A Phylogeographic Study of *Arenaria ciliata* and *Arenaria norvegica* in Ireland and Europe

Emma Howard-Williams

NUI Maynooth

February 2013

A Thesis submitted for the degree of doctor of Philosophy

Supervisor:

Dr. Conor Meade

Co-supervisor:

Dr. Colin Kelleher

Head of Department:

Prof. Paul Moynagh

Declaration

This thesis has not been submitted in whole or in any part to this or any other univerity for any degree and is except where otherwise stated the original work of the author.

Signed _____

Date _____

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Conor Meade for the amazing opportunities that he has given me and for his guidance and support that I have received throughout the project. Special thanks also to my co-supervisor Dr. Colin Kelleher, who has been a great source of advise and support.

Thanks to SFI who provided funding for the PhD project (SFI/08/RFP/E0B1545), and to Synthesys (ES-TAF-1599) who provided funding for the lab work carried out in the Royal Botanic Gardens in Madrid.

Thanks to Xiaodong for being such great company in the lab and on all the mountains that we scaled, through sun, rain and midges!

I would like to say big thanks to Andreas Tribsch, who let me join his wonderful excursions to the Alps, and to Ursula who kindly let me stay with her. Thanks also to the fellow excursionists from the University of Salzburg who helped me with sampling. I would also like to say a big thanks to everyone who made me feel so welcome in the Royal Botanic Gardens in Madrid, and especially Pablo Vargas who made this trip such a valuable and amazing experience. Thanks also to Emilio for technical assistance, and to the other students in the RJB who made me feel so at home, I didnt want to leave! I also wish to say thanks to Kevin Walker who was extremely helpful and contributed greatly to the sampling of *Arenaria norvegica*. I am grateful to the NPWS and Mike Jackson for the background information for the Irish populations. I would like to say thanks to everyone who assisted with sampling: Peter Gruber, Pernille Bronken Eidesen and Lorna Little and to the students who assisted in field and lab work: Matthew Brocklebank, Brian Collogy and Dale O Flaherty.

A big thanks to everyone in the herbarium in the National Botanic Gardens Dublin, who helped me in every way possible and thanks in particular to Christina, Noeleen, Matthew, Emer, and Evelyn. All of you gave me such good advice and made my experience there so enjoyable. I would also like to say thanks to Joan, Edel, Alex and Colette.

I owe so much to my amazing friends who have provided me with huge support especially Rose (also a brilliant housemate and car lender, thank you!), Rebecca, Keira, Maria, Eavan and James. Also a big thanks to Gemma and Aine for being the best housemates in the world, and for putting up with me the last few months! Special thanks to George, who has really been there in thick and thin, literally every step of the way.

To all the gang in the Biology department for the brilliant banter, thanks for giving me such a good laugh! To Eoin, Nessa, Steven, Judy, Lisa, Kathy, Aisling, Grainne, Rebecca and to the medical mycology lab, thanks! I especially want to say thanks to my office buddies, Gail, Eimear and Paddy who gave me great encouragement and advice. Eimear and Gail you have been invaluable, thanks for everything!

Thanks to my family who have always given me great support throughout my studies, especially Paul for his endless lifts and bicycle repairs and to Kathy, Rowena and Siobhan for their constant encouragement. I owe a huge amount to my mother, who has given me such good advice, support and encouragement over the years.

I have to say a gigantic thank you to Chris, I dont think I could have done this without you! For your constant support, patience, advice and proofreading, you have been amazing and I cant thank you enough.

Finally, this thesis is dedicated to my father, who was the best role model, father and friend I could ask for. He was a pinnacle of support and encouragement and truly inspired me and instilled me with an appreciation for the natural world ever since I can remember.

Contents

List of Figures

List of Tables

Abbreviations

AB - Applied Biosystems
AFLP - Amplified Fragment Length Polymorphism
AMOVA - Analysis of Molecular Variance
BIC - Bayesian Information Criterion
bp - base pair
BP - before present (where 1 January 1950 is the origin of the age scale)
BS - bootstrap
BSA - Bovine Serum Albumin
°C - degrees celsius
c - circa
cm - centimetre
concat. - concatenated
cp - chloroplast
CTAB - Cetyltrimethylammonium Bromide
DD - double-distilled
DNA - Deoxyribonucleic acid
dNTPs - Deoxynucleotide Triphosphates
EDTA - Ethylenediamine tetra-acetic acid
g - gram
h - hours
H_20 - water
HCL - Hydrochloric Acid
HPD - highest posterior density
Hz - Hertz
ITS - internal transcribed spacer
kb - kilobase
km - kilometre
L - Litre
LCL - lower calibration limit
LGM - last glacial maximum
μ - Micro

m - mili
M - Mole
MCC - maximum clade credibility
MCMC - Markov chain Monte Carlo
ml - mililitre
ML - Maximum Likelihood
$MgCl_2$ - Magnesium Chloride
MRCA - most recent common ancestor
mt - mitochondrial
My - million years
n - nano
$NaCl_2$ - Sodium Chloride
nDNA - nuclear DNA
NJ - Neighbour-Joining
NN - Neighbour-Net
No. - Number
spp. - species
stdev - standard deviation
temp. - temperature
TE - Tris Ethylenediamine tetra-acetic acid
GPS - Global Positioning System
PCA - Principal Components Analysis
PCoa - Principal Co-ordinates Analysis
PCR - Polymerase Chain Reaction
RPM - Revolutions per minute
SAC - Special Area of Conservation
TAE - Tris-Acetate EDTA buffer
UCL - upper calibration limit
UPGMA - Unweighted Pair Group Method Using Arithmetic Averages
UV - Ultraviloet
v - version
V - Volts

Abstract

The unusual distribution of Irish flora and fauna has intrigued and eluded biogeographers for centuries and remains the subject of ongoing debate. The conventional hypothesis for the postglacial colonization of Ireland across an Irish-British land-bridge has come under increasing scrutiny, with a growing body of evidence suggesting Ireland may have been a refugium during the last glacial cycle. In addition a strong affinity to Iberian populations is evident among much of the islands's native flora and fauna. This study focused on the disjunctly distributed arctic-alpine plant *Arenaria ciliata*, and its close arctic relative, *A. norvegica*, in an effort to characterise and date the earliest links between Ireland and Continental Europe and to investigate the possibility for *in situ* survival of populations of these species in Ireland during the last glacial maximum.

Twenty-nine populations of the target species were sampled throughout their range in Europe. Four separate chloroplast regions were sequenced: *psb*A-*trnH*, *rpl32*-*trnL*, *trnK*-*matK* and *matK*, and the internal transcribed spacer region of the nuclear ribosomal DNA (ITS). The chloroplast and ITS sequences were analyzed using maximum likelihood, haplotype networks, and a molecular clock analysis using Bayesian inference. Amplified fragment length polymorphism (AFLP) markers were generated for all populations sampled using three selective primers and the obtained data were analysed using neighbour-joining analysis, networks, AMOVAs and direct analysis of the fragments.

The results indicate a complex glacial history for these two species, where the consensus of the analysis demonstrates phylogeographic

patterns indicative of Pleistocene glacial survival *in situ*, with evidence for cryptic refugia in Ireland, Rum and Svalbard, and multiple colonization events for *A. ciliata*. In contrast, the phylogeographic patterns obtained for *A. norvegica* are, for the most part, typical of a recent post-glacial expansion, with very low sequence divergence between individuals and populations.

Evidence for *in situ* survival of the Ben Bulben population of *A. ciliata* is presented. This population contains multiple unique haplotypes, many private AFLP fragments, distinct genetic structuring and the molecular clock analysis indicates an ancient divergence for the Ben Bulben population. This indicates an origin for this resident population much earlier than the last glacial maximum, predating the origin of the *A. norvegica* haplotypes. AFLP genotypes show distinct grouping for the Ben Bulben and Svalbard populations. Similarities in the genetic distinctiveness of populations from Rum, Svalbard and Ben Bulben indicate the possiblity that these locations represent cryptic Pleistocene refugia for *A. ciliata*.

Taxonomic distinctions between *A. ciliata* and *A. norvegica* are complex, though, a distinct genetic grouping for individuals of *A. norvegica* is demonstrated, albeit as a monophyletic group within *A. ciliata*. Incongruences between the chloroplast and nuclear ITS sequences confound this issue, and highlight the possibility of hybridisation events between these species. AFLP markers show a distinct clustering for *A. norvegica* genotypes and demonstrate the suitability of AFLP markers in discriminating between closely related taxa. None of the recognized subspecies of the *A. ciliata* complex could be clearly discriminated based on the molecular markers implemented in this study.

This study provides the first evidence for the survival of a terrestrial plant species *in situ* in Ireland throughout the Pleistocene glacial cycles.

1

Introduction

1.1 Postglacial biogeography in Europe

1.1.1 Genetic consequences of the Quaternary glacial cycles

Since the beginning of the Quaternary Period (2.5 My - present), large parts of the world have been dominated by climatic oscillations brought on by fluctuations in the Earth's orbit and resulting in major glacial cycles (Webb and Bartlein, 1992). The ice-sheets began to grow in the Northern Hemisphere at the beginning of the Quaternary approximately 2.5 million years ago (Webb and Bartlein, 1992), however it was only in the last 930-870,000 years that the first extensive ice-sheets developed (Toucanne *et al.*, 2009). During this time, several glacial cycles, each lasting approximately 100,000 years, were interrupted by shorter, warmer periods called interglacials. These lasted approximately 10-20 thousand years (Webb and Bartlein, 1992), and were similar to, or slightly warmer than, the present day climate (Willis and Niklas, 2004). The climate during the cold glacial cycles was very different from today's climate however, with temperatures approximately 10 °C - 25 °C lower than current temperatures (Birks and Willis, 2008). At the full extent of the last glacial maximum (LGM), which occurred 18-21,000 years ago (Webb and Bartlein, 1992), the European ice-sheet extended South to 52 °N

(Hewitt, 2004), including Ireland. Large areas South of 47 °N were in a state of permafrost (Hewitt, 2004), with tundra habitat occurring close to the ice-sheet margins (Birks, 2008).

These major fluctuations in climate resulted in large-scale shifts in species distributions; where some species became extinct in large parts of their range, others survived in refugia (areas to which the ice-sheet did not extend) and then expanded again (Hewitt, 1996). The glacial cycles not only affected the current geographical distribution of populations, but also the amount and distribution of genetic variation and patterns of genetic diversity in plants and animals (Hewitt, 1996; Taberlet *et al.*, 1998). The effects of the glacial cycles and climatic changes are best documented for the last glacial cycle which was at its coldest 18-21,000 years ago, at the last glacial maximum (LGM) (Webb and Bartlein, 1992).

Analysis of the genetics of extant populations can consequently offer us insights into historical events and inferences of processes such as range fragmentation, isolation in refugia, colonization, range expansion and molecular divergence (Comes and Kadereit, 1998). The combination of analyses of current genetics along with the comparison of fossil remains and pollen records can offer powerful insights into past events and previous distributions of species, which were shaped by the glaciation periods of the Quaternary.

The effects of the glacial cycles in Europe in particular have been studied extensively for a range of species e.g. Petit (2002); Schmitt and Haubrich (2008); Taberlet *et al.* (1998). Many studies have focused their attention on the Alps (e.g. Schonswetter *et al.* (2005, 2006)), and these studies have played a major part in the understanding of the shaping of the phylogeography of European species. The Alps have acted as a refuge and a barrier during and after glaciation, and they support a wide range of diverse biota with varied and complex histories (Hewitt, 2004). These vast mountain ranges provide an excellent model to compare the effects of the glacial cycles on current species as paleoenvironmental data is assessable for many regions, and have been studied in parallel with population genetics (Tribsch and Schonswetter, 2003). Comparisons using these methods have identified putative glacial refugia located along the South-Western, Southern, Eastern and Northern border of the Alps and also the presence of refugia on ice free mountaintops, or nunataks (isolated rocky peaks that protrude above

an ice-sheet and were not glaciated (Gugerli and Holderegger, 2001)), present in central Alpine areas (Schonswetter *et al.*, 2005).

Consensus among these studies suggest that, during the maximum extent of the last glacial cycle, approximately 18-21,000 years ago (Webb and Bartlein, 1992) the distribution of many organisms were restricted to refugia in Southern Europe (Taberlet *et al.*, 1998). Figure 1.1 shows the extent of the ice-sheet during the last glacial maximum (LGM) and a summary of some potential refugia in Europe for temperate and cold-adapted species is also depicted. Taberlet *et al.* (1998) inferred putative postglacial migration routes for 10 taxa, demonstrating that Northern regions of Europe were generally colonised from Iberic and Balkanic refugia after the retreat of the ice-sheets (Figure 1.2). The extant remains of these putative refugial populations in Southern Europe are expected to contain high levels of genetic diversity relative to their likely descendant populations (Comes and Kadereit, 1998) and contractions of a species range into refugia would promote genome reorganization (Hewitt, 1996).

Taxa that vastly increased their range Northward however after the melting of the ice-sheets did so from a small subset of their full glacial population (Bennett and Provan, 2008). Those populations which existed at the rapidly expanding Northern limits of the refugial range were probably long distance dispersers which set up colonies before other organisms arrived (Hewitt, 1999, 2000). These founding events would have been repeated many times over a long colonization route, thus leading to the loss of alleles and reduced genomic diversity. This reduction in genetic diversity is to be expected as a result of repeated population bottlenecks and is consistent with recent post-glacial expansion (Hewitt, 1996). Consequently, populations in Northern Europe are predicted to contain reduced genetic diversity as a result of these rapid expansions of population range, which can be observed in the reduction in the number of species, the extent of subspecific division, and allelic variation (Hewitt, 1999).

Although common-source refugia have been found for many locations in Europe, it is now apparent, for the most part, that migration routes are distinct for each species, demonstrating that individual species responded differently to the Quaternary cold periods (Taberlet *et al.*, 1998). In addition to this, the expansion of populations which subsequently meet in a common geographic area

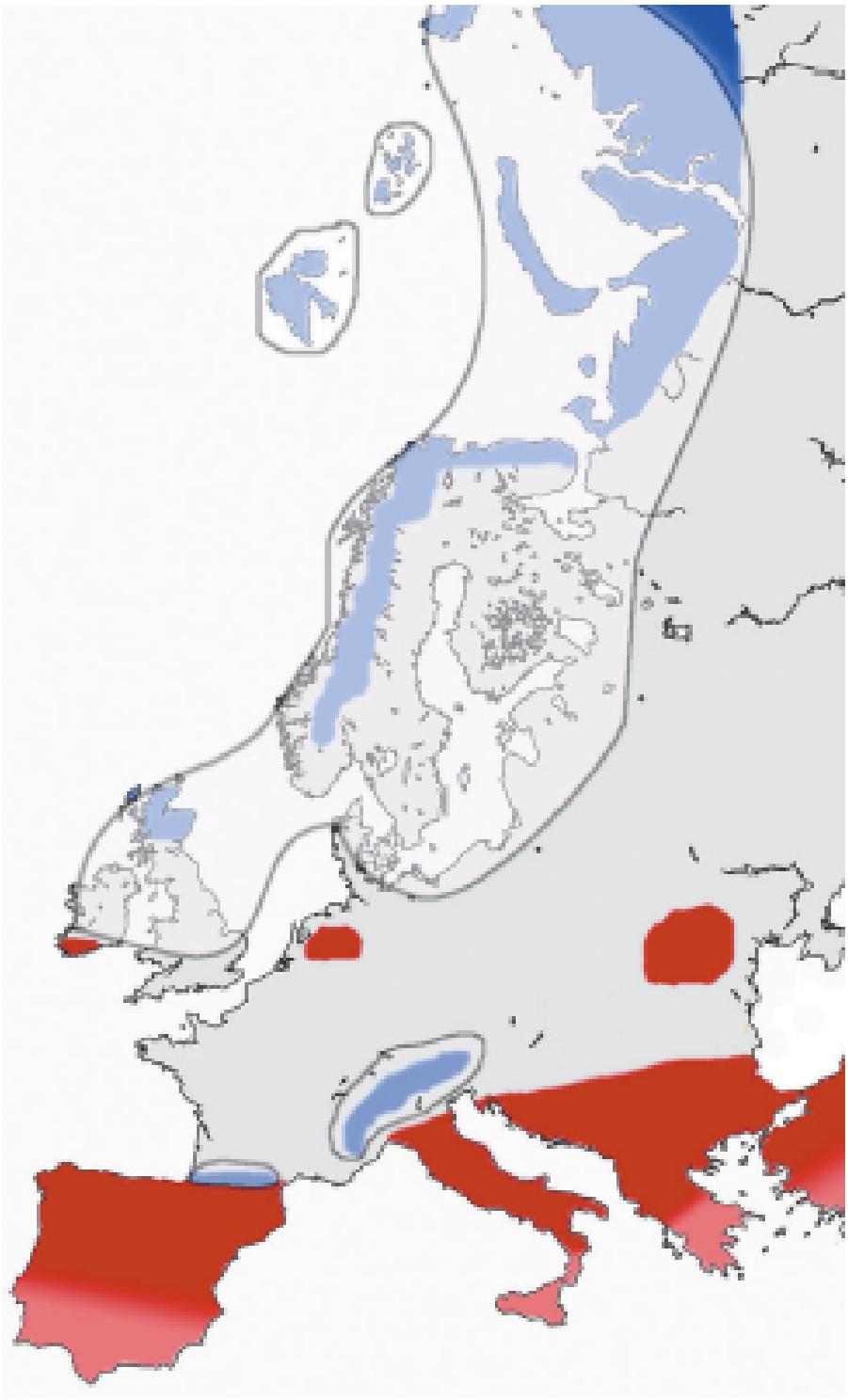


Figure 1.1: Map of Europe modified from Stewart *et al.* (2010) which shows different types of refugia and the limits for the ice-sheet during the LGM. The blue areas indicate interglacial refugia for cold-adapted species, red areas indicate glacial refugia for temperate species where the darker colours represent refugia that were inhabitated throughout at least one cycle. The grey line indicates the extent of the ice-sheet during the LGM.

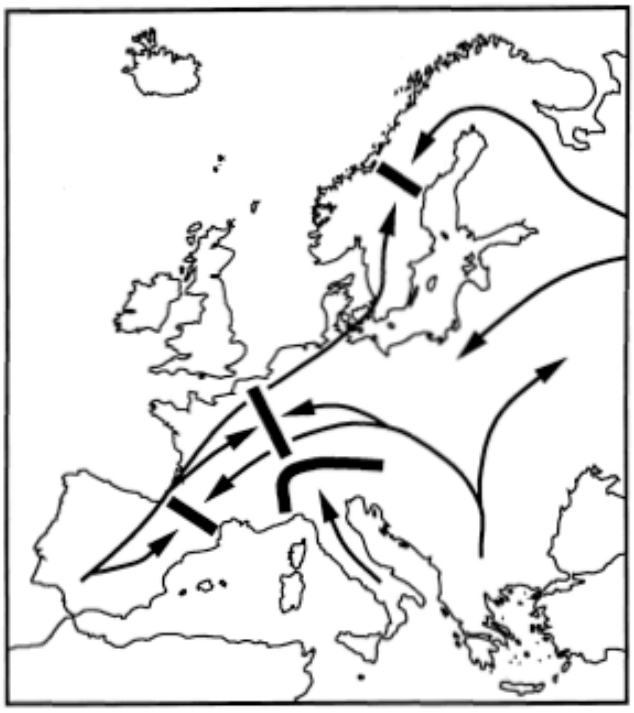


Figure 1.2: Map of Europe from Taberlet *et al.* (1998) showing the main post-glacial colonization routes indicated by the arrows and suture zones indicated by the thick lines.

may lead to the formation of suture-zones. The term suture-zone was introduced by Remington (1968), who described suture-zones as a band, whether narrow or broad, of geographical overlap between major biotic assemblages, including some pairs of species which hybridize in the zone. Phylogeographical and paleobotanical analyses have subsequently demonstrated the presence of suture-zones (Hewitt, 1996; Taberlet *et al.*, 1998) and Figure 1.2 shows some of the main suture-zones inferred by Taberlet *et al.* (1998).

1.1.2 Cryptic refugia

The classical view of postglacial expansion from Southern refugia may be incomplete (Birks and Willis, 2008) and it is clear that each species responded to the Quaternary glacial cycles differently (Taberlet *et al.*, 1998). Variations in this

general pattern of postglacial migration have been demonstrated in phylogeographical and paleobotanical analyses (Birks and Willis, 2008), particularly for cold-adapted species. Each species responds to the glacial cycles differently due to variations in each species adaptation and response to environmental changes (Bhagwat and Willis, 2008; Stewart and Lister, 2001; Stewart *et al.*, 2010).

Arctic-alpine plants which would have occupied large areas of the Northern European plain retreated to higher latitudes and altitudes as the climate warmed during interglacials (Webb, 1983). Cold tolerant species would have maintained large populations widely distributed across large non-glaciated areas of the European mainland during glacial periods (Hewitt, 2001; Stewart *et al.*, 2010) thus potentially displaying little reduction in genetic diversity e.g. Hamill *et al.* (2006). The glacial history of a species is distinct and depends on many factors, including whether it is a temperate or cold-adapted species, its competitive ability, mode of dispersal and generation time as well as different rates of colonization (Taberlet *et al.*, 1998; Willis and Niklas, 2004).

Recent evidence from phylogeographical and paleobotanical analyses is pointing towards the possibility of cryptic Southern refugia (Bhagwat and Willis, 2008; Birks and Willis, 2008; Provan and Bennett, 2008; Stewart and Lister, 2001; Stewart *et al.*, 2010; Teacher *et al.*, 2009; Westergaard *et al.*, 2011). Cryptic refugia, which are situated outside of the expected latitudinal and/or longitudinal range of glacial refugia often consist of areas of climatic stability which differ from the surrounding areas (Stewart and Lister, 2001). The definition of a cryptic Southern refugia is outlined in (Stewart *et al.*, 2010) as "interglacial refugia for cold-adapted species situated at lower latitudes". These refugia facilitate the survival of relict populations of formerly widespread cold-adapted species (Stewart *et al.*, 2010). Phylogeographic studies indicate that some of these refugia are situated within supposed limits of the ice-sheet during the LGM (Provan and Bennett, 2008). Large areas of the Arctic and Subarctic in North-West America and Eastern Siberia are now thought not to have been completely covered by ice-sheets (Abbott, 2000) and areas in North-Eastern Europe appear to contain many small Pleistocene refugia where boreal or cold tolerant species survived (Taberlet *et al.*, 1998; Hewitt, 2004). The possibility of many small Pleistocene refugia has also been suggested for Central and Northern Europe based on genetic

and paleobotanical evidence (Stewart and Lister, 2001). These sheltered regions would have remained relatively climatically stable and are recognized for the UK, Hungary, Ireland and Slovakia (Stewart and Lister, 2001) (e.g. Figure 1.1).

Despite current knowledge on the genetics of refugial populations and post-glacial recolonisation pathways (Abbott, 2000; Hewitt, 1996; Schonswetter *et al.*, 2005; Stehlik, 2003; Taberlet *et al.*, 1998), the exact mechanism by which the assemblage and distribution of the Northern European biota established in the first place is currently unknown.

1.2 Origins of the Irish Flora and Fauna

1.2.1 Ireland during the ice-ages

At the full extent of the LGM approximately 18-21,000 years ago (Webb and Bartlein, 1992), large volumes of accumulated ice reduced sea levels by at least 85 metres (CLIMAP, 1976), producing land bridges in several parts of the world, and increased the land area by circa (c.) 20% (Hewitt, 2004). This reduction of sea level occurred throughout the Pleistocene and sea levels fluctuated depending on the severity of the glaciation. A map demonstrating the extent of ice-sheet cover and the subsequent reduction in sea level, dating from the LGM and during the most extensive glaciation period, the Elsterian glaciation, c. 450,000 years ago, is shown in Figure 1.3, modified from Toucanne *et al.* (2009).

A thick ice-sheet covered the land mass of Ireland and extended onto the adjacent coastal shelf during the LGM, though the extent of the ice-sheet cover, and resulting sea level is still a matter of ongoing debate (Brooks *et al.*, 2008; Clark *et al.*, 2012; Devoy, 1985; Lambeck and Purcell, 2001; O' Cofaigh *et al.*, 2012; Watts, 1977). Some models such as those proposed by Watts (1977) and Brooks *et al.* (2008) show that some areas of Ireland may have remained ice free. Ballantyne *et al.* (2008) believe that the ice-sheet extended beyond the traditionally delineated Southern Irish end Moraine Limit and that models such as those used by Brooks *et al.* (2008) underestimate the thickness of the ice-sheet.

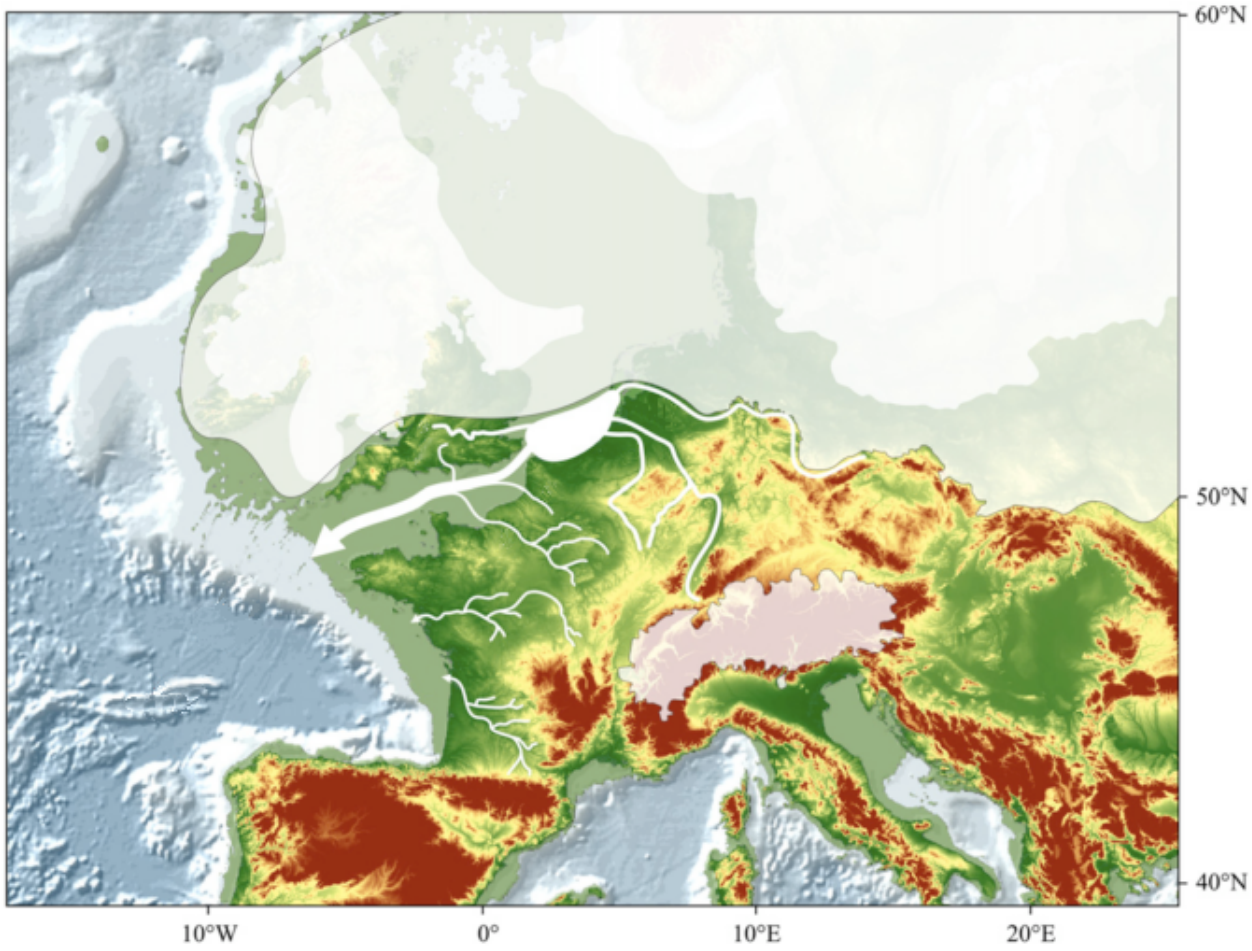


Figure 1.3: Map of Europe modified from Toucanne *et al.* (2009) demonstrating the extent of ice-sheet cover and its effect on sea levels during the Pleistocene cold-periods. The extent of the ice-sheet during the maximum Pleistocene ice-advance from the Elsterian, c. 450,000 years ago is depicted in the shaded white areas, and the white areas over Britain, Ireland and Scandinavia show the extent of the ice-sheet at the last glacial maximum c. 20-30,000 years ago for comparison. The white arrows indicate the main European rivers; the large white arrow is the 'Fleuve Manche' paleoriver. Note the green areas which protruded above sea level during the Elsterian Pleistocene and could indicate potential land bridges.

Recent models indicate Ireland was completely covered by a thick ice-sheet (Clark *et al.*, 2012; O' Cofaigh *et al.*, 2012). Despite this, it is clear from the map shown in Figure 1.3 (Toucanne *et al.*, 2009), that sea levels were significantly reduced for large areas off the North-West and Southern coast of Ireland and along the Celtic sea to the West of France, and that the ice-sheet extent varied depending on the severity of the glacial cycle. The hypothesis of an Irish-British land bridge, however, is difficult to confirm at this time because of the lack of reliable sea level data (Brooks *et al.*, 2008; Lambeck and Purcell, 2001). The extent of the ice-sheet in Ireland is still unresolved and as a result, we cannot ascertain when exactly the land bridge(s) disappeared and Ireland became an island (Mitchell, 2006).

1.2.2 "The Irish Question" and the conventional hypothesis

The unusual biogeography of the Irish flora and fauna, for which the term "The Irish Question" has been coined (Moore, 1987), has eluded and intrigued botanists, biologists and geologists and the origin of the Irish Flora and Fauna has spawned a debate which has been ongoing for centuries.

The conventional hypothesis for the origin of species in Ireland is the Northward migration of species from a British-Continental land bridge, followed by dispersal into Ireland after the retreat of the ice-sheet, via a land bridge across the Irish Sea (Mitchell and Ryan, 2007; Webb, 1983; Wingfield, 1996). This land bridge would have connected Carnsore Point in Ireland to Devon in Britain and crept gradually Northwards, facilitating the migration of plants and animals into Ireland via this route until it was cut off by rising sea levels approx 9,500 years ago (Mitchell and Ryan, 2007). If the organisms that are currently in Ireland today migrated across this land bridge after the last ice-age, European biogeography theory would predict that genetic variation should be reduced in the Irish populations, with the loss of alleles and reduced genomic diversity in current populations (Hewitt, 1999).

The relatively small number of current Irish species (compared to Britain and mainland Europe), certainly fits with the conventional theory of postglacial expansion from Britain. For example, the number of flowering plant species in Ireland is recorded at 815, compared to France at 3,500 and Britain at 1,172 (Webb, 1983). It is not the number of species, however, that has intrigued researchers of the history of Irish Biogeography - it is rather the unusually disjunct distribution of some species that reveal incongruences with this conventional theory of postglacial migration (Praeger, 1932; Synnott, 1981; Webb, 1983).

One of the major problems encountered with the conventional hypothesis of Northward migration into Ireland from Britain is centred around a group of flora and fauna known as the 'Lusitanian' species, which have disjunct distributions (Webb, 1983). These species, not present in Britain, are notably present in Ireland and the Iberian peninsula (Webb, 1983). Some species found in Ireland, such as *Arenaria ciliata* L. and *Minuartia recurva* (All.) Schinz & Thell., have their closest neighbouring populations in Switzerland and Spain, respectively. In total there are 15 Irish flowering plant species which do not occur in Britain, and for some species such as *Arenaria ciliata*, the location of the closest sister population is as far as 3,250 km distant (Webb, 1983). If the Irish plant and animal populations did not migrate across the supposed land bridge, or this land bridge was cut off much sooner than previously thought, the 'Irish Question' would require a different answer to explain the biogeographical pattern we currently observe.

1.2.3 Alternative hypotheses for the origin of Irish species

There are many lines of evidence that indicate major flaws with the conventional hypothesis of postglacial colonization. Consequently, alternative hypotheses have been proposed to account for these problems. A number of hypotheses on the origin of Irish organisms have been suggested based on phylogeographical studies, pollen core analysis and current distributions. Three main alternatives to the Irish-British land bridge hypotheses can be advanced:

1. The existence of a land bridge connecting Ireland to France over the Celtic sea which facilitated the postglacial migration of organisms into Ireland from refugia located in Southern Europe

2. Dispersal over relatively short stretches of sea facilitating postglacial migration

3. *In situ* survival of populations in Ireland in areas off the coast that were exposed due to lowered sea levels in glacial periods and/or in localized refugia on nunataks protruding through the ice-sheet covering Ireland

Anomalies in the distributions of the Lusitanian elements of the Irish biota have led to the suggestion of another route of postglacial migration, this time a land bridge across the Celtic Sea along the Western Fringe of the European Continent and Northward to Ireland. Studies on sea level changes during the LGM have highlighted the exposure of landmass in these areas (Toucanne *et al.*, 2009) (Figure 1.3). Palynological data collected for major tree species in Ireland indicate a direct migration from the South, bypassing expansion across the Irish Sea or migration through Britain (Mitchell, 2006).

The accumulation of genetic work on the phylogeography of various Irish species in a European context has provided insight into postglacial processes that facilitated the presence of these organisms in Ireland today. Phylogeographic studies for a range of animal species demonstrate a strong divergence between Irish and British populations, including stoats (Martinkova *et al.*, 2007), mountain hares (Hamill *et al.*, 2006) and frogs (Teacher *et al.*, 2009). This distinction in genetic variation between Irish and British populations has been attributed to contrasting migration routes and the possibility of survival throughout glacial cycles in an Irish refugium (Teacher *et al.*, 2009). If an Irish-British postglacial land bridge did exist, then we would expect little genetic divergence between the Irish and British populations. These studies reinforce inadequacies found with the conventional hypothesis as an explanation for the origins of current genetic structure in Irish and British populations and require other explanations (Hamill *et al.*, 2006; Mascheretti *et al.*, 2003; Martinkova *et al.*, 2007; Teacher *et al.*, 2009).

Phylogeographical studies on a wide range of species including terrestrial plant species such as *Daboecia cantabrica* (Heather) (Beatty and Provan, 2012), *Quercus robur* and *Q. petraea* (Oak) (Kelleher *et al.*, 2004) and *Colchicum autumnale* L. (Autumn Crocus) (Smith and Waldren, 2009) and in mammals such as the mountain hare (Hamill *et al.*, 2006), are providing accumulating evidence for a

post glacial land bridge across the Celtic Sea, reinforcing this Lusitanian linkage of Irish and Spanish species. Many phylogeographical studies infer this route as an explanation for the genetic patterns observed for Irish and Southern populations in Iberia and France. However, it is difficult to infer from these studies whether the relationship between the Irish and Spanish species are due to a land bridge or reflect long distance dispersal.

1.2.4 *In situ* survival of species in Ireland during the ice-ages

The possibility of a refugium or multiple refugia in Ireland has been suggested as an alternative to the Irish-British land bridge for some time, due to the disjunct distributions of Irish species (Forbes, 1846; Webb, 1983). In particular, the North West Mountain range of Ben Bulben in Co. Sligo has been suggested as a potential refugium in Ireland on the basis of the presence of disjunctly distributed plant species (Praeger, 1932; Webb, 1983), paleobotanical evidence (Colhoun *et al.*, 1972) and evidence that the Truskmore summit (which forms part of the Ben Bulben uplands) remained ice-free during the LGM (Coxon, 1988).

Paleobotanical evidence has shown that arctic-alpine and arctic species were more widespread in Ireland before the LGM than they are today. Evidence of open tundra vegetation communities is documented in pollen fossils at Derryvree, and radiocarbon dating indicated the lake where the pollen fossils were found to be 30,500 years old (Colhoun *et al.*, 1972). Pollen analysis carried out at this lake showed the occurrence of plant species such as the arctic-alpine *Arenaria ciliata* L. or *A. norvegica* Gunn. and the circumpolar arctic plant *Silene acaulis*. Synge (1969) notes that Ben Bulben, which is located 50 km from Derryvree, and which contains the only population of *Arenaria ciliata* in Ireland presently, may have acted as a nunatak during the LGM which facilitated the persistence of the Irish population. Webb (1983) also suggests that *Minuartia recurva* and *Arenaria ciliata* occurred in Ireland long before the last glaciation. The Irish plant fossil record also indicates the possibility of the survival of species in Ireland, where

plants with rare and disjunct distributions are documented since the Gortian, c. 428-302, 000 years BP (Coxon and Waldren, 1995).

Molecular studies have also indicated the possibility of refugia in Ireland during the LGM for a range of species, including marine organisms: the brown seaweed *Fucus serratus* (Hoarau *et al.*, 2007), red seaweed *Palmaria palmate*, (Provan *et al.*, 2005), brown trout (*Salmo trutta*) (McKeown *et al.*, 2010), terrestrial animals: the common frog (*Rana temporaria*) (Teacher *et al.*, 2009) and the land snail *Cepaea nemoralis* (Davison, 2000). These studies suggest the possibility of refugia in the South West and West coasts of Ireland, based on the presence of unique haplotypes in the Irish populations and/or high genetic diversity.

Despite evidence pointing towards the possibility of the survival of species *in situ* during the full glacial periods, this hypothesis is still in doubt due to the extent of the ice-sheet cover and sea level changes, as well as the lack of contemporary flora records from pollen or macrofossils in Ireland until around 13,000 BP (Kelleher *et al.*, 2004; Mitchell, 2006; Rowe *et al.*, 2006). However it is clear that an alternative explanation to postglacial expansion is essential in order to explain the current distribution and genetic structuring for some of the Irish species.

1.3 Phylogeography

1.3.1 Overview of Phylogeography

The term phylogeography was first coined by Avise *et al.* (1987) and has since been implemented in a vast array of studies on a variety of organisms and using a range of molecular markers. A search of the ISI Web of Science (www.isiknowledge.com, February 2013) returned 13,191 studies with the search keyword phylogeography, 238 of which included the search terms phylogeography, ice-age and Europe. The development of this research field has lead to the investigation of a range of topics, such as the inference of Pleistocene refugia e.g. Maggs *et al.* (2008); Schonswetter *et al.* (2005); Valtuena *et al.* (2012) and patterns of

postglacial colonization e.g. Martinkova *et al.* (2007); Schonswetter *et al.* (2003); Taberlet *et al.* (1998).

Advances in the development of molecular techniques used in biogeographic studies have enabled researchers to infer genetic relatedness among individuals and populations, therefore allowing insight into past events based on the spatial distribution of genetic variation (Hewitt, 2000). The study of phylogeography is a sub-discipline of biogeography, which is focussed on the processes which govern the geographic distributions of genealogical lineages, especially those within and among closely related species (Avise, 2000).

Phylogeographical methods provide a way to examine the history of genetic exchange among populations and can help distinguish biogeographic patterns of genetic variation resulting from current gene flow processes or common ancestry (Schaal *et al.*, 1998). Distinct genetic lineages in a population can be traced back to a common ancestor that they shared at some point in history (Avise, 2009), since populations which have recent common ancestry will be genetically more similar than those populations which are more distantly related (Schaal *et al.*, 1998).

1.3.2 Molecular markers and loci used in phylogeography

Molecular techniques for phylogeographic studies can be divided into two categories (i) sequence based markers where variations at loci can reveal variation between individuals (ii) PCR-based markers with non-specific primers (e.g. Amplified Fragment Length Polymorphism (AFLPs)). PCR-based markers with non-specific primers are a common choice among population genetic studies as they do not require prior sequence information for the target species (Nybom, 2004). This is an important consideration, especially for studies which do not include a model organism or wild species which has not been previously investigated. Some of these techniques may provide a large amount of polymorphic markers and therefore high levels of variation. This allows the detection of variation on an individual level within and between populations. Differences between individuals in a population, however, may not be detected as frequently using single locus sequencing methods. Low variability in a marker means that locus may

have evolved too slowly to provide informative polymorphisms for phylogenetic analysis (Lowe *et al.*, 2007). Sequencing methods, however, are fast and reliable and many studies have found sufficient variation to detect variability among individuals using this single-locus approach (Olsen *et al.*, 2004; Fernandez-Mazuecos and Vargas, 2011; Rodriguez-Sanchez *et al.*, 2009). Many studies now favour the combination of using single locus sequences and techniques such as AFLPs (Brochmann and Brysting, 2008; Schonswetter *et al.*, 2006; Valtuena *et al.*, 2012; Westergaard *et al.*, 2011).

A vast range of genetic tools are available for phylogeographical analyses and the characteristics of each molecular technique must be considered carefully when choosing for any given study. Other considerations, such as the cost of the technique used and the variability of the sequence region or markers investigated should also be considered. The region of the genome from which a marker will be amplified (for example nuclear, chloroplast and mitochondrial) is also an important factor. Each of these different genomes differ in their mode of inheritance and substitution rate (Wolfe *et al.*, 1987), and so the selection of each marker region must be appropriate for the hypothesis being investigated. For example, the chloroplast genome is inherited maternally through seeds in most angiosperms and so can be considered as a single, non-recombining unit of inheritance (Schaal *et al.*, 1998). As a result, the mutation rate of chloroplast DNA (cpDNA) is slower than that of nuclear DNA (nDNA), where the nDNA can evolve twice as fast as cpDNA in plants (Wolfe *et al.*, 1987). The use of cp DNA in phylogeographical studies is very common as cpDNA is more likely to retain historical sequences over longer time periods than nuclear DNA markers where gene flow occurs through both seed and pollen (Comes and Kadereit, 1998). Mitochondrial DNA is widely implemented in animal phylogeographical studies, however the use of this marker in plants is limited due to the low rates of nucleotide substitution resulting in inadequate variation to estimate intraspecific phylogenies (Schaal *et al.*, 1998).

The uniparental mode of inheritance of chloroplasts also means, however, that genealogical patterns revealed by plastid markers are not necessarily representative of the entire population history (Hare, 2001). The additional inclusion of nuclear markers, however, can convey information on the phylogeographical history of the target organism that the chloroplast data cannot provide, such as more

recent gene flow processes. A commonly used region in the nuclear genome that has been successfully sequenced and implemented in phylogeography studies is the nuclear ribosomal internal transcribed spacer (ITS) region (Eidesen *et al.*, 2007). Methods apart from direct sequencing such as AFLP provide a whole genome assessment and many markers sampled per individual, a description of this technique is provided in Chapter 5. Ideally, a combination of multiple loci sampled from the chloroplast and nuclear genome and the comparison of gene trees reflecting both deep and shallow phylogeographical histories, produced from both types of markers will reveal insights into many historical and biological processes as well as into current and past gene flow (Eidesen *et al.*, 2007; Schaal *et al.*, 1998).

1.3.3 Methods of analysis implemented in phylogeography - intraspecific phylogenies and networks

Phylogenetic trees have been used to understand evolutionary relationships between organisms since the first depiction of an evolutionary tree by Charles Darwin in 1859 (Darwin, 1859). Since then scientists have sought to classify all known organisms within this biological framework of common ancestry. Many research methods have been developed to model phylogenetic trees based on algorithms that analyze and compare genetic sequence data. Phylogenetic principles (including neighbour joining (NJ), maximum likelihood (ML) and bayesian inference) have been widely applied to intraspecific molecular data for a range of organisms (Hamill *et al.*, 2006; Jolly *et al.*, 2006; Hoarau *et al.*, 2007; Kelleher *et al.*, 2005; Schonswetter, Tribsch and Niklfeld, 2004*b*; Teacher *et al.*, 2009).

1.3.3.1 Neighbour-joining

Neighbour-joining (NJ) is a distance based phylogenetic tree building method which has been implemented widely in phylogeographic studies and performs well when sequence divergence is low (Holder and Lewis, 2003). The principle

behind this method is to find pairs of operational taxonomic units that minimize the total branch length at each stage of sequence clustering and a final tree is produced under the principle of minimum evolution (Saitou and Nei, 1987). The raw data are first converted into a distance matrix where evolutionary distances are calculated for all pairs of terminal taxa, and a tree is constructed based on the relationships between these distance values (Lowe *et al.*, 2007)

Disadvantages of using neighbour-joining methods is that information is lost in compressing sequences into distances and reliable estimates of pairwise distances can be more difficult to obtain for divergent sequences (Holder and Lewis, 2003). For intraspecific studies, however, sequences are less divergent and the NJ method has been extensively used for constructing trees for AFLP analyses, for example Chung *et al.* (2004); Ehrich *et al.* (2007); Kelleher *et al.* (2005); Kropf *et al.* (2008); Schonswetter *et al.* (2002, 2006); Skrede *et al.* (2006); Westergaard *et al.* (2008). This technique is very fast and easy to implement and the tree generated may also provide a starting point for model-based analyses, such as maximum likelihood or Bayesian analyses (Lowe *et al.*, 2007).

1.3.3.2 Maximum Likelihood

One of the most widely used methods of tree construction is maximum likelihood (ML), which has been implemented in a range of phylogeographic studies using genetic sequence data (Chung *et al.*, 2004; Eidesen *et al.*, 2007; Fernandez-Mazuecos and Vargas, 2011; Hamill *et al.*, 2006; Hoarau *et al.*, 2007; Jolly *et al.*, 2006; Rowe *et al.*, 2006; Teacher *et al.*, 2009; Valtuena *et al.*, 2012). Maximum likelihood can depict evolutionary relationships by finding the tree which maximises the probability of obtaining the observed data (Lowe *et al.*, 2007; Huelsenbeck and Crandall, 1997; Wiley and Lieberman, 2011; Xiong, 2006). It does this given a specific model of evolution adopted by the investigator (Huelsenbeck and Crandall, 1997; Wiley and Lieberman, 2011; Xiong, 2006).

Assumptions are first made about the rates of base pair change which can be determined using a statistical test to choose the best-fit model of nucleotide substitution (Lowe *et al.*, 2007). When a tree search is conducted, the fit of the trees constructed is assessed using the substitution model and the tree that

results in the highest probability (likelihood) of predicting the observed sequences is chosen as the ML tree (Felsenstein, 1981; Holder and Lewis, 2003).

Advantages of using maximum likelihood include its potential to infer population subdivision and geneflow estimates (Hare, 2001). The major disadvantage to maximum likelihood is that it can be computationally demanding. However in the case of intraspecific studies this is generally not an issue as sequence divergence is reduced and this method is less computationally demanding than it would be for an interspecific analysis.

1.3.3.3 Bayesian inference

Bayesian inference is the newest method of phylogenetic tree construction and the implementation of this technique has proved useful in many phylogenetic, phylogeographic and molecular clock analyses. One of the advantages of using Bayesian inference include that it allows implementation of complex models of sequence evolution which may be too difficult for ML (Holder and Lewis, 2003). The primary Bayesian analysis also produces both a tree and measures of uncertainty for groups on the tree (Holder and Lewis, 2003), unlike maximum likelihood and NJ methods where bootstrap values need to be calculated to assess the robustness of support.

Bayesian methods of phylogenetic tree construction are statistically similar to maximum likelihood in that the optimal hypothesis for the creation of the tree is one that maximizes the posterior probability (the posterior probability is proportional to the product of the prior probability and the likelihood i.e. the probability that the tree is correct (Huelsenbeck *et al.*, 2001)). The most recent methods of Bayesian inference implemented in the program BEAST (Drummond and Rambaut, 2007) allow the user to specify prior divergence age distributions for selected nodes in the tree (Ho and Phillips, 2009). Such priors allow the user to input additional information, such as the taxonomy of the group (Huelsenbeck *et al.*, 2001), or date estimates derived from fossil data, to give the trees more weight for the prior probability (e.g. normal, uniform, exponential). The priors are drawn from a distribution, and therefore the application of this prior information is more appropriate for assigning distribution of fossil dates for example,

rather than point calibrations, which more appropriately incorporate calibration uncertainties (Drummond *et al.*, 2006). This also allows a high degree of flexibility when incorporating a time scale into a phylogenetic analysis.

Bayesian inference is commonly used for molecular clock analysis and an advantage of using Bayesian inference over maximum likelihood is that this method takes topological and branch length uncertainties into account, whereas age estimated using maximum likelihood are determined using a single input tree (Frajman *et al.*, 2009).

1.3.3.4 Networks

Evolutionary relationships above and below species level are different from each other and traditional tree-based methods such as maximum likelihood may not depict genealogical relationships between individuals from the same species adequately or may have poor resolution (Clement *et al.*, 2000; Posada and Crandall, 2001). The development of marker data at the population level and advances in population genetics theory has led to the generation of powerful tools for intraspecific analysis for testing a variety of evolutionary and population genetic hypotheses (Posada and Crandall, 2001). One of these tools is the generation of networks which are thought to be more appropriate in inferring gene genealogies than tree-based methods (Pleines *et al.*, 2009). Networks may account effectively for processes that may be common on a population level, e.g. reticulation, and could be able to incorporate predictions from population genetics theory (Posada and Crandall, 2001).

An ongoing debate on the advantages and disadvantages of using traditional tree constructing methods or networks for use in intraspecific analyses is evident in the literature (Posada and Crandall, 2001; Salzburger *et al.*, 2011). Salzburger *et al.* (2011) simulated data sets to test the performance of the traditional phylogenetic alogorthims and network constructing methods and found that the traditional phylogenetic methods are applicable to haplotype data and show reasonable performance for both accuracy and robustness. However, they also found that the statistical parsimony network approach performed poorly. Implementing a variety of methods such as maximum likelihood and network analysis that

may complement each other can aid in compensating the disadvantages of each method.

1.4 Description, taxonomy and distribution of *Arenaria ciliata* and *A. norvegica*

The target species in this study are *Arenaria ciliata* L. (Fringed Sandwort) and *Arenaria norvegica* Gunnerus (Arctic Sandwort). They are both small, tufted, white-flowered angiosperm species belonging to the Carnation (Caryophyllceae) family.

Arenaria ciliata L. is an arctic-alpine montane perennial which occurs in open habitats on basic soils (Tutin *et al.*, 1964). *Arenaria norvegica* Gunnerus is strictly an Arctic perennial or annual, and, like *A. ciliata*, occurs in open habitats on basic soils. *Arenaria norvegica*, however, is more confined to exposed, finer and shallower soils than *A. ciliata*, which frequently grows alongside other vegetation in patchy upland grasslands, scree slopes along disturbed pathways and cliffs (Curtis and McGough, 1998).

1.4.1 Taxonomic classification of the target species

The family Caryophyllaceae is a large family consisting of 86 genera and 2,200 species (Bittrich, 1993). The *Arenaria* genus contains approximately 160 species, which occur mainly in arctic and temperate regions in the Northern hemisphere (Wyse Jackson and Parnell, 1987).

McNeill (1962) describes a section of the *Arenaria* genus, *Rariflorae*, which contains approximately 14 species, including *Arenaria ciliata* L. and *A. norvegica* Gunnerus. Other species which are included in the original description of this genus include *Arenaria gothica* Fries, *A. pseudofrigida* (Ostenf. & Dahl) Juz., *A. cinera* DC. and *A. humifusa* (Swartz) Wahlb. The species occurring in this section *Rariflorae* are widely distributed arctic-alpines which occur throughout most of Northern Europe and America and extend into the Central European mountains, Spain, the Balkans, Anatolia and Iran (McNeill, 1962).

Arenaria ciliata sensu lato is a taxonomically difficult group which includes the closely related species: *Arenaria ciliata* L., *A. norvegica* Gunnerus and *A. gothica* Fries (*A. moehringioides*) (Wyse Jackson and Parnell, 1987). The current taxonomic status for *Arenaria ciliata* and *A. norvegica* in the literature based on morphological characters is quite confusing. These two species belong to a complex with a large amount of variation and overlap in morphology, ploidy number and geographic distribution (Wyse Jackson and Parnell, 1987). *Arenaria ciliata* and *A. norvegica* are both recognized as distinct species, however, each of which contains a variety of subspecies. Species and subspecies distinctions are still not clear in the literature for this species complex and the classification varies in the literature (Castroviejo *et al.*, 1990; Tutin *et al.*, 1964; Wyse Jackson and Parnell, 1987). Webb (1983) comments on this discrepancy in the classification of this species and subspecies within this complex, where the Irish plants from Ben Bulben are "generally allowed to be distinct at subspecific rank, but as to which other populations are to be included in the same species every author has different views".

The most comprehensive taxonomic study of the *Arenaria ciliata* complex is that of Wyse Jackson and Parnell (1987) which studied 25 morphological characters to analyze species distinctions within this complex using biometric analysis. The authors show that there is little evidence for strong distinctions between any of the species or subspecies in the *A. ciliata* complex (*A. ciliata* L, *A. norvegica* Gunnerus and *A. gothica* Fries). Despite this, the authors provide a description of the subspecies which include: *A. ciliata* subsp. *ciliata*, *A. ciliata* subsp. *pseudofrigida* (Ostenf. & Dahl) Juz., *A. ciliata* subsp. *bernensis* Favarger and *A. ciliata* subsp. *tenella* (Kit.). *Arenaria gothica* Fries includes subsp. *gothica* and subsp. *moehringioides* (J. Murr). M.B. Wyse Jackson & J. Parnell. The subspecies for *Arenaria norvegica* includes subsp. *norvegica* (Gunnerus) Fries and subspecies *anglica* Halliday.

Wyse Jackson and Parnell (1987) found that *Arenaria moehringioides* Murr (= *A. multicaulis* L.) did not form a distinct group and therefore classify *A. moehringioides* as *A. gothica* subsp. *moehringioides*. This change is not included in the Flora Iberica (Castroviejo *et al.*, 1990), however, which still refers to this taxon as *A. moehringioides*. This taxon is also referred to as *A. moehringioides*

in Fior *et al.* (2006) who conducted the only combined genetic and morphological analysis to include *Arenaria ciliata* in a phylogenetic context. Results from their study show the consistent pairing of *Arenaria ciliata* and *A. moehringioides* with strong bootstrap support for nuclear ribosomal internal transcribed spacer (ITS) sequences, chloroplast *matK* sequences and morphological characters. Their closest sister species is indicated as *A. nevadensis*.

It is difficult to asses the taxonomic status of *Arenaria ciliata* and *A. moehringioides* (*A. gothica* subsp. *moehringioides*) based on the results from Fior *et al.* (2006) as only two taxa from the *A. ciliata* complex are included in this study (*A. ciliata* and *A. moehringioides*). To date, no genetic analysis has been carried out on the populations of these species and subspecies. Thus, for this study we grouped the populations into two broad groups of *Arenaria ciliata* and *A. norvegica* and *A. ciliata* subsp. *pseudofrigida* for the individuals of *A. ciliata* sampled in Svalbard.

1.4.2 Distribution of the target species in Europe

According to the Atlas Florae Europeae (Jalas and Suominen, 1983) and Flora Europea (Tutin *et al.*, 1964), *A. norvegica* subsp. *norvegica* is a perennial which occurs in Scotland, Ireland, Iceland, Norway, Sweden and Finland. *A. norvegica* subsp. *anglica* Halliday is recorded as being an annual or biennial and of a laxer habit than subsp. *norvegica* (Tutin *et al.*, 1964) and occurs amongst limestone fragments and in shallow solution hollows with thin soils in the Craven area of Mid-West Yorkshire (Walker, 2000).

Arenaria ciliata is widely distributed among the major European mountain ranges Southwards to Central Spain and Northern Italy (Tutin *et al.*, 1964). According to the Atlas Florae Europaeae (Jalas and Suominen, 1983) this species has been recorded in the central and Eastern Alps, in the Carpathians, in Finland, North Norway, arctic Russia, Svalbard and Ireland. *Arenaria gothica* subsp. *moehringioides*, which is classed as *A. multicaulis L.* in Jalas and Suominen (1983), is distributed in Central and Northern Spain, Pyrenees, Jura, Alps Eastwards to Vorarlberg and the North Apennines.

1.4.3 Site descriptions and distribution of the target species in Ireland

Arenaria ciliata and *A. norvegica* are both red listed species in Ireland which means they are legally protected under the Flora Protection Order 1987. Photographs of *Arenaria ciliata* and *A. norvegica* growing in their respective Irish sites are shown in Figure 1.4.

Arenaria ciliata L. is found in only one location in Ireland, in the Ben Bulben Mountain range in County Sligo, where it is occurs frequently on the exposed limestone rocks, cliffs and alpine grasslands. This species is absent from Britain and the next closest population occurs in the Alps, separated by a distance of 1,350 kilometres (Webb, 1983). The individuals from Ben Bulben were originally grouped as an endemic sub-species (subsp. *hibernica* (Ostenf. & O. C. Dahl) G. C. Druce) which is thought to constitute part of the *Arenaria ciliata* complex, though it is no longer considered a sub-species (Wyse Jackson and Parnell, 1987).

The Ben Bulben uplands stand in a high plateau, 350-400 meters above the surrounding vicinity (NPWS, 2003) and consist of Carboniferous limestone, capped in areas by shale. This site has been designated as a Special Area of Conservation (SAC) (site code: 000623). The plateau in Ben Bulben is recognized as being one of the botanically richest sites in Ireland and provides the country's best example for alpine and arctic-alpine vegetation, including two flowering vascular plant species which are not known to occur anywhere else in Ireland (*Arenaria ciliata* and *Saxifraga nivalis*) (NPWS, 2003). Ben Bulben is also the home to the rare disjunctly distributed moss species *Barbula maxima* which also occurs in the sub-arctic permafrost region in the East side of the Mackenzie Mountains in Canada, 5,600 km away (Synnott, 1981).

Arenaria norvegica is likewise only found in one location in Ireland, on the Karst Carboniferous limestone pavement at Blackhead, in the Burren, Co. Clare. The occurrence of this species in the Burren is of special interest because the rediscovery of this single population in 2007 followed a period of the species going unreported for 47 years (Walker *et al.*, 2008). This area is also a designated SAC (site code: 000020) (NPWS, 2003). This site is located within the main area of the Burren that was glaciated during previous glacial maxima and contains erratics

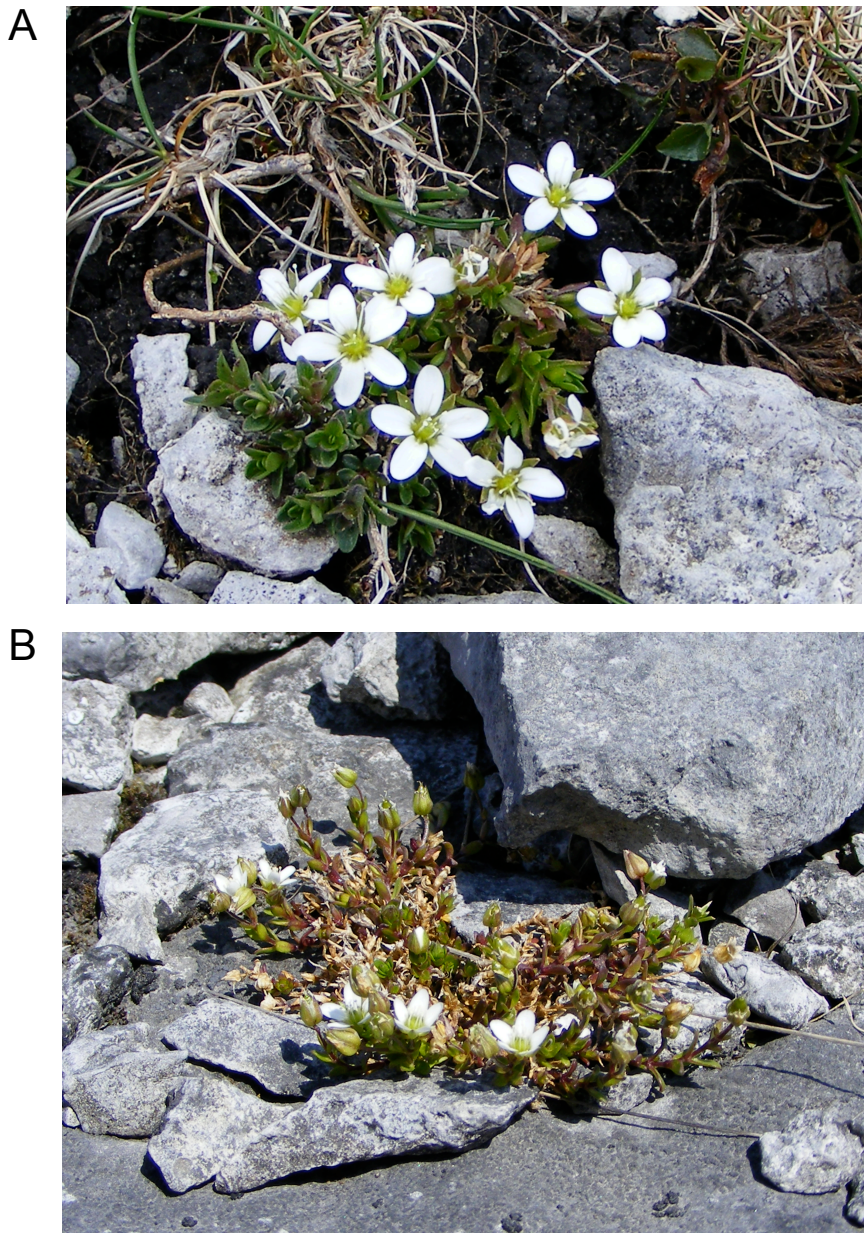


Figure 1.4: Close up of *Arenaria ciliata* growing among limestone scree in Ben Bulben, Co. Sligo (A) and *A. norvegica* growing in the cracks of the limestone pavement in the Burren, Co. Clare (B). Note the shorter, more squat appearance and smaller inter-node length of the branchlets of *A. norvegica* compared to *A. ciliata*

from Galway granite (NPWS, 2003). The area is of international scientific interest and the bare limestone pavement is interspersed with species-rich dry, calcareous grasslands that contain a high proportion of rare and noteworthy species (NPWS, 2003).

1.5 Aims and chapter layout

The aim of this study was to investigate the glacial history of the Irish populations of *Arenaria ciliata* and *A. norvegica* in a European context. A phylogeograpical analysis involving a range of molecular markers and methods of analysis was carried out in order to determine putative postglacial migration routes to Ireland and investigate the possibility of *in situ* survival during the last glacial maximum (LGM). Populations of the target species were sampled across Europe to cover a wide geographical distribution and the molecular markers used to compare the genotypes of the Irish populations with the continental European metapopulations. Chapters presented here are divided on the basis of the molecular marker and analysis method used.

Chapter 2 presents the materials and methods associated with sample collection and all wet-lab procedures for PCR, DNA sequencing and AFLP analysis. The data analysis which was carried out subsequently are also detailed in this chapter.

Chapter 3 focuses on the phylogeographical analysis of chloroplast and nuclear sequences of *Arenaria ciliata* and *A. norvegica* using maximum likelihood and haplotype networks. The relationship between *Arenaria ciliata* and *A. norvegica* was investigated using maximum likelihood phylogenetic analysis of the Caryophyllaceae family for the *matK* chloroplast sequence in order to determine if the classification as distinct species is supported by molecular markers.

In Chapter 4 a molecular clock analysis using Bayesian inference was conducted in order to infer the timing of the key phylogeographic events and relationships for the target species. This analysis was also used to identify an age estimate for the node determining the expansion of the *A. ciliata* lineage and its closest sister species that was used to calibrate a node for the intraspecific

analysis of the target species to estimate dates for the divergence of the major lineages.

Finally Chapter 5 presents a genome wide assessment of the populations of the target species using AFLP genotyping method. Relationships between the populations were determined using neighbour-joining trees and neighbour-net networks. Partitioning of genetic structure within and among populations was determined using analysis of molecular variance (AMOVA). AFLP data were also used to determine the taxonomic association between *Arenaria ciliata* and *A. norvegica*.

The investigation of these rare and disjunctly distributed cold-tolerant plant species provides a good model to examine hypotheses of post-glacial colonization and survival *in situ* for the Irish Arctic-Alpine Flora. The populations of both species are widely distributed in Europe and are distributed in areas which are thought to have been key refugia during Glacial Cycles, such as the Alps, but also in more Arctic regions such as Svalbard. The mounting evidence from a range of disciplines suggests that Ben Bulben is a glacial refugium (Coxon, 1988; Synge, 1969; Synnott, 1981; Webb, 1983). This, coupled with the cold-tolerant nature of *Arenaria ciliata* and *A. norvegica*, places these species as key candidates for glacial relicts which survived the LGM *in situ*. Of key importance in the postglacial history of the Irish flora, these species have not as yet been studied in any phylogeographic context and there exists a considerable knowledge gap in relation to the effects of the glacial cycles on the Irish Flora.

2

Materials and Methods

2.1 Chemicals

The following chemical reagents and stock solutions were used in wet-lab analysis.

2.1.1 List of chemicals

- Agarose: Invitrogen, Paisley, Scotland
- β-mercaptoethanol: AnalR, BDH, Poole, England
- Chloroform: Merk KGaA, Darmstadt, Germany
- CTAB powder: Sigma-Aldrich Co. LLC., Dorset, UK
- Iso Amyl Alcohol: AnalR, BDH, Poole, England
- EDTA: Sigma-Aldrich Co. LLC., Dorset, UK
- Formamide: Sigma-Aldrich Co. LLC., Dorset, UK
- Silica Gel : Sigma-Aldrich Co. LLC., Dorset, UK
- Sodium chloride: Sigma-Aldrich Co. LLC., Dorset, UK
- TAE Buffer (Modified Tris-Acetate EDTA Buffer): Milipore, MA, USA

- Tris-HCL, pH 8: Sigma-Aldrich Co. LLC., Dorset, UK

2.1.2 Preparation of reagents

2% CTAB Buffer 200ml
4g CTAB powder
16.364g $NaCl_2$
20ml 1M Tris-HCL, pH 8
8ml 0.5M EDTA
All of the chemicals were first dissolved in 70ml ddH_20 and then diluted with ddH_20 to 200ml.

SEVAG 250ml
240ml Chloroform
10ml Isoamyl Alcohol

$TE_{0.1}$ Buffer 500ml
5ml 1M Tris-HCL
1ml 0.5M EDTA
494ml ddH_20
10ml of this solution was then diluted with 90ml of ddH_20.

2.2 Population surveys for the red listed target species in Ireland

The Irish collections involved detailed site descriptions, compared to simple site observations everywhere else. Surveys to map the populations and describe associated vegetation for *Arenaria ciliata* (AC) in Ireland were carried out in June 2009. *Arenaria norvegica* (AN), a species previously considered extinct in Ireland, was also included in the survey as it was recently rediscovered in a single population in the Burren, Co. Clare (Walker *et al.*, 2008). A detailed survey of

each of these populations was carried out incorporating: GPS coordinates (Table 2.1 and 2.2), altitude, habitat/associated geology and approximate number of plants in each population (Table 2.3 and 2.5). The approximate area covered by each population was also recorded.

To characterise the vegetation in which the target species occurred, quadrats of $1m^2$ were used to record the cover abundance of plant species. In the case of *Arenaria ciliata*, five quadrats were placed at each population, three of which included the target species in the quadrats and two quadrats without the target species. The average percentage cover for each plant species recorded was determined separately among the quadrats that included and excluded the target species (Table 2.4). One quadrat was used to record the species that co-occurred with *Arenaria norvegica*, which was recorded by Kevin Walker (BSBI, North Yorkshire, England). The quadrats were also used to record the percentage cover of bare earth, bare rock, total vegetation cover, bryophytes and grasses. The average values, also calculated among the quadrats that included and excluded the target species, are shown in Table 2.4.

