LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Genetics, climate and phylogeography; understanding the Hiberno-Iberian distribution of *Daboecia cantabrica* (Huds.) K. Koch and *Erica mackaiana* Bab. (Ericaceae)

Thesis submitted for the degree of

Doctor of Philosophy

Trinity College Dublin

Sarah Rosarine Pene Eftonga

May 2013

TRINITY COLLEGE
1 9 JUN 2013
LIBRARY DUBLIN

Thesis 10124

DECLARATION

I, Sarah Eftonga, declare that this thesis has not been submitted as an exercise for a degree at this, or any other university, and that it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

"The Irish heathers have long been as troublesome as they are beautiful."

R. L. Praeger

(Irish Naturalists Journal, 1946, vol. 8, p324)

SUMMARY

Daboecia cantabrica (Huds.) K. Koch and *Erica mackaiana* Bab. are two ericaceous species whose disjunct distribution in Europe is best described as Hiberno-Iberian. Both species have their origins in the Atlantic coastal areas of the Iberian peninsula, are absent from Britain, and are found only in certain areas of the west of Ireland. How and when the current populations of these species became established in Ireland has been a question of debate for over 150 years. To gain an understanding of the processes and events that have shaped this unusual distribution a combined approach of molecular analysis and species distribution modelling was carried out.

Location data of extant populations of both species across their Irish and continental ranges were linked to the bioclimatic factors and soil characteristics at these locations. From this species distribution models were constructed and used to predict the full extent of the potential range of each species under current climatic conditions, conditions at the Last Glacial Maximum (LGM) and at the Last Interglacial, and conditions in the 2080s.

The Irish populations occupy a range of climatic conditions which are substantially different from those occupied by the continental populations. The potential range of each species under current climatic conditions is more extensive than the actual range on the continent as well as in Ireland. There is strong evidence that soils play a part in limiting range expansion. Hindcasting the model to the LGM shows the potential range of both species occurring further north than their current distributions, and in the case of *E. mackaiana* reaching the coast of present-day Ireland. However recent advances in dating and delimiting the extent of the ice sheet over Ireland preclude definitively the possibility of any in-situ or proximal refuge in that area. Forecasting to the future shows that even under conservative climate change scenarios, the range of suitable conditions on the continent will drastically diminish, whilst those in Ireland remain largely unchanged.

DNA sequencing was carried out on two chloroplast regions and one nuclear region of the genome, all of which were non-coding and therefore selectively neutral. From these sequences individual haplotypes were identified. Haplotype networks were constructed to demonstrate the relationships between haplotypes, and the distribution of these haplotypes across the geographical range of each species was mapped.

Across the two species 17 chloroplast haplotypes and 30 nuclear haplotypes were found, consistent with the higher rate of substitution in nuclear DNA. For *D. cantabrica* 13

chloroplast haplotypes and 26 nuclear (ITS) haplotypes were identified. Six of these ITS haplotypes were unique to Ireland. The number of haplotypes identified for *E. mackaiana* was much lower, consisting of four chloroplast and four ITS haplotypes. Of these, one chloroplast haplotype and two ITS haplotypes were unique to Ireland.

Amplified Fragment Length Polymorphisms (AFLPs) were used to determine the extent of genetic diversity and genetic differentiation in continental and Irish populations of both species. The AFLPs were also used to determine genetic distances between individuals and between populations, and these distances were depicted on UPGMA and Neighbour-Joining trees. Genetic differentiation between Irish and continental populations was evident in both species. For *D. cantabrica*, population structure highlighted a differentiation between eastern and western Iberian populations, with Irish populations showing a greater similarity to the eastern lineages than to those in the west.

Taken in total the results of the study indicate that although they have not persisted in or near Ireland throughout the last ice age, both species have been present in Ireland for a considerable time period, precluding the possibility of human introduction, although the status of the Kerry and Mayo *E. mackaiana* populations is less clear. A colonisation route into Ireland directly from the Iberian peninsula and not through Britain is the most likely scenario for their establishment in Ireland. More fossil evidence and more detailed molecular dating of DNA sequence divergence is needed to narrow down the timescale of when this occurred. However, the fact that they almost certainly do not have their origins in anthropogenic introduction indicates that they are native to Ireland and therefore an important component of the Irish flora.

