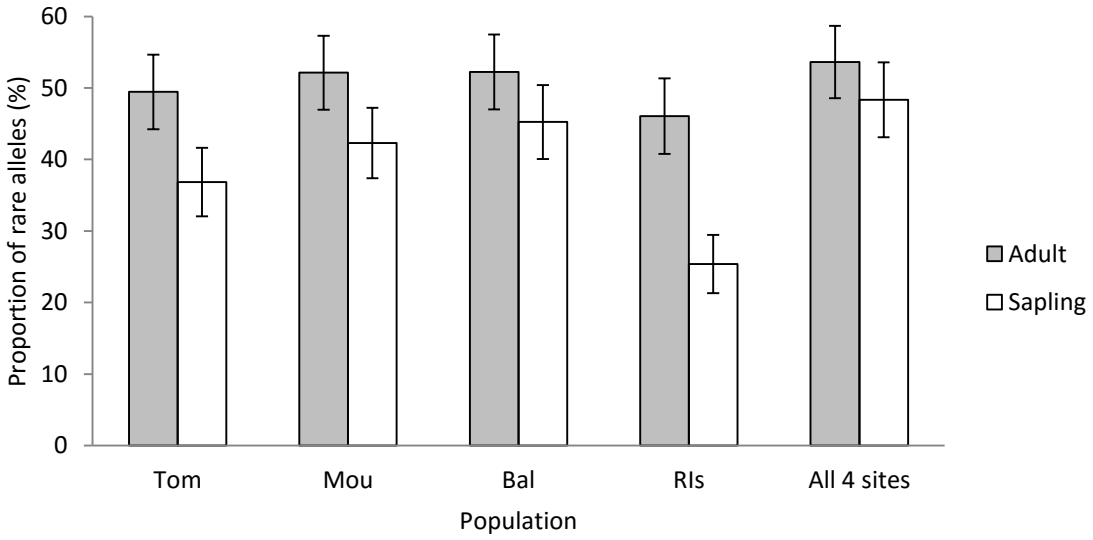

Figure 2.2 The percentage of rare alleles (<0.05 frequency) ± standard errors between the adult and sapling generations within each population, and across all four populations.

2.4.4 Maternity analysis

No saplings were found to share the same maternal parent, at either the 95% or the 80% confidence level. Of the 167 saplings, 42 (25%) had at least one potential parent sampled from across all four populations. Of these 42 saplings 16 were only assigned one parent and were therefore excluded from further analysis as it was unknown whether it was the maternal or paternal parent. Consequently, a total of 26 maternal parent assignments were made (16% of the total sapling generation), seven with 95% confidence and 19 with 80% confidence. No mismatched alleles between sapling and maternal parent were

recorded. All subsequent analyses are based on the 26 maternal parent assignments assuming ≥80% confidence.

Effective seed dispersal distances were recorded between 0 m – 2.5 km (mean 641 ±67 m). The mean distance between the maternal parent and locally sired sapling (*i.e.* within the same population) was 42 ±9 m (range 0 m – 98 m). In comparison, seed dispersal originating from a maternal parent in a neighbouring population had a mean dispersal distance of 1,239 ±218 m (range 300 m – 2.5 km) (Figure 2.3).

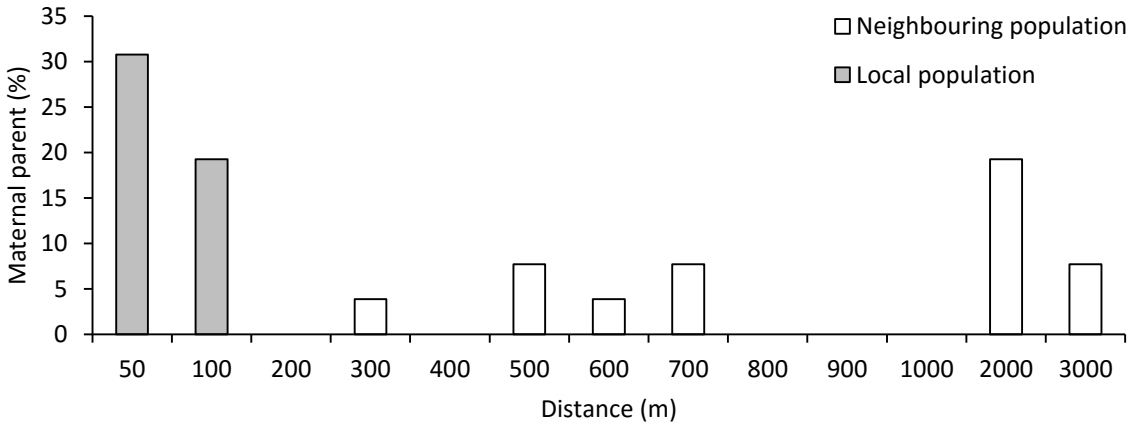

Figure 2.3 The percentage of local seed dispersal events compared to seed dispersal from a neighbouring (upstream) population. Based on the 26 saplings assigned a maternal parent with ≥80% confidence following parentage analysis of a total of 167 saplings in Cervus (Kalinowski *et al*. 2007).

Across all four sites, 50% of maternal parents were local to their sired sapling and 50% of maternal parents were located in a neighbouring population to that of their respective sapling offspring. In the Tomdachoille population all seed dispersal was local as it was the most upstream location sampled in this study. However, in the three other study populations most (between 57% - 75%) seed dispersal was found to have originated in a neighbouring population (Figure 2.4).

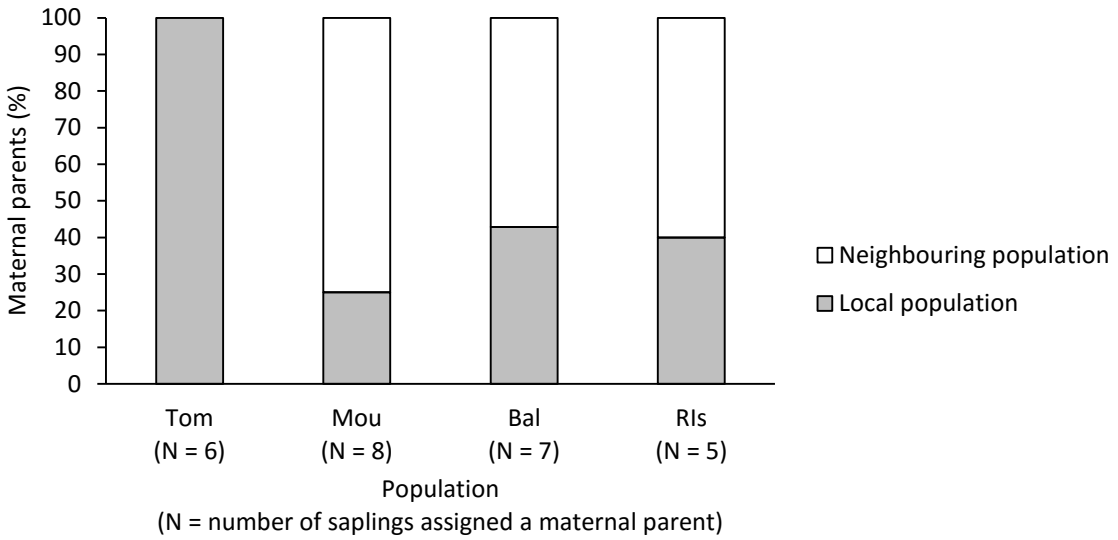

Figure 2.4 Proportion of local seed dispersal versus seed dispersal from a neighbouring population for each study population. Based on the 42 saplings assigned a maternal parent with ≥80% confidence following parentage analysis of a total of 167 saplings in Cervus (Kalinowski *et al*. 2007).

Across all four populations the analysis demonstrated a leptokurtic, fat tailed dispersal curve (Figure 2.5a). However, the pattern of seed dispersal originating from local and neighbouring populations was quite different. Locally, all seed dispersal events occurred within 100 m, with most (62%) occurring within 50 m of the maternal parent (Figure 2.5b). In comparison, although seed dispersal originating from a neighbouring population did not span the full distance of the study area, no discernible pattern in dispersal distances was observed (Figure 2.5a).

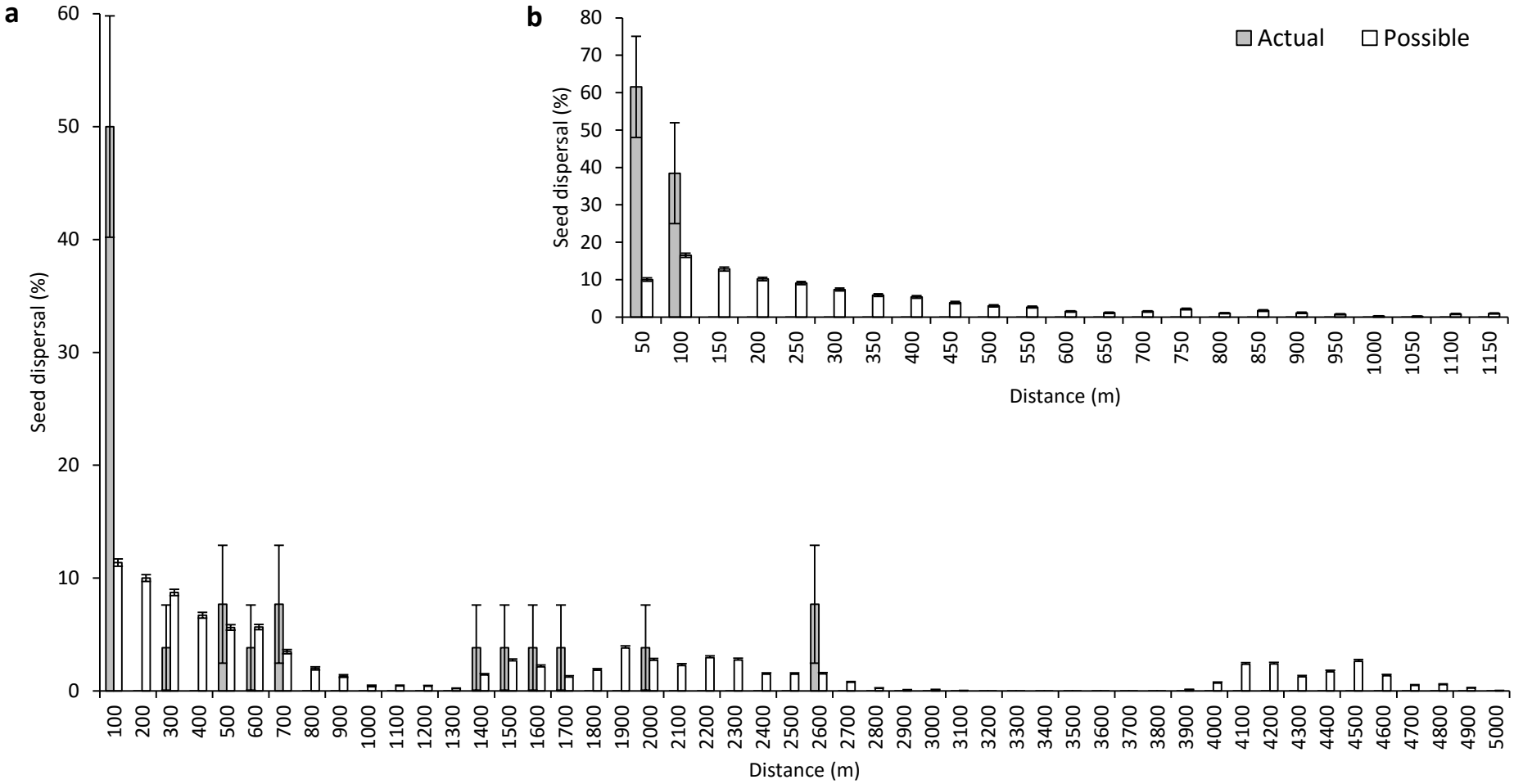

Figure 2.5 Comparison of possible and observed seed dispersal distances with ±SE within and among riparian *Alnus glutinosa* populations of the River Tummel. Seed dispersal distances assumed under random mating (white) against actual (grey) mating events between a) the 653 possible parents and 167 saplings sampled and b) local pairwise adult-sapling distances only (*i.e.* within population adult-sapling pairs). Results based on 26 maternal assignments identified in parentage analysis implemented in Cervus (Kalinowski *et al*. 2007).

2.4.5 Paternity analysis

Depending on the proportion of candidate parents sampled, the ML paternity analysis assigned between 3 and 65 fathers with 95% confidence and between 19 and 394 fathers with 80% confidence (Figure 2.6a). The proportion of assigned paternity was not consistent with the predictions of the simulations (Figure 2.6a). The number of observed paternity assignments was typically considerably less than expected although when assuming ≤10% of parents had been sampled this difference was relatively low (Figure 2.6a). With 95% confidence, as the proportion of candidate parents sampled decreased the number of assigned fathers steadily decreased (Figure 2.6a). With 80% confidence, the number of assigned fathers was initially almost identical to that expected from the simulation when all parents were assumed to have been sampled. However, a sharp decline was observed once it was assumed that not all candidate parents were sampled (Figure 2.6a).

In comparison, the exclusion paternity analysis assigned between 3 and 83 fathers with 95% confidence and between 23 and 110 fathers with 80% confidence, not dissimilar to the number of ML assignments identified (Figure 2.6b). As with the ML paternity analyses, assigned and expected paternity were not consistent with each other and both declined as the proportion of candidate parents sampled decreased (Figure 2.6b). At both 80% and 95% levels of confidence, the number of assigned fathers steadily decreased as the proportion of candidate parents sampled decreased (Figure 2.6b). A similar number of paternal parent assignments were made by both the ML and exclusion paternity analysis when ≤25% of candidate parents were assumed to have been sampled (Figure 2.6).

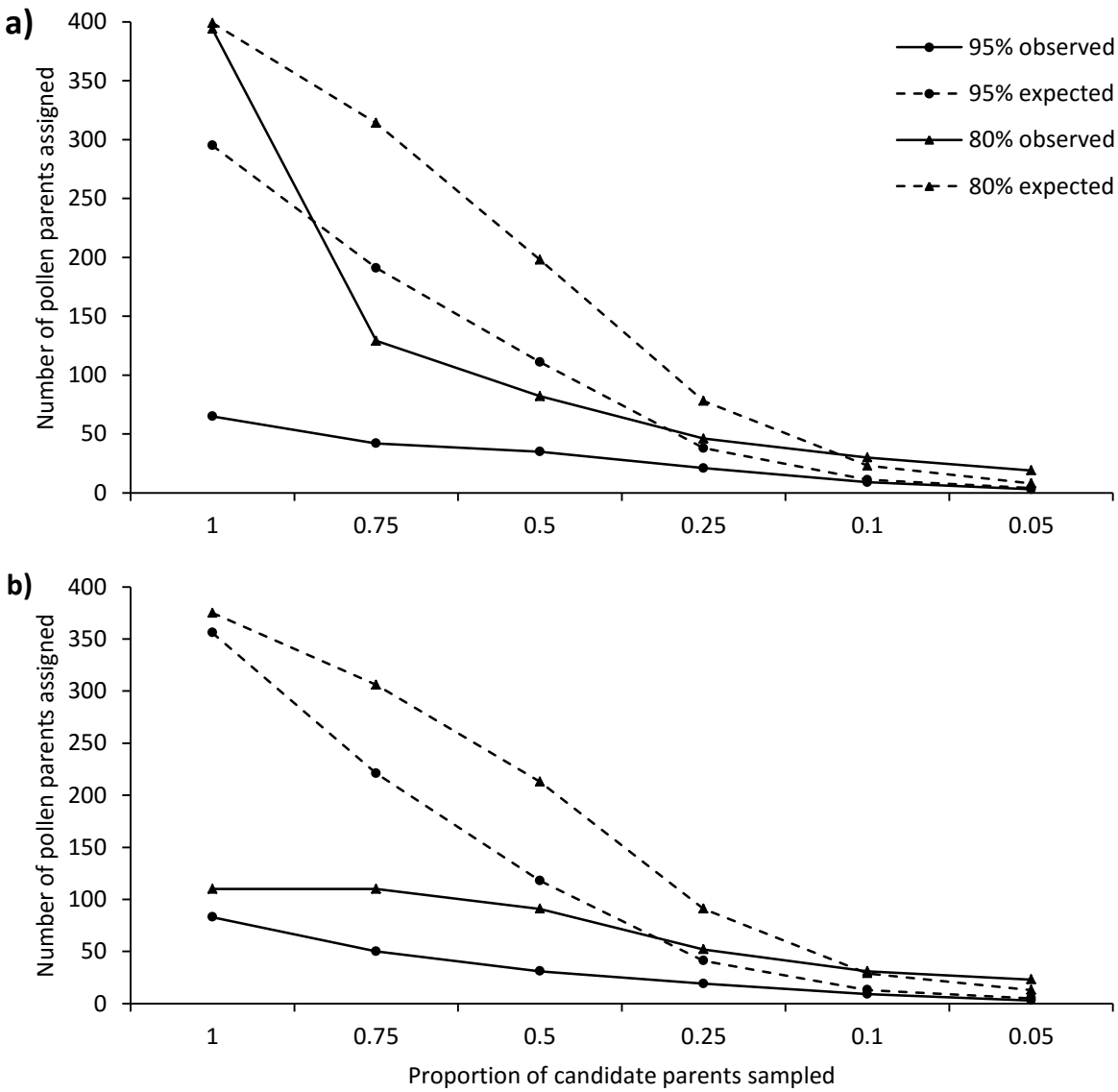

Figure 2.6 Results obtained from (a) maximum likelihood paternity analysis, and (b) exclusion paternity analysis assuming different proportions of candidate parents sampled. The number of assigned (solid line) and expected (dashed line) fathers obtained with 80% confidence (triangle) and 95% (circle) is shown. The results are based on analyses implemented in Cervus (Kalinowski *et al*. 2007).

Overall, it is clear that both the ML and exclusion paternity analyses were sensitive to the (unknown) proportion of parents sampled. Further to this, the disparity between the number of assignments observed and the number of assignments expected highlights the fact that not all potential parents were sampled. As a consequence the non-exhaustive sampling undertaken here has weakened the power of the paternity analyses. Importantly however, whilst the incorrect specification of the proportion of candidate parents sampled

will result in incorrect estimates of confidence in assignment, the rank order of compatible parents is not affected (Jones *et al*. 2010). On this basis, the paternity analysis results obtained were used to assess *A. glutinosa* pollen dispersal distances.

2.4.6 Contemporary pollen-mediated gene flow

Clearly, assessment of *A. glutinosa* pollen dispersal distances will vary depending on which paternity analysis data set assessment is based on. Here, assessment of pollen dispersal is based on the exclusion analysis obtained with 80% confidence and assuming 25% of candidate parents were sampled. Two steps were taken in selecting this data set to assess pollen dispersal distances. Firstly, the data set based on 25% of candidate parents was selected because the number of assignments made was closer to, but not greater than, the number of assignments expected (Figure 2.6b). Secondly, based on Oddou-Muratorio *et al*. (2003), the presence of type I and type II errors was considered, type I error being the incorrect identification of pollen immigration where actually the father has been sampled and, type II errors being the incorrect assignment of true pollen immigration to a sampled father. To assess gene flow Oddou-Muratorio *et al*. (2003) propose minimising type II error by using the 80% confidence level, and to assume no scoring error, as in the exclusion analysis implemented here. Finally, although the paternity analyses do not account for the majority of pollination events, and the selected data set undoubtedly contains type I and type II errors, the results from each analysis type produce consistently similar patterns of pollen dispersal distances (data not presented).

Pollen parents were located in all four populations, with pollen dispersal occurring between 1 m – 4.31 km (mean 549 ±131 m) (Figure 2.7a). The mean pollen dispersal

distance between the maternal tree and pollen parents located in a neighbouring population was 1.85 ±0.37 km (range 205 m – 4.31 km). The pollen dispersal distance between the maternal tree and pollen parents located in the same population was 160 ±25 m (range 1 m – 587 m, Figure 2.7b). Most pollen parents (77%) were from the same population as the seed maternal trees, a level higher than expected (25%) under random mating. Most (54%) pollen dispersal occurred within 200 m of the seed (37% within 100 m), with an additional 12% of pollen dispersal occurring with 200 m – 300 m of the maternal seed tree (Figure 2.7a).

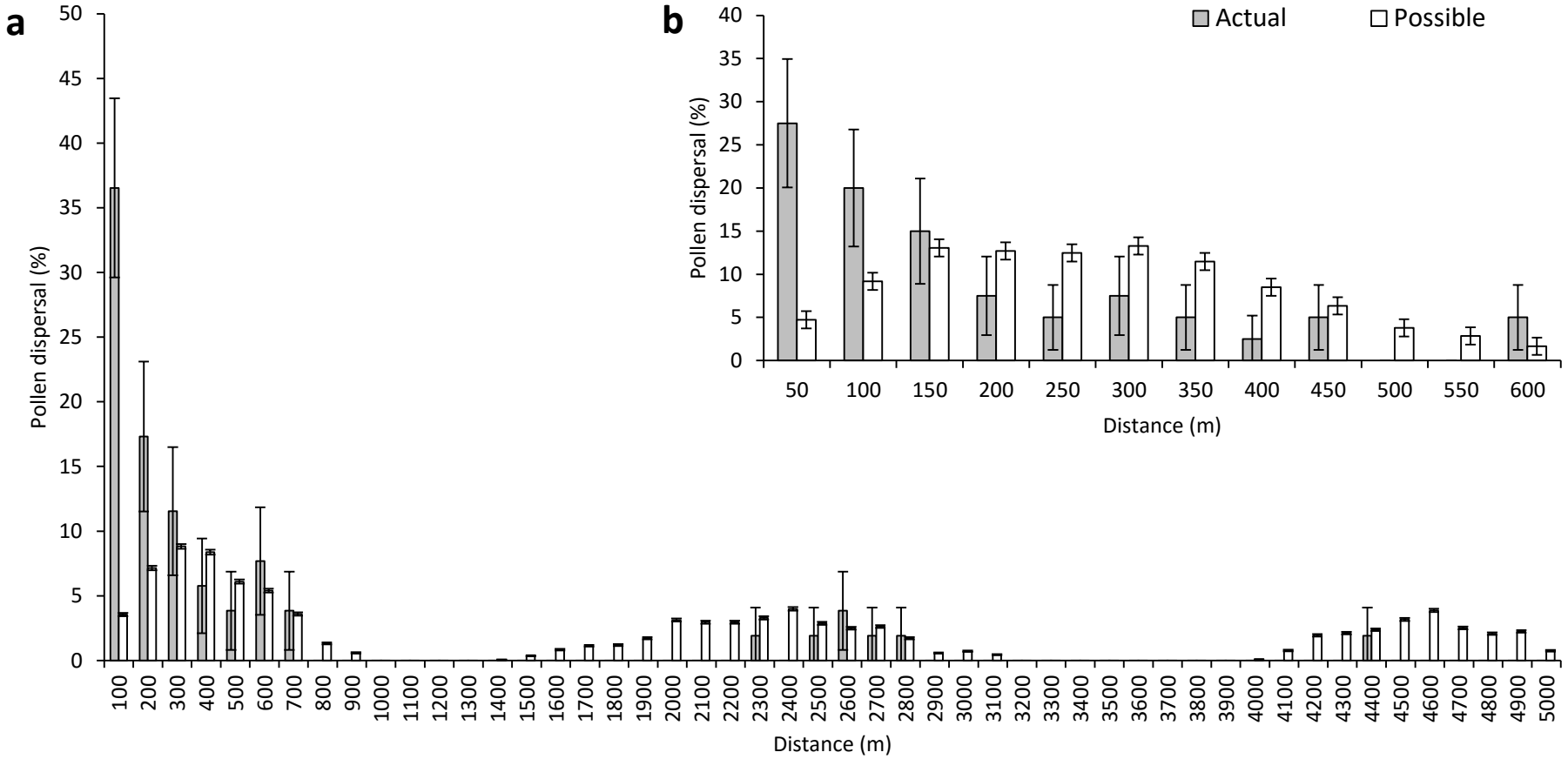

Figure 2.7 Comparison of possible and actual pollen dispersal distances with ±SE within and among riparian *Alnus glutinosa* populations of the River Tummel. Pollen dispersal distances assumed under random mating (white) against actual (grey) mating events between a) the 820 possible pollen parents and 399 seeds sampled and b) local pairwise maternal-paternal parent distances only (*i.e.* within population maternal-pateranl pairs). Results based on 52 paternal assignments identified in parentage analysis implemented in Cervus (Kalinowski *et al*. 2007).

2.4.7 Spatial genetic structure

No SGS was observed in any of the study populations although positive SGS was observed in the first distance class (0 m – 20 m) in the Tomdachoille and Richard's Island populations (Figure 2.8). Across all four populations significant positive SGS in the adult generation was observed in the smallest distance class of 20 m and at the 40 m, 80 m and 500 m distance classes (Figure 2.9a). Additional analysis of the adult dataset, based on smaller distance classes, revealed significant SGS in the 0 – 5 m distance class (P = 0.0000) and in the 10 – 15 m distance class (P = 0.0145), but not in the 5 – 10 m distance class (data not shown). Although not significant, F_{ij} values in the sapling generation were negative up to 60 m before generally levelling out around zero with the exception of F_{ij} values peaking just outside the negative 95% confidence interval at 700 m and outside the positive 95% confidence interval at 900 m (Figure 2.9b). The between-generation SGS obtained by analysing pairs including one adult tree and one sapling revealed a similar pattern to that of the adult generation, with significant SGS identified at the 20 m and 80 m distance classes (Figure 2.9c). The paternal seed allele revealed significant SGS at 0 – 20 m, followed by a reasonably smooth decline with F_{ij} values consistently below zero beyond 60 m (Figure 2.9d). Use of the *Sp* statistic as a measure of SGS intensity showed slightly more intense SGS occurring in the adult generation than that in the paired adult-sapling analysis (Table 2.5). The most intense SGS was seen in the paternal allele of the seed cohort (Table 2.5).

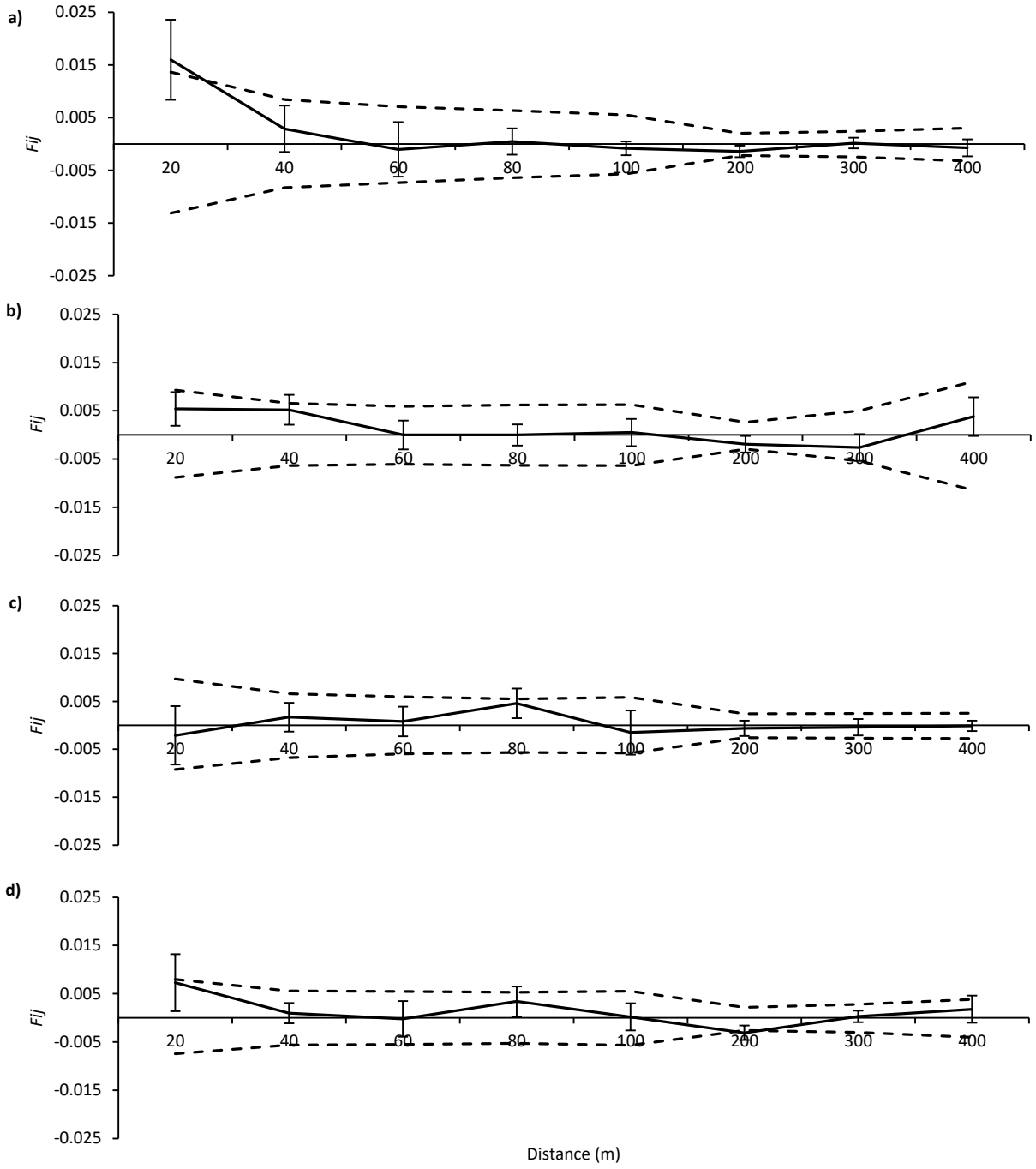

Figure 2.8 Spatial genetic structure present in the four study populations a) Tomdachoille; b) Moulinearn; c) Ballinluig; and d) Richard's Island. Relatedness between individuals is based on Nason's F_{ij} kinship coefficient, implemented in SPAGeDi. Broken lines represent the 95% upper and lower confidence intervals, based on 10,000 simulations.