Table 2.1: Location of sampled *Arenaria ciliata* populations with date surveyed, date collected and latitude/longitude

Pop.	Site Location of Population	Date Surveyed	Date Col.	Latitude/Longitude
AC1	Kings Mountain, Ben Bulben, Ireland	8.6.09	25.8.09	N54 °20.672' W8 °27.373'
AC2	Gowlaun Valley, Ben Bulben, Ireland	8.6.09	25.8.09	N54 °21.353' W8 °27.290'
AC3	Glencarbury Mine, Ben Bulben, Ireland	9.6.09	24.8.09	N54 °21.549' W8 °24.044'
AC4	Glendarragh Valley, Ben Bulben, Ireland	9.6.09	24.8.09	N54 °21.879' W8 °25.705'
AC6	Rifugio Mongioie, Ligurische Alpen, Piemonte, Italy	28.6.09	28.6.09	N44 °09.864' E7 °47.201'
AC7	Sent. Frassati, Cottische Alpen, Piemonte, Italy	30.6.09	30.6.09	N44 °27.227' E6 °55.313'
AC8	Colle dell'Agnello, Cottishe Alpen, Piemonte, Italy	1.7.09	1.7.09	N44 °40.709' E6 °59.484'
AC9	Col Agnel, Cottische Alpen, Hautes-Alpes, France	2.7.09	2.7.09	N44 °46.88' E6 °40.638'
AC10	Hospital de Benasque, Huesca, Pyrenees, Spain	28.6.10	28.6.10	N42 °40.957' E0 °36.177'
AC11	Mine, Picos de Europa, Cantabrica, Spain	1.7.10	1.7. 10	N43 °09.533' W4 °49.302'
AC13	Hospital de Benasque, Huesca, Pyrenees, Spain	28.6.10	28.6.10	N42 °40.957' E0 °36.177'
AC14	Cabana Veronica, Picos de Europa, Spain	1.7.09	1.7.09	N43 °10.644' W4 °49.967'
AC15	Hacendida de Corarrobres, Picos de Europa, Spain	3.7.09	3.7.09	N43 °09.374' W4 °48.213'
AC16	Niedere Tauren/Wolzer Tauren, Steiermark, Austria	12.7.10	12.7.10	N47 °16.27' E14 °21.21'
AC17	Karawanken, Karten, Austria	13.7.10	13.7.10	N47 °30.20' E14 °29.12'
AC18	Leukerbad, Switzerland	20.6.11	20.6.11	N46 °23.883' E7 °34.596'
AC19	Leukerbad, Switzerland	21.6.11	21.6.11	N46 °25.165' E7 °37.478'
AP1	Midtre Lovnbreen, Norway	21.7.11	21.7.11	N78 °54.28.8' E12 °04.417'
AP2	Bohemanflya, Norway	27.7.11	27.7.11	N78 °23.25' E14 °44.26'

AP - *Arenaria ciliata* subsp. *pseudofrigida*

Table 2.2: Location of sampled *Arenaria norvegica* populations with date collected and latitude/longitude

Pop.	Site Location of Population	Date Col.	Latitude/Longitude
AN1	Black Head, The Burren, Co. Clare, Ireland	1.6.09	N53 °08.243' W9 °16.048'
AN2	Ribblesdale, Yorkshire, England	June 2009	N43 °38.194' W116 °14.289'
AN3	Ribblesdale, Yorkshire, England	June 2009	N43 °38.194' W116 °14.289'
AN4	Ruinsival, lower NW slopes and Loch Fiachanis, Rum, Scotland	29.7.09	N43 °38.194' W116 °14.289'
AN5	Cnoc Eilid Mathain, Inchnadamph, Scotland	11.08.09	N58 °07.493' W4 °55.374'
AN6	Unst, Shetland, Scotland	August 2009	N43 °38.194' W116 °14.288'
AN7	Valley of Skafta river, Herobreio, Iceland	4.07.10	N65 °26.15.78' W16 °51.109'
AN8	Fjallsjokull Glacier, Hofn, Iceland	14 June 2012	N64 °00.23' W16 °23.10'

2.3 DNA sampling protocol

Twenty-nine populations of the two target species were sampled across Europe. Maps showing the distribution of the sampled populations in Europe, which were correctly identified and subsequently used for genetic analyses, are shown in Figure 2.2. Maps showing the regional distribution of the sampled populations in The Alps and Spain are shown in Figure 2.3. A detailed distribution for the areas covered in Ireland are shown in Chapter 3. A total of eight populations were collected for *Arenaria norvegica* and 21 for *A. ciliata*. Leaf tissue samples were also acquired from other collectors in Scotland (*Arenaria norvegica* - Kevin Walker), England (*A. norvegica* - Kevin Walker), Iceland (*A. norvegica* - Peter Gruber and Conor Meade). Populations were also sampled from Svalbard for *Arenaria ciliata* subsp. *pseudofrigida* (AP) (Pernille Bronken Eidesen and Lorna Little). In most cases leaf tissue was collected from at least ten individuals per population. In some populations, such as the Irish populations and for certain continental European populations, up to 30 individuals were sampled. In other cases only a few individuals could be obtained due to small population sizes. The number of individuals sampled for each population are recorded in Tables 2.3 and 2.5.

Individuals at each site were sampled at least ten metres apart in order to obtain a thorough representation of the spatial distribution of individuals within each local population and to decrease the likelihood of sampling sibling groups. Single branchlets with a common basal stem from individual plants were removed and placed in separate envelopes and dried with silica gel. Voucher specimens of individuals from the Irish populations were collected and deposited in the National Botanic Gardens Herbarium, Glasnevin, Dublin. DNA was extracted from the dried tissue samples, divided into duplicate aliquots, and then stored at $-20\,^{\circ}\mathrm{C}$ (short term - <1 month) and $-80\,^{\circ}\mathrm{C}$ (long term > 1 month) in the Molecular Ecology Laboratory, NUI Maynooth. Taxonomic identification of all samples was based on an initial examination of European and Irish herbarium material at the herbaria in The National Botanic Gardens, Dublin and the Real Jardin Botanico (RJB), Madrid. Seeds were also sampled by removing a few mature capsules from approximately 10% of each sampled population. These

capsules were placed in an envelope and the seeds were removed and cleaned at a later stage.

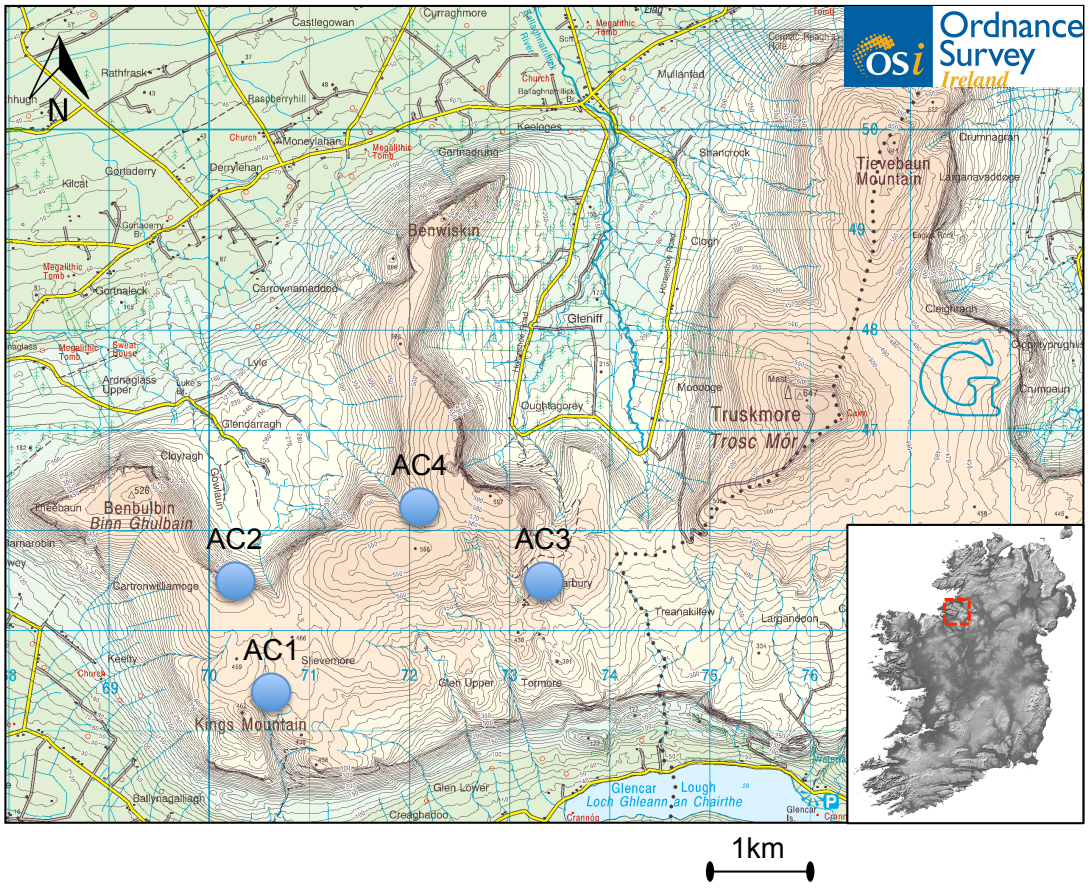


Figure 2.1: Ordnance Survey map of the Ben Bulben mountain range in Co. Sligo with the sites sampled for *Arenaria ciliata* indicated by the blue circles. Copyright Permit No. MP 002513.

2.4 Site details and distribution maps

2.4.1 *Arenaria ciliata*

The populations of *Arenaria ciliata* sampled in Europe are shown in Figure 2.2 and 2.3. A map showing the Ben Bulben Mountain range in Co. Sligo is shown in Figure 2.1.

Four sub-populations were identified for *Arenaria ciliata* in Ben Bulben, Co. Sligo: Kings Mountain (AC1), Gowlaun Valley (AC2), Glencarbury Mine (AC3) and Glendarragh Valley (AC4) (Table 2.1). Each of these sub populations were chosen as they were separated by $>$ 1km of upland blanket bog which did not contain any suitable habitat for *Arenaria ciliata*. The area that the surveyed populations covered are shown in Figure 2.4.

At the Ben Bulben sites *Arenaria ciliata* generally occurs on open mountain grassland interspersed with limestone scree and rocks on basic soil (an example of *A. ciliata* in this habitat at Ben Bulben is shown in Figure 2.5). It was also found to occur on a number of the limestone cliffs and outcrops and it was very frequent on some patches of gravel on the rough paths which are interwoven throughout the Glencarbury mine complex. The typical habitats that *Arenaria ciliata* occurred are shown in Figure 2.6.

Glencarbury mine (AC3) contained the largest population of *Arenaria ciliata* with an abundance of very large clumps of *A. ciliata* (Figure 2.6 panel D). The habitat here is quite heterogeneous with an array of moderately grazed grassland containing rocky limestone outcrops and gravel pathways which are interwoven throughout the mine (Figure 2.6 panels A,B,E,F). Other populations also contained many individuals of *A. ciliata* in a complex of open mountain grassland interspersed with limestone scree and rocks on basic soils. The smallest population occurred on Kings Mountain (AC1) where it mostly occurred on limestone

outcrops and cliffs in shallow soils on the north and west slopes (Figure 2.4 panel C) and Figure 2.6 panel C).

Altitude, habitat, associated geology, population size and the number of sampled individuals for each population for *Arenaria ciliata* is recorded in Table 2.3. Other plant species which were recorded in the same quadrat as *A. ciliata* are shown in Table 2.4 along with species from the surrounding area. The vegetation is dominated by *Festuca* spp. and *Sesleria caerulea* and *Thymus polytrichus* was also commonly recorded. *Silene acaulis* (an arctic-alpine cushion plant) is also very common and co-occurs with *Arenaria ciliata* but is more common in the surrounding area.

The diversity of vegetation found occurring with *A. ciliata* at Ben Bulben was generally much lower compared to sites in the Alps, Pyrenees and Picos De Europa. A photograph of this diversity of alpine flora from the Pyrenees is shown in Figure 2.7. Figure 2.8 shows a close up image of the ciliated leaves of *A. ciliata* from the Pyrenees. The level of ciliation in the leaves varies between regions and populations. Typical habitats that *A. ciliata* occurred in Picos De Europa and the Austrian Alps are shown in Figure 2.9.

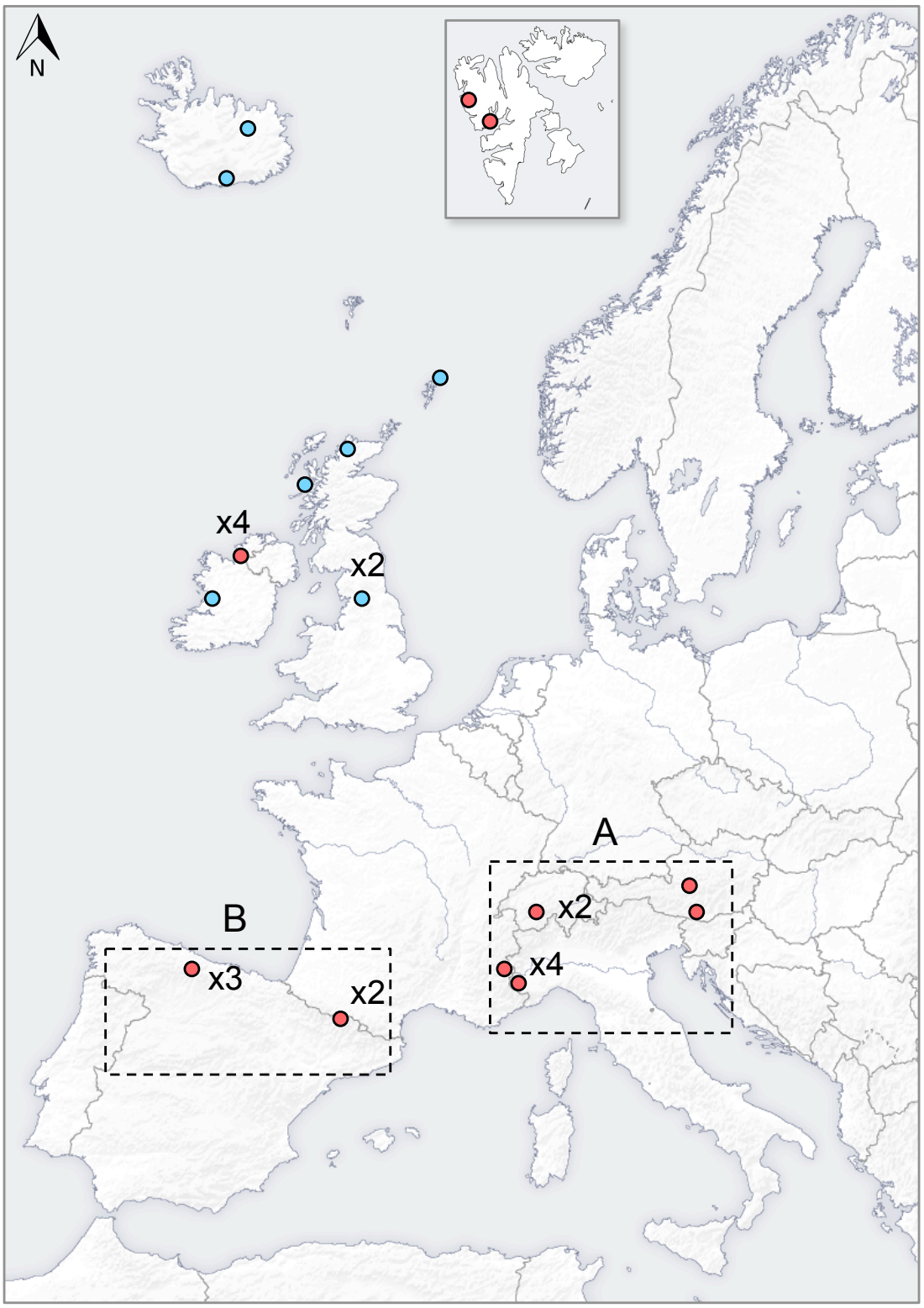


Figure 2.2: Distribution map of the populations sampled for *Arenaria ciliata*, (red circles) and *A. norvegica* (blue circles) with Svalbard inset. Where more than one population was sampled from an area this is indicated. Detailed regional maps are provided for the Alps (Box A) and Spain (Box B) in Figure 2.3.

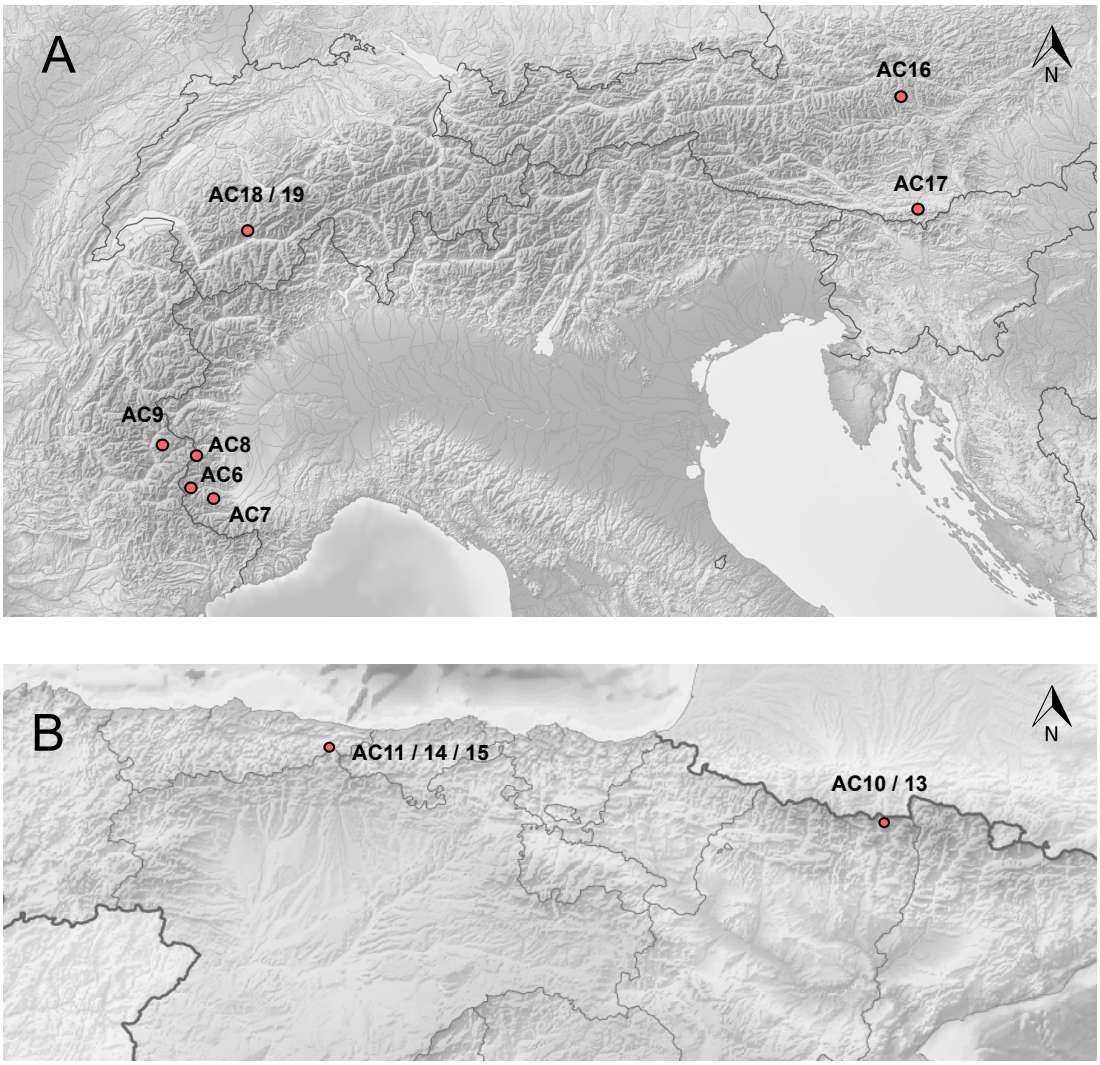


Figure 2.3: Regional maps showing the distribution of the populations sampled of *Arenaria ciliata* in The Alps (A) and Spain (B).

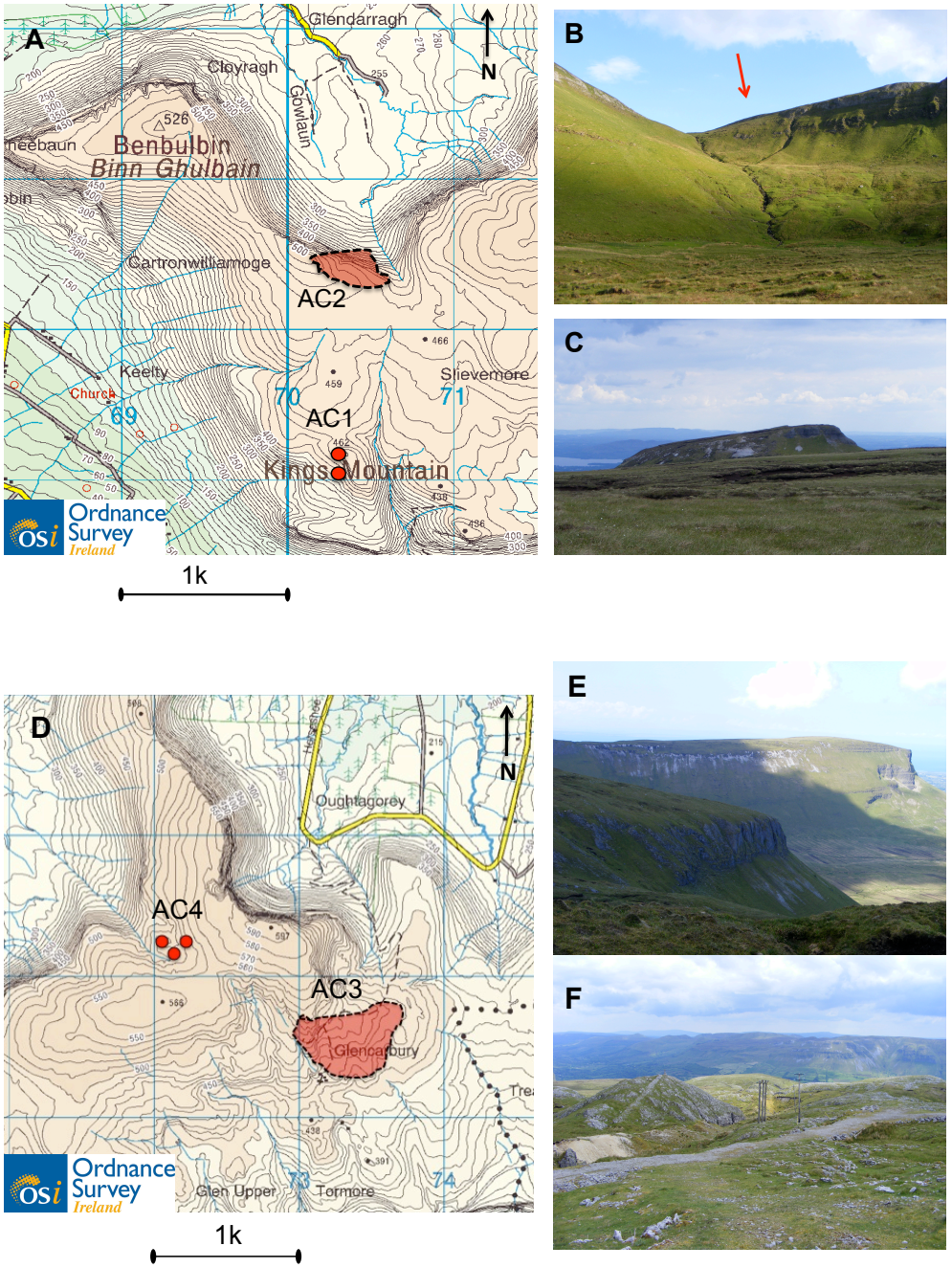


Figure 2.4: A - Map of populations AC1 and AC2 sampled in Ben Bulben. The red areas show the extent that the surveyed populations covered. B - Gowlaun Valley seen from the north where AC 2 occurs. The red arrow shows where *Arenaria ciliata* was first recorded on the limestone scree slopes and cliffs. C - AC1 at Kings Mountain seen from the north. D - Map of populations AC3 and AC4 sampled in Ben Bulben. E - Ben Bulben Plateau (in distance) viewed looking west from Ben Wiskin, with upper reaches of Gowlaun valley visible in middle foreground where AC2 occurs. F - AC3 at Glencarbury Mine. Copyright Permit No. MP 002513.

Figure 2.5: *Arenaria ciliata* growing in shallow calcareous soil with limestone scree at Ben Bulben, Co. Sligo.

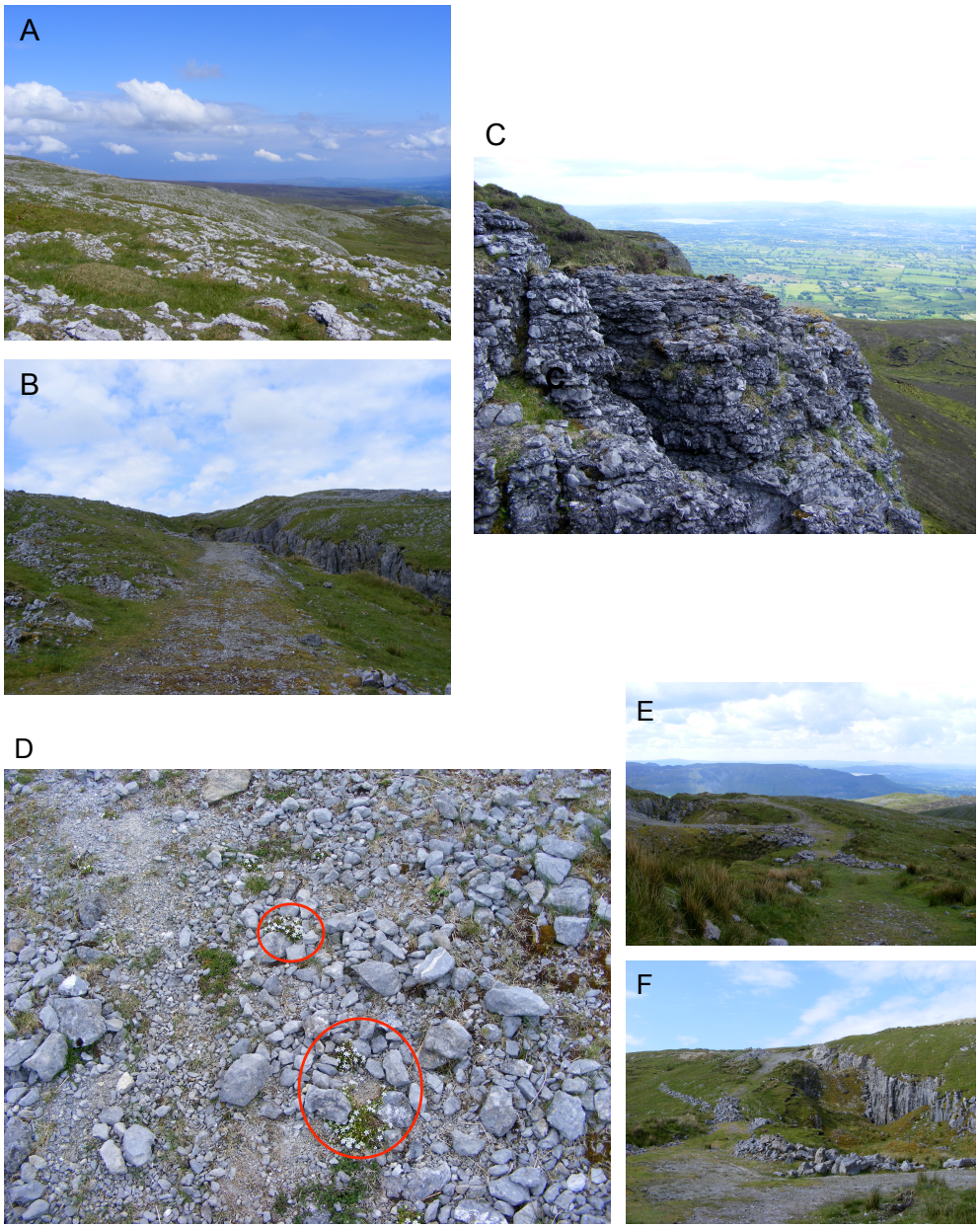


Figure 2.6: Habitats and landforms where *Arenaria ciliata* occurred frequently in Ben Bulben. A - Open mountain grassland interspersed with limestone scree and rocks on basic soil (Glencarbury mine - AC 3). B - Rough limestone gravel pathways in Glencarbury Mine with shallow soils. C - Limestone outcrops and cliffs (Kings Mountain AC1). D - Close up of *A. ciliata* growing on the gravel limestone paths in Glencarbury Mine (AC3). The red circles indicate some plants of *A. ciliata* and there are at least eight individuals in this photograph. E - The gravel pathways which are interwoven throughout Glencarbury Mine and surrounding habitat which includes the open upland grassland. F - Hills and small cliffs cut into the limestone at Glencarbury Mine where *A. ciliata* also occurs.

Table 2.3: Description of the sampled *Arenaria ciliata* populations

Pop.	Alt.	Habitat and Associated Geology	Pop. size	[a]Sample Size
AC1(Ireland)	463m	Steep limestone cliffs	approx. 60	29
AC2 (Ireland)	465m	Rocky limestone outcrops and steeper parts of cliffs in grassland	>1000	30
AC3 (Ireland)	499m	Grows abundantly on gravelly limestone pathways in mine and also on rocks in grassland	>>1000	30
AC4 (Ireland)	538m	Rocky grassland and limestone outcrops	>200	30
AC6 (Italy)	2057m	Stony pathway and adjacent in rocky grassland on limestone	>1000	23
AC7 (Italy)	2386m	Gravelly limestone pathway	>50	1
AC8 (Italy)	2338m	Patchy grassland interrupted by large rocks on limestone	>50	4
AC9 (France)	2506m	Stony pathway and adjacent in rocky grassland on limestone	>50	6
AC10 (Spain)	1804m	Consolidated roadside material beside road - shitz quartz gravel	>50	20
AC11 (Spain)	2050m	Bare karst limestone with pockets of shallow soil >50	20	
AC13 (Spain)	1834m	Limestone ridge	> 100	16
AC14 (Spain)	2325m	Bare karst limestone with extensive pockets of shallow soil	>50	16
AC15 (Spain)	1933m	Bare karst limestone with extensive pockets of shallow soil	>50	20
AC16 (Austria)	2000-2135m	Limestone outcrop with underlying silieous red sandstone	>100	27
AC17 (Austria)	2100-2139m	Limestone stable scree pathways with intermediate alpine grassland and many cushion plants. High diversity of alpines. Red sandstone and limestone	>100	20
AC18 (Switzerland)	2307m	Alpine grassland interspersed with limestone stable scree pathways	>50	>30
AC19 (Switzerland)	2257m	Alpine grassland interspersed with limestone stable scree pathways	>100	>30
[b]AP2 (Norway)	-	*Salix polaris*, *S.oppositifolia* dominated community - *Silene acaulis* very common	-	2
AP2 (Norway)	-	*Salix polaris*, *S.oppositifolia* dominated community - *Silene acaulis* very common	-	5

[a] The number of individuals sampled in each population

[b] *Arenaria ciliata* subsp. *pseudofrigida*

Figure 2.7: *Arenaria ciliata* (indicated in the red circle) growing alongside *Minuartia verna* (white petal tufted plant in the foreground) and many other alpines on Limestone shallow soils in the Pyrenees.

Figure 2.8: Leaf habit *Arenaria ciliata*. Obovate leaves with marginal ciliate hairs extending from base. The coverage and density of hairs varies between populations and regions. The leaves of this individual contain many glandular hairs which are present on the sides of most leaves extending to the mid point. The extent of ciliation can vary between regions and populations.

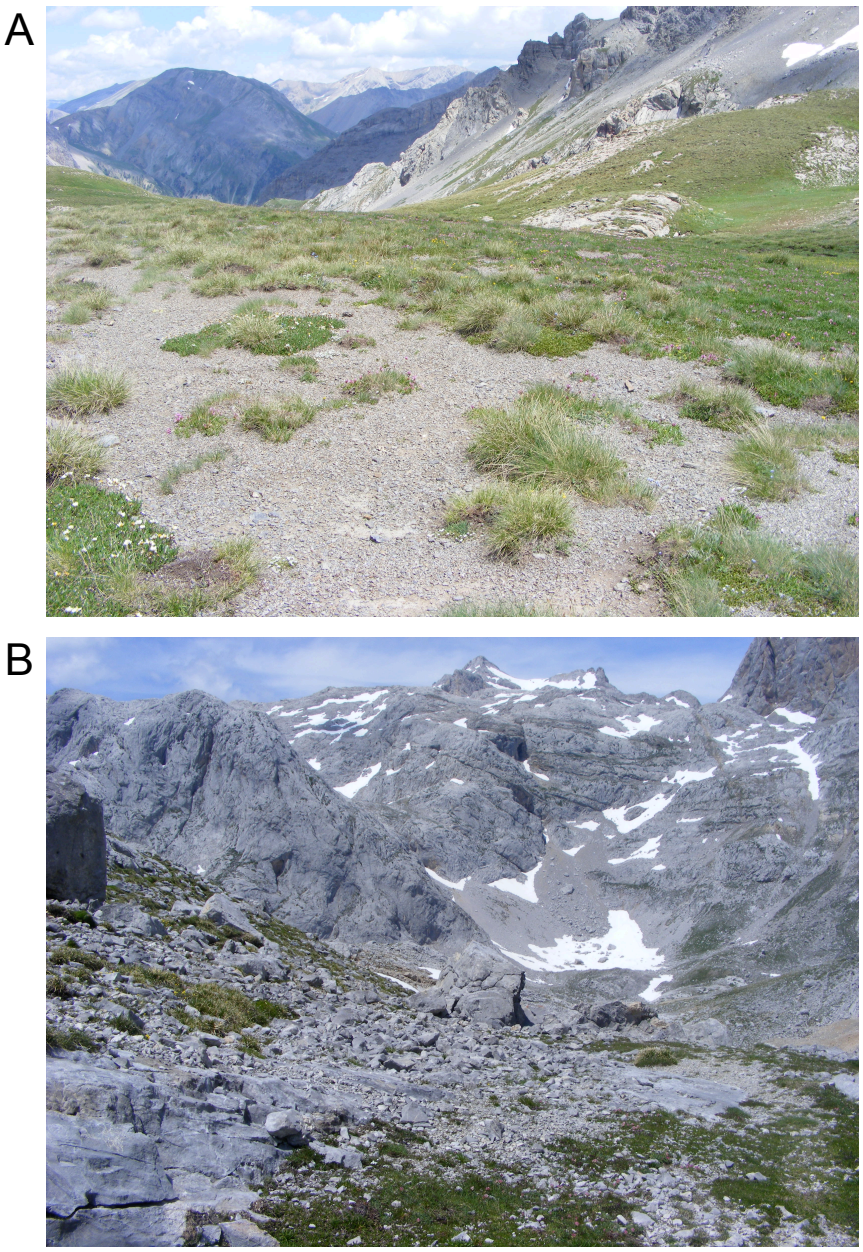


Figure 2.9: Typical habitats and landforms where *Arenaria ciliata* is found in the major mountain ranges in Europe. A - High altitude alpine grassland with exposed and shallow soils on limestone scree bedrock in the Austrian Alps co occurring with plants such as *Dryas octopetala* c. 2500m altitude. B - Sparsely vegetated limestone scree slopes in Picos De Europa c.2300m.

Table 2.4: Average % cover for plant species recorded with and without *Arenaria ciliata* from the four sampled populations at Ben Bulben

Species/cover	***A. ciliata* Quadrat % cover**	**[a]C. Quadrat % Cover**
Agrostis spp.	2.4	3.8
Alchemilla vulgaris	0.1	0.3
Asplenium spp.	0.2	0.0
Anthoxanthum odoratum	1.8	0.0
Arenaria ciliata	1.9	0.0
Bellis perennis	1.3	1.2
Calluna vulgaris	0.4	0.0
Carex spp.	0.1	0.3
Cerastium fontanum	0.3	0.3
Cirsium spp.	0.0	0.0
Epilobium brunnescens	2.3	4.8
Euphrasia sp.	0.3	0.2
Festuca spp.	11.5	13.0
Galium saxatile	0.0	0.0
Linum catharticum	0.1	0.0
Luzula sp.	0.0	0.2
Oxalis acetosella	0.9	1.3
Prunella vulgaris	0.7	1.3
Sagina nodosa	0.2	0.1
Saxifraga aizoides	0.4	0.0
Saxifraga tridactylites	0.2	0.3
Sesleria caerulea	7.9	2.2
Silene acaulis	1.4	5.0
Thymus polytrichus	6.9	6.3
Viola spp.	1.1	1.2
Taraxacum spp.	0.3	0.0
Bare earth	5.3	3.8
Bare rock	53.6	50.8
Total vegetation	46.8	45.8
Bryophytes	17.5	20.8
Grasses	6.9	4.0

[a] control quadrat - quadrat which did not contain *A. ciliata*

2.4.2 *Arenaria norvegica*

Altitude, habitat, associated geology, population size and the number of sampled individuals for each population for *Arenaria norvegica* are shown in Table 2.5. A distribution map showing the location of the site where *A. norvegica* is found in the Burren, Co. Clare are shown in Figure 2.10. Typical limestone pavement habitat where *A. norvegica* occurs on Black Head in the Burren is shown in Figure 2.11. Close up photographs of *A. norvegica* growing in shallow soils in the cracks of limestone pavement in the Burren and unconsolidated glacial debris in Iceland are shown in Figure 2.12.

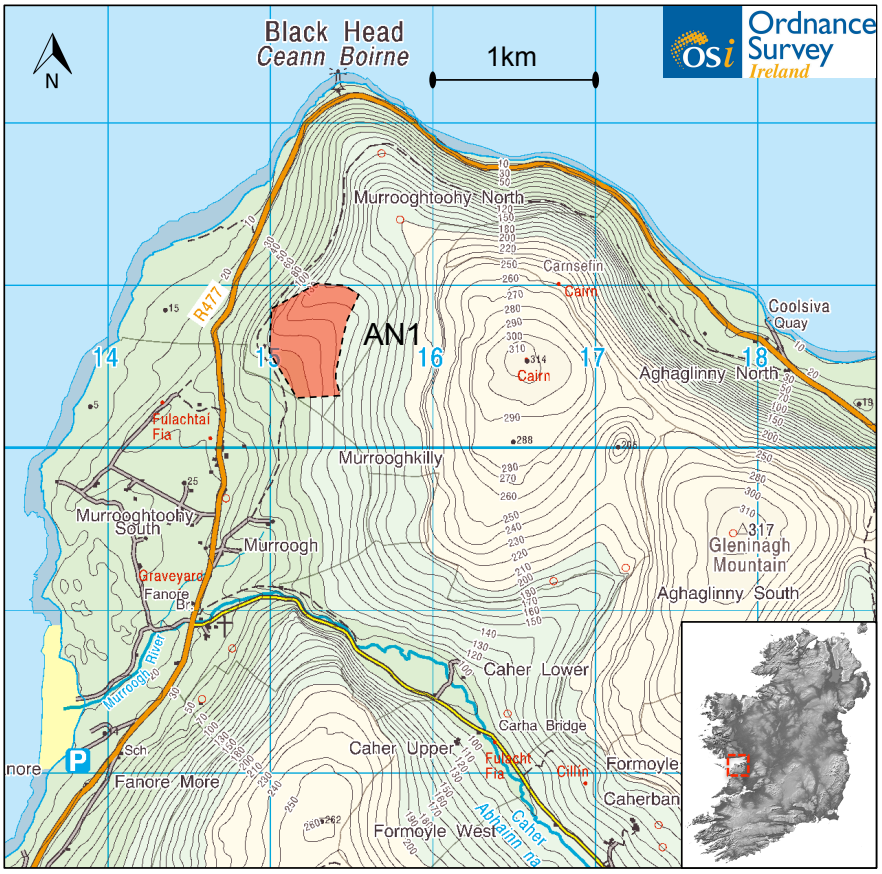


Figure 2.10: Ordnance Survey map of the Irish population of *Arenaria norvegica* (AN1), sampled in Black Head, The Burren, Co. Clare. The red area shows the extent that the surveyed population of *Arenaria norvegica* covered. Copyright Permit No. MP 002513.

Table 2.5: Altitude, habitat, associated geology, population size and no. of sampled individuals for *Arenaria norvegica*

Pop.	Alt.	Habitat and Associated Geology	Pop. size	Indivs.[a]
AN1 (Ireland)	79m ± 16.5m	pockets of gravelly shallow soils in limestone pavement with surrounding area sparsely vegetated (Limestone - Carboniferous)	approx. 60	31
AN2 (England)	-	shallow skeletal soils on sparsely vegetated level or slightly inclined limestone exposures	>1000	14
AN3 (England)	-	shallow skeletal soils on sparsely vegetated level or slightly inclined limestone exposures	>1000	7
AN4 (Rum, Scotland)	-	fine gravel on low terraces or shattered bed-rock (Ultrabasic - Peridotite)	>1000	20
AN5 (Scotland)	378.3 ± 43.4	fine gravel on low terraces or shattered bed-rock (Limestone - Durness)	>100	27
AN6 (Shetland)	-	fine gravel on low terraces or shattered bed-rock and crevices in rock outcrops (Ultrabasic - Serpentine)	>5000	32
AN7 (Iceland)	565m	Volcanic tuff and aeolian sediments	-	2
AN8 (Iceland)	-	Exposed pockets of shallow soil on accumulation of unconsolidated glacial debris (Basalt)	-	20

[a] The number of individuals sampled in each population

A

B

Figure 2.11: Typical habitats where *Arenaria norvegica* is found in Ireland (A) and Iceland (B). A - Karst carboniferous limestone pavement typical on Black Head where *Arenaria norvegica* was sampled for population AN1. B - Exposed pockets of shallow soil on accumulation of unconsolidated glacial debris in Iceland, where population AN8 was sampled.

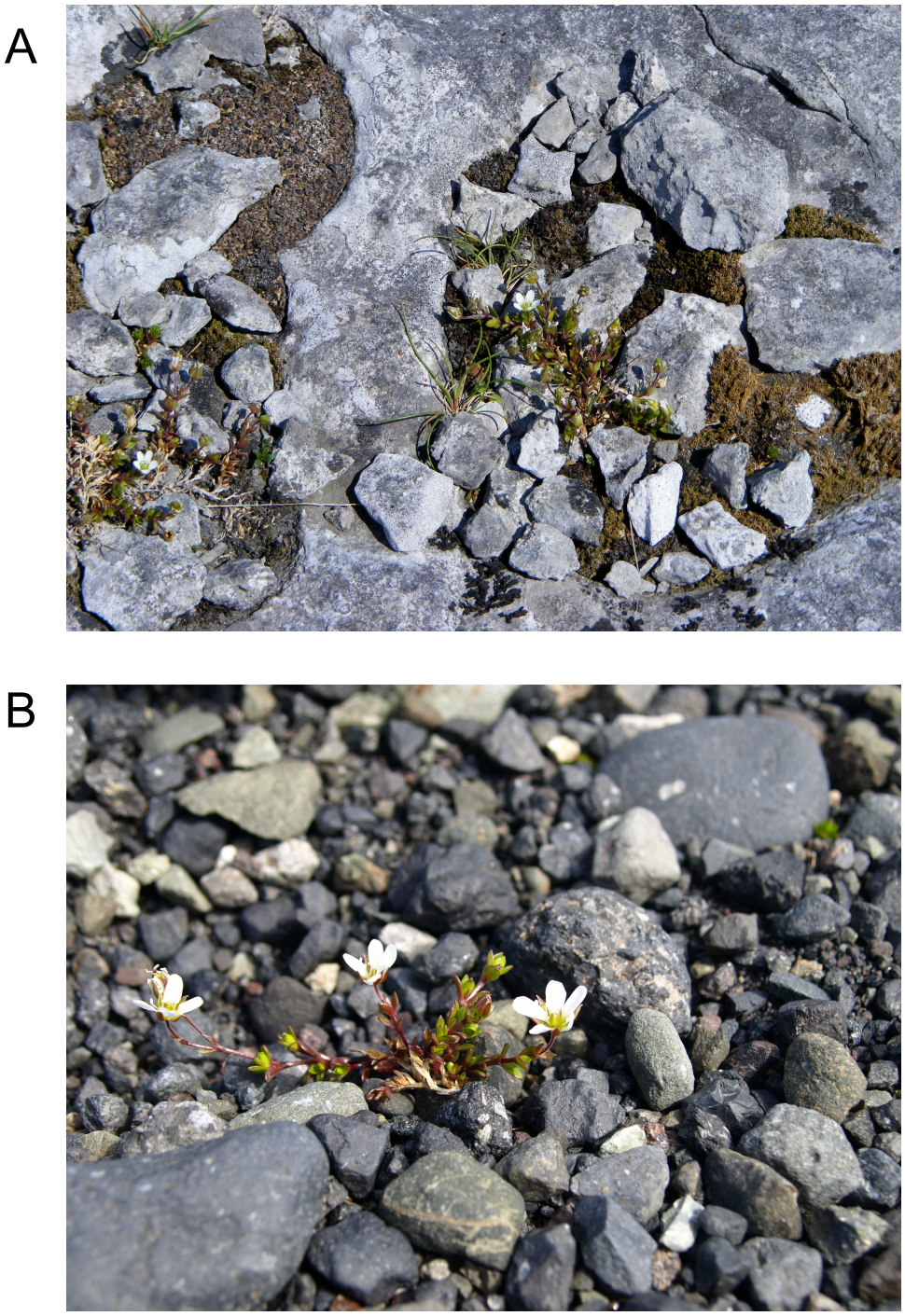


Figure 2.12: *Arenaria norvegica* in the site sampled from Ireland (A) and Iceland (B). A - Typical habitat that this species occupies in The Burren, with shallow exposed soils in the limestone pavement. B - Close up of *A. norvegica* occurring on exposed unconsolidated glacial debris in Iceland.

2.5 General molecular biology techniques

2.5.1 DNA extraction

A list of the chemicals used and how they were prepared is provided in Section 2.1. Total genomic DNA was extracted from approximately 20mg of leaf tissue dried with silica gel using a modified 2 x CTAB extraction (Doyle and Doyle, 1987). Dried leaf tissue was powdered in 2ml microfuge tubes for 30 seconds at 30 Hz on a mixer mill (Retsch Mixer Mill 300) using two ball bearings. 750μl of 2% CTAB and 3μl of β-mercaptoethanol was added to each sample, mixed and pre-heated at 65 °C for 10 minutes. 750μl of SEVAG was added to each sample and each tube was placed on a rocker (New Burnswick Scientific, Edison, N.J., U.S.A.) for 10 minutes at 200RPM. The samples were then centrifuged at 7, 500 x g for 10 minutes. The supernatent containing DNA (approximately 500μl) was transferred to new 1.5ml eppendorf tubes.

In order to precipitate the DNA, 800μl of 100% ethanol (pre-cooled at -20 °C) was added to each sample and left in a -20 °C freezer for at least 40 minutes. The samples were then centrifuged at 7, 500 x g for 10 minutes. All liquid was poured off, leaving pelleted DNA at the bottom of the tubes. 750μl of 70% ethanol was added to wash the DNA pellets. The pellets were dislodged in the ethanol and the tubes were agitated for at least 40 minutes at 200 RPM on a rocker (New Burnswick Scientific, Edison, N.J., U.S.A.). The samples were centrifuged at 7, 500 x g for 5 minutes. The ethanol was poured out from each tube, taking care not to pour off the DNA pellets. The tubes were left open for at least 1 hour in the fume cupboard, or with ventilation off overnight. Once the tubes were free of ethanol, the DNA was re-suspended in 200μl TE $_{0.1}$ buffer and left aside at room temperature for at least 4 hours (or 1-2 days at 4 °C) to get all the nucleic acid into solution. Approximately 10% of the extracts were quantified photometrically using a nanodrop (Nanodrop 1000, Mason) and the DNA concentration and purity estimated. The quality of the extracted DNA was checked on 1.5% TAE-agarose gels and visualized with a UV transilluminator (Syngene G:Box, Mason Technology).

2.5.2 Preparation of agarose gel

1.5% Agarose gels were prepared by adding 3g of agarose to 200ml of 1 x TAE buffer (diluted from 50 X TAE buffer with ddH_20). This solution was heated in a microwave for approximately two minutes until the agarose was completely dissolved. The gel was allowed to cool slightly and 5μl of SybrSafe(Invitrogen) stain was added to 80ml of the molten agarose. Gels which were used to test the preselective AFLP products, were stained with 5μl of Ethidium Bromide (10mg/ml). A comb was placed in the gel casting tray and the tray ends taped with masking tape. The agarose was then poured into the casting tray and allowed to set.

2.6 Polymerase chain reactions for amplifying chloroplast and nuclear DNA

2.6.1 PCR conditions for *Arenaria ciliata* and *A. norvegica*

Four separate chloroplast regions: (*psbA-trnH*, *rpl*32-trnL,*trnK-matK* and *matK*, (highlighted in the chloroplast genome map for *Arabidopsis* in Figure 2.13) and the Internal Transcriber Spacer regions of the nuclear ribosomal DNA locus (ITS1-5.8S-ITS2) (Baldwin *et al.*, 1995) were amplified for *Arenaria ciliata* and *A. norvegica* using PCR. PCR primers used for these analyses are shown in Table 2.6. A portion of the PCR work was carried out in the Molecular Systematics Lab., Real Jardin Botanico (RJB), Madrid and the remainder carried out in the Molecular Ecology Lab, NUI Maynooth at a later stage. The chloroplast regions surveyed in this investigation generated polymorphic loci suitable for phylogeographic analysis. A list of individuals which were sequenced including the region that was sequenced and which lab the work was carried out in, is shown in the appendices in Tables D.1 and D.2. In total, 112 *Arenaria ciliata* and *A. norvegica* individuals were sequenced for *matK*, 33 individuals were sequenced for the chloroplast regions: *psbA-trnH*, *rpl*32-trnL,*trnK-matK*, and 33 individuals were sequenced for the ITS region.

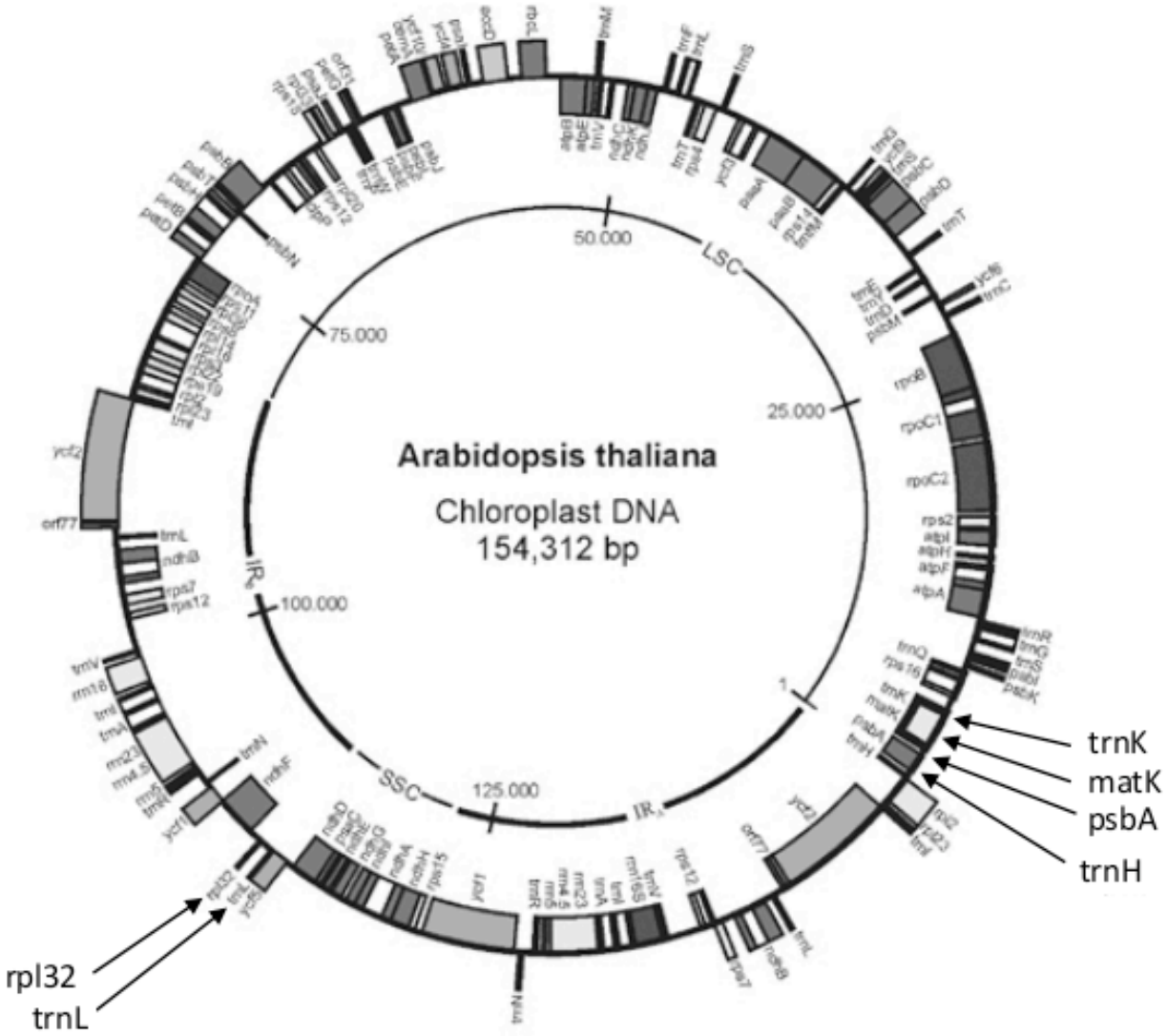


Figure 2.13: Map of the chloroplast genome of *Arabidopsis thaliana* modified from Schweer *et al.* (2006) with the relevant gene regions used in this study indicated by the arrows.

PCR master mixes were prepared in 1.5ml reaction tubes (Eppendorf) according to Table 2.7 and reactions were carried out in 96 well plates (Sorenson Bioscience). 1μl of genomic DNA extract was added to 24μl of the PCR master mix in each well. To optimize the PCR reaction, the annealing temperatures and extension time was varied in separate PCR reactions (Table 2.6). PCRs were carried out according to Table 2.8 on an EP Gradient Mastercycler (Eppendorf).

5μl of PCR product was run with 1μl of 5 x loading dye (Bioline) on a 1% agarose gel (Invitrogen) at 120V for 15 minutes. Gels were visualized with a UV transillluminator (Syngene G:Box, Mason Technology). The PCR products were sent off to MacroGen Sanger Sequencing Service, The Netherlands, where the PCR products were purified and sequenced.

14 individuals of *Arenaria ciliata* and *A. norvegica* were sequenced at a later date and the *matK* region (390f - 1440r) was amplified for these individuals in the Molecular Ecology Lab., NUI Maynooth using the components outlined in Table 2.7. All other procedures for PCRs for these individuals are outlined in Table 2.8. The reactions were carried out in 0.2ml Thermowell Gold PCR tubes (Corning, USA) and the PCRs carried out in a Peltier Thermal Cycler (PTC-200, Bio-Sciences, Dublin). These individuals are indicated in the appendices in Tables D.1 and D.2.

Table 2.6: PCR conditions and primer sequences used for the amplification of variable DNA regions sequenced for *Arenaria ciliata* and *A. norvegica*

Primer	Sequence	Annealing Temp.	Annealing Time	Reference
psbA f	CGA AGC TCC ATC TAC AAA TGG			
trnH r	ACT GCC TTG ATC CAC TTG GC	52 °C	1 min.	Hamilton (1999)
*rpl*32 f	CAG TTC CAA AA A AAC GTA CTT C			
TrnL r	CTG CTT CCT AAG AGC AGC GT	54 °C	1 min.	Shaw *et al.* (2007)
trnK (3914f)	TGG GTT GCT AAC TCA ATG G			
matK (1470r)	AAG ATG TTG AT(CT) GTA AAT GA	52 °C	2 min.	Johnson and Soltis (1994)
matK (390f)	CGA TCT ATT CAT FIC AAT ATT TC			Cuenoud *et al.* (2002)
matK (1440r)	GTG TTT ACG AGC YAA AGT TC	52 °C	1 min.	Fior *et al.* (2006)
ITS4 f	TCC TCCG CTT ATT GAT ATG C			
ITS5 r	GGA AGG AGA AGT CGT AAC AAG G	58 °C	min.	White *et al.* (1990)

Table 2.7: Reagents for PCR amplification of *Arenaria ciliata* and *A. norvegica*

Reageant	Manufacturer RJB / NUIM	Concentration	Volume
H_20			26.75μl
Taq Buffer	Bioline / Applied Biosystems	10 X	10μl
dNTP's	Bioline / Applied Biosystems	2.5mM	2μl
Forward Primer	Stab Vida / MWG	10μM	1μl
Reverse Primer	Stab Vida / MWG	10μM	1μl
BSA	Ecogen / New England Biolabs	1mg/ml	1μl
DNA		0.05 - 0.5 ng/μl	1μl
BioTaq Polymerase	Bioline / Applied Biosystems	5U/μl	0.3μl
$MgCl_2$	Bioline / Applied Biosystems	50mM	1.4μl

Table 2.8: Thermal cycler parameters for PCR amplification of *Arenaria ciliata* and *A. norvegica* DNA

Hold	PCR Amplification Cycle			No. of Cycles
	Denature	Anneal	Extension	
95 °C	-	-	-	1
1 min.	-	-	-	
-	95 °C	refer to Table 2.6	72 °C -	-
-	1 min.	refer to Table 2.6	1 min.	30
72 °C	-	-	-	1
1 min.	-	-	-	
4 °C	-	-	-	1
forever	-	-	-	

2.7 Amplified Fragment Length Polymorphism (AFLP)

The AFLP work was carried out in three different batches, each of which used a slightly different set of source reagents. Each batch was treated separately and scored accordingly.

1. Preliminary AFLP analysis was carried out for *Arenaria norvegica* and *A. ciliata* in the Molecular Ecology Laboratory, NUI Maynooth using the starter kit provided by Applied Biosystems. This was carried out to screen for informative primer combinations and to complete an initial analysis on a subset of populations. Carried out: 2009-2010. Batch name: NUIM 2009

2. AFLP analysis for a subset of populations including *Arenaria ciliata* using an established protocol with in-house reageants in the Molecular Systematic Laboratory in the Real Jardin Botanico, Madrid. Carried out: March 2011. Batch name: RJB

3. AFLP analysis carried out for *Arenaria norvegica* and *A. ciliata* in the Molecular Ecology Laboratory NUI Maynooth using the protocol established and implemented in the Molecular Systematic Laboratory in Madrid with in-house reagents. Carried out: summer 2012. Batch name: NUIM 2012

2.7.1 Screening for selective primers

An initial screening for selective primer pairs was performed on individuals from a range of taxa using the Selective Amplification Start-Up Module for Regular Genomes (Applied Biosystems). The selective primers consisted of the same sequence as the preselective primers with an additional three bases added at the 3' end. In the first screening, 32 primer pairs were used on a range of taxa from the Caryophyllaceae family which were sampled from the National Botanic Gardens Dublin. The species which were used in the preliminary screening include

Minuartia verna, *Spergularia marina*, *Cerastium fontanum*, *M. recurva*, *Silene alpestris*, *S. saxifraga*, *M. kashmirica*, *M. capillaceae* and *Arenaria ciliata*.

Seventeen selective primer pairs resulted in an electropherogram which contained more than 50 peaks. These primers were used in a second screening which were tested on a greater number of individuals and replicates in order to test the reproducibility of the selective primers. Three primer combinations that gave clear and reproducible bands and that showed variation within and between populations were chosen for further analysis. A sample electropherogram from GeneScan in Figure 2.14 depicts a representative AFLP profile demonstrating many clear AFLP markers which would have been screened further in the selective primer screening process for the presence of reproducible polymorphic peaks.

The selective primer combinations that were chosen for all 3 batches for *Arenaria ciliata* and *A. norvegica* are: *Eco*RI ACT - *Mse*I CAG, *Eco*RI AGC - *Mse*I CTG and *Eco*RI AGC - *Mse*I CAA. The selective primers used for all batches of work in NUIM2009, RJB and NUIM2012 are the same except for the fluorescent label attached to the *Eco*RI primer.

For the 1st and 2nd batch (NUIM2009 / RJB), *Eco*RI AGC was labelled with NED, and the *Eco*RI ACT labelled with 6-FAM. The fluorescent dyes labelled to the selective *Eco*RI primers for the 3rd batch (NUIM 2012) are *Eco*RI ACT with 5-FAM and *Eco*RI AGC with Cy3. The full primer sequence for the selective primers along with all primers involved in every step of AFLP analysis is detailed in Table 2.9.

2.7.2 Capillary electrophoresis

Once the selective AFLP PCR primers were chosen a full AFLP analysis was carried out using capillary electrophoresis. A list of the 212 individuals that were subjected to capillary electrophoresis for AFLP analysis is detailed in the Appendix Tables F.1, F.2 and F.3. Duplicate samples were also included in the analysis (approx. 5% of the total number of samples) for each primer combination.

The selective amplification products generated for the first batch (NUIM2009) were subjected to capillary electrophoresis on an ABI 310 Prism Genetic Analyzer. 0.5μl of each amplification product was placed in a 200μl 0.2ml Thermowell Gold

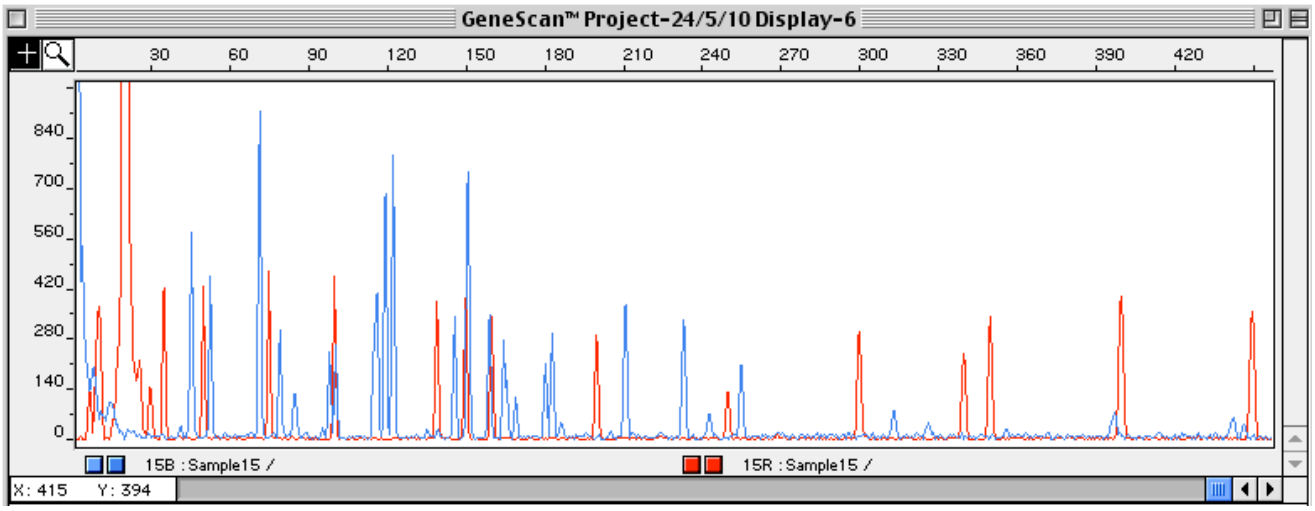


Figure 2.14: An example AFLP profile for an individual of *Arenaria ciliata* from Ben Bulben (AC3.1) taken from GeneScan from the primer screening analysis for the selective primer combination *Eco*RI - ACT *Mse*I - CAG (blue peaks) and Rox 500bp size standard (red peaks) . On the x-axis is the size of the fragment (in bp) and the y-axis is the amplitude of the fragment.

Table 2.9: AFLP *Eco*RI and *Mse*I primers used in all steps for the AFLP analysis.

AFLP reaction step	Primer sequence
Restriction ligation	Forward *Eco*RI 5'-CTC GTA GAC TGC GTA CC-3'
Restriction ligation	Reverse *Eco*RI 5'-AAT TGG TAC GCA GTC TAC-3'
Restriction ligation	Forward *Mse*I 5'-GAC GAT GAG TCC TGA G-3'
Restriction ligation	Reverse *Mse*I 5'-TAC TCA GGA CTC AT-3'
Preselective PCR	Forward *Eco*RI 5'-GAC TGC GTA CCA ATT C-3'
Preselective PCR	Reverse *Mse*I 5'-GAT GAG TCC TGA GTA A-3'
Selective PCR	Forward *Eco*RI 5' - GAC TGC GTA CCA ATT C ACT - 3'
Selective PCR	Reverse *Mse*I 5' - GAT GAG TCC TGA GTA A CAG - 3'
Selective PCR	Forward *Eco*RI 5' - GAC TGC GTA CCA ATT C AGC - 3'
Selective PCR	Reverse *Mse*I 5' - GAT GAG TCC TGA GTA A CAA - 3'
Selective PCR	Forward *Eco*RI 5' - GAC TGC GTA CCA ATT C AGC - 3'
Selective PCR	Reverse *Mse*I 5' - GAT GAG TCC TGA GTA A CTG - 3'

PCR tube (Corning, USA) with 12μl of formamide and 0.5μl of the size standard (GeneScan-500 ROX, Applied Biosystems). The tubes were mixed gently and placed in a centrifuge at 7, 500 x g for 5 seconds. The tubes were then placed in a Peltier Thermal Cycler (PTC-200, Bio-Sciences, Dublin) and heated to 95 °C for

5 minutes. The samples were then transferred into 0.5ml genetic analyser sample tubes (Applied Biosystems) and a septa (Applied Biosystems) placed on top of each tube. These tubes were placed on an automated sequencer (ABI Prism 310 Genetic Analyzer, Applied Biosystems) for capillary electrophoresis. Raw data were collected and aligned with the internal size standard (GeneScan-500 ROX, Applied Biosystems) using the ABI Prism GeneScan Analysis Software (Version 2.1.1., Applied Biosystems).

The AFLP reactions using generic reagents for batches 2 (RJB) and 3 (NUIM2012) were analyzed on a 3730 capillary sequencer (Applied Biosystems). The size standard used for the RJB reactions was Liz500 and for the NUIM2012 reactions the Rox500 size standard was used.

2.7.3 Scoring AFLP fragments

The results from each batch of work were scored separately from NUIM2009, RJB and NUIM2012 for each selective primer combination. The NUIM2009 results were imported from GeneScan into Genotyper (Version 2.5.2) for scoring of the fragments. For the RJB and NUIM2012 batch, the raw data were imported and scored in GeneMapper (Version 4.0).

In Genotyper the scoring step was performed using a macro which specified the conditions of the peaks to be selected: peaks in the size range 0.00 to 50.00bp were omitted; peaks with an amplitude height less than 32% of the highest peak amplitude in a categorys range were omitted; peaks within 0.00 to 1.60bp by a higher, labelled peak were omitted; and finally peaks succeeded by a higher, labelled peak within 0.00 to 3.00bp were also omitted. In Genemapper an analysis macro was also set up to discriminate which peaks to be labelled. An analysis range from 50-500bp was set with a bin width of 1.0 bp. The peak amplitude height threshold was set to 200 and the max peak width set to 1.5bp. In both scoring programs however it was impossible to generate a correct automated system of scoring, so for every primer combination in each batch of work the peaks were checked manually. A peak presence/absence binary matrix of samples for all valid peak fragments was generated for each primer combination and every valid peak for each individual was checked visually and added to the matrix.

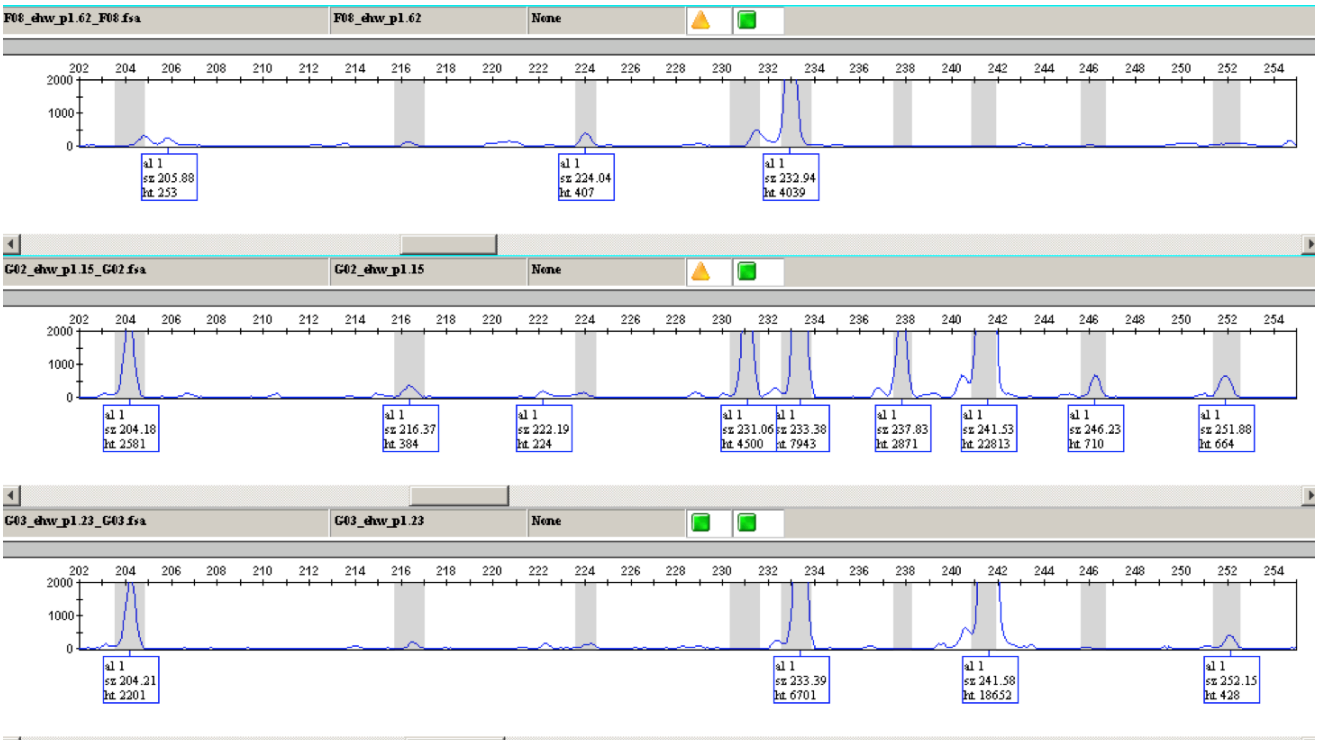


Figure 2.15: A screen shot taken for three samples viewed in GeneMapper for the selective primer combination *Eco*RI - ACT *Mse*I - CAG for the range 202-254 bp. On the x-axis is the size of the fragment (peak) and the y-axis is the amplitude of the peak.

Any electrophoregrams which displayed ambiguous profiles were omitted. Duplicate samples were used to asses any ambigous alleles which were omitted from the analysis. Samples with markedly higher or lower fragment presences were excluded on the basis that they indicated sample contamination and PCR error, respectively. Peaks of low intensity (<50 in amplitude height) were included in the analysis when unambiguous scoring was possible. Duplicate samples were included in this scoring process (approx. 5% of the total number of samples) for each primer combination. These were used as a reference to identify AFLP fragments that exhibited ambiguous peaks which were subsequently omitted from the analysis. Figure 2.15 shows a screen shot taken from Genemapper which shows electrophoregrams for 3 individuals for the selective primer combination *Eco*RI ACT - *Mse*I CAG. The grey areas are called bins which are manually created to help guide where the peak fragments are positioned. In this figure the presence of polymorphic fragments (e.g. at position bp 204), peaks which are present in all 3 individuals (233 bp) and ambigous fragments at 216 bp.