Their continental range is shown to be threatened by future climate change, and their capacity for range expansion is already limited, raising the possibility that Ireland could become a stronghold for both species in Europe. It is therefore recommended that both species be given some measure of specific conservation protection in Ireland, which currently neither of them has. This study has elucidated some of the dynamics underlying the distribution of these two species of the Irish flora, as well as contributed to the wider body of knowledge on the postglacial floral recolonisation of Western Europe.

ACKNOWLEDGEMENTS

This project was funded by a postgraduate scholarship from the Irish Research Council for Science, Engineering and Technology (IRCSET).

First and foremost I would like to thank my supervisor Professor John Parnell for my initial introduction to Trinity as a research assistant, his encouragement of my plans to undertake a PhD, and his guidance and supervision throughout this project.

Many thanks go as well to Dr Trevor Hodkinson, not only for his expertise with all things molecular, but also for his unwavering positivity and support over the last three and half years. Other staff members of the TCD Botany Department from whom I received advice and guidance are Dr Fraser Mitchell, Dr Stephen Waldren and Dr David Byrne.

For their assistance in the field in Spain and Portugal I would like to thank Dr Jaime Fagundez and Dr Marisa Vera, as well as Tony Rieser.

After fire and flooding decimated the Botany Department lab, the completion of my molecular work was only made possible with the invaluable service provided by Source Biosystems, under the expert (and good-humoured) management of Dr Barry Murphy and Dr Eleanor Cunningham.

A very big thank you to the Botany technical staff, Siobhán McNamee, Trish Coughlan and Jacqui Stone, not only for ensuring that the Department doesn't fall apart from one day to the next, but also for their fantastic friendship and support over the years.

Very special thanks go to Marcella Campbell, for putting a roof over my head during my first week in Ireland, and for sharing with me her astute insights on love, life, and surviving a PhD in the Trinity College Botany department.

To my first Irish office buddy, JP Corcoran, a very big thank you for such a fun year, where we hatched all sorts of crazy plans (e.g. doing a PhD). And to Cinzia, for sharing the boredom (and the travel perks!) of the scanning project and for all the laughs we had on the way.

A huge thank you to the awesome crew of Botany postgrads (past and present) and postdocs; Karen, Langan, Miles & Aoife, Ruth, Wuu, Niamh, Brendan, Nat, Chloe & Sam, Dave, Caroline, Jesko & Sarah, Muriel, Paola, Anke, Eileen and Annelies, for all the good

times and general mayhem. A special shout-out to my fellow sufferers of AFLP torment; Emer and Christina - may your peaks rise clear and strong above the noise!

Massive gratitude also goes particularly to Nova, Dara, Max, Ellen and Caoimhe for their combined efforts at getting me through the horrific writing-up stage, with a highly efficient strategy of unflagging encouragement, relentless nagging, culinary treats, and threats of violence.

A very special acknowledgement to Ursula King and Maximilian von Sternburg, my two comrades-in-arms of the "Phylogeography research group" for keeping me sane, and thanks to whom the long and winding road was never a lonely one.

Special thanks as well to Danny, for always insisting this was something I should do, and for trudging with me through the bogs (literally as well as figuratively!). And to my friends from home: Adele, Losana, Zaif, Shirleen and Anju, who have always been there with encouragement when I needed it, despite my woeful efforts at staying in touch. See you all again soon!

A PhD just isn't a PhD unless you have spent a significant amount of time avoiding it. Hence I'd like to thank all the gradient-chasing crazies in the Irish Mountain Running Association (especially the hardcore members of the Vulgar Callunas team), the fantastic yoga teachers at the Elbow Room and Open Minds Project, the Irish Film Institute and the Lighthouse Cinema, my fellow committee members and teachers of the Trinity Arts Workshop, the organisers and coaches of the Trinity Triathlon Club and Athletics Club, and the organisers of the Mighty Deerstalker, the Achill Roar and the Dublin and Berlin marathons. Procrastination rocked!