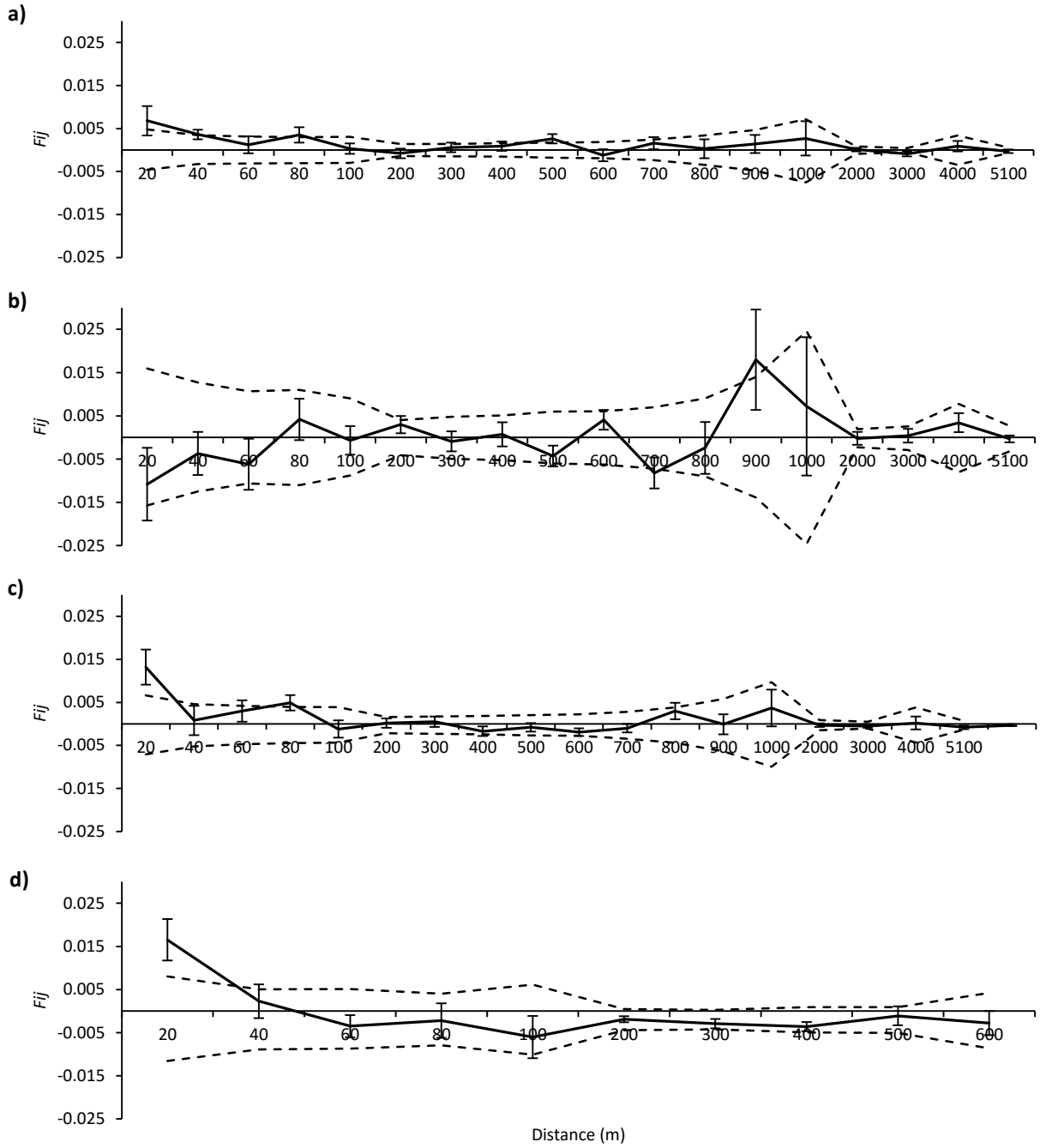

Figure 2.9 Spatial genetic structure present in a) the adult generation; b) the sapling generation; c) the paired adult and sapling generations; and d) the paternal alleles of seed cohort. Relatedness between individuals is based on Nason's F_{ij} kinship coefficient, implemented in SPAGeDi. Broken lines represent the 95% upper and lower confidence intervals, based on 10,000 simulations.

Table 2.5 SGS parameters showing the kinship coefficient for the first distance class of 0 – 15 m ($F_{(1)}$) and the rate of decrease of pairwise kinship with distance (*Sp*) ± standard error (SE). Significant *p* values are shown as $^{**}p < 0.01$, $^{***}p < 0.001$, based on 1-sided values, as calculated in SPAGeDi.

Dataset	Generation	$F_{(1)}$	*Sp* ± SE
All 4 sites	Adult	0.0068**	0.0006 ± 0.0002
All 4 sites	Adult & sapling	0.0132***	0.0004 ± 0.0002
Paternal allele	Seed	0.0165***	0.0022 ± 0.0007

2.5 Discussion

Genetic diversity between generations

Heterozygosity and the inbreeding coefficient between the sapling and adult generations were practically identical, both within and amongst study populations. However, compared to the adult generation, the sapling generation consistently held fewer rare alleles, with a significant difference occurring in the Richard's Island population (V = 4, *p* value = 0.002). Fewer rare alleles in the younger generation may be attributable to genetic drift, as rare alleles are the first to be lost under drift (Lande 1988, Young *et al*. 1996). Furthermore, in long-lived organisms such as trees, it can take several generations for the impacts of genetic drift to become apparent in measures of heterozygosity and inbreeding coefficient (Bacles and Jump 2011). Nevertheless, of the four study populations, only the Richard's Island population showed significantly less rare alleles in the sapling generation. Furthermore, allelic richness, a suitable measure for assessing short-term diversity loss (Lowe *et al*. 2005), remained very similar between generations, both within and among populations. Taking these genetic measures together, the observed lack of difference between generations suggests that high outcrossing rates and high gene flow, typical of tree species (Petit and Hampe2006), is maintaining similar levels of genetic diversity between *A. glutinosa* generations. Further, it is feasible that the difference in rare alleles is a consequence of the greater variation in age of the adult generation compared to the sapling generation therefore the adult generation has acquired more rare alleles over a longer period of time. Finally, aside from revealing significantly positive inbreeding coefficient measures, reflecting pseudo replication

outlined in Section 2.3.7, measures of genetic diversity in the seed cohort revealed no significant differences with either the sapling or adult generations (data not shown).

Maternity analysis of saplings

Despite the low number (15%) of maternal assignments obtained through parent-pair analysis, an important initial insight into the effective dispersal of *A. glutinosa* seed was obtained. It is likely that the considerable number of maternal parents that could not be assigned is due to the fact that not all trees were sampled, whether from within each local population or from other, unsampled populations. Clearly, given the landscape-scale occurrence of *A. glutinosa*, it was not possible to sample all possible parents at the scale studied, thus the results presented here are based on sampling feasibility. The maternal parent assignments made indicated a high level of genetic exchange between populations. In fact, when upstream *A. glutinosa* populations were present, seeds were more likely to have originated from a neighbouring population than from the same population as the sapling (Figure 2.4). The high level of between-population seed dispersal is higher than that reported elsewhere, with most studies reporting seed immigration rates of less than 20% (reviewed in Ashley 2010). However, approximately 50% seed immigration is reported between isolated fragments of *Fraxinus excelsior* forest (Bacles *et al.* 2006). Importantly, Bacles *et al.* (2006) highlights the potential of landscape features, including rivers, for enabling the observed (long-distance) high rate of seed dispersal between forest fragments. Indeed, studies using tracers to simulate plant dispersal along the River Tummel revealed that all tracers placed in backwaters and the main channel will move from their original position, under both base-flow and flood conditions (Keruzoré 2012). Keruzoré (2012) shows backwaters are an important river-

landscape feature, both for within-site movement of plant material and for acting as a source habitat of plant material to the main channel. In this study, although most sampled trees were located away from the main river channel, backwater channels were present at all sites, maintaining a connection to the main river channel that would enable both within-population dispersal and downstream between-population dispersal of seed.

At the local, within-population scale, both wind and hydrochorous seed dispersal may influence the movement of seed. In comparison to the between-population dispersal described, local within-population seed dispersal was not random with all local seed dispersal occurring within 100m of the maternal parent. Most local seed dispersal (62%) occurred within 50 m of the maternal parent, consistent with wind-dispersal distances of up to 60 m described by McVean (1955). Two reasons to explain local seed dispersal occurring over distances greater than 60 m are suggested. Firstly, McVean's (1955) expected wind-dispersal distance of 60 m is based on field observations of saplings i.e. no genetic evidence was available to confirm kinship. Secondly, as previously discussed, 100% of plant material occurring in backwaters is found to move (Keruzoré 2012), hence it seems feasible that within-population *A. glutinosa* seed movement may be facilitated by both wind and water dispersal. The mean local seed dispersal distance of 42 ±31 m is remarkably similar to the within-backwater dispersal distances recorded by Keruzoré (2012). Based in the same study system, and using independent methods based on tracers to simulate plant propagule dispersal, Keruzoré (2012) recorded mean within-backwater dispersal distances of 45 m under mean base-flow conditions, although under flood-conditions the mean dispersal distance increased to 160m. Finally, it is acknowledged that assessment of seed dispersal distances was partly based on excluding

maternal parent assignments located more than 60 m away from the relevant sapling. Revisiting the parentage analysis shows four local, within population maternal assignments were excluded on this basis. Incorporating these additional four maternal parents does not influence the overall pattern of seed dispersal observed. However, it does indicate that local, within-population seed dispersal may occur over distances up to 260 m although only two of the four maternal assignments indicate seed dispersal occurring over 100 m (data not shown).

The overall mean dispersal distance of *A. glutinosa* seed was 641 ±67 m, 15 times greater than local dispersal distances. With seed dispersal occurring over a range of distances the empirical data observed here reflects previous findings where seed dispersal is frequently found to fit a leptokurtic, fat-tailed dispersal kernel, indicative of long-distance dispersal (Nathan 2001). Only one other comparable study investigating the hydrochorous seed dispersal of riparian tree species was found. In populations of *Populus nigra* L., a pioneer riparian tree with seed dispersal occurring via wind and water, located on the Morava River, Czech Republic, seed dispersal was found to occur over distances of up to 370 m, spanning the full length of the study site (Pospíškova and Šálková 2006). Again, the between-population seed dispersal distances (i.e. within the main river channel) identified in this study are similar to the within-main channel dispersal distances found in Keruzoré's (2012) dispersal simulations. Keruzoré (2012) examined dispersal in the main river channel (River Tummel and, further downstream, the River Tay) by tracking dispersal originating from either the main channel, or from a backwater. Under base-flow conditions, and when tracer dispersal originated from a backwater, the mean dispersal distance was 1.4 km, increasing to 1.7 km when tracer dispersal originated in the main

channel (Keruzoré 2012). These simulated dispersal distances are very similar to the mean, between-population seed dispersal distance of 1.2 km observed here in *A. glutinosa*, particularly when considering the smaller span of river distance studied here (5 km, compared to 33 km). Still, under flood flow conditions, simulated dispersal distances increased to 3.5 km when material originated from a backwater and 12.2 km for material originating in the main river channel itself (Keruzoré 2012).

Although not hydrochorously dispersed, parentage analysis of two Northern Irish *Fraxinus excelsior* populations revealed a mean wind dispersal distance of seed to be 42 m (Beatty *et al.* 2015), identical to the mean local dispersal distance of *A. glutinosa* identified in this study. Beatty *et al.'s* (2015) use of both nuclear and (maternally inherited) chloroplast markers provides an unambiguous insight into seed dispersal distances, including the occurrence of dispersal over hundreds of metres. Of note, Beatty *et al.'s* (2015) plot of effective seed dispersal distances revealed a Janzen-Connell recruitment process which, alongside a lack of evidence for SGS, is considered indicative of density-dependent mortality close to the mother plant. The difference in the shape of *A. glutinosa* and *Fraxinus excelsior* seed dispersal curves may be a consequence of the differing dispersal mechanisms although further research would be required to ascertain this.

Identifying seed parents is challenging, and seed dispersal is a complicated and multifaceted process (Ashley 2010). The assumptions made in this study regarding the proportion of parents sampled, and how maternal and paternal parents were distinguished, whilst enabling a maternity analysis to be undertaken, may also bias the results obtained. Although the approach taken may have caused an increase in Type I

error (where a seed matches a candidate parent by chance, when the true parent has not been sampled) (Ashley 2010), the insight gained into seed dispersal distances is consistent with the results of the SGS analysis, and remarkably similar to dispersal distances identified by Keruzoré (2010) using independent methods. In addition, as described earlier, seed dispersal patterns outlined here represent effective seed dispersal, reflecting both seed dispersal *per se* and establishment processes. Despite concerns around the low power of the maternal parentage, analysis undertaken corroborates the leptokurtic dispersal kernel expected for seed dispersal. Whilst long distance seed dispersal was not observed beyond 2.5 km, it is clear that hydrochorous seed dispersal is at least as important as wind mediated seed dispersal and that hydrochorous seed dispersal plays an important role in maintaining genetic connectivity between riparian *A. glutinosa* populations.

Paternity analysis of seeds

That paternity analyses may be sensitive to the proportion of parents sampled was clearly demonstrated here, in both the ML and exclusion analyses undertaken. Nevertheless, the pattern of pollen dispersal distances is consistent across each implemented analysis. The analyses revealed a fat-tailed, leptokurtic pollen dispersal curve, with most pollen dispersal occurring over relatively short distances (37% within 100 m), although 33% of pollen parents were located in neighbouring populations. Pollen dispersal occurred up to distances of 4.31 km, confirming riparian populations of *A. glutinosa* are genetically connected to downstream populations via pollen dispersal. Hence, despite concerns around genotyping error and uncertainty around partial sampling of the reproductive

population, the pattern of pollen dispersal distances is consistent with that reported in studies of other tree species.

As reported elsewhere (Ashley 2010), *A. glutinosa* pollen dispersal was not random, displaying a highly skewed contribution of pollen sources with most pollen dispersal occurring over short distances. This finding concurs with previous studies showing, in wind-pollinated trees, most pollination events result from near-neighbour pollen donors (Ashley 2010). Although no comparable *Alnus* studies were found, examples in other wind-pollinated tree studies revealed over 80% of *Fraxinus excelsior* pollen dispersal occurring within 100 m of the maternal tree (Bacles and Ennos 2008); 54% of *Quercus macrocarpa* pollinations occurring within 70 m (Dow and Ashley 1998); and most pollination events of the riparian tree *Cercidiphyllum japonicum* were found to occur within 50 m of the maternal tree (Sato *et al*. 2005). Nevertheless, all of these studies also reported pollen dispersal occurring across the distance of the study area, as reported here, as well as significant pollen immigration.

Whilst this study is thought to be the first to investigate *A. glutinosa* pollen dispersal, other studies clearly indicate the presence of pollen immigration over long wind-dispersed distances (Ashley 2010). In *F. excelsior* pollen dispersal occurred up to distances of 2.9 km, with pollen immigration accounting for over 40% of effective pollination (Bacles and Ennos 2008) and in *Populus trichocarpa* a mean dispersal distance of 7.6 km was reported with approximately a third of observed pollinations resulting from pollen immigration travelling over 16 km (Slavov *et al*. 2009). In fact, extensive pollen dispersal is

expected in wind-pollinated species, such that populations of wind-pollinated trees are effectively panmictic across large spatial scales (Ashley 2010).

Spatial genetic structure

This study demonstrated that SGS did not remain constant over time, with different life stages of A*. glutinosa* revealing varying patterns of fine-scale SGS. Although the paternal alleles of the seed cohort revealed the most intense SGS, no evidence of SGS was observed within the sapling generation. However, although weak, significant SGS was observed in the adult generation. Consequently, the hypothesis that SGS in riparian *A. glutinosa* populations would decline over time, and therefore higher SGS would be observed in the sapling generation than the adult generation, was not supported here.

The genetic structure of the paternal allele may represent correlated mating, the generation of full siblings in the aggregates of half siblings sampled from the same maternal tree. In *Quercus salicina*, significant SGS observed in the paternal alleles of seed is thought to be a consequence of correlated mating, mediated by limited pollen dispersal (Nakanishi *et al*. 2009). Here, the seed sample size obtained is not sufficient to investigate the level of correlated mating in *A. glutinosa*. However, the assessment of pollen dispersal distances clearly showed that pollen dispersal was not random, with significantly more pollen dispersal than expected occurring over relatively short distances (Figure 2.7). It is therefore likely that this non-random pollen dispersal contributes to the weak, but significant, spatial genetic structure observed in the paternal allele of the seed cohort. Clearly, non-dispersed *A. glutinosa* seed will maintain significant SGS due to the presence of half siblings and the results obtained here suggest that SGS within the

paternal allele will further contribute to levels of SGS within any seed cohort. Nevertheless, here, no evidence of SGS was observed in the sapling generation, indicating a loss of genetic structure in the transition from seed to sapling.

A number of processes could explain the lack of SGS observed in the sapling generation. Few other empirical studies report a loss and / or absence of SGS in the juvenile stage when SGS is simultaneously observed in the seed and / or adult stage. However, in the perennial herb *Trillium grandiflorum*, a lack of SGS in the juvenile cohort has been assigned to a combination of random mating, high mortality of seedlings, and moderate seed dispersal distances (Kalisz *et al*. 2001). In the Neotropical tree *Jacaranda copaia* a loss of SGS between the seed and sapling generation was attributed to a combination of low recruitment, density-dependent seedling mortality, and spatial and temporal variation in seed production (Jones and Hubbell 2006).

A combination of low seedling recruitment and high seedling mortality, as well as long-distance dispersal of both seed and pollen are considered contributory factors to the loss of SGS between the seed and sapling generations observed here. Successful establishment of *A. glutinosa* seedlings requires at least 20% of the above canopy light and a high level of moisture, and consequently no natural regeneration tends to occur within woodlands, except where there is a canopy opening >1,000 m^2 (Claessens *et al*. 2010). Indeed, riparian plant communities may frequently be recruitment limited (Nilsson *et al.* 2010). Field observations undertaken here confirm this. In addition to very few seedlings being observed, seedlings that did occur were located in very open, gravel areas close to the main river channel. Consequently, the habitat requirements necessary for the

successful recruitment of *A. glutinosa* seedlings may reduce relatedness in the sapling generation. In addition to this, Kalisz *et al.* (2001) demonstrated that thinning of post-dispersed *T. grandiflorum* seeds to single juvenile individuals led to a decrease in SGS. In *A. glutinosa*, concentrations of seed occur as a consequence of wind and water action (McVean 1955). Given this clump-like stranding of dispersed seeds it seems likely that a high level of mortality-driven rarefaction of seedlings will occur, weakening any pattern of SGS.

Most SGS forms as a result of limited gene dispersal (Vekemans and Hardy 2004), however when both pollen dispersal and seed dispersal are random, or when pollen dispersal is highly localised but seed dispersal is random, SGS will not develop (Kalisz *et al*. 2001). Most (57 – 75%) *A. glutinosa* seed dispersal, at least where upstream populations occurred, was shown to have originated from a neighbouring population (Figure 2.45). Thus, despite local *A. glutinosa* seed dispersal appearing limited to within 100 m of the maternal tree (Figure 2.5b), the high level of between-population seed dispersal observed should have a homogenising effect on SGS. In addition to this, although most pollination events were local, pollen dispersal occurred across the whole span of the population and, between populations dispersal distances up to 4.3 km were observed. Thus, whilst hydrochorous seed dispersal is considered a major factor in reducing SGS between the seed and sapling generations of *A. glutinosa*, it is likely that long-distance pollen dispersal also contributes to the loss of SGS.

It was expected, based on the lack of SGS in the sapling generation that no SGS would be apparent within the adult life stage of *A. glutinosa*. It was therefore with some surprise

that although weak, significant SGS was observed in the adult generation (Figure 2.9a). Possible explanations for an increase in genetic structure from the sapling to adult generation include overlapping generations, selection, historical influences, and non-equilibrium dynamics (Jones and Hubbell 2006). A development of SGS, or an increase in SGS, from juveniles to adults has been attributed to a historical bottleneck or micro-environmental selection in perennial herb *T. grandiflorum* (Kalisz *et al*. 2001); selection for mycorrhizal interactions and overlapping seed shadows in the orchid *Orchis purpurea* (Jacquemyn *et al*. 2006); life history attributes and low survival rates in early life stages in the Neotropical tree *Jacaranda copaia* (Jones and Hubbell 2006); and micro-environmental selection as well as overlapping seed shadows in *Dalbergia nigra* of the Atlantic Forest (Leite *et al.* 2014).

Historical factors offer a potential explanation for the processes underlying the increase in genetic structure from the sapling to adult generation in the *A. glutinosa* study populations. The dynamic nature of the River Tummel results in a shifting mosaic of riparian zones (Gilvear and Willby 2006). As a consequence it is likely that the study populations have, over time, developed from initially small aggregations of *A. glutinosa*. For example, *A. glutinosa* is one of only a few species that are able to undergo vegetative regeneration following uprooting (Francis *et al.* 2009). Livewood from uprooted and later deposited trees grows faster than seeds, enabling establishment before any subsequent disturbance (Francis *et al.* 2009). Further to this, an increased availability of local habitat, suitable for seedling establishment, may have meant that wind-dispersal of seed enabled local expansion of *A. glutinosa* populations. These more limited wind-dispersal distances would result in an overlap of *A. glutinosa* ages including, for example, a parent or

grandparent and its offspring growing in close proximity. In this study, the adult generation encompassed individual trees of diverse ages, potentially reflecting the above scenario, and as a result revealing a weak but significant level of SGS. Under this historical hypothesis, as succession and the spread of *A. glutinosa* occurred, a loss of habitat suitable for *A. glutinosa* seedling establishment would follow, limiting suitable habitat for future recruitment to new open, gravel areas close to the main channel where contemporary, between-population hydrochorous seed dispersal may dominate, as revealed in the maternity analysis implemented here. Thus, SGS may be observed in the adult generation but not the contemporary sapling generation.

Although not possible to investigate here, due to the use of neutral, microsatellite markers, selection for related individuals, adapted to local micro-habitat conditions may also contribute to the increase of SGS in the adult generation. Assuming gene flow is widespread, as observed here, strong micro-environmental selection could generate predictable local changes in gene frequencies as the stand matures (Epperson 1992). For example, *A. glutinosa* typically benefit from root symbiosis with *Frankia* however, *A. glutinosa* shows resistance to ineffective *Frankia* strains (Van Dijk and Sluimer-Stolk 1990) and, root nodule symbiont effectiveness has been shown to vary depending on the host genotype (Douglas 1998). The dynamic nature of the River Tummel may also lead to disturbance-mediated selection, as hypothesised by Banks *et al.* (2013). The process of selection assumes that, following the successful establishment of seedlings, subsequent selection would favour the survival of locally adapted saplings resulting in increasing levels of relatedness with increasing tree age. Future work, utilising non-neutral genetic markers, will be required to gain further insight into selection effects.

The processes underlying the increase in genetic structure from sapling to adult stages remain, however, unclear. In particular, the SGS observed between the sapling and adult generations suggest that saplings were more related to the individuals within the adult generation than they were to neighbours within the sapling generation (Figure 2.9c). This result conflicts with the key finding obtained from the maternity analysis that most seed dispersal occurs between populations, therefore genetically homogenising the seed and subsequent sapling cohort. One explanation for this apparent conflict in results is that the adult generation may be dominated by relatively young adult trees. If this were the case it is possible that the saplings and young adults are more related to each other than saplings may be to older adults, thus the relative abundance of young adults within the adult generation dataset may skew the SGS analysis. Further investigation, incorporating age cohorts within the adult generation, is described in the following chapter.

2.6 Conclusion

The lack of difference in measures of genetic diversity between the sapling and adult generations of *A. glutinosa* suggested that high outcrossing rates and high gene flow occur, typical of tree species. The high level of dispersal identified through parentage analyses confirmed that the movement of both pollen and seed contribute significantly to within and between population gene flow. Importantly, the evidence reported here suggests that most seed dispersal occurs between populations, highlighting the importance of hydrochory in maintaining between-population genetic connectivity and long-distance gene flow via seed dispersal. The parentage analyses undertaken here suggest that between-population seed dispersal is greater than between-population

pollen dispersal. However, as previously acknowledged, caution in comparing differences between pollen and seed dispersal is exercised due to the low power of both the maternity and paternity analyses. Nevertheless, although extensive local pollen and seed dispersal was observed, it is clear that between-populations pollen and seed dispersal are both widespread and occur over long distances. The lack of SGS observed in the sapling generation corroborated evidence of extensive pollen and seed dispersal. However, the processes underlying the observed increase in SGS between the sapling and adult generations remain unclear.

This study is thought to be the first to investigate seed and pollen dispersal in *A. glutinosa*. Alongside the parentage analyses, the complementary assessment of fine scale spatial genetic structure illustrates the additional insight into the processes shaping genetic diversity and structure within *A. glutinosa* populations. In addition, by classifying individuals by life stage this study illustrates that the effects of ecological and evolutionary processes are likely to vary between different life stages of *A. glutinosa*.

2.7 Acknowledgements

I thank the landowners involved, the Scottish Wildlife Trust as land manager, and Scottish Natural Heritage for statutory permissions; Antoine Keruzoré for help with site information; Alexander Flint, Armorel Flint, Roy Flint, Rachel Lintott, and Stewart Lothian for help with fieldwork; Patricia González Diáz for help with labwork; CE Timothy Paine for help with R script; and Rachael Cooper-Bohannon for help with GIS graphics.

Chapter 3

Detection of demographic and genetic structure in the riparian *Alnus glutinosa* woodlands of a dynamic river system

3.1 Abstract

Demographic and genetic structure within populations is influenced by the landscapes they occur in. The identification and description of species-environment interactions informs our understanding of ecosystem functioning and guides management and conservation. This study focuses on *Alnus glutinosa*, a key European riparian tree species, within four riparian populations of a dynamic river. Woodland inventory data was used to describe demographic structure, and dendrochronology methods implemented to develop a size-age standard for *A. glutinosa* in the study area. Genotyping of 820 individual *A. glutinosa* trees, based on 11 SSR loci, was used to examine genetic diversity and structure. Comparison between mature and young woodland revealed significant differences, notably a paucity of *A. glutinosa* seedling regeneration in mature riparian woodland. Multi-stemmed 'sprouting' growth of *A. glutinosa* indicated the occurrence of site disturbance and generalised linear mixed models identified distance to the main river channel as an explanatory factor for the temporal woodland structure identified. Genetic analyses revealed no differences in genetic diversity between age cohorts although differences in the pattern of spatial genetic structure, but not temporal genetic structure, were revealed. These results indicate interaction between riverine features and riparian vegetation. This interaction results in demographic structure within riparian *A. glutinosa* populations however minimal between-generation genetic variation or differentiation was observed, indicative of widespread gene flow.

3.2 Introduction

The occurrence of species, and the variation within and between populations, is affected by the spatial and temporal heterogeneity of the landscape they occur in. The

identification and description of the variation and structure in species distributions across different spatial and temporal scales is, therefore, central to informing our understanding of species-environment interactions (Levin 1992; Legendre 1993). Examining the strength and significance of species-environment interactions is of key importance, informing our understanding of species ecology and ecosystem functioning, as well as enabling the development of management and conservation principles (Levin 1992; de Knegt *et al.* 2010). Within terrestrial landscapes, riparian corridors are among the most dynamic and diverse of habitats (Gregory *et al.* 1991; Naiman *et al.* 1993), where riparian vegetation and hydrogeomorphic processes are intimately connected (Stoffel and Wilford 2012).

Hydrogeomorphological influences include the formation of bar and floodplain habitat mosaics suitable for plant establishment; the hydrochorous dispersal of plant propagules; and hydrological disturbance events, resulting in, for example, plant breakage or damage, loss of suitable habitat, and burial by sediment and / or debris (Francis 2006). In turn, riparian vegetation influences hydrogeomorphological process. The development of root structures physically reinforces the riparian substrate, thus improving bank and bar stability, as well as resistance to fluvial erosion. Vegetation also provides resistance to water flow, reducing the capacity of the flow to carry sediment, thus resulting in deposition of fine sediment in the proximity of the vegetation (Francis 2006). In addition, erosion / deposition, and the lateral movement of the river channel, are important to understanding patterns in riparian vegetation, with spatial zonation of vegetation often occurring along a transverse gradient perpendicular to the river channel (Naiman and Décamps 1997).

Spatial zonation of vegetation may also be reflected in the non-random distribution of genotypes, resulting in genetic differentiation across both spatial and temporal scales. Gene dispersal is often identified as the principal agent leading to the non-random distribution of genotypes, with restricted pollen and seed movement resulting in spatial genetic structure (SGS). Although reproductive biology (life form, dispersal mode, outcrossing rate) often explains the SGS observed (Vekemans and Hardy 2004), local conditions can also influence SGS (Hoban *et al.* 2014) with any or all of the hydrogeomorphic-riparian vegetation interactions described above having the potential to influence SGS. For example, in the temperate butternut tree (*Juglans cinerea* L.) of North America, habitat, aggregation, and site history were found to result in different patterns of SGS between riparian and upland sites, with site demography (based on diameter at breast height(DBH)) also found to influence population SGS (Hoban *et al.* 2014).

Most previous SGS studies do not distinguish between age classes in populations (Hossaert-McKey *et al.* 1996; Qiu *et al.* 2013), yet the successional stage of individuals can influence levels of genetic differentiation (Hossaert-McKey *et al.* 1996). Whilst the genetic structure of young trees represents the processes of recent time periods, the pattern of SGS in the older adult generation will reflect the accumulation of past and present ecological and evolutionary effects (Hossaert-McKey *et al.* 1996; Kalisz *et al.* 2001; Qie *et al.* 2013). Comparisons of genetic diversity and SGS between different age cohorts is one way to detect changes in SGS over time (Kalisz *et al.* 2001; Fuchs and Hamrick 2010), especially when identifying the age of individuals is not feasible (Hossaert-McKey *et al.* 1996). In temperate tree species tree age can be measured by counting the number of

annual rings revealed on an increment core. By cross-dating more than one core taken from the same tree, greater accuracy in tree age estimates are achieved, particularly when compared to diameter at breast height (DBH, measured at 1.3 m height) measurements which may be a poor predictor of age (Ogden 1981).

In this study the demographic and genetic structure in riparian *Alnus glutinosa* (L.) Gaertn populations of the River Tummel in the eastern Highlands of Scotland were investigated. The River Tummel, a sizeable river by UK standards, has previously been the focus for a programme of research on riverine landscape diversity (e.g. Gilvear and Winterbottom 1992; Winterbottom 2000; Parsons and Gilvear 2002; Gilvear and Willby 2006; Keruzoré *et al.* 2012) and *A. glutinosa* is a key tree species of the riparian habitat (e.g. Parsons and Gilvear 2002; Gilvear and Willby 2006). Although dynamic in nature, attempts to control flooding in the lower 10 km of the River Tummel resulted in channel confinement following the construction of embankments during the 18th and 19th centuries (Gilvear and Winterbottom 1992; Winterbottom 2000). However, subsequent abandonment of the flood embankments in 1903 has seen the river has return to a more natural state with further channel change and movement occurring. Historical maps and aerial photography show how vegetation succession has followed channel change (Winterbottom 2000) and that changes in fluvial landforms and patterns of vegetation are related to flooding events and fluvial disturbance (Parsons and Gilvear 2002). *Alnus glutinosa*, along with herbaceous species, is one of the first species to colonise bare gravel areas (Parsons and Gilvear 2002) where vegetation succession, following flood and disturbance events, may be rapid with *A. glutinosa* / *Salix* woodland establishing within 30 years (Gilvear and Willby 2006). Thus, by identifying and describing patterns of demographic and genetic

structure within *A. glutinosa* populations, a key tree species of European riparian habitat, this study aims to extend current knowledge of the hydrogeomorphological influences on riparian vegetation. Following identification of the woodland tree species present and genetic characterisation of the *A. glutinosa* populations studied, following hypotheses were tested:

I. Mature riparian woodland is comprised of larger, older *A. glutinosa* adult trees, and has a lower density of individuals than young riparian woodland.

II. In comparison with young riparian woodland, little or no *A. glutinosa* seedling regeneration occurs within mature riparian woodland.