2.7.4 AFLP analysis using Applied Biosystems protocol - 1st Batch: NUIM 2009

2.7.4.1 Digestion

The Applied Biosystems Starter Kit (for regular genomes) plus additional reagents required for preselective and selective amplification were used to screen and optimize the Amplified Fragment Length Polymorphism analysis of *Arenaria ciliata* and *A. norvegica*. An enzyme master mix was made up according to Table 2.10 and added to each sample of 5.5μl of genomic DNA ranging from a concentration of 10 - 200ng/μl to make up a final volume of 50μl with BSA in each tube (or well in a 48 well plate (ThermoScientific, UK)). The samples were incubated for 1 hour at 37 °C in a Peltier Thermal Cycler (PTC-200, Bio-Sciences, Dublin).

Table 2.10: Digestion step of AFLP analysis.

Reageant	Manufacturer	Concentration	Volume
EcoRI	New England Biolabs	40U/	1μl
Mse1	New England Biolabs	10U/	1μl
10 x EcoRI buffer	New England Biolabs	1:10	5μl
10 x MseI buffer	New England Biolabs	1:10	5μl
BSA	New England Biolabs	1:100	add to make up 50μl

2.7.4.2 Restriction-Ligation

The enzyme master mix for the restriction-ligation step was prepared in a 0.2ml Thermowell Gold PCR tube (Corning, USA) as detailed in Table 2.11. Sterile water was added to bring the total volume to 100μl for 100 samples. The solution was mixed gently and spun down in a microcentrifuge (Eppendorf) for 10 seconds at 1,400 x g. The master mix was stored on ice until ready to be aliquoted into the individual reaction tubes (or later, into wells on a 96-plate).

10μl of the digestion product was added to a 0.2ml Thermowell Gold PCR tube (Corning, USA) or well (AB PCR plates, ThermoScientific, UK), along with 5.5μl the T4 Ligase/adaptor mix which was prepared as detailed in Table 2.12.

Table 2.11: Master mix for AFLP restriction ligation - volumes listed here are for 100x samples

Reageant	Manufacturer	Concentration	Volume
T4 Liagse Buffer	New England Biolabs	10X	10
NaCl	Sigma	0.5M	10μl
BSA	New England Biolabs	1mg/ml	5μl
EcoRI	New England Biolabs	500 Units	25μl
MseI	New England Biolabs	100 Units	10μl
T4 DNA Ligase	New England Biolabs	100 Weiss Units	0.25μl

Table 2.12: Reagents for AFLP restriction ligation

Reageant	Manufacturer	Concentration	Volume
T4 Liagse Buffer	New England Biolabs	10X	1μl
NaCl	Sigma	0.5M	1μl
BSA	New England Biolabs	1mg/ml	0.5μl
EcoRI adaptor	Applied Biosystems	-	1μl
MseI adaptor	Applied Biosystems	-	1μl
Enzyme Master Mix	-	-	1μl

Adaptor pairs (Applied Biosystems) were prepared by heating at 95 °C for 5 minutes and cooled down to room temperature over a 10 minute period. The samples were then spun in a microcentrifuge (Eppendorf) for 10 seconds at 1,400 x g.

The samples were mixed thoroughly and placed in a microcentrifuge for 10 seconds at 1,400 x g. The samples were then incubated in a Peltier Thermal Cycler (PTC-200, Bio-Sciences, Dublin) at 37 °C for 2 hours. 189μl of $TE_{0.1}$ was added to each restriction-ligation reaction and mixed thoroughly. The samples were then stored at 4 °C for up to one month, or at −20 °C for long term storage.

2.7.4.3 Preselective Amplification

To prepare the samples for preselective amplification the following components were combined in a 0.2ml Thermowell Gold PCR tube (Corning, USA) (or in 96-well plate): 4μl diluted DNA prepared by restriction-ligation, 1μl AFLP pre-

selective primer pairs (Applied Biosystems) and 15μl AFLP Core mix (Applied Biosystems). The samples were placed in a Peltier Thermal Cycler (PTC-200, Bio-Sciences, Dublin) and the parameters for this amplification step are outlined in Table 2.13.

Table 2.13: Thermal cycler parameters for pre-selective amplication

Hold	PCR Amplification Cycle			No. of Cycles
	Denature	Anneal	Extension	
72 °C	-	-	-	1
2 min.	-	-	-	
-	94 °C	56 °C	72 °C -	
-	20 sec.	30 sec.	2 min.	30
60 °C	-	-	-	1
30 min.	-	-	-	
4 °C	-	-	-	1
forever	-	-	-	

5μl of each reaction was run on a 1.5% agarose gel in TAE buffer at 70V for 1-2 hours and viewed on a UV transilluminator (Syngene G:Box, Mason Technology) where a smear of product from 100-1500bp should be visible. 10μl of the preselective amplification product was diluted with 190μl of $TE_{0.1}$ and stored at 4 °C and at −20 °C for long term storage.

2.7.4.4 Selective Amplification

The following components were combined in a 200μl 0.2ml Thermowell Gold PCR tube (Corning, USA) (or 48 well plate (ThermoScientific, UK)): 3μl of diluted preselective amplification reaction product, 1μl of Mse1 Selective Primer at 5μM, 1μl of *Eco*RI Selective labelled Primer at 1μM and 15μl of core mix (Applied Biosystems).

Selective PCR was run on a PTC-200 thermocycler according to the parameters shown in Table 2.14. The selective products were stored at 4 °C after amplification.

Table 2.14: Thermal cycler parameters for selective amplification

Hold		Cycle		No. of Cycles
	Denature	Anneal	Extension	
94 °C	-	-	-	1
2 min.	-	-	-	
-	94 °C	66 °C	72 °C	10
		reduce by 1°C per cycle		
-	20 sec.	30 sec.	2 min.	
-	94 °C	56 °C	72 °C	20
	20 sec.	30 sec.	2 min.	
60 °C	-	-	-	1
30 min.	-	-	-	
4 °C	-	-	-	1
forever	-	-	-	

2.7.5 AFLP Analysis using generic reagent mixes - 2nd and 3rd batch: RJB and NUIM 2012

AFLPs were generated using a similiar protocol, at the Molecular Systematics Laboratory, Real Jardin Botanico (RJB), Madrid and the Molecular Ecology Laboratory in NUIM, for *Arenaria ciliata* and *A. norvegica* using generic reagent mixes from a protocol optimized in RJB Madrid.

2.7.5.1 Restriction-Ligation

The forward and reverse adaptor pairs were prepared by adding 25μl of EcoR1 forward adaptor at 100μM to 25μl of Mse1 reverse adaptor at 100μM and mixed with 450μl of TE buffer in a 1.5ml tube. The adaptors were then heated at 95 °C for 5 minutes and cooled down to room temperature over a 10 minute period. The samples were then spun in a microcentrifuge (Eppendorf) for 10 seconds at 1,400 x g. The restriction ligation enzyme master mix was prepared according to Table 2.15. 5.5μl of this mix was added to 5.5μl of template genomic DNA in each well of a 96 well PCR plate (Sorenson Bioscience) (RJB) or 48 well plate (ThermoScientific, UK) (NUIM2012). The samples were incubated for 2 hours at 37 °C in the EP Gradient Mastercycler (Eppendorf) (RJB) or Peltier Thermal

Cycler (PTC-200, Bio-Sciences, Dublin) (NUIM2012). The samples were then diluted with 90μl sterilized ddH_20.

Table 2.15: Reagents for AFLP restriction ligation

Reageant	Manufacturer RJB / NUIM2012	Concentration	Volume
H_20	-	-	0.325μl
T4 Liagse Buffer	TakaRa / New England Biolabs	10X	1.1μl
NaCl	Ecogen / Sigma	0.5M	1.1μl
BSA	Ecogen / Roche	1mg/ml	0.55μl
EcoRI Adaptor	Stab Vida / MWG	5μM	1μl
MseI Adaptor	Stab Vida / MWG	50μM	1μl
EcoRI	New England Biolabs	40 U/μl	0.125μl
MseI	New England Biolabs	10 U/μl	0.1μl
T4 DNA Ligase	TakaRa / New England Biolabs	340 U/μl	0.2μl

2.7.5.2 Preselective Amplification

Master mixes for the preselective amplification were prepared in 1.5ml tubes according to Table 2.16. 2.5μl of template DNA subjected to resctriction ligation was added to 10μl of the preselective master mix in a 96 well PCR plate (Sorenson Bioscience) (RJB) or 48 well plate (Thermo Scientific) (NUIM2012). The samples were then placed in the EP Gradient Mastercycler (Eppendorf) (RJB) or PTC (Bio-Sciences) for preselective PCR. The parameters for the thermocycling conditions for preselective PCR is detailed in Table 2.17 for both batches.

2.7.5.3 Selective Amplification

Master mixes for the selective amplification step were prepared in 1.5ml tubes according to Table2.18. 2.5μl of template DNA prepared by preselective amplification was added to 10μl of the selective master mix in a 96 well PCR plate (Sorenson Bioscience) in RJB or 48 well PCR plate (ThermoScientific) NUIM2012. The samples were then placed in the EP Gradient Mastercycler (Eppendorf) (RJB) or PTC-200 (Bio-Sciences). The parameters used for the thermocycling conditions for selective PCR for both batches are detailed in Table 2.19.

Table 2.16: Reagents for AFLP preselective amplification

Reageant	Manufacturer RJB / NUIM2012	Concentration	Volume
H_20	-	-	7.15μl
Ampli Taq Buffer	Applied Biosystems	10X	1.25μl
$MgCl_2$	Applied Biosystems	25mM	0.75μl
dNTPs	Ecogen / Promega	10mM	0.25μl
EcoRI Primer	Stab Vida / MWG	10μM	0.25μl
MseI Primer	Stab Vida / MWG	10μM	0.25μl
Ampli Taq	Applied Biosystems	5 U/μl	0.1μl

Table 2.17: Thermal cycler parameters for preselective amplication

Hold	PCR Amplification Cycle			No. of Cycles
	Denature	Anneal	Extension	
72 °C	-	-	-	1
2 min.	-	-	-	
-	94 °C	56 °C	72 °C -	
-	30 sec.	30 sec.	2 min.	30
60 °C	-	-	-	1
10 min.	-	-	-	
4 °C	-	-	-	1
forever	-	-	-	

Table 2.18: Reagents for AFLP selective amplification

Reageant	Manufacturer RJB / NUIM2012	Concentration	Volume
H_20	-	-	6.55μl
Ampli Taq Buffer	Applied Biosystems	10X	1.25μl
$MgCl_2$	Applied Biosystems	25mM	0.75μl
dNTPs	Ecogen / Promega	10mM	1μl
EcoRI NNN + Dye Primer	Stab Vida / MWG	10μM	0.1μl
MseI NNN Primer	Stab Vida / MWG	10μM	0.25μl
Ampli Taq Gold	Applied Biosystems	5 U/μl	0.1μl

Table 2.19: Thermal cycler parameters for selective amplification

Hold	Cycle			No. of Cycles
	Denature	Anneal	Extension	
95 °C	-	-	-	1
10 min.	-	-	-	
-	94 °C	65 °C	72 °C	13
		reduce by 0.7°C per cycle		
-	30 sec.	1 min.	2 min.	
-	94 °C	56 °C	72 °C	24
	30 sec.	1 min.	2 min.	
72 °C	-	-	-	1
10 min.	-	-	-	
4 °C	-	-	-	1
forever	-	-	-	

2.8 Data analysis

This section outlines how the data analysis was carried out for the DNA sequence data and AFLPs generated for *Arenaria ciliata* and *A. norvegica*.

2.8.1 Chapter 3 - Matrices for maximum likelihood analysis

Sequences returned from MacroGen Sanger Sequencing Service and MWG were aligned and manually edited using the program Geneious Pro (v. 5.3.6) (Drummond *et al.*, 2010). Alignments were generated using ClustalW (Larkin *et al.*, 2007) (within the Geneious Pro program) for (i) the four concatenated chloroplast (cp) regions (*psbA*-trnH, *rpL*32-*trnL*, *trnK*-*matK* and *matK*) (ii) *matK*, and (iii) the Internal Transcriber Spacer region (ITS) of nuclear ribosomal DNA. The matrices used for analysis are listed in Table 2.20, along with the number of characters and taxa in each matrix, the substitution models chosen (using the best fit DNA substitution model of evolution in MEGA (v. 5.04), Tamura *et al.* (2011)) for maximum likelihood analysis and the figures that they are subsequently used in.

2.8.1.1 *MatK*

A selection of sequenced individuals for *matK* which represent the full range of haplotypes for *Arenaria ciliata*, *A. norvegica* and *A. serpyllifolia* were aligned with a range of other species from Caryophyllaceae (Valente *et al.*, 2010). This dataset consists of *matK* sequences for members of Caryophyllaceae originally used in the phylogenetic analysis carried out by Fior *et al.* (2006), supplemented with extra sequences for *Dianthus*, one species of *Velezia*, and two species of *Petrorhagia*. *Silene rothmaleri* was the only species excluded from the alignment of Fior *et al.* (2006).

Two other alignments were generated for the *matK* region: one containing all of the *Arenaria ciliata* individual sequences (113 individuals) using *Arenaria serpyllifolia* as an outgroup. Network analysis using the computer program TCS

was carried out using this matrix to generate the haplotypes. A representative sequence from each haplotype was subsequently used to generate the other alignment using *Arenaria nevadensis* as an outgroup. This alignment containing individual haplotypes only was used in the maximum likelihood phylogenetic analysis.

2.8.1.2 Concatenated chloroplast sequences

Aligned matrices including the four concatanated chloroplast sequences were created for the following four groupings: (i) all the sampled *Arenaria ciliata* individuals; (ii) a subset representing only the *A. ciliata* haplotypes recorded in this study; (iii) all the sampled *A. ciliata* and *A. norvegica* individuals; and (iv) a subset representing the sampled *A. ciliata* and *A. norvegica* haplotypes. *A. serpyllifolia* was specified as an outgroup in all alignments. The alignments for each of the four chloroplast regions used in the concatenated dataset were manually concatenated in Mesquite (v.2.75) (Maddison and Maddison, 2011).

2.8.1.3 ITS

Three matrices were generated for the intraspecific analysis of the nuclear ITS region: (i) all individuals of ITS with *Arenaria serpyllifolia* as an outgroup for maximum likelihood analysis for comparison with the cp. concatenated sequences and for network analysis to generate nucleotypes for subsequent maximum likelihood analysis, (ii) ITS nucleotypes with *A. nevadensis* as an outgroup, (iii) ITS nucleotypes with *A. serpyllifolia* as an outgroup.

Table 2.20: Alignment matrices used with models of substitution chosen and non-uniformity of evolutionary rates among sites, number of taxa and characters for maximum likelihood phylogenetic analysis

Matrix	**Substitution Model**	**No. of Taxa**	**No. of bp.**	**Figure**
Cp. concat. *A. ciliata* and *A. norvegica* haplotypes	T92 + G	22	2,577	Fig. 3.6
Cp. concat. *A. ciliata* and *A. norvegica* individuals	T92 + G	34	2,577	Fig. 3.7, 3.12
matK Caryophyllaceae *spp.* from Valente *et al.* (2010)	GTR + G	103	484	Figs. 3.2
matK all *Arenaria* individuals	NA	113	689	Figs. 3.4, 3.5
matK all *Arenaria* haplotypes	T92	20	705	Fig. 3.3
ITS *A. ciliata* and *A. norvegica* individuals	JC	34	716	Figs. 3.12, 3.10
ITS *A. ciliata* complex nucleotypes outgroup *A. serpyllifolia*	K2	9	717	Fig. 3.9
ITS *A. ciliata* complex nucleotypes outgroup *A. nevadensis*	K2	9	526	Fig. 3.9

T92: Tamura 3-Parameter, JC: Jukes-Cantor, GTR: General Time Reversible, G: Gamma distribution

Arenaria nevadensis was indicated as the closest sister species to *A. ciliata* and *A. norvegica* using maximum likelihood analysis by Fior *et al.* (2006) and Valente *et al.* (2010) (section 3.4.2.1). When *Arenaria serpyllifolia* was included however the matrix contains 717 characters as this individual was collected and sequenced as part of this project. When *Arenaria nevadensis* is included as an outgroup the matrix contains 526 characters as the individual of *A. nevadensis* was downloaded from GenBank (Accession number: AY936280) and contained less characters. Maximum likelihood analysis was performed on both matrices containing each outgroup species to determine if there was any differences using more characters and different sister species.

A maximum likelihood tree was constructed using all individuals from the chloroplast concatenated sequences and separately using the same individuals for ITS. These two trees were constructed for both *A. ciliata* only and *A. ciliata* and *A. norvegica*.

2.8.2 Network analysis

Haplotypes for each DNA sequence locus were generated using TCS (v.1.21) (Clement *et al.*, 2000). TCS is a software which estimates genealogical relationships based on the stastical parsimony alogorithm described by Templeton *et al.* (1992). TCS collapses sequences into haplotypes and estimates the frequencies of these haplotypes in the overall sample group and within individual populations, which are subsequently used to calculate outgroup probabilities correlated with the age and sequential migration possibilities of each haplotype (Clement *et al.*, 2000). The maximum number of differences resulting from single substitiutions among haplotypes was calculated with 95% confidence limits, treating gaps as a 5th state. Haplotypes identified with TCS were subsequently used for maximum likelihood phylogenetic analysis.

2.8.3 Maximum Likelihood phylogenetic analysis

Maximum likelihood phylogenetic anlaysis was carried out on the 11 assembled matrices for all sampled loci in the four concatenated sequences, ITS and *matK* datasets. Maximum Likelihood fits for 24 different nucleotide substitution models were tested in MEGA (v. 5.04) (Tamura *et al.*, 2011) and are shown in Table 2.20. Models with the lowest BIC scores (Bayesian Information Criterion) were chosen to describe the optimum substitution model. Non-uniformity of evolutionary rates among sites was also modeled in MEGA and estimates of gamma shape parameter and/or the estimated fraction of invariant sites were determined (Table 2.20). Maximum likelihood trees were generated in MEGA (v. 5.04) and bootstrap replicates were resampled 1000 times for each locus. Most of the phylgoenetic trees were visualized using MEGA (v. 5.04) except for the maximum likelihood tree generated for *matK* using a range of species from Fior *et al.* (2006) and Valente *et al.* (2010) which was visualized using FigTree (available at http://tree.bio.ed.ac.uk/software).

2.8.4 Chapter 4 - Bayesian molecular clock analysis

2.8.4.1 Interspecific Bayesian analysis - generation of calibration times using outgroups

The estimation of divergence times for *Arenaria nevadensis*, *A. serpyllifolia* and *A. ciliata* were generated using *matK* sequences. Sequences for *matK* for the Caryophyllaceae family were obtained from Luis Valente - this dataset was used for the Bayesian analysis carried out in Valente *et al.* (2010) to estimate the times and rates of divergence for the *Dianthus* genus. This dataset comprises *matK* sequences for members of the Caryophyllaceae originally used in the phylogenetic analysis of the family implemented in Fior *et al.* (2006), supplemented with extra sequences for *Dianthus*, one species of *Velezia*, and two species of *Petrorhagia*.

Two datasets were constructed to include *Arenaria ciliata* for this study in addition to the sequences of *matK* from the Caryophyllaceae family: (i) one individual of *Arenaria ciliata* from Ben Bulben and (ii) a dataset including one haplotype from each of the 5 main lineages identified in the maximum likelihood

analysis in section 3.4.2. (based on *matK*19, *matK*17, *matK*1, *matK*14, *matK*9). These datasets were constructed (one haplotype versus 5 haplotypes) in order to establish the most reliable estimate for the crown node age for *Arenaria ciliata*, using the dates inferred for *Dianthus* as a positive control, to provide the date estimates which will be implemented in the intraspecific analysis of *A. ciliata*. These sequences were aligned and manually edited in Geneious Pro (v. 5.3.6) (Drummond *et al.*, 2010) and imported into BEAUti (Drummond and Rambaut, 2007) which was used to construct an input file for the Bayesian analysis conducted in BEAST (Drummond and Rambaut, 2007). All models, priors and calibrations chosen for each dataset were implemented in BEAUti.

As implemented in Valente *et al.* (2010), an age constraint of 34-45 my was assigned to the crown node of the Caryophylloideae and Alsinoideae (Figure 2.16). A minimum age estimate of 25 million years was also calibrated to decrease the minimum age of the fossil by 10 million years and test the possibility of a very young age of the fossil. A maximum age was also calibrated for the fossil increasing the maximum age estimate of the fossil by 10 million years to incorporate uncertainty regarding the dating of the fossil.

A selection of clades were grouped together to determine divergence times of the key crown nodes. Figure 2.16 shows a maximum clade credibility (MCC) tree constructed in BEAST, modified from Valente *et al.* (2010) which shows the fossil node which is calibrated at 45 my. This tree also shows the key crown nodes to be dated which are represented as coloured arrows. The crown nodes to be dated in the current analysis and the purpose for each grouping are:

- To date the ancestral node for *Arenaria ciliata* and its closest sister species *A. nevadensis* (*A. nevadensis* is identified as a sister species to *A. ciliata* from the maximum likelihood analysis - Figure 3.2). The date generated here would be subsequently used as a calibration point in the intraspecific *A. ciliata* analysis when estimating the times of divergence between the lineages.

- To date the divergence time of *A. ciliata*/*A. nevadensis* from *A. serpyllifolia*. This will also be used as a calibration point in the intraspecific analysis.

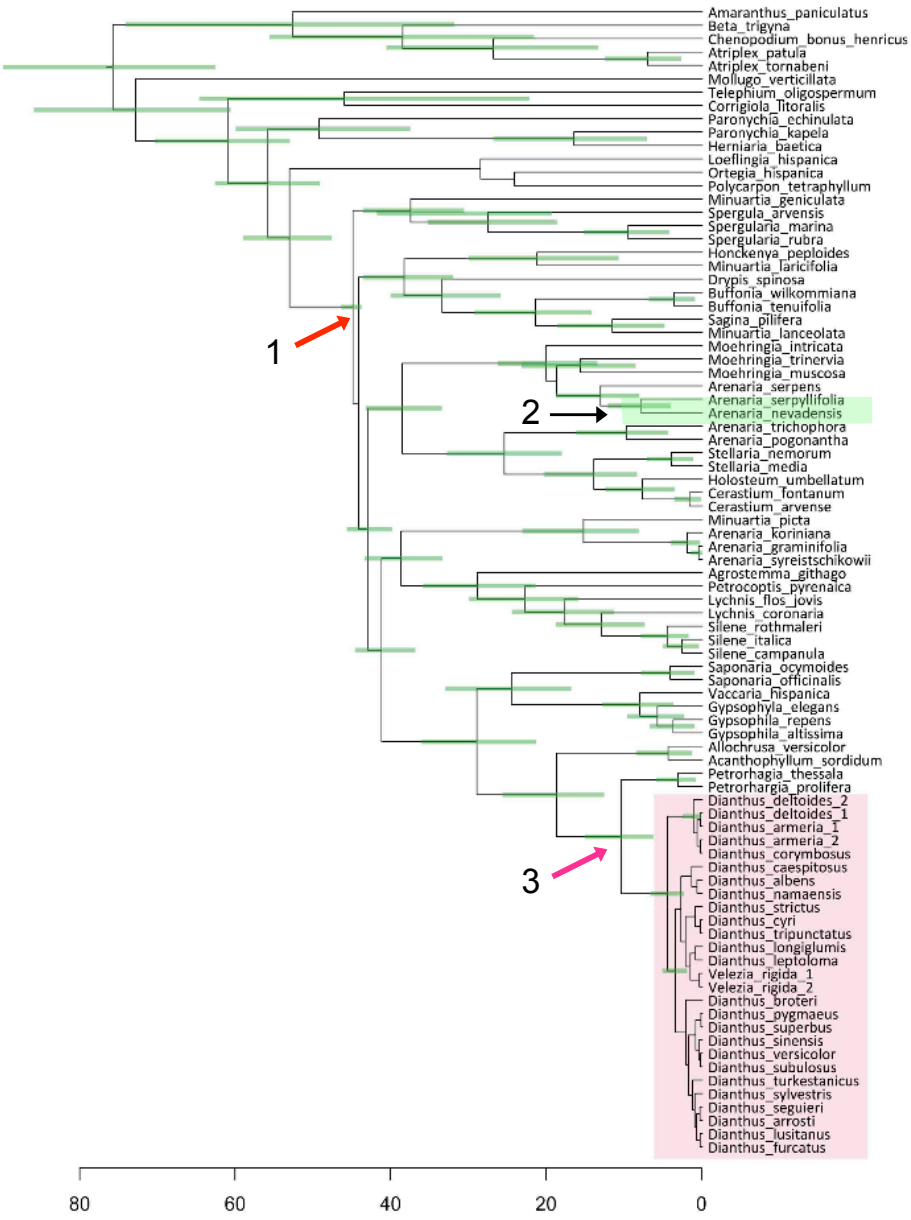


Figure 2.16: Maximum clade credibility (MCC) tree generated in Beast modified from Valente *et al.* (2010) using *matK* sequences. The nodes for calibration in this analysis are indicated by the numbered arrows: 1 - The red arrow shows the crown node of the Caryophylloideae and Alsinoideae to which the fossil was calibrated. 2 -The pink arrow indicates the node for *Dianthus* and 3 - the black arrow shows the node calibrated for *A. ciliata* and *A. nevadensis* and the highlighted green areas shows the terminal branch where the Irish individual of *A. ciliata* will be included. The time scale shows age in million years from present.

- To date the crown node for *Dianthus* - This dating estimate will act as a positive control to compare with the estimates obtained for the crown node for *Dianthus* by Valente *et al.* (2010)

- To date the crown node for *Arenaria ciliata* - This is the calibration point to be included in the dataset comprising the 5 additional haplotypes from *Arenaria ciliata* to estimate the time of divergence of these haplotypes. This date may then be used as an alternate basal node date in calibrating the intraspecific analysis.

- Basal calibration point - This is the crown node of the Caryophylloideae and Alsinoideae to which the Caryophyllaceae fossil inflorescence from the Middle-Late Eocene was calibrated.

The optimal model of evolution was inferred using jModelTest (v.12.9.0) (Posada, 2008) and a general time-reversible (GTR) model with gamma distributed rate variation was suggested as the most appropriate model for this dataset. The prior for the distribution node heights was given as a Birth-Death speciation process as this is model is appropriate for interspecific datasets which contain sequences from different species. The uncorrelated lognormal relaxed clock was chosen with uniform distribution. The starting tree was randomly generated under the coalescent process. The calibrated node prior was set to lognormal with mean = 0.5, stdev = 0.5 and offset at (the number of years in millions). These priors were chosen to replicate the molecular clock analysis carried out in Valente *et al.* (2010). The chain length was set at 20,000,000 generations, sampling every 1,000 trees with the first 10% of trees discarded as burn-in. The first Trees were summarized in a maximum clade credibility (MCC) tree obtained in TreeAnotator (v.1.6.1) (Drummond and Rambaut, 2007) and viewed with FigTree (v.1.3.1 available at http://tree.bio.ed.ac.uk/software/figtree/). The statistical estimates for each crown node from the Bayesian analysis run with BEAST were visualised in TRACER (http://tree.- bio.ed.ac.uk/software/tracer/). The mean age and effective sample size (ESS) were recorded for each key node.

2.8.5 Intraspecific Bayesian analysis - estimation of divergence times for lineages within the *Arenaria ciliata* complex

2.8.5.1 Datasets for the intraspecific Bayesian analysis

The mean estimated ages determined from section 4.2.1. using the interspecific Bayesian anlaysis for the *matK* haplotypes (Table 4.1) were subsequently used to calibrate an intraspecific Bayesian analysis of individual sample sequences from the *Arenaria ciliata* complex for the *matK*, concatenated chloroplast and ITS regions. The dates that were used as calibrations in the intraspecific analyses for all regions are: 3.328, 4.582, 5.874 and 7.217 my for *A. nevadensis* ; 5.425, 7.47, 9.621 and 11.808 my for *A. serpyllifolia*; and 2.725, 3.655, 4.675 and 5.81 my for the *A. ciliata* haplotypes only. This technique (of using the ages estimated from the interspecific analysis to calibrate an intraspecific tree) was implemented for all chloroplast and nuclear ITS sequences.

Most of the estimated dates for the key nodes from section 4.2.1. were chosen from dataset (i), which included one individual of *Arenaria ciliata*. This was chosen over the 5 haplotype date estimates, as the dates estimated for *Dianthus* were (marginally) closer to the dates estimated in Valente *et al.* (2010). The only exception to this was an estimate derived from dataset (ii) with five haplotypes of *A. ciliata* for the estimation of dates for the *A. ciliata* only group. The dates estimated using the dataset including five sequences of *Arenaria ciliata* may be more inaccurate than using a single sequence of *A. ciliata*, as the optimum model of evolution in terms of substitution rate may alter if there are multiple sequences included from one species and therefore the rates of substitution may alter among terminal clades within the dataset. However, the dates estimated for the expansion of the *Arenaria ciliata* complex (from dataset (ii)) were subsequently used in the intraspecific analysis to construct datasets which include *A. ciliata* sequences only. Therefore an age for the radiation of the *A. ciliata matK* haplotypes can be tested without an outgroup. Eight datasets were constructed for the intraspecific analysis consisting of alignments of sequences in the *Arenaria ciliata* complex for the chloroplast and nuclear ITS sequences which aim to test the inclusion of

different outgroup species available in the Bayesian dating analysis. The eight datasets constructed for the intraspecific analysis are shown in Table 2.21.

Table 2.21: Mean estimate for the time of divergence for the crown nodes obtained from BEAST and viewed in TRACER. Age is estimated in million years (my). Age estimates for *Dianthus* are also included from Valente *et al.* (2010)

Dataset	Sequence region	Outgroups included
(i) *MatK* AN	*MatK*	*A. nevadensis*
(ii) *MatK* AS	*MatK*	*A. serpyllifolia*
(iii) *MatK* AC	*MatK*	*A. ciliata* only
(iv) * conat AS	concatenated cp.	*A. serpyllifolia*
(v) concat AC	concatenated cp.	*A. ciliata* only
(vi) ITS AN	ITS	*A. nevadensis*
(vii) ITS AS	ITS	*A. serpyllifolia*
(viii) ITS AC	ITS	*A. ciliata* only

*As concatanated chloroplast sequence data was not available for *A. nevadensis*, it was not possible to include it as an outgroup in this analysis.

When using fossil data to calibrate nodes in molecular clocks, the geological age and taxonomic relationship between the exant taxa will contain an element of uncertainty and these need to be taken into account (Graur and Martin, 2004). Therefore the dates generated from the 4 basal calibration points (estimated from the 4 ages calibrated for the fossil in the interspecific analysis) were used to generate date estimates in each dataset.

The datasets (alignments) were constructed as outlined in section 4.2.1. The sequence of *Arenaria nevadensis* for the ITS region was downloaded from GenBank (Accession number: AY936280). The chloroplast haplotypes were identified for the *Arenaria ciliata* complex in section 3.4 and multiple sample sequences of each haplotype were not included in the analysis. However, a preliminary Bayesian analysis run for the ITS region in BEAST showed that when the nucleotypes only alignment was included (which contains 8 representative sequences), the MCMC (Markov chain Monte Carlo) chain did not converge so this alignment

was unsuitable for analysis. An alignment containing all sequences sampled for ITS was therefore included in the Bayesian analysis.

2.8.5.2 Intraspecific Bayesian analysis

A preliminary Bayesian analysis carried out in BEAST (using an uncorrelated clock with a calibration date estimated from the fossil at 45 my) was conducted to establish the topology of the maximum clade credibility (MCC) tree for each dataset and consequently determine what nodes to estimate an age for. The posterior values of the nodes were viewed in FigTree and any nodes which were over 0.5 were estimated and assigned a letter to group the key nodes to estimate a time of divergence. Each of the datasets for the different gene regions were then run in BEAST (as outlined in section 4.2.1) using the appropriate age estimate given, which was determined in section 4.2.1 using the fossil calibrations for the interspecific analysis (Table 4.1).

The best fit substitution model HKY (Hasegawa, Kishino & Yano, 1995) was determined using jModelTest (v.12.9.0) (Posada, 2008). The prior established for the distribution of nodes heights was set to coalescent constant size. The coalescent constant size was chosen as this is appropriate for datasets which contain sequences from the same species and are therefore less divergent from each other. The uncorrelated lognormal relaxed clock was chosen with uniform distribution. Methods of molecular clock analysis which assume a clock-like model of evolution may not be appropriate for modelling the process of molecular evolution as variation in the rates of substitution within and among lineages may be prevalent (Britten, 1986; Ho and Larson, 2006; Rutschmann, 2006). Models which allow the rate of template evolution to vary on an evolutionary tree, and so 'relax' the molecular clock, may therefore be more appropriate (Drummond *et al.*, 2006) and reflect true evolutionary scenarios more accurately. Therefore the uncorrelated lognormal relaxed clock was used in both the inter and intraspecific molecular clock analyses for this study.

The starting tree was randomly generated under the coalescent process. The calibrated node prior was set to exponential with mean = 0.5, offset = the number of years in millions estimated from section 4.2.1 (Table 4.1), stdev = 1.

This prior was chosen as both the exponential and lognormal distributions are appropriate for fossil calibrations (Ho, 2007), however, it was not possible to implement the lognormal prior for the intraspecific analysis. The chain length was set at 10,000,000 generations, sampling every 1,000 trees with the first 10% of trees discarded as burn-in. The summary statistics including mean age, the standard error, the effective sample size (ESS) and the 95% highest posterior density (HPD) upper and lower intervals were viewed in TRACER for each key node.

A MCC tree was constructed using the output from BEAST for the age estimates derived from the fossil calibrated to 34my. This calibration implements a conservative approach as it estimates younger dates than the other calibrations (excluding 25my which is added for comparison but which may not be a realistic estimate for the fossil calibration). The MCC tree was obtained in TreeAnotator (v.1.6.1) (Drummond and Rambaut, 2007) and viewed with FigTree (v.1.3.1 available at http://tree.bio.ed.ac.uk/software/figtree/) for each of the datasets including the different outgroups for each of the gene regions.

2.8.6 Chapter 5 - Amplified Fragment Length Polymorphisms (AFLPs)

A list of the binary matrices created for each batch of work (NUIM2009, RJB and NUIM2012) and for each individual selective primer combination, are detailed in Table 5.1. Some of the datasets within each batch were concatenated to include 2 and/or 3 selective primers. The RJB Madrid matrices were divided into two datasets; one which contained all of the individuals analyzed and one which contained a reduced number of Irish individuals so that the number of individuals analyzed was equivalent across all the sampled populations (dataset RJB less Irish indiv.).

Principal coordinate analyses (PCoA) and Analysis of molecular variance (AMOVA) were carried out using GenALeX (v.6.5) (Peakall and Smouse, 2012). The binary AFLP matrix was analysed directly in GenALeX for the number of fragments per population (Nf), the number of fragments with frequency >5% (Nf freq > 5%), the number of private fragments (Pf) and the number of private

fragments per individual (Pf / N) occurring in each population. PCoA were used to compare pairwise similarity among the AFLP genotypes on three axes.

Clustering methods based on genetic distances were used to create trees in order to visualise the relationships between the AFLP genotypes among populations. Neighbour-joining trees (NJ) were performed in PAUP (v. 4.0b10) (DL, 2002) based on Nei & Li's distance measure. The robustness of the AFLP data was assessed for the NJ trees by calculating bootstrap support values for 1000 replicates using a NJ analysis in PAUP. The trees were rooted using individuals from populations in the East and West Alps. Format conversions of each raw-data matrix for input into PAUP were carried out with the AFLPdat R-script (Ehrich, 2006).

Differentiation of genetic variation within and among populations and groups was quantified using analysis of molecular variance (AMOVA) based on pairwise distances for the majority of the datasets. The ratio between the *within* to *among* variation in the populations can provide insight into how much of the genetic variation that occurs in a population is accounted for by geneflow events between populations. For example, outbreeding species with less genetic differentiation between populations record high within variation and low PhiPT value, wheras restricted geneflow between populations would record high among variation and high PhiPT values which are characteristic for annual, selfing species (Hamrick and Godt, 1996).

The populations in each datasets were grouped in three ways: (i) *Arenaria ciliata* and *A. norvegica* according to geographic region (reg), (ii) *Arenaria ciliata* individuals according to geographic region (AC) and (iii) *A. norvegica* individuals according to geographic region (AN). Some of the datasets were also grouped according to the main clusters identified in the PCoA and the neighbour-net and NJ analysis (clu).

3

Phylogenetic and Phylogeographic Analysis of *Arenaria ciliata* and *A. norvegica* using DNA Sequence Data

3.1 Introduction

3.1.1 Taxonomy of the carnation family and the genus *Arenaria*

The carnation family (Caryophyllaceae) is a large family of flowering plants which consist of 86 genera and 2,200 species (Bittrich, 1993). These species are distributed predominately in the holarctic but concentrated in the Mediterranean and Irano-Turanean region (Fior *et al.*, 2006).

Taxa and genera within this family have been difficult to classify due to com-

plex and possible homoplasious morphological characters. Attempts have been made using morphological and molecular characters to resolve the taxonomy of this family (Fior *et al.*, 2006; Smissen *et al.*, 2002). Phylogenetic analyses have also highlighted the difficulties in discriminating between species of some genera such as *Moehringia* (Fior and Ola, 2007) and *Arenaria* (Valcarcel *et al.*, 2006).

The *Arenaria* genus contains approximately 160 species, which occur mainly in arctic and temperate regions in the Northern hemisphere (Wyse Jackson and Parnell, 1987). The difficulties in the classification of this genus have been noted by McNeill (1962) based on morphological characters and reinforced by Fior *et al.* (2006) based on a combination of morphological, nuclear and chloroplast markers. One of the major problems with the taxonomy of this genus is that it does not form a distinct monophyletic lineage. Figure 3.1 highlights the polyphyletic nature of this genus, where different species of *Arenaria* fall into relatively divergent lineages (Fior *et al.*, 2006). The red boxes highlight where different species of *Arenaria* occur in the phylogenetic tree.

Discordance between plastid and nuclear DNA phylogenetic trees and even between two plastid regions have been demonstrated for the *Arenaria* section *Plinthine* (Valcarcel *et al.*, 2006). Factors such as rampant aneuploidy, polyploidy, hybridization, lineage sorting, and concerted evolution of ITS sequences have been used to account for the discordance between the different loci (Valcarcel *et al.*, 2006).

The present study investigates the phylogeography of two species which belong to the *Arenaria ciliata* complex: *A. ciliata* and *A. norvegica*. In section 1.4.1, a more detailed account of the taxonomy of this complex is presented and the difficulties in discriminating species within this complex is discussed. As the current study includes two species within this taxonomically difficult complex and genus, it is important to investigate species relationships between these taxa and ensure that sampled populations in this study contain the correct taxa. This is also necessary to decide whether these taxa need to be analysed together, or treated separately for the phylogeographical analyses. The discordance between different regions of the genome reported for the *Arenaria* section *Plinthine* (Valcarcel *et al.*, 2006), also conveys the need for the use of sequences from multiple regions from the chloroplast and nuclear genome.

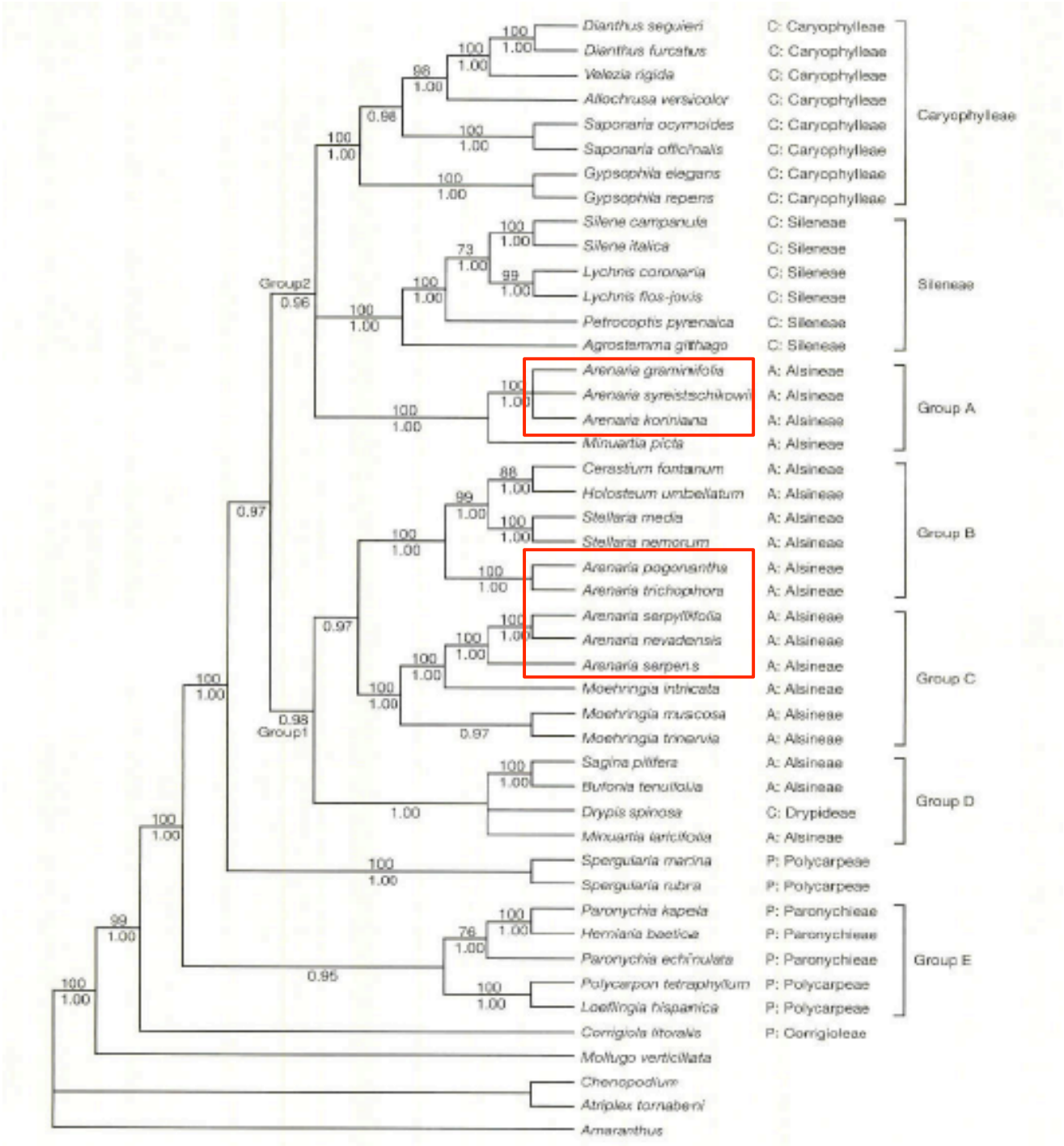


Figure 3.1: Parsimony strict-consensus tree of Caryophyllaceae based on simultaneous analysis of ITS and *matK* data modified from Fior *et al.* (2006). The red boxes indicate species of *Arenaria*.

3.1.2 Sequence data analysis in phylogeography

The advances in the field of phylogeography since it first emerged in 1987 (Avise, 1998) have led to a proliferation of studies investigating the effects of Pleistocene glaciations on the genetics of extant populations. Many of these studies have successfully analyzed sequence data to investigate processes of the Pleistocene glaciations governing putative Pleistocene refugia, re-colonized areas and migration routes (Pleines *et al.*, 2009). Sequence data is a powerful tool which allows the output of easily scored, high-quality markers, while automated sequencing and high-powered computer facilities can generate and process these data (Lowe *et al.*, 2007).

The use of sequence data to examine variation between, and within, populations of individuals is very informative on the basis of mutational differences among variants which indicate relationships among these individuals and provide the information required to infer genealogical relationships (Schaal and Olsen, 2000). In order to analyse polymorphic variability between samples, prior knowledge of sequence variability and potential primer sites are needed, and this usually requires the screening of multiple potential loci (Borsch and Quandt, 2009). The appropriate region of the genome to be screened for these primers can depend on the nature of the organism being studied and the hypotheses that are being investigated i.e. the phylogeographic signal in nuclear versus organelle genes.

A key goal in phylogeography is the determination of gene genealogies and the reconstruction of ancient histories. Plastid markers are ideal for this as they are slowly evolving (Schaal *et al.*, 1998), non-recombining and are inherited maternally (in most angiosperms) and so contain a very low mutation rate and as a result retain ancient historical structures (Avise, 2009; Comes and Kadereit, 1998). As chloroplasts are transmitted through seeds they are also very informative in phylogeographical analyses and can reveal patterns of seed migration within species. Despite the low variation observed in chloroplast DNA loci (Schaal *et al.*, 1998), phylogeographic studies use non-coding variable regions from the chloroplast genome successfully, for example, the *matK* region (Olsen *et al.*, 2004), *psbA*-trnH (Marr *et al.*, 2008; Ronikier *et al.*, 2008), *rpl32-trnL* (Westergaard *et al.*, 2011; Fernandez-Mazuecos and Vargas, 2011), and *trnK-matK*

(Fernandez-Mazuecos and Vargas, 2011; Rodriguez-Sanchez *et al.*, 2009).

Criticism has been placed on studies overly reliant on a single gene as a marker of evolutionary descent (Bermingham and Moritz, 1998; Doyle, 1992). The necessity of using sequences from different regions of the genome and combining chloroplast and nuclear data has also been highlighted (Eidesen *et al.*, 2007). Many studies incorporate multiple genes and in many plant studies the inclusion of chloroplast non-coding sequences and nuclear sequences is very common (Dixon *et al.*, 2009; Eidesen *et al.*, 2007; Provan *et al.*, 2005; Valcarcel *et al.*, 2006). Chloroplast DNA differ from nuclear DNA in that it is (in most angiosperms) inherited maternally and shows reduced recombination where genomes from different lineages apparently never recombine (Birky, 2001). The nuclear genome however, is not uniparentally inherited and is recombined in every generation (Birky, 2001).

DNA markers for non-coding regions which show higher levels of variation such as nuclear DNA have the potential to reveal a more recent history than chloroplast DNA which are more slowly evolving. Sequencing of the nuclear ribosomal internal transcribed spacer (ITS) region has been carried out in many phylogeographical and phylogenetic studies successfully (Baldwin *et al.*, 1995; Eidesen *et al.*, 2007; Fior *et al.*, 2006; Guo *et al.*, 2004; Valcarcel *et al.*, 2006). The nuclear ribosomal ITS region is repeated frequently in the plant nuclear genome and other features such as its small size and highly conserved flanking regions, all contribute to the efficient amplification, sequencing and subsequent analysis of ITS data (Baldwin *et al.*, 1995). By comparing both the nuclear and chloroplast regions, a variety of biological processes can be revealed between plant species like the potential of hybridisation and the relative role of pollen versus seed migration as sources of gene flow (Ennos, 1994).

3.1.3 Glacial history of Irish flora and fauna

Different plants and animals in Ireland have responded to the glacial cycles in different ways, where each species contains a unique history. A variety of scenarios for the biogeographical history of Irish plants and animals have been presented for a range of species for example, recent post glacial colonization and survival

in a refugium or refugia. Evidence for recent postglacial colonization to Ireland has been indicated in a range of phylogeographical studies including St. Dabeoc's heath (*Daboecia cantabrica*) (Beatty and Provan, 2012), the mountain hare (Hamill *et al.*, 2006), stoats (Martinkova *et al.*, 2007) and oak trees (Kelleher *et al.*, 2004). Some of these studies also indicate the divergence of the Irish populations from the British populations which may indicate colonization route via Britain (Hamill *et al.*, 2006; Martinkova *et al.*, 2007; Mascheretti *et al.*, 2003). These studies indicate no unique haplotypes and reduced genetic diversity in the Irish populations which is suggestive of a more recent post glacial immigration and rapid range expansion (Schaal *et al.*, 1998).

In contrast however, there are studies which suggest that Ireland may have contained refugia that enabled the movement of haplotypes northwards. Hoarau *et al.* (2007) claimed that for the brown seaweed *Fucus serratus*, the Irish refugium was the source that enabled the movement of a single haplotype to Scandinavia via northern Scotland.

Sampling of the mtDNA for the land snail, *Cepaea nemoralis*, showed a unique mitotype from Mullaghmore (Davison, 2000). The authors suggest that Ireland was colonized separately from Britain, however, it is impossible to know based on the samples from Britain and Ireland alone, whether this could be indicative of a refugium, or colonization via an alternative route from Spain. mtDNA reveals that Irish individuals for the common frog (*Rana temporaria*) group within two main lineages, one which is unique to Ireland and one which is grouped in a clade which contains haplotypes widely distributed in Western Europe (Teacher *et al.*, 2009). The authors attribute this anomaly to survival in a refugium in Ireland or dual colonization where some frogs survived in a refugium in Ireland and others which colonized through a different route.

3.1.4 Aims and Objectives:

In this chapter a range of sequence data from the chloroplast and nuclear genomes were analysed in a phylogenetic and phylogeographic context using maximum likelihood and haplotype networks using statistical parsimony methods. The

results of these analyses are presented in four components: (i) the *matK* chloroplast region (section 3.4.1), (ii) 4 sequence regions from the chloroplast genome which were concatenated (*psbA*-trnH, *rpL32-trnL*, *trnK-matK* and *matK*) (section 3.4.2), (iii) the nuclear ribosomal internal transcribed spacer (ITS) (section 3.4.3) and (iv) a comparative maximum likelihood analysis for the ITS and concatenated chloroplast region (section 3.4.4).

The *matK* region was chosen for separate analysis because a fossil-calibrated dataset containing *matK* sequences from Caryophyllaceae was available from Fior *et al.* (2006) and Valente *et al.* (2010). Both *Arenaria ciliata* and *A. norvegica* are included in this dataset to conduct a phylogenetic analysis.

For each of the four components of data outlined, a maximum likelihood and a haplotype network using statistical parsimony was carried out to achieve the following objectives:

- Infer the phylogenetic relationship between *Arenaria ciliata* and *A. norvegica* to determine if the classification as distinct species, based on morphological markers, is supported by the molecular markers investigated in this study
- Infer the phylogeographic relationships of *Arenaria ciliata* and *A. norvegica* from the distribution of chloroplast haplotypes and ITS nucleotypes in Europe
- Determine the implied most likely origin and nearest sister populations for the Irish samples

These objectives were implemented to elucidate phylogeographic relationships between *Arenaria ciliata* and *A. norvegica*, in an effort to determine putative postglacial migration routes to Ireland for these species.

3.2 Results

3.2.1 *MatK* chloroplast region

3.2.1.1 Maximum Likelihood interspecifc phylogenetic analysis of the Caryophyllaceae family using the *matK* locus

In order to generate an accurate measure of phylogeographic structure in these plant populations, both inter-and intra-specific genetic data are needed. A maximum likelihood phylogenetic tree was constructed including a range of species from the Caryophyllaecae family in order to understand the relationship between *Arenaria ciliata* and *A. norvegica* in a broader phylogenetic context using the chloroplast region *matK*. The maximum likelihood phylogenetic tree was generated for 103 taxa using 484 characters. Eighty five species were included from Fior *et al.* (2006) and Valente *et al.* (2010) and 18 sequences of *Arenaria ciliata*, *A. norvegica* and *A. ciliata* subsp. *pseudofrigida* which were sequenced in this study.

The resulting ML phylogenetic tree is shown in Figure 3.2. The phylogenetic analysis revealed that for the most part, DNA extracted from individual samples from these taxa are monophyletic. The only individuals that did not occur in this group include the *Arenaria* individuals which were sampled by other collectors: AC 5 from Romania and AC12 from Switzerland (highlighted in red) which are suggested to be most closely related to *Arenaria serpens*. These individuals were omitted from subsequent analysis as these were deemed not to be part of the *Arenaria ciliata* complex.

The closest sister species to the *Arenaria ciliata* and *A. norvegica* group is *A. nevadensis* (highlighted in blue) and the second closest species is *A. serpyllifolia* (highlighted in yellow). The individuals of *Arenaria serpyllifolia* that were sampled in this study group with the individual of *A. serpyllifolia* from Fior *et al.* (2006) so these are correctly placed and are therefore the correct species and may be used for subsequent analysis.

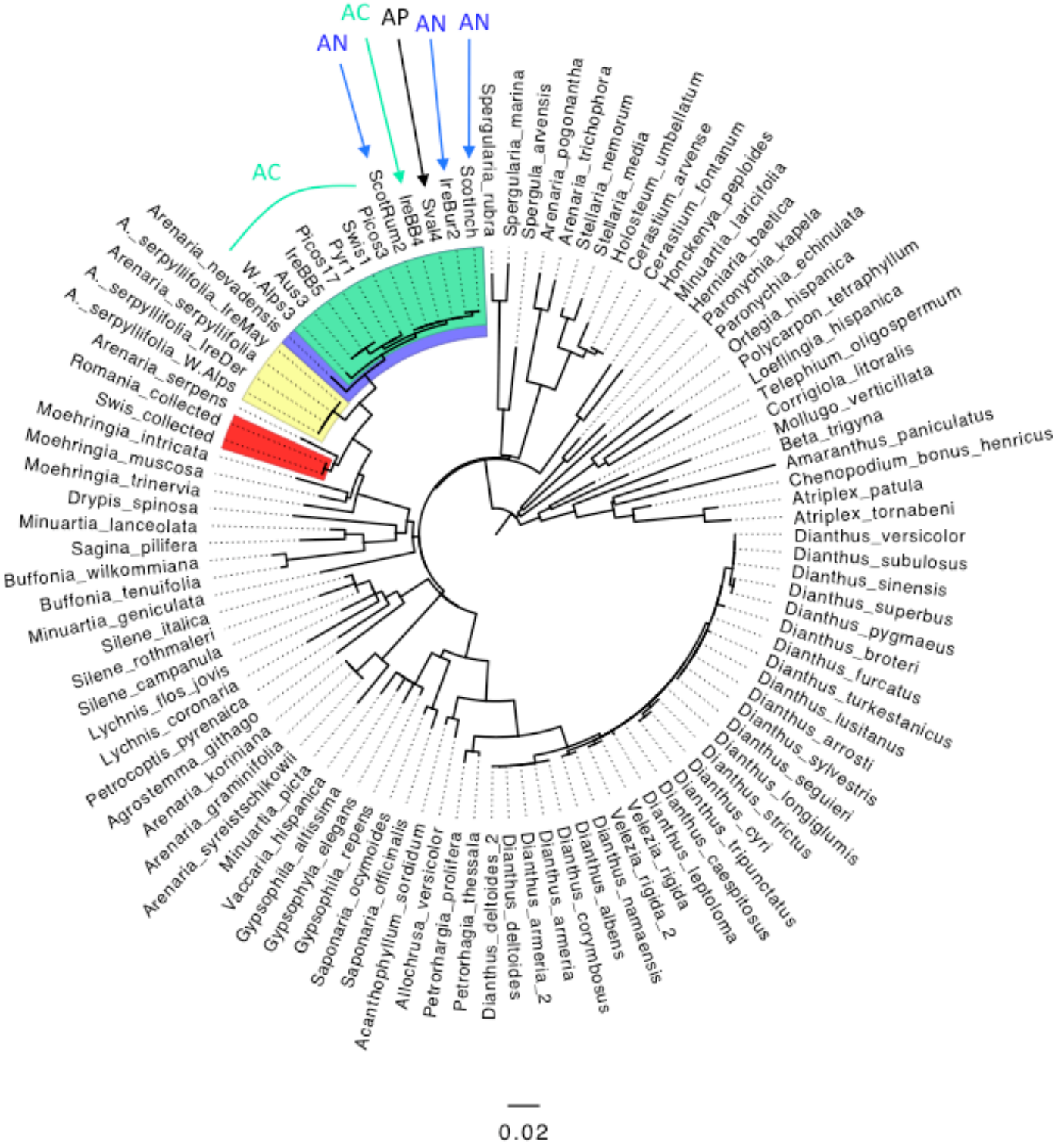


Figure 3.2: Maximum likelihood phylogenetic tree based on *matK* analysis for species in the Caryophyllaceae family from Fior *et al.* (2006); Valente *et al.* (2010) including individuals of *Arenaria ciliata* (AC - indicated by the green line and arrow), *A. ciliata* subsp. *pseudofrigida* (AP - black arrow) and *A. norvegica* (AN - Blue arrows). The highlighted colours represent different groupings: red - *Arenaria* spp. collected from Romania and Switzerland, yellow - *A. serpyllifolia*, blue - *A. nevadensis* and green - *A. ciliata* complex. Visualized using FigTree (available at http://tree.bio.ed.ac.uk/software).

Constructing this phylogenetic tree demonstrates that for *matK* the sampled individuals of *Arenaria ciliata*, *A. norvegica* and *A. ciliata* subsp. *pseudofrigida* all occur in the same lineage (highlighted in green). There is a complex taxonomic association between these taxa, therefore, it is important to include these taxa together where possible when conducting phylogeographic analysis. Because of this, these taxa will be collectively referred to as the *Arenaria ciliata* complex.

3.2.1.2 Maximum Likelihood Intraspecifc phylogenetic analysis for the *Arenaria ciliata* complex haplotypes

Nineteen haplotypes were identified for the *Arenaria ciliata* complex using *matK* sequence data (Table 3.1). The maximum likelihood tree generated with sequences represented from these haplotypes is shown in Figure 3.3 with bootstraps indicated on the branches. As *Arenaria nevadensis* is identified as the sister species to the *A. ciliata* complex (Figure 3.2), this was used as the outgroup in this maximum likelihood phylogenetic analysis.

There are three major lineages identified which are shown in Figure 3.3. Group A haplotypes occur over a broad European distribution which includes the Eastern Alps, Pyrenees, Switzerland, Svalbard, Picos De Europa and Ireland. Group B haplotypes occur in Iceland, England, Scotland, Ireland, Pyrenees, Picos De Europa, East Alps, West Alps, Scotland and Svalbard.

Group C haplotypes occur only in the Eastern and Western Alps. Haplotype 19 is basal to all three groups and is located in the Eastern and Western Alps only. There is uncertainty as to the placement of group C, which contains the haplotypes from the East and West Alps and which is placed with group B with a low bootstrap value of 53.

Table 3.1: Haplotypes generated for the *Arenaria ciliata* complex using *matK* with individuals identified for each haplotype and location

Haplotype	Individuals	Location	Lineage
1	AC1.25, AC1.11, AC1.5, AC1.7, AC1.8, AC16.10, AC16.13, AC16.5, AC2.1 AC3.1, AC3.14, AC3.4, AC4.1, AC4.13, AC4.15, AC4.3, AC4.7, AC6.13 AN4.1, AN4.10, AN4.14, AN4.15, AN4.8, AP2.3, AC4.10	IreBB; E.Alps; W. Alps; Rum, Scot; Sval	B
2	AP2.5, AP2.2, AP2.4, AP2.1	Svalbard	B
3	AN8.2, AN3.1	Iceland; Yorkshire	B
4	AN1.14, AC10.15, AN7.2, AN6.28, AN5.4, AN4.4, AN4.19 AN2.5, AN1.8, AN1.5, AN1.3, AN1.2, AN1.19, AC10.10	IreBur; Pyr; Ice; Scot; Eng	B
5	AC14.13	Picos	B
6	AC14.10, AC15.5, AC15.3, AC15.15, AC15.13, AC15.10 AC15.1, AC14.5, AC14.1, AC11.7, AC11.15, AC10.5	Picos ; Pyrenees	B
7	AC19.30	Swiss Alps	A
8	AP1.1R	Svalbard	A
9	AC3.23, AC3.17, AC2.7, AC2.5, AC2.4, AC2.25, AC2.12, AC17.1, AC14.7 AC14.15, AC13.5, AC13.13, AC11.3, AC11.10, AC11.1, AC1.2, AC1.10	IreBB; E. Alps; Picos ; Pyrenees	A
10	AC11.13	Picos	A
11	AC16.3	E. Alps	A
12	AC19.6, AC19.24, AC19.18, AC18.4, AC18.1	E. Alps	A
13	AC10.13	Pyr	A
14	AC16.15, AC16.1, AC13.7, AC13.10, AC13.1, AC10.6, AC10.4, AC10.1	E. Alps; Pyr.	A
15	AC17.3, AC17.10	E. Alps	C
16	AC8.3	W. Alps	C
17	AC9.5, AC9.4, AC9.3, AC8.4, AC8.2, AC8.1, AC7.1, AC6.5, AC6.18, AC6.16	W. Alps	C
18	AC6.3	W. Alps	C
19	AC9.2, AC9.1, AC17.8, AC17.5, AC17.2	W. Alps, E. Alps	Outgroup
20	AS1.17	Kildare, Ireland	Outgroup

The details for the individuals are in Tables 2.1 2.3

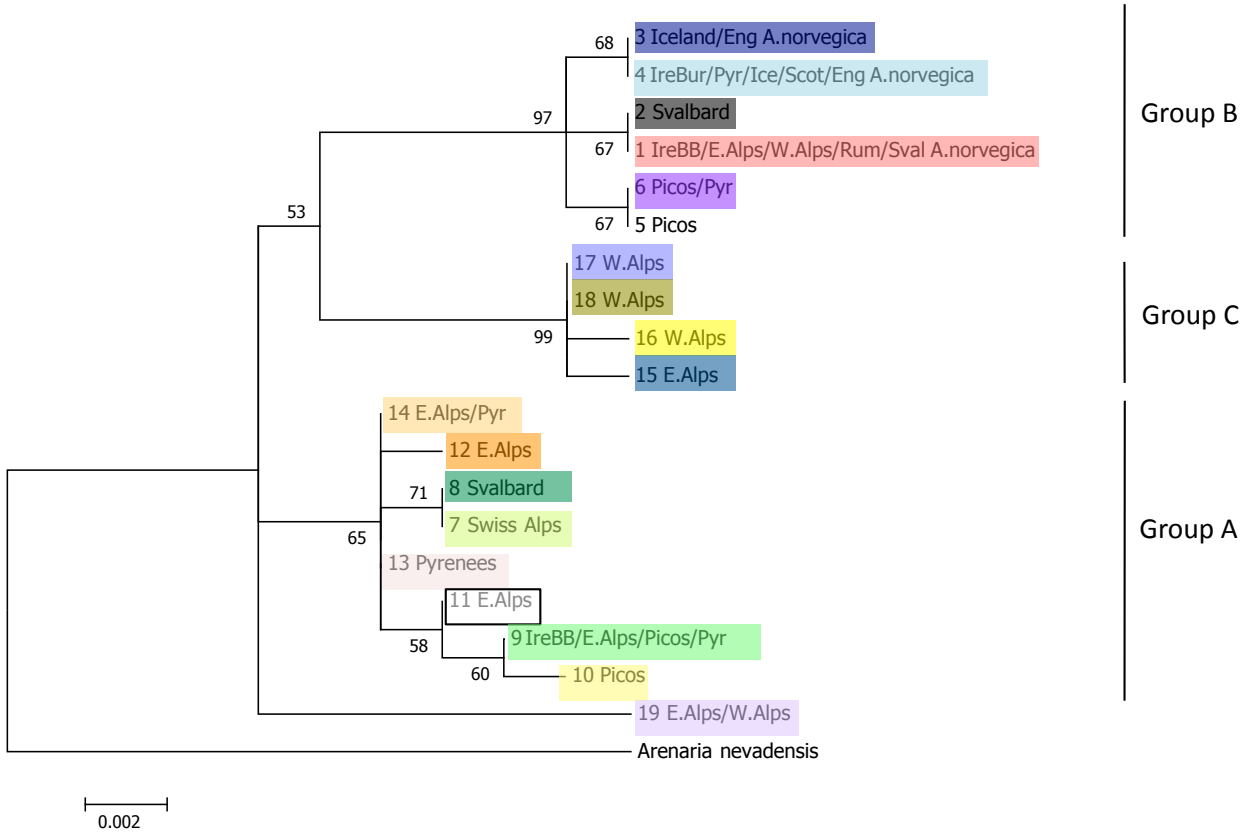


Figure 3.3: Maximum likelihood phylogenetic tree based on *matK* haplotypes for the *Arenaria ciliata* complex with bootstrap support values on the branches. Each label contains haplotype numbers and is colour coded to correspond to the haplotypes in Figure 3.4.

It is not clear whether this lineage is a basal group to the other groups or whether it is a sister group to group B. However, it is clear that one of the haplotypes from both the concatenated region and *matK* contains an ancestral haplotype from the Alps (West Alps in concatenated dataset and East and West Alps in the *matK* dataset) which is basal to all three groups.

Three haplotypes occur in Ireland. These haplotypes are grouped in two major lineages with haplotype 4 and 1 placed in group B and haplotype 9 placed in group A. Group B includes the Burren types of *Arenaria norvegica*. There are two haplotypes which include *Arenaria norvegica* - haplotype 4 from Ireland (The Burren), Iceland, England and also *A. ciliata* from the Pyrenees. The other *Arenaria norvegica* haplotype is assigned to a more predominately *A. ciliata* group and includes Rum (Scotland) and *A. ciliata* from Ireland (Ben Bulben), East

Alps, West Alps and Svalbard. The population sampled from Rum in Scotland is represented in both of these haplotypes, the *Arenaria norvegica* dominated haplotype (4) and the *A. ciliata* dominated haplotype (1).

The East Alps is the only population that contains haplotypes in the three major groupings and this is only seen with the *matK* dataset as more individuals were included than the concatenated dataset. Populations from Ireland, The Pyrenees, Picos De Europa and Svalbard also contain haplotypes which occur in groups A and B. The haplotypes from Switzerland are assigned only to group A and its sister group is identified as Svalbard.

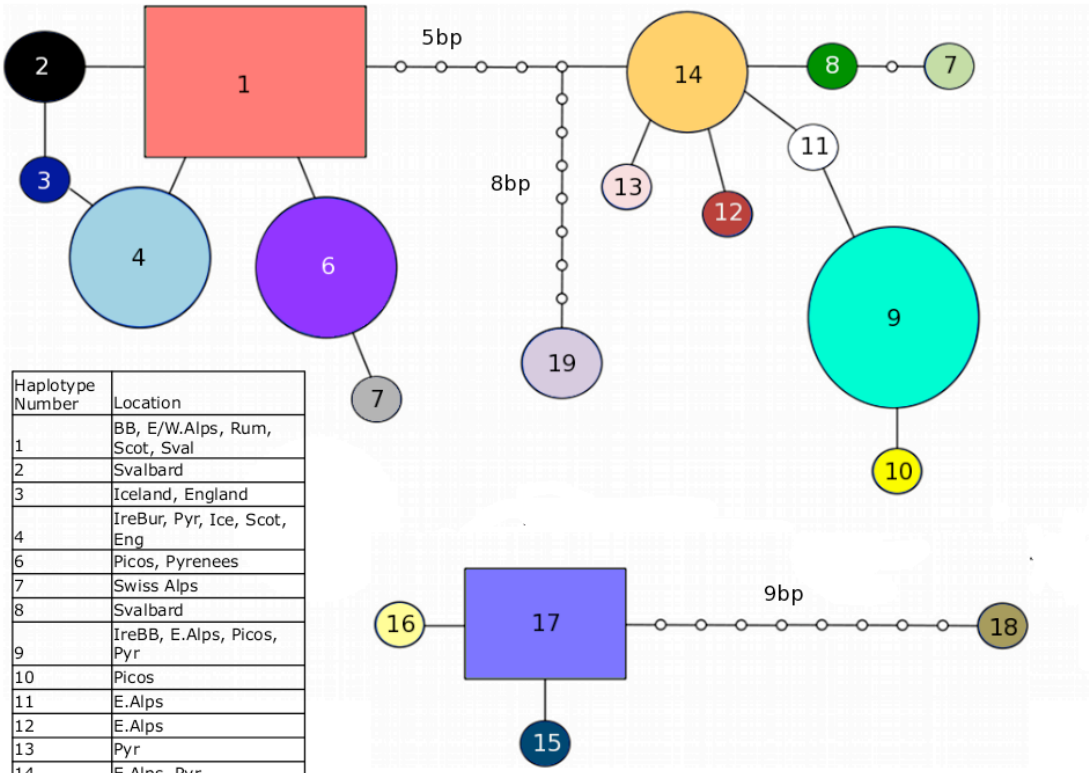
3.2.1.3 Haplotype network for the *Arenaria ciliata* complex for *matK*

A haplotype network for all species including *Arenaria serpyllifolia* as an outgroup with an alignment of 113 samples and 689 characters from the *matK* chloroplast region. The haplotypes generated are in Table 3.1 and the identical sequences which differ by missing or ambiguous characters are shown. The frequency that the haplotypes occur is shown in Appendix A. Nineteen haplotypes were identified which were contained in two main networks; one of which contained the Alpine group only and the rest of the haplotypes were constrained to the other network.

The two haplotype networks are shown in Figure 3.4. The most parsimonious way of connecting the haplotypes is indicated. The haplotype with the highest outgroup probability is displayed as a rectangle and others are displayed as circles. Each colour indicates a haplotype which corresponds to the haplotype frequency distribution shown in Figure 3.5. The size of the haplotype is proportional to the frequency that the haplotype is present. The larger the circle or rectangle, the more frequent the haplotype is present.