To the English branch of the family, especially Lucy, Tom, Bea and Jerry, thank you for all the kindness and happy times, it's been great having you just a hop across the water.

Lastly and most importantly to the family who were a bigger hop away across a slightly bigger water; my big brother Conway, and my parents, Kafoa and Frances, for all the support (both emotional and practical) during the PhD, and all the way back to the beginning too – thank you and fại'ȧk se'ea.

DEDICATION

This thesis is dedicated to the memory of

Maximilian von Sternburg

my colleague, my teacher, and my dear friend,

whose companionship and support echo on every page,

and whose keen wit and gentle heart are missed every single day.

TABLE OF CONTENTS

CHAPTER 1 General Introduction 1

1.1 The Irish Flora and the 'Lusitanian' Element 1
1.1.1 The Irish Flora 1
1.1.2 The 'Lusitanian' Flora and Fauna 1
1.1.3 Hiberno-Iberian Heathers Studied in this Thesis 3
1.2 Introduction to the Ericaceae and the species *Daboecia cantabrica* and *Erica mackaiana* 3
1.2.1 The family Ericaceae 3
1.2.2 *Daboecia cantabrica* (Huds.) K. Koch 5
1.2.3 *Erica mackaiana* Bab. 11
1.3 Theories to explain the disjunct distribution 14
1.3.1 Disjunct distributions – an introduction 14
1.3.2 The fossil record 14
1.3.3 Postglacial Recolonisation of Europe – an overview 15
1.3.4 Hypotheses for the current disjunct distribution of the study species 19
1.4 Approaches and techniques selected for this study 24
1.4.1 Species Distribution Modelling 25
1.4.2 Comparative DNA sequencing 26
1.4.3 Amplified Fragment Length Polymorphism (AFLPs) 27
1.5 Aims of this study 27
CHAPTER 2 Species Distribution Modelling 29

2.1 Introduction 29
2.1.1 Climatic and edaphic requirements of study species 30
2.1.2 Species Distribution Modelling 32
2.1.3 Aims 35
2.2 Methodology 36
2.2.1 Occurrence data and georeferencing 36
2.2.2 Environmental data 36
2.2.3 Soil Data 38
2.2.4 Selection of variables for analysis 39
2.2.5 Description of the niche 39
2.2.6 Construction of the SDM 39
2.2.7 Principle of MaxEnt 40
2.2.8 Model validation 40

2.2.9 Temporal projection of the model 41
2.3 Results 41
2.3.1 Current actual distribution 41
2.3.2 Bioclimatic ranges for study species 51
2.3.3 Selection of environmental variables for SDM 51
2.3.4 Visualising the climatic envelope 52
2.3.5 Model validation 56
2.3.6 Modelling Results 57
2.4 Discussion 69
2.4.1 Insights into climatic ranges tolerated by the two species 69
2.4.2 Continental range over-predictions 69
2.4.3 Irish range predictions 70
2.4.4 Range in Britain 71
2.4.5 Irish refugium 72
2.5 Long distance dispersal in postglacial times 72
2.5.1 Climate change consequences 73
CHAPTER 3 Comparative DNA Sequencing and Haplotype Analysis 75
3.1 Introduction 75
3.1.1 Molecular markers 75
3.1.2 Comparative DNA sequencing 75
3.1.3 Haplotypes and networks 77
3.1.4 Molecular dating 77
3.1.5 Hybrid investigation 78
3.1.6 DNA sequence analysis and infraspecific phylogeography 78
3.1.7 Aims 80
3.2 Methodology 81
3.2.1 Sampling 81
3.2.2 DNA Extraction 81
3.2.3 Amplification of markers 82
3.2.4 Haplotype networks and mapping 87
3.2.5 Phylogenetic trees 87
3.2.6 Molecular clock test 88
3.3 Results 89
3.3.1 Chloroplast haplotype distribution 89
3.3.2 Nuclear ribosomal DNA (ITS) haplotype distribution 96
3.3.3 Molecular clock test 104