III. Distance from the main river channel will influence the temporal structure of riparian *A. glutinosa* woodland.

IV. Differences in the pattern of spatial genetic structure between different *A. glutinosa* age cohorts will be observed, indicative of different influences occurring between generations.

V. Temporal genetic structure will reflect the demographic structure identified.

3.3 Materials and methods

3.3.1 Study species

Alnus glutinosa (L.) Gaertn, commonly known as black alder, is an important temperate riparian tree species distributed across Europe. Growing alongside the edge of rivers and standing water, *A. glutinosa* can act as a pioneer species and forms a key element of

dynamic river systems (McVean 1953). A monoecious tree, the male and female flowers form as catkins during February / March and are wind-pollinated (McVean 1955). Following fertilisation, the female flower forms as a woody cone-like fruit containing approximately 60 seeds (strictly achenes) that are dispersed in autumn (McVean 1953). The seeds, which have lateral cork-like float chambers and an oily outer coat, are principally dispersed by water, although seed dispersal by wind can occur up to a distance of 60m (McVean 1955). Trees mature between the ages of three and 30 years and are considered self-incompatible (Steiner and Gregorius 1999). Growth habit is described as variable by McVean (1953), from low multiple-stemmed bush form to tall single-bole trees. At the locations studied here, some trees appeared as multi-stemmed trees characterised by trunks growing individually, but so close to one another at the base that is was not possible to distinguish whether they were the same tree or not. Future reference to multi-stemmed trees refers to the growth characteristic described here.

3.3.2 Study sites

Sampling of *A. glutinosa* trees took place within four riparian sites located alongside the River Tummel in the eastern Highlands of Scotland. The study area (Figure 3.1) covers the River Tummel from just south of the Pitlochry dam to its confluence with the River Tay. It is approximately 60 m wide and flows within a wandering gravel-bed channel and has a mean discharge, at the River Tay confluence, of $70m^3s^{-1}$ (Gilvear and Willby 2006). Analyses of old maps and documentary sources show that major changes to the river planform have occurred over the last 200 years (Gilvear and Winterbottom 1992). Flood protection embankments of the 19^{th} and 20^{th} centuries transformed unstable river sections to narrower, single-channel reaches, with limited lateral migration (Gilvear and

Winterbottom 1992) however the embankments fell into disrepair following a large flood event in 1903 (Parsons and Gilvear 2002). Subsequently, the river has returned to a more natural morphology and although the river still occupies one main channel, lateral instability is prevalent (Winterbottom 2000), secondary channels are activated at high flow, and backwaters are present (Parsons and Gilvear 2002). Each of the four study sites encompassed areas of dynamic river shingle, where river action deposits and re-arranges shingle, and a wide range of successional vegetation communities occur. The *A. glutinosa* woodland occurs on shingle and other alluvial soils alongside areas of bare shingle, neutral grassland, and open water, including abandoned river channels and backwaters. Here, the term 'backwater' refers to a former river channel that, through the deposition of alluvial or woody debris, has lost its upstream connection with the main river channel but maintains a downstream connection. The four sites, referred to here, in upstream to downstream order as Tomdachoille, Moulinearn, Ballinluig, and Richard's Island, occur over an approximate 6 km stretch of the River Tummel, with Richard's Island located on the confluence of the River Tummel and the River Tay (Figure 3.1). Each site varied in size, ranging from 6.5 Ha to 19.3 Ha (Table 3.1). These four sites are of national importance, designated collectively as the Shingle Sands Site of Special Scientific Interest (SSSI), forming a series of extensive and dynamic river shingle areas in various stages of colonisation (Scottish Natural Heritage 2013). The study sites are also of European conservation importance, forming part of the Shingle Islands Special Area of Conservation (SAC) for Annex I priority feature 'alluvial forests with *Alnus glutinosa* and *Fraxinus excelsior*' (JNCC 2014). All necessary permissions were gained from the landowners, land manager, and the statutory authority prior to fieldwork being undertaken in 2011.

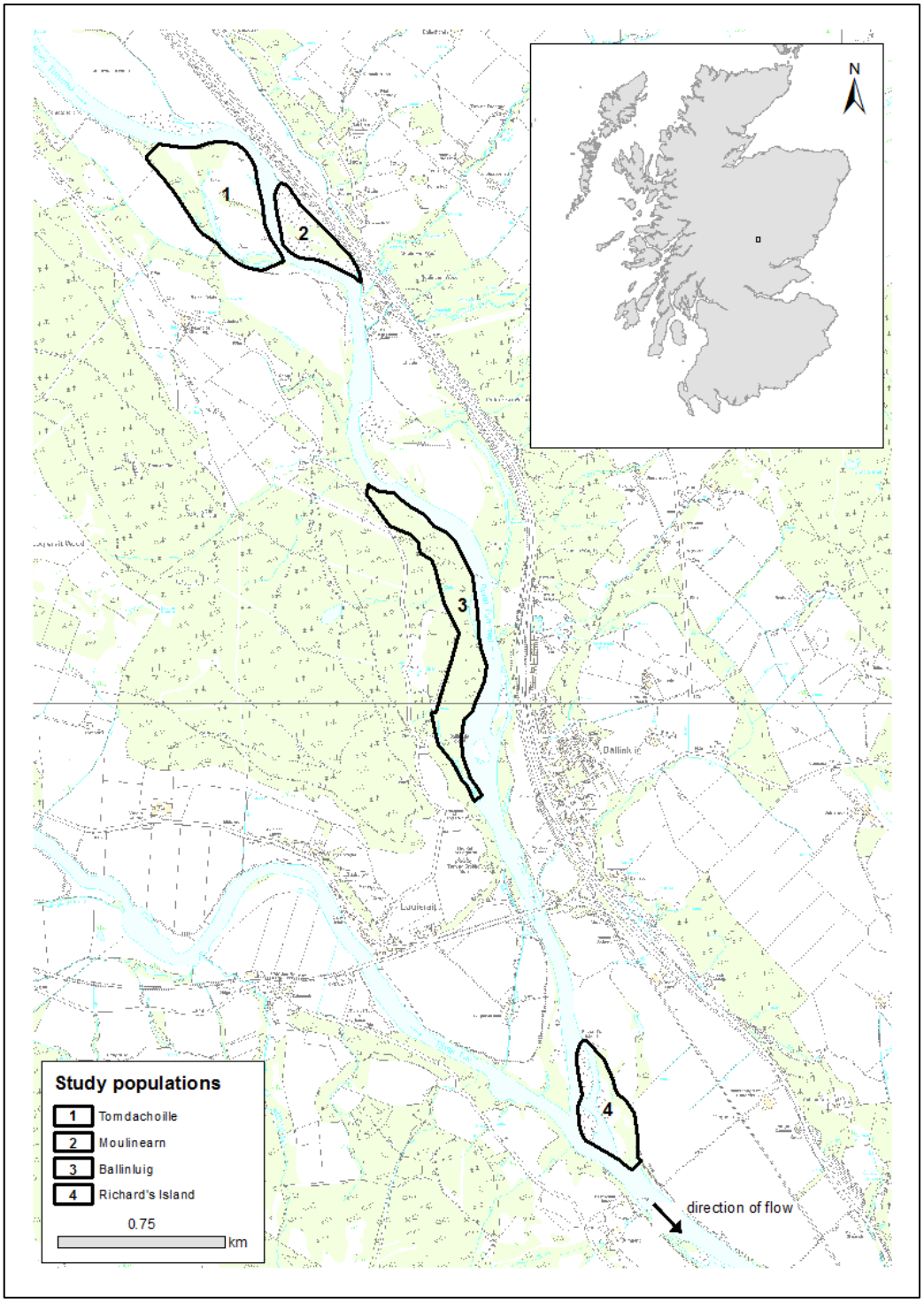

Figure 3.1 Map showing the location of the four study sites on the River Tummel. Inset map shows the location of the River Tummel in Scotland.

3.3.3 Sample collection and preparation

3.3.3.1 Tree core sampling and preparation

Tree cores were taken from 60 *A. glutinosa* trees located at the Ballinluig site. Straight, single stemmed trees with a DBH >5 cm, free of any apparent injury, were selected for sampling and the diameter at breast height (DBH) and diameter at 30 cm height was recorded. Cores were taken using a standard 4.3 mm increment corer at 30 cm above ground. One core per tree was taken except for in the case where the pith was missed in which case two cores, perpendicular to each other, were taken to improve the likelihood of intercepting the pith. The increment corer was sterilised prior to coring each tree and all cores were stored and air-dried in paper straws. Dried samples were mounted, sanded and polished up to 1200 grit following standard procedures (Speer 2010), then scanned at 1600 dpi, using an Epson GT-20000 flatbed scanner, and saved as .jpg files. Ring width was measured to an accuracy of 0.001 mm using CooRecorder v.7.4 (Larsson 2003a) and, where cores had been taken from the same tree, cross-referenced using CDendro v.7.4 (Larsson 2003b). Bark width was measured following the same methodology.

3.3.3.2 Riparian woodland inventory data collection

At each of the four study sites woodland inventory data were recorded in four 20 m x 20 m plots, within mature woodland and young woodland / scrub habitats. Areas of mature and young woodland were identified by walk overs of each site, aerial photography and maps. Maps of each site were then overlaid with a grid, representing 20 m x 20 m plots, and potential inventory plots from each stratified section selected randomly using a random number generator. On site, assuming accessibility, inventory plots were sampled according to the randomly generated points. All trees were recorded and classed as

seedlings based on their height (<30 cm), as saplings based on their height (>30 cm) and size (<20 cm circumference at 30 cm height) and as adult when ≥20 cm circumference at 30 cm height. Adult trees were identified as multi-stemmed or single stemmed and the DBH and at 30 cm height was measured.

3.3.3.3 Leaf sample collection, DNA extraction and microsatellite analysis

A total of 902 leaf samples were collected from adult and sapling trees across the four study sites (Table 3.1), randomly sampling individuals situated at a range of distances apart. Trees were classed as adults where ≥20 cm circumference at 30 cm height. Trees were identified as saplings based on their height (>30 cm high) and size (<20 cm circumference at 30 cm height). Occasionally, in the case of multi-stemmed trees, leaves were collected from more than one stem. In this situation each stem was recorded as an individual sample, but shared the same geographical location as the other stems of the same multi-stemmed tree. The geographical location of each sampled tree was recorded using a Garmin GPSMAP 62s handheld navigator and, in the adult cohort, DBH and at 30 cm height was measured for subsequent woodland structure analyses. Leaf samples were immediately placed in silica gel (Chase and Hill 1991) and subsequently stored at room temperature.

Genomic DNA from the seedling samples was obtained using the Isolate Plant DNA Mini Kit (Bioline), according to manufacturer instructions, and eluted into a final volume of 30 µL. Genomic DNA from the leaf samples was obtained using the DNeasy 96 Plant Kit (Qiagen) following manufacturer instructions. All extracted DNA was stored at -20 ^{o}C until further use. DNA concentration was quantified using a NanoDrop 2000

Spectrophotometer (Thermo Fisher Scientific) and samples were adjusted to 10 ng/ µL for PCR amplification. DNA was amplified using the multiplex of 12 nuclear microsatellite markers (Ag01, Ag05, Ag09, Ag10, Ag13, Ag14, Ag20, Ag23, Ag25, Ag27, Ag30, Ag35) of Lepais and Bacles (2011). Multiplex reactions were carried out in a total volume of 5 µL using 1X Type-it Microsatellite PCR Kit (Qiagen) and 0.5 µL of template DNA and performed in a Veriti thermocycler (Applied Biosystems). PCR conditions followed those described in Lepais and Bacles (2011): 5 min denaturation at 95 ^{o}C followed by 30 cycles of 95 ^{o}C for 30 s, 58 ^{o}C for 180 s, and 72 ^{o}C for 30 s, finishing with a final elongation step of 60 ^{o}C for 30 min. Following test amplicon success on 2% agarose gel 1 x TBE electrophoresis, samples were sent to DNA Sequencing and Services (Dundee, UK) for fragment analysis on a Biosystems 3730 capillary sequencer at a 1:50 dilution using GeneScan 500 LIZ size standard (Applied Biosystems). The resulting electropherograms were analysed using GeneMarker v.2.4.0 software (Softgenetics) and the correct assignment of allele size class checked in FlexiBin (Amos *et al*. 2007).

Based on Bonin *et al*. (2004), repeat amplification (including some blind samples) and fragment analysis was undertaken to estimate allele and genotype mismatch errors within each population and across all populations. Across all four populations a total of 160 leaf samples (20% of the total genotyped sample size) were repeated. Genotyping errors and null alleles were quantified using two methods, as advocated by Dąbrowski *et al*. (2014), using Micro-Checker (Van Oosterhout *et al.* 2004) and Cervus v.3.0.7 (Kalinowski *et al*. 2007).

Genotypic linkage equilibrium between loci pairs within each population and across all four populations was checked using FSTAT v.2.9.3.2 (Goudet 1995). Significant associations between loci were identified by randomly associating genotypes at pairs of loci over 1100 and 8800 permutations respectively, based on the 5% nominal level after Bonferonni correction.

Table 3.1 Details of the four study sites and the number of *A. glutinosa* adult and sapling trees sampled at each site.

Site name	Site code	Latitude	Longitude	Size of site (Ha)	No. adults sampled	No. saplings sampled
Tomdachoille	Tom	3° 41' 43''	56° 40' 38''	19.3	190	38
Moulinearn	Mou	3° 41' 20''	56° 40' 35''	6.5	128	70
Ballinluig	Bal	3° 40' 36''	56° 39' 37''	18.5	232	45
Richard's Island	RIs	3° 39' 52''	56° 38' 26''	9.6	180	19

3.3.4 Data analysis

3.3.4.1 Developing a size-age standard for *A. glutinosa*

To estimate the age of sampled *A. glutinosa* trees, the numbers of annual rings revealed on each increment core were counted. For individuals where the core did not include the pith an age correction procedure adapted from Duncan (1989) was implemented as follows. The number of missing years not represented by the core was estimated by first identifying the length of core missing (based on the diameter measured in the field and the length of the sampled core), then estimating the average annual growth ring width based on the first five years of the sample, the number of missing years then estimated by dividing the missing length by the mean ring width. The number of tree rings measured and the estimated missing years were summed for each core and the final individual tree age estimated based on the mean number of tree rings from each core. No correction

factor to adjust for tree rings lost due to coring height was applied as it was not possible to measure tree height. Consequently all subsequent estimates of *A. glutinosa* tree age are based on a 'greater than' age.

Regression analysis was used to describe the relationship between diameter at 30 cm field measurements and tree age based on the above ring counts. To test whether a straight-line model or non-linear regression model provided the best description of the data the *F* test using anova was used. Following Crawley (2013), the straight-line linear model was consequently compared to a linear model bound by zero, and to a polynomial regression. Using AIC, the straight-line linear regression was also compared to an asymptotic regression. All analyses were conducted using R (R Core Team 2014). Subsequently, all *A. glutinosa* trees sampled in the field (as part of woodland inventory survey and leaf collection) were assigned an estimated age based on diameter at 30 cm height.

3.3.4.2 Woodland structure

3.3.4.2.1 Riparian woodland inventory

For each species, the mean number of adults, saplings and seedlings recorded in both mature and young woodland habitat at each site, and across all four sites was calculated. For *A. glutinosa*, the mean DBH, age, stems per adult and per sapling, and density per hectare for adults, saplings, and seedlings, was calculated for mature and young woodland habitat at each site, and across all four sites. The age of adult *A. glutinosa* trees, based on the circumference at 30 cm height measured in the field, was estimated using the size-age standard described in (Sections 3.3.4.1 and 3.4.1). In the case of multi-

stemmed trees, DBH and age were taken from the largest stem. To assess the level of multi-stemmed trees the total number of individual adult and sapling stems was divided by the number of adult and sapling trees respectively, giving a number of stems per adult and per sapling for the mature and young woodland habitats. Welch's T-test was used to test for differences in the mean DBH, age, and, number of stems per adult and per sapling between mature and young woodland habitat at each site, and across all four sites.

3.3.4.2.2 *Alnus glutinosa* woodland structure

Analysis of *A. glutinosa* woodland structure was investigated by examining the location of individual sapling and adult *A. glutinosa* trees in relation to the main river channel. For each individual tree three Euclidean distance variables were considered: 'east-west distance' represented the distance between the location of each tree and the centre line of the River Tummel along an east-west gradient; 'near-distance' represented the shortest distance between each tree and the centre line of the River Tummel, regardless of direction; and, at each site, the 'distance-south' represented the distance between the location of each tree and the most southerly location of each site. The distance-south measure accounted for the most downstream point of each site as, within the study reach, the River Tummel flows in an approximately north to south direction. All distances were measured using ArcGIS v.10.2.2 (ESRI Inc. 2014), based on digital ordnance survey maps (EDINA Digimap® 2015).

Initial analysis tested if saplings occurred closer to the main channel, and at the upstream location of each site, in comparison to adults. Significant difference between saplings and adults for each distance measure were tested using the Wilcoxon rank sum test,

implemented in R (R Core Team 2014). Secondly, the relationship between the age of individual *A. glutinosa* adults and their location within each site in relation to the main river channel was investigated. To account for the Poisson distribution of tree age generalized linear mixed models were implemented. Analysis was undertaken using R (R Core Team 2014), using the *glmer* function in the lme4 R package (Bates *et al.* 2014a; Bates *et al.* 2014b). The fixed predictor variables of *A. glutinosa* tree age were the three distance measures described, and each site was included as a random effect. As continuous values, all the numerical predictor variables were centred and scaled by subtracting the mean of each variable from each variable value, and then dividing by the standard deviation. Prior to undertaking statistical analysis, the data was examined, following Zuur *et al.* (2010), to identify any potential data problems. Consequently, collinearity between the east-west distance and near distance measures was identified (correlation coefficient = 0.76). Thus, subsequent model comparisons were used to identify which distance measure was the most explanatory. Model selection was based on using Akaike information criterion (AIC). The best model was selected based on the lowest AIC value and the alternative models were assessed using differences in AIC (Δ_i), and associated Akaike weights (w_i) following Burnham and Anderson (2002). To describe the amount of variance explained by each model, the marginal R^2 (proportion of variance explained by fixed factors) and conditional R^2 (proportion of variance explained by both the fixed and random factors) was estimated following Nakagawa and Schielzeth (2013), implemented using the MuMIn R package (Barton 2015).

3.3.4.3 Genetic analysis

Initial *A. glutinosa* genetic analysis sought to first, identify the level of clonality across the four study sites, and second, to examine whether major genetic discontinuities existed either within or between the study sites. Subsequent analysis focused on comparison between different *A. glutinosa* age cohorts. Age cohorts were determined based on the age estimates obtained for individual *A. glutinosa* adults. Three age cohorts were defined: saplings (n = 166), the 200 youngest adult trees, and the 200 oldest adult trees, with individuals from each site represented within each age cohort (Table 3.2). Restricting the youngest and oldest cohorts to 200 individual *A. glutinosa* trees ensured that there was no crossover between upper 95% confidence interval of the young cohort and the lower 95% confidence interval of the old cohort.

Table 3.2 Details of the three *A. glutinosa* age cohorts examined.

		No. individuals from each study site					
Cohort	**n**	**Tom**	**Mou**	**Bal**	**RIs**	**Mean age (years)**	**Age range (years)**
Sapling	166	36	66	45	19	-	-
Young	200	53	57	56	34	26	6 - 36
Old	200	56	12	96	36	96	68 - 233

3.3.4.3.1 Occurrence of clones

Initial analysis sought to clarify whether leaves collected from more than one stem of the same multi-stemmed tree shared the same genotype. Implemented in GenClone v.2.0 (Arnaud-Haond and Belkhir 2007), genotypes from all sampled trees were compared and, where identical genotypes were revealed, field records were consulted to determine whether they were sampled from the same multi-stemmed tree (*i.e.* shared the same

geographical coordinates). Where clones were detected, only one individual was retained for subsequent analysis.

3.3.4.3.2 Identification of major spatial genetic structure

To identify whether major genetic discontinuities existed, either within or between the four study sites, genetic clustering methods were used to investigate population structure. Using the individual-based Bayesian genetic assignment method in STRUCTURE v.2.3.4 (Pritchard *et al.* 2000; Falush *et al.* 2003) to infer the number of distinct genetic clusters, analysis was first undertaken with no *a priori* geographic location provided, using the admixture model (α, allowed to vary, based on the data and initialised at 1) to account for mixed ancestry among populations, and the correlated allele frequency model (λ = 1), which assumes that the K populations have undergone independent drift away from any ancestral population (Pritchard *et al.* 2010). For each analysis, 10 iterations were run for each K = 1 to K = 8 (River Tummel reach) and K = 1 to K = 4 (each individual site). Each run was composed of a burn-in of 100,000 followed by 200,000 Markov Chain Monte Carlo (MCMC) iterations, which was sufficient to reach convergence. Due to the low level of population structure detected all analysis were repeated using the LOCPRIOR model (coding each population as a different integer), with the aim of improving STRUCTURE performance (Hubisz *et al.* 2009). All outputs were evaluated by first pooling parameter estimates for each run to identify the lowest mean likelihood L(K) and variance per K value, implemented in STRUCTURE HARVESTER (Earl and vonHoldt 2012); and second, by inspecting the assignment of individuals to populations for the most appropriate value of K (Pritchard *et al.* 2010).

Validation of clustering-based analysis, such as STRUCTURE, is important (Guillot *et al.* 2009); therefore further analysis to investigate the presence of genetic structure was undertaken using Geneland v.4.0.4 (Guillot *et al*. 2005a; Guillot *et al*. 2005b), another Bayesian clustering program. As recommended, analysis was started using the uncorrelated allele frequency model (Geneland Development Group 2012), as setting K as an unknown in the correlated model can lead to an overestimation of K (Guillot *et al*. 2014). Using the *mcmc* function each analysis was based on 1,000,000 Markov Chain Monte-Carlo (MCMC) iterations, thinning set to 1,000, and a burn-in of 200 for each value of K = 1 to K = 10. Runs were performed 10 times for each model to compare average posterior probabilities for each value of K. As the correlated model is better at detecting structure in the case of low differentiation (Guillot 2008) each analyses was then re-run using the correlated model, fixing K at the value obtained from the uncorrelated frequency model. Analysis was used to investigate population structure across the four study sites.

3.3.4.3.3 Genetic diversity between age cohorts

Genetic diversity statistics for each age cohort, including the mean number of alleles per locus (N_A), allelic richness (A_R) (Petit *et al*. 1998), gene diversity (H_E) (Nei 1978) and the inbreeding coefficient (F_{IS}) (Weir and Cockerham 1984) were calculated using SPAGeDi 1.4c (Hardy and Vekemans 2002). Rarefaction analysis of A_R was based on 300 gene copies and significance of F_{IS} values were obtained following 10,000 permutations of gene copies within individuals relative to each population. Significant differences between N_A, A_R, H_E, and F_{IS} between age cohorts were tested using the Kruskal-Wallis test,

implemented using R (R Core Team 2014). The presence of private alleles was detected using GenAlEx v.6.501 (Peakall and Smouse 2006; Peakall and Smouse 2012).

3.3.4.3.4 Spatial genetic structure

Assessment of differences in spatial genetic structure (SGS) within the youngest and oldest cohorts of *A. glutinosa* trees was undertaken in SPAGeDI v.1.4c (Hardy and Vekemans 2002). SGS within the sapling cohort was previously assessed in 2 (see Sections 2.3.10 and 2.4.7). The extent of SGS between each of the three cohorts was also assessed. Kinship coefficients (F_{ij}) between individuals *i* and *j* were estimated using Nason's kinship coefficient (Loiselle *et al*. 1995), as it is found to be statistically robust (Vekemans and Hardy 2004). Nason's kinship coefficient is based on the probability that a random gene from *i* is identical to a random gene from *j*, and defined as $F_{ij} = (Qij - Qm) / (1 - Qm)$, where *Qij* is the average probability of identity by state for random gene copies from individuals *i* and *j*, and *Qm* is the average probability of identity by state for gene copies coming from random individuals from the reference population (Vekemans and Hardy 2004). In each dataset, the association between all pairs of F_{ij} and spatial distances (*r*) was characterised by averaging the pairwise statistics to a set of predefined distance intervals. Preliminary tests were undertaken to establish suitable distance classes that would enable comparison between three age cohorts, and to ensure as close to >100 pairwise comparisons within each distance class, as advised by Hardy and Vekemans (2013). The tests revealed that very few pairwise comparisons occurred between distances of 600 m and 2,000 m, and that between 2,000 m and 5,000 m mean F_{ij} values showed no significant change in slope. Consequently, subsequent analysis were restricted to a maximum distance of 550 across nine distance classes set at 20m, 40m, 60m, 80m, 100m,

200m, 300m, 400m and 550m. For further investigation, and due to the larger dataset, additional analyses for the oldest and youngest adult cohorts were undertaken using smaller distance classes (smallest distance class 5 m in the oldest cohort, 10 m in the youngest cohort). In every analysis over 100 pairwise comparisons occurred within each distance class except in the between youngest adult-sapling cohort analysis where only 75 pairwise comparisons occurred in first distance class (20 m) and, in the between oldest adult-youngest adult analysis where no pairwise comparisons occurred until approximately 100 m. Averaged F_{ij} were regressed to the natural logarithm of the distance ln(r_{ij}) to provide the regression slope (b). To allow comparison between analyses relatedness values were calculated using the same allele frequencies, based on all *A. glutinosa* saplings and adults (n = 820). To test for SGS, and to obtain 95% confidence intervals, the regression slope was compared to that obtained following 10,000 random permutations of the spatial positions of individuals under the null hypothesis that F_{ij} and d_{ij} are uncorrelated. Standard errors and mean multilocus F_{ij} estimates within each distance class were obtained through jackknifing over loci following Sokal and Rohlf (1995).

To compare the extent of SGS among three age cohorts, the *Sp* statistic was used (Vekemans and Hardy 2004), as it accounts for differences in SGS due to variation in sampling schemes. The Sp statistic was calculated as $-b / (1 - F_{(1)})$, where b is the regression slope of F_{ij} on the natural logarithm of the distance classes, and $F_{(1)}$ is the mean F_{ij} between individuals belonging to the first distance interval. Thus *Sp* considers average kinship across individuals relative to the extent of the decrease in *F* across distance

intervals. The standard error of *b* is given as an estimate of the variability of *Sp*, calculated by jackknifing over loci (Hardy *et al*. 2006).

Finally, to assess the level of temporal genetic structure, comparison between the genotype and age of individual adult *A. glutinosa* trees was made. Using the same methodology described for examining SGS, pairwise genetic distances were compared to pairwise age differences. Twenty five age classes were applied automatically using SPAGeDi, at 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 24, 27, 29, 32, 35, 38, 42, 46, 51, 56, 61, 69, 78, 92, and 227 years to ensure an approximately equal number, and over 100, pairwise comparisons within each age class.

3.4 Results

3.4.1 Tree cores

All adult trees sampled, bar 13 individuals which could not be measured in the field, were assigned an estimated age (n = 640). In total, 88% of all tree ages were adjusted to compensate for missing rings from the pith of the tree. Following exploratory analysis a straight-line model, based on diameter at 30 cm and tree age, was found to best describe the data, indicating a mean age increase of 2.6 (±SE 0.14) years for every centimetre increase in diameter (R^2 = 0.84) (Figure 3.2).

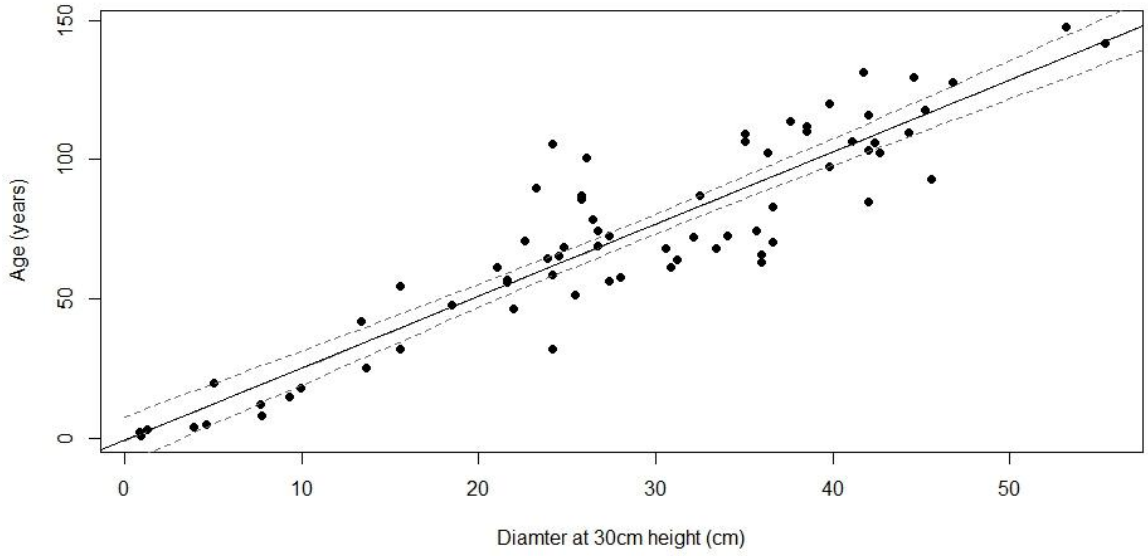

Figure 3.2 Linear regression analysis showing the relationship between diameter at 30 cm height and tree age (number of growth rings) for *A. glutinosa* (n = 60) within the lower River Tummel. Dotted line represents 95% confidence interval.

3.4.2 Riparian woodland structure

A total of 13 tree species were recorded across the four study sites (Appendix 3.1). Species composition varied between sites, and between woodland habitat types. *Alnus glutinosa* and *Salix sp*. were the most frequently occurring species, with other species including *Acer pseudoplatanus* and *Betula pendula* (Figure 3.3). Across all four study sites, *A. glutinosa* accounted for 51% and 52% of all adult trees counted in the mature and young woodland habitat respectively (Figure 3.3). In the mature woodland, *A. glutinosa* accounted for 1% of all saplings and none of the seedlings counted (Figure 3.3). By comparison, in the young woodland, *A. glutinosa* accounted for 34% of all saplings and 33% of all seedlings observed (Figure 3.3). Although the overall mean number of individual *A. glutinosa* adult trees in the mature and young woodland habitats was practically identical, the size-structure of *A. glutinosa* in mature and young woodland differed (Figure 3.4). Significantly larger stem DBH in adult *A. glutinosa* trees was found in

mature woodland compared to young woodland at every study site, and across all four sites (Table 3.3). Similarly, *A. glutinosa* trees at every site, and across all four study sites, were significantly older in the mature woodland (Table 3.3). No discernible pattern in the number of *A. glutinosa* stems per adult, or the number of *A. glutinosa* stems per sapling, between the mature and young woodland habitats was apparent and across all four sites the number of stems per adult and per sapling was practically identical (Table 3.3). However, at the Ballinluig site, significantly more *A. glutinosa* stems per adult occurred in the young woodland, compared to the mature woodland (Table 3.3). Comparison between the number of stems per adult and the number of stems per sapling (i.e. regardless of woodland habitat type) consistently revealed a higher number of stems per adult at each site although this difference was only significant at the Richard's Island site (Table 3.4). No obvious pattern in *A. glutinosa* adult tree density, or adult stem density was apparent: higher densities of *A. glutinosa* adults were found in the mature woodland at all sites except Tomdachoille; but higher densities of *A. glutinosa* adult stems were found in the young woodland at the Tomdachoille and Ballinluig sites (Table 3.3). However, the density of *A. glutinosa* saplings, sapling stems, and seedlings was consistently higher in the young woodland at every study site (Table 3.3). The total number of *A. glutinosa* stems per hectare (i.e. adults plus saplings plus seedlings) was consistently higher in the young woodland compared to the mature woodland.