Two very frequent haplotypes which are separated by eight steps occur in the two Irish populations from Ben Bulben. The biggest outgroup probability was specified as haplotype 1 which contains haplotypes from Ireland (Ben Bulben), E.Alps, W.Alps, Scotland (Rum) and Svalbard. This is the most frequent haplotype and also one of the most widely distributed. It also conveys a link with the *Arenaria norvegica* haplotypes found on Rum. This haplotype is equally

related to haplotypes 2 (Svalbard), 4 (Ireland (Burren), Pyrenees, Iceland, Scot-



Haplotype Number	Location
1	BB, E/W.Alps, Rum, Scot, Sval
2	Svalbard
3	Iceland, England
4	IreBur, Pyr, Ice, Scot, Eng
6	Picos, Pyrenees
7	Swiss Alps
8	Svalbard
9	IreBB, E.Alps, Picos, Pyr
10	Picos
11	E.Alps
12	E.Alps
13	Pyr
14	E.Alps, Pyr
15	E.Alps
16	W.Alps
17	W.Alps
18	W.Alps
19	E/W.Alps

Figure 3.4: Haplotype network for the *Arenaria ciliata* complex using the *matK* chloroplast region. The circles represent haplotypes and the rectangles indicate the haplotype with the highest outgroup probability and each line represents a mutation. The size of the circles and rectangles are proportionate to the frequency of the haplotype present and each colour represents a haplotype corresponding to the haplotypes in Figure 3.4.

Haplotype 19 is the most distantly related haplotype and is connected to haplotype 1 by 13 steps and haplotype 14 by 9 steps. Although it includes haplotypes from the East and West Alps, it is not connected to the other network which contains the East and West Alps only. Haplotype 14 (E.Alps, Pyrenees) is connected to 4 haplotypes by one step each: haplotype 13 (Pyrenees), 12 (E.

Alps), 11 (E.Alps), 8 (Svalbard). 7 (E.Alps) is connected to haplotype 8 by two steps.

The haplotype 9 is very frequent and widely distributed. It occurs in Ireland (Ben Bulben), E. Alps, Picos De Europa and Pyrenees. This is connected to haplotypes 11 (E.Alps) and 10 (Picos De Europa) by one step each. *MatK*17, which contains a large frequency of haplotypes from West Alps, is connected to haplotypes 16 (W.Alps) and 15 (E.Alps) by one step, and haplotypes 18 (W.Alps) by 9 steps.

3.2.1.4 Haplotype frequency distribution map for the *Arenaria ciliata* complex for *matK*

The haplotype frequencies for each population (Appendix B.1) were represented as pie charts for every population, where a different colour represents each haplotype (Figure 3.5). The colours are coordinated with the haplotypes generated in Figure 3.4. Each of these populations with the haplotype frequencies represented as sections in pie charts were plotted on a map of Europe (Figure 3.4).

It is clear from this map that there are populations which contain many unique haplotypes, such as in the East Alps and there are other populations which there is a large frequency of a haplotype in a few populations, for example haplotype 6 in Picos De Europa. There are other haplotypes which in some populations are the only haplotype represented, but yet are widely distributed. This is the case for haplotype 4 which is distributed in Ireland (Burren), Pyrenees, Iceland, Scotland and England. Surprisingly, this haplotype which predominately occurs in *Arenaria norvegica* populations, also occurs in some *A. ciliata* individuals sampled from the Pyrenees.

The most widely distributed haplotypes appear to be 1 and 9 which both occur in Ben Bulben and Svalbard where they are the two most frequent haplotypes. Ireland and Svalbard are the only two populations where these haplotypes occur together in the same population. They also occur together in the East Alps (Austria) however, they appear in separate populations here.

The population from Rum which is thought only to contain *Arenaria norvegica* contains the haplotype 1 which occurs in populations of *A. ciliata* and this hap-

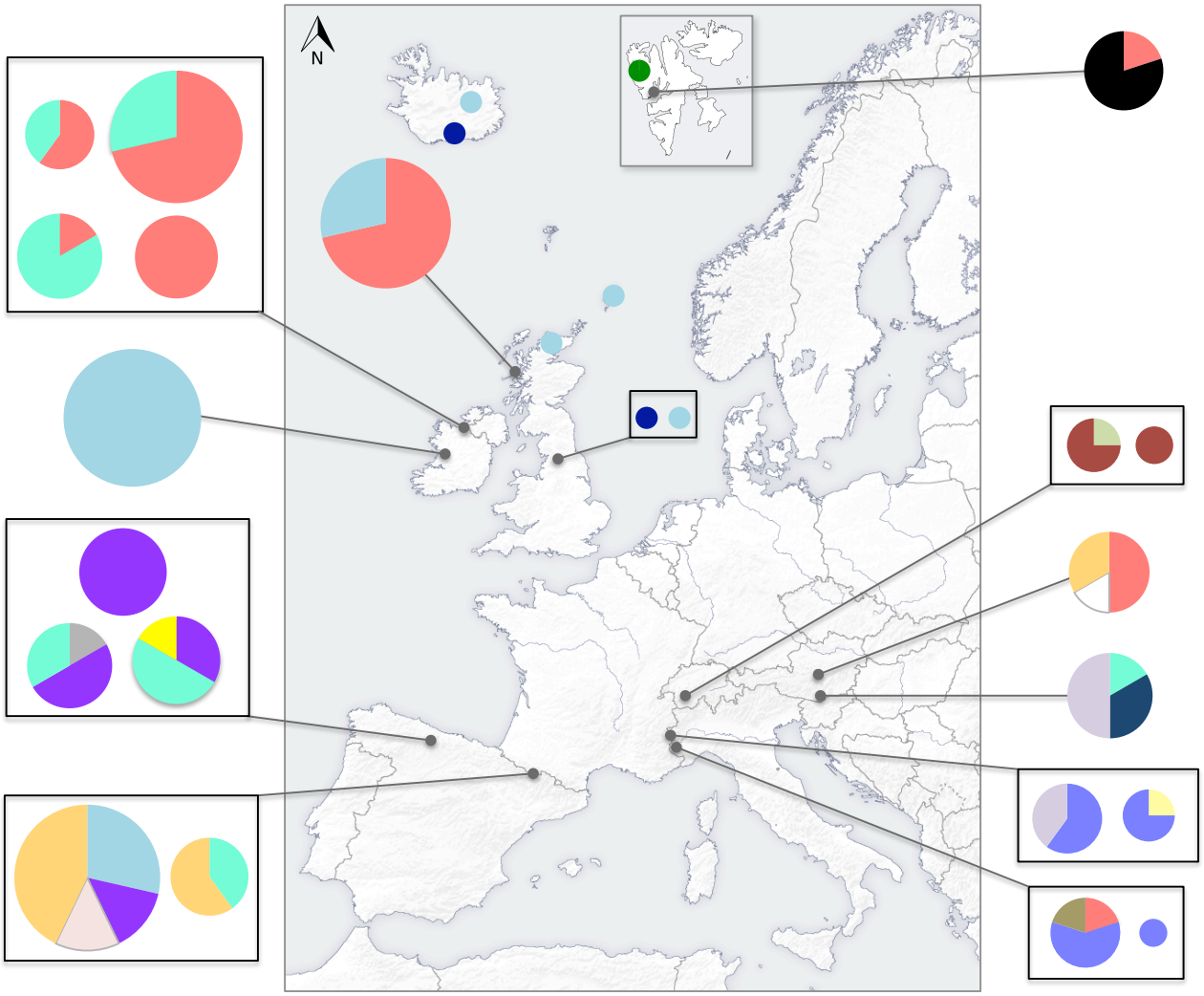


Figure 3.5: Distribution of haplotype frequency in Europe for the *matK* chloroplast region. Each circle represents a population and it is divided by haplotype frequency where each colour represents a haplotype corresponding to the haplotypes in Figure 3.4. Size of the circles is proportional to the frequency of individuals sampled.

lotype is more frequent in this population than 4, which is more predominately associated with *A. norvegica* populations. This population appears to be a mix of haplotypes in between the two Irish populations from Ben Bulben and The Burren.

The E. Alps (Austria) populations contains the most divergence of haplotypes and each population contains different haplotypes. Population AC16 contains haplotype 1, 11 and 2 and AC17 contains haplotypes 9, 15 and 19. The Swiss populations contain two haplotypes which are confined to these populations.

3.2.2 Concatenated cp regions

3.2.2.1 Maximum Likelihood phylogenetic analysis for *Arenaria ciliata* and *A. norvegica* haplotypes

Twenty-one haplotypes were identified for *Arenaria ciliata* and *A. norvegica* (Table 3.2). The maximum likelihood tree generated with sequences represented from these haplotypes is shown in Figure 3.6 with branch lengths measured in the number of substitutions per site.

The bootstrap support values shown on the branches indicate the percentage of bootstrap resampled trees in which the associated taxa clustered together. There are three main lineages (groups A,B,C) with strong bootstrap support values in most branches >80 .

In Figure 3.6 group A represents a broad European distribution including haplotypes from Ireland, The Eastern Alps (Austria) and Spain. Group B represents Picos De Europa (Spain), Ireland and all of the haplotypes from the *A. norvegica* sample group. Group C represents haplotypes from The Eastern and Western Alps.

Table 3.2: Haplotypes generated for *Arenaria ciliata* and *A. norvegica* for the concatenated cp. regions with individuals identified for each haplotype, location and the lineages that they are assigned to according to the ML analysis

Haplotype name	**Individuals**	**Location**	**Lineage**
1 - IreBB	AC1.25, AC3.1, AC4.15	Ireland, Ben Bulben	Group B
2 - IreBB	AC4.1	Ireland, Ben Bulben	Group B
3 - Picos	AC14.1	Picos De Europa, Spain	Group B
4 - Picos	AC15.1, AC11.15, AC15.15	Picos De Europa, Spain	Group B
5 - IreBB	AC2.25, AC3.23, AC2.4	Ireland, Ben Bulben	Group A
6 - Pyr	AC13.1, AC10.1	Pyrenees, Spain	Group A
7 - Aus	AC16.1	East Alps, Austria	Group A
8 - Aus	AC16.15	East Alps, Austria	Group A
9 - Picos	AC11.1	Picos De Europa, Spain	Group A
10 - Picos	AC14.15	Picos De Europa, Spain	Group A
11 - Aus	AC17.1	East Alps, Austria	Group A
12 - W.Alps	AC9.2	West Alps, France	Group C
13 - W.Alps	AC6.16	West Alps, Italy	Group C
14 - W.Alps	AC9.5	West Alps, France	Group C
15 - W.Alps	AC8.2	West Alps, Italy	Group C
16 - W.Alps	AC6.5	West Alps, Italy	Group C
17 - W.Alps	AC7.1	West Alps, Italy	Group C
18 - Aus	AC17.10	East Alps, Austria	Group C
19 - ScotRum	AN4.1, AN4.10	Rum, Scotland	Group B
20 - IreBur/Eng/Scot/Shet	AN2.5, AN5.4, AN1.14, AN6.28	Ireland, England, Scotland	Group B
21 - Ice/IreBur	AN7.2, AN1.3	Iceland, Ireland	Group B
Arenaria serpyllifolia	AS1.17	Kildare, Ireland	Outgroup

The details for the individuals are in Tables 2.1 2.3

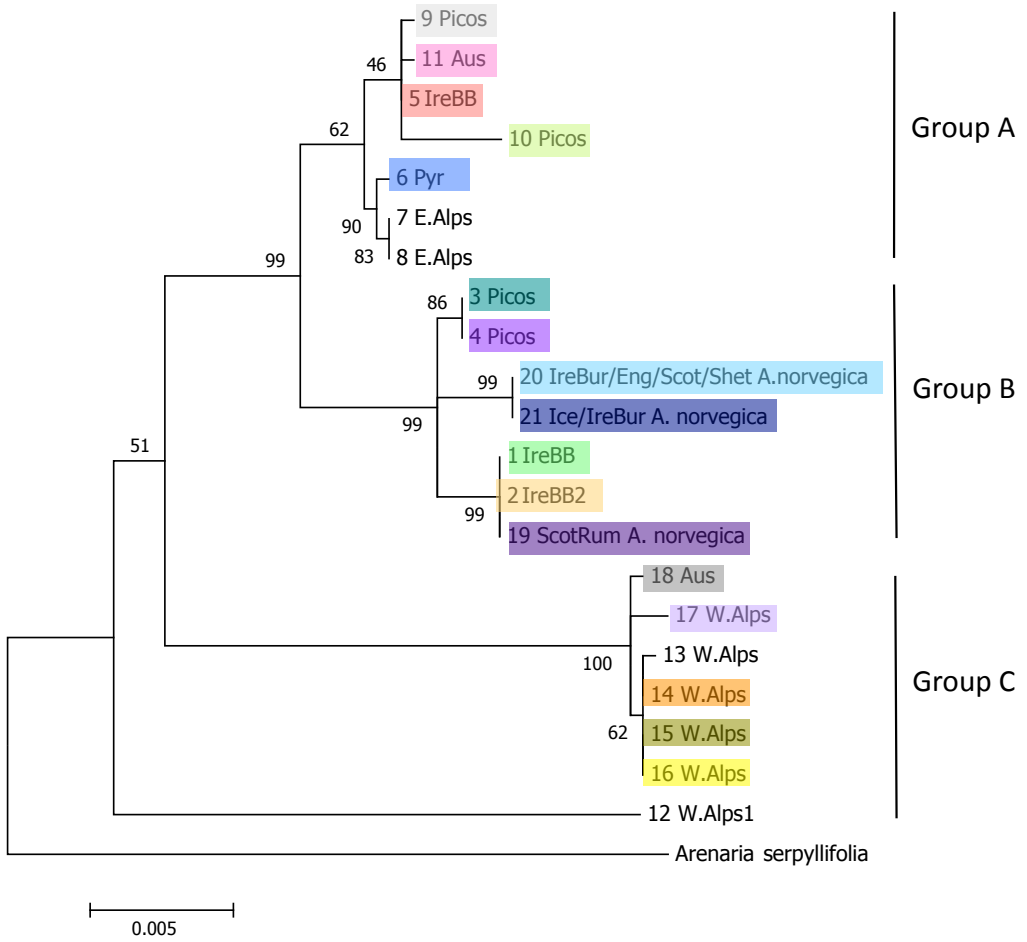


Figure 3.6: Maximum likelihood phylogenetic tree for *Arenaria ciliata* and *A. norvegica* for the concatenated cp. haplotypes with bootstrap support values on the branches. Each label contains haplotype numbers and is colour coded to correspond to the haplotypes in Figure 3.7.

Three haplotyes are unique to Ireland for *Arenaria ciliata* (Figure 3.6). These haplotyes are in two major lineages with haplotype placed in Group A and the haplotypes 1 and 2 placed in group B. The haplotypes in group B are contained within their own distinct lineage and their sister group contains haplotypes 3 and 4 from Picos De Europa (Spain). There is very strong support for this grouping with a bootstrap value of 99. The three haplotypes identified for *A. norvegica* also emerge in two separate clades, one (19 from Rum, Scotland) associated with haplotypes 1 and 2 (*A. ciliata* haplotypes from Ben Bulben). The widespread haplotypes 20 and 21 form their own terminal clade in Group B, and the division between these two clades has 99% bootstrap support.

3.2.2.2 Haplotype network *Arenaria ciliata* and *A. norvegica*

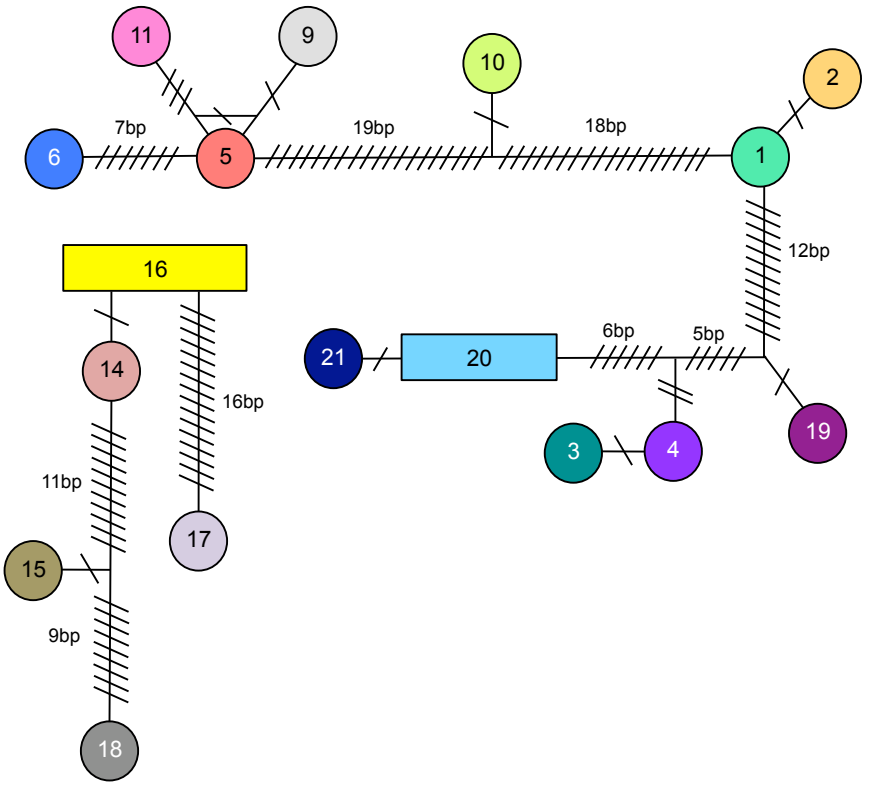
A haplotype network was generated for the two species using *A. serpyllifolia* as an outgroup, with a total of 2,577 characters concatenated from the four chloroplast loci in 34 individual sequences. The haplotypes generated are shown in Table 3.2. The identical sequences assigned to each haplotype (Table 3.2), (other than for the presence of missing or ambiguous characters), were identified by TCS.

Six networks were identified; the two networks containing the majority of the haplotypes are shown in Figure 3.7. The other networks contain individuals haplotypes which were each constrained to their own network including haplotypes: 12 (AC9.2), 13 (AC6.16) and *Arenaria serpyllifolia* (AS1.17). The last network contained two individuals from Austria each assigned to a haplotype and separated by one mutation (haplotypes 7 and 8 (AC16.1, AC16.15)).

The most parsimonious way of connecting the haplotypes is indicated. The haplotype with the highest outgroup probability is displayed as a rectangle and others are displayed as circles. Each colour represents a geographical region; green: Ireland and Britain, blue: Picos De Europa, orange: East Alps, Purple: West Alps, Yellow: Pyrenees. The grouping of the three major lineages is congruent with the maximum likelihood phylogenetic tree is shown in Figure 3.6. This analysis demonstrates in more detail how many mutational steps exist between haplotypes and how exactly these haplotypes are related in the optimal network.

The greatest outgroup probability was specified as haplotype 20 (*Arenaria norvegica*) in the largest network and in the network containing the Alpine haplotypes only is 16. The other networks include an Austrian only group with two individuals from the Austrian alps, *Arenaria serpyllifolia* only, and the haplotypes 12 and 13 were each constrained to their own network.

The three distinct Irish haplotypes identified in *Arenaria ciliata* are congruent with the placement of these haplotypes in Figure 3.6 where haplotype 5 is placed within a broad Europan group. Here it is connected to haplotype 9 by one mutation, haplotype 11 by three mutations and haplotype 6 by seven mutations. The other two Irish haplotypes, 1and 2, are separated from the Irish haplotype 5 by 37 and 38 mutations and between these two groupings of the Irish haplotypes is a Spanish haplotype, (haplotype 10), from Picos De Europa.



Hap. Number	Location
1	Ben Bulben
2	Ben Bulben
3	Picos De Europa
4	Picos De Europa
5	Ben Bulben
6	Pyrenees
9	Picos De Europa
10	Picos De Europa
11	East Alps
14	West Alps
15	West Alps
16	West Alps
17	West Alps
18	East Alps
19	Rum
20	Burren, Eng, Scotland, Shet
21	Iceland, Burren

Figure 3.7: Haplotype network for *Arenaria ciliata* and *A. norvegica* for the concatenated chloroplast regions. The circles represent haplotypes and the rectangles indicate the haplotype with the highest outgroup probability and each dash represents a mutation. Each colour represents a distinct haplotype which also correspond with the haplotypes displayed in Figure 3.8. The haplotype numbers with the corresponding location are also shown.

Arenaria norvegica is also placed within the largest network containing *A. ciliata* and three haplotypes are identified here, two of which are Irish. The haplotype containing individuals from Ireland, England, Scotland (haplotype 20) is connected to another *A. norvegica* haplotype (haplotype 21) from Ireland and Iceland by one mutation.

The other *A. norvegica* haplotype is distinct to Rum, Scotland (haplotype 19) and this haplotype is more closely related to the *Arenaria ciliata* haplotypes from Picos De Europa (8 and 9 mutations) than to the other *A. norvegica* haplotypes (12 and 13 mutations).

3.2.2.3 Frequency distribution map for the concatenated chloroplast haplotypes

The concatenated chloroplast haplotype frequencies for each population (Appendix A.1) were represented as pie charts for every population, where a different colour represents a distinct haplotype corresponding to the haplotypes generated in Figure 3.7. Each of these populations with the haplotype frequencies represented as sections in pie charts were plotted on a map of Europe in Figure 3.8. The size of the circles are proportional to the frequency of individuals sampled.

The presence of many unique haplotypes is clear in this map, and the uniformity of the haplotypes for *Arenaria norvegica* is particularly evident where there are only two haplotypes for most of the populations sampled for *A. norvegica*. The exception to this is the population sampled in Rum, which contains a unique haplotype. The populations from the Burren contain a haplotype which is also shared with a population in Iceland, and the other more widespread haplotype which occurs in England and Scotland.

The presence of three unique haplotypes in the Ben Bulben populations is also particularly striking, where the number of haplotypes is comparable with that of continental European meta-populations such as in Picos De Europa where 4 unique haplotypes are sampled. Two of the Ben Bulben haplotypes are frequently present in at least two populations, where one haplotype is restricted to a single population.

The Alps contain a particularly diverse assemblage of haplotypes, where the highest diversity is found in the Western Alps where six unique haplotypes were sampled.

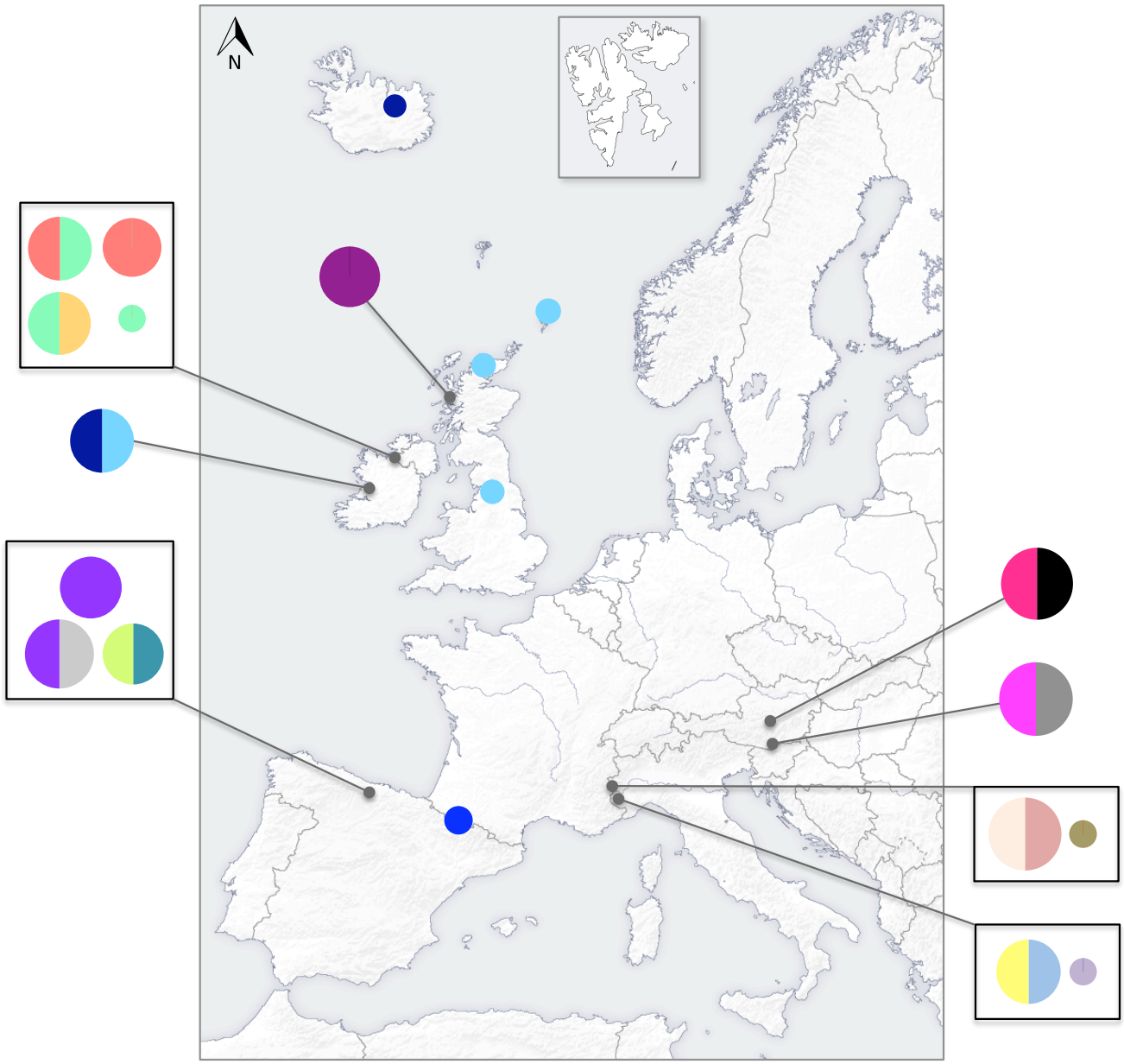


Figure 3.8: Distribution of haplotype frequency in Europe for the concatenated chloroplast region. Each circle represents a population and it is divided by haplotype frequency where each colour represents a haplotype corresponding to the haplotypes in Figure 3.7. Size of the circles are proportional to the frequency of individuals sampled.

3.2.3 ITS nuclear region

3.2.3.1 Maximum Likelihood intraspecifc phylogenetic analysis for the *Arenaria ciliata* complex nucleotypes using ITS sequence data

Eight nucleotypes were identified from ITS sequence data from 34 individuals of *Arenaria ciliata* and *A. norvegica* (Table 3.3). A representative sequence from each haplotype was used in the maximum likelihood phylogenetic analysis. The two phylogenetic trees constructed using *Arenaria nevadensis* and *A. serpyllifolia* as outgroups are shown in Figure 3.9. Bootstrap values are shown on the branches.

The topology in both trees is conserved, however, the tree which includes *Arenaria serpyllifolia* and consequently 191 more base pairs demonstrates better resolution in some of the branches and an increased bootstrap support. The trees here demonstrate a clear geographic association between the nucleotypes, i.e. the nucleotypes are clustered together depending on their location.

In the tree which includes *Arenaria serpyllifolia* as an outgroup, the Spanish lineage, which contain The Pyrenees and Picos De Europa is placed as the basal nucleotype. The West Alps contain the greatest number of haplotypes (3), whereas all other locations contain only one representative nucleotype. ITS3 which contains haplotypes from the *Arenaria norvegica* populations is suggested as the sister group to the Irish *A. ciliata* nucleotype, ITS4. The placement of ITS7 (Austria) is not well resolved and falls in between the West Alps lineage and the Irish Ben Bulben nucleotype (ITS4) and the *A. norvegica* group (ITS3).

Table 3.3: Individuals assigned to haplotypes for the ITS region for the *Arenaria ciliata* complex

Haplotype Name	**Individual samples carrying haplotype**	**Location**
ITS1 Picos	AC11.15, AC14.1, AC14.15, AC15.15, AC15.1, AC11.1	Picos De Europa, Spain
ITS2 Pyr	AC10.15, AC13.1, AC10.1	Pyrenees, Spain
ITS3 A.norvegica	AN6.28, AN7.2, AN6.16, AN5.4, AN4.10, AN4.1, AN2.5, AN1.3, AN1.14	Iceland, Scotland, England, Ireland (Burren)
ITS4 IreBB	AC2.25, AC1.8, AC3.23, AC1.25, AC2.4, AC4.1, AC3.1, AC4.15	Ireland (Ben Bulben)
ITS5 W.Alps1	AC6.16, AC6.5, AC9.2, AC8.1, AC9.5	West Alps (Italy, France)
ITS6 W.Alps2	AC7.1	West Alps (Italy)
ITS7 W.Alps3	AC8.2	West Alps (Italy)
ITS8 Aus	AC16.15, AC17.10, AC17.1, AC16.1	East Alps (Austria)
Arenaria serpyllifolia	AS1.17	Ireland (Kildare)

3.2.3.2 Nucleotype network for the *Arenaria ciliata* complex for ITS

A nucleotype network for all species, including *Arenaria serpyllifolia* as an outgroup, was generated with an alignment of 34 taxa and 715 characters from the ITS nuclear region. The nucleotypes generated are in Table 3.3 and the identical sequences which differ by missing or ambiguous characters are shown. One main network was constructed, the output from TCS is shown in Figure 3.10. The maximum number of steps connecting the nucleotypes parsimoniously is indicated. The nucleotype with the highest outgroup probability is displayed as a rectangle and others are displayed as circles.

The biggest outgroup probability was specified as ITS1 which contains nucleotype from Picos De Europa. This nucleotype is connected to ITS2, Pyrenees, by one step. The West Alps contain three nucleotypes connected by one step and the other Alpine population in Austria, ITS8, is more distantly related, connected to the Spanish nucleotypes by 8 steps. Distinct nucleotypes are assigned to the populations for *Arenaria norvegica* seen in ITS3 (Iceland, Scotland, England) and this is the most distantly related nucleotype. The most closely related nucleotype is ITS4 from Ben Bulben, Ireland. ITS3 is the only nucleotype which contains individuals from different locations.

3.2.3.3 Frequency distribution map for the ITS nucleotypes

The ITS nucleotype frequencies for each population (Appendix C.1) were represented as pie charts for every population, where a different colour represents a distinct nucleotype corresponding to the nucleotypes generated in Figure 3.10. Each of these populations with the nucleotype frequencies represented as sections in pie charts were plotted on a map of Europe in Figure 3.11. The size of the circle depicting the nucleotype is proportional to the frequency that the nucleotype is present.

It is very clear that there is much less diversity of nucleotypes represented for the ITS region than for the chloroplast region. There is very clear geographical structure for the nucleotypes of *Arenaria ciliata*, where each region contains a distinct nucleotype. In most regions however, there is only one ITS nucleotype represented in the populations.

The region which contains the highest diversity of nucleotypes is the West Alps where there are three ITS nucleotypes compared to just the 1 nucleotype observed in every other population, and two of these nucleotypes are contained in a single population (AC7).

The distribution pattern for the ITS nucleotypes for *Arenaria norvegica* is similar to that observed for the cp. haplotypes where the nucleotypes are not structured according to geographic region. There is one ITS nucleotype found in all of the sampled populations of *A. norvegica*. In contrast to the haplotype distribution map for the concatenated cp. region, Rum (West Scotland) does not contain a unique nucleotype.

3.2.4 Maximum likelihood analysis for comparison of the ITS nuclear and concatenated chloroplast regions

It is clear that there is a large discrepancy between the nuclear ITS data and the chloroplast data. In order to visualise this more clearly, a simultaneous maximum likelihood analysis was carried out for the individuals for the 4 concatenated chloroplast regions and the nuclear ITS regions with matching individuals. This was carried out for both *Arenaria ciliata* and *A. norvegica* and the resulting trees are shown in Figure 3.12. The names for the individuals are shown in Table 3.4.

Branches on the two trees generated for ITS and the cp. concatenated region were orientated to match up as closely as possible and lines were drawn to connect the equivalent individuals in both trees. The lines were coloured to represent geographical areas. In Figure 3.12 the green lines indicate Ireland and populations of *Arenaria norvegica* from Britain and Iceland, orange represents Spain and purple represents the Alps.

It is clear from these trees that the ITS sequences form distinct geographic groupings, whereas the chloroplast sequences do not depict a clear geographic association and some individuals from the same population occur in two or more different lineages (e.g. Austria).

The West Alps appear to be the most congruent geographic group and the chloroplast regions demonstrate a relatively conserved association with the ITS

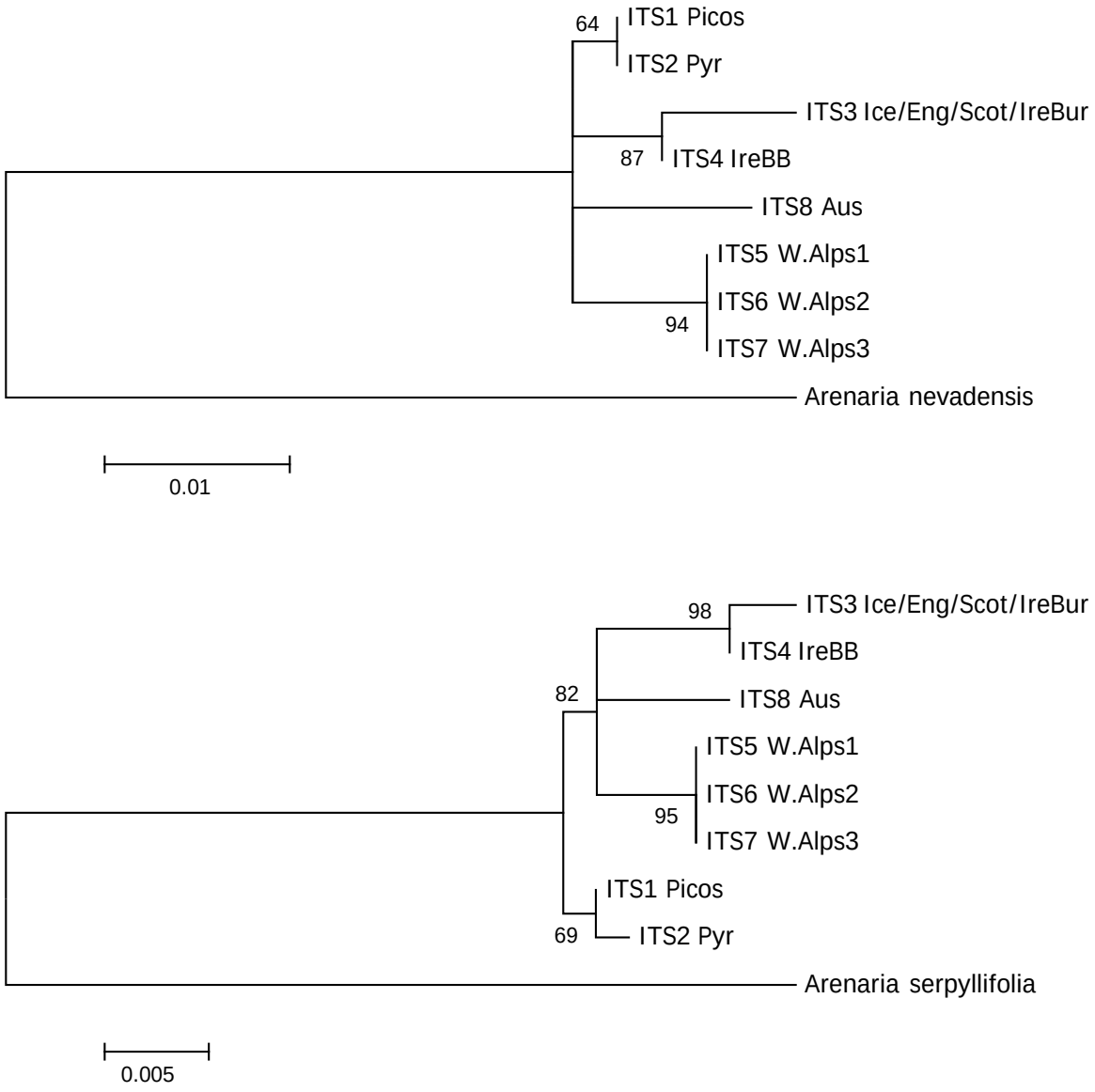
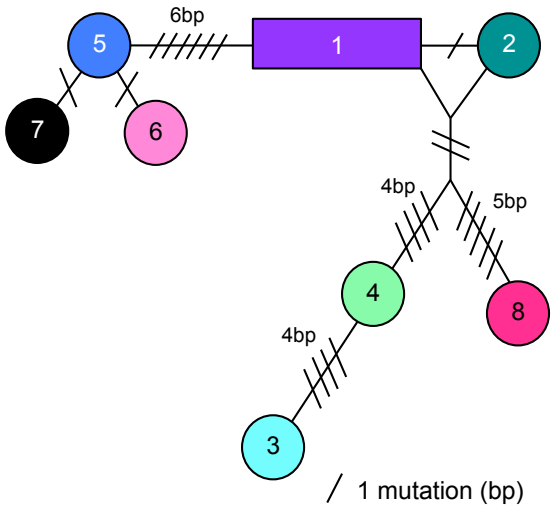


Figure 3.9: Maximum likelihood phylogenetic trees for the nuclear ITS nucleotypes for the *Arenaria ciliata* complex. In the top tree *A. nevadensis* is included as an outgroup and the bottom tree includes *A. serpyllifolia* as an outgroup. Bootstrap support values are shown on the branches.

sequences. Many of the individuals from Picos De Europa also match up well for both regions, however, two of these individuals fall into a different lineage in the chloroplast group. The two major Spanish locations, Picos De Europa and the



Haplotype Number	Location
1	Picos De Europa
2	Pyrenees
3	Ice, Scot, Eng, IreBur (*A. norvegica*)
4	Ben Bulben
5	W.Alps
6	W.Alps
7	W.Alps
8	E.Alps

Figure 3.10: Haplotype network for the *Arenaria ciliata* complex using the ITS nuclear region. The circles represent haplotypes and the rectangles indicate the haplotype with the highest outgroup probability and each crossed line represents a single bp. difference. The size of the circles and rectangles are proportionate to the frequency of the haplotype present.

Pyrenees, are grouped together in the ITS tree, however, in the chloroplast tree they are segregated into two separate lineages, the Pyrenees being grouped with Austria, and Picos being associated with multiple lineages.

The individuals sampled from Ireland also group together in the ITS region and are divided into two separate lineages in the chloroplast tree. The inclusion of *A. norvegica* reveals that, except in the case of Rum (Scotland), the chloroplast haplotypes are congruent with the nucleotypes. This is the only congruent group which extends over multiple geographic locations where Iceland, Ireland, Shetland, England and Scotland are all represented.

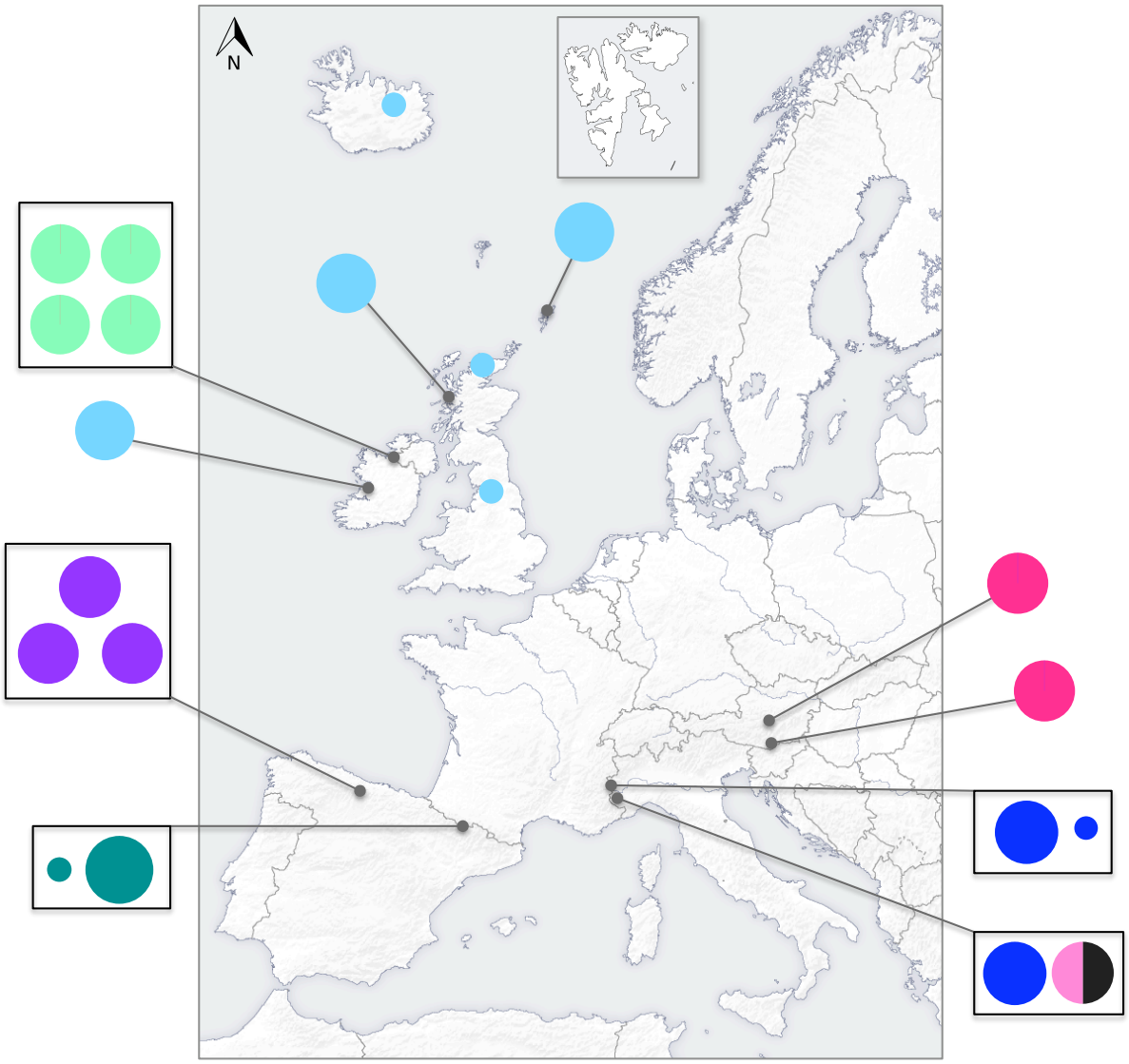


Figure 3.11: Distribution of nucleotype ITS frequency in Europe. Each circle represents a population and it is divided by nucleotype frequency where each colour represents a nucleotype corresponding to Figure 3.10. Size of the circles is proportional to the frequency of individuals sampled.

Table 3.4: Names for individuals for concatenated regions and ITS for *Arenaria ciliata* and *A. norvegica*

Taxa name	Individual name	Location
Aus1	AC16.1	Austria
Aus2	AC16.15	Austria
Aus3	AC17.1	Austria
Aus4	AC17.10	Austria
EngYork (*A. norvegica*)	AN2.5	Yorkshire, England
Ice (*A. norvegica*)	AN7.2	Iceland
IreBB1	AC2.25	Ben Bulben, Ireland
IreBB2	AC3.23	Ben Bulben, Ireland
IreBB3	AC4.15	Ben Bulben, Ireland
IreBB4	AC1.25	Ben Bulben, Ireland
IreBB5	AC2.4	Ben Bulben, Ireland
IreBB6	AC3.1	Ben Bulben, Ireland
IreBB7	AC4.1	Ben Bulben, Ireland
IreBur1 (*A. norvegica*)	AN1.14	Burren, Ireland
IreBur2 (*A. norvegica*)	AN1.3	Burren, Ireland
Picos1	AC11.15	Picos De Europa, Spain
Picos2	AC14.15	Picos De Europa, Spain
Picos3	AC15.1	Picos De Europa, Spain
Picos4	AC11.1	Picos De Europa, Spain
Picos5	AC14.1	Picos De Europa, Spain
Picos6	AC15.15	Picos De Europa, Spain
Pyr1	AC10.1	Pyrenees, Spain
Pyr2	AC13.1	Pyrenees, Spain
ScotInch (*A. norvegica*)	AN5.4	Inchnadamph, Scotland
ScotRum1 (*A. norvegica*)	AN4.1	Rum, Scotland
ScotRum2 (*A. norvegica*)	AN4.10	Rum, Scotland
ScotShet (*A. norvegica*)	AN6.28	Shetland, Scotland
W.Alps1	AC6.16	West Alps, Italy
W.Alps2	AC9.5	West Alps, France
W.Alps3	AC6.5	West Alps, Italy
W.Alps4	AC7.1	West Alps, Italy
W.Alps5	AC8.2	West Alps, Italy
W.Alps6	AC9.2	West Alps, France
Arenaria serpyllifolia	AS1.17	Kildare, Ireland

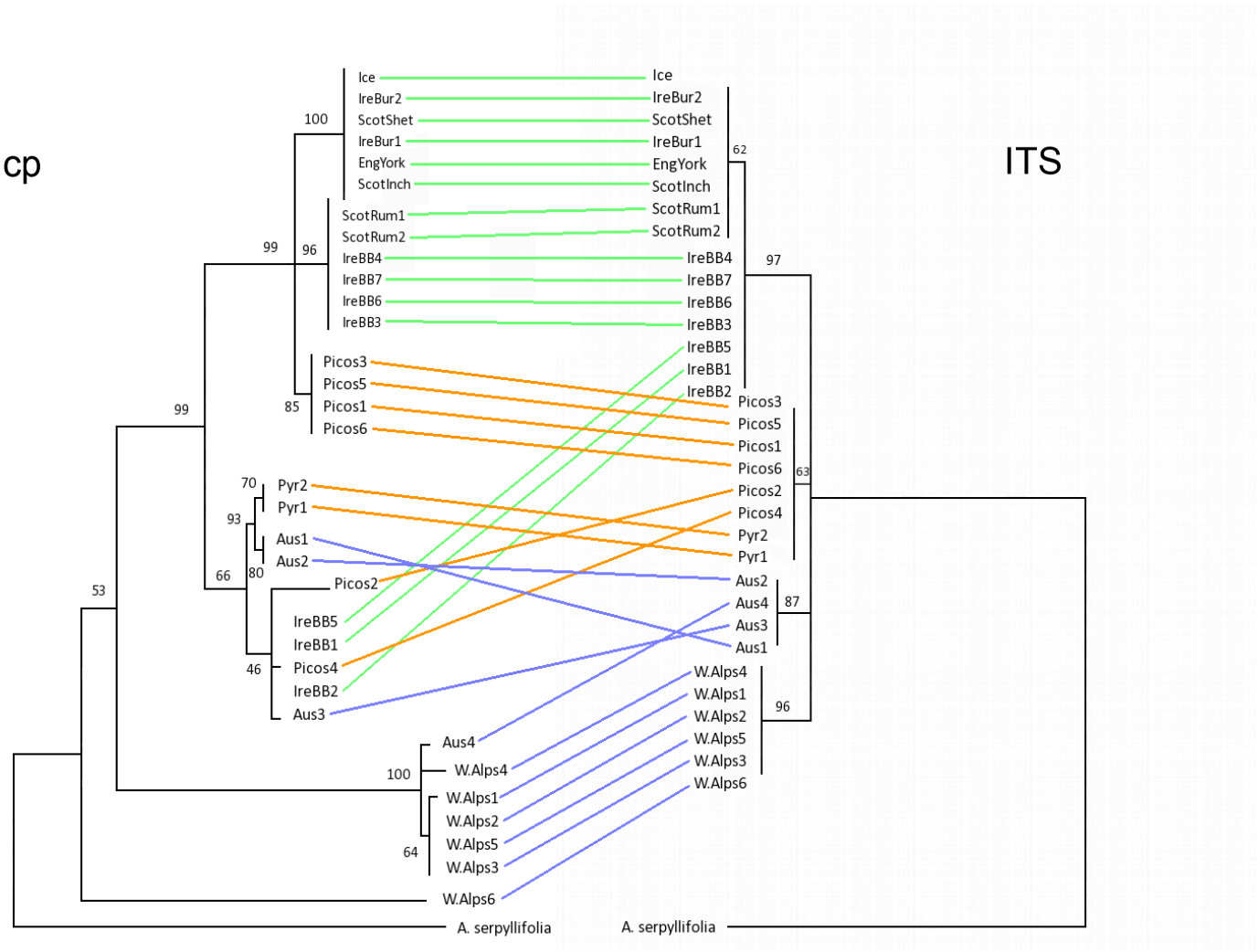


Figure 3.12: Maximum likelihood trees based on simultaneous analysis of the concatenated cp. regions and ITS for *Arenaria ciliata* and *A. norvegica* with BS values on the branches. The coloured lines represent geographical areas: Green-Ireland/Britain/Iceland, Orange-Spain, Purple-Alps. The tree to the left contains the concatenated chloroplast haplotypes and the tree to the right contains the ITS nucleotypes.

3.3 Discussion

3.3.1 Taxonomic relationship between *Arenaria ciliata* and *A. norvegica*

A complicated taxonomic relationship between *Arenaria norvegica* and *A. ciliata*, where both taxa are consistently grouped together in the same lineage, is evident when analyzing either chloroplast or nuclear markers. This relationship is also observed when many other taxa from Caryophyllaceae are included. The difficulties in discerning taxonomic relationships among the *Arenaria* complex has also been noted using morphological characters (Wyse Jackson and Parnell, 1987).

Another complication in the resolution of these two species is the grouping of the *Arenaria norvegica* Rum (West Scotland) population with the *A. ciliata* population from Ben Bulben. According to the TCS network using the concatenated chloroplast regions, the *A. norvegica* Rum haplotype is intermediate between the Ben Bulben and Picos De Europa *A. ciliata* haplotypes, and the other *A. norvegica* haplotype (which is common in all other *A. norvegica* populations) is more closely related to the Picos De Europa haplotype. The *A. norvegica* haplotypes are therefore polyphyletic within this species complex and have potentially arisen from different ancestors. When more individuals are included using the *matK* chloroplast region, the Rum *A. norvegica* haplotype is grouped with *A. ciliata* from Ben Bulben, and also from Austria, West Alps and Svalbard. The common *A. norvegica* haplotype is observed in a population of *A. ciliata* from the Pyrenees. The ITS region demonstrates that *A. norvegica* is also monophyletic within *A. ciliata*, and it is placed in the same lineage as the *A. ciliata* individuals from Ben Bulben. If these taxa are to be recognised as two distinct species, they should theoretically emerge in two distinct lineages.

Evolutionary rates in Arctic plants may be lower than temperate species as a result of extreme environmental constraints (Hulten, 1937). This may explain the conservative nature of the *Arenaria ciliata* complex compared to the high speciation rates observed in more temperate genera such as *Dianthus* (Valente *et al.*, 2010). In several arctic plant groups, speciation has occurred quite recently, probably within the period of the major glacial cycles, less than one million years

ago (Brochmann and Brysting, 2008). Low within population variation is suggestive of a very young flora and is both inadequate for phylogenetic reconstruction and a feature of many arctic species (Brochmann and Brysting, 2008). In some prominent arctic plant genera such as *Artemisia*, *Papaver* and *Salix* predominant markers which have been suggested as barcodes for land plants such as *trnL* intron, *rpo*C1, *rpo*B, *matK* and *trnHpsbA* are insufficient in resolving species distinctions in these genera as all or most of these arctic species demonstrated identical sequences (Brochmann and Brysting, 2008). The authors suggest that this is due to speciation occurring in these groups both rapidly and recently. This appears also to be the case for *Arenaria norvegica* and therefore the markers used in this study may be insufficient in resolving species phylogenies between these taxa.

Other studies have shown similar observations in the evolution of arctic flora such as Ronikier *et al.* (2008)who also found that the distribution of haplotypes do not coincide with the taxonomic varieties and the authors attribute this to recent divergence or environmental influence through phenotypic plasticity.

DNA phylogenies can often be discordant with organismal phylogenies and factors which can account for this are sampling error, hybridization, concerted evolution, incomplete lineage sorting, convergence and evolutionary rate heterogeneity (Acosta and Premoli, 2010; Avise, 2000; Doyle, 1992; Eidesen *et al.*, 2007; Riesberg and Soltis, 1991; Riesberg *et al.*, 1996). Figure 3.13 demonstrates some of these processes which result in incongruent organismal and gene phylogenies and individuals which are polyphyletic. ITS is particularly prone to concerted evolution which can confound the interpretation of sequences at the intraspecific level (Buckler *et al.*, 1997). The difficulties in discerning the taxonomic relationship between *Arenaria ciliata* and *A. norvegica* could likely be accounted for by a combination of these processes resulting in the production of a gene phylogeny predating the origin of *A. norvegica* but which may not reflect the extant organismal phylogeny.

3.3.2 Postglacial history of *Arenaria ciliata* and *A. norvegica*

3.3.2.1 The Irish haplotypes - two distinct lineages

One of the most striking results to emerge from this study is the separation of the two *Arenaria ciliata* haplotypes on Ben Bulben into two relatively divergent lineages. If the population in Ben Bulben was a recent post glacial immigrant, it would be expected to contain few unique haplotypes and low genetic diversity compared with populations which remained *in situ* in refugia (Comes and Kadereit, 1998; Hewitt, 1996; Schaal *et al.*, 1998). This pattern is not evident in the data presented here - unique haplotypes are observed in the concatenated dataset and multiple haplotypes which belong to divergent lineages are observed for both *matK* and the concatenated regions. Ancient haplotypes should be located at the centre of the gene tree and be geographically widespread (Templeton *et al.*, 1995) which is observed here for the Irish haplotypes (Figure 3.12). This corroborates other phylogeographical studies which present evidence for an Irish refugium (Davison, 2000; Hoarau *et al.*, 2007; McKeown *et al.*, 2010; Provan *et al.*, 2005; Sinclair *et al.*, 1998; Teacher *et al.*, 2009).

The anomaly where Irish individuals of *Arenaria ciliata* from the same population separate into divergent lineages has also been documented for the common frog (Teacher *et al.*, 2009). Two major lineages of mtDNA are identified for the common frog, one unique to Ireland and one which contains haplotypes widely distributed in Western Europe (Teacher *et al.*, 2009). This is similiar to the pattern seen for *Arenaria ciliata* for the concatenated data, where some individuals are grouped in a distinct clade in Ireland and others are assembled in a clade that is widely distributed in Europe.

3.3.2.2 Ireland and Scotland - A north Atlantic western refugium?

Another striking observation evident from this study is that the Irish *Arenaria ciliata* haplotype which belongs to Group B is more closely related to *A. norvegica* than the other Irish haplotype belonging to the same population in Ben Bulben. Huntley and Birks (1983) and Birks (1989) acknowledge the possibility of a refugium based on the early Holocene movement of pollen in, or near, West

Britain. Maps in Huntley and Birks (1983) (pg. 650) indicate potential refugia in the Outer Hebrides, the north west tip of Skye and in the west central part of Ireland. Birks (1989) suggest that the populations in west Ireland or northern Scotland may contain an independent glacial refugia. Pollen evidence indicates the spread of Scots Pine from Ireland (Birks, 1989) and these authors suggest the possibility of sheltered refugia in or near northwest Scotland during the LGM.

The analysis of mitochondrial DNA (Sinclair *et al.*, 1998), isozyme analysis (Kinloch *et al.*, 1986) and the anomalous distribution of pollen (Bennett, 1984; Birks, 1989; Huntley and Birks, 1983) in Scots Pine (*Pinus sylvestris*) indicate that north western Scottish populations in Wester Ross are very distinct from other British and continental populations and they were derived from multiple refugia during the LGM. The individuals from Wester Ross are also grouped in multiple lineages similar to *Arenaria norvegica* from Rum. When more populations from Europe were analysed for mtDNA and isoenzyme variation and genetic diversity in Scots Pine, both mitotypes from Scotland are widespread throughout Europe, however, like the distribution of haplotypes observed in *Arenaria ciliata*, the European distribution for each haplotype are different and belong to multiple lineages (Sinclair *et al.*, 1999; Prus-Glowacki *et al.*, 2012). Prus-Glowacki *et al.* (2012) also found that peripheral populations from the Iberian Peninsula and Scotland contained a slightly higher level of heterozygosity than the other regions.

The chloroplast data from *Arenaria ciliata* and *A. norvegica* may indicate that these species were co-occurring in a suture zone at some point in the past. This north-west Scottish/Irish association suggests a distinct grouping that arose independently and may indicate a glacial refugium where *Arenaria norvegica* and potentially other species such as *Pinus sylvestris* could have expanded from this population. The presence of two unique *matK* haplotypes in Svalbard may also indicate a cryptic northern refugium for *Arenaria ciliata*. This corroborates the recent claim for the survival of the arctic-alpine Caryophyllace species *Sagina caespitosa* and *Arenaria humifusa* in Svalbard (Westergaard *et al.*, 2011). Westergaard *et al.* (2011) demonstrated using AFLPs that patterns of genetic diversity and distinctiveness indicate the survival of these species in Svalbard. The presence of *matK* haplotype 1 which is present in Ireland, Rum, Svalbard and the

Alps, which all contain a high diversity of haplotypes, could be indicative of an ancient haplotype which survived the LGM in these likely glacial refugia.

3.3.2.3 Postglacial expansion of *Arenaria norvegica*

Arenaria norvegica contains nuclear ITS sequences from different countries grouped in the same lineage whereas *A. ciliata* contains distinct lineages which reflect the individuals geographic origin. This pattern could suggest a more recent colonisation for *A. norvegica* that radiated from the same founder populations and subsequently formed a bottleneck resulting in decreased genetic diversity and the production of the same chloroplast sequence for many regions. It is possible that this founder population originated from *A. ciliata* as *A. norvegica* is monophyletic within *A. ciliata*, even according to the nuclear sequences which suggest a more recent phylogenetic history. It is also possible that it originated from a population in Ireland or Spain as these are the closest sister lineages identified from the concatenated chloroplast sequences and according to *matK* there are many potential origins in Ireland, Austria, West Alps, Svalbard and Spain.

The reduced genetic diversity of *Arenaria norvegica* compared to *A. ciliata* reflects a pattern typical for postglacial expansion also observed in Ireland for species such as St. Dabeoc's heath (*Daboecia cantabrica*) (Beatty and Provan, 2012), the mountain hare (Hamill *et al.*, 2006), stoats (Martinkova *et al.*, 2007) and oak trees (Kelleher *et al.*, 2004).

The biogeography of *Arenaria norvegica* reflects that seen in other arctic plants - that is low genetic diversity with identical sequences over vast northern areas. This has been observed in other arctic plants such as *Arabis alpina* (Ehrich *et al.*, 2007) which consist of a single or very few AFLP genotypes in the north, but contain higher diversity levels in their southern populations as a result of repeated bottlenecks during postglacial colonization.

3.3.3 Incongruence between nuclear and chloroplast sequence data

Inconsistencies between the nuclear and chloroplast phylogenies observed in the *Arenaria ciliata* complex may provide insight into potential origins for *A. norvegica* and reveal ancient hybridization and chloroplast capture events that occurred in this complex. Differences may arise through chloroplast capture, and bottlenecks alone can create discordance between cytoplasmic and nuclear DNA (Fay and Wu, 1999). In this study the nuclear ITS region reflects the geographic associations among the individuals more accurately than the chloroplast DNA for *A. ciliata*. The geographic association for the chloroplast sequences were more complex where individuals in populations from Ireland, Austria and Spain were polyphyletic and did not cluster according to location.

Homogenisation and recent gene flow events in the nuclear ITS region can account for these distinct geographic groupings in *Arenaria ciliata*. The congruence observed between the geographical structuring of the ITS nucleotypes indicate a longstanding pattern of restricted gene flow (Schaal *et al.*, 1998). As chloroplast DNA is much more slowly evolving than the nuclear region, this may give us insight into ancient interactions among populations and between these two species, so that anomalies that are observed between the nuclear and chloroplast regions may be explained by past colonisation, speciation and hybridisation processes. These anomalies include the discrepancy observed for the nuclear signature in the *Arenaria norvegica* Rum (West Scotland) population which is grouped with the other *A. norvegica* individuals for the ITS nuclear region, but contains chloroplast similar to or the same as individuals of *A. ciliata* from Ben Bulben (concatenated and *matK*). When more individuals are included from *matK* this Rum/Ben Bulben haplotype is also evident in many other locations including: Austria, West Alps and Svalbard. This phenomenon could be indicative of an *Arenaria ciliata* haplotype that was previously more widespread many years ago and is now retained in the population of *A. norvegica* on Rum.

Chloroplast capture and hybridization have been commonly used to account for discrepancies between plastid and nuclear phylogenetic trees (Acosta and Premoli, 2010; Guo *et al.*, 2004; Kim and Donoghue, 2008; Okuyama *et al.*, 2005;

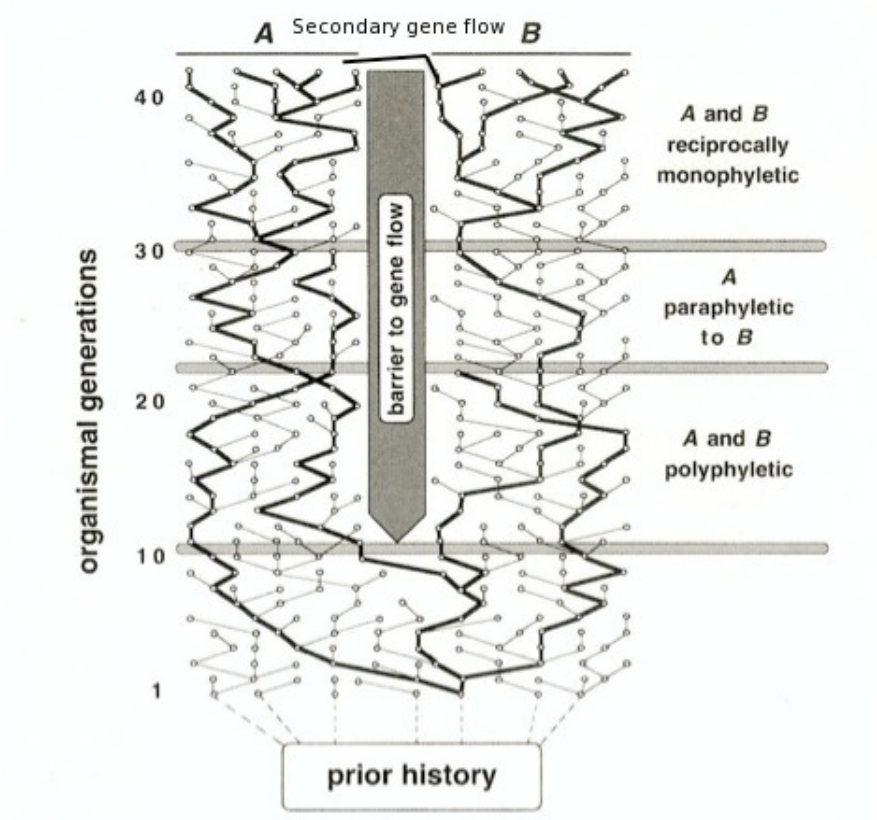


Figure 3.13: Schematic representation of the sorting process of genes in two populations A and B separated by a gene flow barrier (modified from Avise (2000)). The black lines represent key lineages that help define the phylogenetic categories.

Riesberg and Soltis, 1991; Rieseberg, 1995; Stegemann *et al.*, 2012; Tsitrone *et al.*, 2003; Van Droogenbroeck *et al.*, 2006; Van Raamsdonk *et al.*, 1997). Chloroplast capture, the process in which genomes of chloroplasts can be laterally transmitted between organisms, has been commonly documented among closely related species and within species complexes (Riesberg and Soltis, 1991). Replacement of the host plastid genome by the foreign genome can occur in the absence of nuclear gene flow and intergenomic recombination (Stegemann *et al.*, 2012). This would result in populations, or individuals of a species, characterized by the morphology or nuclear markers of one species, but the chloroplast genome of another species (Rieseberg *et al.*, 1995). These processes may occur at a remarkable speed (Riesberg and Soltis, 1991; Tsitrone *et al.*, 2003) and have recently been demonstrated

through grafting sexually incompatible species (Stegemann *et al.*, 2012).

Hybridization may occur in-between species such as *Arenaria ciliata* and *A. norvegica* where individuals of one species are in the minority and receive foreign pollen belonging to the related taxa as a result of interspecific pollen competition (Acosta and Premoli, 2010). This could explain the phenomenon where *Arenaria norvegica* in Rum (West Scotland) contains the chloroplast genome from *A. ciliata*. Likewise, the occurrence of an *Arenaria norvegica matK* haplotype in the Pyrenees, which also occurs in *A. norvegica* populations in Ireland, Iceland, England and Scotland, may suggest that the chloroplast was donated from a few individuals of *A. norvegica* which consequently became fixed in a Pyrenees population of *A. ciliata*.

Male sterility may promote chloroplast capture as energy normally invested in pollen production could be allocated to ovules and increasing female fitness (Tsitrone *et al.*, 2003), thus promoting the presence of the maternally inherited chloroplast genome in the offspring. Continual introgressions would lead to a majority of individuals containing the *A. ciliata* cp DNA, but with a *A. norvegica* nuclear signature, as seen in the subgenus *Nothofagus* (Acosta and Premoli, 2010). Individuals from this founder population could subsequently colonise new locations potentially expanding the distribution of the species (Acosta and Premoli, 2010).

3.3.4 Conclusion

It is clear that Ireland contains a complex postglacial history for *Arenaria ciliata* and that species distinctions between *A. ciliata* and *A. norvegica* remain unresolved. However, it is evident from this analysis from the occurrence of highly divergent haplotypes in the Irish population and the presence of unique haplotypes observed in the concatenated chloroplast regions are indicative of an Irish refugium and potential expansion outwards from this refugium. Concordant patterns also observed for other cold adapted species such as *Pinus sylvestris* in West Scotland which is thought to have previously migrated from Ireland may also indicate the possibility of a north-west Atlantic refugium which give rise to the genetic patterns seen in populations in the west of Ireland and Scotland

today. Multiple colonisations and survival in an Irish refugium are also inferred for species including *Fucus serratus* (a brown seaweed) (Hoarau *et al.*, 2007), *Palmaria palmate* (a red seaweed), common frog (Teacher *et al.*, 2009),(Provan *et al.*, 2005) and brown trout (McKeown *et al.*, 2010).

Taxonomic distinctions are complex between *Arenaria ciliata* and *A. norvegica* and this may suggest a couple of factors: 1) *A. norvegica* may have evolved from *A. ciliata* and therefore remains monophyletic within this complex; 2) Insufficient markers were used to determine species distinctions as *A. norvegica* portrays low sequence divergence, indicative of recently evolved arctic flora; 3) Previous suture zones in which *A. ciliata* and *A. norvegica* co-occurred in a refugium which survived the LGM which resulted in chloroplast capture and hybridization of these two species.

4

Dating divergence times within the *Arenaria ciliata* complex using Bayesian molecular clock analysis

4.1 Introduction

4.1.1 Divergence of the Caryophyllaceae family

The results from Chapter 3 indicate that the *Arenaria ciliata* Irish population from Ben Bulben contains distinct haplotypes which belong to two relatively divergent lineages. It is possible to date the time of divergence between these and other lineages using a molecular clock analysis. By identifying the time of divergence between the key lineages, this will help to assess the biogeographic history of these populations and establish a timeline for their origin. The results from Chapter 3 also reveal a complex history between the populations of *A. ciliata* and *A. norvegica*. This chapter aims to further investigate past relationships of *A. ciliata* and *A. norvegica* populations by estimating times of divergence

between the major lineages identified in Chapter 3. A relatively new method of phylogenetic inference, Bayesian analysis, will be used to carry out molecular clock analyses to date the divergence of these key lineages.

Molecular clock analyses have indicated times of divergence for major clades within the Caryophyllace family, such as the subfamilies Alsinoideae and Caryophylloideae (Jordan and Macphail, 2003). The dates derived from these analyses are particularly well resolved because a fossil inflorescence dating from the middle-late Eocene, c. 34-45 million years ago, was used to calibrate and date the molecular clock analysis (Jordan and Macphail, 2003). This fossil was identified as a new genus and species (*Caryophylloflora paleogenica*), and is the oldest known record from the Caryophyllaceae family (Jordan and Macphail, 2003). Jordan and Macphail (2003) carried out a phylogenetic analysis using molecular and morphological data to determine the relationship of the fossil to extant lineages within this family.

The date estimates obtained (for the clades Alsinoideae and Caryophylloideae) using this fossil have subsequently been used in other molecular clock analyses to date lineages such as the genus *Dianthus* (Valente *et al.*, 2010). Using this date estimate, Valente *et al.* (2010) carried out a BEAST molecular clock analysis based on *matK* sequences using an uncorrelated lognormal relaxed clock model to infer a time for the divergence of the Caryophyllaceae genus *Dianthus*. The authors assigned an age constraint at the crown node of the Alsinoideae and Caryophylloideae using age extremes for this fossil at 34 and 45 million years before present (bp). Using these calibrations, the authors found that the age of the crown node for *Dianthus* was 1.9-7 million years bp.

The Caryophyllaceae fossil described by Jordan and Macphail (2003) was used in the present study to calibrate a molecular clock analysis of lineage divergence within the *A. ciliata* complex. The application of this fossil calibration is preferable to a more generic fossil calibration outside of the Caryophyllaceae, as a within-family calibration should lead to a more powerful molecular clock analysis with a smaller error range. A date estimate using this fossil can subsequently be used as calibration point for the determination of intraspecific lineages within the *Arenaria ciliata* complex.

4.1.2 Molecular clock analysis

The assumption behind molecular clocks is that genes evolve at a rate that is constant with time and across evolutionary lineages (Arbogast *et al.*, 2002). The idea of a molecular clock was first proposed in the 1960s by Zuckerkandl and Pauling in a study of haemoglobin evolution (Zuckerkandl and Pauling, 1965). The authors found that when using a molecular clock analysis, the timing for the origin of the major lineages of vertebrates correlated with paleontological estimates. If genetic divergence accumulates consistently over time, then time scales for the evolutionary distances between organisms may be developed in the absence of fossil data (Arbogast *et al.*, 2002).