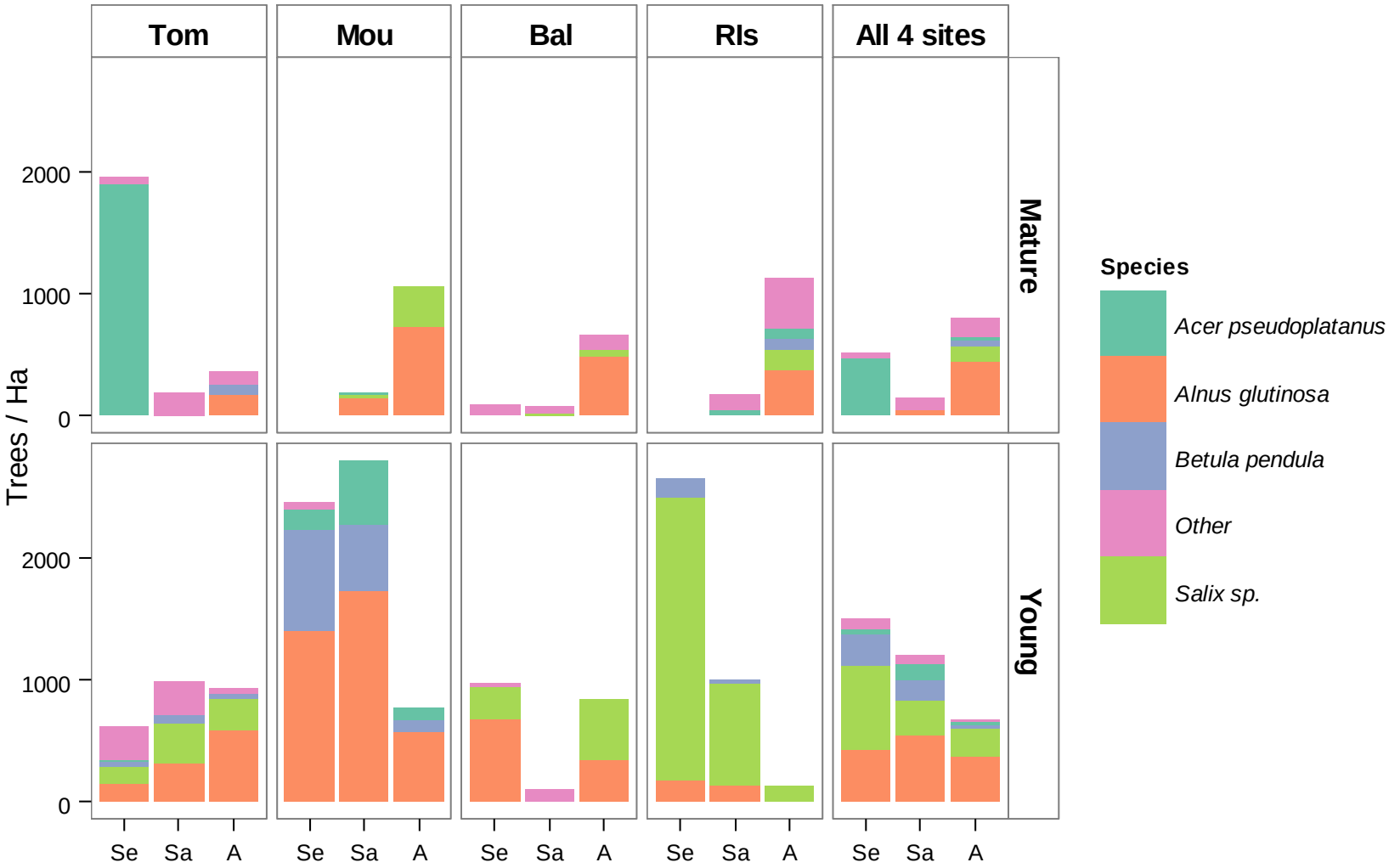

Figure 3.3 Mean number of seedling (Se), sapling (Sa), and adult (A) individuals within the mature and young woodland habitats at each study site, and across all four sites.

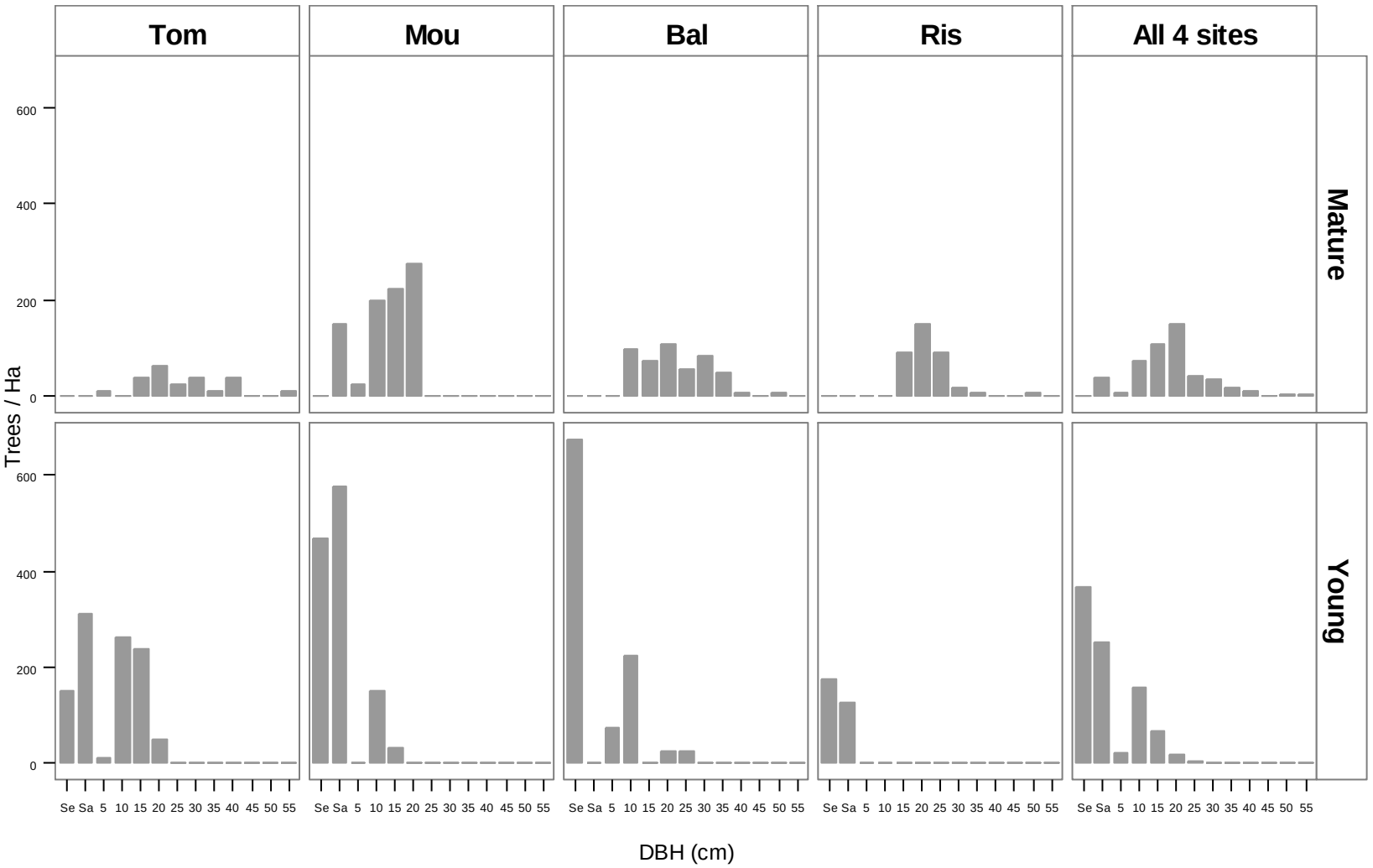

Figure 3.4 Size-structure of *A. glutinosa,* including seedlings (Se), Saplings (Sa), and adults (based on diameter at breast height (DBH)) within mature and young woodland habitats at each study site, and across all four sites.

Table 3.3 Mean *Alnus glutinosa* values ± standard errors for forest inventory quadrats within mature (MW) and young woodland (YW) habitat at each site, and across all four sites. Significant values for differences in DBH, age, stems per adult tree, and stems per sapling tree between mature and young woodland are shown next to the young woodland values as *$P < 0.05$ and ***$P < 0.001$.

	Tomdachoille		Moulinearn		Ballinluig		Richard's Island		All 4 sites	
	MW	YW	MW	YW	MW	YW	MW	YW	MW	YW
No. quadrats	**2**	**2**	**1**	**3**	**3**	**1**	**3**	**1**	**9**	**7**
DBH	24.5 ± 11.7	10.9 ± 3.6***	12.8 ± 4.1	8.4 ± 1.8***	19.8 ± 9.1	8.6 ± 5.0***	18.8 ± 6.4	-	18.8 ± 8.7	9.8 ± 3.7***
Age	79.5 ± 36.8	33.6 ± 11.0***	39.4 ± 12.6	26.6 ± 5.4***	62.9 ± 27.4	28.2 ± 16.7***	59.5 ±	-	59.5 ± 26.6	30.8 ± 11.4***
Stems per adult tree	2.0 ± 1.0	1.9 ± 1.2	2.6 ± 2.4	2.4 ± 1.6	1.5 ± 1.0	2.6 ± 1.74*	2.3 ± 1.7	-	2.0 ± 1.5	2.1 ± 1.4
Stems per sapling tree	-	1.8 ± 1.2	2.0 ± 1.7	2.0 ± 1.8	-	-	-	1.2 ± 0.5	2.0 ± 1.7	1.9 ± 1.6
Density adults / Ha	237.5 ± 17.7	587.5 ± 795.5	725.0 ± NA	190.0 ± 332.0	500.0 ± 253.7	350.0 ± NA	3.7 ± 200.5	-	472.2 ± 222.7	300.0 ± 437.1
Density adult stems / Ha	475.0 ± 141.3	1112.5 ± 1538.0	1875.0 ± NA	450.0 ± NA	765.0 ± 316.6	925.0 ± NA	825.0 ± 390.5	-	916.7 ± 487.3	642.9 ± 904.9
Density saplings / Ha	-	312.5 ± 123.7	150.0 ± NA	575.0 ± NA	-	-	-	125.0 ± NA	16.7 ± NA	353.6 ± 237.0
Density sapling stems / Ha	-	550.0 ± 353.6	275.0 ± NA	1150.0 ± 409.3	-	-	-	150.0 ± NA	30.6 ± NA	671.4 ± 524.1
Density seedlings / Ha	-	150.0 ±	-	465.0 ± 245.4	-	675.0 ± NA	-	175.0 ± NA	-	364.3 ± 218.7

Table 3.4 Mean number of *Alnus glutinosa* stems per adult tree ± standard errors and stems per sapling tree ± standard errors at each study site, and across all four sites. Significant differences between the adult and sapling trees are shown next to the stems per sapling values as **$P < 0.01$.

Site	Stems per adult tree	Stems per sapling tree
Tomdachoille	1.9 ± 1.1	1.8 ± 1.2
Moulinearn	2.5 ± 1.9	2.0 ± 1.7
Ballinluig	1.7 ± 1.2	-
Richard's Island	2.3 ± 1.7	1.2 ± 0.5**
All 4 sites	2.1 ± 1.5	1.9 ± 1.6

3.4.3 *Alnus glutinosa* woodland structure

The Wilcoxon rank sum test indicated significant difference between the location of *A. glutinosa* sapling and adult trees in relation to the main river channel along both the east-west distance and the near distance, with saplings occurring closer to the river than adults (Table 3.5). No significant difference was observed between the location of *A. glutinosa* sapling and adult trees along the distance south measure (Table 3.5).

Table 3.5 Results of the Wilcoxon rank sum test analysing differences between the location of *A. glutinosa* sapling and adult trees in relation to the main river channel based on three distance measures.

	Wilcoxon rank sum test		Mean distance (m) ±SE	
Distance	**W**	***P* value**	**Saplings**	**Adults**
East-west	69531	0.000	139 ±6	180 ±3
Near	77719	0.000	95 ±4	145 ±2
South	49407	0.156	478 ±24	426 ±10

Model comparison examining the relationship between *A. glutinosa* tree age and distance from the main river channel revealed all three fixed predictor distance measures considered explained more than the null model (Table 3.6). Comparison between the east-west distance and the near distance predictor measures identified east-west distance to provide a better model (Table 3.6). Subsequent model comparison, based on Δ_i and w_i, and considering only distance south and east-west distance as predictor variables, identified the predictors of the best-supported model for explaining the demographic structure of *A. glutinosa* trees as both distance south and east-west distance (Table 3.7). The variance explained by this best model was identified as marginal $R^2 = 0.1655$ (fixed factors), and conditional $R^2 = 0.3657$(fixed and random factors).

Table 3.6 AIC values, in ascending order, for each fixed predictor distance measure considered in GLMM examining the relationship between *A. glutinosa* tree age and distance from the main river channel .

GLMM factors		
Fixed	**Random**	**AIC**
South distance + east-west distance	Site	12,002.31
South distance + near distance	Site	12,285.59
East-west distance	Site	12,724.43
South distance	Site	12,900.34
Near	Site	12,955.09
Null	Site	13,525.87

Table 3.7 GLMM model comparison results considering south distance and east-west distance as fixed predictors for temporal variation *A. glutinosa* adults across the four study sites (random factor).

GLMM factors				
Fixed	**Random**	**AIC**	Δ_i	w_i
South distance + east-west distance	Site	12,002.31	0.00	1.00
East-west distance	Site	12,724.43	722.12	0.00
South distance	Site	12,900.34	898.03	0.00

3.4.4 DNA extraction and microsatellite analysis

A total of 884 individuals (711 adults and 173 saplings) were successfully genotyped at 12 microsatellite loci. Missing data was recorded at seven loci, with loci Ag14 reporting 7.5% mean missing data, Ag25 and Ag27 reporting 1.2% mean missing data and all other loci reporting <1% missing data. Repeat genotyping of samples revealed low levels of allele and genotype mismatch errors. The mean mismatch error rate per allele varied within each study site but was consistently <1% within each site (Table 3.8). The mean mismatch genotype error rate was higher and variable between each study site, but low across all four populations at 2.50% (Table 3.8). Null alleles were consistently revealed at locus Ag14 at every study site in both Micro-Checker and Cervus analyses. Other null alleles were identified in each study site but with no consistency between sites or between analysis methods. All subsequent analysis is based on accepting the error checking results which were consistently positive between different analysis methods, as recommended by Dąbrowski *et al.* (2014). Therefore, locus Ag14 was excluded from subsequent analysis

(as previously reported in Lepais and Bacles 2011). Removal of locus Ag14 resulted in an overall lower allele and genotype mismatch error rate (Table 3.8). All loci pairs, within each study site, and across all four sites, were found to be in linkage equilibrium.

Table 3.8 Mean mismatch error rates per allele and per genotype based on repeat amplification and genotyping of individuals within each study site. Results are shown for all 12 amplified loci, and for 11 loci following exclusion of locus Ag14 due to the presence of null alleles.

		12 loci		11 loci (excluding Ag14)	
Population	**Repeat samples (%)**	**Allele error rate (%)**	**Genotype error rate (%)**	**Allele error rate (%)**	**Genotype error rate (%)**
Tom	38	0.18	2.63	0.20	2.63
Mou	11	0.00	0.00	0.00	0.00
Bal	18	0.00	0.00	0.00	0.00
RIs	10	0.96	11.11	0.00	0.00
All 4 sites	20	0.19	2.50	0.09	1.25

3.4.5 Population genetic analysis

3.4.5.1 Occurrence of clones

Of the 887 individual *A. glutinosa* stems genotyped 815 genotypes were identified with 65 genotypes occurring more than once, and a total of 137 stems sharing a genotype with at least one other stem. Of these 137 stems, 129 stems shared the same geographical co-ordinates with at least one other stem, and eight stems, although sharing a genotype, did not share any geographical coordinates. Where stems shared the same genotype and geographical coordinates, only one individual was retained for subsequent analysis, representing 62 genotypes. All eight stems sharing the same genotype but not the same geographical coordinates were retained, representing three genotypes. Consequently, unless stated otherwise, all future analysis is based on 820 stems (653 adult and 167

saplings), representing 815 genotypes. Overall, where multi-stemmed trees occurred there was a 0.76 likelihood of the stems being genetically identical.

3.4.5.2 Genetic structure

Cluster analysis in STRUCTURE revealed no distinct genetic clusters within any site, or across all four sites. In each analysis the lowest mean log likelihood L(K) and variance values were for K = 1, and inspection of the assignment of individuals showed that the proportion of the sample assigned to each cluster was symmetric and all individuals were admixed. Output from analysis using the LOCPRIOR model gave less clear results. Individual sites, as well the aggregate of all four sites, resulted in K = >1 having the lowest mean log likelihood values. However, convergence was not clear and, in all cases, K = 1 showed little variance in output while each K = >1 output showed high variance between iterations. Inspection of the assignment of individuals to populations revealed admixture in all individuals. Consequently LOCPRIOR output was viewed with some caution as non-convergence may point towards spurious results (Guillot *et al.* 2009). Analysis with Geneland was concurrent with the output from STRUCTURE, with no genetic clustering revealed.

3.4.5.3 Genetic diversity

A total of 97 alleles were revealed across the 11 loci, ranging from 3 – 15 alleles per locus, across all three age cohorts. Genetic diversity measures were very similar between each age cohort although a similarly low but significant departure from Hardy-Weinberg genotypic proportions was identified within the oldest tree cohort (Table 3.9). No significant differences between age cohorts for the number of alleles per locus (N_A), allelic

richness (A_R), expected heterozygosity (H_E) or inbreeding coefficient (F_{IS}) were revealed by the Kruskall-Wallis test. Six private alleles were identified, two in the sapling cohort, three in the youngest tree cohort, and one in the oldest tree cohort. All loci pairs, within each age cohort, were found to be in linkage equilibrium at the 5% nominal level after Bonferroni correction.

Table 3.9 Multilocus genetic diversity for each age cohort: N, number of samples; N_A, mean number alleles per locus; A_R, allelic richness; H_E, gene diversity; and F_{IS}, inbreeding coefficient.

	Saplings	Youngest adults	Oldest adults
N	166	200	200
N_A	8.27	8.45	8.36
A_R	8.21	8.35	8.20
H_E	0.64	0.65	0.64
F_{IS}	0.01	0.00	0.03*

*$p < 0.05$.

3.4.6 Spatial genetic structure

Analysis of the oldest 200 *A. glutinosa* adult trees revealed positive F_{ij} values occurring in pairwise comparisons up to pairwise distances of 40 m although kinship values were low (maximum F_{ij} 0.0031), and not significant (Figure 3.5a). Testing of this SGS pattern with smaller distance classes (first distance class 5 m, then 10 m) revealed the same pattern in F_{ij} values, although significant positive SGS was observed in the 20 m – 30 m distance class in both analyses. In contrast, the youngest 200 *A. glutinosa* trees revealed negative F_{ij} values in pairwise distances up to 40 m before generally levelling out around zero, with the exception of F_{ij} values peaking just outside the positive 95% confidence interval at 200 m (Figure 3.5b). Testing of this SGS pattern with smaller distance classes (first distance class 10 m) revealed the same pattern of F_{ij} values. The pattern of F_{ij} values revealed in the sapling cohort were very similar to that reported previously (Section 2.4.7), and

shown in Figure 3.5c, with negative F_{ij} observed for pairwise distances up to 60 m prior to levelling out around zero. Between-generation SGS, obtained by analysing pairs between the oldest adults and the saplings, revealed no obvious pattern, with low variation in SGS (F_{ij} values between -0.0030 and 0.0028) except in the 80 m – 100 m distance class where significant positive SGS was observed (Figure 3.5e). Similarly, analysis of SGS between the youngest adults and saplings also revealed no obvious pattern, although significant positive SGS was observed in the 20 m – 40 m distance class (Figure 3.5f). Between-generation comparison among the old and young cohort was not possible as no pairwise comparisons occurred in the first four distance classes (<100 m). Use of the *Sp* statistic as a measure of SGS intensity showed that cohort of the oldest 200 *A. glutinosa* trees revealed the most intense, and only positive, SGS (Table 3.10). The *Sp* values for the youngest 200 adults, and for the between cohort comparisons (oldest-sapling and youngest-sapling) were very similar (Table 3.10). Finally, comparison of pairwise genetic distances against pairwise age differences revealed very little variation in F_{ij} values, although significant negative SGS was observed in the 10 – 12 years age difference class and significant positive SGS was observed in the 14 – 16 years age difference class (Figure 3.6).

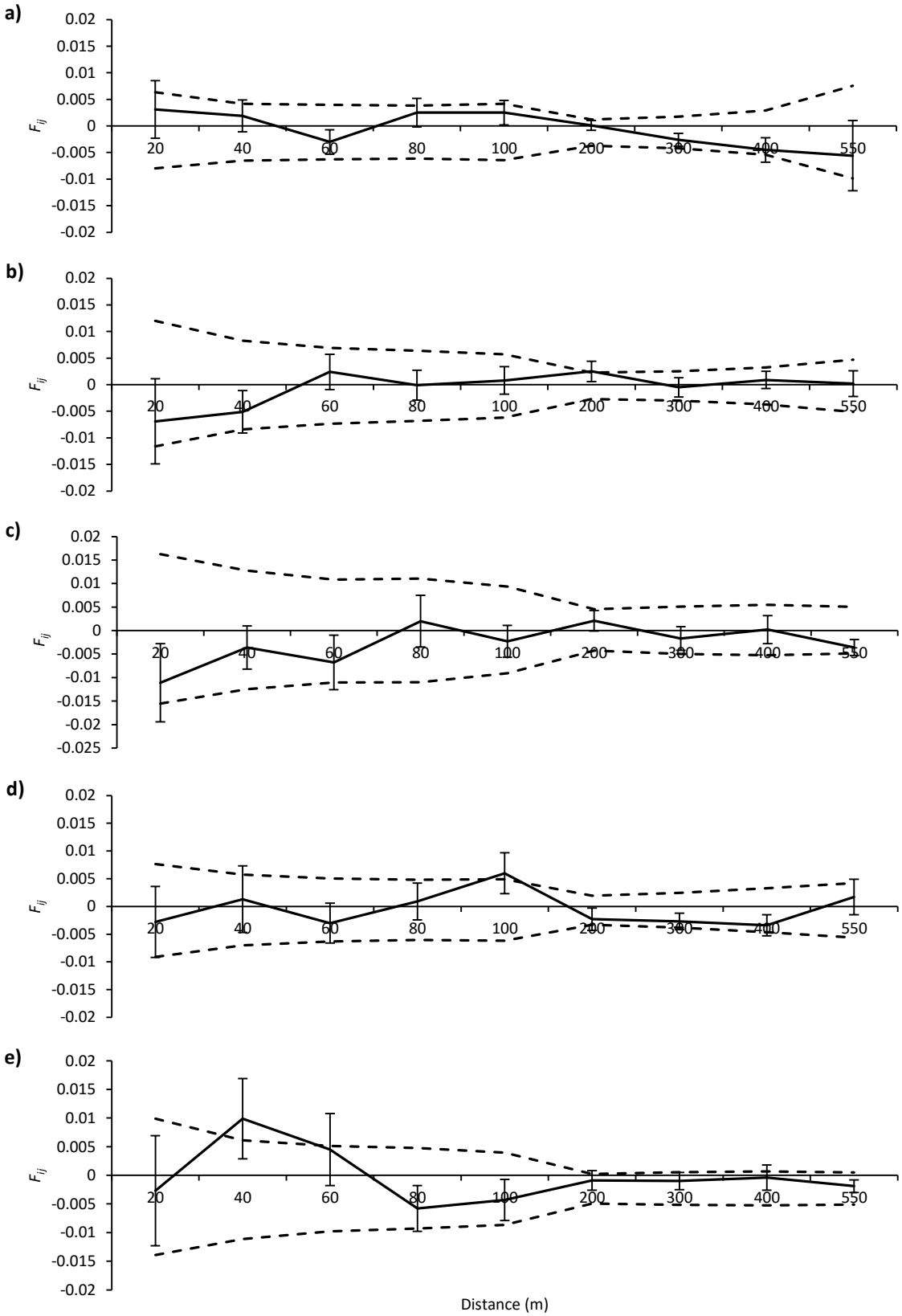

Figure 3.5 Spatial autocorrelograms for a) the oldest 200 adults; b) the youngest 200 adults; c) saplings; d) the paired oldest adult and sapling cohorts; and e) the paired youngest adult and sapling cohorts of *A. glutinosa*. Relatedness between individuals is based on Nason's F_{ij} kinship coefficient, implemented in SPAGeDi. Broken lines represent the 95% upper and lower confidence intervals, based on 10,000 simulations.

Table 3.10 SGS parameters showing the kinship coefficient for the first distance class of 0 m – 20 m (F(1)) and the rate of decrease of pairwise kinship with distance (Sp) ± standard error (SE).

Cohort	F(1)	*Sp* ±SE
Oldest 200 adults	0.0031	0.0023 ± 0.0009
Youngest 200 adults	-0.0069	-0.0009 ± 0.0009
Saplings	-0.0111	-0.0050 ± 0.0013
Oldest & saplings	-0.0028	0.0008 ± 0.0012
Youngest & saplings	-0.0027	0.0010 ± 0.0018

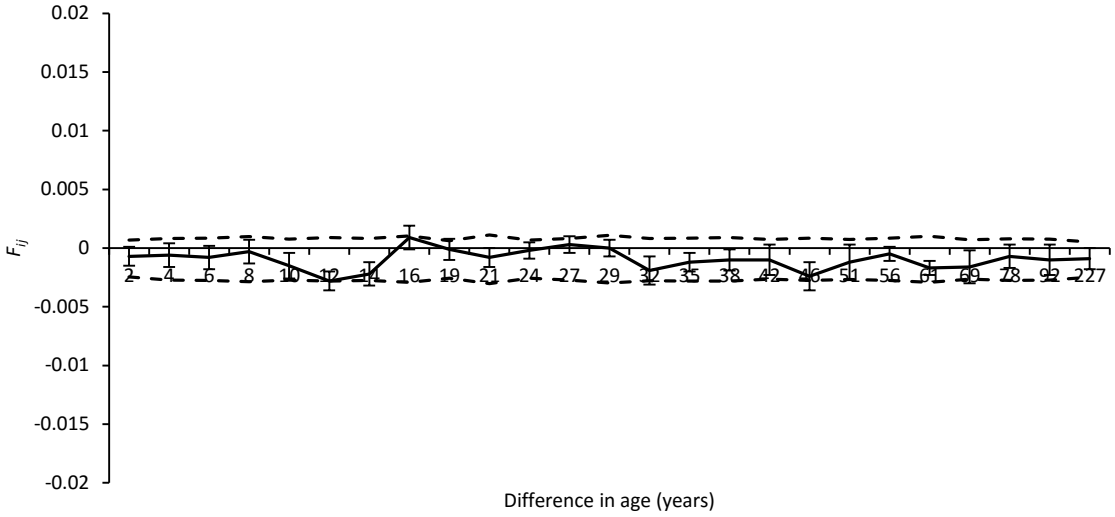

Figure 3.6 Temporal autocorrelogram of the adult generation (n = 640) based on Nason's kinship coefficient (F_{ij}), implemented in SPAGeDi. Broken lines represent the 95% upper and lower confidence intervals, based on 10,000 location permutations. Error bars around mean F_{ij} values represent standard errors obtained through jackknifing over loci to obtain multilocus estimates.

3.5 Discussion

Woodland structure

Significant differences between mature and young riparian *A. glutinosa* woodland inventory plots were identified. Despite differences between study sites, *A. glutinosa* adult trees, as anticipated, were significantly larger (DBH) and older in mature woodland compared with *A. glutinosa* adult trees in the young woodland. Similarly, the mature woodland was found to contain fewer *A. glutinosa* stems per hectare than the young woodland. Most notable however was that although *A. glutinosa* seedlings and saplings

were consistently observed in young woodland, few, sometimes no, seedlings or saplings were recorded in the mature woodland plots. This lack of *A. glutinosa* regeneration within mature woodland is reported elsewhere (McVean 1956; Claessens *et al.* 2010), and is considered a consequence of low under-storey light levels as well as a lack of moisture (Claessens *et al.* 2010), with regeneration only found to occur in forest openings greater than 1,000 m^2, or following disturbance events such as flooding (Claessens *et al.* 2010). Consistent with many pioneer species, *A. glutinosa* requires relatively high light levels for successful regeneration.

Analyses based on the age and location of individual trees, in relation to the main river channel, clearly identified a spatio-temporal pattern within the riparian *A. glutinosa* woodlands studied here. Taking account of differences between each study site, the location of adult trees in relation to the main river channel explained up to 37% of the variance in tree age. The east-west distance between each tree and the main river channel was the variable explaining the most variation in *A. glutinosa* age, with older trees located further away from the main river channel than younger adult trees, and significantly further away from the main river channel than sapling trees. This finding supports the hypothesis that there is an interaction between hydrogeomorphic processes and the riparian *A. glutinosa* woodland studied here. In particular, the spatio-temporal pattern of tree ages, along a transverse gradient perpendicular to the main river channel, suggests that the historical, lateral east-west movement of the River Tummel has influenced *A. glutinosa* stand development. Lateral channel migration is an important factor in influencing the demography of riparian vegetation (Naiman and Decamps 1997), and the insight gained here is consistent with other studies. For example, riparian trees,

including *Alnus rubra* (red alder), along the Queets River, Washington, USA, a dynamic alluvial river, were found to have a smaller basal area if located within the active floodplain, with larger trees located further away from the river on the mature terrace (Balian and Naiman 2005). This study is believed to be the first to use tree cores to age *A. glutinosa* trees in order to examine the temporal structure of *A. glutinosa* woodland, to identify cohorts of *A. glutinosa* adults based on their age, and to subsequently relate tests of genetic diversity and genetic structure to age cohorts.

The resultant insight gained is consistent with the empirical evidence previously obtained for the same reach of the River Tummel, presented earlier, and allowed investigation into temporal genetic variation in *A. glutinosa* for the first time.

Alnus glutinosa multi-stemmed growth

An unexpected finding from the woodland inventory work was that the number of stems per adult was consistently higher than the number of stems per sapling, although this difference was only significant at the Richard's Island site (Table 3.4). Sprouting of woody species has been observed elsewhere, as a response to stressful environments, including, for example, coastal dunes (Nzunda *et al.* 2007) and steep slopes with shallow soils (Sakai *et al.* 1995) and is advantageous in that a tree maintains its presence in woodland through the persistence niche (Bond and Midgley 2001). In this study it had been considered that basal regeneration of *A. glutinosa* within the study sites may occur as a response to the deposition of silts / gravel on top of seedlings / young saplings following flood events, or simply the continued growth of more than one stem, originating from the same tree that escaped burial, as observed by Gilvear and Willby (2006). However, the

consistent pattern across all four study sites of adults having more stems per tree than saplings suggests that, while sprouting may occur during the sapling stage, sprouting must also occur in the later, adult life-stage of *A. glutinosa*. Interestingly, a review of sprouting in woody species found that sprouting ability may increase with tree size, and that adult sprouting is indicative of species persistence following disturbance events, such as flooding or drought, particularly when recruitment is low (Bond and Midgley 2001), as observed in the mature woodland studied here. The mean number of 2.1 stems per *A. glutinosa* adult tree reported here is lower than the mean of 3.2 stems reported for *A. glutinosa* trees located in the forested wetlands of the Ibero-Atlantic region of Portugal and Western Spain (Rodríguez-González *et al.* 2010). Rodríguez-González *et al.* (2010) identified hydrology as the dominant driver for the number of stems per tree, with a higher number of stems per *A. glutinosa* tree observed in more saturated sites, with sprouting interpreted as a response to flooding stress and a means to persisting at sites with poor seedling recruitment. In the study undertaken here no clear pattern in the number of stems per *A. glutinosa* tree can be discerned, particularly given the mosaic of habitat types occurring within the four study sites. Finally, the genetic analysis revealed that where *A. glutinosa* growth was multi-stemmed there was a 76% likelihood of stems from the same tree being genetically identical. Believed to be the first genetic analysis of multi-stemmed *A. glutinosa* growth, this result suggests the possibility of different factors influencing or driving multi-stemmed growth. Whilst most multi-stemmed trees are genetically identical, some apparent multi-stemmed trees are genetically distinct individuals. The cause of genetically distinct multi-stemmed trees is unknown although the stranding of multiple seeds following hydrochorous dispersal of seed would feasibly lead to genetically distinct individuals occurring close to each other.

Population genetic analysis

Heterozygosity between the sapling, youngest adult and oldest adult cohorts were practically identical, with levels of expected heterozygosity in line with that reported for North African (Lepais *et al.* 2013), Irish (Cubry *et al.* 2015), and Northern Irish (Beatty *et al.* 2015) *A. glutinosa* populations. In particular, allelic richness, a suitable measure for assessing short-term genetic diversity loss (Lowe *et al.* 2005), was also more or less identical between cohorts. This observed lack of difference between generations suggests that high outcrossing rates and high gene flow, typical of tree species (Petit and Hampe 2006), maintains similar levels of genetic diversity between *A. glutinosa* generations.

Of note, the cohort of oldest adults revealed a significantly positive inbreeding coefficient, indicating a departure from Hardy-Weinberg proportions due to an excess of homozygotes. Although many forest tree species show no departure from Hardy-Weinberg equilibrium (Hamrick *et al.* 1992), where an excess of homozygotes has been identified, it has typically been at the seed stage and not in later adult life stages (e.g. Yazdani *et al.* 1985; Alvarez-Buylla *et al.* 1996). In these studies, the occurrence of thinning, or selection, between the seed and adult life stages, are posited as possible mechanisms resulting in null or negative F_{IS} values in the adult generation (Yazdani *et al.* 1985; Alvarez-Buylla *et al.* 1996). Positive F_{IS} values in the adult generations are reported in populations of *Fraxinus excelsior* L. in north-eastern France (Morand *et al.* 2002), in *Dalbergia nigra* located in the Brazilian Atlantic Forest (Leite *et al.* 2014), and in a number of Northern Irish *A. glutinosa* populations (Beatty *et al.* 2015). In all studies no firm explanation accounted for the F_{IS} values observed although the presence of null alleles

and assortative mating (Morand *et al.* 2002), a Wahlund effect (Leite *et al.* 2014, Beatty *et al.* 2015), and inbreeding (Beatty *et al.* 2015) are considered. Similarly, no firm explanation for the positive F_{IS} values observed in the oldest *A. glutinosa* adults in this study are available. While undetected null alleles may be present within any of the 11 loci included here it seems unlikely that any effect of null alleles on F_{IS} would be apparent in the oldest adult cohort but not in the youngest adult or sapling cohorts. Based on the spatio-temporal variation described above a Wahlund effect is possible, due to the presence of breeding subunits within the study population and / or uneven patterns of recruitment, as suggested by Beatty *et al.* (2015). One further possibility is historical mating among relatives in the oldest generation. Initial, small aggregations of *A. glutinosa* at each study site, now represented by the oldest *A. glutinosa* trees, may have resulted in overlapping generations and increased mating between related individuals, leading to a subsequent increase in homozygosity not observed in the contemporary generations. It may also be possible that small historical *A. glutinosa* populations may have been restricted by grazing when the river was more managed by flood embankments, again resulting in increased mating between related individuals. Finally, although no clear explanation is identified to account for the significantly positive F_{IS} value in the oldest adult cohort, the overall effect is low (F_{IS} 0.03), and lower than F_{IS} values reported for Northern Irish populations (maximum F_{IS} 0.155) (Beatty *et al.* 2015).

Despite clear evidence of spatio-temporal structure within the riparian *A. glutinosa* woodlands surveyed, no significant SGS was identified in the first distance classes, either within or between the age cohorts studied. Nevertheless distinct differences between the oldest adult cohort and the younger cohorts were revealed. In the first two distance

classes (pairwise distances up to 40 m) the young adult and sapling cohorts both reveal a similar pattern of negative F_{ij} values, however the oldest adult cohort is different, showing positive F_{ij} values. These patterns are consistent with the *Sp* statistic for each age cohorts examined, with the oldest adult cohort showing the only positive *Sp*, as well as the most intense *Sp* across all cohort comparisons and between-cohort comparisons. These findings are also consistent with the between-cohort analyses where the *Sp* values indicate that individuals in the sapling and youngest adult cohorts are more related to each other than individuals in the sapling and oldest adult cohorts are.

Although not significant, the SGS pattern for the oldest adult cohort is consistent with the SGS pattern, and *Sp* statistic, identified for the adult generation in (2) and the SGS pattern in the young adult cohort is consistent with the SGS pattern of the sapling cohort. Further, the differing SGS patterns are consistent with the significant F_{IS} identified in the oldest adult cohort but not in the young adult or sapling cohorts. Given the regeneration requirements of *A. glutinosa*, particularly suitable light and water levels and disturbance-driven areas of open gravel habitat, it seems unlikely that seeds would successfully establish in proximity to a parent tree. Hence, no or negative SGS would be expected in short distance classes, particularly where seed and pollen dispersal is high (as reported in Chapter 2). In the oldest adult cohort, the SGS pattern and F_{IS} values identified here suggest an underlying process influencing higher relatedness in the oldest cohort of *A. glutinosa* trees studied here, consistent with the findings reported and discussed previously in 2.

Despite evidence that there is spatio-temporal demographic structure within the riparian *A. glutinosa* woodlands studied, a remarkable lack of temporal genetic structure across the adult generation was identified. Trees with two years difference in age between them are no more related to each other than trees with a difference in age of over 200 years. The absence of temporal genetic structure is consistent with the absence of a reduction in gene diversity in young cohorts, further indicating extensive gene flow and high outcrossing rates.

3.6 Conclusion

Forest inventory work revealed evident differences between young and mature riparian woodland, with *A. glutinosa* accounting for approximately 50% of all recorded trees. Notably, virtually no *A. glutinosa* regeneration occurred within the mature woodland, with seedling regeneration occurring predominantly in the young woodland. *Alnus glutinosa* tree density was also higher in the young riparian woodland. Widespread, multi-stemmed 'sprouting' growth of sapling and adult *A. glutinosa* trees indicated different factors influencing multi-stemmed growth, including site disturbance. The development of a size-age standard for *A. glutinosa*, based on tree ring counts, enabled identification of a spatio-temporal structure within the riparian sites studied. In the dynamic river system studied here, riparian *A. glutinosa* woodland appears to be structured along a transverse gradient perpendicular to the main river channel, with older adult trees located further away from the river and younger adult and sapling trees located closer to the river. This finding illustrates an interaction between riparian woodland and hydrogeomorphic processes, particularly the historical, lateral, east-west movement of the River Tummel. Despite evidence for demographic structure within *A. glutinosa* woodlands, little

indication of genetic diversity or structure was identified. Measures of genetic diversity between *A. glutinosa* saplings, youngest adult trees, and oldest adult tree showed no difference between cohorts, indicative of high outcrossing and high gene flow. The inbreeding coefficient suggested an excess of homozygotes in the oldest adult cohort which is not definitively explained and, although low, may indicate a Wahlund effect or historical influences. Similarly, although no significant SGS was observed in any of the *A. glutinosa* age cohorts, the spatial autocorrelograms revealed positive SGS in the oldest adult cohort and negative SGS in the sapling and youngest adult cohorts.

Although some spatial genetic pattern was observed within each age cohort, no temporal genetic differentiation was observed, indicating random gene flow occurs across generations.

3.7 Acknowledgements

I thank the landowners involved, the Scottish Wildlife Trust as land manager, and Scottish Natural Heritage for statutory permissions; Antoine Keruzoré for help with site information; Alexander Flint, Armorel Flint, Roy Flint, Rachel Lintott, and Stewart Lothian for help with fieldwork; Liam Cavin and Sarah Greenwood for dendrochronology advice; Patricia González Diáz for help with molecular laboratory work; Paloma Ruiz-Benito for sharing her knowledge of mixed models; and Martin Clarke for help with ArcGIS analyses.

Appendix 3.1 Tree species recorded in woodland inventory quadrats

Table 3.11 Tomdachoille tree species recorded in four 20 m x 20 m woodland inventory quadrats.

Quadrat	Habitat	Species	No. adults	No. adult stems	No. saplings	No. sapling stems	No. seedlings
Tom1	Mature	*Acer pseudoplatanus*	0	0	0	0	71
		Alnus glutinosa	10	16	0	0	0
		Betula pendula	4	4	0	0	0
		Crataegus monogyna	0	0	1	1	0
		Fraxinus excelsior	2	2	13	15	6
		Prunus padus	3	6	3	3	0
		Salix sp.	1	2	1	1	0
		Ulmus glabra	1	1	1	1	0
Tom2	Mature	*Acer pseudoplatanus*	0	0	0	0	5
		Alnus glutinosa	10	23	0	0	0
		Betula pendula	6	8	0	0	0
		Fraxinus excelsior	2	2	1	1	1
		Prunus padus	2	2	2	8	0
		Sambucus nigra	1	2	1	3	0
Tom3	Young	*Acer pseudoplatanus*	0	0	0	0	9
		Alnus glutinosa	1	1	16	32	12
		Betula pendula	1	1	6	6	3
		Fagus sylvatica	0	0	0	0	3
		Fraxinus excelsior	0	0	0	0	10
		Salix sp.	1	1	9	31	11
Tom4	Young	*Alnus glutinosa*	46	88	9	12	0
		Betula pendula	2	5	0	0	0
		Corylus avellana	0	0	2	8	0
		Fraxinus excelsior	1	1	5	5	6
		Prunus padus	1	3	15	29	3
		Salix sp.	20	29	17	19	0
		Ulmus glabra	1	1	0	0	0

Table 3.12 Moulinearn tree species recorded in four 20 m x 20 m woodland inventory quadrats.

Quadrat	Habitat	Species	No. adults	No. adult stems	No. saplings	No. sapling stems	No. seedlings
Mou1	Young	*Acer pseudoplatanus*	1	1	1	1	2
		Alnus glutinosa	0	0	25	32	13
		Betula pendula	0	0	15	18	24
		Quercus robur	0	0	0	0	1
		Salix sp.	3	22	2	24	2
Mou2	Mature	*Acer pseudoplatanus*	1	1	1	1	0
		Alnus glutinosa	29	75	6	11	0
		Salix sp.	13	29	1	1	0
Mou3	Young	*Acer pseudoplatanus*	3	6	20	41	5
		Alnus glutinosa	23	54	30	64	13
		Betula pendula	4	9	2	3	1
		Salix sp.	14	32	31	87	0
Mou4	Young	*Alnus glutinosa*	0	0	14	42	30
		Betula pendula	0	0	5	7	8
		Pinus sylvestris	0	0	0	0	1
		Salix sp.	0	0	0	0	2

Table 3.13 Ballinluig tree species recorded in four 20 m x 20 m woodland inventory quadrats.

Quadrat	Habitat	Species	No. adults	No. adult stems	No. saplings	No. sapling stems	No. seedlings
Bal1	Mature	*Alnus glutinosa*	17	25	0	0	0
		Fraxinus excelsior	0	0	0	0	9
		Pinus sylvestris	2	2	0	0	0
		Salix sp.	2	3	2	12	0
Bal2	Mature	*Alnus glutinosa*	31	45	0	0	0
		Fraxinus excelsior	2	4	0	0	0
		Prunus padus	1	3	4	11	0
		Salix sp.	4	4	0	0	0
Bal3	Mature	*Alnus glutinosa*	11	21	0	0	0
		Crataegus monogyna	4	5	0	0	0
		Fraxinus excelsior	2	2	1	4	0
		Pinus sylvestris	1	1	0	0	0
		Prunus padus	1	3	1	15	1
Bal4	Young	*Alnus glutinosa*	14	37	0	0	27
		Fraxinus excelsior	0	0	2	3	0
		Prunus padus	0	0	2	3	0
		Quercus robur	0	0	0	0	1
		Salix sp.	20	34	0	0	11

Table 3.14 Richard's Island tree species recorded in four 20 m x 20 m woodland inventory quadrats.

Quadrat	Habitat	Species	No. adults	No. adult stems	No. saplings	No. sapling stems	No. seedlings
RIs1	Mature	*Acer pseudoplatanus*	3	3	1	1	0
		Alnus glutinosa	14	29	0	0	0
		Betula pendula	1	1	0	0	0
		Crataegus monogyna	14	17	4	4	0
		Fraxinus excelsior	5	5	2	2	0
		Prunus padus	5	13	1	3	0
		Salix sp.	16	26	1	1	0
		Ulmus glabra	3	3	0	0	0
RIs2	Young	*Alnus glutinosa*	0	0	5	6	7
		Betula pendula	0	0	1	1	6
		Salix sp.	5	15	34	95	93
RIs3	Mature	*Acer pseudoplatanus*	3	3	1	1	0
		Alnus glutinosa	7	22	0	0	0
		Betula pendula	7	10	0	0	0
		Crataegus monogyna	1	1	1	1	0
		Fraxinus excelsior	5	5	1	1	0
		Prunus padus	3	10	0	0	0
		Salix sp.	4	8	0	0	0
RIs4	Mature	*Acer pseudoplatanus*	5	6	3	3	0
		Alnus glutinosa	23	48	0	0	0
		Betula pendula	2	4	0	0	0
		Crataegus monogyna	2	2	3	5	0
		Fraxinus excelsior	5	5	1	1	0
		Prunus padus	1	2	0	0	0
		Salix sp.	1	1	0	0	0
		Ulmus glabra	6	11	2	3	0

Chapter 4

Landscape genetics of a key riparian tree species

Alnus glutinosa at a river catchment scale

4.1 Abstract

Rivers and their terrestrial corridors are among the most diverse and complex of terrestrial landscapes. River systems may act as important corridors for plant dispersal and gene flow over large landscapes. This study takes a landscape genetics approach to investigate, at a river catchment scale, the genetic structure of *Alnus glutinosa* (L) Gaertn, a widespread European tree and keystone species of riparian ecosystems. Leaves from 1,457 adult trees from 49 populations, across six rivers within the River Tay catchment (Scotland), were genotyped at 12 microsatellite loci to test landscape-based hypotheses, including the dispersal mechanisms of windborne pollen and waterborne seed dispersal. No downstream accumulation of genetic diversity and no genetic structure were found within any river or across the catchment despite significant differentiation between populations. Populations connected by (waterborne) seed dispersal and (overland) pollen dispersal showed significantly lower F_{ST} values than populations only connected by pollen dispersal. No isolation by distance (IBD) was found for overland Euclidean distances however significant IBD was found at hydrological distances >25 km at the catchment scale. This study shows that wind dispersal of pollen appears to be the main dispersal factor however the dispersal of seed via rivers influences the genetic structure of riparian *A. glutinosa* populations.

4.2 Introduction

Landscape genetics seeks to understand how geographic and environmental heterogeneity influence spatial genetic variation, population structure and gene flow (Manel *et al*. 2003; Manel and Holderegger 2010). Populations occur in a landscape mosaic with patterns of population genetic differentiation often reflecting the spatial

variation of dispersal, the movement of individuals or propagules that can sustain gene flow (Ronce 2007). Consequently, gene flow is a function of the dispersing individual, the habitat in which the population is located, and the intervening landscape (Sork *et al*. 1999; Baguette *et al.* 2013). Understanding how landscape features influence genetic variation within and between populations therefore has important implications for ecology, evolution, and conservation biology (Sork *et al.* 1999; Holderegger and Wagner 2006; Storfer *et al*. 2007; Segelbacher *et al*. 2010). Nevertheless, incorporating the complexity of heterogeneous landscape and other biotic factors (e.g. climate, elevation, geography) alongside observed genetic variation to assess interactions is challenging (Holderegger and Wagner 2006; Balkenhol *et al.* 2009; Storfer *et al.* 2010).

Within terrestrial landscapes, riparian corridors are among the most dynamic, diverse and complex of landscapes, adding disproportionately to both terrestrial and aquatic ecosystem function and diversity (Gregory *et al.* 1991; Naiman *et al.* 1993). Situated at the interface of terrestrial and aquatic zones, riparian vegetation influences, and is influenced by, hydrogeomorphic processes (e.g. Corenblit *et al*. 2007; Stoffel and Wilford 2012), typically forming spatial and temporal vegetation mosaics (Naiman and Décamps 1997). As well as appearing as clearly defined landscape features, enabling the identification of spatially distinct populations (e.g. populations on different rivers), rivers can influence patterns of gene flow both by acting as a physical barrier to the movement of some species and a conduit for the dispersal of other species (e.g. Sork *et al*. 1999; Storfer *et al*. 2007).

A defining characteristic of rivers and their riparian habitat is their linearity. River systems may, therefore, act as important corridors of passive plant movement, facilitating dispersal of individuals and genes across landscapes, and maintaining connectivity between upstream and downstream populations (e.g. Naiman and Decamps 1997). In plants, gene flow is maintained via dispersal of pollen and seed (Ennos 1994), with both pollen and seed being influenced by interactions with abiotic or biotic dispersal agents. Hydrochory, the passive dispersal of organisms by water, is an important biotic dispersal agent of plants, influencing the population dynamics and geographic distribution of plant species (Nilsson *et al*. 2010). Plant adaptations such as hydrochorous propagules (e.g. Johansson and Nilsson 1993) and cork-like seed tissue (e.g. McVean 1955) enable plant survival in water, facilitating the downstream dispersal of plant material within a river catchment.

Structural connectivity, described as the linking of habitats and populations by a spatial structure (Manel and Holderegger 2013), is clearly provided by rivers. In the case of aquatic and riparian plant species, the structural connectivity provided by river catchments potentially facilitates the dispersal of plants and the functional connectivity of plant populations across large, heterogeneous landscapes. Functional connectivity, the response of individuals to landscape features (Taylor *et al.* 1993), will influence gene flow and shape spatial genetic variation across landscapes. Clearly, spatial genetic variation will also be shaped by the ecology of individual species, such as different dispersal mechanisms, and it is therefore important to consider species-specific life-history attributes in landscape genetic studies (Bolliger *et al*. 2014). For example, because rivers generally have unidirectional flow, the ‘unidirectional diversity hypothesis’ infers there

will be a downstream accumulation of genetic diversity, and upstream paucity, as a consequence of the downstream direction of seed dispersal (e.g. Ritland 1989; Markwith and Scanlon 2007). In contrast, the movement of plant propagules via animals (zoochory) or wind (anemochory) can occur across river catchments, in both an upstream and downstream direction (Werth and Scheidegger 2014).

Alnus glutinosa (L.) Gaertn, commonly known as black alder, is an important temperate riparian tree species distributed across Europe. Growing in small woodlands or alongside the edge of rivers and standing water *A. glutinosa* may act as a pioneer species and forms a key element of dynamic river systems (McVean 1953). A monoecious tree, the male and female flowers form as catkins during February / March and are wind-pollinated (McVean 1955). Following fertilisation the female flower forms as a woody cone-like fruit which contains the seed, dispersed in autumn (McVean 1953). The seeds, which have lateral cork-like float chambers and an oily outer coat, are principally dispersed by water, although seed dispersal by wind can occur up to a distance of 60m (McVean 1955; Chambers and Elliot 1989). *Alnus glutinosa* is considered self-incompatible (Steiner and Gregorius 1999).

In this study, the objectives were to assess the genetic diversity and structure in widely occurring *A. glutinosa* populations within the River Tay catchment in the eastern Highlands of Scotland. Key to the approach this study took was to use the differing dispersal mechanisms of *A. glutinosa*, overland pollen dispersal and river-mediated seed dispersal, to test landscape-based hypotheses. Following genetic characterisation of *A.*

glutinosa populations across the River Tay catchment the following hypotheses were tested:

I. An accumulation of downstream genetic diversity (and upstream paucity) will be observed as a consequence unidirectional downstream dispersal of *A. glutinosa* seed.

II. Populations located on the same river will be more genetically similar to each other, due to increased landscape connectivity via rivers, than to populations located on different rivers where connectivity between populations via river flow is not possible.

III. The effect of isolation by distance (IBD) on pairwise population differentiation will differ between overland Euclidean distance and hydrological distance along rivers.

4.3 Materials and methods

4.3.1 Study site

Sampling of riparian *A. glutinosa* trees took place within the River Tay catchment, situated in the eastern Highlands of Scotland, encompassing an area over 5,000 km^2. The River Tay itself is the longest river in Scotland, flowing 193 km from source to outflow. The catchment has a number of large tributaries, has a predominantly upland catchment, and is characterised by semi-natural floodplains and by gravel-bed channels. The main rivers encompass bare shingle to mixed woodland habitat, including abandoned river channels, with some sites of European conservation importance, designated under the 'Shingle Islands' Special Area of Conservation (SAC) for Annex I priority feature 'alluvial

forests with *Alnus glutinosa* and *Fraxinus excelsior*' (JNCC 2014). Approximately 95 km of the River Tay was sampled for *A. glutinosa*, as well as five of its tributaries including the Rivers Tummel, Braan, Almond, Earn and the Lunan Burn (Figure 4.1, Table 4.1).

4.3.2 Sample collection

In July - August 2011 and 2012 leaf material was collected from 49 *A. glutinosa* populations located across the River Tay catchment, including the River Tummel (n = 12), the River Tay (n = 16), the River Braan (n = 3), the Lunan Burn (n = 4), the River Almond (n = 5), and the River Earn (n = 9) (Figure 4.1). Future reference to the river catchment refers to the inclusion of all populations across all rivers unless otherwise stated.

Sampled populations occurred in a range of habitats including river bank trees, riparian woodland, alder carr, floodplain, and slopes with wet flushes, and ranged between 20 – 260 m above sea level. All populations were located adjacent to a main river channel with the exception of some loch populations along the Lunan Burn and Lochs Earn, Tay and Tummel and three further populations adjacent to the historical river route but now disconnected from the main channel (Figure 4.1 and Table 4.1). Sampled populations were located approximately 5 km apart along the river channel / waterbody except where access was not possible or no *A. glutinosa* was present. The geographical location of each sampled tree was recorded using a Garmin GPSMAP 62s handheld navigator. Leaf material was collected from up to 30 evenly sampled trees within each population, with at least 10 m between each sampled tree. Leaf material was collected from a total of 1,457 adult trees. Leaf samples were immediately placed in silica gel (Chase and Hill 1991) and subsequently stored at room temperature.

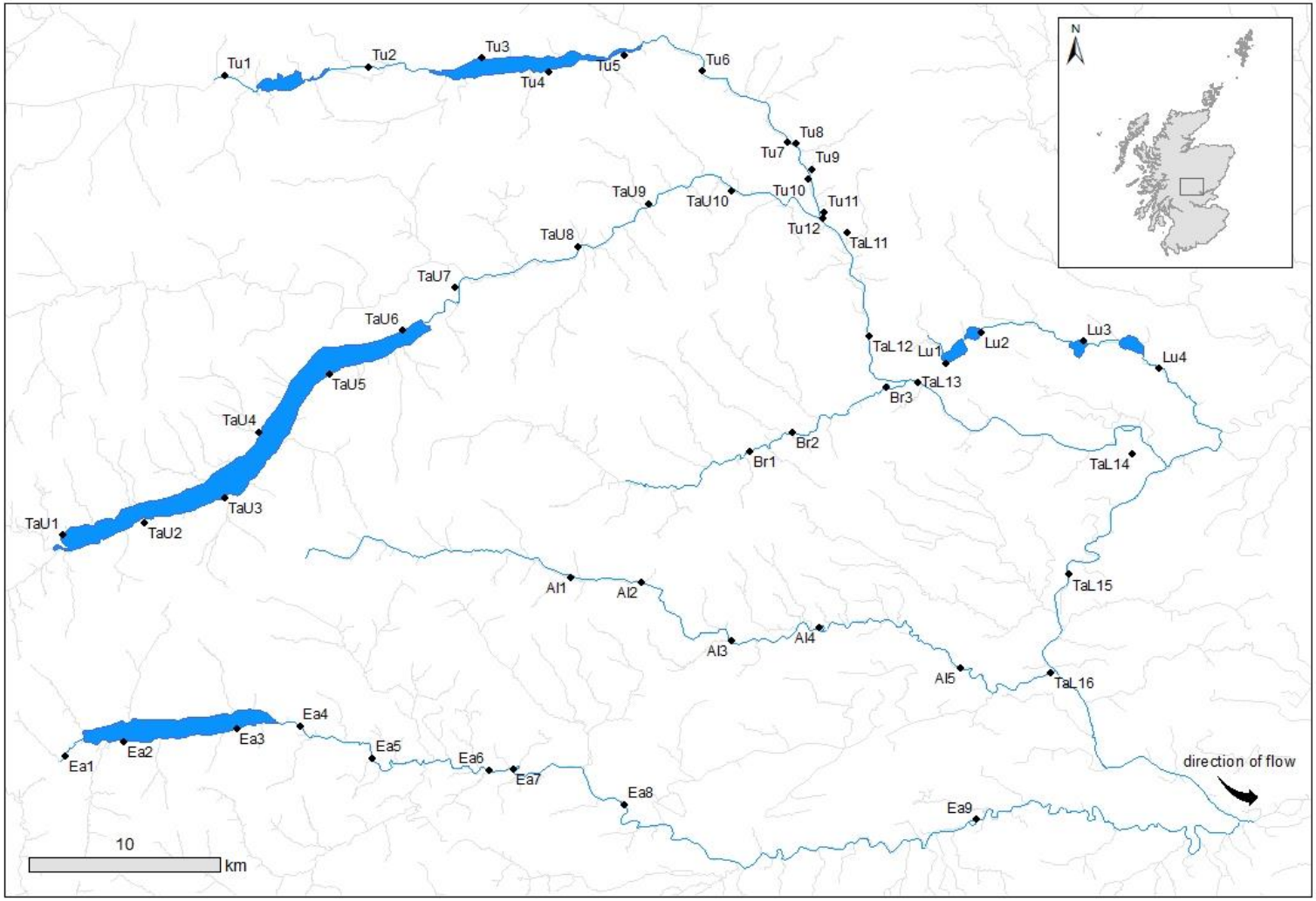

Figure 4.1 Map showing the location of the 49 *A. glutinosa* populations located along the Rivers Tummel, Tay, Braan, Lunan, Almond, and Earn. Site details are given in Table 4.1. Inset map shows location of River Tay catchment in Scotland, UK.

Table 4.1 Details of the 49 *A. glutinosa* populations. 'Tay upper' and 'Tay lower' refer to populations located upstream or downstream respectively of the River Tummel confluence (Figure 4.1). 'Population code' letters refer to the river, and number refers to the upstream – downstream location. Coordinates identify centre of each population, based on individual coordinates.

River (length sampled)	Population name	Population code	Population location	Latitude	Longitude	Altitude (m)
Tummel	Kinloch Rannoch	Tu1	River	56.70182	-4.17930	203
(40 km)	Dalriach	Tu2	River	56.70785	-4.05638	164
	Aldcharmaig	Tu3	Loch	56.71377	-3.95905	144
	Lick Ford	Tu4	Loch	56.70775	-3.90194	144
	Coille Mhòr	Tu5	Loch	56.71702	-3.83807	144
	Faskally	Tu6	River	56.71026	-3.77033	91
	Tomdachoille	Tu7	River	56.67708	-3.69576	70
	Moulinearn	Tu8	River	56.67630	-3.68843	68
	Tynereich Island	Tu9	River	56.66413	-3.67448	64
	Ballinluig	Tu10	River	56.65919	-3.67759	63
	Richard's Island	Tu11	Disconnected	56.64349	-3.66340	60
	Richard's Island	Tu12	River	56.64094	-3.66458	59
Tay_upper	Loch Tay Marshes	TaU1	Loch	56.47781	-4.30409	107
(50 km)	Fiddlers Bay	TaU2	Loch	56.48491	-4.23480	105
	Ardeonaig	TaU3	Loch	56.49832	-4.16741	105
	Lawers	TaU4	Loch	56.53021	-4.13964	105
	Callelochan	TaU5	Loch	56.55918	-4.08163	105
	Dalerb	TaU6	Loch	56.58139	-4.02072	105
	Newhall Bridge	TaU7	River	56.60278	-3.97714	98
	Aberfeldy	TaU8	River	56.62407	-3.87272	85
	Edradynate	TaU9	River	56.64553	-3.81293	83
	Grandtully	TaU10	River	56.65265	-3.74298	64
Tay_lower	Kindallachan	TaL11	Disconnected	56.63415	-3.64365	57
(45 km)	Tom Ban	TaL12	River	56.58452	-3.62276	53
	Dunkeld	TaL13	River	56.56267	-3.58034	51
	Bloody Inches	TaL14	Disconnected	56.53057	-3.39534	34
	Cambusmichael	TaL15	River	56.47167	-3.44782	17
	Denmarkfield	TaL16	River	56.42441	-3.46110	11
Braan	Dullator	Br1	River	56.52775	-3.72123	208
(10 km)	Drumour Bridge	Br2	River	56.53730	-3.68592	178
	Inver	Br3	River	56.56008	-3.60672	62
Lunan	Loch of Lowes	Lu1	Loch	56.57227	-3.55634	105
(22 km)	Loch of Butterstone	Lu2	Loch	56.58763	-3.52638	104
	Loch of Clunie	Lu3	Loch	56.58425	-3.43938	55
	Burnside	Lu4	River	56.57176	-3.37433	46
Almond	Conichan	Al1	River	56.46482	-3.87119	262
(48 km)	Newton Bridge	Al2	River	56.46323	-3.81105	231
	Buchanty	Al3	River	56.43596	-3.73279	170
	Glenalmond	Al4	River	56.44356	-3.65887	129
	Methven Woods	Al5	River	56.42539	-3.53835	75
Earn	Edinchip Wood	Ea1	River	56.37142	-4.29529	133
(80 km)	Coille Criche	Ea2	Loch	56.37945	-4.24585	96
	Ardtrostan Wood	Ea3	Loch	56.38711	-4.15106	96
	Dundern Mill	Ea4	River	56.38933	-4.09733	93
	Drumlochlan	Ea5	River	56.37487	-4.03466	67
	Lennoch	Ea6	River	56.37056	-3.93596	47
	Strowan Wood	Ea7	River	56.37161	-3.91518	46
	Haughs of Pittentian	Ea8	River	56.35568	-3.82009	30
	Dupplin	Ea9	River	56.35265	-3.52115	7

4.3.3 DNA extraction and microsatellite analysis

Genomic DNA from the leaf samples was obtained using the DNeasy 96 Plant Kit (Qiagen) following manufacturer instructions and subsequently stored at -20 ^{o}C until further use. DNA concentration was quantified using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific) and samples were adjusted to 10 ng/ µL for PCR amplification. DNA was amplified using the multiplex of 12 microsatellite markers (Ag01, Ag05, Ag09, Ag10, Ag13, Ag14, Ag20, Ag23, Ag25, Ag27, Ag30, Ag35) of Lepais and Bacles (2011). Multiplex reactions were carried out in a total volume of 5 µL using 1X Type-it Microsatellite PCR Kit (Qiagen) and 0.5 µL of template DNA and performed in a Veriti thermocycler (Applied Biosystems). PCR conditions followed those described in Lepais and Bacles (2011): 5 min denaturation at 95 ^{o}C followed by 30 cycles of 95 ^{o}C for 30 s, 58 ^{o}C for 180 s, and 72 ^{o}C for 30 s, finishing with a final elongation step of 60 ^{o}C for 30 min. Following test amplicon success on 2% agarose gel 1 x TBE electrophoresis, samples were sent to DNA Sequencing and Services (Dundee, UK) for fragment analysis on a Biosystems 3730 capillary sequencer at a 1:50 dilution using GeneScan 500 LIZ size standard (Applied Biosystems). The resulting electropherograms were analysed using GeneMarker v.2.4.0 software (Softgenetics) and the correct assignment of allele size class checked in FlexiBin (Amos *et al*. 2007). Repeat amplification and fragment analysis was undertaken on 77 samples (5% of the total) to assess genotyping error due to allelic dropout (E1) and other genotyping error (E2) in Pedant (Johnson and Haydon 2007). Null alleles across the whole data set were identified using Cervus (Kalinowski *et al*. 2007). Genotypic linkage equilibrium between loci pairs was checked using FSTAT v.2.9.3.2 (Goudet 1995), with significant associations identified by randomly associating genotypes at pairs of loci 53,900 times and using a 5% nominal level after Bonferonni correction.