Molecular clocks have been used to date key times of divergence for a wide range of organisms from the early Eukaryotes (Parfrey *et al.*, 2011) to kingdoms of organisms (Doolittle *et al.*, 1996; Wang *et al.*, 1999). Within kingdoms, they have been used to identify divergence times for mammals and birds (Hedges *et al.*, 1996) and plants (Sanderson *et al.*, 2004). Some concerns about the reliability of this method remain because of disparities which are sometimes apparent between molecular and paleontological divergence dates (Ho and Larson, 2006). The proliferation of sophisticated models implemented in molecular clock analysis, however, has led to an increased agreement between divergence dates estimated using molecular clocks and fossil data (Donoghue and Benton, 2007).

There are many methods available for molecular dating analysis that incorporate various assumptions on the rate of substitution across lineages. These include methods that implement a molecular clock using a global rate of substitution (e.g. maximum likelihood clock optimization Felsenstein (1981)), methods that correct for rate heterogeneity (e.g. local molecular clock (Hasegawa *et al.*, 1989)) and methods that incorporate substitution rate heterogeneity (Bayesian variable rate models in BEAST (Drummond and Rambaut, 2007)).

The program BEAST implements Bayesian inference and Markov chain Monte Carlo (MCMC) methods to co-estimate phylogeny and divergence times where the relaxed molecular clock model can be implemented (Drummond *et al.*, 2006) and therefore lineage rate heterogeneity can be incorporated across the tree. This program also enables the user to assign probabilistic prior distributions (such as

log-normal, normal, exponential, gamma or uniform distributions) to parameter values (Drummond and Rambaut, 2007). This allows flexibility in the incorporation of the calibration of fossil dates instead of relying on single estimates (Drummond *et al.*, 2006). BEAST can circumvent calculation difficulties that arise with molecular clock analysis, such as the problem of assigning single point calibrations, by allowing the user to assign a distribution to the calibration. This, in combination with the implementation of the relaxed clock model helps to improve the accuracy of the molecular clock analysis carried out for the present study.

4.1.3 Aims and Objectives

The primary aim of this chapter is to infer the timing of the key phylogeographic events within the evolution of the *Arenaria ciliata* complex. Using the fossil age constraints derived in Jordan and Macphail (2003) and the Bayesian parameters for dating analysis implemented in Valente *et al.* (2010), a molecular clock analysis for the *Arenaria ciliata* complex was carried out with the following objectives:

- To identify the most reliable estimate for the crown node age for *Arenaria ciliata* by testing the addition of five *A. ciliata* haplotypes versus the addition of a single haplotype to the *matK* sequence dataset implemented in Valente *et al.* (2010)
- Date the ancestral node for *Arenaria nevadensis* and *A. serpyllifolia* (the closest sister species to *A. ciliata*), by including all *A. ciliata* haplotype/s in the dataset of *matK* sequences implemented in Valente *et al.* (2010)
- Implement the calibration point estimated from the interspecific analysis of *matK* sequences to estimate times of divergence for the lineages within the *Arenaria ciliata* complex using (i) *matK*, (ii) concatenated chloroplast and (iii) Internal Transcribed Spacer sequences

These objectives were implemented to ultimately date: (1) divergence time separating the two divergent lineages separating the Ben Bulben *Arenaria ciliata* chloroplast haplotypes, (2) the time of divergence between *A. ciliata* and *A.*

norvegica chloroplast haplotypes and ITS nucleotypes and (3) the divergence time for the unique *A. ciliata* concatenated chloroplast haplotypes.

4.2 Results

4.2.1 Divergence times for *Arenaria nevadensis*, *A. serpyllifolia* and *A. ciliata* using *matK* sequences

Mean age estimates for the divergence of *Arenaria ciliata* with each outgroup, *A. nevadensis* and *A. serpyllifolia*, are presented in Table 4.1 for each calibrated date of 25, 34, 45 and 55 million years. Age estimates are shown for each dataset which included either a single *A. ciliata* haplotype or five *A. ciliata* haplotypes. Divergence times for *Dianthus* are also shown in Table 4.1 which were used as a positive control, to compare with the results from Valente *et al.* (2010). These dates were subsequently used to calibrate the intraspecific molecular clock analysis. The dates that were chosen to calibrate the intraspecific analysis for the four calibrated ages (25, 34, 45 and 55 my) for *A. nevadensis* are: 3.328, 4.582, 5.874 and 7.217 my; 5.425, 7.47, 9.621 and 11.808my for *A. serpyllifolia* and 2.725, 3.655, 4.675 and 5.81my for the *A. ciliata* haplotypes only.

Results outputted from TRACER showed that the ESS (effective sample size) values were all acceptable - most parameters indicated values much greater than 200 which is a reasonable sampling size (Drummond and Rambaut, 2007). The results obtained for both datasets (one haplotype of *Arenaria ciliata* versus five haplotypes) were very similar, however, when the age of the calibration was increased, the difference between the age estimate increased and larger age estimates were given for the dataset containing five haplotypes of *A. ciliata*. Therefore calibration dates from dataset (i) containing one haplotype of *Arenaria ciliata* were used for the most part, except for when no outgroups were included and the calibration dates were taken from dataset (ii), which contained five haplotypes of *A. ciliata*.

The age estimates given for *Dianthus* in this analysis were lower than those given in Valente *et al.* (2010). Age estimates were only slightly lower for the 34my

Table 4.1: Mean estimate for the time of divergence for the crown nodes obtained from BEAST and viewed in TRACER. Age is estimated in million years (my). Age estimates for *Dianthus* are also included from Valente *et al.* (2010)

Crown node	**25my**	**34my**	**45my**	**55my**
One AC haplotype - dataset (i)				
A.nevadensis	3.328	4.582	5.874	7.217
A.serpyllifolia	5.425	7.47	9.621	11.808
Dianthus	2.383	3.169	4.122	5.012
Five AC haplotypes - dataset (ii)				
A.nevadensis	3.734	4.983	6.38	7.929
A.serpyllifolia	5.938	7.951	10.141	12.62
Dianthus	2.359	3.153	4.079	4.911
A. ciliata	2.725	3.655	4.675	5.81
Estimates for *Dianthus* from Valente *et al.* (2010)				
95% HPD* estimates *Dianthus*	-	1.85-5.11	2.49-6.95	3.06-8.47
Mean age estimate *Dianthus*	-	3.47	5	6.27

*Highest posterior density

fossil calibration and the difference between ages estimated increase as the age of the fossil calibration is increased. The dataset which contains one individual of *Arenaria ciliata* (i), gave an age for the crown node of *Dianthus* which was closer to the mean age given in Valente *et al.* (2010) than the age given for *Dianthus* using the dataset containing five haplotypes of *A. ciliata* (ii). The difference in the ages given for each dataset however were negligible and the mean ages estimated for the crown node of *Dianthus* were consistent within the 95% highest posterior density (HPD) intervals given for the age of *Dianthus* in Valente *et al.* (2010) of 1.85-8.47my (Table 4.1).

The age estimates given for the divergence of *Arenaria ciliata* from *A. nevadensis* according to the dataset containing one individual of *A. ciliata* is 2.2-7.2my for the 95% HPD upper and lower limit for the fossil calibrated at 34my, and 2.7-9.1my for the fossil calibrated at 45my. The age estimates given for the time of divergence between *A. ciliata* and *A. serpyllifolia* were older where the fossil calibrated to 34my gave HPD 95% upper and lower intervals at 4.3-11.1my and

the fossil calibrated to 45my gave estimates of 5.5-14.3my.

A maximum clade credibility (MCC) tree constructed in BEAST which includes the *matK* sequences from the Caryophyllaceae family and the five haplotypes for *Arenaria ciliata* (dataset (ii)), where the fossil node was calibrated to 45 my, is shown in Figure 4.1. The *Arenaria ciliata* haplotypes are highlighted in green (*matK*19, *matK*17, *matK*1, *matK*14, *matK*9) and the nodes that were estimated are shown by the arrows - green: *A. ciliata* haplotypes; black: *A. nevadensis* and *A. ciliata*; blue: *A. ciliata*, *A. nevadensis* and *A. serpyllifolia*. The pink arrow indicates the node for *Dianthus* and the red is the crown node of the Caryophylloideae and Alsinoideae to which the fossil was calibrated. The topology of this tree is the same as the tree constructed with one individual of *Arenaria ciliata* (dataset (i)). The topology of the tree is also equivalent to the tree generated in Valente *et al.* (2010) and the placement of *Arenaria ciliata* is consistent with the maximum likelihood tree in section 3.4.2.

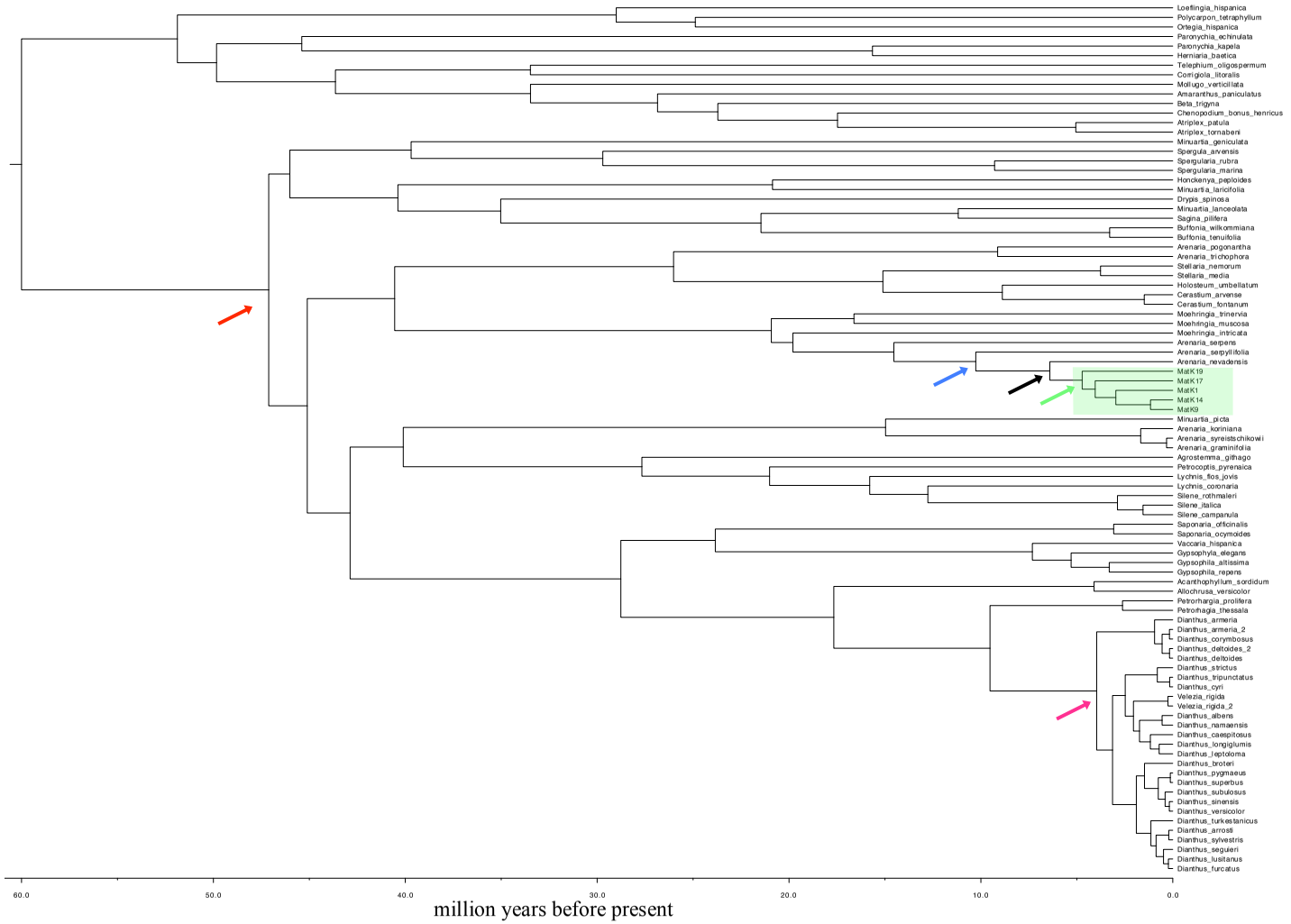


Figure 4.1: Maximum clade credibility tree generated in BEAST based on *matK* sequences for species in the Caryophyllaceae family which includes the five haplotypes of *Arenaria ciliata* (*matK*19, *matK*17, *matK*1, *matK*14, *matK*9 highlighted in green). The nodes for calibration and estimates of age divergence are indicated by the arrows: green includes the *A. ciliata* haplotypes; black includes *A. nevadensis* and *A. ciliata*; blue includes both these taxa and *A. serpyllifolia*. The pink arrow indicates the node for *Dianthus* and the red is the crown node of the Caryophylloideae and Alsinoideae to which the fossil was calibrated at 45my.

4.2.2 Summary of results for the *matK*, concatenated chloroplast and ITS sequences

The summary statistics estimated from the Bayesian analysis were viewed in TRACER and the mean age for each node, the standard error of the mean (stderr) and an upper and lower limit for the 95% highest posterior density (HPD) interval for each dataset are shown in Appendix E. These values were recorded for each of the four time points used to calibrate the intraspecific analysis, determined in section 4.2.1 from the fossil calibrations for 25, 34, 45 and 55 million years (Table 4.1). A summary of the mean ages estimated for the key Irish nodes of interest for the chloroplast *matK* and concatenated regions is shown in Table 4.2.

Maximum clade credibility (MCC) trees constructed in BEAST (using the calibration point of 34my) and assembled in TreeAnnotater are shown in Figures 4.2, 4.3 and 4.4 for the *matK* sequences, Figures 4.5 and 4.6 for the chloroplast concatenated sequences and Figures 4.7, 4.8 and 4.9 for the ITS sequences. The nodes which were dated are labelled with letters which correspond to the nodes labelled in Tables E.1, E.2 and E.3 for the *matK* sequences, Tables E.5 and E.4 for the chloroplast concatenated sequences and Tables E.6, E.7 and E.8 for the ITS sequences. A scale for the time in million years is shown from present. The location for each haplotype is shown in Table 4.3 for the *matK* sequences; Table 4.4 for the concatenated chloroplast sequences and Table 4.5 for the ITS sequences.

The most realistic dates to use in the intraspecific analyses are those derived from the fossil calibrated to 34my and 45my and these results will be referred to for the most part. The age estimates and summary statistics are also included for 25my and 55my as these will provide the lowest and highest theoretical ages that the fossil could be dated for. The mean estimated age and the upper and lower limits for the 95% highest posterior density (HPD) are reported here in million years (my) for the fossil calibrated to 34my and 45my. The age estimate used to calibrate the intraspecific analysis derived from the 34my fossil calibration will be referred to as the lower calibration limit (LCL) and the 45my fossil calibration as the upper calibration limit (UCL). The datasets which contain *Arenaria*

Table 4.2: Summary of the mean ages (my) estimated for the key Irish nodes of interest for the chloroplast *matK*, concatenated and ITS sequences for the four calibrated dates.

Node estimated and dataset	**34my**	**45my**
Separation of the two Irish BB haplotypes		
matK (node A) *A.nevadensis* (AN)	2.754	3.467
matK (node A) *A.serpyllifolia* (AS)	4.061	5.137
interspecific estimation (root calibration for matK AC)	3.655	4.675
concat (node G) *A.serpyllifolia* (AS)	2.051	2.59
concat (node G) *A.ciliata* only (AC)	1.343	1.712
Expansion of *A.norvegica*, IreBB, Rum		
matK (node F) *A.nevadensis* (AN)	0.655	0.82
matK (node F) *A.serpyllifolia* (AS)	0.976	1.23
matK (node D) *A.ciliata* only (AC)	0.917	1.153
concat (node F) *A.serpyllifolia* (AS)	0.696	0.875
concat (node F) *A.ciliata* only (AC)	0.446	0.572
ITS (node B) *A.nevadensis* (AN)	0.4603	0.5994
ITS (node D) *A.serpyllifolia* (AS)	0.6752	0.8878
ITS (node A) *A.ciliata* only (AC)	1.7849	2.2391
Expansion of IreBB from E.Alps, Picos, Pyr		
matK (node D) *A.nevadensis* (AN)	0.32	0.402
matK (node D) *A.serpyllifolia* (AS)	0.471	0.597
matK (node A) *A.ciliata only* (AC)	0.387	0.486
concat (node K) *A.serpyllifolia* (AS)	0.223	0.283
concat (node K) *A.ciliata* only (AC)	0.14	0.176

nevadensis are referred to as dataset AN, *A. serpyllifolia* as dataset AS and the dataset containing *A. ciliata* haplotypes only as dataset AC.

It is clear that as the age is increased for the calibration point, the age for the estimated nodes also increases (i.e. from the LCL at 34my to the UCL at 45my). When using different outgroups the age for the nodes varies where the age increases when using deeper divergence times for *A. serpyllifolia*.

The topologies of the Bayesian phylogenetic trees constructed for the chloroplast *matK* region were more similar to the concatenated chloroplast sequences, than to the nuclear ITS sequences, as observed in Chapter 3. The topologies slightly differed between the *matK* and concatenated chloroplast trees as the

Table 4.3: Location of each *matK* haplotype generated for the *Arenaria ciliata* complex (modified from section 3.4.2.2).

Haplotype	**Location**
MatK1	Ben Bulben, Ireland; East Alps; West Alps; Rum, Scotland; Svalbard
MatK2	Svalbard
MatK3	Iceland; Yorkshire
MatK4	Burren, Ireland; Pyrenees; Iceland; Scotland; England
MatK5	Picos De Europa
MatK6	Picos De Europa; Pyrenees
MatK7	Swiss Alps
MatK8	Svalbard
MatK9	Ben Bulben, Ireland; East Alps; Picos De Europa; Pyrenees
MatK10	Picos De Europa
MatK11	East Alps
MatK12	East Alps
MatK13	Pyrenees
MatK14	East Alps; Pyrenees
MatK15	East Alps
MatK16	West Alps
MatK17	West Alps
MatK18	West Alps
MatK19	West Alps, East Alps
MatK20	Kildare, Ireland

matK haplotypes were not as variable as the concatenated chloroplast haplotypes and did not infer any unique Irish *Arenaria ciliata* haplotypes.

The topologies of the ITS Bayesian phylogenetic trees varied from the chloroplast sequence trees where there was clear geographic grouping of the nucleotypes. Consequently there was no division of the *Arenaria ciliata* Ben Bulben haplotypes into two divergent lineages, therefore this node could not be dated for the ITS analyses. There is a clear division, however, between the *Arenaria ciliata* and *A. norvegica* nucleotypes for the ITS sequences and this could be dated and compared with the chloroplast sequences dates for this node.

The overall topology of the trees constructed for the *matK* sequences in each dataset (AN, AS, AC) was the same where four major lineages were identified: (1)

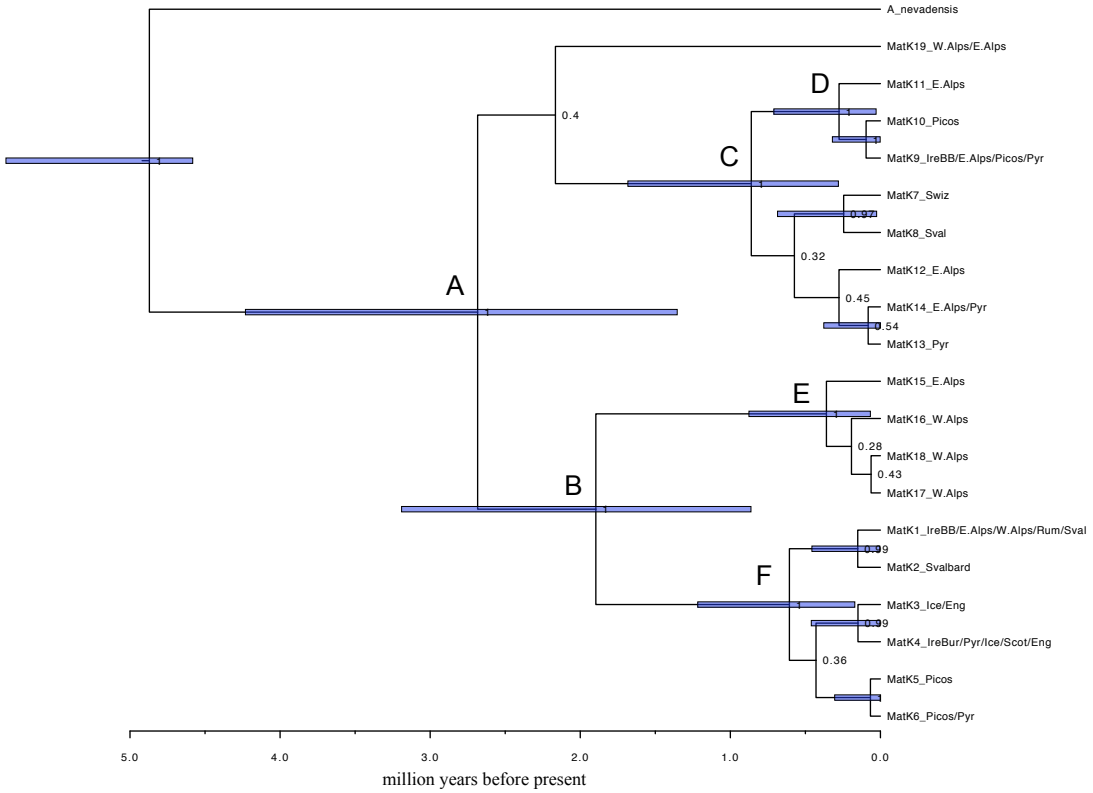


Figure 4.2: Maximum clade credibility (MCC) trees of the *matK* haplotypes for the *Arenaria ciliata* complex with *A. nevadensis* constructed in BEAST and viewed in FigTree. Node bars represent the 95% highest posterior density (HPD) intervals for the divergence time estimates of clades with posterior probabilities above 0.5. Numbers at the nodes indicate the Bayesian posterior probabilities. The time scale shows age in million years from present. Letters represent the nodes which were dated.

West Alps (2) a widely distributed group including haplotypes from: East Alps, Picos De Europa, Ben Bulben Ireland, Pyrenees, Switzerland and Svalbard (3) East and West Alps and (4) another widely distributed group which contains the other Irish haplotype from Ben Bulben, East Alps, West Alps, Rum, Svalbard, Iceland, England, Pyrenees and Picos De Europa (Figures 4.2, 4.3 and 4.4). There is poor resolution (according to the low posterior probability of 0.4) for the node assigning the placement of *matK19* from the West and East Alps, which appears to be the most divergent haplotype.

The topology of the concatenated chloroplast tree was exactly the same using both datasets. There were four major lineages identified from: 1. the West Alps, 2. West and East Alps, 3. a group which contained widespread distributions

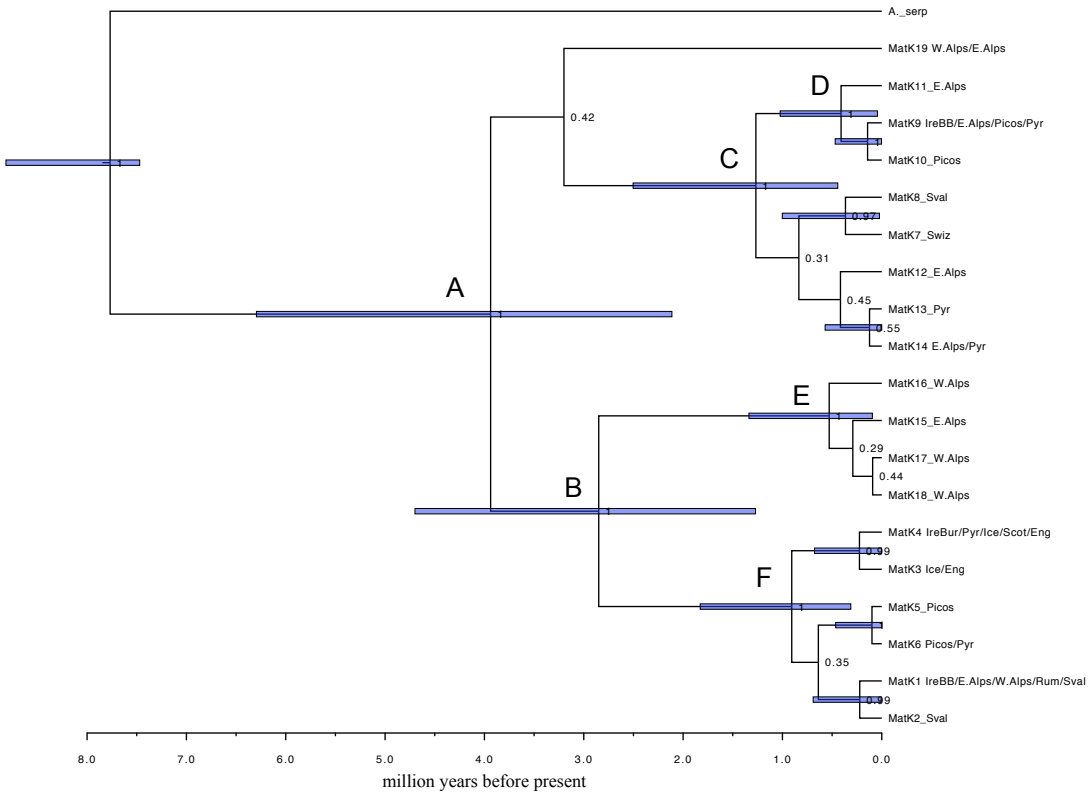


Figure 4.3: Maximum clade credibility (MCC) trees of the *matK* haplotypes for the *Arenaria ciliata* complex with *A. serpyllifolia* constructed in BEAST and viewed in FigTree. Node bars represent the 95% highest posterior density (HPD) intervals for the divergence time estimates of clades with posterior probabilities above 0.5. Numbers at the nodes indicate the Bayesian posterior probabilities. The time scale shows age in million years from present. Letters represent the nodes which were dated

from Ben Bulben Ireland, Scotland, Iceland, England and Picos De Europa and 4. another group which included the other haplotype from Ben Bulben and the Pyrenees, E. Alps and Picos De Europa.

The topology of the three trees constructed for the ITS sequences slightly deviate because of the low resolution in determining the placement of the Alpine sequences. There were five main lineages identified in all three trees which for the most part correspond to the geographic origins of the ITS sequences: 1. West Alps, 2. East Alps, 3. Spain (Picos De Europa and Pyrenees), 4. Ireland Ben Bulben, 5. Ireland Burren, Scotland, Iceland, England. Lineage 5 is the only lineage which contains ITS sequences from multiple locations. According to dataset AN and AC the ancestral ITS group is identified as the West Alps,

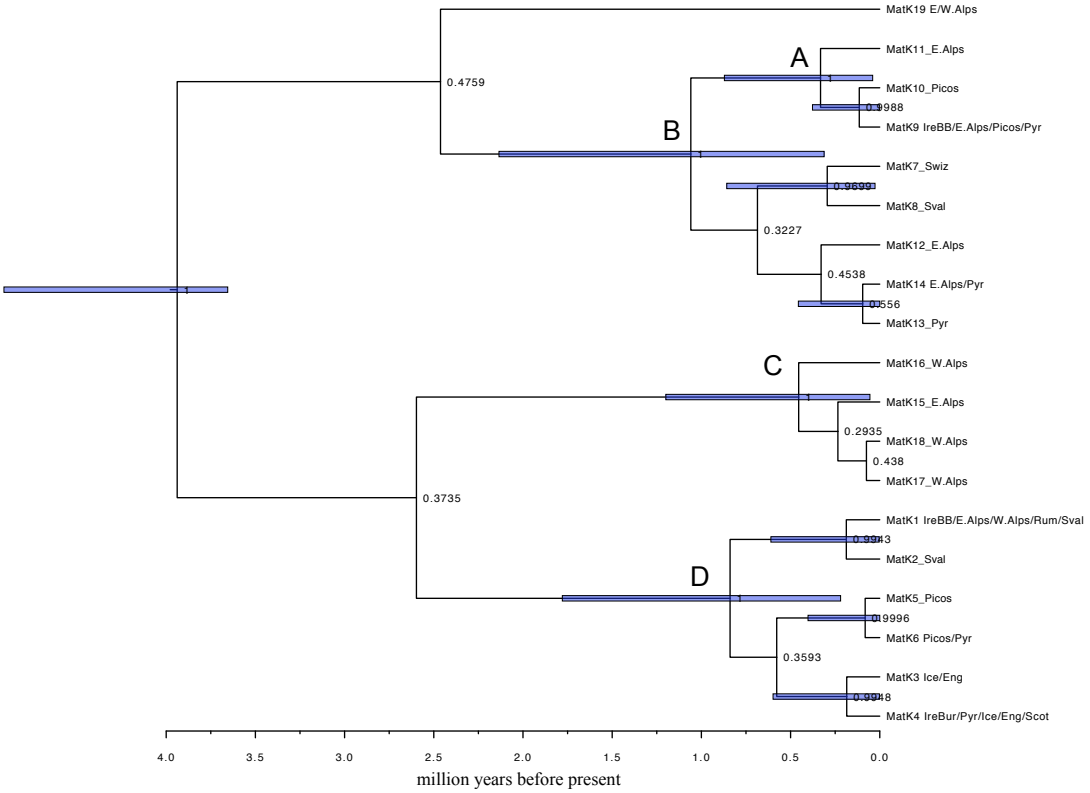


Figure 4.4: Maximum clade credibility (MCC) trees of the *matK* haplotypes for the *Arenaria ciliata* complex constructed in BEAST and viewed in FigTree. Node bars represent the 95% highest posterior density (HPD) intervals for the divergence time estimates of clades with posterior probabilities above 0.5. Numbers at the nodes indicate the Bayesian posterior probabilities. The time scale shows age in million years from present. Letters represent the nodes which were dated.

however according to dataset AS the ancestral ITS group is identified as Spain. The poor resolution in the placement of the East Alps could be a likely cause in the discrepancies observed between the trees. The East Alps haplotypes were paraphyletic in the chloroplast regions, and one of the haplotypes for the East Alps was placed within the West Alps group in the concatenated dataset and the *matK* dataset. This placement is also conflicted with lower posterior probability values in both chloroplast datasets using all possible outgroups.

The results from all of the dated nodes for each of the chloroplast and nuclear sequences are shown in Appendix E. Emphasis will be placed here on the dates for the key nodes of interest for this study. These are the nodes showing the division of the Irish Ben Bulben *Arenaria ciliata* haplotypes for the chloroplast sequences, the divergence of the unique concatenated chloroplast haplotypes from

Table 4.4: Location of each concatenated chloroplast haplotype generated for the *Arenaria ciliata* complex (modified from chapter 3).

Haplotype	Location
con1 IreBB1	Ireland, Ben Bulben
con2 IreBB2	Ireland, Ben Bulben
con3 Picos1	Picos De Europa, Spain
con4 Picos2	Picos De Europa, Spain
con5 IreBB3	Ireland, Ben Bulben
con6 Pyr	Pyrenees, Spain
con7 Aus1	East Alps, Austria
con8 Aus2	East Alps, Austria
con9 Picos3	Picos De Europa, Spain
con10 Picos4	Picos De Europa, Spain
con11 Aus3	East Alps, Austria
con12 W.Alps1	West Alps, France
con13 W.Alps2	West Alps, Italy
con14 W.Alps3	West Alps, France
con15.W.Alps4	West Alps, Italy
con16 W.Alps5	West Alps, Italy
con17 W.Alps6	West Alps, Italy
con18 Aus4	East Alps, Austria
con19 ScotRum	Rum, Scotland
con20 IreBur/Eng/Scot/Shet	Ireland, England, Scotland
con21 Ice/IreBur	Iceland, Ireland
Arenaria serpyllifolia	Kildare, Ireland

their most recent common ancestor, and the division between *A. ciliata* and *A. norvegica*.

4.2.3 Time of divergence for the separation of the *Arenaria ciliata* Ben Bulben chloroplast haplotypes

The node which represents the division of the Ben Bulben *Arenaria ciliata* haplotypes is depicted as node A in the *matK* trees containing the outgroups *A. nevadensis* and *A. serpyllifolia* (Figures 4.2 and 4.3) and as the calibrated node

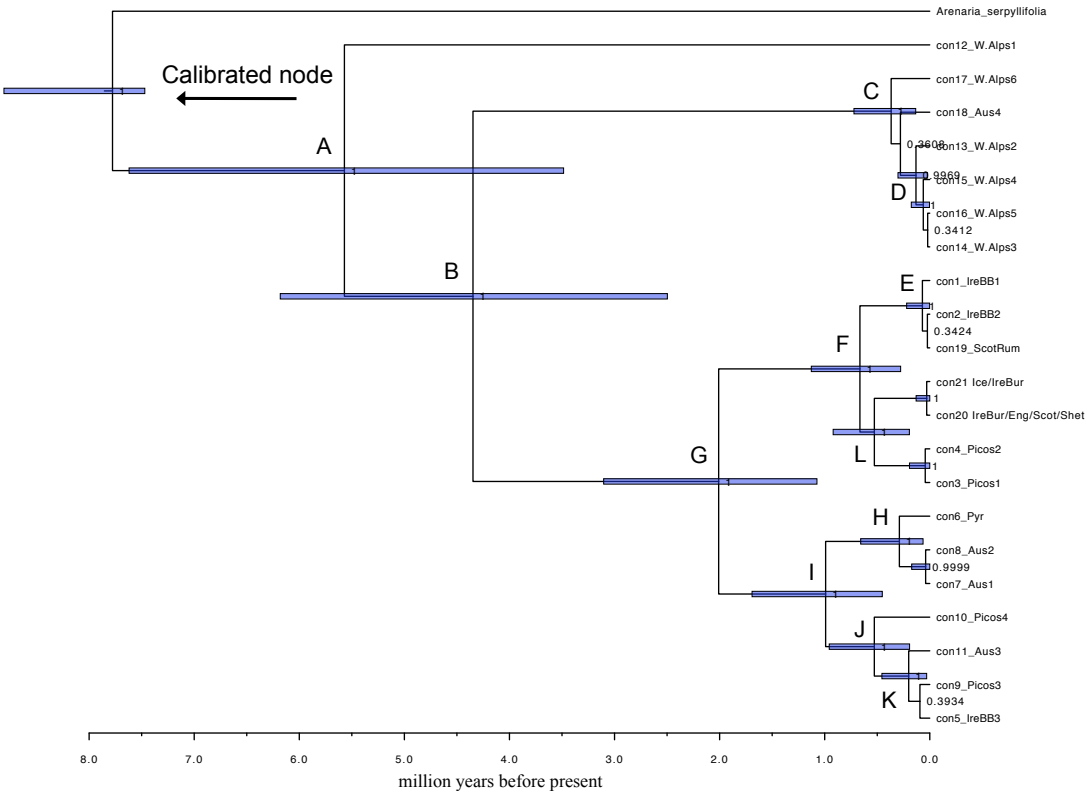


Figure 4.5: Maximum clade credibility (MCC) trees of the concatenated chloroplast haplotypes for the *Arenaria ciliata* complex with *A. serpyllifolia* constructed with BEAST. Numbers at the nodes are Bayesian posterior probabilities. Letters represent the groupings of those clades which have posterior probabilities above 0.5 which are to be estimated for the various age estimates given for the calibrated node.

in the tree containing no outgroups (Figure 4.4). In the concatenated trees, this node is represented as node G for both trees containing *A. serpyllifolia* (Figure 4.5) and the tree with no outgroups (Figure 4.6). Although the topology of the trees for the *matK* and concatenated chloroplast sequences are slightly different, both convey considerably old times of divergence for this node, and this is evident in the phylogenetic trees, where in some cases this split is the oldest divergence for any of the haplotypes (Figures 4.2 and 4.3).

According to the *matK* analyses, the separation of the *Arenaria ciliata* Ben Bulben haplotypes into two divergent lineages occurs in the oldest split of all the lineages, in each of the phylogenetic analyses including *A. nevadensis* (Figure 4.2 - node A), *A. serpyllifolia* (Figure 4.3 - node A) and no outgroups (Figure 4.4 - calibrated node). The times of divergence for this node are summarized in Table

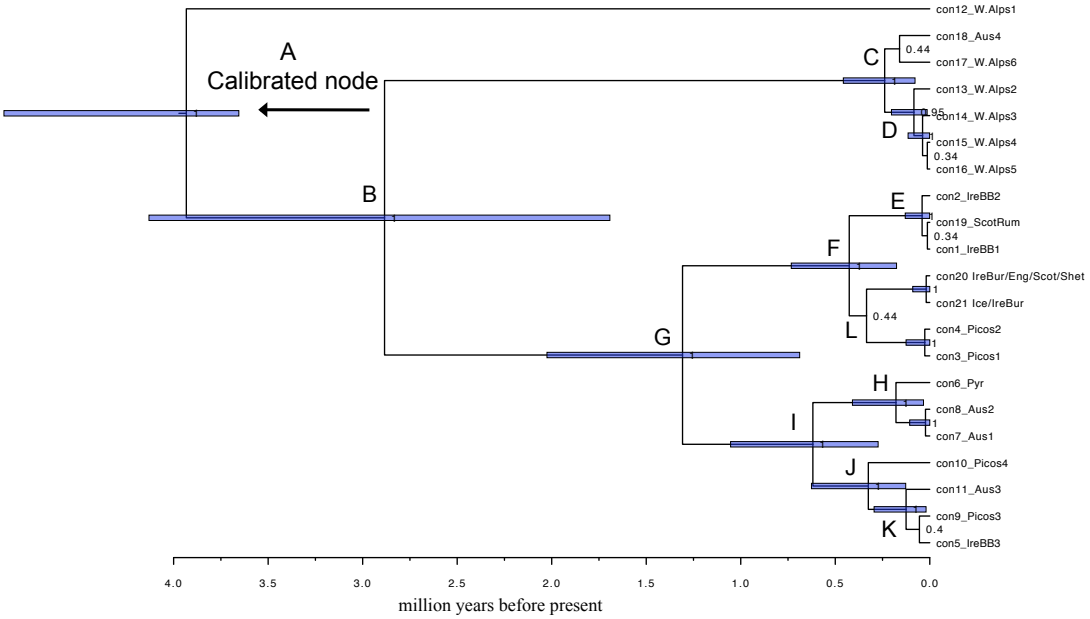


Figure 4.6: Maximum clade credibility (MCC) trees of the concatenated chloroplast haplotypes for the *Arenaria ciliata* complex with no outgroups constructed with BEAST. Numbers at the nodes are Bayesian posterior probabilities. Letters represent the groupings of those clades which have posterior probabilities above 0.5 which are to be estimated for the various age estimates given for the calibrated node.

4.2 for the calibrated dates of 34 and 45 my. The dates estimated are all within a reasonable range for all of the datasets, even though they were calculated using different outgroups and sequence data. The minimum mean time of divergence for this node (for the two calibration points at 34 and 45 my) is 1.343 my (for the concatenated AC only dataset) and the oldest date estimated for this node was as old as 5.137my (for the *matK* AS dataset).

According to the *matK* dataset containing *Arenaria nevadensis* (dataset - AN), the mean time of divergence for this node is 2.754my (95% HPD: 1.35 - 4.2my) for the LCL and 3.467my (95% HPD: 1.7-5.4my) for the UCL. The concatenated chloroplast sequences revealed younger ages than the *matK* sequences where expansion of this node occurred 2.051my (95% HPD: 1.1-3.1my) according to dataset AS for the LCL and 2.59my (95% HPD: 1.41-3.9my) for the UCL.

Table 4.5: Names for individuals for ITS for *Arenaria ciliata* and *A. norvegica*

Haplotype	**Location**
Aus1	Austria
Aus2	Austria
Aus3	Austria
Aus4	Austria
EngYork	Yorkshire, England
Ice	Iceland
IreBB1	Ben Bulben, Ireland
IreBB2	Ben Bulben, Ireland
IreBB3	Ben Bulben, Ireland
IreBB4	Ben Bulben, Ireland
IreBB5	Ben Bulben, Ireland
IreBB6	Ben Bulben, Ireland
IreBB7	Ben Bulben, Ireland
IreBur1	Burren, Ireland
IreBur2	Burren, Ireland
Picos1	Picos De Europa, Spain
Picos2	Picos De Europa, Spain
Picos3	Picos De Europa, Spain
Picos4	Picos De Europa, Spain
Picos5	Picos De Europa, Spain
Picos6	Picos De Europa, Spain
Pyr1	Pyrenees, Spain
Pyr2	Pyrenees, Spain
ScotInch	Inchnadamph, Scotland
ScotRum1	Rum, Scotland
ScotRum2	Rum, Scotland
ScotShet	Shetland, Scotland
W.Alps1	West Alps, Italy
W.Alps2	West Alps, France
W.Alps3	West Alps, Italy
W.Alps4	West Alps, Italy
W.Alps5	West Alps, Italy
W.Alps6	West Alps, France
Arenaria serpyllifolia	Kildare, Ireland

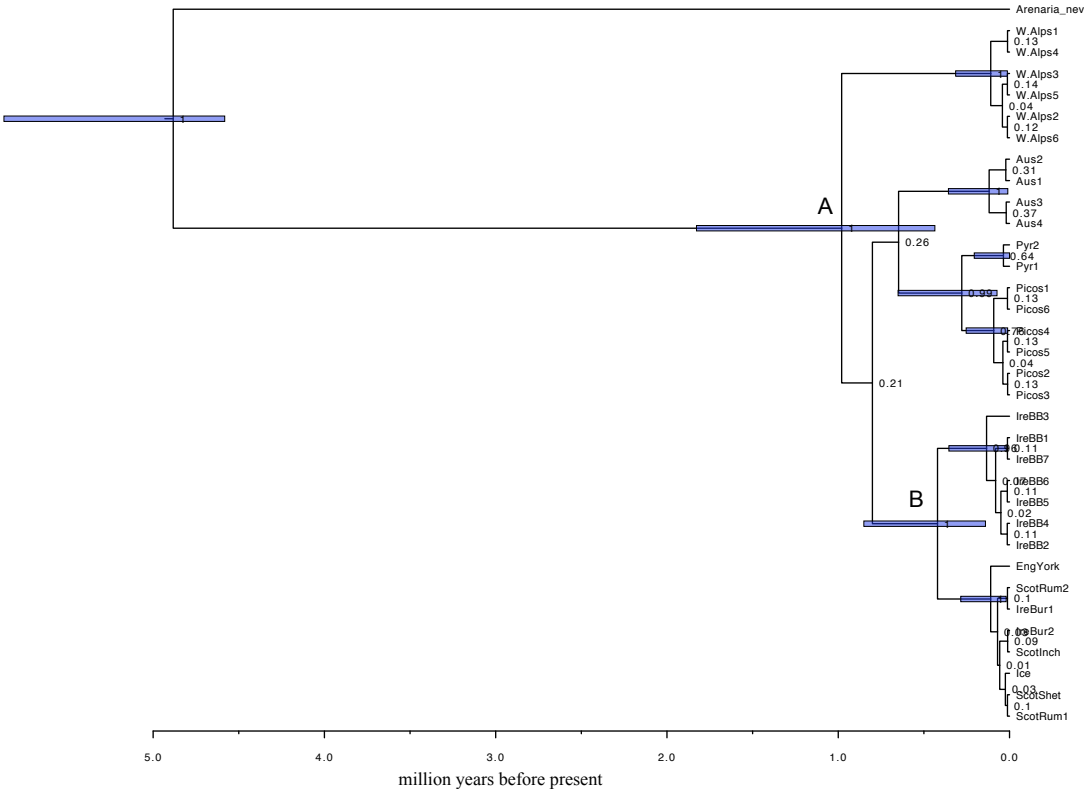


Figure 4.7: Maximum clade credibility (MCC) tree of the ITS sequences for the *Arenaria ciliata* complex with *A. nevadensis* constructed with BEAST. Numbers at the nodes are Bayesian posterior probabilities. Letters represent the groupings of those clades which have posterior probabilities above 0.5 which are to be estimated for the various age estimates given for the calibrated node.

According to dataset AC this expansion occurred 1.343my (95% HPD: 0.69-2my) for the LCL and 1.712my (95% HPD: 0.9-2.5my) for the UCL. In the dataset containing *Arenaria serpyllifolia* for the *matK* sequences (dataset - AS), the estimated ages are 4.061my (95% HPD: 2.1-6.3mya) for the LCL, and 5.137my (95% HPD: 2.6-7.98my) for the UCL. The calibration time implemented in the dataset containing *A. ciliata* haplotypes only (dataset - AC) which corresponds to this node which was estimated in the interspecific analysis are 3.655my (95% HPD: 2-5.6my) for the LCL and 4.675my (95% HPD: 2.5-6.9my) for the UCL.

Although there is variability in the ages estimated for this node, depending on the outgroup and type of sequence data analysed, a common consensus is found in the analyses that demonstrates an ancient split which predates the LGM by millions of years.

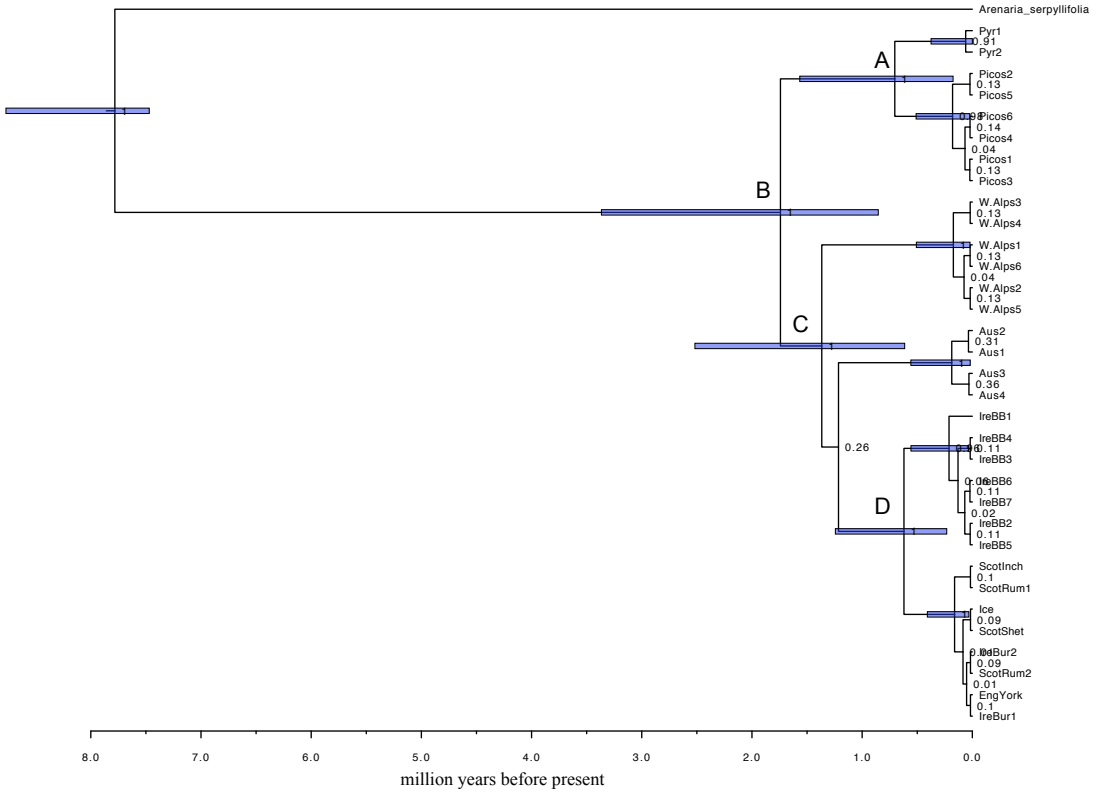


Figure 4.8: Maximum clade credibility (MCC) trees of the ITS sequences for the *Arenaria ciliata* complex with *A. serpyllifolia* constructed with BEAST. Numbers at the nodes are Bayesian posterior probabilities. Letters represent the groupings of those clades which have posterior probabilities above 0.5 which are to be estimated for the various age estimates given for the calibrated node.

4.2.4 Time of divergence for the division *Arenaria ciliata* and *A. norvegica*

The node which separates the haplotypes and nucleotypes of *Arenaria ciliata* and *A. norvegica* is represented as node F in the *matK* phylogenetic trees for the trees including *A. nevadensis* (Figure 4.2) and *A. serpyllifolia* (Figure 4.3) and node D in the tree with no outgroups (Figure 4.4). In the concatenated phylogenetic tree this node is represented as node F for all trees (Figures 4.5 and 4.6). In the ITS trees the node is designated node B for the tree including *Arenaria nevadensis* as an outgroup (Figure 4.7), node D for the tree containing *A. serpyllifolia* as an outgroup (4.8) and node A for the tree with no outgroups (Figure 4.9).

In Table 4.2 a summary of the ages for this node is shown for the concatenated, *matK* and ITS analyses. The youngest mean age estimated for this node for the

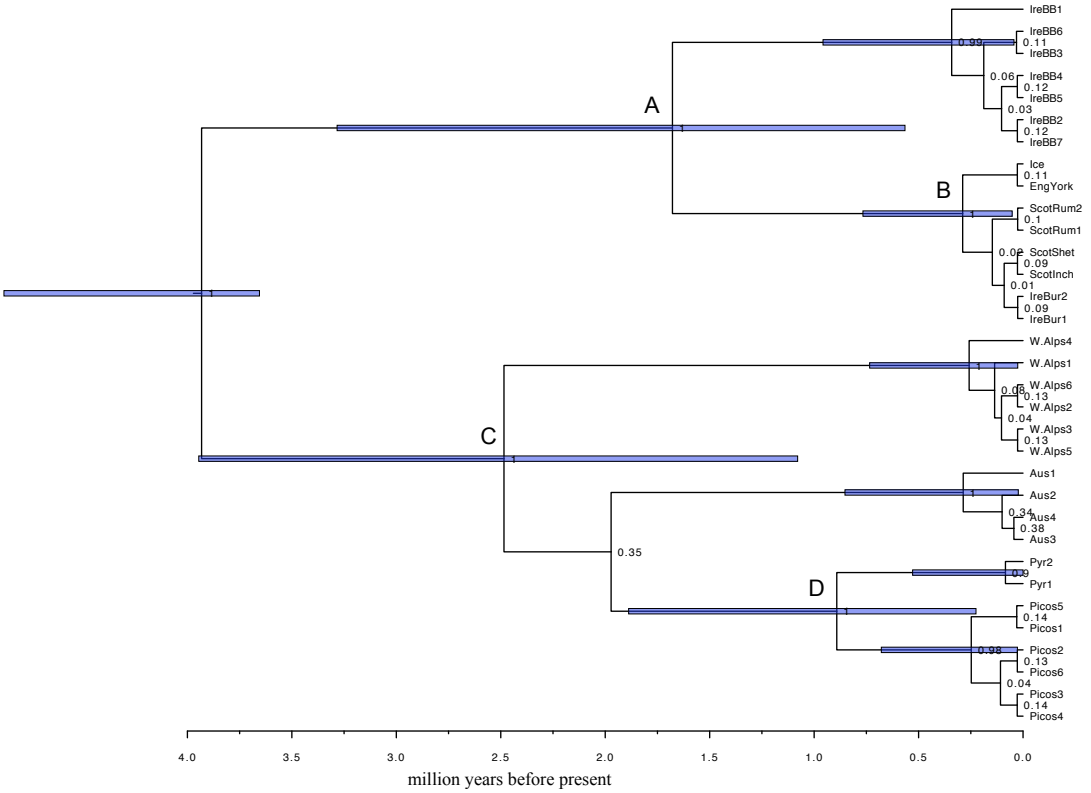


Figure 4.9: Maximum clade credibility (MCC) trees of the ITS sequences for the *Arenaria ciliata* complex with no outgroup constructed with BEAST. Numbers at the nodes are Bayesian posterior probabilities. Letters represent the groupings of those clades which have posterior probabilities above 0.5 which are to be estimated for the various age estimates given for the calibrated node.

two calibrated dates (34 and 45my) is 0.446my for the concatenated chloroplast sequences with no outgroup and the oldest age was estimated for the ITS 45my calibration for the dataset containing no outgroups. The dates estimated using the chloroplast and ITS regions are consistent with each other and most estimated dates are less than 1 million years old, but great than 500,000 years old. These dates are younger than the lineage separating the *Arenaria ciliata* Ben Bulben haplotypes, however they also occur much earlier than the LGM.

The *matK* analyses depicts an age of divergence for the node separating *Arenaria ciliata* and *A. norvegica* at 0.655my (95% HPD: 0.17-1.2my) for the LCL and 0.82my (95% HPD: 0.2-1.5my) for the UCL for dataset AN; dataset AS: 0.976my (95% HPD: 0.31-1.8my) (LCL), 1.23my (95% HPD: 0.4-2.4my) (UCL) and AC: 0.917my (95% HPD: 0.2-1.8my) (LCL), 1.153my (95% HPD: 0.27-2.2) (UCL). This node also includes the time of divergence of the haplotypes from

the Burren Ireland, Iceland, England and the Pyrenees from its MRCA in Spain (Picos De Europa and Pyrenees). Although the posterior probability value is less than 0.5 the MCC trees indicates approximate mean ages for this node at 0.4-0.5my for dataset AN, 0.6-0.7my for dataset AS and 0.5-0.6my for dataset AC using the LCL.

The molecular clock analyses using the concatenated sequences again depicts younger ages than the *matK* region. Dates assigned for the expansion of haplotypes from this node for dataset AS: 0.696my (95% HPD: 0.28-1.13my) for the LCL and 0.875my (95% HPD: 0.38-1.4my) for the UCL for dataset AS. The dates assigned for dataset AC: 0.446my (95% HPD: 0.18-0.7my) for the LCL and 0.572my (95% HPD: 0.2-0.95my) for the UCL. The minimum possible mean age from the fossil calibrated at 25my is 0.512my for dataset AS and 0.338my for dataset AC. These dates also show the MRCA for node L which contain haplotypes from the Burren, England, Scotland, Shetland, Iceland and Picos De Europa.

Although the ITS sequences do not depict as much variability as the chloroplast regions, there is a considerable amount of time estimated for the time of divergence for the *Arenaria ciliata* and *A. norvegica* nucleotypes. The time of expansion according to the AN dataset (node A) is 1.05my (95% HPD: 0.44-1.8my) for the LCL and 1.4my (95% HPD: 0.59-2.5my) for the UCL. According to dataset AS (node B) the time of expansion was 1.89my (95% HPD: 0.8-3.4my) for the LCL and 2.5my (95% HPD: 1.1-4.6my) for the UCL.

4.2.5 Time of divergence for the unique Irish *Arenaria ciliata* haplotypes

The unique Irish *Arenaria ciliata* haplotypes identified from the concatenated chloroplast sequences, occur with Rum in node E and with Picos de Europa in node K (Figures 4.5 and 4.6). The time of expansion of the Irish *Arenaria ciliata* haplotypes with the Rum *A. norvegica* haplotypes occurred at a mean age of 0.087my for the LCL and 0.113my for the UCL for dataset AN and a mean age of 0.051my for the LCL to 0.065 my for the UCL for dataset AC.

The other Irish *Arenaria ciliata* haplotype is paired with Picos De Europa and according to the MCC trees this expansion occurred at least 30k years ago. The exact ages could not be estimated as the posterior probability was less than 0.5 however an estimate can be established using the MCC trees for both datasets. The time of expansion for this group which also includes East Alps (node K) contained a high posterior value and the estimates for this expansion occurred 0.223my (95% HPD: 0.03-0.46my) (LCL dataset AS) and 0.283my (0.047-0.598my) (UCL dataset AS). For dataset AC the LCL is 0.14my (95% HPD: 0.02-0.295my) and the UCL is 0.176my (95% HPD: 0.023-0.37my).

4.3 Discussion

4.3.1 Divergence of the *Arenaria ciliata* chloroplast haplotypes from Ben Bulben

The *matK* and concatenated chloroplast molecular analyses demonstrate that one of the earliest expansions of the *Arenaria ciliata* complex haplotypes includes a deeply divergent lineage which separates the two Irish Ben Bulben haplotypes. This separation is not evident for the ITS nucleotypes, however, as all of the Irish *A. ciliata* nucleotypes are grouped together in the same lineage. The *matK* and concatenated chloroplast molecular clock analyses demonstrate, however, that the expansion of the *A. ciliata* haplotypes for this node could be as old as 5.137 million years (*matK* analyses including *A. serpyllifolia* as an outgroup). Some of the dates estimated from the dataset containing *Arenaria serpyllifolia* without an outgroup, however, are very old and more accurate estimates might be generated using the dates approximated from the interspecific analysis from the closest sister relative, *A. nevadensis*. The dates generated from this dataset are still considerably old even when set to the lowest calibration at 25my, where the mean time of divergence is 2.085 my. It is clear that the dates estimated for this lineage predate the LGM and that the time of divergence for these haplotypes is in the range of millions of years.

4.3.2 The separation of *Arenaria ciliata* and *A. norvegica*

A key node in understanding the biogeographic history of the *Arenaria ciliata* complex (particularly the historical relationships between *A. ciliata* and *A. norvegica*), is the node which shows the expansion of the *A. norvegica* haplotypes and the separation into two clades. One of these clades is specific to Rum and is closely related to Ben Bulben and another which is widely distributed in Scotland, England, Iceland, Pyrenees and the Burren, Ireland. The pattern is observed in both the *matK* and concatenated chloroplast haplotypes. Again this is in contrast to the pattern depicted for the ITS region which shows the expansion of *Arenaria norvegica* expanding from a node which also groups the Irish Ben Bulben haplotypes.

Both the *matK* and concatenated chloroplast analyses suggest the expansion of *Arenaria norvegica* within the Pleistocene less than 1 million years ago with a likely expansion at a minimum age for the LCL of 446k years for the concatenated chloroplast region and a maximum age for the UCL at 1.23my for *matK* (Table 4.2). The origin of the *Arenaria norvegica* haplotypes show a more recent history than some of the more divergent *A. ciliata* lineages which is typical of Arctic species which undergoes a more recent expansion (Hewitt, 2004). The majority of the lineages in the *A. ciliata* complex also expanded within the last 2 million years, so it is evident that the full extent of the Pleistocene period played a major role in shaping the genetics of these populations, and not just the LGM.

The *Arenaria norvegica* haplotypes, for the most part (except for the divergence of the Rum haplotype), reflect what is typically seen in the phylogeography of Arctic species; a shallow distinct clade with common geographical boundaries (Hewitt, 2004). The pattern for the Irish haplotypes (and the Rum *A. norvegica* haplotype) is much more divergent and its old divergence times are indicative of the survival of these haplotypes over many glacial periods (Hewitt, 2004). The ages estimated from this analysis for the *A. ciliata* lineages are comparable to the divergence of species such as the radiation of *Dianthus* in the last 1.9-7 million years (Valente *et al.*, 2010). There are many phylogeographic studies which also demonstrate very old divergence times for lineages within populations which

expanded long before the last glacial maximum (LGM), (Paulo *et al.*, 2001; Fernandez Mazuecos and Vargas, 2012; Hoarau *et al.*, 2007).

Estimated ages vary considerably depending on whether chloroplast or nuclear ITS sequences are used: *matK* gives the oldest ages and the nuclear ITS region the most recent ages. This is also evident in the short branch lengths for the ITS region compared to the chloroplast regions in the maximum likelihood phylogenetic trees in Chapter 3. The more recent dates estimated and strong geographic clustering of populations for the nuclear ITS region could reflect more recent gene flow and homogenization events, compared to the older biogeographic histories depicted for the chloroplast haplotypes. However, this is not the case for all of the lineages. In the case of the divergence of *Arenaria ciliata* and *A. norvegica* nucleotypes, the time frame in which this expansion occurs is similar to that recorded from the concatenated region. Ages estimated for the ITS nucleotypes range (for the two calibrations at 34 and 45my) from 1.05 my to 2.5 my. These date estimates are very old considering ITS displays much less variation than the chloroplast regions. This could reflect restricted gene flow between *Arenaria ciliata* and *A. norvegica* which results in a more distinct clustering of geographic groups. It is noteworthy that the Irish Ben Bulben *Arenaria ciliata* lineage is grouped with all of the *A. norvegica* nucleotypes rather than the other continental *A. ciliata* nucleotypes. This could be a result again of the nuclear nature of the ITS region which shows gene flow occurring between populations which are situated more closely to each other.

4.3.3 Divergence of the unique Irish *Arenaria ciliata* haplotypes

The expansion of the *Arenaria ciliata* haplotypes from Ben Bulben from the Rum (West Scotland) *A. norvegica* haplotypes (node E, Figures 4.5 and 4.6) is a key biogeographic event as this shows the last time that these haplotypes co-existed and subsequently diverged from each other. Compared to the separation of the more widely distributed *Arenaria norvegica* haplotypes from *A. ciliata*, this divergence occurred more recently. Estimated dates for this expansion range from

0.051my to 0.113my. It is clear from the MCC trees in particular that these haplotypes expanded from each other approximately 50,000 - 100, 000 years ago and remained in isolation with each other for a considerable period of time. These Ben Bulben *A. ciliata* and Rum *A. norvegica* haplotypes remained in isolation from other *Arenaria ciliata* and *A. norvegica* haplotypes for approximately 400,000 - 800,000 years (node F).

The node which shows the expansion of the other *Arenaria ciliata* Ben Bulben haplotype from its MRCA in Picos de Europa contains a posterior value less than 0.5, so the 95% highest density intervals were not calculated for this clade. However, approximate mean ages calculated from the MCC trees show this expansion at approximately 200k years for both sets of analyses. Although exact ages cannot be placed on the times of divergence for each of the unique Irish *Arenaria ciliata* haplotypes, it is clear that they expanded at different times, where the haplotypes that co-existed with the *Arenaria norvegica* haplotypes from Rum (node E) remained in isolation from other *A. ciliata* haplotypes for a much longer period.

The application of molecular clock analyses for populations of a range of species has shown the possibility for the persistence of populations for much longer than the LGM and likely survival throughout multiple glaciations dating back to the mid Pleistocene during the Wurm (120,000-18,000 BP) and Riss glaciations / (200,000-135,000 BP)) (Fernandez-Mazuecos and Vargas, 2011; Hoarau *et al.*, 2007), even as far back as the Pliocene for some organisms such as the Iberian lizard Paulo *et al.* (2001). Demographic expansions of populations for marine organisms in particular appear to have occurred much earlier than the LGM. Expansions of some populations date back to 580,000 and 362,000 years ago for thornback rays (*Raja clavata* L.) and 536,000 years BP for the common goby (*Pomatoschistus microps* (Gysels *et al.*, 2004).

These dates align with one of the longest interglacials in the Quaternary, the Holstenian, and the Cromerian complex which was a succession of long glacial periods followed by short interglacials (Prokopenko *et al.*, 2002; Svendsen *et al.*, 2004). They are comparable with the ages estimated for the expansion of the *Arenaria norvegica* haplotypes and the Spanish populations. Even older estimates are obtained for the expansion of the Rum (Scotland) and Irish Ben Bulben

haplotypes occurring at least 450k years ago and the highest estimate given for the LCL is 1.153my for the *matK* region for dataset AC.

Molecular clock analysis has also demonstrated pre LGM expansion for Irish populations of some marine taxa such as the brown seaweed, *Fucus serratus*, where the largest expansion is thought to have occurred approximately 128k-67k years ago, during the Eem interglacial, much earlier than the LGM (Hoarau *et al.*, 2007). Evidence for Pleistocene refugia off the coast of Ireland has also been generated by dating analysis such as for the common goby (*Pomatoschistus microps* (Gysels *et al.*, 2004), where the estimated time of expansion for the Atlantic haplotypes is 536 ky BP, which may correspond to an interglacial during the Cromerian complex (Prokopenko *et al.*, 2002). Gysels *et al.* (2004) also suggest the isolation of the Irish population from the other populations of the common goby for some time on the basis of a frequent unique Irish haplotype, and the origin for the Irish and West Scottish populations is suggested as originating from a refugium in the North Sea.

The time of expansion for the other Irish Ben Bulben *Arenaria ciliata* haplotypes is more recent than for the haplotype associated with Rum and they show a different biogeographic history. The expansion of the lineage containing haplotypes from Ben Bulben, the East Alps and Spain is estimated to have occurred at a minimum age for the LCL at 140k years for the concatenated region, and a maximum age for the UCL at 597k years for *matK*. These haplotypes share common ancestors in Picos De Europa, the Pyrenees and the East Alps. This clade, which diverged from Picos and the East Alps, also expanded from other locations in Europe from Svalbard and the Swiss Alps. The concatenated chloroplast region also includes haplotypes from these locations, except from the Swiss Alps and Svalbard as these locations were not available for inclusion in the concatenated dataset.

4.3.4 Dates estimated for *Dianthus*

There were slight discrepancies observed from the interspecific analysis between the ages estimated for the crown node of *Dianthus* in this analysis and that estimated in Valente *et al.* (2010) (i.e. the estimated age of *Dianthus* increased when

the age of the fossil calibration was increased). These discrepancies could suggest that there is greater error associated with large age calibrations. The estimated age for the crown node of *Dianthus* in this study is still consistent with the 95% HPD interval given in Valente *et al.* (2010), however. The time of divergence of *Arenaria ciliata* is approximately equivalent to the time of divergence of the genus *Dianthus* and *A. ciliata* probably diverged from *A. nevadensis* before the Pleistocene, possibly during the Pliocene or the Miocene.

4.3.5 Topology and structure of the maximum clade credibility trees

The topology of the maximum clade credibility (MCC) trees generated from the Bayesian analysis using all three regions is very similar to the topology depicted using the maximum likelihood analysis in Chapter 3. The structure of the MCC trees for the two chloroplast datasets (*matK* and the concatenated regions) and the relationships for the major lineages can be identified in both regions and are congruent with each other. As more individuals are included from different locations for the *matK* region, the major relationships depicted using the concatenated region is upheld.

The trees generated for the concatenated chloroplast region contain considerably more characters and therefore a more detailed pattern for the phylogeography, particularly for the Irish populations, can be depicted. The ITS nuclear region depicts an incongruent topology to the chloroplast regions, which is also observed using the maximum likelihood analysis, and is discussed in detail in Chapter 3.

4.3.6 Conclusion

The ages estimated for lineages within the *Arenaria ciliata* complex in this Molecular Clock Bayesian analysis are surprisingly old, especially the age of separation of the two Irish Ben Bulben haplotypes, and the origin of the distinct Ben Bulben haplotypes which also occur in Rum. Although it is difficult to infer an exact age for the expansion of these lineages, it is clear that these expansions occurred

well before the LGM and that the full extent of the Pleistocene in the last 2.5 million years, and not just the last glacial maximum approximately 20,000 years ago, played a major role in the evolution of the haplotypes in the *Arenaria ciliata* complex.

The maximum clade credibility trees inferred from the Bayesian analysis are equivalent to the maximum likelihood phylogenetic trees generated in Chapter 3, and the phylogeographic patterns convey that there were two routes of colonization into Ireland from different origins. In the case of most of the lineages, the chloroplast haplotypes depict older age estimates for the time of divergence, than the nuclear ITS nucleotypes for each lineage. This highlights the nuclear nature of the ITS region which may reflect more recent gene flow events compared to ancient biogeographic histories depicted for the chloroplast haplotypes.

The times of divergence for refugial populations generally exceed hundreds of thousands of years, predating the LGM (Hewitt, 2000; Taberlet *et al.*, 1998) which is an indicator of a refugial population (Stewart and Lister, 2001). The ancient time of expansion for the deeply divergent haplotypes in Ben Bulben and the variable biogeographic histories arising from independent origins and potential different routes of migration indicates the strong possibility of the survival for the populations of *Arenaria ciliata* in Ben Bulben throughout many glacial periods.

5

Amplified Fragment Length Polymorphism (AFLP) Analysis of the *Arenaria ciliata* Complex

5.1 Introduction

5.1.1 Biogeography and taxonomy of *Arenaria ciliata* and *A. norvegica* using AFLPs

In the two previous chapters, the analyses of sequences from the chloroplast and nuclear genomes have revealed complex and unexpected relationships between *Arenaria ciliata* and *A. norvegica*. In addition, the deeply divergent lineages depicted using the maximum likelihood analyses in Chapter 3 and the Bayesian molecular clock analysis in Chapter 4 depicts ancient divergence times between lineages within the *Arenaria ciliata* complex, some of which predate the origin of most of the *A. norvegica* haplotypes. Of key interest to this study, however, is the origin of the *Arenaria ciliata* haplotypes from Ben Bulben and the identification of their closest sister population. In this Chapter further analyses using Amplified Fragment Length Polymorphisms are performed in order to unravel

these complex relationships. The use of AFLPs may be more appropriate than sequence data for discerning species distinctions between recently evolved species such as *Arenaria norvegica* as many more polymorphic markers are available for analysis. In addition to this, AFLP markers are sampled throughout the genome and they are more likely to reveal rare genetic differences in groups with low sequence variation (Meudt and Clarke, 2007).

The DNA genotyping technique AFLP was first introduced by Vos *et al.* (1995) and has been widely used in both plant phylogeographical studies, e.g. Kropf *et al.* (2008); Chung *et al.* (2004); Westergaard *et al.* (2011); Alsos *et al.* (2007); Dixon *et al.* (2009) and also in other organisms, particularly fungi and bacteria (Bensch and Akesson, 2005). AFLP is particularly appropriate and useful for analyzing the genetics of individuals, populations and independently evolving lineages as AFLP fragments are typically variable enough to assess differences on an individual level (Mueller and Wolfenbarger, 1999). It can also be useful in discriminating between closely related species that would otherwise not be distinguished, for example as a result of recent radiations (Bussell *et al.*, 2005), and arctic-alpine species with shallow phylogenies (Skrede *et al.*, 2009). In some phylogeographic studies, it has proven useful in generating discriminating information on the location of refugial populations and the possibility of glacial survival versus long distance colonisation (Eidesen *et al.*, 2007; Kropf *et al.*, 2008; Westergaard *et al.*, 2011).