4.3.4 Genetic diversity

Genetic diversity statistics for each sampling site, including the mean number of alleles per locus (N_A), allelic richness (A_R) (Petit *et al*. 1998), gene diversity (H_E) (Nei 1978) and the inbreeding coefficient (F_{IS}) (Weir and Cockerham 1984) were calculated using SPAGeDi 1.4c (Hardy and Vekemans 2002). Rarefaction analysis of A_R was based on 32 gene copies and significance of F_{IS} values were obtained following 10,000 permutations of gene copies within individuals relative to each population. ADZE 1.0 (Szpiech *et al*. 2008) was used to obtain mean private allelic richness.

Tests based on permutation procedures were carried out to test for differences among loch, river and disconnected populations for allelic richness, observed heterozygosity, gene diversity, and F_{IS} using FSTAT v.2.9.3.2. Differences were checked within each of the three main rivers (Earn, Tay, and Tummel) and across the whole catchment (i.e. populations grouped based on location - loch, river or disconnected). Differences between rivers were also checked by grouping populations by river (i.e. populations grouped based on which river they were associated with).

4.3.5 Testing the unidirectional diversity hypothesis

Genetic diversity parameters A_R and H_E, calculated for each population, were regressed against the distance each population was from the most upstream population along the length of each river. Assuming an accumulation of downstream genetic diversity under the unidirectional diversity hypothesis it can be expected that private alleles would be more likely to occur in upstream populations. Private allelic richness for each population

was therefore regressed against the distance of each population, along the river, from the most upstream population to further test the unidirectional diversity hypothesis.

Regression analyses were undertaken twice for rivers Tummel, Tay, Braan and Lunan. Firstly, including only populations located on each river (excluding disconnected populations), and secondly also including downstream populations on the River Tay_lower (as shown in Figure 4.1). Regressions were implemented using R (R Core Team 2014).

4.3.6 Examining genetic structure

Initial testing for the presence of genetic structure across the river catchment used F_{ST} (Weir and Cockerham 1984) to investigate pairwise population differentiation. Analysis was implemented in FSTAT v.2.9.3.2 (Goudet 1995) to describe pairwise comparisons between all populations (n = 49) and between each river (n = 7, i.e. individuals from the same river were analysed as one population) and significance was evaluated following permutation tests (10,000) and strict Bonferroni correction.

Pairwise population F_{ST} values were then used to test whether river-connected populations were more similar to each other than pairwise populations only connected via overland pollen dispersal. Differences in pairwise F_{ST} values were tested using the Wilcoxon signed-rank test, comparing pairwise river-connected population F_{ST} values (n = 166, excluding Tay_lower populations) and pairwise overland-connected population F_{ST} values (n = 737, excluding Tay_lower populations), implemented using R (R Core Team 2014).

To assess the hierarchical distribution of genetic variation within and among rivers an analysis of molecular variance (AMOVA) was conducted in Arlequin (Excoffier and Lischer 2010) at three hierarchical levels (among rivers, among populations within rivers, and within populations). Statistical significance was tested by nonparametric permutations of individual genotypes among populations and among rivers.

After detecting evidence of pairwise population structure, both between sites and between rivers, genetic clustering methods were used to investigate population structure within the River Tay catchment as well as within each individual river. Using the individual-based Bayesian genetic assignment method in STRUCTURE v.2.3.4 (Pritchard *et al*. 2000; Falush *et al.* 2003) to infer the number of distinct genetic clusters, analysis was first undertaken with no *a priori* geographic location provided, using the admixture model (α, allowed to vary, based on the data and initialised at 1) to account for mixed ancestry among populations, and the correlated allele frequency model (λ = 1), which assumes that the K populations have undergone independent drift away from any ancestral population (Pritchard *et al.* 2010). For each analysis, 10 iterations were run for each K = 1 to K = 8 (River Tay catchment) and K = 1 to K = 4 (each individual river). Each run was composed of a burn-in of 100,000 followed by 200,000 Markov Chain Monte Carlo (MCMC) iterations, which was sufficient to reach convergence. Due to the low level of population structure detected all analyses were repeated using the LOCPRIOR model (coding each population as a different integer), with the aim of improving STRUCTURE performance (Hubisz *et al.* 2009). All outputs were evaluated by first pooling parameter estimates for each run to identify the lowest mean likelihood L(K) and variance per K

value, implemented in STRUCTURE HARVESTER (Earl and vonHoldt 2012); and second, by inspecting the assignment of individuals to populations for the most appropriate value of K (Pritchard *et al*. 2010).

Validation of clustering-based analysis, such as STRUCTURE, is important (Guillot *et al.* 2009); therefore further analysis to investigate the presence of genetic structure was undertaken using Geneland v.4.0.4 (Guillot *et al*. 2005a; Guillot *et al*. 2005b), another Bayesian clustering program. As recommended, analysis was started using the uncorrelated allele frequency model (Geneland Development Group 2012), as setting K as an unknown in the correlated model can lead to an overestimation of K (Guillot *et al*. 2014). Using the *mcmc* function each analysis was based on 1,000,000 Markov Chain Monte-Carlo (MCMC) iterations, thinning set to 1,000, and a burn-in of 200 for each value of K = 1 to K = 5. Runs were performed 10 times for each model to compare average posterior probabilities for each value of K. As the correlated model is better at detecting structure in the case of low differentiation (Guillot 2008) each analyses was then re-run using the correlated model, fixing K at the value obtained from the uncorrelated frequency model. Analysis was completed for each individual river, and for the whole catchment.

Finally, because MCMC convergence was not obtained at the catchment scale using Geneland, further validation was sought using a discriminant analysis of principal components (DAPC) (Jombart *et al*. 2010), implemented in the R package Adegenet (Jombart 2008). As an alternative to Baysian-based analysis, DAPC is a multivariate method that uses sequential K-means and model selection to infer genetic clusters by

first transforming the data using principal component analysis (PCA), and then undertaking discriminant analysis (DA) where differentiation among groups is maximised and variation within groups minimised. This process offers a good alternative to Bayesian-based analysis, notably by not assuming any population genetic model, such as Hardy-Weinberg equilibrium or linkage equilibrium (as assumed in STRUCTURE and Geneland) (Pritchard *et al.* 2000; Jombart *et al.* 2010). The *find.clusters* function was used to transform the genetic data using a PCA, keeping all principle components to maximise the variation. The best number of clusters K was chosen interactively, as recommended in the Adegenet user guide, by looking at both the lowest Bayesian Information Criterion and the difference in slope from the optimal K (as observed in stepping stone models) (Jombart *et al.* 2014). For DA analysis, the *dapc* function was applied to describe the diversity between the (now pre-defined) clusters K. Each DA analysis was run twice, once using 40 PCs and then again, using fewer PCs to account for the trade-off between power of discrimination and over-fitting, following use of the *optim.a.score* function. In both DA analyses all discriminant functions were retained (equal to the best number of clusters, K, minus 1). Finally, the membership probability of each individual to each cluster was computed.

4.3.7 Isolation by distance

The Mantel test (Mantel 1967) was used to test if genetic differentiation between populations of *A. glutinosa* were correlated with spatial distances. Two spatial distances were used, Euclidean and hydrological distance. Hydrological distance, the distance between two populations with measurement restricted to the river network, was calculated to take account of whether populations were connected by downstream river

flow or not. For flow-connected populations, hydrological distance was calculated by measuring distance along the river between populations. Where populations were not flow-connected (i.e. on different rivers) hydrological distance was measured to a shared downstream confluence. Loch populations were considered flow connected and the distance between them calculated as the shortest, straight line distance across the loch, from one population to another. The two spatial distance measures were used to test the effect of distance, for each dispersal mechanism, on genetic differentiation; Euclidean distance representing overland wind dispersal of pollen and hydrological distance representing the dispersal of seed via the river catchment. Correlation between Euclidean and hydrological distance was tested and Mantel tests were performed to test the effect of isolation by distance (IBD), via both overland and river dispersal, on population genetic differentiation, firstly across population pairs within each river and then across population pairs within the river catchment.

IBD models were tested using the *mantel* function in the R package Vegan (Oksanen *et al.* 2014) using 10,000 permutations. Genetic differentiation was quantified using multiallelic pairwise F_{ST} values following Weir and Cockerham (1984), as calculated in FSTAT v.2.9.3.2 (Goudet 1995), and unmodified and log-transformed spatial distances were used. To assess the presence and intensity of IBD at the catchment scale the maximum F_{ST}-spatial distance correlation was investigated by repeated analysis on three subsets of population comparisons (van Strien *et al.* 2015), each subset limited to a maximum pairwise hydrological distance of 75km, 50km and 25km. Finally, to compare hydrological and Euclidean distance models partial Mantel tests were implemented in the R package Vegan with significance assessed following 10,000 permutations.

4.4 Results

4.4.1 DNA extraction and microsatellite analysis

A total of 1457 *A. glutinosa* individuals were genotyped at 12 microsatellite loci; 27 individuals failed to amplify at >6 loci and were excluded from further analysis. Based on 1430 samples from 49 populations loci Ag14 showed significant presence of null alleles, as previously reported (Lepais and Bacles 2011), and was removed from subsequent analysis. The remaining 11 loci showed 3.40% missing data (range 0.00% - 1.47%), and repeated blind genotyping on 5% of the total sample gave very low genotyping errors with a mean allelic dropout (E1) probability across loci of 0.17% (range 0.00% - 0.80%) and a probability of other genotyping (stochastic) error of 0.25% (range 0.00% - 0.76%) across loci. All loci pairs, within each population and across all populations, were found to be in linkage equilibrium.

4.4.2 Genetic diversity

A total of 108 alleles were revealed across the 11 loci, with an average of 9.8 (range 4 – 19) alleles per locus. At the population level, the mean number of alleles per locus (N_A) was 6.26 (range 5.09 – 7.00), with a mean allelic richness (A_R) of 5.43 (range 4.61 – 6.18), and gene diversity corrected for sample size (H_E) of 0.631 (range 0.566 – 0.671) (Table 4.2). Positive F_{IS} values indicated a significant departure from Hardy-Weinberg genotypic proportions in three populations Tu3, TaU4, and TaL14, indicating heterozygote deficiency (Table 4.2). Ten private alleles were revealed, in nine different populations (Tu3, Tu6, TaU2 (x2), TaU6, Br1, Lu4, Al4, Al5, Ea7), and private allelic richness (A_P) ranged from 0.00 – 0.077 (Table 4.2).

Table 4.2 Multilocus genetic diversity at the within population level: N, number of samples; N_A, mean number alleles per locus; A_R, allelic richness; A_P, private allelic richness; H_E, gene diversity; and F_{IS}, inbreeding coefficient. *$p < 0.05$, **$p < 0.01$.

Population	N	N_A	A_R	A_P	H_E	F_{IS}
Tu1	29	5.82	5.32	0.002	0.647	0.011
Tu2	28	6.09	5.43	0.002	0.632	-0.043
Tu3	30	6.09	5.30	0.024	0.603	0.083**
Tu4	28	6.18	5.48	0.000	0.626	0.012
Tu5	29	6.09	5.41	0.000	0.613	0.039
Tu6	29	6.36	5.66	0.042	0.636	0.048
Tu7	30	6.36	5.62	0.001	0.617	-0.042
Tu8	30	6.91	5.88	0.018	0.639	-0.048
Tu9	30	5.91	5.28	0.000	0.648	-0.021
Tu10	30	6.55	5.64	0.000	0.642	-0.034
Tu11	30	6.27	5.51	0.002	0.654	-0.017
Tu12	30	6.55	5.75	0.000	0.661	-0.032
TaU1	30	6.64	5.62	0.018	0.620	0.032
TaU2	30	6.45	5.35	0.077	0.590	-0.028
TaU3	29	6.64	5.79	0.001	0.649	-0.023
TaU4	30	6.18	5.56	0.003	0.658	0.069*
TaU5	30	6.36	5.64	0.018	0.618	-0.005
TaU6	30	6.45	5.59	0.036	0.646	-0.006
TaU7	30	5.64	5.03	0.018	0.597	-0.011
TaU8	27	6.64	5.99	0.001	0.624	0.040
TaU9	26	7.00	6.24	0.036	0.640	0.006
TaU10	29	6.64	5.84	0.001	0.643	-0.006
TaL11	30	6.64	5.79	0.012	0.648	-0.039
TaL12	30	5.91	5.20	0.003	0.603	-0.004
TaL13	30	5.82	5.26	0.000	0.626	-0.004
TaL14	30	5.91	5.28	0.001	0.608	0.068*
TaL15	29	6.45	5.48	0.000	0.615	-0.010
TaL16	16	6.18	6.18	0.003	0.671	-0.035
Br1	30	6.45	5.69	0.027	0.634	0.033
Br2	30	6.18	5.62	0.001	0.658	-0.033
Br3	30	5.91	5.25	0.001	0.614	0.013
Lu1	29	6.00	5.37	0.000	0.621	0.015
Lu2	30	5.09	4.61	0.000	0.566	-0.028
Lu3	30	6.45	5.55	0.002	0.625	0.016
Lu4	28	5.91	5.31	0.024	0.635	-0.049
Al1	28	5.73	5.17	0.000	0.629	-0.038
Al2	30	6.18	5.31	0.001	0.616	-0.034
Al3	29	6.27	5.63	0.000	0.651	-0.002
Al4	29	6.36	5.48	0.024	0.629	-0.018
Al5	30	6.09	5.40	0.023	0.635	0.012
Ea1	29	5.91	5.37	0.001	0.639	-0.011
Ea2	30	6.36	5.64	0.018	0.665	0.018
Ea3	30	6.64	5.77	0.018	0.619	0.030
Ea4	30	6.55	5.84	0.001	0.651	0.059
Ea5	30	5.82	5.19	0.000	0.604	0.009
Ea6	30	6.55	5.80	0.001	0.647	-0.036
Ea7	30	6.82	5.79	0.024	0.608	-0.012
Ea8	30	6.45	5.71	0.003	0.648	0.037
Ea9	29	6.27	0.64	0.003	0.637	0.041

Significant differences between loch and river populations were found for observed heterozygosity and F_{IS} on the River Tummel, but no other significant differences were found either within rivers, across the river catchment, or between rivers (Table 4.3).

Table 4.3 *P* values of permutation tests for allelic richness (A_R), observed heterozygosity (H_O), expected heterozygosity (H_E) and inbreeding coefficient (F_{IS}); between loch (L), river (R) and disconnected (D) populations within the Rivers Earn, Tay and Tummel, and across the river catchment; and between rivers within the river catchment. *P* values based on 10,000 permutations.

	Earn			**Tay**			**Tummel**			**Catchment**			**Catchment**						
	L	**R**	**D**	**L**	**R**	**D**	**L**	**R**	**D**	**L**	**R**	**D**	**Al**	**Br**	**Ea**	**Lu**	**TaU**	**TaL**	**Tu**
N	2	7	0	6	8	2	3	8	0	14	32	3	5	3	9	4	10	6	12
A_R	0.643			0.838			0.367			0.823			0.165						
H_O	0.643			0.959			0.001***			0.260			0.872						
H_E	0.974			0.922			0.067			0.436			0.435						
F_{IS}	0.621			0.835			0.003**			0.429			0.777						

**$p < 0.01$, ***$p < 0.001$.

4.4.3 Testing the unidirectional diversity hypothesis

Linear regression analysis to test the unidirectional diversity hypothesis revealed a significant positive relationship between allelic richness and the position of populations along the River Tummel (but not for River Tummel as well as downstream River Tay populations) ($R^2 = 0.34$, $p = 0.046$). However, no significant relationship was found if (upstream) Loch Tummel populations were excluded from analysis. No significant correlations with allelic richness were revealed for other rivers. No rivers showed a significant correlation between expected heterozygosity, or private allelic richness, and population position (data not shown).

4.4.4 Genetic structure

Although global F_{ST} was low, 0.018 across populations and 0.007 across rivers, there was evidence of significant genetic structure between 429 (out of a possible 1177) pairwise populations (Table 4.4).

Table 4.4 Pairwise F_{ST} population differentiation estimates with significance (above the diagonal).

	Tu1	Tu2	Tu3	Tu4	Tu5	Tu6	Tu7	Tu8	Tu9	Tu10	Tu11	Tu12	TaU1	TaU2	TaU3	TaU4	TaU5	TaU6	TaU7	TaU8	TaU9	TaU10	TaL11	TaL12	TaL13	TaL14	TaL15	TaL16	Br1	Br2	Br3	Lu1	Lu2	Lu3	Lu4	Al1	Al2	Al3	Al4	Al5	Ea1	Ea2	Ea3	Ea4	Ea5	Ea6	Ea7	Ea8	Ea9
Tu1	-	*	**	**	**	NS	*	*	NS	NS	*	*	**	**	*	NS	*	NS	**	NS	NS	NS	NS	**	NS	**	NS	NS	**	**	NS	**	**	*	NS	**	**	*	**	**	NS	*	NS	**	*	*	**	*	**	*
Tu2	0.024	-	*	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	**	NS	NS	**	NS	**	NS	NS	NS	**	**	NS	**	NS	NS	**	**	NS	**	**	NS	*	**	NS	NS	*	NS	**	*	*	NS	NS	NS	NS	NS	NS
Tu3	0.017	0.029	-	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	NS	**	**	NS	NS	**	NS	NS	**	**	*	*	**	NS	*	NS	NS	NS	NS	NS	NS	NS
Tu4	0.023	0.016	0.016	-	NS	**	NS	*	NS	NS	NS	NS	NS	NS	NS	*	NS	NS	**	NS	NS	NS	*	NS	NS	*	NS	NS	*	*	NS	*	**	NS	NS	**	*	*	NS	**	*	NS	NS	*	NS	NS	NS	NS
Tu5	0.020	0.021	0.007	0.011	-	NS	NS	NS	NS	NS	NS	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	**	NS	NS	**	**	NS	**	**	NS	NS	**	*	*	NS	**	NS	NS	NS	NS	NS	NS	NS	NS
Tu6	0.008	0.025	0.003	0.018	0.005	-	NS	NS	NS	NS	NS	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	**	NS	NS	*	NS	NS	**	NS	NS	NS	**	*	*	**	NS	*	NS	NS	NS	NS	NS	NS	NS
Tu7	0.015	0.026	0.019	0.017	0.020	0.017	-	NS	NS	NS	NS	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	NS	*	NS	*	NS	NS	*	NS	NS	**	**	NS	NS	**	*	*	*	NS	**	NS	*	NS	NS	NS	NS	NS
Tu8	0.009	0.018	0.007	0.017	0.003	0.003	0.008	-	NS	NS	NS	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	**	**	NS	**	**	NS	NS	**	NS	NS	*	NS	**	NS	NS	NS	NS	NS	NS	NS
Tu9	0.006	0.012	0.008	0.010	0.007	0.000	0.004	0.001	-	NS	NS	NS	NS	*	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	*	NS	NS	**	**	NS	*	**	NS	NS	**	**	*	*	NS	**	NS	NS	NS	NS	NS	NS	NS
Tu10	0.007	0.017	0.002	0.006	0.003	0.001	0.014	0.000	0.002	-	NS	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	**	NS	NS	**	NS	NS	**	**	NS	NS	**	NS	NS	*	NS	**	NS	NS	NS	NS	NS	NS	NS
Tu11	0.019	0.013	0.014	0.006	0.017	0.006	0.008	0.005	0.000	0.007	-	NS	*	NS	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS	NS	NS	**	NS	NS	**	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	NS	NS
Tu12	0.014	0.024	0.005	0.011	0.009	0.003	0.004	0.002	0.000	0.005	0.005	-	NS	*	NS	NS	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	**	NS	NS	**	**	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS
TaU1	0.014	0.015	0.011	0.011	0.015	0.009	0.016	0.002	0.002	0.008	0.004	0.014	-	NS	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	**	NS	NS	**	**	NS	NS	**	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	NS	NS
TaU2	0.039	0.046	0.050	0.019	0.043	0.041	0.023	0.002	0.028	0.036	0.023	0.031	0.020	-	NS	**	NS	NS	**	NS	NS	*	**	**	**	**	**	**	**	**	**	NS	*	*	**	**	NS	**	**	*	**	NS	NS	NS	NS	NS	NS	NS	NS
TaU3	0.022	0.018	0.014	0.011	0.016	0.014	0.015	0.002	0.006	0.010	0.008	0.011	0.008	0.019	-	NS	NS	NS	**	NS	NS	NS	*	NS	NS	*	NS	NS	*	NS	NS	*	**	NS	NS	**	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS
TaU4	0.006	0.013	0.010	0.015	0.016	0.009	0.013	0.002	0.000	0.007	0.005	0.008	0.002	0.029	0.007	-	NS	NS	**	NS	NS	NS	NS	NS	NS	**	NS	NS	*	NS	NS	NS	**	NS	NS	**	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS
TaU5	0.015	0.036	0.008	0.018	0.012	0.006	0.012	0.002	0.007	0.005	0.016	0.008	0.011	0.020	0.005	0.009	-	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	**	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
TaU6	0.002	0.019	0.002	0.004	0.006	0.003	0.010	0.002	0.000	0.000	0.007	0.004	0.003	0.020	0.003	0.000	0.000	-	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	**	NS	NS	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
TaU7	0.047	0.061	0.060	0.073	0.054	0.037	0.060	0.002	0.046	0.036	0.056	0.052	0.051	0.069	0.044	0.053	0.030	0.045	-	**	*	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	*	**	**	**	**
TaU8	0.009	0.023	0.002	0.007	0.005	0.007	0.012	0.002	0.006	0.002	0.008	0.008	0.003	0.027	0.004	0.007	0.003	0.002	0.048	-	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
TaU9	0.012	0.018	0.014	0.007	0.012	0.004	0.012	0.002	0.000	0.011	0.007	0.011	0.003	0.021	0.000	0.003	0.003	0.000	0.039	0.000	-	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	**	NS	NS	**	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS
TaU10	0.012	0.016	0.011	0.014	0.010	0.007	0.019	0.002	0.006	0.006	0.010	0.009	0.007	0.032	0.013	0.010	0.006	0.004	0.036	0.000	0.000	-	NS	*	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	**	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	*
TaL11	0.011	0.025	0.002	0.009	0.005	0.000	0.009	0.002	0.000	0.000	0.005	0.001	0.011	0.036	0.011	0.008	0.007	0.000	0.053	0.008	0.005	0.008	-	NS	NS	NS	NS	NS	**	*	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
TaL12	0.026	0.038	0.006	0.020	0.019	0.013	0.037	0.002	0.016	0.010	0.023	0.022	0.025	0.053	0.027	0.019	0.024	0.009	0.082	0.021	0.025	0.025	0.010	-	NS	**	NS	NS	**	**	NS	**	**	NS	**	**	*	*	**	NS	NS	NS	NS	NS	*	NS	NS	NS	NS
TaL13	0.005	0.016	0.007	0.011	0.007	0.001	0.002	0.002	0.000	0.001	0.005	0.006	0.009	0.026	0.008	0.006	0.000	0.000	0.034	0.004	0.000	0.004	0.000	0.022	-	NS	NS	NS	**	NS	NS	**	**	NS	NS	**	NS	NS	NS	NS	**	NS	NS	NS	NS	NS	NS	NS	NS	NS
TaL14	0.028	0.045	0.011	0.034	0.019	0.017	0.025	0.002	0.019	0.024	0.029	0.013	0.032	0.051	0.024	0.028	0.017	0.016	0.059	0.016	0.015	0.014	0.010	0.029	0.010	-	NS	*	**	**	**	**	**	NS	NS	**	**	**	**	*	*	**	**	NS	NS	NS	**
TaL15	0.021	0.031	0.002	0.016	0.014	0.006	0.016	0.002	0.008	0.004	0.016	0.003	0.016	0.036	0.003	0.014	0.001	0.002	0.046	0.006	0.007	0.012	0.002	0.019	0.008	0.012	-	NS	NS	NS	NS	NS	NS	**	NS	NS	NS	**	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
TaL16	0.022	0.017	0.023	0.014	0.023	0.021	0.025	0.002	0.016	0.008	0.016	0.014	0.025	0.048	0.012	0.015	0.020	0.009	0.041	0.021	0.018	0.026	0.018	0.035	0.016	0.044	0.017	-	NS	NS	NS	*	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Br1	0.030	0.040	0.022	0.021	0.023	0.025	0.022	0.002	0.017	0.019	0.025	0.013	0.030	0.046	0.016	0.018	0.018	0.014	0.070	0.014	0.020	0.026	0.018	0.026	0.027	0.041	0.012	0.012	-	NS	**	**	**	*	NS	**	**	*	**	**	**	NS	**	NS	NS	NS	NS	NS	**
Br2	0.022	0.020	0.022	0.019	0.022	0.019	0.020	0.002	0.014	0.016	0.015	0.016	0.022	0.049	0.005	0.011	0.023	0.012	0.062	0.014	0.008	0.016	0.015	0.038	0.019	0.035	0.016	0.009	0.012	-	NS	**	**	**	*	**	**	**	**	**	**	**	NS	**	NS	NS	NS	*	**	NS	*
Br3	0.002	0.013	0.004	0.010	0.006	0.006	0.008	0.002	0.001	0.000	0.013	0.008	0.008	0.032	0.010	0.006	0.008	0.000	0.035	0.006	0.005	0.010	0.002	0.023	0.000	0.019	0.010	0.007	0.022	0.018	-	**	**	NS	NS	**	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Lu1	0.041	0.046	0.018	0.022	0.034	0.019	0.032	0.002	0.020	0.023	0.017	0.015	0.033	0.044	0.028	0.033	0.025	0.017	0.078	0.030	0.027	0.024	0.012	0.033	0.025	0.043	0.018	0.035	0.038	0.031	0.031	-	**	**	**	**	**	**	**	**	**	**	**	**	*	NS	**	**
Lu2	0.051	0.045	0.034	0.043	0.029	0.044	0.036	0.002	0.037	0.044	0.037	0.030	0.045	0.062	0.032	0.047	0.039	0.035	0.088	0.025	0.033	0.029	0.035	0.054	0.028	0.024	0.028	0.060	0.059	0.046	0.041	0.053	-	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**	**
Lu3	0.015	0.022	0.010	0.010	0.012	0.017	0.008	0.002	0.010	0.008	0.012	0.008	0.010	0.019	0.006	0.013	0.005	0.000	0.046	0.002	0.011	0.010	0.014	0.024	0.011	0.020	0.007	0.009	0.012	0.018	0.007	0.033	0.029	-	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Lu4	0.009	0.019	0.003	0.011	0.000	0.000	0.008	0.002	0.000	0.003	0.009	0.002	0.007	0.031	0.009	0.007	0.002	0.000	0.045	0.001	0.004	0.001	0.000	0.017	0.000	0.006	0.001	0.017	0.014	0.012	0.003	0.026	0.026	0.000	-	**	**	**	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Al1	0.054	0.048	0.046	0.037	0.040	0.056	0.058	0.002	0.048	0.031	0.045	0.036	0.049	0.067	0.027	0.031	0.042	0.032	0.090	0.038	0.042	0.046	0.043	0.047	0.055	0.064	0.032	0.019	0.025	0.022	0.048	0.060	0.067	0.034	0.043	-	**	*	*	*	**	**	**	**	**	**	*	**
Al2	0.031	0.017	0.039	0.020	0.024	0.026	0.023	0.002	0.020	0.016	0.012	0.028	0.020	0.031	0.014	0.022	0.025	0.019	0.052	0.024	0.019	0.027	0.019	0.039	0.015	0.046	0.030	0.022	0.041	0.029	0.019	0.048	0.053	0.028	0.026	0.041	-	NS	NS	NS	**	**	**	**	*	*	**	**
Al3	0.022	0.026	0.029	0.017	0.022	0.027	0.017	0.002	0.020	0.008	0.020	0.017	0.024	0.028	0.007	0.019	0.020	0.011	0.047	0.021	0.019	0.030	0.016	0.037	0.016	0.044	0.022	0.006	0.021	0.019	0.010	0.038	0.053	0.014	0.022	0.025	0.006	-	NS	NS	**	*	*	*	**	*	**	NS	**
Al4	0.025	0.020	0.035	0.014	0.025	0.030	0.020	0.002	0.020	0.014	0.013	0.026	0.021	0.028	0.017	0.016	0.018	0.011	0.063	0.024	0.015	0.022	0.015	0.033	0.015	0.046	0.028	0.010	0.031	0.022	0.017	0.038	0.059	0.021	0.021	0.032	0.002	0.009	-	NS	**	*	*	**	**	NS	NS	NS	**
Al5	0.021	0.013	0.026	0.014	0.017	0.017	0.011	0.002	0.009	0.012	0.007	0.014	0.012	0.026	0.012	0.009	0.016	0.013	0.055	0.018	0.015	0.018	0.012	0.035	0.012	0.039	0.025	0.012	0.032	0.021	0.009	0.033	0.049	0.020	0.017	0.041	0.001	0.009	0.002	-	**	**	**	NS	NS	NS	NS	**
Ea1	0.021	0.047	0.012	0.047	0.017	0.012	0.038	0.002	0.024	0.014	0.041	0.020	0.028	0.063	0.024	0.022	0.012	0.014	0.037	0.020	0.025	0.020	0.021	0.031	0.021	0.027	0.010	0.031	0.037	0.034	0.023	0.045	0.051	0.025	0.011	0.050	0.053	0.036	0.048	0.043	-	NS	NS	**	NS	NS	**	**	**
Ea2	0.008	0.024	0.009	0.015	0.007	0.004	0.006	0.002	0.000	0.003	0.007	0.000	0.014	0.042	0.008	0.007	0.011	0.005	0.054	0.005	0.008	0.013	0.003	0.020	0.006	0.015	0.008	0.009	0.014	0.011	0.007	0.032	0.037	0.011	0.002	0.036	0.026	0.017	0.019	0.016	0.016	-	NS	NS	NS	NS	NS	NS	NS
Ea3	0.017	0.029	0.011	0.017	0.013	0.009	0.035	0.002	0.011	0.017	0.019	0.017	0.006	0.031	0.010	0.005	0.012	0.004	0.056	0.006	0.007	0.014	0.020	0.020	0.018	0.025	0.015	0.029	0.035	0.024	0.018	0.037	0.042	0.015	0.011	0.039	0.038	0.038	0.034	0.029	0.018	0.015	-	NS	NS	NS	NS	NS	NS
Ea4	0.011	0.021	0.004	0.014	0.012	0.007	0.017	0.002	0.005	0.008	0.008	0.002	0.015	0.047	0.008	0.006	0.014	0.006	0.059	0.000	0.008	0.009	0.006	0.015	0.011	0.019	0.004	0.017	0.010	0.011	0.012	0.029	0.025	0.010	0.006	0.028	0.030	0.025	0.031	0.025	0.020	0.001	0.010	-	NS	NS	NS	NS	NS
Ea5	0.014	0.025	0.004	0.019	0.000	0.000	0.014	0.002	0.005	0.001	0.013	0.009	0.013	0.035	0.010	0.009	0.000	0.005	0.031	0.002	0.009	0.008	0.007	0.017	0.000	0.015	0.003	0.020	0.020	0.021	0.005	0.035	0.032	0.007	0.000	0.045	0.023	0.025	0.029	0.017	0.010	0.010	0.013	0.008	-	NS	NS	NS	NS
Ea6	0.017	0.014	0.002	0.011	0.006	0.000	0.014	0.002	0.000	0.007	0.004	0.003	0.009	0.040	0.006	0.006	0.013	0.006	0.048	0.009	0.003	0.009	0.000	0.019	0.002	0.014	0.008	0.018	0.027	0.019	0.004	0.019	0.030	0.019	0.005	0.052	0.024	0.024	0.026	0.010	0.019	0.003	0.011	0.006	0.008	-	NS	NS
Ea7	0.019	0.026	0.001	0.010	0.009	0.000	0.014	0.002	0.002	0.007	0.011	0.006	0.016	0.035	0.011	0.013	0.004	0.000	0.056	0.010	0.006	0.012	0.000	0.008	0.001	0.012	0.004	0.016	0.016	0.014	0.006	0.015	0.035	0.008	0.000	0.046	0.030	0.026	0.023	0.020	0.027	0.011	0.015	0.012	0.004	0.006	-	NS
Ea8	0.020	0.013	0.002	0.011	0.004	0.007	0.012	0.002	0.000	0.005	0.004	0.001	0.009	0.036	0.002	0.002	0.009	0.006	0.053	0.003	0.004	0.007	0.003	0.013	0.007	0.015	0.002	0.012	0.003	0.008	0.009	0.024	0.028	0.009	0.002	0.026	0.024	0.020	0.021	0.013	0.020	0.001	0.012	0.000	0.003	0.000	0.005	-	NS
Ea9	0.016	0.022	0.012	0.005	0.015	0.009	0.008	0.002	0.001	0.010	0.001	0.002	0.005	0.020	0.013	0.005	0.019	0.003	0.076	0.012	0.013	0.022	0.006	0.016	0.012	0.026	0.010	0.020	0.021	0.019	0.012	0.023	0.039	0.008	0.005	0.040	0.026	0.025	0.024	0.017	0.033	0.005	0.010	0.009	0.015	0.010	0.008	0.007	-