5.1.2 Overview of the Amplified Fragment Length Polymorphism (AFLP) technique

The AFLP technique is a PCR-based method that generates amplified DNA fragments which can be calibrated so as to estimate levels of relatedness among individuals. The technique is based on the selective amplification of a subset of fragments that are generated by restriction digestion of the whole genome of individual samples. As the amplified fragments are drawn from across the plant total genome (nuclear, chloroplast, mitochondrial), one can expect that below the species level a small but genetically informative portion of these will vary in size between individual samples, reflecting the slight mutational differences that

exist between these individuals. Because the greater portion of fragments are invariant, this allows the calibration of differences between individuals. Thus, by compiling the total signal contained in the selectively amplified fragments it is possible to estimate the overall genetic relatedness among samples.

Although the AFLP technique is a blind genotyping method (in that the template origin of each fragment is not known), the number of AFLP fragments produced can be controlled to some extent in the selection of the number of bases added to the selective primer. The choice of additional bases is also important in order to construct appropriately sensitive primers which will amplify polymorphic peaks able to discriminate between individuals and to assess the level of diversity that may be present. If the selective primers are too sensitive it's possible to end up with conflicting homoplasic noise, and no resolution; if too insensitive it's possible to end up with poor signal and no resolution. The selection of informative selective primer combinations is important, therefore, in the determination of reproducible fragments which contain informative polymorphisms to discriminate mutational differences between individuals and populations.

An overview of the steps involved in this technique is shown in Figure 5.1. There are 4 main steps involved: (i) Digestion, (ii) Restriction-Ligation, (iii) Preselective Amplification and (iv) Selective amplification. The first step is the digest of the whole genome using restriction enzymes to break the genome into many fragments. In this study the restriction enzymes *Eco*RI and *Mse*I were used. The second step (ii) Restriction-ligation, usually occurs in the same reaction as (i) Digestion, where the sticky ends of the fragments generated from the digest are ligated to *Eco*RI and *Mse*I adaptor sequences. These adaptors are complementary to the preselective primers which are used in the preselective amplification (iii). The subset of samples which bind to the preselective primers are amplified via PCR. Finally, in step (iv) selective PCR amplification, the fragments which were amplified in step (iii) are subjected to another round of PCR, this time with more selective primers containing additional bases at the 3' end. The number of bases added can range from 1-4; the more bases added, the more selective the group of fragments amplified. The combination of selective bases can be varied as can the type of fluorescent dye which is attached to the *Eco*RI selective primer to enable visualization of amplified target fragments when run

with capillary electrophoresis. For each genome analyzed this results in a unique, reproducible profile with many markers (peaks corresponding to fragment length complementary to the selective primer) for each individual.

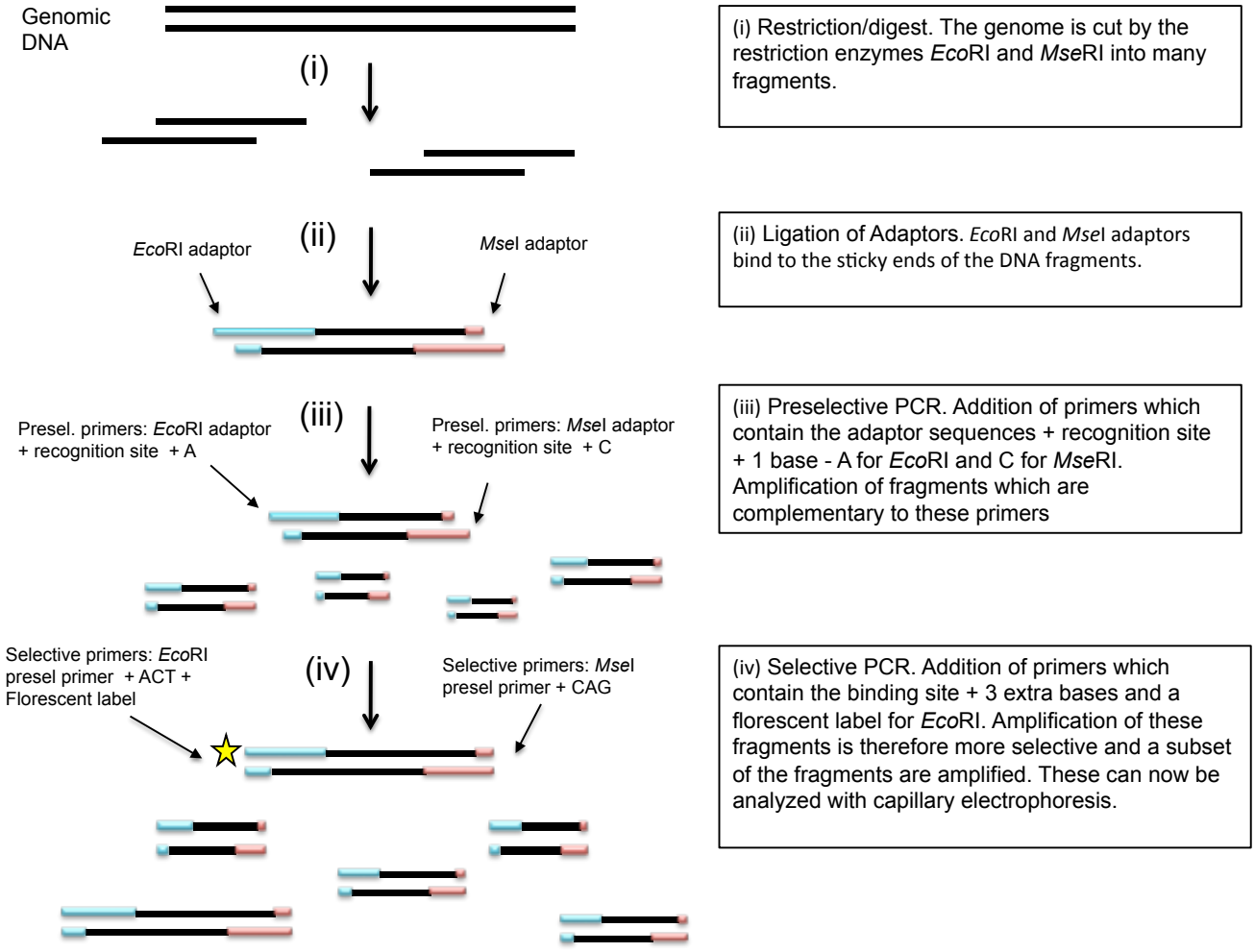


Figure 5.1: An overview of the steps involved for AFLP.

5.1.3 Advantages and Disadvantages of AFLP

Many authors have commended the use of AFLP genotyping for a range of applications (Bensch and Akesson, 2005; Meudt and Clarke, 2007; Mueller and Wolfenbarger, 1999; Vos *et al.*, 1995). One of the major advantages is that no prior information about the nucleotide sequences is needed. Compared to other genotyping techniques such as microsatellites and multigene sequencing, a short start up time is needed and a large amount of loci can be analyzed (>1000). With other methods, <50 loci can be assessed, because of the long start up times and relatively large costs per loci that can be produced and analyzed (Bensch and Akesson, 2005). However, in a study to test the reproducibility of AFLP fragments in a network of European laboratories, the authors found that this technique can be initially difficult to perform (Jones *et al.*, 1997). Once there is greater familiarity with the protocol the AFLP profiles appear to show extremely high reproducibility (Jones *et al.*, 1997).

The AFLP technique involves PCR amplification so only a small amount of DNA is needed. The large number of fragments with high variability compared to sequences from e.g. the chloroplast genome means that it is particularly useful for discriminating between individuals which are not very variable, for example between individuals within a population or taxa with shallow phylogenies. The fragments that make up the fingerprint are widely distributed throughout the genome, allowing an assessment of whole genome variation. The predominately nuclear origin of AFLP fragments may also be significant as plastid genomes may not be sufficiently variable for this technique (Meudt and Clarke, 2007).

One of the major disadvantages of using this technique is that AFLP generates only dominant markers and the heterozygous state cannot be identified at any locus. Therefore, estimates of heterzygosity cannot be calculated as with codominant markers. Techniques have been implemented however to estimate the frequency of the null allele so that average heterzygosity can be determined (Meudt and Clarke, 2007). Size homoplasy is also problematic where markers of the same length although not homologous, are scored incorrectly (Bensch and Akesson, 2005). This happens when non-homologous fragments migrate to the

same position or when different mutations lead to the loss of the same fragment (Meudt and Clarke, 2007).

5.1.4 Aims and Objectives

The objectives of the AFLP fingerprinting work are as follows:

- To evaluate the accepted taxonomic relationship between *Arenaria ciliata* and *A. norvegica* using AFLP data
- To evaluate the genetic relatedness between individuals and populations using neighbour-joining and neighbour-net analyses
- To identify the structure of the AFLP genotypes using Principal Coordinate Analysis (PCoA)
- To analyze the AFLP fragments directly to asses the number of AFLP fragments and private fragments occurring in each population
- To quantify the differentiation of genetic differentiation within and among populations using analysis of molecular variance (AMOVA)

These objectives were carried out to supplement the analyses carried out in Chapter 3 and Chapter 4 - to understand the taxonomic relationships between *Arenaria ciliata* and *A. norvegica* and to observe the grouping of the Irish Ben Bulben *A. ciliata* genotypes in relation to their continental European populations using AFLP data.

5.2 Results

5.2.1 Overview of results

A list of the datasets indicating where the results for the neighbour-joining (NJ), neighbour-net (NN), principal coordinate analysis (PCoA) and analyses of fragments in each matrix, are detailed in Table 5.1. The names for the individual genotypes in each figure correspond to the name originally assigned to individuals of *Arenaria ciliata* and *A. norvegica*. The two letters represent the taxa name; AC - *Arenaria ciliata*, AN-*Arenaria norvegica* and AP - *Arenaria ciliata* subsp. *pseudofrigida*; the first number represents the population and the second number represents the individual sampled from that population (i.e. AC1.28 is *Arenaria ciliata* from population 1, individual 28 from Ben Bulben). Details of the sites where the populations were sampled from are summarized in Table 5.2.

The NJ trees demonstrated a similar pattern to the NN and the PCoA in each of the datasets. The results from the NJ and neighbour-net analyses are presented together, and the PCoA are presented together with the results from the fragment analyses. Bootstrap values over 50 are shown on the branches for the major lineages on the NJ tree.

The NUIM 2009 dataset contains mostly individuals from Ireland, the West Alps and a high proportion of individuals of *Arenaria norvegica*. Dataset RJB contains many individuals of *Arenaria ciliata* from Ireland, the West Alps and the Pyrenees, and dataset NUIM 2012 contains a few individuals from every population sampled representing a broad geographic spread of sampled AFLP genotypes.

Table 5.1: Details of the datasets used for AFLP analyses

Dataset	**No. Indivs.**	**No. Loci**	**NJ / NN Analysis**	**PCoA / Fragment Analysis**
NUIM 2009 3 sel. primers: ACT-CAG / AGC-CAA / AGC-CTG	30	146	Figure 5.2	Figure 5.8
NUIM 2009 2 sel. primers: AGC-CAA / AGC-CTG	41	81	Figure 5.3	Figure 5.9
NUIM 2009 ACT-CAG	58	65	Appendix G.1	Appendix H.1
NUIM 2009 AGC-CAA	72	47	Appendix G.2	Appendix H.2
NUIM 2009 AGC-CTG	83	34	Appendix G.3	Appendix H.3
RJB All indiv. 2 sel. primers: AGC-CAA / AGC-CTG	82	176	Figure 5.4	Figure 5.10
RJB less Irish indiv. 2 sel. primers: AGC-CAA / AGC-CTG	46	176	Figure 5.5	Figure 5.11
NUIM 2012 3 sel. primers: ACT-CAG / AGC-CAA / AGC-CTG	20	221	Figure 5.6	Figure 5.12
NUIM 2012 2 sel. primers: ACT-CAG / AGC-CAA	31	146	Figure 5.7	Figure 5.13
NUIM 2012 ACT-CAG	45	91	Appendix G.4	Appendix H.4
NUIM 2012 AGC-CAA	38	55	Appendix G.5	Appendix H.5
NUIM 2012 AGC-CTG	35	75	Appendix G.6	Appendix H.6

No. Indivs.: number of individuals; No. Loci.: number of Loci

5.2.2 Neighbour-joining (NJ) and neighbour-net (NN) Analyses

The neighbour-joining (NJ) and neighbour-net (NN) analyses identified similar groupings of AFLP genotypes. These groups were more strongly depicted in the datasets which contained fragments obtained from combining 2 or 3 selective primers and therefore more loci per dataset. Because more complex evolutionary scenarios may be poorly described when using tree building methods such as neighbour-joining, phylogenetic networks can aid visualising more complex evolutionary events such as hybridization, horizontal gene transfer and recombination (Huson and Bryant, 2006). For this reason Neighbour-net (NN) networks based on Nei-Li distances were also constructed using SplitsTree 4 (Huson and Bryant, 2006) to visualize the relationship of the AFLP genotypes among populations.

The bootstrap values on the NJ trees were higher using these datasets where most of the lineages are supported by strong bootstrap values >60. With the exception of the *Arenaria norvegica* individuals, in the NUIM 2009 (Figures 5.2 and 5.3) and RJB datasets (Figures 5.4 and 5.5) the lineages correspond to location, where Ireland and the West Alps are constrained to their own group. In the NUIM 2009 dataset *Arenaria norvegica* is placed within its own group, which is distinct from the other *A. ciliata* lineages (Figure 5.2). This is also observed in the NUIM 2012 datasets where *A. norvegica* is clearly placed within its own lineage (Figures 5.6 and 5.7). In all datasets which contain *A. norvegica* (Figures 5.2, 5.3, 5.6 and 5.7), the NJ trees identify the individuals of *A. ciliata* from Ben Bulben as the closest sister group to the *A. norvegica* lineage. When more individuals are included from populations from a broader range of locations in dataset NUIM 2012, this pattern is still evident, where a strong bootstrap value links the two groups in the dataset of 2 selective primers. Another strong pattern which emerges when more locations are included is the placement of Svalbard with Ben Bulben in the NJ analyses, indicated by strong bootstrap values >80 (Figures 5.6 and 5.7). This is also very clear in the neighbour-net diagrams where the individuals from Ben Bulben and Svalbard are grouped together.

For the most part, the Irish Ben Bulben individuals are constrained to their own lineage (and Svalbard in datasets NUIM 2012). The only exception is in

Table 5.2: Location of each population sampled for *Arenaria ciliata* and *A. norvegica*

Population name	Location of Population Sampled
AC1	Ben Bulben, Ireland
AC2	Ben Bulben, Ireland
AC3	Ben Bulben, Ireland
AC4	Ben Bulben, Ireland
AC6	West Alps
AC7	West Alps
AC8	West Alps
AC9	West Alps
AC10	Pyrenees, Spain
AC11	Picos De Europa, Spain
AC13	Pyrenees, Spain
AC14	Picos De Europa, Spain
AC15	Picos De Europa, Spain
AC16	East Alps
AC17	East Alps
AC 18	Swiss Alps
AC19	Swiss Alps
AP1	Svalbard
AP2	Svalbard
AN1	The Burren, Co. Clare, Ireland
AN2	Ribblesdale, Yorkshire, England
AN3	Ribblesdale, Yorkshire, England
AN4	Rum, Scotland
AN5	Inchnadamph, Scotland
AN6	Shetland, Scotland
AN7	Iceland
AN8	Iceland

some of the datasets which contain one selective primer which either contain less resolution in dataset NUIM 2009 AGC-CTG, which contains only 34 fragments and no bootstrap values > 50 so there is much less resolution and more noise in this dataset. The only other dataset in which the Ben Bulben individuals were not constrained to their own lineage was in dataset NUIM 2012 AGC-CAA. This dataset contains 55 fragments and no bootstrap values over 50. However

interestingly, the pattern for the grouping of the Irish *A. ciliata* individuals here is similar to that depicted for the chloroplast regions where one group is related to the East Alps and the other group is more distantly related and is associated with Svalbard and *A. norvegica*. This is evident in both the NJ and neighbour-net analyses (Appendix G.5). In the RJB datasets which contain two selective primers there is one individual from the West Alps (AC6.17) which is grouped with the Irish Ben Bulben genotypes (Figures 5.4 and 5.5). This also occurs in the dataset NUIM 2009 AGC-CAA where AC6.20 is grouped with *A. norvegica* and in dataset NUIM 2009 AGC-CTG AC6.21 is grouped with *A. norvegica* (Appendix G.6). This could be due to the presence of homologous fragments which cause the incorrect placement of these individuals in the NJ and neighbour-net analysis.

In the dataset which contains the most locations, NUIM 2012 (Figure 5.7), five main groups are identified: (i) Ireland/Svalbard, (ii) East Alps, (iii) West Alps, (iv) Spain/Swiss Alps and (v) *Arenaria norvegica*. These are strongly supported by high bootstrap values in the NJ trees (Figures 5.6 and 5.7) and are also evident in the neighbour-net diagrams. These groupings are also observed in the NJ and neighbour-net figures for the individual selective primers (Appendix G.4, G.5 and G.6).The two datasets constructed for the RJB datasets using two selective primers showed similar patterns and no difference was observed when the number of Irish individuals was reduced. In both datasets, there are three main groups observed, Ireland, Pyrenees and the West Alps.

There is no strong evidence for an internal structuring within the *Arenaria norvegica* lineage. When more *A. norvegica* individuals are included in the NUIM 2009 datasets, there is a split into two groups, one which contains individuals from Yorkshire (AN2 / AN3) and Inchnadamph, Scotland (AN5) and another group which contains individuals from The Burren, Ireland (AN1), Yorkshire (AN2/3), Rum (AN4) and Inchndamph (AN5) (Figures 5.2 and 5.3). This pattern is somewhat reflected in the individual selective primers especially for ACT-CAG, however this pattern the other two datasets AGC-CAA and AGC-CTG is not as clear.

The branch lengths in the neighbour-net diagrams may be an indicator of less divergence between individuals and this is particularly evident for the individuals of *Arenaria norvegica* where the branch lengths are much shorter than seen for

the individuals of *A. ciliata*. The branch lengths for the European genotypes of *Arenaria ciliata* and the branch lengths for the Irish Ben Bulben branch lengths are variable. The Ben Bulben branch lengths are generally longer than those seen for *Arenaria norvegica* but shorter than the European individuals (e.g. Appendix G.4). In some datasets however they remain just as long as the European genotypes (e.g. Figures 5.2 and 5.7).

5.2.3 Principal Coordinate Analysis (PCoA)

The principal coordinate analysis (PCoA) identified groupings which were congruent with the groupings found using the NJ and neighbour-net analyses. The majority of the percentage of variation was accounted for by axis 1 and axis 2. Generally the inclusion of a higher number of fragments using more selective primers, provided a stronger grouping of individuals, however some indiviual primers also showed a clear pattern such as NUIM 2009 ACT-CAG (Appendix H.1 and H.6). A very distinct grouping is evident for the clustering of *Arenaria norvegica* in all datasets which include this taxon. These genotypes are separated from *Arenaria ciliata* according to Axis 1, which accounts for the majority of the variation, (>30% in every dataset). The Irish Ben Bulben genotypes are also found forming a distinct cluster together in most datasets. In the RJB datasets a genotype from the West Alps is found to cluster with the Ben Bulben population and this reflects the anomaly also observed in the NJ and neighbour-net analyses.

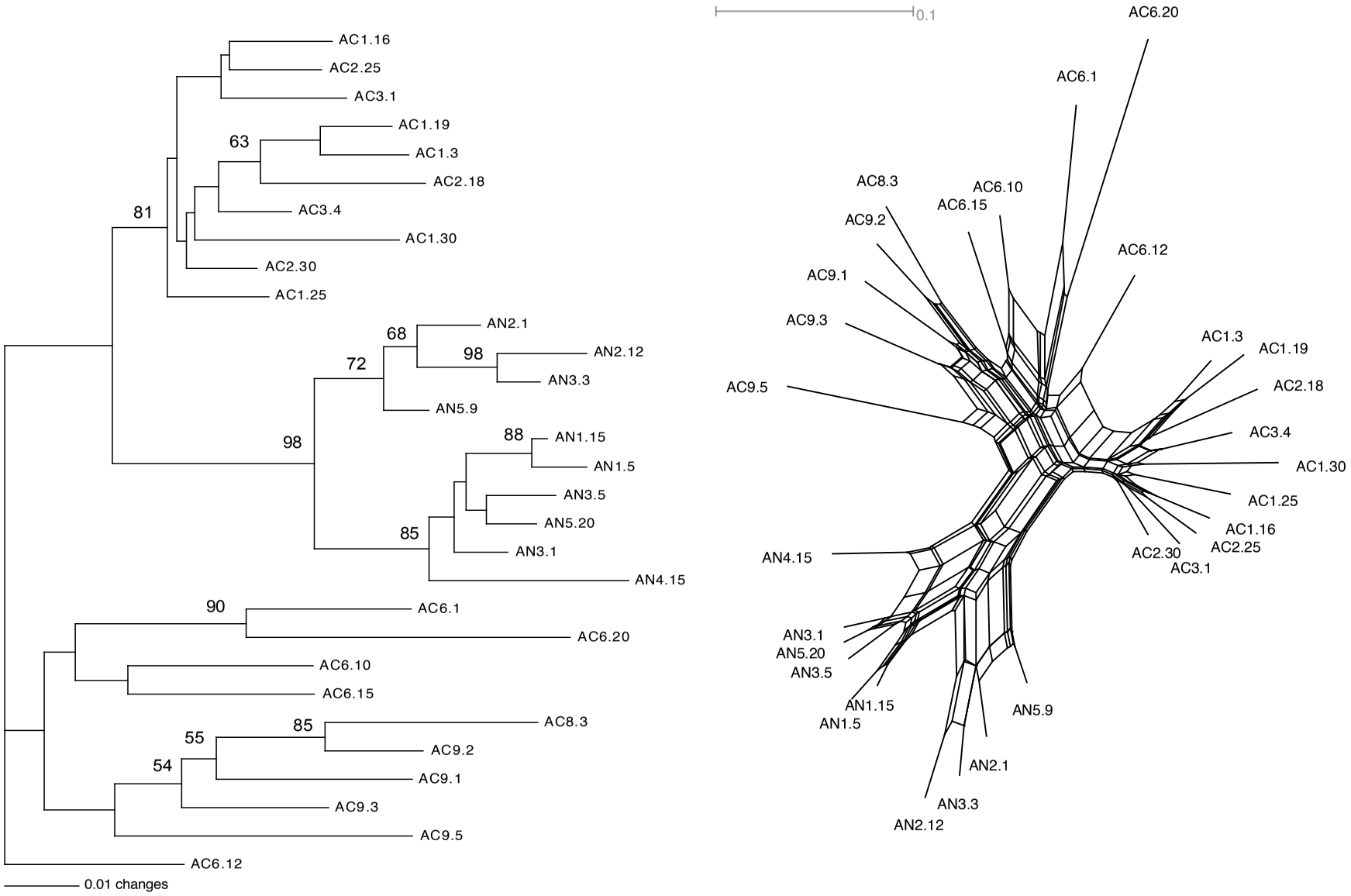


Figure 5.2: Neighbour-Joining (NJ) tree and neighbour-net network for dataset NUIM 2009 for three selective primers

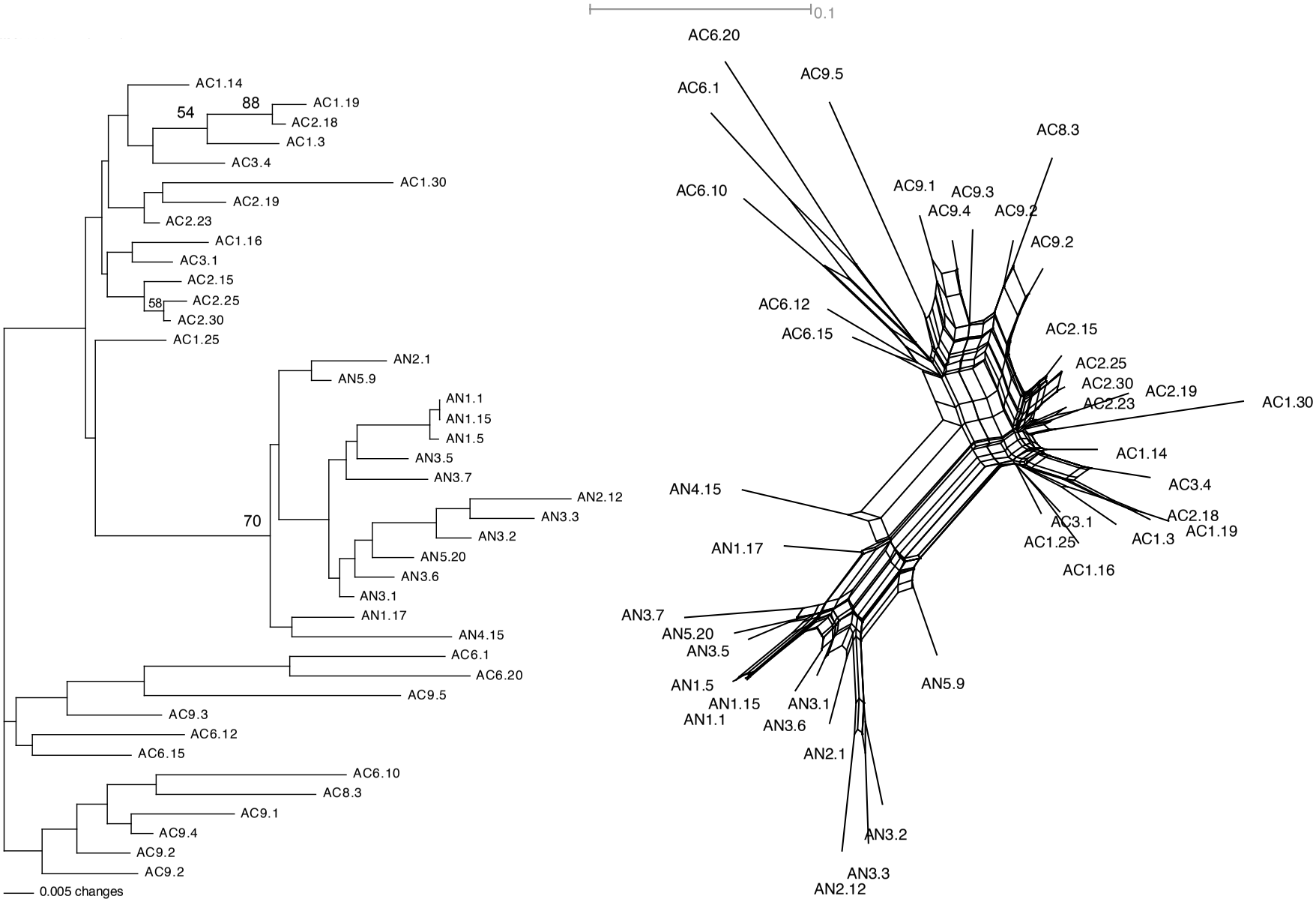


Figure 5.3: Neighbour-Joining (NJ) tree and neighbour-net network for dataset NUIM 2009 for the selective primers AGC-CAA / AGC-CTG

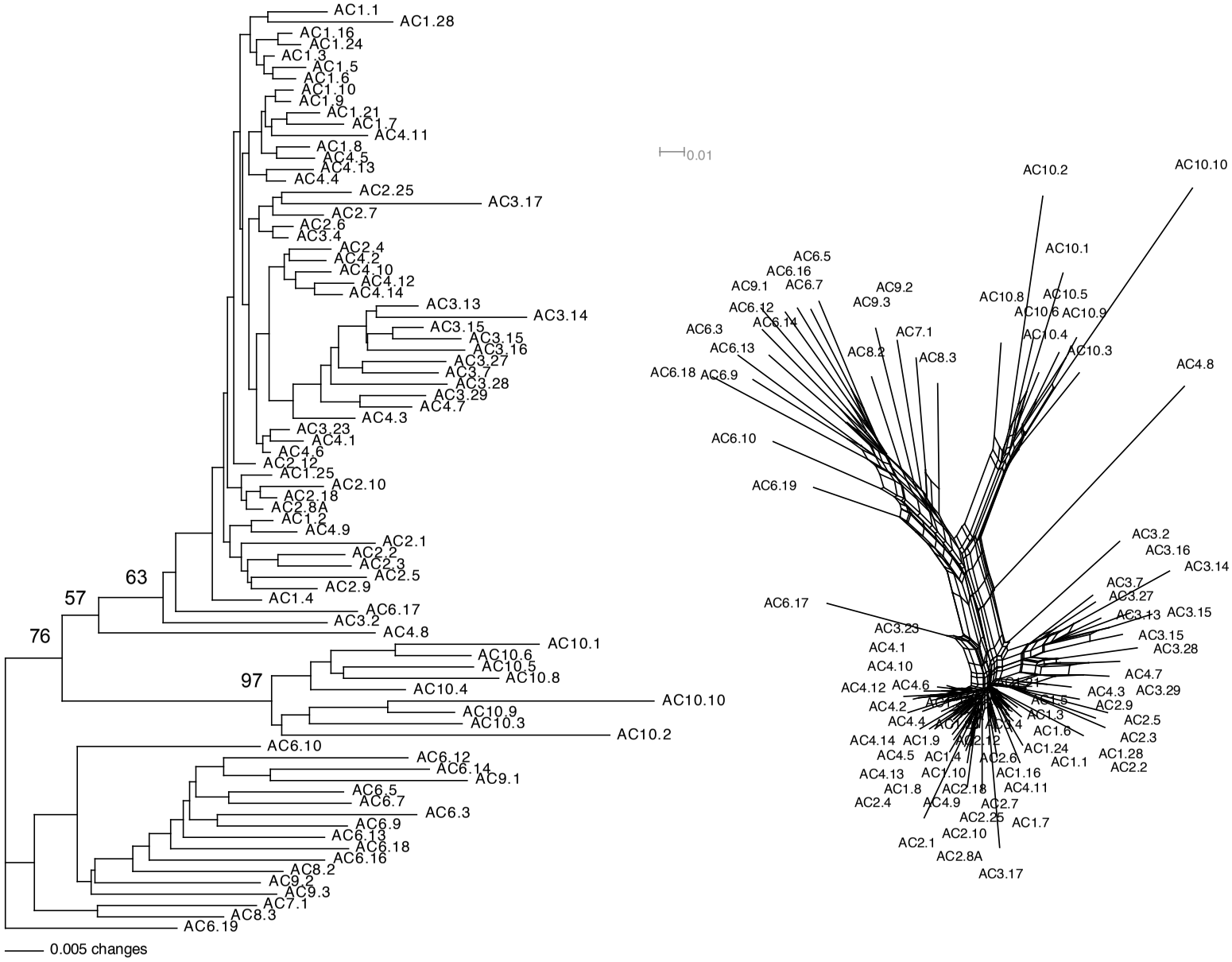


Figure 5.4: Neighbour-Joining (NJ) tree and neighbour-net network for dataset RJB with all individuals included for the selective primers AGC-CAA / AGC-CTG

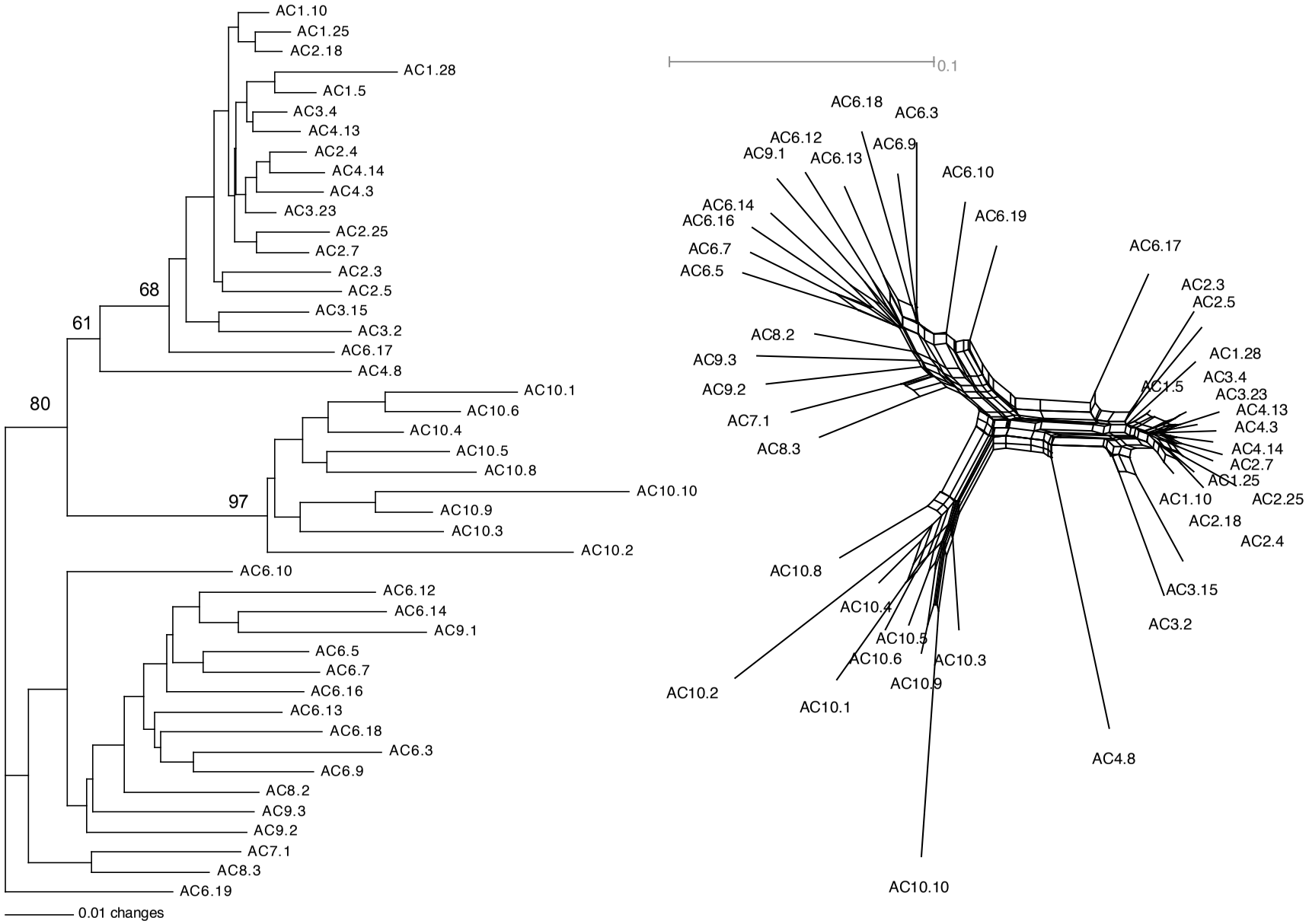


Figure 5.5: Neighbour-Joining (NJ) tree and neighbour-net network for dataset RJB with reduced number of Irish individuals for the selective primers AGC-CAA / AGC-CTG

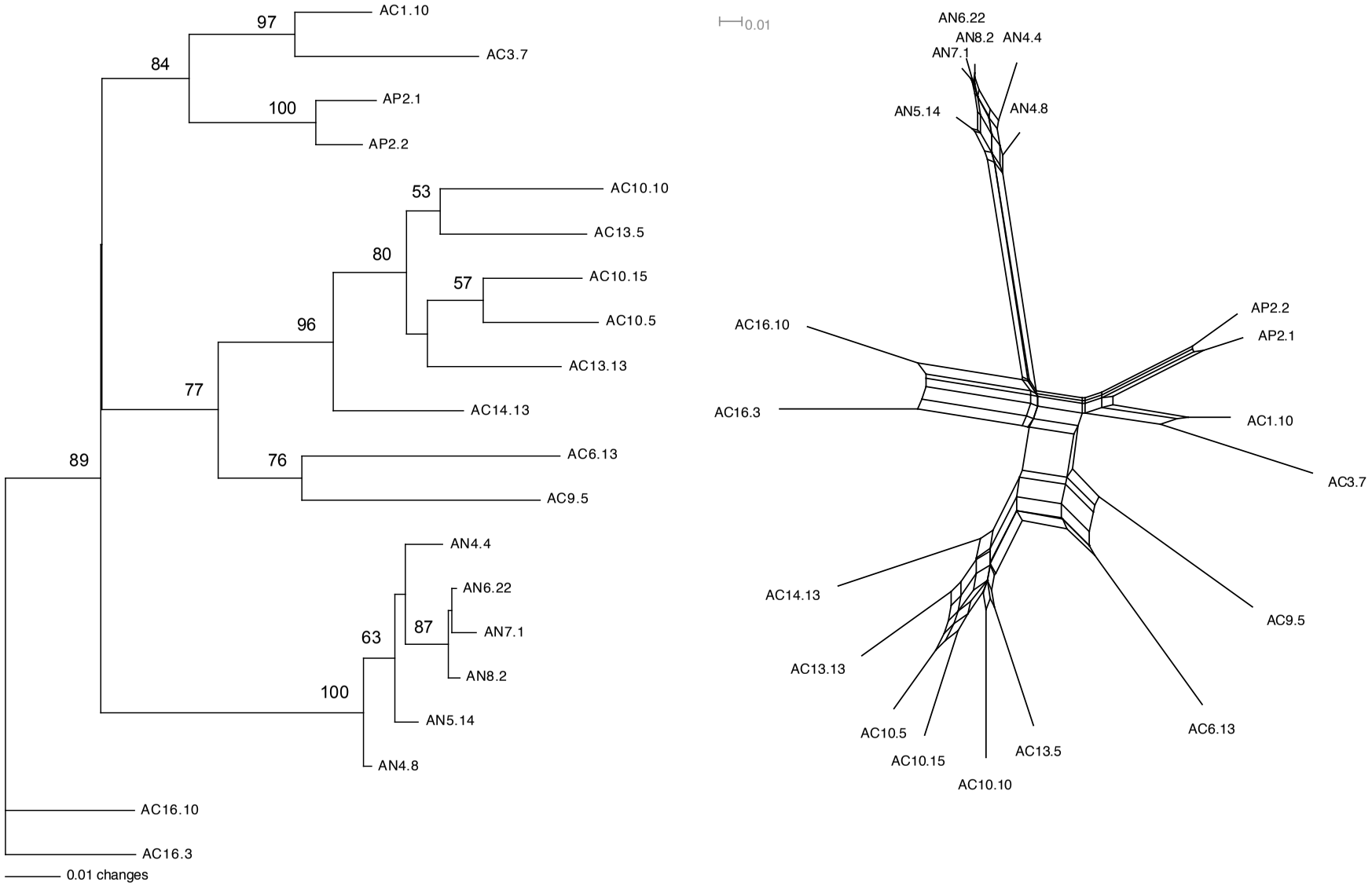


Figure 5.6: Neighbour-Joining (NJ) tree and neighbour-net network for dataset NUIM 2012 for the three selective primers

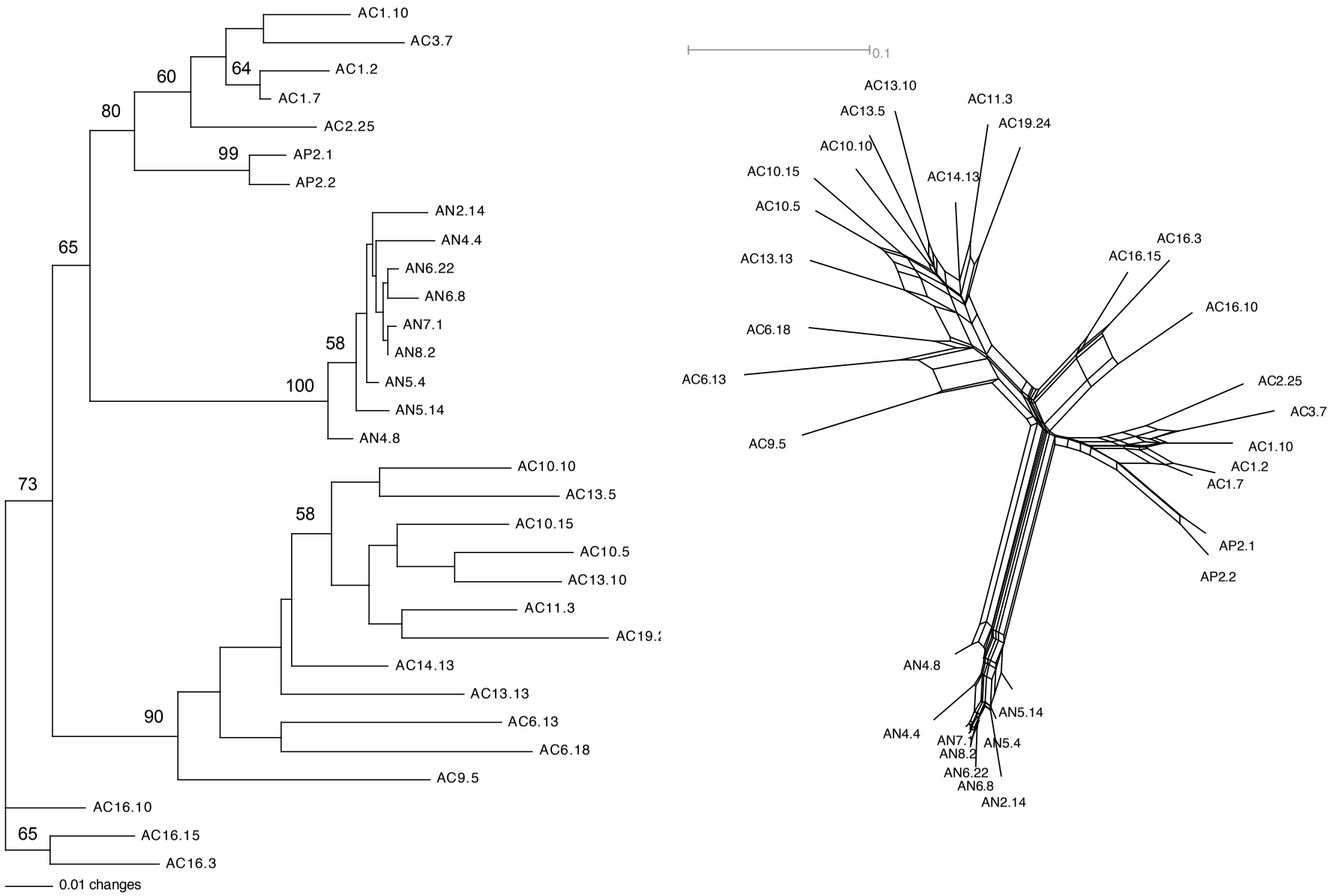


Figure 5.7: Neighbour-Joining (NJ) tree and neighbour-net network for dataset NUIM 2012 for the selective primers ACT-CAG / AGC-CAA

In the NUIM 2012 datasets including 3 primer combinations, the Irish Ben Bulben genotypes are grouped with Svalbard, this pattern is also observed in the NJ and neighbour-net analyses. In the PCoA however the Irish genotypes are also placed in a broad group with the East and West Alps (Figure 5.12). When more individuals are included in dataset NUIM 2012 for 2 selective combinations, the Irish and Svalbard still form a distinct grouping, however intermediate between the Spanish/Swiss group and the Irish group are individuals from the East and West Alps (Figure 5.13). This pattern is not as evident in the NJ and neighbour-net analyses.

A pattern which is evident in the NJ trees and the neighbour-net diagrams but not in the PCoA is the relationship between the Irish Ben Bulben genotypes and *Arenaria norvegica*. According to the PCoA the *A. norvegica* are not grouped closely to the individuals unlike the NJ trees and neighbour-net figures which indicate that *A. norvegica* is grouped with *A. ciliata* from Ben Bulben.

5.2.4 Assessment of fragments in each dataset - presence of private fragments

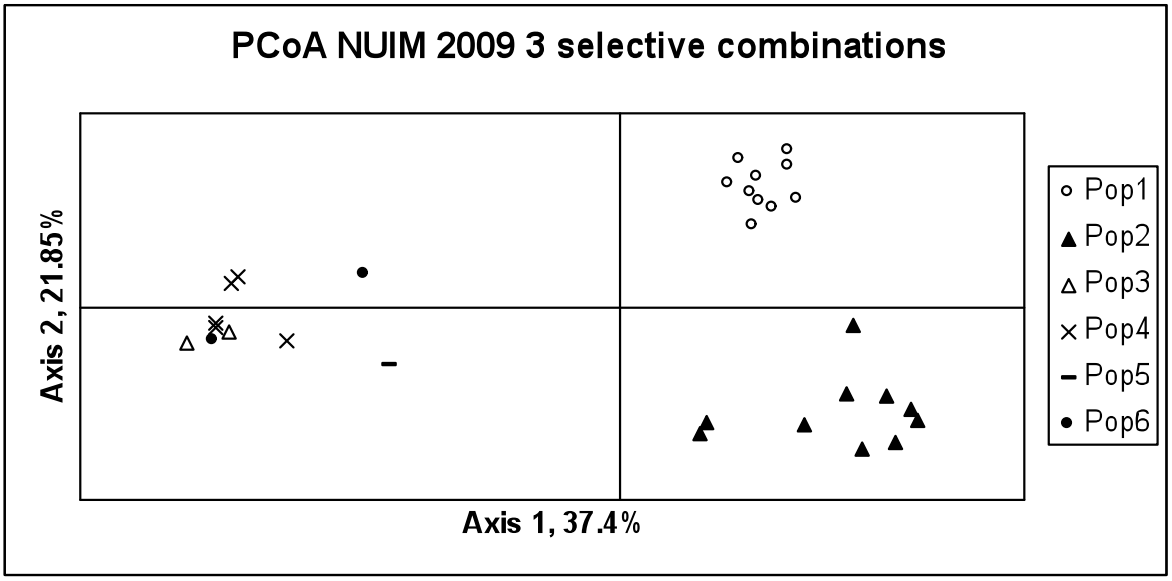
The highest number of private fragments occurring per individual (Pf / N) (fragments exclusive to one population), according to the datasets NUIM 2009 and RJB is the West Alps. The only exceptions in these datasets is RJB where all individuals are included and the proportion of private fragments per individual is the same for the West Alps and the Pyrenees (Figure 5.10) and NUIM 2009 AGC-CTG, where the Pf / N is 0.4 compared to 0.3 for the West Alps (Appendix H.3). When more individuals are included, from a broader range of locations, in the NUIM 2012 dataset, the West Alps is still indicated as having the highest proportion of private fragments per individual, where the East Alps and The Pyrenees are the next highest (Figure 5.12 and Figure 5.13).

The Ben Bulben population contains the next highest proportion of private fragments per individual after the Alps and The Pyrenees. This is observed in every dataset where at least two selective primer combinations are included (Figures: 5.8, 5.9, 5.10, 5.11, 5.12 and 5.13). A high proportion of private alleles

for the Irish and Svalbard populations are observed in the NUIM 2012 datasets in particular where there are 4 private alleles per individual recorded for the Ben Bulben population for 3 selective combinations, and 3.5 for the Svalbard population, compared to 9 in the West Alps, 6 in the East Alps and 5.6 in the Pyrenees.

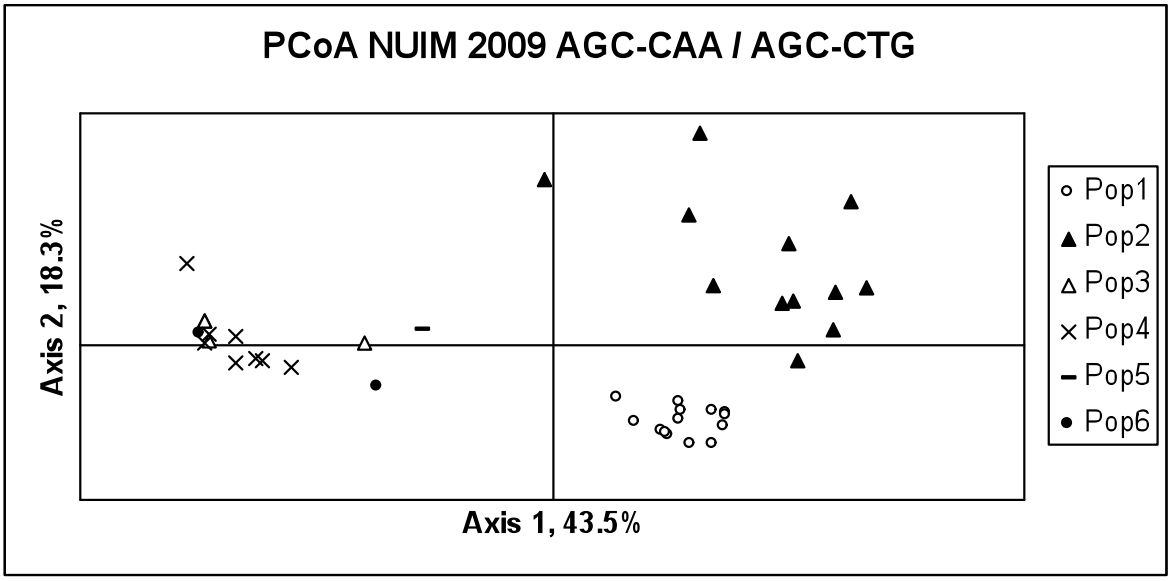
The number of private fragments occurring in the *Arenaria norvegica* populations are much lower than *A. ciliata* and in some cases no private fragments are recorded in the population of *A. norvegica* (e.g. Figure 5.13 and 5.8). The populations of *Arenaria norvegica* where private alleles were found, (at least 0.3 Pf per individual), in Yorkshire (dataset NUIM 2009 3 selective combinations - 0.4 Pf per individual), Rum (dataset NUIM 2012 3 selective combinations - 1 Pf per individual; dataset NUIM 2009 AGC-CTG - 0.4 Pf per individual; dataset NUIM 2012 ACT-CAG 0.3 Pf per individual) and Iceland (dataset NUIM 2012 3 selective combinations - 1 Pf per individual).

The number of fragments per population (Nf), the number of fragments with a frequency $> 5\%$, the number of private fragments (Pf) and the number of private fragments per individual (Pf) were calculated for all populations including at least 2 individuals. Any population that contained only one individual were omitted from this analysis. The proportion of fragments occurring $> 5\%$ occurred in most populations. The exceptions are in datasets RJB for the Irish Ben Bulben population where the number of fragments occurring $> 5\%$ is reduced when all individuals are included (Figure 5.10). However when the number of Irish individuals are reduced, all of the fragments identified in this population occur $> 5\%$ of the time (Figure 5.11). The only other dataset where this occurs is in dataset NUIM 2009 AGC-CTG (Appendix H.3).



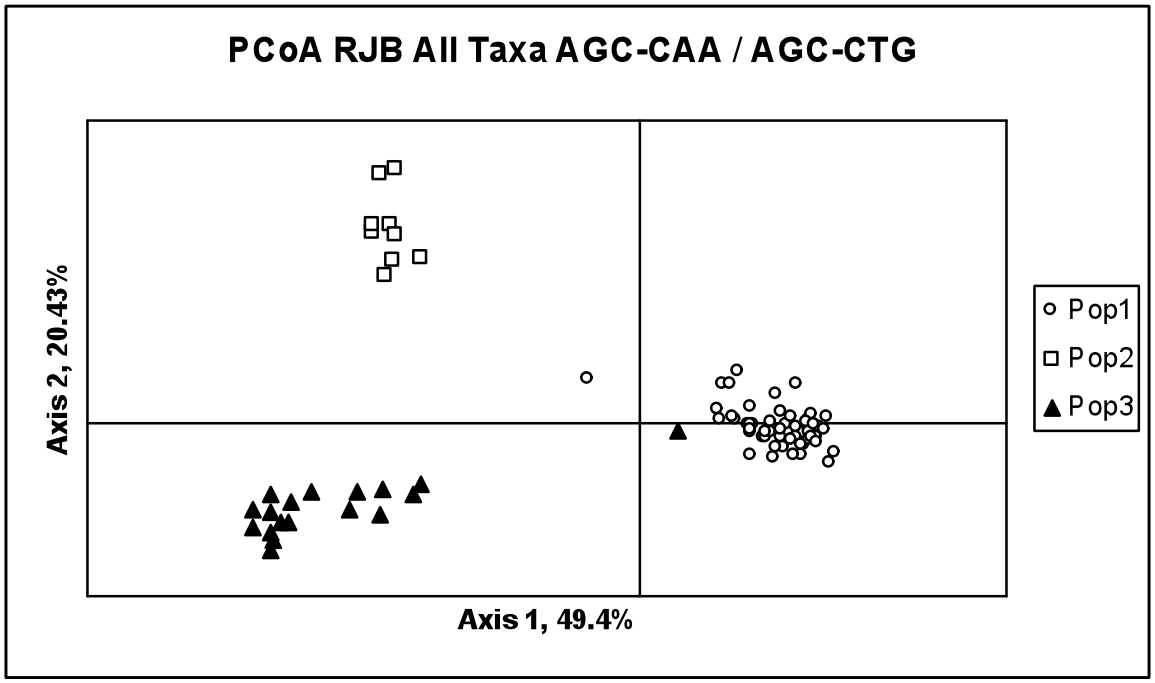
Dataset: NUIM 2009 3 selective combos - 146 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	10	68	68	5	0.5
Pop2 - West Alps	10	98	98	27	2.7
Pop3 - The Burren, Ireland	2	47	47	0	0.0
Pop4 - Yorkshire, England	5	68	68	2	0.4
Pop5 - Rum, Scotland	1	na	na	na	na
Pop6 - Inchnadamph, Scotland	2	58	58	0	0.0

Figure 5.8: Principal coordinate analysis (PCoA) for dataset NUIM 2009 for 3 selective primers. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.



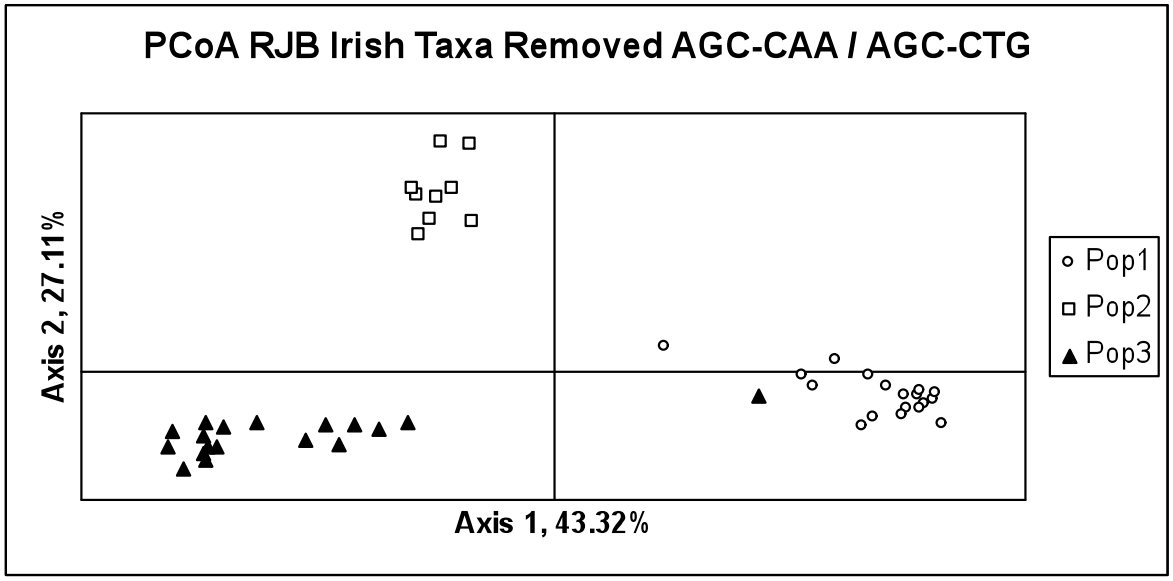
Dataset: NUIM 2009 2 AGC-CAA/AGC-CTG - 81 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	14	37	37	4	0.3
Pop2 - West Alps	12	50	50	11	0.9
Pop3 - The Burren, Ireland	4	29	29	0	0.0
Pop4 - Yorkshire, England	8	42	42	1	0.1
Pop5 - Rum, Scotland	1	na	na	na	na
Pop6 - Inchnadamph, Scotland	2	32	32	0	0.0

Figure 5.9: Principal coordinate analysis (PCoA) for dataset NUIM 2009 for the selective primers AGC-CAA / AGC-CTG. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.



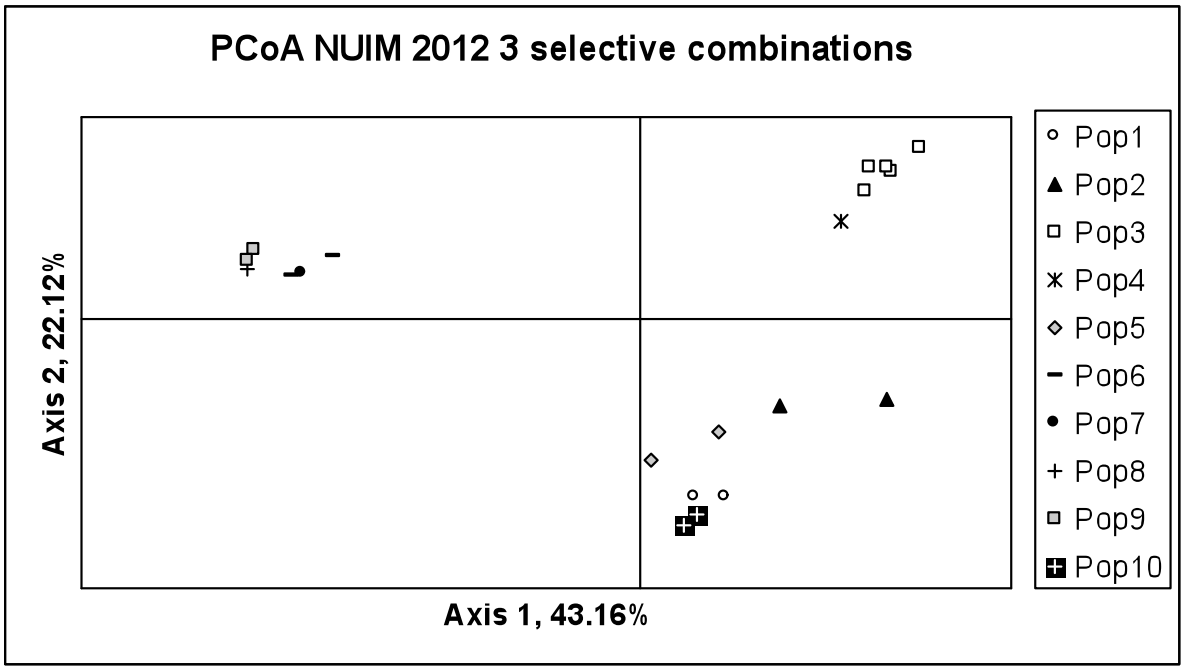
Dataset: RJB all taxa - AGC-CAA/AGC-CTG - 176 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	55	109	61	20	0.4
Pop2 - Pyrenees, Spain	9	81	81	14	1.6
Pop3 - West Alps	18	117	117	29	1.6

Figure 5.10: Principal coordinate analysis (PCoA) for dataset RJB including all individuals for the selective primers AGC-CAA / AGC-CTG. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.



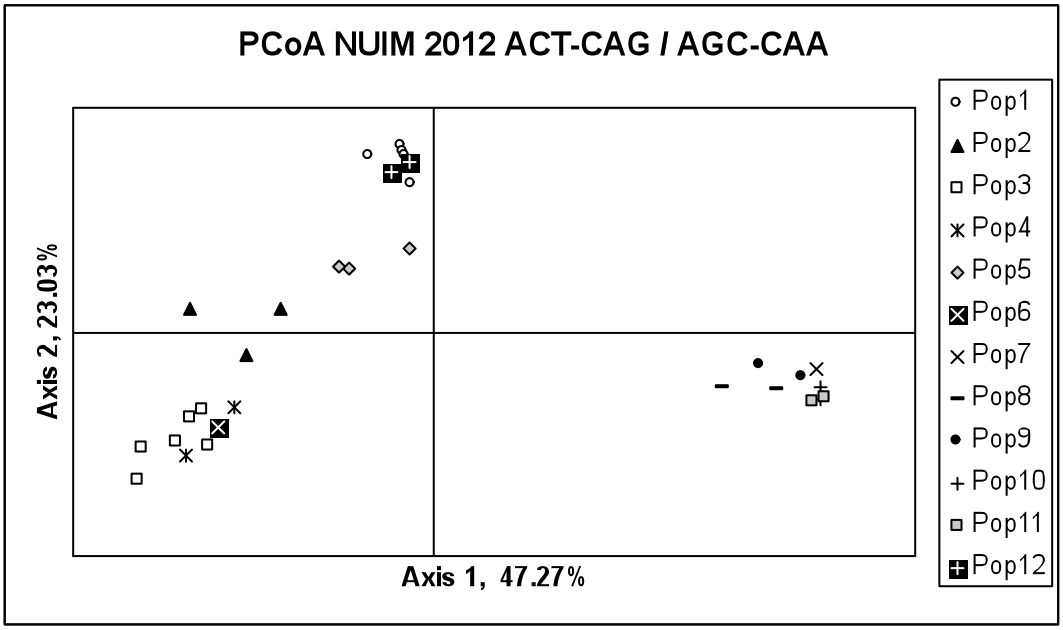
Dataset: RJB less Ire taxa - AGC-CAA/AGC-CTG - 176 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	18	90	90	13	0.7
Pop2 - Pyrenees, Spain	9	81	81	16	1.8
Pop3 - West Alps	18	117	117	36	2.0

Figure 5.11: Principal coordinate analysis (PCoA) for dataset RJB with a reduced number of Irish individuals for the selective primers AGC-CAA / AGC-CTG. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.



Dataset: NUIM 2012 3 combos - 221 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	2	58	58	8	4.0
Pop2 - West Alps	2	73	73	18	9.0
Pop3 - Pyrenees, Spain	5	89	89	28	5.6
Pop4 - Picos De Europa, Spain	1	na	na	na	na
Pop5 - East Alps	2	76	76	12	6.0
Pop6 - Rum, Scotland	2	72	72	2	1.0
Pop7 - Inchnadamph, Scotland	1	na	na	na	na
Pop8 - Shetland, Scotland	1	na	na	na	na
Pop9 - Iceland	2	66	66	2	1.0
Pop10 - Svalbard	2	55	55	7	3.5

Figure 5.12: Principal coordinate analysis (PCoA) for dataset NUIM 2012 for the 3 selective primers. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.



Dataset: NUIM 2012 ACT-CAG/AGC-CAA - 146 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	5	52	52	7	1.4
Pop2 - West Alps	3	58	58	10	3.3
Pop3 - Pyrenees, Spain	6	66	66	10	1.7
Pop4 - Picos De Europa, Spain	2	44	44	2	1.0
Pop5 - East Alps	3	56	56	6	2.0
Pop6 - Swiss Alps	1	na	na	na	na
Pop7 - Yorkshire, England	1	na	na	na	na
Pop8 - Rum, Scotland	2	53	53	0	0.0
Pop9 - Inchnadamph, Scotland	2	48	48	0	0.0
Pop10 - Shetland, Scotland	2	52	52	0	0.0
Pop11 - Iceland	2	47	47	0	0.0
Pop12 - Svalbard	2	42	42	4	2.0

Figure 5.13: Principal coordinate analysis (PCoA) for dataset NUIM 2012 for the selective primers ACT-CAG / AGC-CAA. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.

5.2.5 Analysis of molecular variance (AMOVA)

The source of variation for the datasets which were subjected to AMOVA and the PhiPT values are recorded in Table 5.3. How the populations in each dataset were partitioned into groups are also indicated in this Table 5.3 where "reg" represents the grouping of the individuals of *Arenaria ciliata* and *A. norvegica* according to geographic region. These grouping were also used for the PCoA and are shown in Tables 5.8, 5.9, 5.10, 5.11, 5.12, 5.13 and in Appendix H. Clu represents the individuals of *A. ciliata* and *A. norvegica* which were grouped according to the main clusters identified in the PCoA, NN and NJ analysis; AC represents the grouping of *Arenaria ciliata* individuals according to geographic region, and AN represents the grouping of *A. norvegica* individuals according to geographic region. The datasets that were not included in this analysis contained a low number of individuals per population (NUIM 2012 3 selective primers), or a low resolution and loci number (NUIM 2009 AGC-CTG).

There was only a slight difference in the within/among population variation when the individuals in the datasets were partitioned according to geographic regions (reg) or clusters according to the NJ, NN and PCoA (clu). There was a large difference observed in the within versus among variation however when each dataset was partitioned into groups for both taxa versus the analysis of *Arenaria ciliata* and *A. norvegica* separately.

In every dataset the percentage of within population variation increases when each taxon is analyzed separately. This occurs for both taxa however the proportion of within population variation increases when individuals of *A. norvegica* only are analyzed. This high ratio of within population variation is observed across all *A. norvegica* only datasets where the lowest within population variation is 64% in NUIM AGC-CTG and the highest is 99% in NUIM 2009 3 selective primers. The lowest within population variation for the *Arenaria ciliata* only dataset occurs in dataset RJB all individuals 2 selective primers at 52% and the highest in datasets 2009 at 71%.

The ratios of within versus among population variation differ between the batches of work for NUIM 2009, RJB and NUIM 2012. In the NUIM 2009 datasets, generally there is a higher ratio of within population variation. The

exceptions to this are dataset NUIM 2009 2 selective primers for both taxa (56:44, within: among) and NUIM AGC-CAA both taxa (51:49, within: among). In the RJB datasets the ratio for within to among population variation is 52:48 for all individuals included and 58:42 for the trimmed dataset. For the NUIM 2012 datasets, the proportion of within population variation decreases and the among population variation increases in the datasets containing both taxa. The exception to this is dataset NUIM 2012 where the ratio for within to among population variation is 51:49.

Table 5.3: Results of the analysis of molecular variance (AMOVA)

Dataset	Within group variation (%)	Among population variation (%)	PhiPT	p-value
NUIM 2009 3 sel. - reg	63	37	0.372	0.001
NUIM 2009 3 sel. - clu	60	40	0.395	0.001
NUIM 2009 3 sel. - AC	71	29	0.294	0.001
NUIM 2009 3 sel. - AN	99	1	0.012	0.482
NUIM 2009 2 sel. - reg	56	44	0.437	0.001
NUIM 2009 2 sel. - clu	57	43	0.427	0.001
NUIM 2009 2 sel. - AC	71	29	0.287	0.001
NUIM 2009 2 sel. - AN	80	20	0.203	0.011
NUIM 2009 ACT-CAG - reg	62	38	0.376	0.001
NUIM 2009 ACT-CAG - clu	60	40	0.398	0.001
NUIM 2009 ACT-CAG - AC	71	29	0.295	0.001
NUIM 2009 ACT-CAG - AN	81	19	0.192	0.001
NUIM 2009 AGC-CAA - reg	51	49	0.489	0.001
NUIM 2009 AGC-CAA - clu	51	49	0.489	0.001
NUIM 2009 AGC-CAA - AC	66	34	0.339	0.001
NUIM 2009 AGC-CAA - AN	65	35	0.346	0.001
RJB All indivs. 2 sel.	52	48	0.481	0.001
RJB Less Irish indivs. 2 sel.	58	42	0.423	0.001
NUIM 2012 2 sel. - reg	45	55	0.551	0.001
NUIM 2012 2 sel. - clu	43	57	0.565	0.001
NUIM 2012 2 sel. - AC	56	44	0.442	0.001
NUIM 2012 2 sel. - AN	80	20	0.202	0.029
NUIM 2012 ACT-CAG - reg	51	49	0.49	0.001
NUIM 2012 ACT-CAG - clu	49	51	0.505	0.001
NUIM 2012 ACT-CAG - AC	56	44	0.44	0.001
NUIM 2012 ACT-CAG - AN	87	13	0.13	0.035
NUIM 2012 AGC-CAA - reg	37	63	0.632	0.001
NUIM 2012 AGC-CAA - AC	57	43	0.427	0.001
NUIM 2012 AGC-CAA - AN	69	31	0.306	0.004
NUIM 2012 AGC-CTG - reg	42	58	0.58	0.001
NUIM 2012 AGC-CTG - AC	59	41	0.41	0.001
NUIM 2012 AGC-CTG - AN	64	36	0.36	0.002

How the datasets were partitioned are indicated; reg - populations divided by geographic region; AC - populations of *Arenaria ciliata* only divided by geographic region; AN - populations of *A. norvegica* only divided by geographic region; clu - populations of both taxa grouped according to the main clusters identified in the PCoA, NN and NJ analysis. 3 sel. indicates datasets with fragments generated with 3 selective primers, and 2 sel. indicates datasets with fragments generated from 2 selective primers.

5.3 Discussion

5.3.1 Taxonomic relationship and grouping of the *Arenaria norvegica* AFLP genotypes

A distinct group is evident in each dataset which includes all of the *Arenaria norvegica* AFLP genotypes. This group is evident using neighbour-joining (NJ), neighbour-net (NN) and principal coordinate analysis (PCoA). In the NJ analysis the bootstrap support values are very strong for this grouping and in some NJ trees this bootstrap value is 100.

Similar to the pattern observed for the ITS sequences (maximum likelihood analysis - Chapter 3 and Bayesian analysis - Chapter 4), *Arenaria ciliata* populations group according to geographic region, whereas individuals of *A. norvegica* all cluster together, irrespective of geographic origin. Unlike the pattern observed for the chloroplast haplotypes (e.g. the association of haplotypes from Rum and Ben Bulben) there is no distinct association for any of the *Arenaria norvegica* genotypes with *A. ciliata* AFLP genotypes. The similar patterns observed for the AFLP genotypes and the ITS sequences may reflect the predominately nuclear nature of AFLP markers.

In contrast to the clear phylogeographical structuring observed for the AFLP genotypes of *Arenaria ciliata*, there is no clear internal structure of groups within *A. norvegica*. Branch lengths appear to be quite short in the NJ and NN analyses, indicating that genetic variation may be restricted between individuals of this species. The branch lengths are longer however than those observed for the ITS region and the numerous polymorphic AFLP markers provide a clearer resolution between *Arenaria ciliata* and *A. norvegica*.

Populations which have recently colonised can be detected using AFLP markers where there is little genetic diversity, weak structure and few private fragments in populations (Ehrich *et al.*, 2007; Schonswetter, Tribsch and Niklfeld, 2004*a*). The short branch lengths observed in *Arenaria norvegica* and low number of private fragments demonstrates a pattern consistent with the leading edge colonization model of post glacial colonization where reduced genetic diversity occurs due to the recolonisation of an area from an individual population and repeated

bottlenecks result in a loss of genetic diversity Hewitt (1996, 1999). This pattern using AFLPs has also been observed in other Arctic plant species whose current range may span vast areas in the Arctic and North Atlantic region (Ehrich *et al.*, 2007; Eidesen *et al.*, 2007; Skrede *et al.*, 2009, 2006).

Based on the analysis of AFLP markers it is clear that *Arenaria norvegica*, although closely related to *A. ciliata*, forms a distinct group which appears to be of more recent origin than *A. ciliata*, based on the branch lengths and the lack of private fragments in these populations. The partition of genetic diversity between and among populations can be explained by the breeding system (Hamrick and Godt, 1996) and the patterns observed in the analysis of molecular variance (AMOVA) indicate that both species are primarily outcrossing indicated by high levels of within population diversity. When both taxa are included in the AMOVA, there is a higher among population variation indicated in the majority of the datasets. As each taxa is analyzed individually, the proportion of within population variation increases. This demonstrates that there is reduced gene flow events between the two taxa, and the majority of the genetic variation is accounted for in both taxa by within population variation.

There is a higher within population variation in *Arenaria norvegica* and this corresponds to the high similarity seen between genotypes of *A. norvegica*. Although populations are separated in some cases by relatively large distance (e.g. Iceland versus The Burren), this only demonstrates the capacity of Arctic species for long distance dispersal that occurs in many arctic species even which may have presumed lack of long distance dispersal adaptations (Ehrich *et al.*, 2007; Schonswetter, Tribsch and Niklfeld, 2004*b*; Tribsch *et al.*, 2002; Westergaard *et al.*, 2011).

5.3.2 Grouping of Irish *Arenaria ciliata* AFLP genotypes

Structuring of *Arenaria ciliata* AFLP genotypes into distinctive groups corresponding to their geographic origin is evident in the majority of the datasets and using all methods of analyses (NJ, NN, PCoA). The Irish Ben Bulben genotypes consistently form a distinct cluster in each of the datasets and when more individuals are included from a broader geographic range, the Irish genotypes group

with the Svalbard genotypes. The distinct phylogeographical structuring of the groups are indicative of older pre-glacial populations, unlike recent post-glacial colonizers which contain very little structure (Ehrich *et al.*, 2007; Skrede *et al.*, 2009; Schonswetter, Tribsch and Niklfeld, 2004*a*; Westergaard *et al.*, 2011).

According to the NJ analysis, the Ben Bulben group is closely related to the *Arenaria norvegica* group. This is also evident for both nuclear ITS and chloroplast sequences where individuals of *A. norvegica* consistently form a monophyletic group with the *A. ciliata* individuals from Ben Bulben (maximum likelihood analysis - Chapter 3 and the Bayesian analysis - Chapter 4). This pattern however is less evident for the AFLP genotypes in the PCoA and neighbour-net analyses where the two taxa are separated more distinctly. More complex evolutionary processes may not be well described by classical tree building methods (Huson and Bryant, 2006) so networks can therefore be used to help explore conflicting signal (Meudt and Clarke, 2007) and may be particularly useful in the case of *Arenaria ciliata* and *A. norvegica* which display complicated phylogeographic and taxonomic histories.

The presence of private fragments can provide useful information on biogeographic questions such as the identification of refugial areas. Private fragments are evident in the Ben Bulben and Svalbard populations, the occurrence of these fragments are nearly as high as the private fragments occurring in the continental European populations in some of the datasets. The presence of fragments exclusive to one population may indicate old divergence and refugial populations as rare markers are expected to accumulate in isolated populations (Schonswetter, Tribsch and Niklfeld, 2004*b*; Schonswetter and Tribsch, 2005; Skrede *et al.*, 2009; Tribsch *et al.*, 2002). The occurrence of private fragments, in both the *Arenaria ciliata* Ben Bulben and Svalbard populations may therefore indicate the possibility of a refugial population for these two northerly populations.

5.3.3 Grouping of the continental European *Arenaria ciliata* AFLP genotypes

Each of the continental European populations contain clear phylogeographic structuring which contain a high proportion of private fragments in each population. The only exception to this is the Swiss population, which is grouped with individuals from Spain. This distinct grouping is clear in all analyses and particularly in the NJ trees and NN networks. The PCoA also demonstrates clear groups based on geographic origin, however in the NUIM 2012 dataset, there is a continuum of genotypes (or broad grouping) between the Irish/Svalbard genotypes and East and West Alps genotypes. This could be an indication of a source population for the Irish/Svalbard group, or on the other hand the PCoA may not indicate common ancestry. This is not clear using the NN or NJ analysis so it is difficult to infer a source population using these AFLP fragments alone.

The presence of a high number of private fragments, particularly in the West and East Alps is not surprising and is indicative of refugial populations with high levels of genetic diversity and private fragments. Many regions in the Alps have been proposed as hotspots for diversity and endemism (Tribsch and Schonswetter, 2003) and have been demonstrated for many species as key locations for Pleistocene refugia (Ehrich *et al.*, 2007; Huck *et al.*, 2009; Ronikier *et al.*, 2008; Schonswetter *et al.*, 2003; Schonswetter, Tribsch and Niklfeld, 2004*b*; Schonswetter *et al.*, 2005; Stehlik, 2003; Tribsch *et al.*, 2002). However the survival of species throughout Pleistocene glacial cycles, particularly in the interior of the Alps, is still species dependent and even for presumed endemic species, intraspecific genetic structure and genetic diversity can reveal patterns revealing recent postglacial immigration (Schonswetter *et al.*, 2002).

5.3.4 Conclusion

There is a distinct separation of the two taxa, *Arenaria ciliata* and *A. norvegica*, based on the analysis of AFLP markers. The analysis of AFLP fragments can be a powerful genome wide fingerprinting technique for discriminating between closely related species as demonstrated here. The combined use of the neighbour-joining,

neighbour-net, principal coordinate analysis and analysis of molecular variance is a powerful assessment of phylogeographical structure and the inference of a relict refugial signature versus postglacial colonizers for populations. Comparison of these techniques is also very useful in the discrimination between these closely related species.

The distinct clustering and presence of private fragments in the Ben Bulben population and the association with genotypes in Svalbard indicates the possibility of long term pre-glacial survival in refugia for these two populations. In contract, the patterns observed for *Arenaria norvegica* are typical for post-glacial colonizers. The patterns observed for the AFLP genotypes are congruent for the major lineages identified in the ITS region and reflects the predominately nuclear nature of the AFLP markers.

6

General Discussion

6.1 Summary

This study focused on the investigation of the glacial history of *Arenaria ciliata* and *A. norvegica*, using a phylogeographic analysis of extant European populations of these closely related species. By implementing a range of phylogeograpical analyses using the nuclear internal transcribed spacer (ITS) sequences, chloroplast (cp) sequences and amplified fragment length polymorphism (AFLP) markers, a very complex history emerges for these two species, which appears to have been shaped by a combination of survival *in situ*, post-glacial colonization and chloroplast capture processes. Although the histories of these two species are connected and species distinctions are difficult using some of the markers implemented in this study, for the most part *A. ciliata* and *A. norvegica* both show distinct genetic groupings, and separate responses to the Pleistocene glacial cycles. *Arenaria norvegica* depicts a pattern typical of a recent post-glacial expansion from a single source population and *A. ciliata* demonstrates patterns typical of populations which survived in glacial refugia. A discussion of the major findings of this study relative to the objectives laid out in the general introduction is presented in this Chapter.

6.2 Taxonomic relationship and subspecies distinctions for *A. ciliata* and *A. norvegica*

A complicated taxonomic relationship between *A. ciliata* and *A. norvegica* is observed when analyzing either the chloroplast, nuclear ITS or AFLP markers used in this study. Incongruent patterns observed between the chloroplast and nuclear ITS/AFLP analyses highlight the difficulties in unravelling species distinctions between these two currently recognised taxa. Some patterns are clear, however, and these may be able to shed some light on the discrimination of subspecies that are currently recognised within the *A. ciliata* complex.

Arenaria ciliata and *A. norvegica* are clearly grouped within the same monophyletic lineage for all ITS and chloroplast sequences (Figures 3.6, 3.3 and 3.9). The interspecific maximum likelihood (ML) analysis in Chapter 3 shows the monophyletic grouping of the two species within the same lineage when many other members of the Caryophyllaceae family are included. This pattern is consistent in all of the intraspecific trees for all chloroplast and nuclear sequences in both the ML analysis in Chapter 3, and the Bayesian analysis in Chapter 4. Although these two species are recognized as distinct taxa in the literature (Wyse Jackson and Parnell, 1987), the *A. norvegica* lineage is closely associated with the Irish *A. ciliata* individuals from Ben Bulben. This is clearly observed in the nuclear ITS region and AFLPs, where the individuals of *A. norvegica* group with the Irish individuals from Ben Bulben.

It is clear from the dating analysis that the divergence of the *A. norvegica* ITS nucleotypes from the Ben Bulben *A. ciliata* nucleotypes, is more recent than the divergence of other lineages within *A. ciliata*. Mean estimate for the separation of the Ben Bulben and *A. norvegica* nucleotypes based on the ITS region is estimated at less than 1 million years ago. Dates estimated from the chloroplast region for the divergence of this Ben Bulben/*A. norvegica* node include minimum mean age estimates of 0.338-1.084my for the concatenated region and 0.497- 1.505 my for the *matK* sequences.

This is consistent with the recent rapid speciation seen for many arctic plant groups where low variation between populations is a feature of recently evolved

flora. This makes widely recognized barcodes for land plants inappropriate for resolving species distinctions in arctic species (Brochmann and Brysting, 2008; Murray, 1995). The shallow lineages observed for *A. norvegica* indicated in the ML, Bayesian, NJ analysis and short branch lengths in the NN analysis, and the widespread distribution of individual haplotypes also indicate phylogeographical patterns typical for Arctic species (Hewitt, 2004). The different phylogeographical patterns observed between *A. ciliata* and *A. norvegica* may also indicate that these are separate taxa, and that perhaps the gene phylogenetic tree predates the origin of *A. norvegica*, which consequently may not reflect the extant organismal phylogeny of this species complex.

Results from the AFLP analyses depict a pattern that reflects that observed for the ITS nuclear sequences. Although the taxonomic relationship between *A. ciliata* and *A. norvegica* is quite complicated in the chloroplast region, the AFLP markers show distinct groupings that separate these species. However, similar to the pattern depicted for the ITS region, the *A. norvegica* genotypes, although distinct from the *A. ciliata* genotypes, are consistently grouped with the Irish Ben Bulben genotypes in the neighbour-joining (NJ) phylogenetic tree analyses. This is not as clear in the principal coordinate analysis (PCoA) where the *A. norvegica* genotypes consistently group together and do not overlap with the *A. ciliata* individuals. On this basis, it is difficult to disregard the separation of *A. ciliata* and *A. norvegica* as separate taxonomic species. Results from the network analysis from the neighbour-net (NN) trees also demonstrate a clear grouping for genotypes of *A. norvegica* and the long branch separating *A. norvegica* from *A. ciliata* mirrors the distinct groupings also observed in the NJ trees and PCoA.

The haplotype networks constructed emphasize large divergences in the *A. norvegica* haplotypes occuring on Rum and the haplotype that is widely distributed in the sampled populations, and highlights the difficulties that can arise using chloroplast data for inferring taxonomic relationships. According to the haplotype networks, the haplotype from Rum is distinct and is situated in the network positioned between Irish haplotypes from Ben Bulben, Spain, and not other haplotypes from *A. norvegica*. In the concatenated cp. region, a unique haplotype is observed for the Rum population, unlike any other population of *A. norvegica*. The haplotype network constructed for the *matK* region shows

less mutations occuring between the *A. norvegica* haplotypes, however, the haplotypes for the *A. norvegica* are still closely connected to the haplotypes of *A. ciliata*. The nucleotype network constructed for the ITS region shows a distinctive haplotype for the *A. norvegica* group which occurs in all of the *A. norvegica* populations.

The combination of the consistent separation of *A. norvegica* into a distinct monophyletic lineage, albeit within *A. ciliata* lineages, could suggest that *A. norvegica* can be regarded as a separate species, although very closely related to *A. ciliata*. The different glacial histories of the two taxa clearly could also suggest that although these two taxa have clear intertwined histories, they represent separate taxa. It is clear that *A. norvegica* is more recently derived than *A. ciliata*.

The molecular markers and sequences used in this study do not support the currently recognized sub-species in the literature. *Arenaria norvegica* subsp. *norvegica* and *A. norvegica* subsp. *anglica* do not demonstrate any distinct groupings or differences within the *A. norvegica* lineage that could constitute their discrimination as subspecies using ITS, chloroplast or AFLP markers. The only distinct group within *A. norvegica* was represented by the *A. norvegica* individuals from Rum which contained their own distinct haplotype according to the concatenated chloroplast region. The ITS sequences and the AFLP signatures, however, do not demonstrate any distinct grouping for the Rum individuals.

The recognition of the species *Arenaria gothica* subsp. *moehringiodes* (Wyse Jackson and Parnell, 1987) (or *A. moehringiodes* (Castroviejo *et al.*, 1990; Fior and Ola, 2007)), does not stand on the basis of the nuclear, chloroplast or AFLP markers used in this investigation and the results presented here do not warrant taxonomic status as a species distinct from *A. ciliata*. The only known genetic analysis including *A. ciliata* in a phylogenetic context is Fior *et al.* (2006) where *A. ciliata* and *A. moehringiodes* are consistently grouped together with strong bootstrap support for both the nuclear ITS sequences and chloroplast *matK* sequences, the closest sister species indicated as *A. nevadensis*. However, as not every species from the *A. ciliata* complex was included, it is impossible to infer whether the individuals sampled in this study for *A. ciliata* and *A. moehringioides* are distinct taxa.

Finally the polyphyletic placement of the individuals from Svalbard using the chloroplast data and the distinct grouping of the Svalbard AFLP genotypes with those from Ben Bulben would strongly suggest against the subspecies status of *Arenaria ciliata* subsp. *pseudofrigida* for the individuals from Svalbard. Individuals from Ben Bulben and Svalbard have been noted as being cytologically and chemically very similar (Wyse Jackson and Parnell, 1987). Wyse Jackson and Parnell (1987) note that an individual *A. ciliata* plant sampled from Ben Bulben had the same chromosome count as an individual from Svalbard (2n=40) Horn (1948) and almost an identical pattern of leaf flavonoids. This is certainly in agreement with the results of this study, where the consistent grouping of the Ben Bulben and Svalbard individuals is particularly evident for the AFLP genotypes.

6.3 Phylogeographical patterns and re-evaluation of the "The Irish Question"

Distinct phylogeographic histories are evident for *A. ciliata* and *A. norvegica*, based on the ML, molecular clock, and AFLP analyses implemented in this study. There is no strong evidence for an obvious post-glacial migration route from a source population for either species. The results, however, suggest a strong possibility that the individuals from Ben Bulben represent a population more ancient than recent glaciation events. This is in contrast to the phylogeographical pattern depicted for *A. norvegica*, where a more recent expansion from Southern Refugia for the populations are inferred, with the exception of Rum.

6.3.1 Phylogeographical evidence of a glacial relict in Ben Bulben

Multiple lines of evidence from this study indicate the possibility that at least part of the *A. ciliata* complex from Ben Bulben are glacial relicts which were present in Ireland well before the LGM (last glacial maximum), and have possibly persisted through several glacial cycles in a refugium close to, or in Ireland. The times of

divergence, and patterns observed for the clustering of *matK* haplotypes, indicate that plant individuals arrived at different times, possibly by different routes via multiple colonizations.

Clustering of the Irish concatenated cp haplotypes into a distinct lineage, with multiple unique haplotypes comparable to the number of haplotypes found in meta-populations in continental Europe, demonstrates the strong likelihood that *A. ciliata* did not arrive into Ireland after the LGM <15,000 bp. It is particularly evident in the network for the concatenated cp haplotypes how distinct the Irish haplotypes are. In this analysis, all Irish haplotypes are separated from their next closely related haplotypes in Rum at 13bp, and Picos De Europa at 19bp. The presence of a high proportion of private AFLP fragments in the Ben Bulben population, similiar to their continental European counterparts, is also indicative of an older population, compared to the few, or no, private AFLP fragments observed in *A. norvegica* populations (except Rum).

If individuals of *A. ciliata* arrived to Ireland after the LGM, they would not have had sufficient time to accumulate mutations to give rise to the unique haplotypes evident in this population. In contrast, genetic uniformity for individuals and populations of *A. norvegica*, with little variation between the individuals observed in most populations, is indicative of a recent post glacial migration and rapid range expansion (Schaal *et al.*, 1998).

The distinct and ancient separation of the Irish haplotypes from Ben Bulben is particularly striking, where dating analysis indicates the separation of the haplotypes between 1.343 and 5.137 million years ago based on the *matK* data. Divergence of this population therefore predates the origin for the *A. norvegica* and *A. ciliata* haplotypes from Picos De Europa and the Pyrenees. This pattern is also clearly evident in the concatenated cp region (in both ML and Bayesian analyses). This data set arguably depicts a more accurate phylogeographical inference than the single *matK* region, as more loci are analyzed from the chloroplast genome offering more reliability on the inference of the distribution of haplotypes and the presence of unique haplotypes.

Evidence from this study for the possibility of *in situ* survival of *A. ciliata* in Ireland are consistent with many lines of evidence based on phylogeographical analyses (Davison, 2000; Hoarau *et al.*, 2007; McKeown *et al.*, 2010; Provan *et al.*,

2005; Teacher *et al.*, 2009), disjunctly distributed plant species (Forbes, 1846; Praeger, 1932; Synnott, 1981; Webb, 1983), pollen analysis (Colhoun *et al.*, 1972), plant fossil record (Coxon and Waldren, 1995) and evidence for ice free nunataks (Coxon, 1988).

6.3.1.1 Postglacial expansion of *A. norvegica*

The phylogeographic pattern depicted for *A. norvegica* is, except in the case of Rum, typical of recent postglacial expansion. This can be inferred from the low number of unique haplotypes, the widespread distribution of a single haplotype/nucleotype and therefore low geographic structuring, low branch lengths observed in the phylogenetic trees and networks and low or no private AFLP fragments.

The traditional consensus regarding the mobility of arctic plants is that they had restricted capabilities for long distance dispersal, and that plants lost their dispersal capacity after repeated isolation and depauperation in small refugial populations (Hulten, 1937). However, recent phylogeographical studies have demonstrated that arctic flora are highly mobile and long-distance dispersal occurs at much higher rates than previously thought from many source regions (Brochmann and Brysting, 2008). This high dispersal ability has also been observed in the fossil record which suggests rapid colonisation following climatic change (Birks, 2008). Phylogeographical studies which confirm the long distance dispersal capacity of arctic plants include Alsos *et al.* (2007) who demonstrate frequent long distant colonization from several source regions into Svalbard, where the predominant source was the most distant region of North-West Russia.

6.3.1.2 The phylogeographic association of the Ben Bulben, Svalbard and Rum populations and *in situ* survival

Evidence from the phylogeographic analyses depicts a clear similarity for the individuals of *A. ciliata* from Ben Bulben and Svalbard. This is particularly evident for the grouping of the AFLP genotypes, where the Ben Bulben and Svalbard populations are consistently grouped together in the NJ, NN and PCo analyses.

In addition to this, the Irish and Svalbard populations share a *matK* haplotype, and a similar biogeographic pattern is found for the grouping of the Irish and Svalbard haplotyes. The haplotypes in Svalbard cluster into two divergent lineages, similar to those found in the Irish Ben Bulben population, where the divergence of these haplotypes occurred at the same time for the Irish *matK* haplotypes, at least 1 million years ago. This could suggest that the origin of the Ben Bulben haplotypes share a common history with the Svalbard population.

This kind of separation event leading to the divergence of these haplotypes into two lineages, where one genotype is unique to Ireland and another belongs to a widespread European distribution, has also been documented in other species such as the common frog (Teacher *et al.*, 2009), the red seaweed *Palmaria palmata* (Provan *et al.*, 2005), and the brown trout *Salmo trutta* (McKeown *et al.*, 2010). Other studies such as the phylogeography of the brown seaweed *Fucus serratus*, which indicate evidence for a refugium in Ireland, also indicate deep divergence separating two types of Irish haplotypes (Hoarau *et al.*, 2007).

Notably, the *matK* haplotype which is common to Ben Bulben, Svalbard and Rum, is also distributed in the West and East Alps, which contain large mountain ranges that remained unglaciated during the LGM (Schonswetter *et al.*, 2005). Both the East and West Alps are documented to contain refugia based on phylogeographical studies (Schonswetter *et al.*, 2002; Schonswetter, Tribsch, Stehlik and Niklfeld, 2004; Schonswetter *et al.*, 2005). The high diversity of unique haplotypes is very evident for the West and East Alps with a very high proportion of private AFLP markers. In the present study, the West Alps was the only region to contain multiple *A. ciliata* nucleotypes for the ITS region and showed considerably relatively high diversity for this region. In addition to this, the highly divergent *matK* haplotypes which are common to the East and West Alps, demonstrate the most divergent grouping in the ML analysis, and haplotypes from the West Alps for the concatenated cp region and ITS belong to the oldest lineage of *A. ciliata* haplotypes/nucleotypes.

The position of the haplotype in a network may also give an indication to the age of the haplotype, where more ancient haplotypes are positioned at internal nodes (Schaal *et al.*, 1998). This is evident in the haplotype network for *matK*,

where the haplotype common to Svalbard, Ireland, Rum and the Alps are positioned at internal nodes. The molecular clock analysis also depicts an older age estimate for this haplotype than the other Ben Bulben *matK* haplotype present in the other deeply divergent lineage which is present in the East Alps, Picos and the Pyrenees. The age of this haplotype also coincides with the divergence of *A. norvegica* haplotypes, which predate haplotypes from Picos De Europa and the Pyrenees. Dates estimated for this node range from a minimum mean age of 338,000 years for the concatenated cp haplotypes, to a maximum age estimate of 1.084 my for the *matK* region.

It is clear from the results in this study that there is a link between the *A. ciliata* and *A. norvegica* populations in Svalbard, Ireland and Rum. There are many lines of evidence which suggest the possibility that these populations belonged to a cryptic refugium; (i) The presence of unique haplotypes in each population, (ii) age estimates for the haplotypes derived from these populations, (iii) chloroplast capture events observed for Ben Bulben and Rum, (iv) the similar patterns for the separation of haplotypes into two very divergent lineages of Svalbard and Ireland, and (v) the presence of private AFLP markers in each population, (vi) the presence of the haplotype common to these populations in other regions which are documented to contain refugia and, (vi) the concordance of a separation event between two divergent haplotypes occurring in the same population observed in Svalbard, Ireland and this divergent event is also documented for other species (Hoarau *et al.*, 2007; McKeown *et al.*, 2010; Provan *et al.*, 2005; Teacher *et al.*, 2009).

It is becoming apparent that some species may not reflect the classical temperate model of postglacial colonization leading to lower genetic diversity as the result of bottleneck events, and that present communities could be a mix of those that migrated from Southern refugia, supplemented by cryptic refugia populations during the LGM (Stewart and Lister, 2001). Dahl (1946) suggests the possibility of refugia of a coastal mountain type in Western and North-Western Scandanavia and in areas such as Scotland, Iceland, and the Southern half of Greenland. Dahl also postulates the presence of a tundra type refugia which would have included areas of Siberia, Northern Norway and Svalbard. Dahl (1946) suggests the possibility of glacial relicts which he refers to as 'tundra refugees' and includes *Arenaria*

ciliata subsp. *pseudofrigida* Ostf. & Dahl in this group, based on current geographic distributions of the populations. A recent phylogeographical study supports the claims made by Dahl (1946), which demonstrates the survival *in situ* of two arctic-alpine plants, *Sagina caespitosa* and *Arenaria humifusa* in Svalbard (Westergaard *et al.*, 2011).

Phylogeographical studies indicate distinctiveness of some Scottish plant populations and the potential that some species survived in refugia such as Scots pine (Kinloch *et al.*, 1986; Prus-Glowacki *et al.*, 2012; Sinclair *et al.*, 1998) and the arctic-montane plant *Saxifraga hirculus* (Oliver *et al.*, 2006). Pollen analysis also indicate the possibility of refugia in the Outer Hebrides, the North West of Skye and in the West central part of Ireland (Huntley and Birks, 1983).

Evidence for the occurrence of nunataks (ice-free mountain summits) during the LGM has been accumulating in areas such as Scotland and Svalbard (Ballantyne *et al.*, 1998; Brochmann *et al.*, 2003). Ballantyne *et al.* (1998) proposes that a large number of nunataks occurred in North West Scotland during the LGM, and these areas could perhaps account for the persistence of species and explain the genetic uniqueness found in these Scottish populations (Kinloch *et al.*, 1986; Prus-Glowacki *et al.*, 2012; Oliver *et al.*, 2006; Sinclair *et al.*, 1998). Westergaard *et al.* (2011) attribute the persistence of the arctic-alpines *Sagina caespitosa* and *Arenaria humifusa* in Svalbard throughout the LGM to the *in situ* survival on nunataks.

6.3.1.3 Suture zones and chloroplast capture

Contact zones between species or populations, also known as suture zones, may result in hybridization and can occur because of glacial refugia, histories of post-glacial expansion and physical impediments to dispersal (Avise, 2009). Incongruence between the chloroplast and nuclear regions not only suggest ancient hybridization events between *A. ciliata* and *A. norvegica*, but also indicates a previous widespread distribution of these species where there was overlap. Currently, the only region in which there is geographical overlap between *A. ciliata* and *A. norvegica* is Ireland, but the present study found no evidence that these populations were interbreeding at any time.

Chloroplast capture events, where the genomes of chloroplasts are laterally transmitted between species (Riesberg and Soltis, 1991), are evident in *A. ciliata* individuals from the Pyrenees which contain the chloroplast signature from *A. norvegica*, and in individuals from Rum, which contains a higher frequency of the *A. ciliata matK* cp haplotype, than the *A. norvegica* chloroplast haplotype. This might suggest the possibility that individuals from Ben Bulben migrated to Rum, before the LGM at a time when this species was more widely distributed in Europe. This could have led to the formation of hybrids, and subsequent pollen swamping from individuals of *A. norvegica*, to individuals of *A. ciliata* which contained the chloroplast signature from Ben Bulben. This would also explain the nuclear signature for the individuals from Rum, which clearly contain the nuclear signature of *A. norvegica*, indicated in the ITS and AFLP analyses. The presence of the widespread *A. norvegica* haplotype in the Pyrenees also indicates the possibility for ancient hybridization between *A. ciliata* individuals from the Pyrenees and *A. norvegica* and could also be indicative of a previous suture zone along the Celtic Fringe of the Atlantic, during a time when populations of these species were more widely distributed before the LGM.

Again, this reinforces the different histories depicted for these two species, where ancient hybridization events demonstrate previous geographical overlapping ranges that were different from today, and that disregard current 'species distinctions'.

6.3.2 Concluding remarks

The results of this study demonstrate the advantages of sampling a wide range of genomic markers and comparing both nuclear and chloroplast loci. The analyses of these regions have allowed: the inference of chloroplast capture events between species, analysis of recently evolved species with low variation, taxonomic discrimination between closely related species, molecular clock dating analysis and the inference of postglacial colonization versus *in situ* survival through the LGM.

Multiple lines of evidence indicate the possibility of *A. ciliata* populations surviving the last glacial cycle in a refugium in or near Ireland. The dating analysis indicates comparatively early divergence of the haplotypes on Ben Bulben,

from European populations, predating the LGM. Consensus found among the ML, Bayesian and AFLP analyses, for the distinctiveness of the Ben Bulben *A. ciliata* population further strengthen these results.

Consensus was found for phylogeographic patterns depicted in the populations in Rum and Svalbard and these demonstrate the presence of cryptic refugia which could have facilitated the survival of *A. ciliata* and *A. ciliata* throughout the Pleistocene. Incongruence between the nuclear and chloroplast signatures demonstrates possible suture zones between *A. ciliata* and *A. norvegica*, and hybridization events between populations from the Pyrenees, Ben Bulben and Rum.

There are clear taxonomic difficulties between these species and subspecies distinctions are impossible to make based on the molecular markers and analyses used in this study. There is genetic structuring for the individuals of *A. norvegica* which is very clear in the ITS and AFLP analyses, and results from this study show that although very closely related, *A. ciliata* and *A. norvegica* can be separated as distinct taxa. Further analysis of this species complex, sampled from a broader geographic distribution would certainly aid in the understanding of this taxonomically difficult species complex.

Webb (1983) sums up the glacial history for *A. ciliata* based on geographical distribution of its populations, and thinks it "inconceivable that Irish form arrived there since the last glaciation". Despite this, it is possible that climatic conditions in Ireland were so harsh that, during the LGM, only very hardy organisms could have survived (Bennike, 1999). Evidence now places the extent of the LGM ice sheet limit at least 25 km off the current shoreline of Ireland (Ballantyne *et al.*, 2008). However, there is also the possibility that Ben Bulben could have acted as a nunatak, remaining ice-free above the landscape of the ice-sheet (Coxon, 1988; Synnott, 1981). There still seems to be a direct conflict, however, between the genetic and geographic distribution patterns of current biota, and the geological processes governing the glacial cycles in Ireland (Clark *et al.*, 2012; O' Cofaigh *et al.*, 2012). It is difficult to understand how the persistence of plants or animals, however 'hardy', could have occurred if the majority of Ireland was covered by a thick ice-sheet. These conflicts can perhaps be best summed up by Willmott (1930) : "the ice was there, but the plants with definitely 'relict'

type of distribution are there, and while the supposition of extinction is only a deduction, the existence of the plants is a fact“.

Bibliography

Abbott R (2000). Molecular analysis of plant migration and refugia in the Arctic, *Science* **289**, 1343–1346.

Acosta M and Premoli A (2010). Evidence of chloroplast capture in *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae), *Molecular Phylogenetics and evolution* **54**, 235–242.

Alsos I, Eidesen P, Ehrich D, Skrede I, Westergaard K, Jacobsen G, Landvik J, Taberlet P and Brochmann C (2007). Frequent long-distance plant colonization in the changing Arctic, *Science* **316**, 1606–1609.

Arbogast B, Edwards S, Wakeley J, Beerli P and Slowinski J (2002). Estimating divergence times from molecular data on phylogenetic and population genetic timescales, *Annual Review of Ecology and Systematics* **33**, 707–740.

Avise J (1998). The history and purview of phylogeography: a personal reflection, *Molecular Ecology* **7**, 371–379.

Avise J (2000). *Phylogeography. The history and formation of species.*, Harvard University Press, Cambridge, Massachusetts.

Avise J (2009). Phylogeography: retrospect and prospect, *Journal of Biogeography* **36**, 3–15.

Avise J, Arnold J, Ball R, Bermingham E, Lamb T, Neigel J, Reeb C and Saunders N (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics, *Annual Review of Ecology and Systematics* **18**, 489–522.

Baldwin B, Sanderson M, Porter J and Wojciechowski M (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny, *Annals of the Missouri Botanical Garden* **82**, 247–277.

Ballantyne C, Mccarroll D, Nesje A, Dahl S, Stone J and Fifield L (1998). High-resolution reconstruction of the last ice sheet in NW Scotland, *Terra Nova* **10**, 63–67.

Ballantyne C, Stone J and McCarroll D (2008). Dimensions and chronology of the last ice sheet in Western Ireland, *Quaternary Science Reviews* **27**, 185–200.

Beatty G and Provan J (2012). Post-glacial dispersal, rather than in situ glacial survival, best explains the disjunct distribution of the Lusitanian plant species *Daboecia cantabrica* (Ericaceae), *Journal of Biogeography* **40**, 335–344.

Bennett K (1984). Post-glacial history of *Pinus sylvestris* in the British Isles, *Quaternary Science Reviews* **3**, 133–155.

Bennett K and Provan J (2008). What do we mean by 'refugia'?, *Quaternary Science Reviews* **27**, 2449–2455.

Bennike O (1999). Colonisation of Greenland by plants and animals after the last ice age: a review, *Polar Record* **35**, 323–336.

Bensch S and Akesson M (2005). Ten years of AFLP in ecology and evolution: why so few animals?, *Molecular ecology* **14**, 2899–2914.

Bermingham E and Moritz C (1998). Comparative phylogeography: concepts and applications, *Molecular Ecology* **7**, 367–369.

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

Birks H (1989). Holocene isochrone maps and patterns of tree-spreading in the British Isles, *Journal of Biogeography* **16**, 503–540.

Birks H (2008). The Late-Quaternary history of arctic and alpine plants, *Plant Ecology & Diversity* **1**, 135–146.

Birks H and Willis K (2008). Alpines, trees, and refugia in Europe, *Plant Ecology & Diversity* **1**, 147–160.

Birky C (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms and models, *Annual Review of Genetics* **35**, 125–48.

Bittrich V (1993). *The Families and Genera of Vascular Plants, vol. II, Flowering plants, dicotyledons, Magnoliid, Hamamelid and Caryophyllid Families*, Springer-Verlag, Berlin, Germany, chapter Caryophyll, pp. 206–236.

Borsch T and Quandt D (2009). Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA, *Plant Systematics and Evolution* **282**, 169–199.

Britten R (1986). Rates of DNA sequence evolution differ between taxonomic groups, *Science* **231**, 1393–1398.

Brochmann C and Brysting A (2008). The Arctic - an evolutionary freezer?, *Plant Ecology & Diversity* **1**, 181–195.

Brochmann C, Gabrielsen T, Nordal I, Landvik J and Elven R (2003). Glacial survival or tabula rasa? The history of North Atlantic biota revisited, *Taxon* **52**, 417.

Brooks A, Bradley S, Edwards R, Milne G, Horton B and Shennan I (2008). Postglacial relative sea-level observations from Ireland and their role in glacial rebound modelling, *Journal of Quaternary Science* **23**, 175–192.

Buckler E, Ippolito A and Holtsford T (1997). The evolution of ribosomal DNA: divergent paralogues and phylogenetic Implications, *Genetics* **145**, 821–832.

Bussell J, Waycott M and Chappill J (2005). Arbitrarily amplified DNA markers as characters for phylogenetic inference, *Perspectives in Plant Ecology, Evolution and Systematics* **7**, 3–26.

Castroviejo S, Lainz M, Lopez Gonzalez G, Montserat P, Garmendia M, Paiva J and Villar L, eds (1990). *Flora Iberica, Volume II.*, Servicio de Publicaciones del CSIC, Madrid, Madrid.

Chung M, Gelembiuk G and Givnish T (2004). Population genetics and phylogeography of endangered *Oxytropis campestris* var. *chartacea* and relatives: arctic-alpine disjuncts in eastern North America, *Molecular Ecology* **13**, 3657–3673.

Clark C, Hughes A, Greenwood S, Jordan C and Sejrup H (2012). Pattern and timing of retreat of the last British-Irish Ice Sheet, *Quaternary Science Reviews* **44**, 112–146.

Clement M, Posada D and Crandall K (2000). TCS: a computer program to estimate gene genealogies, *Molecular ecology* **9**, 1657–1659.

CLIMAP (1976). The surface of the Ice-Age Earth, *Science* **191**, 1131–1137.

Colhoun E, Dickson J, McCabe A and Shotton F (1972). A Middle Midlandian freshwater series at Derryvree, Maguiresbridge, County Fermanagh, Northern Ireland, *Proceedings of the Royal Society of London B.* **180**, 273–292.

Comes P and Kadereit J (1998). The effect of Quaternary climatic changes on plant distribution and evolution, *Trends in Plant Science* **3**, 432–438.

Coxon P (1988). Remnant periglacial features on the summit of Truskmore, Counties Sligo and Leitrim, Ireland, *Z.Geomorph.N.F* **71**, 81–91.

Coxon P and Waldren S (1995). The floristic record of Ireland's Pleistocene temperate stages, *in Island Britain: a Quaternary perspective*, pp. 243–267.

Cuenoud P, Savolainen V, Chatrou L, Powell M, Grayer R and Chase M (2002). Molecular phylogenetics of Caryophyllales based on nuclear 18S RDNA and plastid RBCL, ATPB, and MatK DNA sequences, *American Journal of Botany* **89**, 132–144.

Curtis T and McGough H (1998). *The Irish Red Data Book 1 Vascular Plants*, The Stationary Office, Dublin.

Dahl E (1946). On different types of unglaciated areas during the ice ages and their signifance to phytogeography, *New Phytologist* **45**, 225–242.

Darwin C (1859). *On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life*, Murray, London.

Davison A (2000). An east-west distribution of divergent mitochondrial haplotypes in British populations of the land snail, *Cepaea nemoralis* (*Pulmonata*), *Biological Journal of the Linnean Society* **70**, 697–706.

Devoy R (1985). The Problem of Late Quaternary land bridges between Britain and Ireland, *Quaternary Science Reviews* **4**, 43–58.

Dixon C, Schonswetter P, Vargas P, Ertl S and Schneeweiss G (2009). Bayesian hypothesis testing supports long-distance Pleistocene migrations in a European high mountain plant (*Androsace vitaliana*, Primulaceae), *Molecular Phylogenetics and Evolution* **53**, 580–91.

DL S (2002), 'PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)'.

Donoghue P and Benton M (2007). Rocks and clocks: calibrating the Tree of Life using fossils and molecules, *Trends in Ecology & Evolution* **22**, 424–31.

Doolittle R, Feng D, Tsang S, Cho G and Little E (1996). Determining divergence times of the major kingdoms of living organisms with a protein clock, *Science* **271**, 470–477.

Doyle J (1992). Gene trees and species trees: molecular systematics as one-character taxonomy, *Systematic Botany* **17**, 144–163.

Doyle J J and Doyle J (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue, *Phytochemical Bulletin* **19**, 11–15.

Drummond A, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones Havas S, Sturrock S, Thierer T and Wilson A (2010), 'Geneious version 5.3.6'.
URL: *http://www.geneious.com*

Drummond A, Ho S, Phillips M and Rambaut A (2006). Relaxed phylogenetics and dating with confidence, *Public Library of Science Biology* **4**, 699–710.

Drummond A and Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees, *BMC evolutionary biology* **7**, 214–222.

Ehrich D (2006). aflpdat: a collection of r functions for convenient handling of AFLP data, *Molecular Ecology Notes* **6**, 603–604.

Ehrich D, Gaudeul M, Assefa A, Koch M, Mummenhoff K, Nemomissa S, Consortium I and Brochmann C (2007). Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains, *Molecular Ecology* **16**, 2542–59.

Eidesen P, Alsos I, Popp M, Stensrud O, Suda J and Brochmann C (2007). Nuclear vs. plastid data: complex Pleistocene history of a circumpolar key species., *Molecular Ecology* **16**, 3902–3925.

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

Fay J and Wu C (1999). A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation, *Molecular Biology and Evolution* **16**, 1003–1005.

Felsenstein J (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach, *Journal of Molecular Evolution* **17**, 368–376.

Fernandez-Mazuecos M and Vargas P (2011). Historical isolation versus recent long-distance connections between Europe and Africa in bifid toadflaxes (*Linaria* sect. Versicolores), *PloS one* **6**, e22234.

Fernandez Mazuecos M and Vargas P (2012). Congruence between distribution modelling and phylogeographic analyses, *New Phytologist* **198**, 1274–1289.

Fior S, Karis P, Casazza G, Minuto L and Sala F (2006). Molecular phylogeny of the Caryophyllaceae (Caryophyllales) inferred from chloroplast MatK and nuclear rDNA ITS sequences, *American Journal of Botany* **93**, 399–411.

Fior S and Ola P (2007). Phylogeny, evolution and systematics of *Moehringia* (Caryophyllaceae) as inferred from molecular and morphological data: a case of homology reassessment, *Cladistics* **23**, 362–372.

Forbes E (1846). On the connexion between the distribution of the existing fauna and flora of the British Isles, and the geological changes which have affected their area, especially during the epoch of the Northern Drift., *Mem. Geol. Surv.* **1**, 336–432.

Frajman B, Eggens F and Oxelman B (2009). Hybrid origins and homoploid reticulate evolution within *Heliosperma* (Sileneae, Caryophyllaceae)-a multigene phylogenetic approach with relative dating, *Systematic biology* **58**, 328–45.

Graur D and Martin W (2004). Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision, *Trends in Genetics* **20**, 80–6.

Gugerli F and Holderegger R (2001). Nunatak survival, tabula rasa and the influence of the Pleistocene ice-ages on plant evolution in mountain areas, *Trends in Plant Science* **6**, 397–398.

Guo Y, Ehrendorfer F and Samuel R (2004). Phylogeny and systematics of *Achillea* (Asteraceae-Anthemideae) inferred from nrITS and plastid trnL-F DNA sequences, *Taxon* **53**, 657–672.

Gysels E, Hellemans B, Pampoulie C and Volckaert F (2004). Phylogeography of the common goby, *Pomatoschistus microps*, with particular emphasis on the colonization of the Mediterranean and the North Sea, *Molecular Ecology* **13**, 403–417.

Hamill R, Doyle D and Duke E (2006). Spatial patterns of genetic diversity across European subspecies of the mountain hare, *Lepus timidus* L., *Heredity* **97**, 355–365.

Hamilton M (1999). Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation, *Molecular Ecology* **8**, 513–525.

Hamrick J and Godt M (1996). Effects of life history traits on genetic diversity in plant species, *Philosophical Transactions of the Royal Society B.* **351**, 1291–1298.

Hare M (2001). Prospects for nuclear gene phylogeography, *Trends in Ecology & Evolution* **16**, 700–706.

Hasegawa M, Kishino H and Yano T (1989). Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in Hominoidea, *Journal of Human Evolution* **18**, 461–476.

Hedges S, Parker P, Sibley C and Kumar S (1996). Continental breakup and the ordinal diversification of birds and mammals, *Nature* **381**, 226–229.

Hewitt G (1996). Some genetic consequences of ice ages, and their role in divergence and speciation, *Biological Journal of the Linnean Society* **58**, 247–276.

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

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

Hewitt G (2001). Speciation, hybrid zones and phylogeography - or seeing genes in space and time, *Molecular Ecology* **10**, 537–549.

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

Ho S (2007). Calibrating molecular estimates of substitution rates and divergence times in birds, *Journal of Avian Biology* **38**(4), 409–414.

Ho S and Larson G (2006). Molecular clocks: when times are a-changin', *Trends in Genetics* **22**, 79–83.

Ho S and Phillips M (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times, *Systematic Biology* **58**, 367–80.

Hoarau G, Coyer J, Veldsink J, Stam W and Olsen J (2007). Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*, *Molecular Ecology* **16**, 3606–3616.

Holder M and Lewis P (2003). Phylogeny estimation: traditional and Bayesian approaches, *Nature Reviews Genetics* **4**, 275–284.

Horn K (1948). Chromosome numbers of Northern plant species, *Rit LandbDeild.* **50**.

Huck S, Budel B, Kadereit J W and Printzen C (2009). Range-wide phylogeography of the European temperate-montane herbaceous plant *Meum athamanticum* Jacq.: evidence for periglacial persistence, *Journal of Biogeography* **36**, 1588–1599.

Huelsenbeck J and Crandall K (1997). Phylogeny estimation and hypothesis testing using maximum likelihood, *Annual Review of Ecology and Systematics* **28**, 437–466.

Huelsenbeck J, Ronquist F, Nielsen R and Bollback J (2001). Bayesian inference of phylogeny and its impact on evolutionary biology, *Science* **294**, 2310–2314.

Hulten E (1937). *Outline of the history of arctic and boreal biota during the Quaternary period*, Aktiebolaget Thule, Stockholm.

Huntley B and Birks H (1983). *An Atlas of Past and Present Pollen Maps for Europe: 0-13,000 years ago*, Cambridge University Press, Cambridge.

Huson D and Bryant D (2006). Application of phylogenetic networks in evolutionary studies, *Molecular Biology and Evolution* **23**, 254–67.

Jalas J and Suominen J (1983). *Atlas Florae Europaeae: Distribution of Vascular Plants in Europe, Volume 6: Caryophyllaceae (Alsinoideae and Paronychioideae)*, Cambridge University Press, Cambridge.

Johnson L and Soltis D (1994). matK DNA Sequences and Phylogenetic Reconstruction in Saxifragaceae s. str., *Systematic Botany* **19**, 143–156.

Jolly M, Viard F, Gentil F, Thiebaut E and Jollivet D (2006). Comparative phylogeography of two coastal polychaete tubeworms in the Northeast Atlantic supports shared history and vicariant events, *Molecular Ecology* **15**, 1841–1855.

Jones C, Edwards K, Castaglione S, Winfield M, Sala F, Van De Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A and Karp A (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories, *Molecular Breeding* **3**, 381–390.

Jordan G and Macphail M (2003). A Middle-Late Eocene inflorescence of Caryophyllaceae from Tasmania, Australia, *American Journal of Botany* **90**, 761–768.

Kelleher C, Hodkinson T, Douglas G and Kelly D (2005). Species distinction in Irish populations of *Quercus petraea* and *Q. robur*: morphological versus molecular analyses, *Annals of botany* **96**, 1237–1246.

Kelleher C, Hodkinson T, Kelly D and Douglas G (2004). Characterisation of chloroplast DNA haplotypes to reveal the provenance and genetic structure of oaks in Ireland, *Forest Ecology and Management* **189**, 123–131.

Kim S and Donoghue M (2008). Incongruence between cpDNA and nrITS trees indicates extensive hybridization within Eupersicaria (Polygonaceae), *American Journal of Botany* **95**, 1122–35.

Kinloch B, Westfall R and Forrest G (1986). Caledonian Scots pine: origins and genetic structure, *New Phytologist* **104**, 703–729.

Kropf M, Comes H and Kadereit J (2008). Causes of the genetic architecture of south-west European high mountain disjuncts, *Plant Ecology & Diversity* **1**(2), 217–228.

Lambeck K and Purcell A (2001). Sea-level change in the Irish Sea since the Last Glacial Maximum: constraints from isostatic modelling, *Journal of Quaternary Science* **16**, 497–506.

Larkin M, G. B, Brown N, Chenna R, McGettigan, P.A., McWilliam H, Valentein F, Wallace I, Wilm A, Lopez R, Thompson J, Gibson T and Higgins D (2007). Clustal W and Clustal X version 2.0., *Bioinformatics* **23**, 2947–2948.

Lowe A, Harris S and Ashton P (2007). *Ecological Genetics. Design, Analysis and Application*, Blackwell Science Ltd, Oxford.

Maddison W and Maddison D (2011), 'Mesquite Version 2.75: a modular system for evolutionary analysis'.
URL: *http://mesquiteproject.org*

Maggs C, Castilho R, Foltz D, Henzler C, Jolly M, Kelly J, Olsen J, Perez K, Stam W, Vainola R, Viard F and Wares J (2008). Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa, *Ecology* **89**, S108–S122.

Marr K, Allen G and Hebda R (2008). Refugia in the Cordilleran ice sheet of western North America: chloroplast DNA diversity in the Arctic-alpine plant *Oxyria digyna*, *Journal of Biogeography* **35**, 1323–1334.

Martinkova N, McDonald R and Searle J (2007). Stoats (*Mustela erminea*) provide evidence of natural overland colonization of Ireland, *Proceedings of the Royal Society B. Biological Sciences* **274**, 1387–1393.

Mascheretti S, Rogatcheva M, Gunduz I, Fredga K and Searle J (2003). How did pygmy shrews colonize Ireland? Clues from a phylogenetic analysis of mitochondrial cytochrome b sequences, *Proceedings of the Royal Society B. Biological Sciences* **270**, 1593–1599.

McKeown N, Hynes R, Duguid R, Ferguson A and Prodohl P (2010). Phylogeographic structure of brown trout *Salmo trutta* in Britain and Ireland: glacial refugia, postglacial colonization and origins of sympatric populations, *Journal of Fish Biology* **76**, 319–347.

McNeill J (1962). Taxonomic studies in the Alsinoideae: I. Generic and infrageneric groups, *Notes from the Royal Botanic Garden Edinburgh* **24**, 79–155.

Meudt H and Clarke A (2007). Almost forgotten or latest practice? AFLP applications, analyses and advances, *Trends in Plant Science* **12**, 106–117.

Mitchell F J (2006). Where did Ireland's trees come from?, *Biology & Environment: Proceedings of the Royal Irish Academy* **106**, 251–259.

Mitchell F and Ryan M (2007). *Reading the Irish Landscape*, 3rd edn, Townhouse, Dublin.

Moore P (1987). Snails and the Irish question, *Nature* **328**, 381–382.

Mueller U and Wolfenbarger L (1999). AFLP genotyping and fingerprinting., *Trends in Ecology & Evolution* **14**, 389–394.

Murray D (1995). Causes of arctic plant diversity, *in Arctic and alpine biodiversity: patterns, causes and ecosystem consequences.*, Springer, Heidelberg, pp. 21–32.

NPWS (2003). *Site Synopsis of Ben Bulben, Gleniff and Glenade Complex*, Technical report.
URL: *www.npws.ie*

Nybom H (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants, *Molecular Ecology* **13**, 1143–1155.

O' Cofaigh C, Telfer M W, Bailey R M and Evans D J (2012). Late Pleistocene chronostratigraphy and ice sheet limits, southern Ireland, *Quaternary Science Reviews* **44**, 160–179.

Okuyama Y, Fujii N, Wakabayashi M, Kawakita A, Ito M, Watanabe M, Murakami N and Kato M (2005). Nonuniform concerted evolution and chloroplast capture: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae), *Molecular Biology and Evolution* **22**, 285–96.

Oliver C, Hollingsworth P and Gornall R (2006). Chloroplast DNA phylogeography of the arctic-montane species *Saxifraga hirculus* (Saxifragaceae), *Heredity* **96**, 222–31.

Olsen J, Stam W, Coyer J, Reusch T, Billingham M, Bostrom C, Calvert E, Christie H, Granger S, la Lumiere R, Milchakova N, Oudot-Le Secq M, Procaccini G, Sanjabi B, Serrao E, Veldsink J, Widdicombe S and Wyllie-Echeverria S (2004). North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina* L., *Molecular Ecology* **13**, 1923–41.

Parfrey L, Lahr D, Knoll A and Katz L (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks, *Proceedings of the National Academy of Sciences of the United States of America* **108**, 13624–9.

Paulo O, Dias C, Bruford M, Jordan W and Nichols R (2001). The persistence of Pliocene populations through the Pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard, *Proceedings of the Royal Society B. Biological Sciences* **268**, 1625–30.

Peakall R and Smouse P (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update, *Bioinformatics* **28**, 2537–9.

Petit R (2002). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence, *Forest Ecology and Management* **156**, 49–74.

Pleines T, Jakob S and Blattner F (2009). Application of non-coding DNA regions in intraspecific analyses, *Plant Systematics and Evolution* **282**, 281–294.

Posada D (2008). jModelTest: phylogenetic model averaging, *Molecular biology and evolution* **25**, 1253–1256.

Posada D and Crandall K (2001). Intraspecific gene genealogies: trees grafting into networks, *Trends in Ecology & Evolution* **16**, 37–45.

Praeger R (1932). Recent views bearing on the problem of the Irish flora and fauna. An address delivered before the Academy, 30 November, 1932, *Proceedings of the Royal Irish Academy. Section B: Biological, Geological and Chemical Science* **41**, 125–145.

Prokopenko A, Williams D, Kuzmin M, Karabanov E, Khursevich G and Peck J (2002). Muted climate variations in continental Siberia during the mid-Pleistocene epoch, *Nature* **418**, 65–68.

Provan J and Bennett K (2008). Phylogeographic insights into cryptic glacial refugia, *Trends in Ecology & Evolution* **23**, 564–71.

Provan J, Wattier R and Maggs C (2005). Phylogeographic analysis of the red seaweed *Palmaria palmata* reveals a Pleistocene marine glacial refugium in the English Channel, *Molecular Ecology* **14**, 793–803.

Prus-Glowacki W, Urbaniak L, Bujas E and Curtu A (2012). Genetic variation of isolated and peripheral populations of *Pinus sylvestris* (L.) from glacial refugia, *Flora* **207**, 150–158.

Remington C (1968). Suture-zones of hybrid interaction between recently joined biotas, *in* T Dobzhansky, M Hecht and W Steere, eds, *Evolutionary biology*, Plenum Press, New York, pp. 321–428.

Riesberg L and Soltis D (1991). Phylogenetic consequences of cytoplasmic gene flow in plants, *Evolutionary Trends in Plants* **5**, 65–84.

Riesberg L, Whitton J and Randal Linder C (1996). Molecular marker incongruence in plant hybrid zones and phylogenetic trees, *Acta Botanica Neerlandica* **45**, 243–262.

Rieseberg L (1995). The role of hybridization in evolution : old wine in new skins, *American Journal of Botany* **82**, 944–953.

Rieseberg L, Desrochers A and Youn S (1995). Interspecific pollen competition as a reproductive barrier between sympatric species of *Helianthus* (Asteraceae), *American Journal of Botany* **82**, 515–519.

Rodriguez-Sanchez F, Guzman B, Valido A, Vargas P and Arroyo J (2009). Late Neogene history of the laurel tree (*Laurus* L., Lauraceae) based on phylogeographical analyses of Mediterranean and Macaronesian populations, *Journal of Biogeography* **36**, 1270–1281.

Ronikier M, Costa A, Aguilar J, Feliner G, Kupfer P and Mirek Z (2008). Phylogeography of *Pulsatilla vernalis* (L.) Mill. (Ranunculaceae): chloroplast DNA reveals two evolutionary lineages across central Europe and Scandinavia, *Journal of Biogeography* **35**, 1650–1664.

Rowe G, Harris D and Beebee T (2006). Lusitania revisited: a phylogeographic analysis of the natterjack toad *Bufo calamita* across its entire biogeographical range, *Molecular Phylogenetics and Evolution* **39**, 335–346.

Rutschmann F (2006). Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times, *Diversity and Distributions* **12**, 35–48.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular Biology and Evolution* **4**, 406–25.

Salzburger W, Ewing G and Von Haeseler A (2011). The performance of phylogenetic algorithms in estimating haplotype genealogies with migration, *Molecular Ecology* **20**, 1952–63.

Sanderson M, Thorne J, Wikstrom N and Bremer K (2004). Molecular evidence on plant divergence times, *American Journal of Botany* **91**, 1656–65.

Schaal B, Hayworth D, Olsen K, Rauscher J and Smith W (1998). Phylogeographic studies in plants: problems and prospects, *Molecular Ecology* **7**(4), 465–474.

Schaal B and Olsen K (2000). Gene genealogies and population variation in plants, *Proceedings of the National Academy of Sciences of the United States of America* **97**, 7024–7029.

Schmitt T and Haubrich K (2008). The genetic structure of the mountain forest butterfly *Erebia euryale* unravels the late Pleistocene and postglacial history of the mountain coniferous forest biome in Europe, *Molecular ecology* **17**, 2194–2207.

Schonswetter P, Paun O, Tribsch A and Niklfeld H (2003). Out of the Alps: colonization of Northern Europe by East Alpine populations of the Glacier Buttercup *Ranunculus glacialis* L. (Ranunculaceae), *Molecular Ecology* **12**, 3373–3381.

Schonswetter P, Popp M and Brochmann C (2006). Rare arctic-alpine plants of the European Alps have different immigration histories: the snow bed species *Minuartia biflora* and *Ranunculus pygmaeus*, *Molecular Ecology* **15**, 709–720.

Schonswetter P, Stehlik I, Holderegger R and Tribsch A (2005). Molecular evidence for glacial refugia of mountain plants in the European Alps, *Molecular Ecology* **14**, 3547–3555.

Schonswetter P and Tribsch A (2005). Vicariance and dispersal in the alpine perennial *Bupleurum stellatum* L. (Apiaceae), *Taxon* **54**, 725–732.

Schonswetter P, Tribsch A, Barfuss M and Niklfeld H (2002). Several Pleistocene refugia detected in the high alpine plant *Phyteuma globulariifolium* Sternb & Hoppe (Campanulaceae) in the European Alps, *Molecular Ecology* **11**, 2637–2647.

Schonswetter P, Tribsch A and Niklfeld H (2004*a*). Amplified Fragment Length Polymorphism (AFLP) reveals no genetic divergence of the Eastern Alpine endemic *Oxytropis campestris* subsp. *tiroliensis* (Fabaceae) from widespread subsp. campestris, *Plant Systematics and Evolution* **244**, 245–255.

Schonswetter P, Tribsch A and Niklfeld H (2004*b*). Amplified fragment length polymorphism (AFLP) suggests old and recent immigration into the Alps by the arctic-alpine annual *Comastoma tenellum* (Gentianaceae), *Journal of Biogeography* **31**, 1673–1681.

Schonswetter P, Tribsch A, Stehlik I and Niklfeld H (2004). Glacial history of high alpine *Ranunculus glacialis* (Ranunculaceae) in the European Alps in a comparative phylogeographical context, *Biological Journal of the Linnean Society* **81**, 183–195.

Schweer J, Loschelder H and Link G (2006). A promoter switch that can rescue a plant sigma factor mutant, *FEBS Letters* **580**, 6617–6622.

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

Sinclair W, Morman J and Ennos R (1999). The postglacial history of Scots pine (*Pinus sylvestris* L.) in western Europe: evidence from mitochondrial DNA variation, *Molecular Ecology* **8**, 83–88.

Sinclair W T, Morman J D and Ennos R A (1998). Multiple origins for Scots pine (*Pinus sylvestris* L.) in Scotland: evidence from mitochondrial DNA variation, *Heredity* **80**, 233–240.

Skrede I, Borgen L and Brochmann C (2009). Genetic structuring in three closely related circumpolar plant species: AFLP versus microsatellite markers and high-arctic versus arctic-alpine distributions, *Heredity* **102**, 293–302.

Skrede I, Eidesen P, Portela R and Brochmann C (2006). Refugia, differentiation and postglacial migration in arctic-alpine Eurasia, exemplified by the mountain avens (*Dryas octopetala* L.), *Molecular Ecology* **15**, 1827–1840.

Smissen R, Clement J, Garnock Jones P and Chambers G (2002). Subfamilial Relationships within Caryophyllaceae as Inferred from 5' NDF Sequences, *American Journal of Botany* **89**, 1336–1341.

Smith R and Waldren S (2009). Patterns of genetic variation in *Colchicum autumnale* L. and its conservation status in Ireland: a broader perspective on local plant conservation, *Conservation Genetics* **11**, 1351–1361.

Stegemann S, Keuthe M, Greiner S and Bock R (2012). Horizontal transfer of chloroplast genomes between plant species, *Proceedings of the National Academy of Sciences of the United States of America* **109**(7), 2434–2438.

Stehlik I (2003). Resistance or emigration? Response of Alpine plants to the ice ages, *Taxon* **52**, 499–510.

Stewart J and Lister A (2001). Cryptic northern refugia and the origins of the modern biota, *Trends in Ecology & Evolution* **16**, 608–613.

Stewart J, Lister A, Barnes I and Dalen L (2010). Refugia revisited: individualistic responses of species in space and time, *Proceedings of the Royal Society B. Biological Sciences* **277**, 661–71.

Svendsen J, Alexanderson H, Astakhov V, Demidov I, Dowdeswell J, Funder S, Gataullin V, Henriksen M, Hjort C, Houmark-Nielsen M, Hubberten H, Ingolfsson O, Jakobsson M, Kjaer K, Larsen E, Lokrantz H, Pekka Lunkka J, Lysa A, Mangerud J, Matiouchkov A, Murray A, Moller P, Niessen F, Nikolskaya O, Polyak L, Saarnisto M, Siegert C, Siegert M, Spielhagen R and Stein R (2004). Late Quaternary ice sheet history of northern Eurasia, *Quaternary Science Reviews* **23**, 1229–1271.

Synge F (1969). The Wurm ice limit in the West of Ireland, *in Quaternary geology and climate*, National Academy of Sciences, Washington, pp. 89–92.

Synnott D (1981). *Barbula maxima* in Canada and its significance for Irish vegetational history, *Irish Naturalists' Journal* **20**, 305–307.

Taberlet P, Fumagalli L, Wust Saucy A and Cosson J (1998). Comparative phylogeography and postglacial colonization routes in Europe, *Molecular Ecology* **7**, 453–64.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Molecular Biology and Evolution* **28**, 2731–2739.

Teacher A, Garner T and Nichols R (2009). European phylogeography of the common frog (*Rana temporaria*): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium, *Heredity* **102**, 490–6.

Templeton A, Crandall K and Sing F (1992). Cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. 111. Cladogram estimation, *Genetics* **132**, 619–633.

Templeton A, Routman E and Phillips C (1995). Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA sequence data. III, *Genetics* **140**, 767–782.

Toucanne S, Zaragosi S, Bourillet J, Gibbard P, Eynaud F, Giraudeau J, Turon J, Cremer M, Cortijo E, Martinez P and Rossignol L (2009). A 1.2Ma record of glaciation and fluvial discharge from the West European Atlantic margin, *Quaternary Science Reviews* **28**, 2974–2981.

Tribsch A and Schonswetter P (2003). Patterns of endemism and comparative phylogeography confirm palaeoenvironmental evidence for pleistocene refugia in the Eastern Alps, *Taxon* **52**, 477–497.

Tribsch A, Schonswetter P and Stuessy T (2002). *Saponaria pumila* (Caryophyllaceae) and the Ice Age in the European Alps, *American Journal of Botany* **89**, 2024–2033.

Tsitrone A, Kirkpatrick M and Levin D (2003). A model for chloroplast capture, *Evolution* **57**, 1776–82.

Tutin T, Heywood V, Burges N, Valentine D, Walters S and Webb D, eds (1964). *Flora Europaea. Volume I. Lycopodiaceae to Platanaceae*, Cambridge University Press, Cambridge.

Valcarcel V, Vargas P and Feliner G (2006). Phylogenetic and phylogeographic analysis of the western Mediterranean *Arenaria* section Plinthine (Caryophyllaceae) based on nuclear, plastid, and morphological markers, *Taxon* **55**, 297–312.

Valente L, Savolainen V and Vargas P (2010). Unparalleled rates of species diversification in Europe, *Proceedings of the Royal Society B. Biological Sciences* **277**, 1489–1496.

Valtuena F, Preston C and Kadereit J (2012). Phylogeography of a Tertiary relict plant, *Meconopsis cambrica* (Papaveraceae), implies the existence of northern refugia for a temperate herb, *Molecular Ecology* **21**, 1423–1437.

Van Droogenbroeck B, Kyndt T, Romeijn-Peeters E, Van Thuyne W, Goetghebeur P, Romero-motochi J and Gheysen G (2006). Evidence of natural hybridization and introgression between *Vasconcellea* species (Caricaceae) from southern Ecuador revealed by chloroplast, mitochondrial and nuclear DNA markers, *Annals of Botany* **97**, 793–805.

Van Raamsdonk L, Smiech M and Sandbrink J (1997). Introgression explains incongruence between nuclear and chloroplast DNA-based phylogenies in *Allium* section *Cepa*, *Botanical Journal of the Linnean Society* pp. 91–108.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M and Zabeau M (1995). AFLP: a new technique for DNA fingerprinting, *Nucleic Acids Research* **23**, 4407–4414.

Walker K (2000). The distribution, ecology and conservation of *Arenaria norvegica* subsp. *anglica* Halliday (Caryophyllaceae), *Watsonia* **2**, 197–208.

Walker K, Ward S and Parr S (2008). *Arenaria norvegica* ssp. *norvegica* rediscovered in Ireland after 47 years, *BSBI News* **109**.

Wang D Y, Kumar S and Hedges S (1999). Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi, *Proceedings of the Royal Society B. Biological Sciences* **266**, 163–71.

Watts W (1977). The Late Devensian Vegetation of Ireland, *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **280**, 273–293.

Webb D (1983). The flora of Ireland in its European context, *Journal of Life Sciences* **4**, 143–160.

Webb T and Bartlein P (1992). Global changes during the last 3 million years: climatic controls and biotic responses, *Annual Review of Ecology and Systematics* **23**, 141–173.

Westergaard K, Alsos I, Ehrich D, Eidesen P, Hollingsworth P and Brochmann C (2008). Genetic diversity and distinctiveness in Scottish alpine plants, *Plant Ecology & Diversity* **1**(2), 329–338.

Westergaard K B, Alsos I G, Popp M, Engelskjon T, Flatberg K and Brochmann C (2011). Glacial survival may matter after all : nunatak signatures in the rare European populations of two west-arctic species, *Molecular Ecology* **20**, 376–393.

White T, Bruns T, Lee S and Taylor J (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, *in* I M., D Gelfand, J Sninsky and T White, eds, *PCR protocols: A guide to Methods and Applications*, San Diego: Academic Press, San Diego, pp. 315–322.

Wiley E and Lieberman B (2011). *Phylogenetics: Theory and Practice of Phylogenetics Systematics*, 2nd ed. edn, Wiley-Blackwell, New Jersey.

Willis K and Niklas K (2004). The role of Quaternary environmental change in plant macroevolution: the exception or the rule?, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359**, 159–72.

Willmott A (1930). Concerning the history of the British flora, *in A l'Etude du Peuplement des Iles Britanniques*, Societe de Biogeographie III, p. 164.

Wingfield R (1996). A model of sea-levels in the Irish and Celtic seas during the end-Pleistocene to Holocene transition. in Island Britain: A Quaternary perspective. In R.C. Preece (ed.), *Geological Society Special Publication* **96**, 209–242.

Wolfe K, Li W and Sharp P (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs, *Proceedings of the National Academy of Science* **84**, 9054–9058.

Wyse Jackson M and Parnell J (1987). A biometric study of the *Arenaria ciliata* L. complex (Caryophyllaceae), *Watsonia* **16**, 373–382.

Xiong J (2006). *Essential Bioinformatics*, Cambridge University Press, Cambridge.

Zuckerkandl E and Pauling L (1965). Evolutionary divergence and convergence in proteins, *in* V Bryson and H Vogel, eds, *Evolving Genes and Proteins*, Academic Press, New York, pp. 97–166.

Appendix A

Tables containing concatenated cp. haplotype frequencies for each population

Table A.1: Concatenated cp. haplotype frequency for each population

Haplotype	AC1	AC2	AC3	AC4	AC6	AC7	AC8	AC9	AC10	AC11	AC13	AC14	AC15
con1	1	-	1	1	-	-	-	-	-	-	-	-	-
con2	-	-	-	1	-	-	-	-	-	-	-	-	-
con3	-	-	-	-	-	-	-	-	-	-	-	1	-
con4	-	-	-	-	-	-	-	-	-	1	-	-	2
con5	-	2	1	-	-	-	-	-	-	-	-	-	-
con6	-	-	-	-	-	-	-	-	1	-	1	-	-
con7	-	-	-	-	-	-	-	-	-	-	-	-	-
con8	-	-	-	-	-	-	-	-	-	-	-	-	-
con9	-	-	-	-	-	-	-	-	-	1	-	-	-
con10	-	-	-	-	-	-	-	-	-	-	-	1	-
con11	-	-	-	-	-	-	-	-	-	-	-	-	-
con12	-	-	-	-	-	-	-	1	-	-	-	-	-
con13	-	-	-	-	1	-	-	-	-	-	-	-	-
con14	-	-	-	-	-	-	-	1	-	-	-	-	-
con15	-	-	-	-	-	-	1	-	-	-	-	-	-
con16	-	-	-	-	1	-	-	-	-	-	-	-	-
con17	-	-	-	-	-	1	-	-	-	-	-	-	-
con18	-	-	-	-	-	-	-	-	-	-	-	-	-
con19	-	-	-	-	-	-	-	-	-	-	-	-	-
con20	-	-	-	-	-	-	-	-	-	-	-	-	-
con21	-	-	-	-	-	-	-	-	-	-	-	-	-

Table A.2: Concatenated cp. haplotype frequency for each population - contd.

Haplotype	AC16	AC17	AN1	AN2	AN4	AN5	AN6	AN7
con1	-	-	-	-	-	-	-	-
con2	-	-	-	-	-	-	-	-
con3	-	-	-	-	-	-	-	-
con4	-	-	-	-	-	-	-	-
con5	-	-	-	-	-	-	-	-
con6	-	-	-	-	-	-	-	-
con7	1	-	-	-	-	-	-	-
con8	1	-	-	-	-	-	-	-
con9	-	-	-	-	-	-	-	-
con10	-	-	-	-	-	-	-	-
con11	-	1	-	-	-	-	-	-
con12	-	-	-	-	-	-	-	-
con13	-	-	-	-	-	-	-	-
con14	-	-	-	-	-	-	-	-
con15	-	-	-	-	-	-	-	-
con16	-	-	-	-	-	-	-	-
con17	-	-	-	-	-	-	-	-
con18	-	1	-	-	-	-	-	-
con19	-	-	-	-	2	-	-	-
con20	-	-	1	1	-	1	1	-
con21	-	-	1	-	-	-	-	1

Appendix B

Tables containing *matK* haplotype frequencies for each population

Table B.1: *MatK* haplotype frequency for each population

Haplotype	AC1	AC2	AC3	AC4	AC6	AC7	AC8	AC9	AC10	AC11	AC13	AC14	AC15	AC16
MatK1	5	1	3	6	1	-	-	-	-	-	-	-	-	-
MatK2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK4	-	-	-	-	-	-	-	-	2	-	-	-	-	-
MatK5	-	-	-	-	-	-	-	-	-	-	-	1	-	-
MatK6	-	-	-	-	-	-	-	-	1	2	-	3	6	-
MatK7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK8	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK9	2	5	2	-	-	-	-	-	-	3	2	2	-	-
MatK10	-	-	-	-	-	-	-	-	-	1	-	-	-	-
MatK11	-	-	-	-	-	-	-	-	-	-	-	-	-	1
MatK12	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK13	-	-	-	-	-	-	-	-	1	-	-	-	-	-
MatK14	-	-	-	-	-	-	-	-	3	-	3	-	-	2
MatK15	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK16	-	-	-	-	-	-	1	-	-	-	-	-	-	-
MatK17	-	-	-	-	3	1	3	3	-	-	-	-	-	-
MatK18	-	-	-	-	1	-	-	-	-	-	-	-	-	-
MatK19	-	-	-	-	-	-	-	2	-	-	-	-	-	-

Table B.2: *MatK* haplotype frequency for each population - contd.