A few populations, notably TaU2, TaU7, Br1, Lu2, Al1 and Ea1 were significantly different to many other populations, both on the same river and on different rivers. Genetic differentiation between all pairwise river F_{ST} values was also found to be significant (Table 4.5). The Wilcoxon signed-rank test reported significant difference between river-connected populations and overland-connected populations (V = 4828, *p* = <0.001), with river-connected populations having a lower mean pairwise F_{ST} value (0.0130) than overland-connected populations (0.0199) (Figure 4.2).

Table 4.5 Pairwise F_{ST} river differentiation estimates with significance (above the diagonal).

	Tummel	**Tay upper**	**Tay lower**	**Braan**	**Lunan**	**Almond**	**Earn**
Tummel	–	***	***	***	***	***	***
Tay upper	0.0045	–	***	***	***	***	***
Tay lower	0.0029	0.0062	–	***	***	***	***
Braan	0.0058	0.0078	0.0069	–	***	***	***
Lunan	0.0054	0.0076	0.0042	0.0101	–	***	***
Almond	0.0135	0.0135	0.0167	0.0120	0.0202	–	***
Earn	0.0021	0.0051	0.0016	0.0058	0.0053	0.0180	–

***p < 0.001.

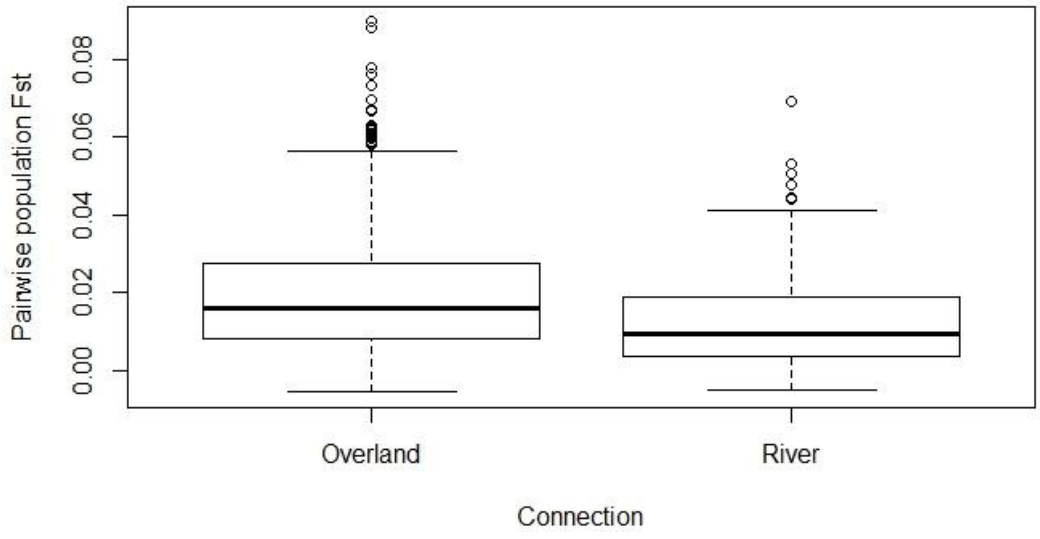

Figure 4.2 Boxplot of mean pairwise population F_{ST} values between river-connected populations overland-connected populations.

AMOVA results indicated significant differences among rivers, among populations within rivers, and within populations, with most (98.1%) of the observed genetic variation attributable to variation within populations (Table 4.6).

Table 4.6 Hierarchical analysis of molecular variance (AMOVA), within populations, within populations on the same river, and among rivers. *P* value based on 1,600 permutations.

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of total variance	*P* value
Among rivers	6	79.32	0.017	0.5	0
Among populations, within rivers	42	265.75	0.049	1.4	0
Within populations	2811	9692.62	3.448	98.1	0

Cluster analysis in STRUCTURE revealed no distinct genetic clusters within any single river or across the river catchment. For each river, and across the whole catchment, the lowest mean log likelihood L(K) and variance values were for K = 1, and inspection of the assignment of individuals showed that the proportion of the sample assigned to each cluster was symmetric and all individuals were admixed (Figure 4.3). Output from analysis using the LOCPRIOR model gave less clear results. Individual river and catchment-wide analysis resulted in K = >1 having the lowest mean log likelihood values. However, convergence was not clear and, in all cases, K = 1 showed little variance in output while each K = >1 output showed high variance between iterations. Inspection of the assignment of individuals to populations revealed admixture in all individuals. Consequently LOCPRIOR output was viewed with some caution as non-convergence may point towards spurious results (Guillot *et al.* 2009).

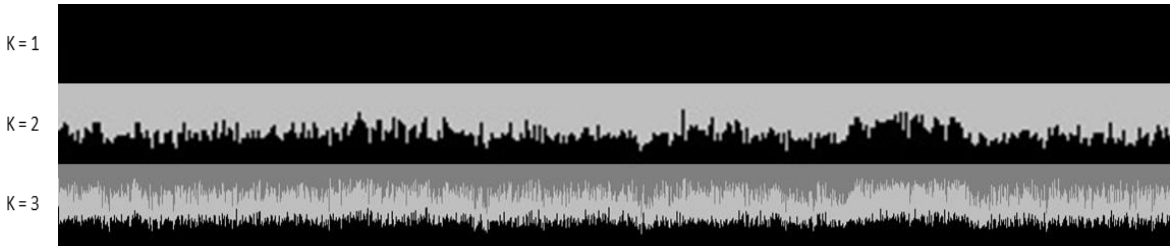

Figure 4.3 Assignment of 1430 individuals to genetic clusters following Bayesian-based clustering analysis implemented in STRUCTURE (showing K = 1 to K = 3 only). Admixture was revealed in all individuals.

Analysis with Geneland was concurrent with the output from STRUCTURE for Rivers Tummel, Almond, Braan, Lunan and Earn with no genetic clustering revealed. However, each of the 10 runs for the River Tay (upper and lower) populations identified two clusters with every run indicating population TaU7 as one cluster and the other 15 populations forming the second cluster (Figure 4.4). Analysis of the catchment-wide data was not possible due to poor MCMC mixing, resulting in a poor convergence, which was considered an effect of the large dataset (>1,000 individuals) (Geneland Development Group 2012).

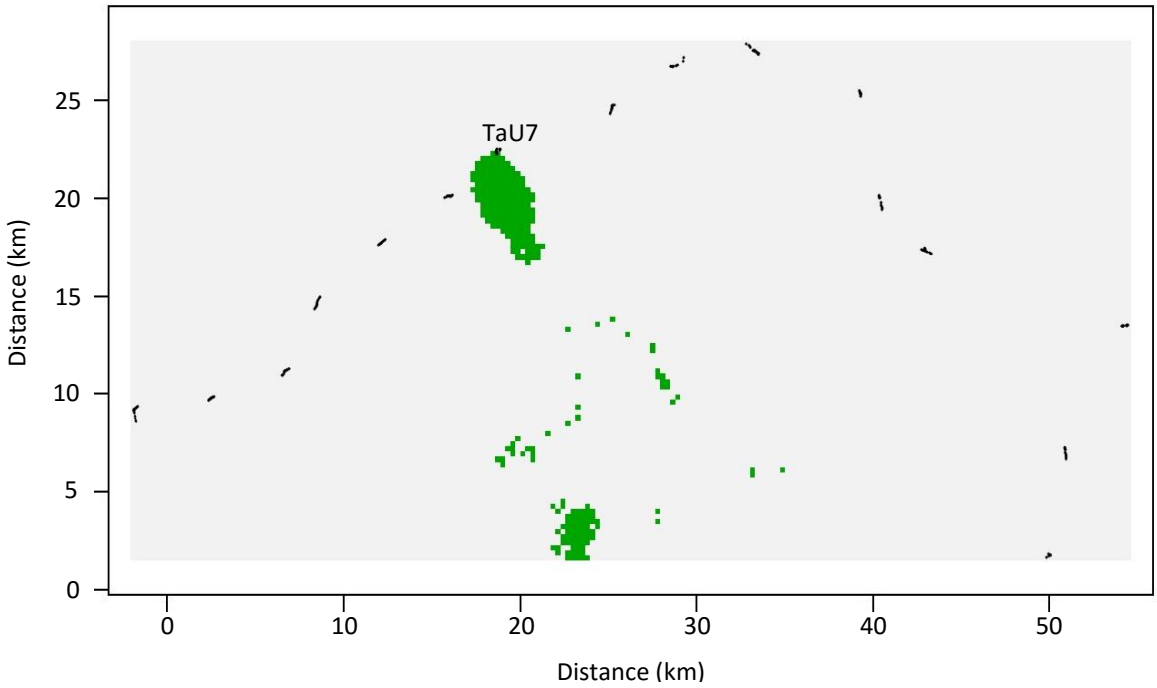

Figure 4.4 Estimated cluster membership of River Tay populations, based on K = 2 Geneland output. Each cluster of black points indicates the location of the 16 sampled populations (as shown in Figure 4.1), all identified as belonging to the same cluster apart from population TaU7 identified as a second cluster, indicated by the green shading.

DAPC analysis was implemented with some caution as selecting genetic clusters based on PCA will not return a solution of K = 1, as found in the Bayesian clustering methods implemented. Analysis of the catchment-wide data revealed the lowest BIC for K = 19, however the difference in slope between K = 2 and K = 19 BIC values was small and so DA analysis was undertaken for both K = 2 and K = 19. The membership probability of each individual, for K = 2 and K = 19, indicated admixture in all individuals (data not shown), concurrent with the STRUCTURE analysis.

4.4.5 Isolation by distance

Hydrological and Euclidean distance were significantly correlated (correlation 0.71, p = <0.001). Mantel tests assessing correlation between either hydrological or Euclidean distance to genetic differentiation (F_{ST}) showed no correlation within individual rivers

(data not shown) however a significant correlation was found at the catchment scale. The Mantel r values were similar for unmodified and modified hydrological and Euclidean distance; results for log-transformed distances are given. Significant correlations between F_{ST} and hydrological distance between populations across the whole catchment, and between populations located up to 75 km and 50 km apart were found but not between populations <25 km apart (Table 4.7). No significant correlations were found between F_{ST} and Euclidean distance (Table 4.7).

Table 4.7 Mantel tests of isolation by distance. Mantel r value and *p* value shown for comparisons of pairwise genetic differentiation (F_{ST}) with log-transformed hydrological distance (HydroDist) and Euclidean distance (EucDist) at different spatial scales.

			Mantel test		**Partial Mantel test**	
Spatial scale	**Median distance (km)**	**Distance matrix**	**r**	***p***	**r**	***p***
<110 km	33.8	log(HydroDist)	0.141	0.011*	0.126	0.024*
	25.1	log(EucDist)	0.069	0.122	-0.028	0.672
<75 km	29.9	log(HydroDist)	0.156	0.004**	0.139	0.009**
	23.8	log(EucDist)	0.077	0.097	-0.031	0.710
<50 km	22.0	log(HydroDist)	0.185	0.009**	0.154	0.026*
	20.2	log(EucDist)	0.104	0.063	-0.005	0.461
<25 km	13.3	log(HydroDist)	0.104	0.217	0.055	0.361
	14.7	log(EucDist)	0.153	0.092	-0.125	0.141

*p< 0.05; **p<0.01.

4.5 Discussion

Unidirectional diversity hypothesis

The unidirectional diversity hypothesis presents a logical argument that an accumulation of genetic diversity will be observed in downstream locations as a consequence of hydrochory. This study found the River Tummel showed weak significance for the downstream accumulation of genetic diversity, with increasing values of (rarefied) allelic richness found over a hydrological distance of 40 km. However, no downstream

accumulation of genetic diversity was observed in the five other study rivers, ranging in distance from 10 – 95 km and, in fact, no significance was found for the River Tummel when analysis was based on river populations only (i.e. upstream loch populations removed from analysis) or when downstream River Tay populations were included in analysis. Empirical evidence shows mixed results for the support of the unidirectional diversity hypothesis (Nilsson *et al*. 2010). Although an increase in genetic diversity in downstream river locations has been found in some studies, e.g. in Potamogeton coloratus in ditches of the Gordano Valley, UK (Gornall *et al*. 1998) in Myricaria laxiflora in the Yangtze river, China (Liu *et al*. 2006), in *Sparganium emersum* in the Niers River in Germany (Pollux *et al*. 2009), and in *Impatiens glandulifera* in the Western Cleddau of Wales and the Tempo/Colebrooke river system of Northern Ireland (Love *et al.* 2013) many studies have found no effect of unidirectional gene flow on the pattern of genetic variation along rivers. A recent meta-analysis of studies reporting genetic structure of riparian and aquatic plant species found no support for the unidirectional dispersal hypothesis (Honnay *et al*. 2010).

A number of reasons have previously been cited to explain an observed lack of increase in downstream genetic diversity including, effective pollen dispersal and other seed dispersal vectors (Nilsson *et al.* 2010), higher seed recruitment opportunities in upstream habitats due to density dependence of recruitment (Honnay *et al.* 2010), and fragmentation of the riparian habitat (Imbert and Lefèvre 2003). Whilst it is possible that wind dispersal of *A. glutinosa* seeds may contribute to some upstream gene flow seed dispersal distance is limited to distances up to 60 m (McVean 1955). However, as a wind-pollinated tree, *A. glutinosa* has high potential for gene flow across large distances (Petit

and Hampe 2006) and, in this case, it seems likely that bidirectional dispersal of pollen over large distances acts as an efficient mechanism of gene flow across the river catchment, counteracting the effect of any downstream accumulation in genetic diversity caused by waterborne seed dispersal. As a comparison, the downstream increase in genetic diversity of *I. glandulifera*, indicative of hydrochorous dispersal of seed, may also be influenced by limited insect pollination dispersal distances (Love *et al.* 2013). The low levels of neutral genetic differentiation observed between *A. glutinosa* populations, ranging from 0.000 to 0.0897 (Table 4.4), as well as the relatively high levels of intrapopulation diversity found (Table 4.2), are also indicative of long distance gene flow between populations.

Genetic structure

Despite indications that a high level of gene flow via pollen dispersal occurs between *A. glutinosa* populations, significantly lower overall F_{ST} values were found between river-connected populations (F_{ST} = 0.0130) compared with overland-connected populations (F_{ST} = 0.0199). Although F_{ST}, as a measure of genetic distance, does not account for differences in seed and pollen dispersal, the lower river-connected F_{ST} values suggest that seed dispersal via rivers, may also influence the genetic structure of *A. glutinosa* populations.

Whilst pairwise estimates of F_{ST} provide insight into historical events, and not current migration (Holsinger and Weir 2009), the difference in F_{ST} values found here suggests connectivity between riparian *A. glutinosa* populations is enhanced by the presence of rivers, at least historically. Lower F_{ST} values between river-connected populations (versus

populations on other rivers) have been reported elsewhere *e.g.* in the wind-pollinated, wind and water dispersed riparian tree *Euptelea pleiospermum* along four rivers within Shennongjia National Nature Reserve, China, inferring that mountain ridges act as a barrier to gene flow between rivers (Wei *et al*. 2013), and in riparian populations of *I. glandulifera* where hydrochorous dispersal of seed is considered to be the primary means of dispersal (Love *et al.* 2013).

However, the significant genetic differentiation found between 36% of pairwise populations across the catchment (Table 4.4), as well as between all rivers (Table 4.5), is indicative, to some extent, of restricted gene flow both across the catchment and within rivers, with even neighbouring populations appearing to be genetically different. Whilst it seems reasonable to expect significant inter-river differentiation due to seed dispersal being restricted to river-connected populations, it is difficult to interpret why some *A. glutinosa* populations were found to be significantly differentiated from nearly all other populations, including neighbouring populations. Some of the highly differentiated populations are the most upstream populations of three rivers (Braan, Almond and Earn) and could conceivably be less connected to other downstream populations and perhaps subject to different influencing factors such as density dependent recruitment. However, overall, the location of the highly differentiated populations appears incidental rather than explanatory as no pattern is apparent across the catchment. It seems, in this case, that care should be taken when interpreting the statistical significance of the genetic differentiation found based on markers with high mutation rates (Hedrick 1999), such as the microsatellites used here.

Although significant variation was found among rivers and among populations, partitioning of the genetic diversity showed most (98.1%) of the observed genetic variation was within populations (Table 4.6), as found in most tree species with extensive gene flow resulting in high heterozygosity (Hamrick *et al*. 1992). Nevertheless, it was with some surprise that based on the concurrent results from two Bayesian-clustering analyses and DAPC analysis, no genetic structure was identified at a river or catchment scale, apart from along the River Tay where STRUCTURE analysis suggested K = 1 and Geneland analysis K = 2 (Figure 4.4). It is interesting to note that the second cluster consisted solely of population TaU7 which was consistently differentiated from all other populations bar nearby upstream TaU5 (Table 4.4), and displayed a lower H_E than average (Table 4.2). It seems unlikely however, that one population would represent a genetic cluster when no other clustering is observed across the catchment and it is concluded that there is no genetic clustering of *A. glutinosa* populations at the spatial scale of the study (maximum Euclidean distance 61 km).

The lack of genetic structure found in this study is different to that reported in many riparian and aquatic shrubs or herbs e.g. in unbranched bur-reed *Sparganium amersum* along the Niers River flowing through Germany and The Netherlands (Pollux *et al*. 2009), in the riparian shrub *Myricaria germanica* in Switzerland (Werth and Scheidegger 2014), and in *I. glandulifera* populations in Wales and Northern Ireland (Love *et al.* 2013). In other wind-pollinated riparian tree species the identification of genetic structure has been attributed to, for example, fragmentation effects on previously widespread populations of *Alnus maritima* (Jones and Gibson 2011), and mountain ridges (highest peak 3105 m) acting as a genetic barrier to gene flow between *E. pleiospermum*

populations located on different rivers within the Shennongjia Mountains of China. In contrast, other studies have also reported no genetic structure. A lack of genetic structuring within and between rivers is considered a consequence of efficient gene flow between *Populus euphratica* populations across an open landscape, as well as the sex ratio in study populations, at least at the scale of the study (maximum Euclidean distance 25 km) (Eusemann *et al*. 2013). Similarly, despite significant population fragmentation, no genetic structure was found within riparian *Fraxinus mandshurica* populations of Maoershan National Forestry Park, China, thought to be a consequence of extensive wind-mediated pollen dispersal and seed dispersal by hydrochory (Hu *et al*. 2010).

Isolation by distance

Despite an apparent lack of genetic structure in *A. glutinosa* populations at the scale studied here, IBD analysis suggests there is a landscape effect on gene flow. Most prior studies of riparian and aquatic plants have not found evidence of IBD (Honnay *et al*. 2010) with only a few studies reporting a significant relationship between genetic and geographic distance (Imbert and Lefèvre 2003; Liu *et al*. 2006; Kondo *et al*. 2009; Werth and Scheidegger 2014). As reported here, IBD was only found in *Ainsliaea faurieana* populations on Yakushima Island, Japan, when all rivers were analysed in combination, but not in any single river (Mitsui *et al*. 2010).

In this study, populations of *A. glutinosa* at hydrological distances >25 km apart are more genetically different the further apart they are along the river yet there is no effect of Euclidean distance. Although the effect of hydrological distance is small, the increase in pairwise population genetic distance as a result of increasing pairwise hydrological

distance remains significant after accounting for Euclidean distance (Table 4.7). An effect of geographical distance on patterns of *A. glutinosa* genetic diversity at a catchment scale has not been found previously. In two recent *A. glutinosa* studies moderate IBD was identified in populations located across 10 European countries, but not at the regional scale across Flanders, Belgium (Cox *et al*. 2011), and no IBD was found between 24 populations across four European countries (De Kort *et al*. 2014). In both of these studies geographic distance was measured as Euclidean distance. This study is thought to be the first to investigate the difference in effect of hydrological distance and Euclidean distance on genetic differentiation between *A. glutinosa* populations as a consequence of river-borne seed dispersal and wind-borne pollen dispersal.

4.6 Conclusion

The tests implemented in this study revealed no evidence for an increase in downstream genetic diversity as a result of hydrochory in any of the six study rivers, suggesting extensive pollen flow at a river or landscape scale. Similarly, across the large, heterogeneous study landscape, no evidence of genetic structure was found, further signifying widespread pollen dispersal across the river catchment. These results initially suggest there is no landscape effect on genetic variation as a consequence of long-distance dispersal of *A. glutinosa* pollen. However, the variable pairwise F_{ST} values and significant genetic differentiation identified between some populations, as well as between rivers, argue against a panmictic effect of pollen dispersal and suggest other factors influence genetic variation of *A. glutinosa*.

When testing landscape effects on the differing dispersal mechanisms of pollen and seed a genetic effect was found. Populations connected by waterborne seed dispersal as well as windborne pollen dispersal were shown to be more genetically similar than populations only connected by windborne pollen dispersal. Despite the evidence for widespread pollen dispersal this test showed that waterborne seed dispersal further increases connectivity between *A. glutinosa* populations. Tests for IBD provided the most interesting result, with no IBD relationship for pollen dispersal found but a significant IBD relationship in seed dispersal at hydrological distances >25 km suggesting that at these distances gene flow via seed dispersal is more constrained by hydrological distance.

The results presented here are of relevance to the management of riparian habitat. As an important keystone species, *A. glutinosa* is crucial to shaping riparian systems. The findings here show that any one *A. glutinosa* population is as important as another population, regardless of location within the river catchment. It is therefore important to manage the overall wider landscape rather than individual populations of *A. glutinosa*.

4.7 Acknowledgements

I thank the landowners involved; Armorel Flint, Roy Flint, and Stewart Lothian for help with fieldwork; Patricia González Diáz and Jose Moreno Villena for help with labwork; and Junhao Sun and Rachael Cooper-Bohannon for help with GIS graphics.

Chapter 5

General discussion

5.1 General discussion

The aim of this thesis was to link gene flow, ecology, and landscape features to examine how they determine the structure of riparian plant populations. The research presented here focusses on *A. glutinosa*, a keystone tree species of European riparian ecosystems. Two key approaches shaped the research undertaken. Firstly, life history attributes of *A. glutinosa* were considered. In particular, the differing dispersal mechanisms of *A. glutinosa* pollen (wind) and seed (water) were linked to land and river features within the study area. Secondly, the study was undertaken at a range of spatial scales, investigating fine-scale within-population effects to between-population landscape scale effects on the structure of *A. glutinosa* populations.

Chapter 2 sought to identify the pattern and distance of *A. glutinosa* pollen and seed dispersal, with the aim of determining the extent of gene flow within and between riparian *A. glutinosa* populations. Parentage analyses identified widespread *A. glutinosa* gene flow. A paternity analysis of *A. glutinosa* seeds revealed a leptokurtic, fat-tailed dispersal curve with most pollen parents (77%) found in the same population as the maternal parent. Although local pollen dispersal was higher than expected under random mating, long distance between-population pollen dispersal was also observed. Maternity analysis of *A. glutinosa* saplings revealed evident differences between wind- and water-dispersed seed. Wind dispersal of seed was limited to distances of less than 100 m, however between-population, hydrochorous seed dispersal was observed up to distances of 2.5 km (within a 6 km river reach). Importantly, most seed-mediated gene flow (57% - 75%) occurred via between-population hydrochorous dispersal rather than within-population wind dispersal. Taken together these results show wind dispersal of pollen

and seed maintains gene flow within populations, however high levels of between-population gene flow also occur, mediated by wind-dispersal of pollen and hydrochorous dispersal of seed.

The extent of gene flow within and between populations influences a range of variables including genetic variation. Analyses in Chapters 2 and 3 revealed relatively high levels of genetic diversity within *A. glutinosa* populations and very little difference in genetic diversity between populations. Genetic diversity measures between populations, and between generations, at the local, river-reach scale were practically identical, typical of tree species (Petit and Hampe 2006), and consistent with the high levels of gene flow described. At the river-catchment scale, measures of genetic diversity between populations were more varied. However, no significant differences were found and no gradient of upstream-downstream genetic diversity was observed, suggesting bi-directional pollen dispersal may counteract the effect of any downstream accumulation in genetic diversity caused by uni-directional, hydrochorous seed dispersal (Chapter 4).

Chapter 4 sought to identify the influence of *A. glutinosa* gene flow at the river catchment scale, and to test landscape-scale effects on genetic connectivity between populations. Across the whole River Tay catchment no genetic clustering of populations was identified, providing further indication of widespread gene flow. Nevertheless, although pollen-mediated gene flow appeared to be unrestricted across the river catchment, evidence for landscape effects on seed-mediated gene flow was apparent. Populations connected by waterborne seed dispersal as well as windborne pollen dispersal were more genetically similar than populations only connected by windborne pollen dispersal, however an

isolation by distance effect was observed for hydrological distances greater than 25 km. Overall these findings indicate that, although gene flow occurs between distant populations, landscape effects on the differing dispersal mechanisms of pollen and seed are evident. Wind-meditated pollen dispersal does not appear to be limited at the spatial scale studied. In contrast, although hydrochorous seed dispersal increases genetic connectivity between populations, this connectivity is limited by the distance between populations. This latter insight is especially relevant because dispersal by seed directly affects the colonisation of new populations.

At the local, within population scale, further riverine landscape – species interactions were identified in Chapter 3. Dendrochronology methods to estimate the age of *A. glutinosa* trees, woodland inventory work, and genetic analyses were combined to provide a unique insight into both the spatial and temporal structure of riparian *A. glutinosa* woodlands. A gradient of *A. glutinosa* tree ages was evident, with young trees located closer to the main river channel and older trees located further away from the main river channel. Despite evident spatio-temporal structuring, no difference in between-generation genetic diversity (Chapters 2 and 3) or temporal genetic structure was apparent (Chapter 3), consistent again with the high levels of gene flow described.

The overall lack of fine-scale spatial genetic structure within *A. glutinosa* populations (Chapters 2 and 3) is thought to further reflect the interplay between the ecology of *A. glutinosa* and features of the riparian / riverine-landscape it occurs within. No seedling regeneration was observed within mature woodland (Chapter 3) and this, combined with the extensive pollen and seed movement (Chapters 2 and 4), as well as the need for

suitable light and water levels and riverine-disturbance to enable seedling establishment (Claessens *et al.* 2010), means it is unlikely that seeds would successfully establish in the vicinity of a parent plant. Thus no or negative fine-scale spatial genetic structure would be observed, as identified in *A. glutinosa* saplings and young trees (Chapters 2 and 3). Nevertheless, spatial genetic structure is not expected to remain constant over time (Kalisz *et al.* 2001; Jones and Hubbell 2006). By utilising the estimated ages of *A. glutinosa* to compare genetic structure between different age cohorts, negative spatial genetic structure was identified in the sapling generation and the cohort of young trees, and positive spatial genetic structure was observed in the oldest tree cohort (Chapter 3). Although the reasons for this difference are discussed in the context of historical population changes (Section 2.5), without further study it remains unclear whether the observed changes in spatial genetic structure are driven by historical factors, local selection or random processes.