Haplotype	AC17	AC18	AC19	AN1	AN2	AN3	AN4	AN5	AN6	AN7	AN8	AP1	AP2
MatK1	-	-	-	-	-	-	5	-	-	-	-	-	1
MatK2	-	-	-	-	-	-	-	-	-	-	-	-	4
MatK3	-	-	-	-	-	1	-	-	-	-	1	-	-
MatK4	-	-	-	6	1	-	2	1	1	1	-	-	-
MatK5	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK6	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK7	-	-	1	-	-	-	-	-	-	-	-	-	-
MatK8	-	-	-	-	-	-	-	-	-	-	-	1	-
MatK9	1	-	-	-	-	-	-	-	-	-	-	-	-
MatK10	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK12	-	2	3	-	-	-	-	-	-	-	-	-	-
MatK11	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK13	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK14	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK15	2	-	-	-	-	-	-	-	-	-	-	-	-
MatK16	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK17	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK18	-	-	-	-	-	-	-	-	-	-	-	-	-
MatK19	3	-	-	-	-	-	-	-	-	-	-	-	-

Appendix C

Tables containing ITS nucleotype frequencies for each population

Table C.1: ITS nucleotype frequency for each population

Haplotype	AC1	AC2	AC3	AC4	AC6	AC7	AC8	AC9	AC10	AC11	AC13	AC14	AC15
ITS1	-	-	-	-	-	-	-	-	-	2	-	2	2
ITS2	-	-	-	-	-	-	-	-	2	-	1	-	-
ITS3	-	-	-	-	-	-	-	-	-	-	-	-	-
ITS4	2	2	2	2	-	-	-	-	-	-	-	-	-
ITS5	-	-	-	-	2	-	1	2	-	-	-	-	-
ITS6	-	-	-	-	-	1	-	-	-	-	-	-	-
ITS7	-	-	-	-	-	1	-	-	-	-	-	-	-
ITS8	-	-	-	-	-	-	-	-	-	-	-	-	-

Table C.2: Concatenated cp. haplotype frequency for each population - contd.

Haplotype	AC16	AC17	AN1	AN2	AN4	AN5	AN6	AN7
ITS1	-	-	-	-	-	-	-	-
ITS2	-	-	-	-	-	-	-	-
ITS3	-	-	2	1	2	1	2	1
ITS4	-	-	-	-	-	-	-	-
ITS5	-	-	-	-	-	-	-	-
ITS6	-	-	-	-	-	-	-	-
ITS7	-	-	-	-	-	-	-	-
ITS8	2	2	-	-	-	-	-	-

Appendix D

Table of individual samples which were sequenced for each region and the location of the lab where they were prepared

Sample	MatK	Concat	ITS	Location	Sample	MatK	Concat	ITS	Location
AC1.2	■			RJB	AC10.10	■			RJB
AC1.5	■			RJB	AC10.13	■			RJB
AC1.7	■			RJB	AC10.15	■	■	■	RJB
AC1.8	■	■	■	RJB	AC13.1	■	■	■	RJB
AC1.10	■			RJB	AC13.5	■			RJB
AC1.11	■			RJB	AC13.7	■			RJB
AC1.25	■	■	■	RJB	AC13.10	■			RJB
AC2.1	■			RJB	AC13.13	■			RJB
AC2.4	■	■	■	RJB	AC11.1	■	■	■	RJB
AC2.5	■			RJB	AC11.3	■			RJB
AC2.7	■			RJB	AC11.10	■			RJB
AC2.12	■			RJB	AC11.13	■			RJB
AC2.25	■	■	■	RJB	AC11.15	■	■	■	RJB
AC3.1	■	■	■	RJB	AC14.1	■	■	■	RJB
AC3.4	■			RJB	AC14.5	■			RJB
AC3.14	■			RJB	AC14.7	■			RJB
AC3.17	■			RJB	AC14.10	■			RJB
AC3.23	■	■	■	RJB	AC14.13	■			RJB
AC4.1	■	■	■	RJB	AC14.15	■	■	■	RJB
AC4.3	■			RJB	AC15.1	■	■	■	RJB
AC4.7	■			RJB	AC15.3	■			RJB
AC4.13	■			RJB	AC15.5	■			RJB
AC4.15		■	■	RJB	AC15.10	■			RJB
AC4.10	■			RJB	AC15.13	■			RJB
AC6.3	■			RJB	AC15.15	■	■	■	RJB
AC6.5	■	■	■	RJB	AC16.1	■	■	■	RJB
AC6.13	■			RJB	AC16.3	■			RJB
AC6.16	■	■	■	RJB	AC16.5	■			RJB
AC6.18	■			RJB	AC16.10	■			RJB
AC7.1	■	■	■	RJB	AC16.13	■			RJB
AC8.1	■	■	■	RJB	AC16.15	■	■	■	RJB
AC8.2	■	■	■	RJB	AC17.1	■	■	■	RJB
AC8.3	■			RJB	AC17.2	■			RJB
AC8.4	■			RJB	AC17.3	■			RJB
AC9.1	■			RJB	AC17.5	■			RJB
AC9.2	■	■	■	RJB	AC17.8	■			RJB
AC9.3	■			RJB	AC17.10	■	■	■	RJB
AC9.4	■			RJB	AC18.1	■			NUIM
AC9.5	■	■	■	RJB	AC18.4	■			NUIM
AC10.1	■	■	■	RJB	AC19.6	■			NUIM
AC10.4	■			RJB	AC19.18	■			NUIM
AC10.5	■			RJB	AC19.24	■			NUIM
AC10.6	■			RJB	AC19.30	■			NUIM

Figure D.1: List of individuals which were sequenced. The shaded cells indicate which region was sequenced and the location of which lab is indicated: RJB - Royal Botanic Gardens, Madrid and NUIM - Molecular Ecology Lab, NUI Maynooth

Sample	MatK	Concat	ITS	Location
AN1.2	■			RJB
AN1.3	■	■	■	RJB
AN1.5	■			RJB
AN1.8	■			RJB
AN1.14	■	■	■	RJB
AN1.19	■			RJB
AN2.5	■	■	■	RJB
AN3.1	■			NUIM
AN4.1	■	■	■	RJB
AN4.4	■			RJB
AN4.8	■			RJB
AN4.10	■	■	■	RJB
AN4.14	■			RJB
AN4.15	■			RJB
AN4.19	■			RJB
AN5.4	■	■	■	RJB
AN6.16			■	RJB
AN6.28	■	■	■	RJB
AN7.2	■	■	■	RJB
AN8.2	■			NUIM
AP1.1	■			NUIM
AP2.1	■			NUIM
AP2.2	■			NUIM
AP2.3	■			NUIM
AP2.4	■			NUIM
AP2.5	■			NUIM
AS1.17_	■	■	■	RJB

Figure D.2: List of individuals which were sequenced. The shaded cells indicate which region was sequenced and the location of which lab is indicated: RJB - Royal Botanic Gardens, Madrid and NUIM - Molecular Ecology Lab, NUI Maynooth

Appendix E

Summary statistics determined from BEAST for the estimated node ages

Table E.1: Summary statistics determined from BEAST and viewed in TRACER for the estimated node ages (in my) for the *matK* haplotypes with *Arenaria nevadensis* calibrated for 4 age estimates

Summary statistic	**A**	**B**	**C**	**D**	**E**	**F**
matK A.nev 3.328my (25my)						
mean	2.085	1.488	0.694	0.241	0.308	0.497
stderr of mean	0.011	0.009	0.005	0.002	0.003	0.004
95% HPD lower	1.063	0.620	0.202	0.023	0.049	0.129
95% HPD upper	3.303	2.442	1.254	0.530	0.679	0.929
matK A.nev 4.582my (34my - LCL)						
mean	2.754	1.962	0.926	0.320	0.408	0.655
stderr of mean	0.014	0.011	0.006	0.003	0.003	0.005
95% HPD lower	1.354	0.863	0.280	0.029	0.067	0.171
95% HPD upper	4.231	3.190	1.683	0.710	0.876	1.217
matK A.nev 5.874my (45my - UCL)						
mean	3.467	2.473	1.157	0.402	0.513	0.820
stderr of mean	0.018	0.015	0.008	0.004	0.004	0.006
95% HPD lower	1.704	1.028	0.352	0.046	0.070	0.232
95% HPD upper	5.359	4.019	2.142	0.895	1.093	1.532
matK A.nev 7.217my (55my)						
mean	4.295	3.069	1.436	0.502	0.637	1.021
stderr of mean	0.020	0.017	0.010	0.004	0.006	0.007
95% HPD lower	2.211	1.429	0.415	0.051	0.080	0.284
95% HPD upper	6.510	4.905	2.602	1.098	1.327	1.858

Table E.2: Summary statistics determined from BEAST and viewed in TRACER for the estimated node ages (in my) for the *matK* haplotypes with *Arenaria serpyllifolia* calibrated for 4 age estimates

Summary statistic	**A**	**B**	**C**	**D**	**E**	**F**
matK A.ser 5.425my (25my)						
mean	2.963	2.137	0.996	0.346	0.439	0.710
stderr of mean	0.016	0.013	0.007	0.003	0.004	0.005
95% HPD lower	1.440	0.915	0.291	0.035	0.055	0.215
95% HPD upper	4.708	3.516	1.808	0.756	0.948	1.330
matK A.ser 7.47my (34my - LCL)						
mean	4.061	2.941	1.365	0.471	0.608	0.976
stderr of mean	0.021	0.017	0.009	0.004	0.005	0.007
95% HPD lower	2.113	1.272	0.443	0.043	0.095	0.312
95% HPD upper	6.295	4.700	2.504	1.023	1.337	1.828
matK A.ser 9.621my (45my - UCL)						
mean	5.137	3.704	1.726	0.597	0.771	1.230
stderr of mean	0.029	0.022	0.013	0.005	0.007	0.009
95% HPD lower	2.613	1.564	0.536	0.078	0.123	0.398
95% HPD upper	7.982	5.891	3.131	1.338	1.675	2.345
matK A.ser 11.808my (55my)						
mean	6.295	4.527	2.118	0.730	0.933	1.505
stderr of mean	0.038	0.030	0.017	0.007	0.008	0.012
95% HPD lower	3.060	1.945	0.679	0.063	0.139	0.439
95% HPD upper	9.768	7.246	3.926	1.600	1.975	2.809

Table E.3: Summary statistics determined from BEAST and viewed in TRACER for the estimated node ages (in my) for the *Arenaria ciliata matK* haplotypes with no outgroup calibrated for 4 age estimates

Summary statistic	**A**	**B**	**C**	**D**
matK A.ciliata only 2.725my (25my)				
mean	0.300	0.889	0.414	0.710
stderr of mean	0.003	0.006	0.003	0.005
95% HPD lower	0.032	0.251	0.052	0.166
95% HPD upper	0.687	1.661	0.940	1.366
matK A.ciliata only 3.655my (34my - LCL)				
mean	0.387	1.143	0.535	0.917
stderr of mean	0.003	0.008	0.004	0.007
95% HPD lower	0.040	0.312	0.055	0.220
95% HPD upper	0.871	2.135	1.199	1.780
matK A.ciliata 4.675my (45my - UCL)				
mean	0.486	1.417	0.661	1.153
stderr of mean	0.004	0.010	0.006	0.009
95% HPD lower	0.046	0.379	0.088	0.266
95% HPD upper	1.096	2.620	1.462	2.222
matK A.ciliata 5.81my (55my)				
mean	0.595	1.750	0.821	1.431
stderr of mean	0.005	0.012	0.007	0.010
95% HPD lower	0.066	0.487	0.094	0.328
95% HPD upper	1.352	3.206	1.808	2.744

Table E.4: Summary statistics from TRACER with the mean age estimate and standard error for each clade for the concatenated chloroplast haplotypes with *A. serpyllifolia*. Also shown are the highest posterior densites (HPD)

Summary statistic	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
concat A.ser 5.425my (25my)												
mean	4.079	3.203	0.293	0.053	0.065	0.512	1.512	0.234	0.753	0.406	0.165	0.405
stderr of mean	0.019	0.017	0.003	0.001	0.001	0.004	0.008	0.002	0.004	0.002	0.001	0.003
95% HPD lower	2.466	1.763	0.107	0.001	0.001	0.216	0.838	0.034	0.328	0.147	0.023	0.138
95% HPD upper	5.688	4.598	0.542	0.126	0.161	0.852	2.317	0.470	1.208	0.716	0.344	0.681
concat A.ser 7.47my (34my - LCL)												
mean	5.568	4.345	0.397	0.074	0.087	0.696	2.051	0.320	1.030	0.553	0.223	0.550
stderr of mean	0.025	0.023	0.004	0.001	0.001	0.005	0.012	0.002	0.006	0.004	0.002	0.004
95% HPD lower	3.485	2.497	0.134	0.003	0.001	0.277	1.074	0.064	0.452	0.192	0.030	0.194
95% HPD upper	7.620	6.181	0.722	0.176	0.220	1.128	3.103	0.657	1.691	0.957	0.456	0.920
concat A.ser 9.621my (45my - UCL)												
mean	6.995	5.479	0.502	0.093	0.113	0.875	2.590	0.403	1.295	0.694	0.283	0.689
stderr of mean	0.036	0.031	0.005	0.001	0.001	0.007	0.016	0.003	0.008	0.005	0.002	0.005
95% HPD lower	4.287	3.080	0.155	0.003	0.002	0.382	1.405	0.077	0.540	0.243	0.047	0.271
95% HPD upper	9.557	7.729	0.898	0.221	0.282	1.433	3.893	0.806	2.048	1.192	0.598	1.180
concat A.ser 11.808my (55my)												
mean	8.598	6.726	0.625	0.112	0.139	1.084	3.188	0.498	1.591	0.850	0.348	0.855
stderr of mean	0.044	0.041	0.006	0.001	0.002	0.007	0.019	0.003	0.009	0.005	0.002	0.006
95% HPD lower	5.414	3.963	0.231	0.003	0.004	0.484	1.794	0.095	0.757	0.294	0.053	0.337
95% HPD upper	11.616	9.429	1.144	0.268	0.350	1.784	4.777	1.004	2.571	1.463	0.726	1.455

Table E.5: Summary statistics from TRACER with the mean age estimate and standard error for each clade for the concatenated chloroplast haplotypes with *A. ciliata* only. Also shown are the highest posterior densites (HPD)

Summary statistic	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**
concat A.ciliata only 2.725my (25my)										
mean	2.218	0.196	0.036	0.040	0.338	1.025	0.155	0.501	0.265	0.107
stderr of mean	0.010	0.002	0.000	0.001	0.002	0.005	0.001	0.003	0.002	0.001
95% HPD lower	1.289	0.067	0.000	0.001	0.146	0.519	0.025	0.210	0.090	0.015
95% HPD upper	3.277	0.360	0.088	0.100	0.570	1.586	0.317	0.825	0.473	0.228
concat A.ciliata only 3.655my (34my - LCL)										
mean	2.907	0.258	0.047	0.051	0.446	1.343	0.200	0.646	0.345	0.140
stderr of mean	0.010	0.002	0.000	0.001	0.003	0.006	0.001	0.003	0.002	0.001
95% HPD lower	1.692	0.078	0.001	0.001	0.176	0.688	0.033	0.273	0.127	0.020
95% HPD upper	4.130	0.457	0.115	0.130	0.733	2.025	0.409	1.054	0.626	0.295
concat A.ciliata only 4.675 (45my - UCL)										
mean	3.674	0.321	0.060	0.065	0.572	1.712	0.253	0.822	0.439	0.176
stderr of mean	0.017	0.003	0.001	0.001	0.005	0.010	0.002	0.005	0.003	0.002
95% HPD lower	2.192	0.099	0.000	0.001	0.220	0.858	0.044	0.371	0.149	0.023
95% HPD upper	5.107	0.574	0.149	0.169	0.948	2.555	0.512	1.357	0.770	0.368
concat A.ciliata only 5.81my (55my)										
mean	4.462	0.396	0.072	0.080	0.687	2.069	0.311	1.000	0.532	0.216
stderr of mean	0.017	0.004	0.001	0.001	0.005	0.010	0.002	0.005	0.003	0.002
95% HPD lower	2.701	0.143	0.002	0.001	0.296	1.127	0.056	0.457	0.192	0.028
95% HPD upper	6.132	0.725	0.173	0.203	1.140	3.136	0.629	1.633	0.940	0.461

Table E.6: Summary statistics from TRACER with the mean age estimate and standard error for each *Arenaria ciliata* clade for the ITS sequences with *A. nevadensis* only (dataset AN). Also shown are the highest posterior densites (HPD)

Summary statistic	**A**	**B**
ITS A.nevadensis 3.328my (25my)		
mean	0.7903	0.3418
stderr of mean	7.27E-03	2.94E-03
95% HPD lower	0.3375	0.1136
95% HPD upper	1.3861	0.6501
ITS A.nevadensis 4.582my (34my - LCL)		
mean	1.0514	0.4603
stderr of mean	1.06E-02	4.58E-03
95% HPD lower	0.4358	0.1405
95% HPD upper	1.8277	0.8507
ITS A.nevadensis 5.874my (45my - UCL)		
mean	1.371	0.5994
stderr of mean	2.87E-02	1.19E-02
95% HPD lower	0.5889	0.1879
95% HPD upper	2.4963	1.1246
ITS A.nevadensis 7.217my (55my)		
mean	1.6196	0.7119
stderr of mean	2.73E-02	1.29E-02
95% HPD lower	0.7527	0.2177
95% HPD upper	2.8177	1.2972

Table E.7: Summary statistics from TRACER with the mean age estimate and standard error for each *Arenaria ciliata* clade for the ITS sequences with *A. serpyllifolia* only (dataset AS). Also shown are the highest posterior densites (HPD)

Summary statistic	**A**	**B**	**C**	**D**
ITS A.ser 5.425my (25my)				
mean	0.5988	1.4382	1.0999	0.5035
stderr of mean	7.06E-03	1.99E-02	1.34E-02	6.20E-03
95% HPD lower	0.1262	0.559	0.4599	0.1556
95% HPD upper	1.2151	2.6335	1.9161	0.9342
ITS A.ser 7.47my (34my - LCL)				
mean	0.7876	1.8999	1.4647	0.6752
stderr of mean	1.05E-02	2.74E-02	1.90E-02	8.85E-03
95% HPD lower	0.1726	0.8434	0.614	0.2422
95% HPD upper	1.5551	3.366	2.52	1.2477
ITS A.ser 9.621my (45my - UCL)				
mean	1.0432	2.5065	1.9281	0.8878
stderr of mean	2.18E-02	6.05E-02	4.30E-02	1.98E-02
95% HPD lower	0.2163	1.0737	0.7485	0.2745
95% HPD upper	2.112	4.5605	3.2745	1.6627
ITS A.ser 11.808my (55my)				
mean	1.2931	3.116	2.3815	1.0869
stderr of mean	2.33E-02	6.08E-02	4.17E-02	1.81E-02
95% HPD lower	0.2737	1.2194	0.9808	0.3303
95% HPD upper	2.6192	5.7747	4.2354	2.0254

Table E.8: Summary statistics from TRACER with the mean age estimate and standard error for each *Arenaria ciliata* clade for the ITS sequences with *A. ciliata* only (dataset AC). Also shown are the highest posterior densites (HPD)

Summary statistic	**A**	**B**	**C**	**D**
ITS A.cil 2.275 (25my)				
mean	1.3741	0.2629	1.9643	0.7635
stderr of mean	7.86E-03	2.95E-03	1.01E-02	5.99E-03
95% HPD lower	0.4064	4.60E-02	0.8341	0.1696
95% HPD upper	2.4823	0.5875	3.0485	1.4687
ITS A.cil 3.655my (34my - LCL)				
mean	1.7849	0.3384	2.5246	0.9751
stderr of mean	1.05E-02	4.12E-03	1.45E-02	7.55E-03
95% HPD lower	0.565	5.22E-02	1.0789	0.2252
95% HPD upper	3.2834	0.7665	3.9458	1.8879
ITS A.cil 4.675 (45my - UCL)				
mean	2.2391	0.4256	3.1801	1.225
stderr of mean	1.25E-02	4.91E-03	1.83E-02	8.98E-03
95% HPD lower	0.6914	7.46E-02	1.3726	0.3239
95% HPD upper	4.0565	0.9498	4.8976	2.3589
ITS A.cil 5.81my (55my)				
mean	2.768	0.5347	3.9411	1.5318
stderr of mean	1.60E-02	5.98E-03	2.04E-02	1.13E-02
95% HPD lower	0.9085	7.00E-02	1.8749	0.3381
95% HPD upper	4.9048	1.1581	6.0932	2.8839

Appendix F

List of individuals which were used in the AFLP analysis

Sample	NUIM 2009			Madrid		NUIM 2012			Sample	NUIM 2009			Madrid		NUIM 2012		
	A	B	C	B	C	A	B	C		A	B	C	B	C	A	B	C
AC1.1									AC2.18								
AC1.2									AC2.19								
AC1.3									AC2.27								
AC1.4									AC2.30								
AC1.5									AC3.1								
AC1.6									AC3.2								
AC1.7									AC3.4								
AC1.8									AC3.7								
AC1.9									AC3.12								
AC1.10									AC3.13								
AC1.14									AC3.14								
AC1.16									AC3.15								
AC1.19									AC3.16								
AC1.21									AC3.17								
AC1.24									AC3.23								
AC1.25									AC3.25								
AC1.28									AC3.27								
AC1.30									AC3.28								
AC2.1									AC3.29								
AC2.2									AC3.30								
AC2.3									AC4.1								
AC2.4									AC4.2								
AC2.5									AC4.3								
AC2.6									AC4.4								
AC2.7									AC4.5								
AC2.8A									AC4.6								
AC2.9									AC4.7								
AC2.10									AC4.8								
AC2.12									AC4.9								
AC2.15									AC4.10								
AC2.18									AC4.11								
AC2.19									AC4.12								
AC2.23									AC4.13								
AC2.25									AC4.14								

Figure F.1: List of individuals from Ben Bulben which were subjected to AFLP analysis. The shaded cells indicate which selective primer combination was used and for which batch of AFLP results.The letters stand for the selective primer combination used where A is ACT-CAG, B is AGC-CAA and C is AGC-CTG.

	NUIM 2009			Madrid		NUIM 2012				NUIM 2009			Madrid		NUIM 2012		
Sample	A	B	C	B	C	A	B	C	Sample	A	B	C	B	C	A	B	C
AC6.1									AC10.2								
AC6.1A									AC10.3								
AC6.2									AC10.4								
AC6.3									AC10.5								
AC6.5									AC10.6								
AC6.7									AC10.7								
AC6.9									AC10.8								
AC6.10									AC10.9								
AC6.12									AC10.10								
AC6.13									AC10.11								
AC6.14									AC10.15								
AC6.15									AC11.3								
AC6.16									AC11.15								
AC6.17									AC13.5								
AC6.18									AC13.10								
AC6.19									AC13.13								
AC6.20									AC14.7								
AC6.21									AC14.13								
AC6.23									AC15.15								
AC7.1									AC16.3								
AC8.1									AC16.10								
AC8.2									AC16.15								
AC8.3									AC17.1								
AC8.4									AC18.1								
AC9.1									AC18.8								
AC9.2									AC19.24								
AC9.3									AP2.1								
AC9.4									AP2.2								
AC9.5									AP2.3								
AC10.1									AP2.4								

Figure F.2: List of individuals from the continental European populations which were subjected to AFLP analysis. The shaded cells indicate which selective primer combination was used and for which batch of AFLP results. The letters stand for the selective primer combination used where A is ACT-CAG, B is AGC-CAA and C is AGC-CTG.

	NUIM 2009			Madrid		NUIM 2012				NUIM 2009			Madrid		NUIM 2012		
Sample	A	B	C	B	C	A	B	C	Sample	A	B	C	B	C	A	B	C
AN1.1	■	■	■						AN4.14	■	■	■			■	■	■
AN1.2	■	■	■			■	■	■	AN4.15	■	■	■			■	■	■
AN1.5	■	■	■			■	■	■	AN4.18	■	■	■					
AN1.8						■	■	■	AN4.19						■	■	■
AN1.9	■	■	■						AN4.20	■	■	■					
AN1.10	■	■	■						AN4.21	■	■	■					
AN1.11	■	■	■						AN4.22	■	■	■					
AN1.10	■	■	■						AN4.24	■	■	■					
AN1.14	■	■	■			■	■	■	AN4.25	■	■	■					
AN1.15	■	■	■						AN4.26	■	■	■					
AN1.17	■	■	■						AN4.27	■	■	■					
AN1.20	■	■	■						AN4.28	■	■	■					
AN1.21	■	■	■						AN4.29	■	■	■					
AN1.23	■	■	■						AN5.1	■	■	■			■	■	■
AN1.24	■	■	■						AN5.3	■	■	■					
AN1.25	■	■	■						AN5.4	■	■	■			■	■	■
AN1.30	■	■	■						AN5.6	■	■	■					
AN1.31	■	■	■						AN5.9	■	■	■			■	■	■
AN2.1	■	■	■						AN5.12	■	■	■					
AN2.2	■	■	■						AN5.15	■	■	■					
AN2.4	■	■	■						AN5.14	■	■	■			■	■	■
AN2.5	■	■	■			■	■	■	AN5.18	■	■	■					
AN2.7	■	■	■						AN5.20	■	■	■					
AN2.9	■	■	■						AN5.23	■	■	■					
AN2.11	■	■	■						AN5.25	■	■	■					
AN2.12	■	■	■						AN5.27	■	■	■					
AN2.13	■	■	■						AN6.1						■	■	■
AN2.14						■	■	■	AN6.3		■						
AN3.1	■	■	■			■	■	■	AN6.5		■						
AN3.2	■	■	■						AN6.8		■				■	■	■
AN3.3	■	■	■						AN6.6		■						
AN3.4	■	■	■						AN6.15						■	■	■
AN3.5	■	■	■			■	■	■	AN6.18						■	■	■
AN3.6	■	■	■						AN6.22						■	■	■
AN3.7	■	■	■						AN6.27		■						
AN4.1	■	■	■			■	■	■	AN6.32		■						
AN4.2	■	■	■						AN7.1		■				■	■	■
AN4.4						■	■	■	AN7.2		■				■	■	■
AN4.7	■	■	■						AN8.2						■	■	■
AN4.8						■	■	■	AN8.10						■	■	■
AN4.10	■	■	■						AN8.15						■	■	■
AN4.12	■	■	■						AN8.18						■	■	■

Figure F.3: List of individuals of *Arenaria norvegica* which were subjected to AFLP analysis. The shaded cells indicate which selective primer combination was used and for which batch of AFLP results. The letters stand for the selective primer combination used where A is ACT-CAG, B is AGC-CAA and C is AGC-CTG.

Appendix G

Neighbour-Joining trees and Neighbour-Net results from the individual selective AFLP primer combinations

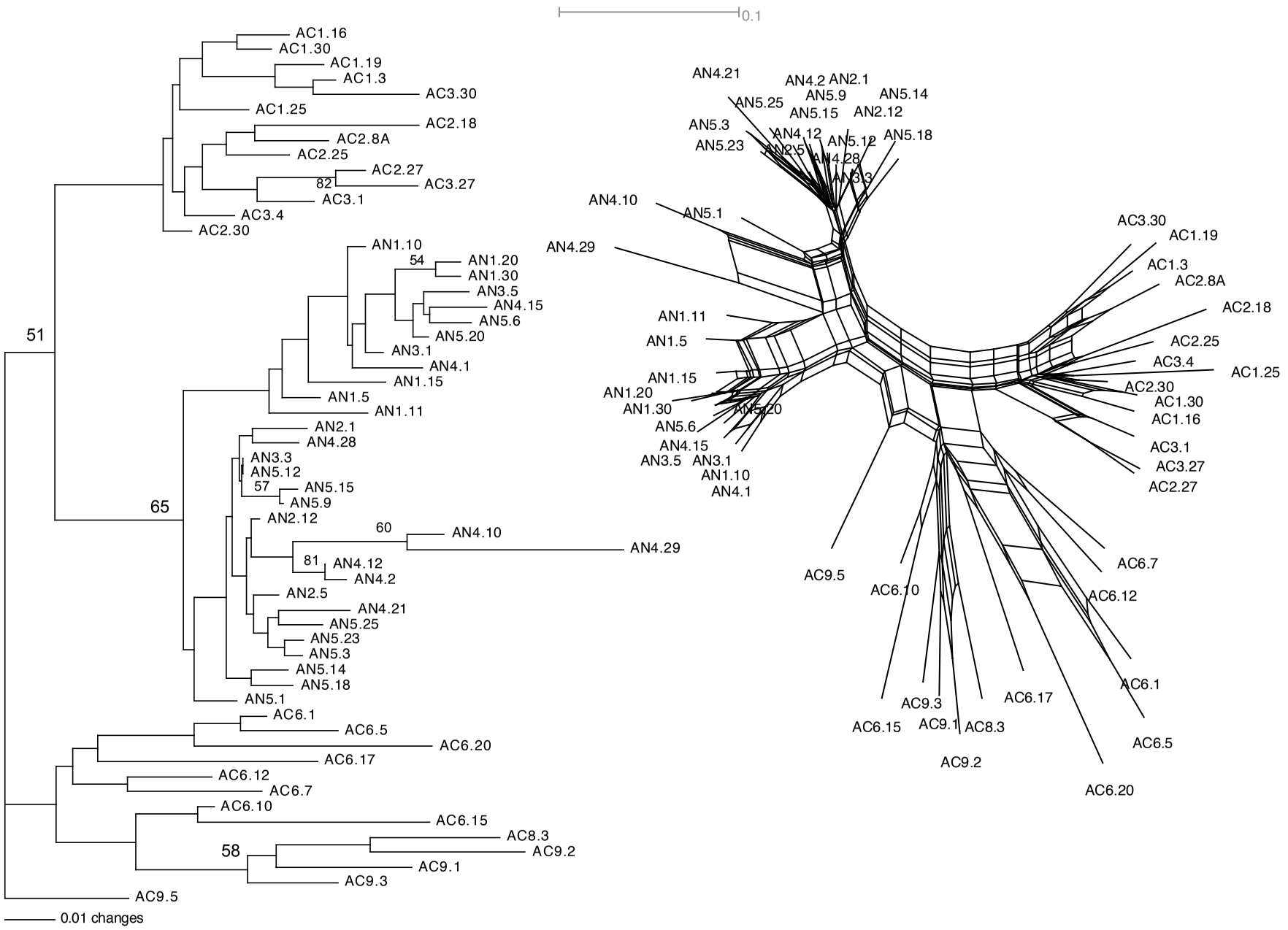


Figure G.1: Neighbour-Joining (NJ) tree and NeighbourNet network for dataset NUIM 2009 for the selective primer ACT-CAG

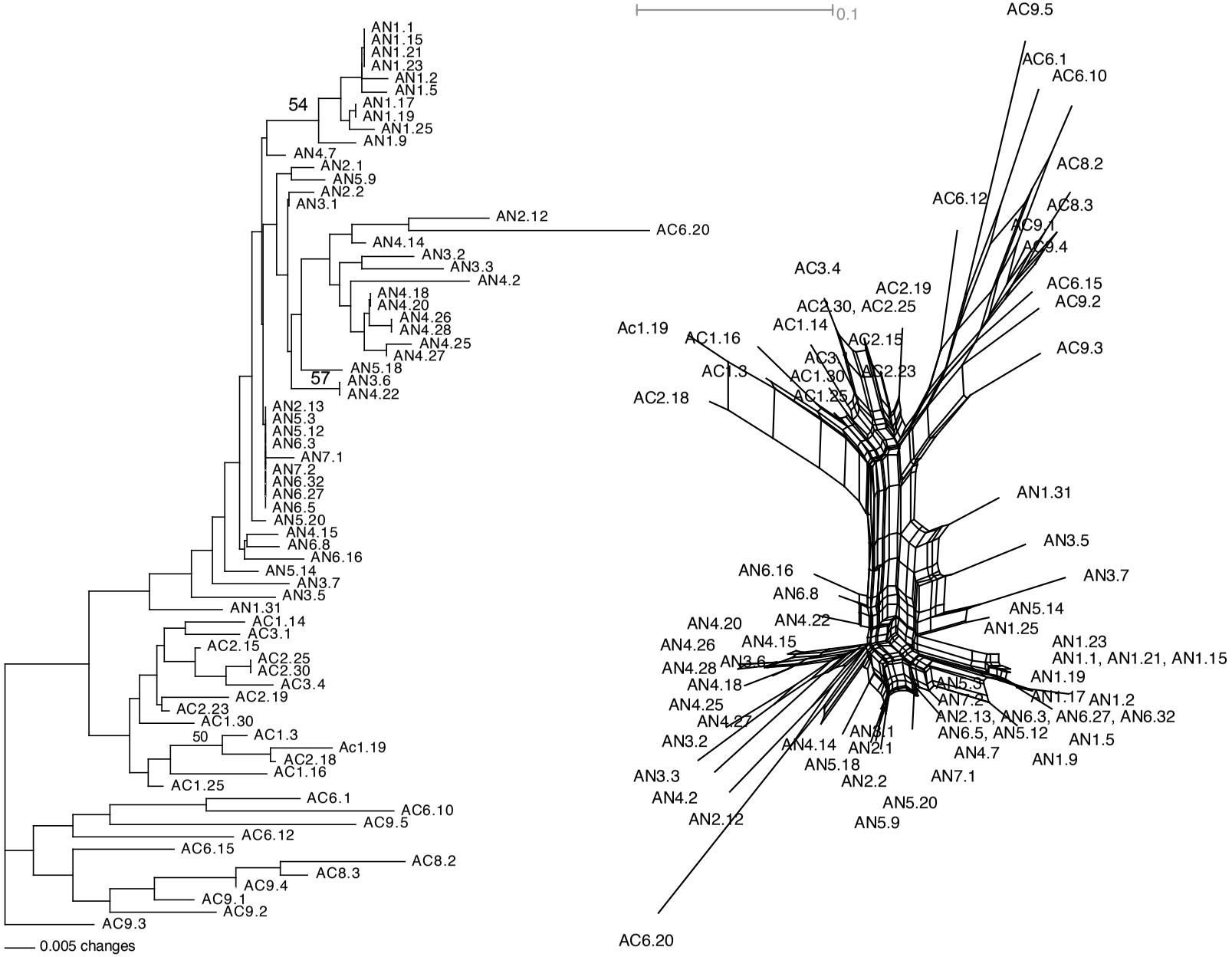


Figure G.2: Neighbour-Joining (NJ) tree and NeighbourNet network for dataset NUIM 2009 for the selective primer AGC-CAA

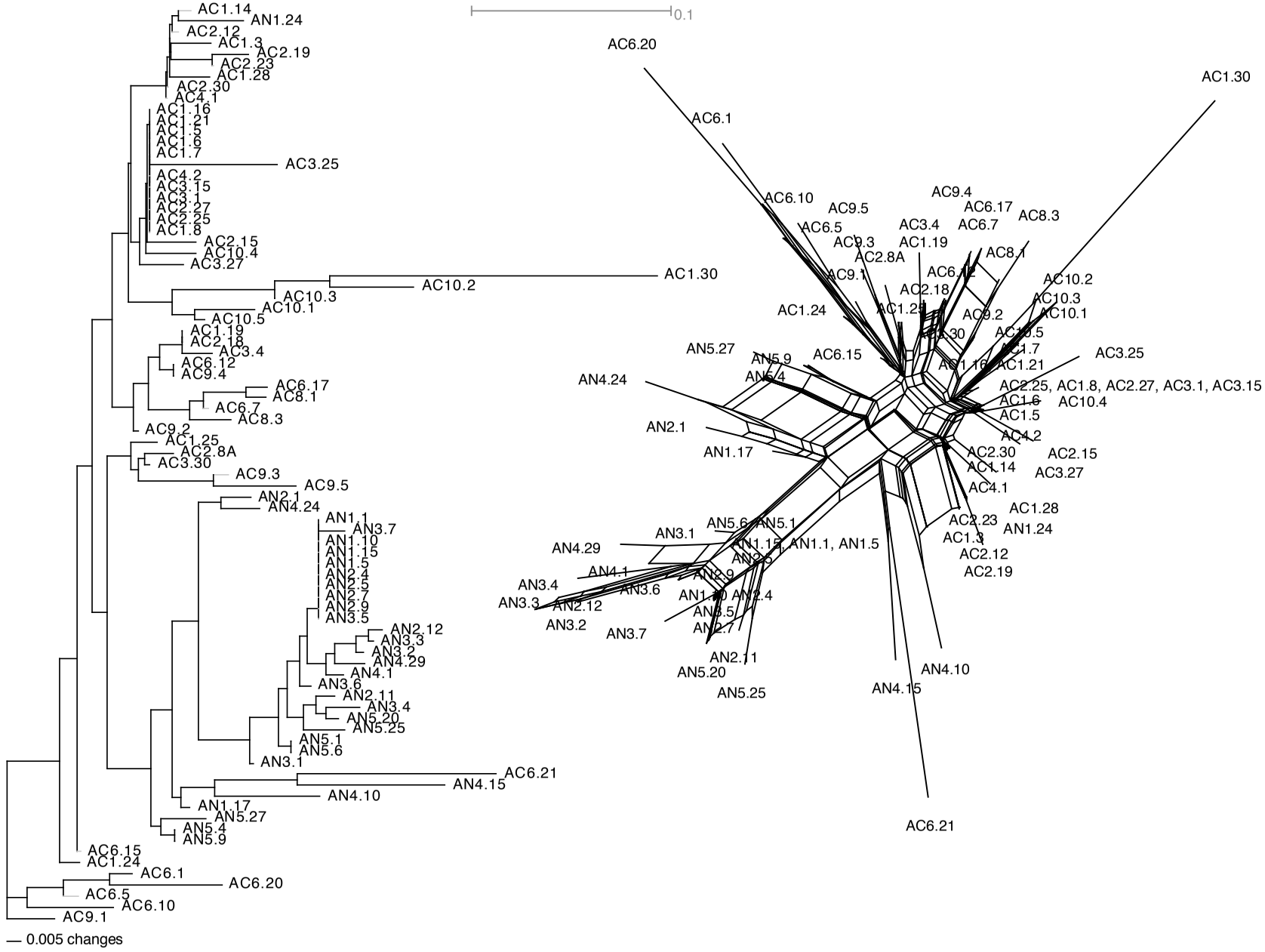


Figure G.3: Neighbour-Joining (NJ) tree and NeighbourNet network for dataset NUIM 2009 for the selective primer AGC-CTG

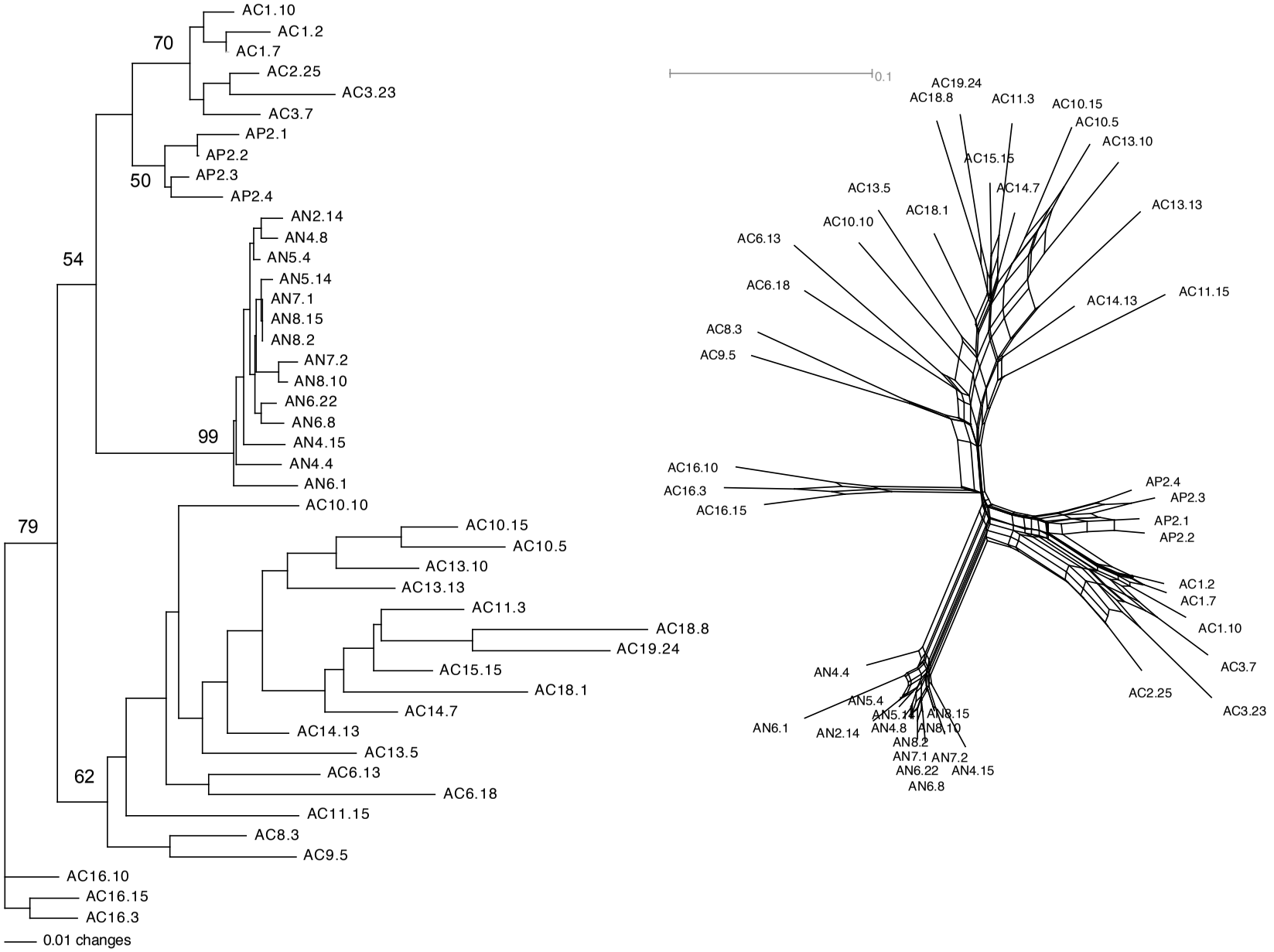


Figure G.4: Neighbour-Joining (NJ) tree and NeighbourNet network for dataset NUIM 2012 for the selective primer ACT-CAG

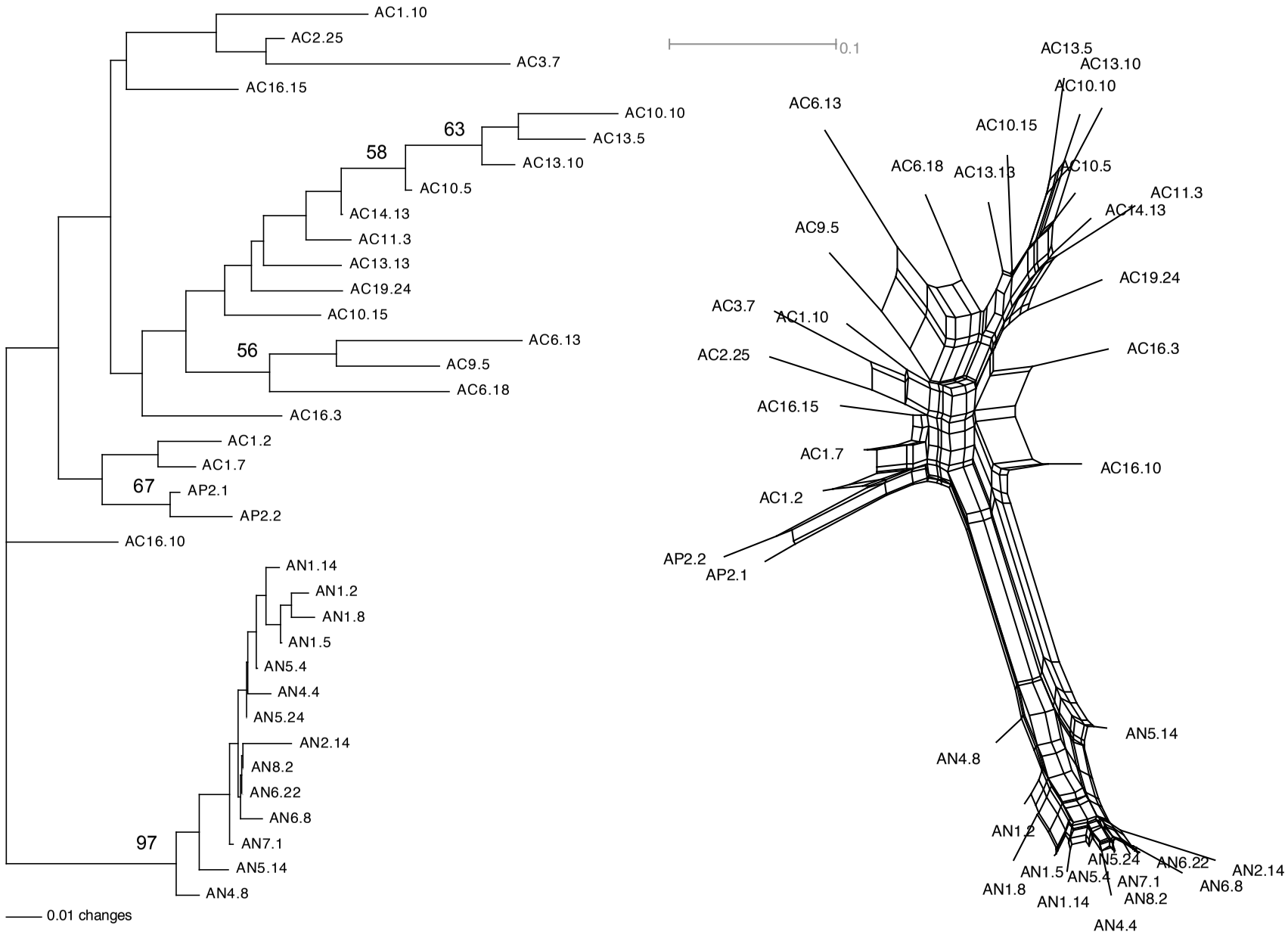


Figure G.5: Neighbour-Joining (NJ) tree and NeighbourNet network for dataset NUIM 2012 for the selective primer AGC-CAA

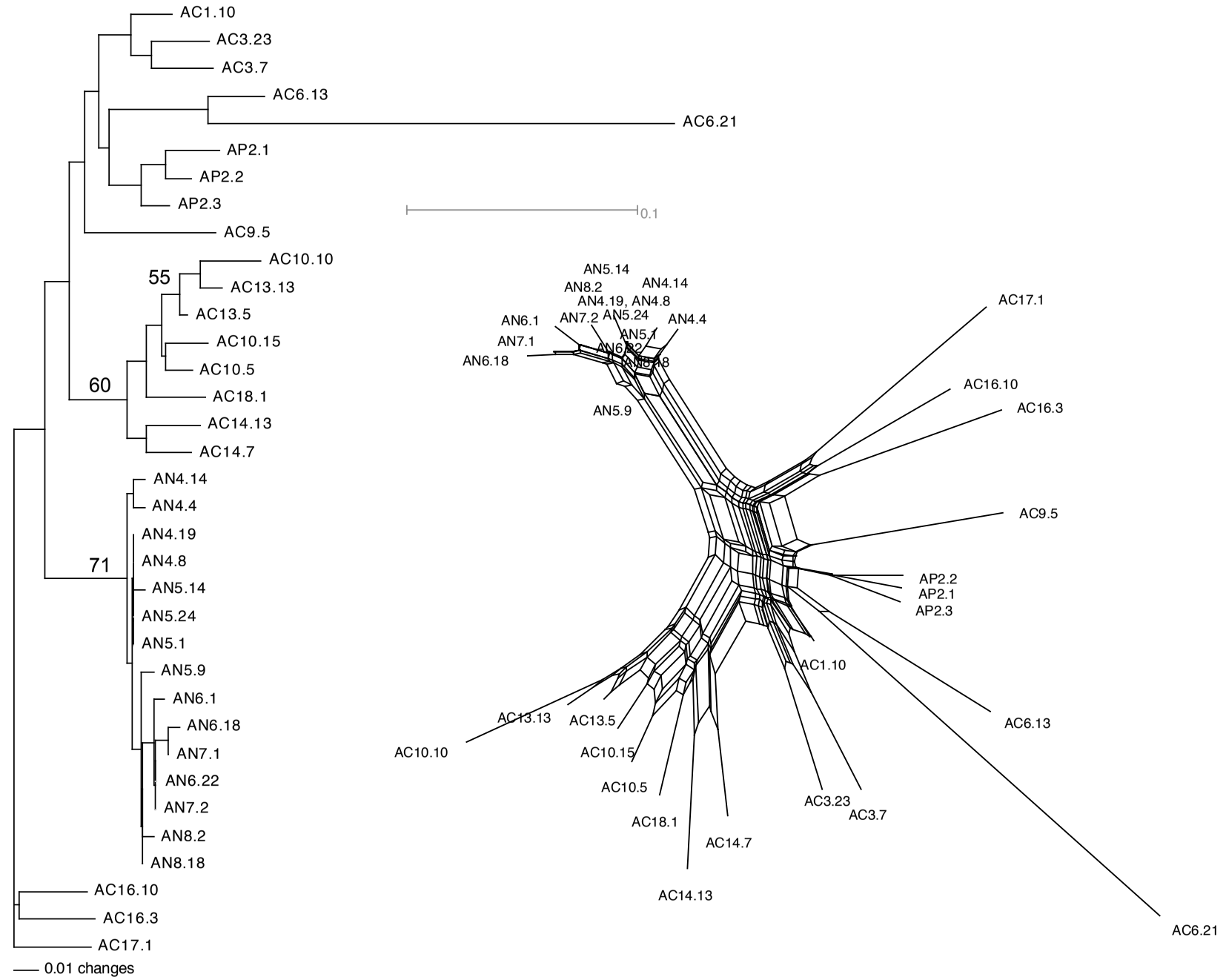
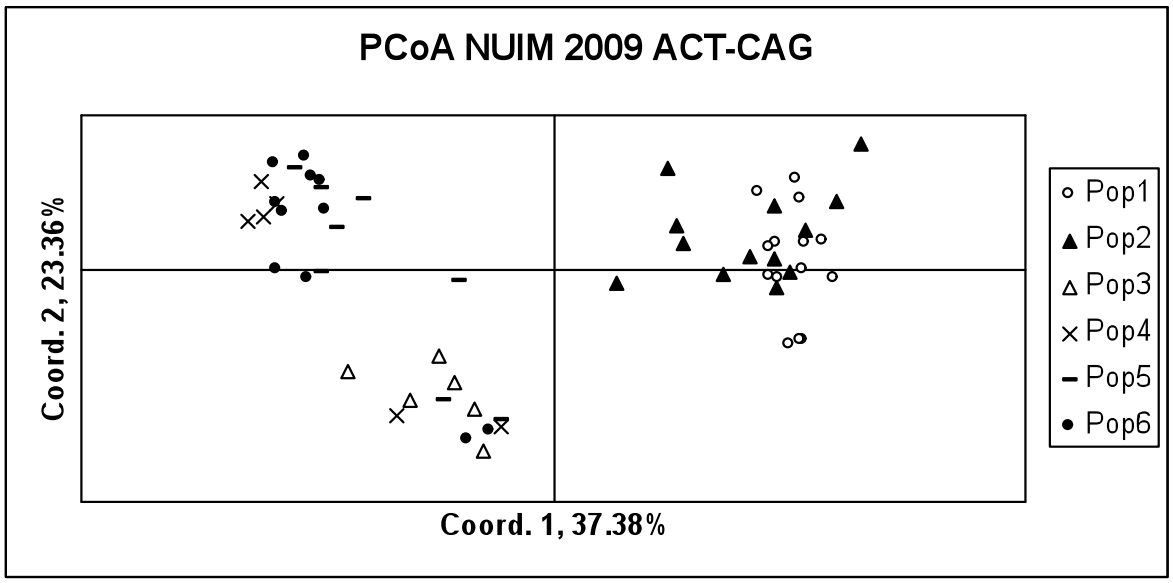


Figure G.6: Neighbour-Joining (NJ) tree and NeighbourNet network for dataset NUIM 2012 for the selective primer AGC-CTG

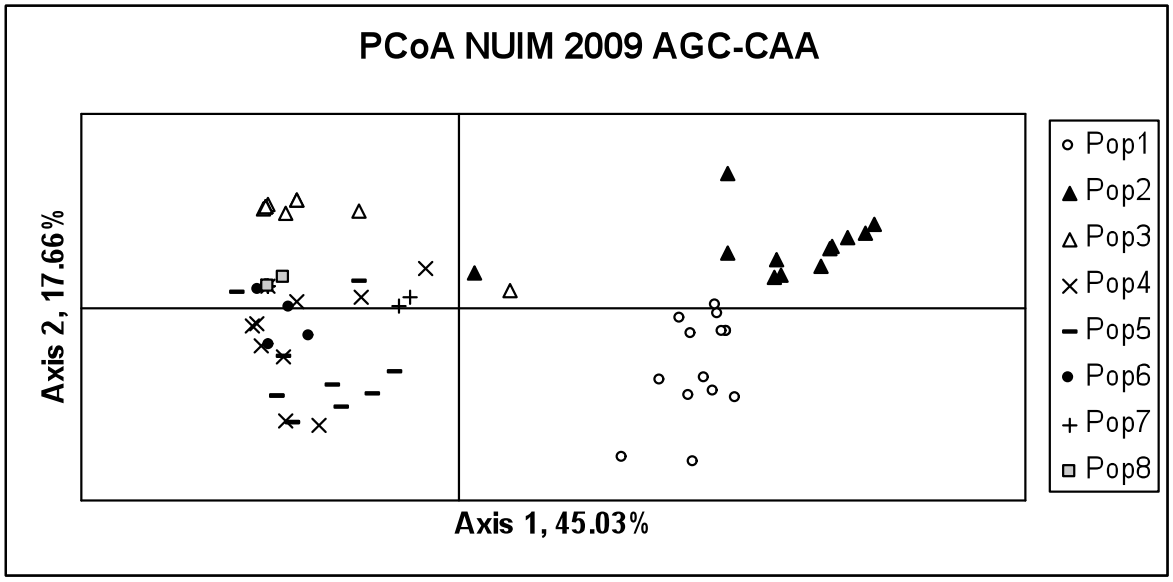
Appendix H

Principal coordinate analysis and the presence of AFLP fragments from the individual selective AFLP primer combinations



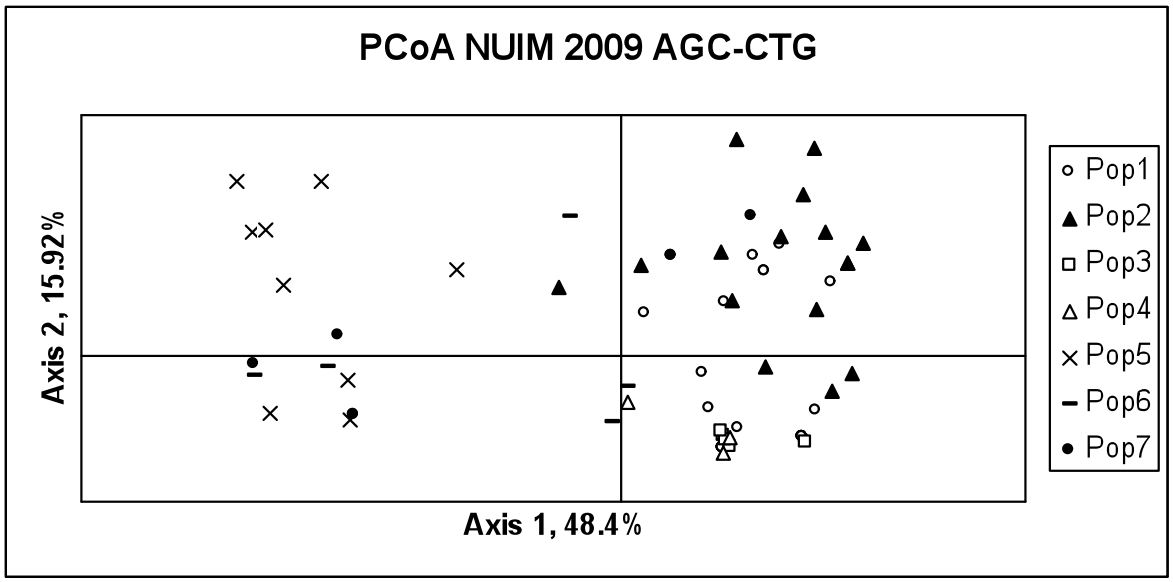
Dataset: NUIM 2009 ACT-CAG - 65 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	14	36	36	1	0.1
Pop2 - West Alps	13	49	49	7	0.5
Pop3 - The Burren, Ireland	6	25	25	0	0.0
Pop4 - Yorkshire, England	6	30	30	0	0.0
Pop5 - Rum, Scotland	8	31	31	0	0.0
Pop6 - Inchnadamph, Scotland	11	32	32	1	0.1

Figure H.1: Principal coordinate analysis (PCoA) for dataset NUIM 2009 for the selective primer ACT-CAG. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.



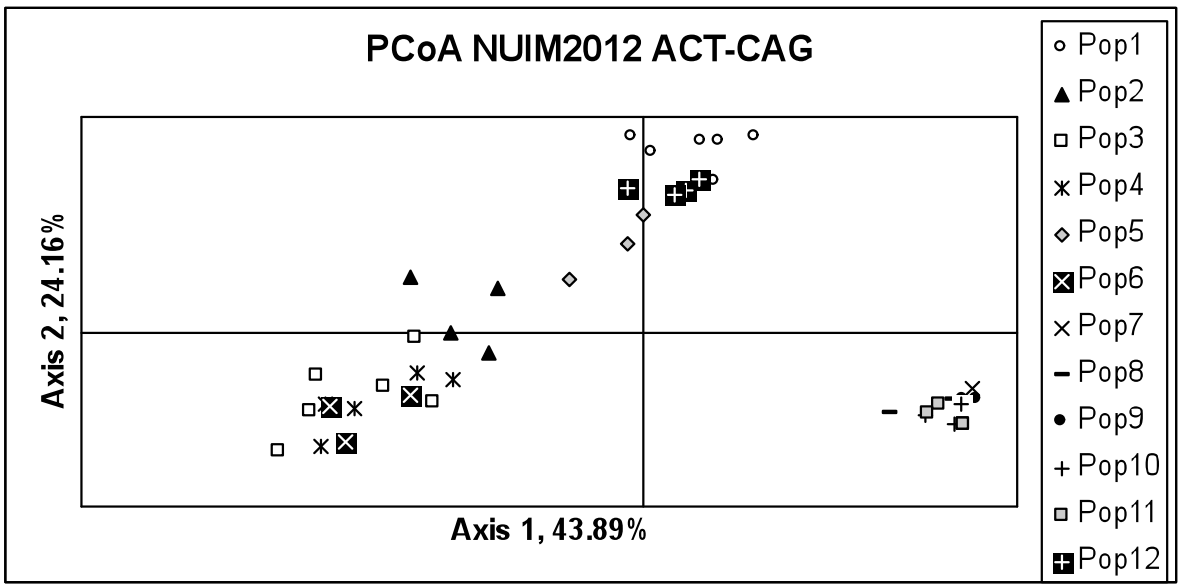
Dataset: NUIM 2009 AGC-CAA - 47 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	14	22	22	0	0.0
Pop2 - West Alps	12	33	33	6	0.5
Pop3 - The Burren, Ireland	11	24	24	0	0.0
Pop4 - Yorkshire, England	10	28	28	1	0.1
Pop5 - Rum, Scotland	11	26	26	1	0.1
Pop6 - Inchnadamph, Scotland	6	22	22	0	0.0
Pop7 - Shetland, Scotland	6	19	19	0	0.0
Pop8 - Iceland	2	18	18	0	0.0

Figure H.2: Principal coordinate analysis (PCoA) for dataset NUIM 2009 for the selective primer AGC-CAA. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.



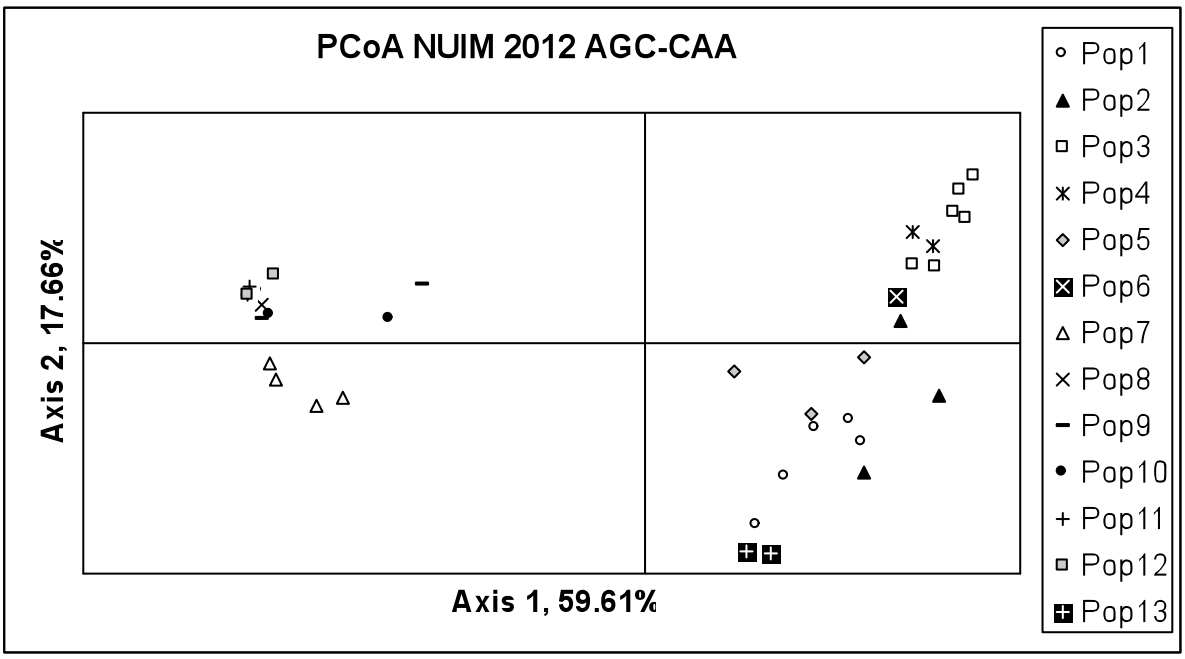
Dataset: NUIM 09 AGC-CTG - 34 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	29	16	10	2	0.1
Pop2 - West Alps	16	21	21	5	0.3
Pop3 - Pyrenees, Spain	5	7	7	2	0.4
Pop4 - The Burren, Ireland	7	11	11	0	0.0
Pop5 - Yorkshire, England	14	16	16	0	0.0
Pop6 - Rum, Scotland	5	17	17	2	0.4
Pop7 - Inchnadamph, Scotland	7	13	13	0	0.0

Figure H.3: Principal coordinate analysis (PCoA) for dataset NUIM 2009 for the selective primer AGC-CTG. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.



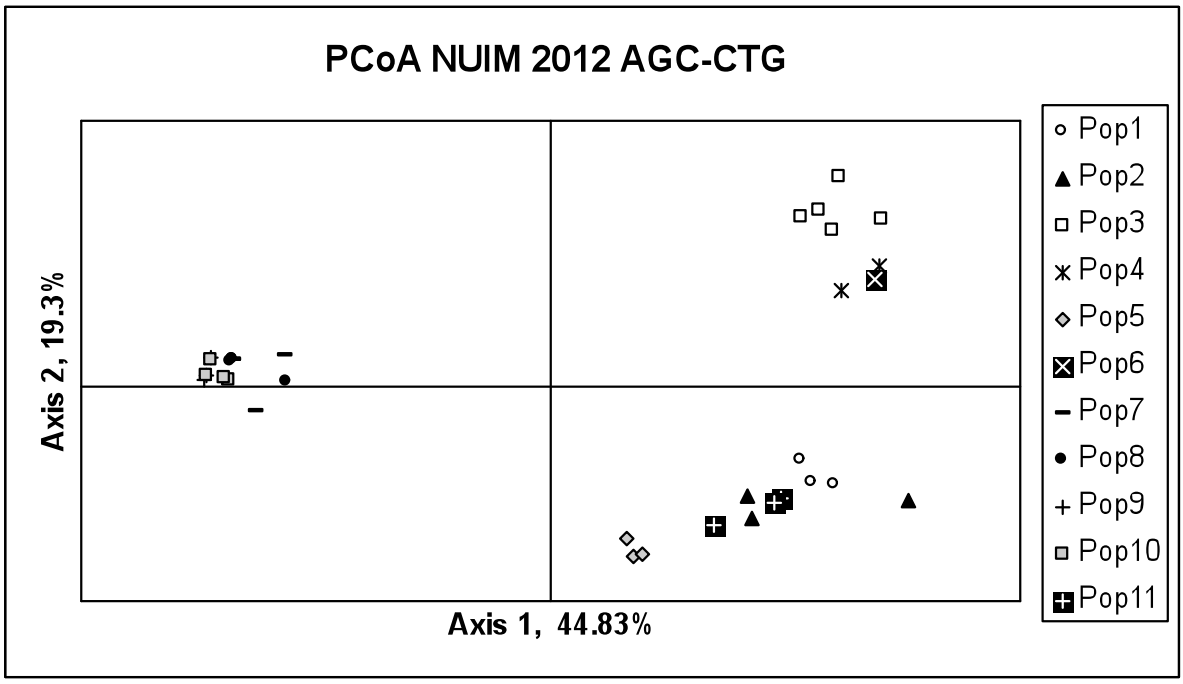
Dataset: NUIM 2012 ACT-CAG - 91 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	6	35	35	7	1.2
Pop2 - West Alps	4	42	42	6	1.5
Pop3 - Pyrenees, Spain	6	43	43	6	1.0
Pop4 - Picos De Europa, Spain	5	36	36	3	0.6
Pop5 - East Alps	3	38	38	4	1.3
Pop6 - Swiss Alps	3	25	25	2	0.7
Pop7 - Yorkshire, England	1	na	na	na	na
Pop8 - Rum, Scotland	3	30	30	1	0.3
Pop9 - Inchnadamph, Scotland	2	23	23	0	0.0
Pop10 - Shetland, Scotland	3	32	32	1	0.3
Pop11 - Iceland	5	24	24	0	0.0
Pop12 - Svalbard	4	26	26	1	0.3

Figure H.4: Principal coordinate analysis (PCoA) for dataset NUIM 2012 for the selective primer ACT-CAG. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.



Dataset: NUIM 2012 AGC-CAA - 55 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	5	19	19	1	0.2
Pop2 - West Alps	3	20	20	4	1.3
Pop3 - Pyrenees, Spain	6	23	23	2	0.3
Pop4 - Picos De Europa, Spain	2	18	18	1	0.5
Pop5 - East Alps	3	18	18	1	0.3
Pop6 - Swiss Alps	1	na	na	na	na
Pop7 - The Burren, Ireland	4	26	26	0	0.0
Pop8 - Yorkshire, England	1	na	na	na	na
Pop9 - Rum, Scotland	2	27	27	0	0.0
Pop10 - Inchnadamph, Scotland	3	26	26	0	0.0
Pop11 - Shetland, Scotland	2	27	27	0	0.0
Pop12 - Iceland	2	25	25	0	0.0
Pop13 - Svalbard	2	19	19	3	1.5

Figure H.5: Principal coordinate analysis (PCoA) for dataset NUIM 2012 for the selective primer AGC-CAA. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.



Dataset: NUIM 2012 AGC-CTG - 75 total fragments					
Populations grouped for PCoA	N	Nf	Nf freq >5%	Pf	Pf / N
Pop1 - Ben Bulben, Ireland	3	22	22	4	1.3
Pop2 - West Alps	3	32	32	10	3.3
Pop3 - Pyrenees, Spain	5	23	23	4	0.8
Pop4 - Picos De Europa, Spain	2	25	25	4	2.0
Pop5 - East Alps	3	32	32	7	2.3
Pop6 - Swiss Alps	1	na	na	na	na
Pop7 - Rum, Scotland	4	19	19	0	0.0
Pop8 - Inchnadamph, Scotland	4	19	19	0	0.0
Pop9 - Shetland, Scotland	3	20	20	0	0.0
Pop10 - Iceland	4	19	19	1	0.3
Pop 11 - Svalbard	3	16	16	3	1.0

Figure H.6: Principal coordinate analysis (PCoA) for dataset NUIM 2012 for the selective primer AGC-CTG. Symbols refer to the AFLP genotypes presented in each individual and the percentage of variation explained by each axis is shown. Detailed are the number of individuals (N), the number of fragments per population (Nf), the number of fragments with a frequency >5 % (Nf freq > 5%), the number of private fragments (Pf) and the number of private fragments per individual (Pf / N) occurring in each population.