5.2 Future research

Further insight into seed-mediated gene flow of *A. glutinosa* and genetic structure at the river catchment scale may be gained with the use of chloroplast microsatellite markers. Chloroplasts are maternally inherited in *A. glutinosa*. Although chloroplast microsatellites typically show less variability, their use alongside the nuclear microsatellites utilised in this thesis may provide additional insight (Provan *et al.* 2001), especially concerning the extent of between-population gene flow via hydrochorous seed dispersal. In addition, because organelle genomes have a lower effective population size, the use of chloroplast microsatellites can reveal more genetic structure, including indicators of founder effects (Provan *et al.* 2001). The use of chloroplast markers at the within-population scale may

therefore provide additional insight into processes resulting in the different patterns of spatial genetic structure between the young and old generations of *A. glutinosa* identified in Chapters 2 and 3.

At the catchment scale, the dendritic network of rivers hosts a range of features, with branches and nodes (i.e. river sections and confluences) arranged in a hierarchical way that dictate the distance and directionality of dispersal (Altermatt 2013). The distinct nature of river catchments including, for example, configuration, changing flow rates, longitudinal connectivity and the location of confluences, all of which influence riparian populations (Naiman and Decamps 1997; Francis 2006), was not directly accounted for within the scope of this thesis. Future work may therefore benefit from the recent development of spatial statistical network models that take account of distinct river features and the spatial autocorrelation among measurements (Peterson and Ver Hoef 2010; Peterson *et al.* 2013; Isaak *et al.* 2014). It may be possible to apply the concepts of these models to riparian populations of *A. glutinosa* based on the hydrochorous seed dispersal and, by incorporating pairwise comparisons of genetic measures, provide a novel insight into river-catchment influences on genetic connectivity between riparian populations not previously undertaken.

5.3 General conclusion

How species interact with the landscapes they occur in is a central question in informing our understanding of how ecosystems function, interact, and respond to change. As linear features, rivers and their associated riparian habitat act as linking features across large landscapes. By taking a landscape genetics approach in this thesis, the influence of river

landscape features on the connectivity and genetic structure of riparian *A. glutinosa* populations was clearly identified. Key findings of this research include (1) gene flow via pollen dispersal is extensive; (2) gene flow via hydrochorous seed dispersal increases connectivity between populations but is limited at distances greater than 25 km; (3) no difference in genetic diversity occurs between populations; (4) riparian *A. glutinosa* woodland shows demographic structuring; (5) despite evident demographic structuring, no fine-scale spatial genetic structure is apparent. Central to the findings gained in this thesis was the incorporation of species ecology, particularly *A. glutinosa* dispersal mechanisms. By directly accounting for the differing dispersal mechanisms when investigating landscape effects on gene flow, the results obtained here highlight the value and importance of incorporating the life history traits of study species in landscape genetics studies.

References

Altermatt (2013) Diversity in riverine metacommunities: a network perspective. *Aquatic Ecology*, **47**, 365-377.

Alvarez-Buylla ER, Chaos A, Pinero D, Garay AA (1996) Demographic genetics of a pioneer tropical tree species: patch dynamics, seed dispersal, and seed banks. *Evolution*, **50**, 1155-1166.

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes*, **7**, 10-14.

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

Ashley MV (2010) Plant parentage, pollination, and dispersal: how DNA microsatellites have altered the landscape. *Critical Reviews in Plant Sciences*, **29**, 148-161.

Bacles CFE, Ennos RA (2008) Paternity analysis of pollen-mediated gene flow for *Fraxinus excelsior* L. in a chronically fragmented landscape. *Heredity*, **101**, 368-380.

Bacles CFE, Jump AS (2011). Taking a tree's perspective on forest fragmentation genetics. *Trends in Plant Science*, **16**, 13-18.

Bacles CFE, Lowe AJ, Ennos RA (2006) Effective seed dispersal across a fragmented landscape. Science, 311, 628-628.

Baguette M, Blanchet S, Legrand D, Stevens VM, Turlure C (2013) Individual Dispersal, landscape connectivity and ecological networks. *Biological Reviews*, **88**, 310-326.

Balian EV, Naiman RJ (2005) Abundance and production of riparian trees in the lowland floodplain of the Queets River, Washington. *Ecosystems*, **8**, 841-861.

Balkenhol N, Gugerli F, Cushman SA, Waits LP, Coulon A, Arntzen JW, Holderegger R, Wagner H, Participants of the Landscape Genetics Research Agenda Workshop 2007 (2009) Identifying future research needs in landscape genetics: where to from here? *Landscape Ecology*, **24**, 455-463.

Banks SC, Cary GJ, Smith AL, Davies ID, Driscoll DA, Gill, AM, Lindenmayer DB, Peakall R (2013) How does ecological disturbance influence genetic diversity. *Trends in Ecology and Evolution*, **28**, 670-679.

Barton K (2015) MuMIn: Multi-model inference. Available at https://CRAN.R-project.org/package=MuMIn.

Bates D, Maechler M, Bolker B and Walker S (2014a) *lme4: Linear mixed-effects models using Eigen and S4*. Available at https://CRAN.R-project.org/package=lme4.

Bates D, Maechler M, Bolker BM and Walker S (2014b) Fitting Linear Mixed-Effects Models using lme4. ArXiv e-print (in press) *Journal of Statistical Software,* available at http://arxiv.org/abs/1406.5823.

Beatty GE, Brown JA, Cassidy EM, Finlay CMV, McKendrick L, Montgomery WI, Reid N, Tosh DG, Provan J (2015) Lack of genetic structure and evidence for long-distance dispersal in ash (*Fraxinus excelsior*) populations under threat from an emergent fungal pathogen: implications for restorative planting. *Tree Genetics & Genomes*, **11**,53. 10.1007/s11295-015-0879-5.

Beatty GE, Montgomery WI, Tosh DG, Provan J (2015) Genetic provenance and best practice woodland management: a case study in native alder (*Alnus glutinosa*). *Tree Genetics & Genomes*, **11**, 92. 10.1007/s11295-015-0919-1.

Bolliger J, Lander T, Balkenhol N (2014) Landscape genetics since 2003: status, challenges and future directions. *Landscape Ecology*, **29**, 361-366.

Bond WJ, Midgley JJ (2001) Ecology of sprouting in woody plants: the persistence niche. *Trends in Ecology and Evolution*, **16**, 45-51.

Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261-3273.

Burnham KP, Anderson DR (2002) *Model selection and multi model inference: a practical information-theoretic approach.* Springer, New York.

Cain ML, Milligan BG, Strand AE (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany*, **87**, 1217-1227.

Chambers FM, Elliott L (1989) Spread and expansion of Alnus Mill. in the British Isles: timing, agencies and possible vectors. *Journal of Biogeography*, **16**, 541-550.

Chase MW, Hills HH (1991) Silica gel: an ideal material for field preservation of leaf samples for SNC studies. *Taxon*, **40**, 215-220.

Chave J (2013) The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecology Letters*, **16**, 4-16.

Chen Z, Li J (2004) Phylogenetics and biogeography of *Alnus* (Betulaceae) inferred rom sequences of nuclear ribosomal DNA region. *International Journal of Plant Sciences*, **165**, 325-335.

Claessens H, Oosterbaan A, Savill P, Rondeux J (2010) A review of the characteristics of black alder (*Alnus glutinosa* (L.) Gaertn.) and their implications for silvicultural practices. *Forestry*, **83**, 163-175.

Convention on Biological Diversity (2015) Convention on Biological Diversity. Available at https://www.cbd.int/.

Corenblit D, Tabacchi E, Steiger J, Gurnell A (2007) Reciprocal interactions and adjustments -between fluvial landforms and vegetation dynamics in river corridors: a review of complementary approaches. *Earth-Science Reviews*, **84**, 56-86.

Council of the European Communities (1992) Council Directive 92/43/EEC. Available at http://www.central2013.eu/fileadmin/user_upload/Downloads/Document_Centre/OP_Resources/HABITAT_DIRECTIVE_92-43-EEC.pdf.

Cox K, Vanden Broek A, Van Calster H, Mergeay J (2011) Temperature-related natural selection in a wind-pollinated tree across regional and continental scales. *Molecular Ecology*, **20**, 2724-2738.

Crawley MJ (2013) *The R Book*. John Wiley & Sons Ltd, United Kingdom.

Cubry P, Gallagher E, O.Connor E, Kellecher CT (2015) Phylogeography and population genetics of black alder (*Alnus glutinosa* (L.) Gaertn.) in Ireland: putting it in a European context. *Tree Genetics & Genomes*, **11**, 1-15.

Dąbrowski MJ, Pilot M, Kruczyk M, Żmihorski M, Umer HM, Gliwicz J (2014) Reliability assessment of null allele detection: inconsistencies between and within different methods. *Molecular Ecology Resources*, **14**, 361-373.

De Knegt HJ, van Langevelde F, Coughenour MB, Skidmore AK, de Boer WF, Heitko IMA, Knox NM, Slotow R, van der Waal C (2010) Spatial autocorrelation and the scaling of species-environment relationships. *Ecology*, **91**, 2455-2465.

De Kort H, Vandepitte K, Bruun HH, Closset-Kopp D, Honnay O, Mergeay J (2014) Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Molecular Ecology*, **23**, 4709-4721.

Douglas AE (1998) Host benefit and the evolution of specialization in symbiosis. *Heredity*, **81**,599-603.

Dow BD, Ashley MV (1998) Factors influencing male mating success in bur oak, *Quercus macrocarpa*. *New Forests*, **15**, 161-180.

Duncan RP (1989) An evaluation of errors in tree age estimates based on increment cores in Kahikatea (*Dacrycarpus dacrydioides*). *New Zealand Natural Sciences*, **16**, 31-37.

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

EDINA Digimap® (2015), OS MasterMap Topography Layer. Available at http://digimap.edina.ac.uk/ [Accessed 4 October 2015].

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250-259.

Epperson BK (1992) Spatial structure of genetic variation within populations of forest trees. *New Forests*, **6**, 257-278.

ESRI Inc. (2014) ArcGIS 10. Available at http://www.esri.com/ [Accessed 4 October 2015].

EUFORGEN (2008) Distribution maps. Available at http://www.euforgen.org/fileadmin/templates/euforgen.org/upload/Documents/Maps/JPG/Alnus_glutinosa.jpg.

Eusemann P, Petzold A, Thevs N, Schnittler M (2013) Growth patterns and genetic structure of *Populus euphratica* Oliv. (Salicaceae) forests in NW China – implications for conservation and management. *Forest Ecology and Management*, **297**, 27-36.

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

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics*, **164**, 1567-1587.

Ford EB (1975*) Ecological Genetics*. Chapman and Hall Ltd, London.

Forest Research (2013)Opportunity mapping for woodland creation to improve water quality and reduce flood risk in the River Tay catchment – a pilot for Scotland. Available at http://www.forestry.gov.uk/pdf/Tay_OM_Report_June13.pdf/$FILE/Tay_OM_Report_June13.pdf.

Francis RA (2006) Allogenic and autogenic influences upon riparian vegetation dynamics. *Area*, **38**, 453-464.

Francis RA, Corenblit D, Edwards PJ (2009) Perspectives on biogeomorphology, ecosystem engineering and self-organisation in island-braided fluvial ecosystems. *Aquatic Sciences*, **71**,290-304.

Fuchs EJ, Hamrick JL (2010) Spatial genetic structure within size classes of the endangered tropical tree *Guaiacum sanctum* (Zygophyllaceae). *American Journal of Botany*, **97**, 1200-1207.

Geneland Development Group (2012) Population genetic and morphometric data analysis using R and the GENELAND program. Available at http://www2.imm.dtu.dk/~gigu/Geneland/.

Gilvear D, Willby N (2006) Channel dynamics and geomorphic variability as controls on gravel bar vegetation; River Tummel, Scotland. *River Research and Applications*, **22**, 457-474.

Gilvear DJ, Winterbottom SJ (1992) Channel change and flood events since 1783 on the regulated River Tay, Scotland: implications for flood hazard management. *Regulated Rivers: Research & Management*, **7**, 247-260.

Gosling PG, McCartan SA, Peace AJ (2009) Seed dormancy and germination characteristics of common alder (*Alnus glutinosa* L.) indicate some potential to adapt to climate change in Britain. *Forestry*, **82**, 573-582.

Gornall RJ, Hollingsworth PM, Preston CD (1998) Evidence for spatial structure and directional gene flow in a population of an aquatic plant, *Potamogeton coloratus*. *Heredity*, **80**, 414-421.

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.

Gregory SV, Swanson FJ, McKee WA, Cummins KW (1991) An ecosystem perspective of riparian zones. *BioScience*, **41**, 540-551.

Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation. The correlated allele frequencies model revisited. *Bioinformatics*, **24**, 2222-2228.

Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261-1280.

Guillot G, Mortier F, Estoup A (2005b) Geneland: a program for landscape genetics. *Molecular Ecology Notes*, **5**, 712-715.

Guillot G, Leblois R, Coulon A, Frantz A (2009) Statistical methods in spatial genetics. *Molecular Ecology*, **18**, 4734-4756.

Guillot G, Santos F, Estoup A (2014) Package 'Geneland': Detection of structure from multilocus genetic data. Available from http://cran.r-project.org/web/packages/Geneland/index.html.

Hampe A (2004) Extensive hydrochory uncouples spatiotemporal patterns of seedfall and seedling recruitment in a 'bird-dispersed' riparian tree. *Journal of Ecology*, **92**, 797-807.

Hampe A, Masri LE, Petit RJ (2010) Origin of spatial genetic structure in an expanding oak population. *Molecular Ecology*, **19**, 459-471.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95-124.

Hardy O, Maggia L, Bandou E, Breyne P, Caron H, Chevallier MH, Doligez A, Dutech C, Kremer A, Latouche-Hallé C, Troispoux V, Veron V, Degen B (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology*, **15**, 559-571.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618-620.

Hardy OJ, Vekemans X (2013) SPAGeDi 1.4 a program for spatial pattern analysis of genetic diversity: user's manual. Available from http://ebe.ulb.ac.be/ebe/SPAGeDi_files/Manual_SPAGeDi%20_1-4.pdf.

Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313-318.

Hoban SM, McCleary TS, Schlarbaum SE, Romero-Severson J (2014) Spatial genetic structure in 21 populations of butternut, a temperate forest tree (*Juglans cinerea* L.), is correlated to spatial arrangement, habitat, and land-use history. *Forest Ecology and Management*, **314**, 50-58.

Hoessaert-McKey M, Valero M, Magda D, Jarry M, Cuguen J, Vernet P (1996) The evolving genetic history of a population of *Lathyrus sylvestris*: evidence from temporal and spatial genetic structure. *Evolution*, **50**, 1808-1821.

Holderegger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends in Plant Sciences*, **15**, 675-683.

Holderegger R, Wagner HH (2006) A brief guide to landscape genetics. *Landscape Ecology*, **21**, 793-796.

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics*, **10**, 639-650.

Honnay O, Jacquemyn H, Nackaerts K, Breyne P, Van Looy K (2010) Patterns of population genetic diversity in riparian and aquatic plant species along rivers. *Journal of Biogeography*, **37**, 1730-1739.

Hu LJ, Uchiyama K, Shen HL, Ide Y (2010) Multiple-scaled spatial genetic structures of *Fraxinus mandshurica* over a riparian-mountain landscape in Northeast China. *Conservation Genetics*, **11**, 77-87.

Hubisz MJ, Falush DS, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322-1332.

Imbert E, Lefèvre F (2003) Dispersal and gene flow of *Populus nigra* (Salicaceae) along a dynamic river system. *Journal of Ecology*, **91**, 447-456.

Isaak DJ, Peterson EE, Ver Hoef JM, Wenger SJ, Falke JA, Torgersen CE, Sowder C, Steel EA, Fortin M-J, Jordan CE, Ruesch AS, Som N, Monestiez P (2014) Applications of spatial statistical network models to stream data. Wiley Interdisciplinary Reviews: Water, **1**,277-294.

Jacquemyn H, Brys R, Vandepitte K, Honnay O, Roldán-Ruiz I (2006) Fine-scale genetic structure of life history stages in the food-deceptive orchid *Orchis purpurea*. *Molecular Ecology*, **15**, 2801-2808.

Johansson ME, Nilsson D (1993) Hydrochory, population dynamics and distribution of the clonal aquatic plant *Ranunculus lingua*. *Journal of Ecology*, **81**, 81-91.

Johnson PCD, Haydon DT (2007) Maximum likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data. *Genetics*, **175**, 827-842.

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403-1405.

Jombart T, Devillard S, Ballous F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jombart T, Collins C, Solymos P, Ahmed I, Calboli F, Cori A (2014) Adegenet: an R package for the exploratory analysis of genetic and genomic data.

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

Jones JM, Gibson JP (2011) Population genetic diversity and structure within and among disjunct populations of *Alnus maritima* (seaside alder) using microsatellites. *Conservation Genetics*, **12**, 1003-1013.

Jones FA, Hubbell SP (2006) Demographic spatial genetic structure of the Neotropical tree, *Jacaranda copaia*. *Molecular Ecology*, **15**, 3205-3217.

JNCC (2014) River Tay. Joint Nature Conservation Committee, Peterborough, England. Available at http://jncc.defra.gov.uk/ProtectedSites/SACselection/sac.asp?EUCode=UK0030312 (accessed September 2014).

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **16**, 1099-1106.

Kalisz S, Nason JD, Hanzawa FM, Tonsor SJ (2001) Spatial population genetic structure in *Trillium grandiflorum*: the roles of dispersal, mating, history, and selection. *Evolution*, **55**, 1560-1568.

Keruzoré AA (2012) *Aquatic vegetation processes in a floodplain-river system and the influence of lateral dynamics and connectivity*. Ph.D thesis, University of Stirling, Scotland. Available at https://dspace.stir.ac.uk/bitstream/1893/7677/1/A.Keruzore.PhD.2012.pdf.

Keruzoré AA, Willby NJ, Gilvear DJ (2012) The role of lateral connectivity in the maintenance of macrophyte diversity and production in large rivers. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **23**, 301-315.

Koch M, Hadfield JD, Sefc KM, Sturmbauer C (2008) Pedigree reconstruction in wild cichlid fish populations. *Molecular Ecology*, **17**, 4500-4511.

Kondo T, Nakagoshi N, Isagi Y (2009) Shaping of genetic strucutre along pleistocene and modern river systems in the hydrochorous riparian azalea, Rhododendron ripense (Ericaceae). *American Journal of Botany*, **96**, 1532-1543.

Koop H (1987) Vegetative reproduction of trees in some European natural forests. *Vegetatio*, **72**, 103-110.

Lande R (1988) Genetics and demography in biological conservation. *Science*, **241**, 1455-1460.

Larsson LA (2003a) CooRecorder: image co-ordinate recording program. Available at http://www.cybis.se.

Larsson LA (2003b) CDendro: Cybis Dendro dating program. Available at http://www.cybis.se.

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659-1673.

Leite FAB, Brandão RL, de Oliveira Buzatti RS, de Lemos-Filho JP, Lovato MB (2014) Fine-scale genetic strucutre of the threatened rosewood *Dalbergia nigra* from the Atlantic Forest: comparing saplings versus adults and small fragment versus continuous forest. *Tree Genetics and Genomes*, **10**, 307-316.

Lepais O, Bacles CFE (2011) De Novo discovery and multiplexed amplification of microsatellite markers for black alder (*Alnus glutinosa*) and related species using SSR-enriched shotgun pyrosequencing. *Journal of Heredity*, **102**, 627-632.

Lepais O, Muller SD, Saad-Limam SB, Benslama M, Rhazi L, Belouahem-Abed D, Daoud-Bouattour A, Gammar AM, Ghrabi-Gammar Z, Bacles C (2013). High genetic diversity and distinctiveness of rear-edge climate relicts maintained by ancient tetraploidisation for *Alnus glutinosa*. *PLoS ONE*, **8**, e75029.

Levin SA (1992) The problem of pattern and scale in ecology. *Ecology*, **73**, 1943-1967.

Levin SA, Muller-Landau HC, Nathan R, Chave J (2003) The ecology and evolution of seed dispersal: a theoretical perspective. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 575-604.

Liu YF, Wang Y, Huang HW (2006) High interpopulation genetic differentiation and unidirectional linear migration patterns in *Myricaria laxiflora* (Tamaricacceae), an

endemic riparian plant in the Three Gorges valley of the Yangtze River. *American Journal of Botany*, **93**, 206-215.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420-1425.

Love HM, Maggs CA, Murray TE, Provan J (2013) Genetic evidence for predominantly hydrochoric gene flow in the invasive riparian plant *Impatiens glandulifera* (Himalayan Balsam). *Annals of Botany*, **112**, 1743-1750.

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

Lowe AJ, Bosheir D, Ward M, Bacles CFE, Navarro C (2005) Genetic resource impacts of habitat loss and defradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity*, **95**, 256-273.

Lowe A, Harris S, Ashton P (2008) *Ecological Genetics*. Blackwell Publishing company, Oxford, UK.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology and Evolution*, **28**, 614-621.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189-197.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209-220. *Trends in Ecology and Evolution*, **18**, 189-197.

Markwith SH, Scanlon MJ (2007). Multiscale analysis of *Hymenocallis coronoaria* (Amaryllidaceae) genetic diversity, genetic structure, and gene movement under the influence of unidirectional stream flow. *American Journal of Botany*, **94**, 151-160.

Marsh TJ, Lees ML (2003) *Hydrological data UK: hydrometric register and statistics 1996 – 2000.* Centre for Ecology and Hydrology, Wallingford.

McVean DN (1953) *Alnus glutinosa* (L.) Gaertn. *Journal of Ecology,* **41**, 447-466.

McVean DN (1955) Ecology of *Alnus glutinosa* (L.) Gaertn. II. Seed distribution and germination. *Journal of Ecology*, **43**, 61-71.

McVean DN (1956) Ecology of *Alnus glutinosa* (L.) Gaertn. III Seedling establishment. *Journal of Ecology*, **44**, 195-218.

Mitsui Y, Isagi Y, Setoguchi H (2010) Multiple spatial scale patterns of genetic diversity in riparian populations of *Ainsliaea faurieana* (Asteraceae) on Yakushima Island, Japan. *American Journal of Botany*, **97**, 101-110.

Morand M-E, Brachet S, Rossignol P, Dufour J, Frascaria-Lacoste N (2002) A generalized heterozygote deficiency assessed with microsatellites in French common ash populations. *Molecular Ecology*, **11**, 377-385.

Naiman RJ, Décamps H (1997) The ecology of interfaces: riparian zones. *Annual Review of Ecology and Systematics*, **28**, 621-658.

Naiman RJ, Décamps H, Pollock M (1993) The role of riparian corridors in maintaining regional biodiversity. *Ecological Applications*, **3**, 209-212.

Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R^2 from generalized linear mixed effect models. Methods in Ecology and Evolution, 4, 133-142.

Nakanishi A, Tomaru N, Yoshimaru H, Manabe T, Yamamoto S (2009) Effects of seed- and pollen-mediated gene dispersal on genetic structure among *Quercus salicina* saplings. *Heredity*, **102**, 182-189.

Nathan R (2001) The challenges of studying dispersal. *Trends in Ecology and Evolution*, **16**, 483-483.

Nathan R (2006) Long-distance dispersal of plants. *Science*, **313**, 786-788

Nathan R, Schurr FM, Spiegel O, Steinitz O, Trakhtenbrot A, Tsoar A (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology and Evolution*, **23**, 638-647.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583-590.

Nilsson C, Brown RL, Jonsson R, Merritt DM (2010) The role of hydrochory in structuring riparian and wetland vegetation. *Biological Reviews*, **85**, 837-858.

Nzunda EF, Griffiths ME, Lawes MJ (2007) Multi-stemmed trees in subtropical coastal dune forest: survival strategy in response to chronic disturbance. *Journal of Vegetation Science*, **18**, 693-700.

Oddou-Muratorio S, Houot M-L, Demesure-Musch B, Austerlitz F (2003) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. I. Evaluating the paternity analysis procedure in continuous populations. *Molecular Ecology*, **12**, 3427-3439.

Ogden J (1981) Dendrochronological studies and the determination of tree ages in the Australian tropics. *Journal of Biogeography*, **8**, 405-420.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2014) Vegan: Community ecology package. R package version 2.2-0. http://CRAN.R-project.org/package=vegan

Parsons H, Gilvear D (2002) Valley floor landscape change following almost 100 years of flood embankment abandonment on a wandering gravel-bed river. *River Research and Applications*, **18**, 461-479.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288-295.

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

Peterson EE, Ver Hoef JM (2010) A mixed-model moving-average approach to geostatistical modeling in stream networks. *Ecology*, **91**, 644-651.

Peterson EE, Ver Hoef JM, Isaak DJ, Falke JA, Fortin M-J, Jordan CE, McNyset K, Monestiez P, Ruesch AS, Sengupta A, Som N, Steel EA, Theobald DM, Torgersen CE, Wenger SJ (2013) Modelling dendritic ecological networks in space: an integrated network perspective. *Ecology Letters*, **16**, 707-719.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin G (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689-701.

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844-855.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 187-214.

Pollux BJA, Luteign A, Van Groenendael JM, Ouborg NJ (2009) Gene flow and genetic structure of the aquatic marcrophyte *Sparganium emersum* in a linear unidirectional river. *Freshwater Biology*, **54**, 64-76.

Pospíškova M, Šálková I (2006) Population structure and parentage analysis of black poplar along the Morava River. *Canadian Journal of Forest Research*, **36**, 1067-1076.

Preston CD, Pearman DA, Dines TD (2002) *New Atlas of the British and Irish Flora*, Oxford University Press, Oxford.

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

Pritchard JK, Wen X, Falush D (2010) Documentation for Structure software: Version 2.3.

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

Qui Y, Lie Y, Kang M, Yi G, Huang H (2013) Spatial and temporal population genetic structure of *Nothotsuga longibracteata* (Pinaceae), a relic conifer species endemic to subtropical China. *Genetics and Molecular Biology*, **36**, 598-607.

R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ritland K (1989) Genetic differentiation, diversity, and inbreeding in the Mountain Monkeyflower (*Mimulus caespitosus*) of the Washington Cascades. *Canadian Journal of Botany*, **67**, 2017-2024.

Rodríguez-González PM, Stella JC, Campelo F, Ferreira MT, Albuquerque A (2010) Subsidy or stress? Tree structure and growth in wetland forests along a hydrological gradient in Southern Europe. *Forest Ecology and Management*, **259**, 2015-2025.

Ronce O (2007) How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 231-253.

Sakai A, Ohsawa T, Ohsawa M (1995) Adaptive significance of sprouting of *Euptelea polyandra*, a deciduous tree growing on steep slopes with shallow soil. *Journal of Plant Research*, **108**, 377-386.

Sato T, Isagi Y, Sakio H, Osumi K, Goto S (2005) Effect of gene flow on spatial genetic structure in the riparian canopy tree *Cercidiphyllum japonicum* revealed by microsatellite analysis. *Heredity*, **96**, 79-84.

Scottish Natural Heritage (2008) Gravel working in the River Tay system – a code of good practice. Available at http://www.snh.gov.uk/publications-data-and-research/publications/search-the-catalogue/publication-detail/?id=1339.

Scottish Natural Heritage (2013) Shingle Islands Site of Special Scientific Interest citation. URL http://gateway.snh.gov.uk/sitelink/siteinfo.jsp?pa_code=1427.

Segelbacher G, Cushman SA, Epperson BK, Fortin MJ, Francois O, Hardy OJ, Holderegger R, Taberlet P, Waits LP, Manel S (2010) Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics*, **11**, 375-385.

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

Sidle RC, Onda Y (2004) Hydrogeomorphology: overview of an emerging science. *Hydrological Processes*, **18**, 597-602.

Slatkin M (1987) Gene flow and the geographical structure of natural populations. *Science*, **236**,787-792.

Slavov GT, Leonardi S, Burczyk J, Adams WT, Strauss SH, Difazio SP (2009) Extensive pollen flow in two ecologically contrasting populations of *Populus trichocarpa*. *Molecular Ecology*, **18**, 357-373.

Sokal RR, Rohlf FJ (1995) *Biometry*. WH Freeman and Company, New York.

Sork VL, Nason J, Campbell DR, Fernandez JF (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology and Evolution*, **14**, 219-224.

Sork VL, Smouse PE (2006) Genetic analysis of landscape connectivity in tree populations. *Landscape Ecology*, **21**, 821-836.

Speer JH (2010) Fundamentals of tree-ring research. University of Arizona Press, Tucson.

Steiner W, Gregorius HR (1999) Incompatibility and pollen competition in *Alnus glutinosa*: evidence from pollination experiments. *Genetica*, **105**, 259-271.

Stoffel M, Wilford DJ (2012) Hydrogeomorphic processes and vegetation: disturbance, process histories, dependencies and interactions. *Earth Surface Processes and Landforms*, **37**, 9-22.

Storfer A, Murphy MA, Evans JS, Golberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128-142.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496-3514.

Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology and Evolution*, **15**, 199-203.

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498-2504.

Taylor PD, Fahrig L, Henein K, Merriam G (1993) Connectivity is a vital element of landscape structure. *Oikos*, **68**, 571-573.

Van Dijk C, Sluimer-Stolk A (1990) An ineffective strain type of *Frankia* in the soil of natural stands of *Alnus glutinosa* (L.) Gaertner. *Plant and Soil*, **127**, 107-121.

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

Van Strien MJ, Holderegger R, Van Heck HJ (2015) Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity*, **114**, 27-37.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921-935.

Vranckx G, Jacquemyn H, Mergeay J, Cox K, Kint V, Muys B, Honnay O (2014) Transmission of genetic variation from the adult generation to naturally established seedling cohorts in small forest stands of pedunculate oak (*Quercus robur* L.). *Forest Ecology and Management*, **312**, 19-27.

Ward JV, Tockner K, Arscott DB, Claret C (2002) Riverine landscape diversity. *Freshwater Biology*, **47**, 517-539.

Wei X, Meng H, Jiang M (2013) Landscape genetic structure of a streamside tree species *Euptelea pleiospermum* (Eupteleaceae): contrasting roles of river valley and mountain ridge. *PLoS ONE*, **8**, e66928.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Werth S, Scheidegger C (2014) Gene flow within and between catchments in the threatened riparian plant *Myricaria germanica*. *PLoS ONE*, 9, e99400.

Wiens JA (2002) Riverine landscapes: taking landscape ecology into the water. *Freshwater Biology*, **47**, 501-515.

Winterbottom SJ (2000) Medium and short-term channel planform changes on the Rivers Tay and Tummel, Scotland. *Geomorphology*, **34**, 195-208.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114-138.

Yazdani R, Muona O, Rudin D, Szmidt AE (1985) Genetic structure of a *Pinus sylvestris* L. seed-tree stand and naturally regererated understory. *Forest Science*, **31**, 430-436.

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

Zurr AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1, 3-14.