3.3.4 Genetic evidence for the parentage of hybrid *E. x stuartii* 104
3.4 Discussion 106
3.4.1 General findings 106
3.4.2 *Daboecia cantabrica* 106
3.4.3 *Erica mackaiana* 109
3.4.4 Genetic evidence for *Erica* x *stuartii* parentage 109
CHAPTER 4 Genetic Diversity and Differentiation 111

4.1 Introduction 111
4.1.1 Principle of fingerprinting techniques and AFLPs 111
4.1.2 Principles of Population Genetics 113
4.1.3 Previous investigations 114
4.1.4 Aims 115
4.2 Methodology 115
4.2.1 Sampling 115
4.2.2 AFLP Protocol 116
4.2.3 Fragment analysis 118
4.2.4 Quality control 119
4.2.5 Scoring 119
4.2.6 Dataset compilation 119
4.2.7 Software Packages Used in Data Analysis 120
4.2.8 Analysis 121
4.3 Results 127
4.3.1 Dataset Overall and Quality Control 127
4.3.2 Genetic diversity 127
4.3.3 Genetic Distance 134
4.3.4 Population Differentiation 144
4.4 Discussion 150
4.4.1 *D. cantabrica* discussion 150
4.4.2 *E. mackaiana* discussion 154
CHAPTER 5 General Discussion 157

5.1 Summary 157
5.1.1 Chapter 2 157
5.1.2 Chapter 3 158
5.1.3 Chapter 4 158
5.2 Synthesis of findings 159

5.2.1 Synthesis of findings for *D. cantabrica* 159
5.2.2 Synthesis of findings for *E. mackaiana* 161
5.2.3 Synthesis of findings for *E.* x *stuartii* 163
5.3 European postglacial colonisation and the Iberian peninsula 163
5.4 Implications 164
5.5 Methodological limitations 165
5.5.1 Sampling 165
5.5.2 Sample sizes 166
5.5.3 Molecular dating methods 166
5.6 Future Research 166
5.7 Concluding remarks 167
References 169

CHAPTER 1 General Introduction

Daboecia cantabrica (Huds.) K. Koch and *Erica mackaiana* Bab. are two ericaceous species that are part of what has come to be known as the 'Lusitanian' element of the Irish flora. Both species have their origins in the Atlantic coastal areas of the Iberian peninsula, are absent from Britain, and are found only in certain areas of the west of Ireland (Webb 1983). How and when the current populations of these species became established in Ireland has been a question of debate for over 150 years.

1.1 The Irish Flora and the 'Lusitanian' Element

1.1.1 The Irish Flora

There are around 1800 vascular plant species in the Irish Flora, if both native and non-native species are considered (Reynolds 2002). Whilst differing estimates exist as to the number of species considered native to Ireland it is in the region of between 815 (Webb 1983) to approximately 1000 (Curtis and McGough 1988). This is a greatly reduced subset of the total number of native species in Britain, which is estimated at around 1350 (Preston *et al.* 2002). This reduction in size of the native flora is attributable to the fact that Ireland became geographically isolated from continental Europe possibly as long as 7000 years before Britain did (Lambeck 1995, Edwards and Brooks 2008).

Whilst there are extremely few species of vascular plants endemic to the island of Ireland (Rich 2005), there are fifteen species of native Irish plants that are absent from Britain, but which are native to other parts of continental Western Europe (Webb 1983).

1.1.2 The 'Lusitanian' Flora and Fauna

The term 'Lusitanian' has come to refer to those species in the Irish flora that are found in greatest abundance in the Iberian peninsula (and sometimes southern France and Mediterranean areas as well), but are either entirely absent from, or found only in small disjunct populations in, the intervening area i.e. the rest of France and the island of Britain (Webb 1983). Table 1 summarises these species, and also includes for comparison those members of the Irish fauna that share a similar distribution.

Table 1: The 'Lusitanian' flora (green) and fauna (red), compiled from: Praeger (1939), Webb (1983) and Vincent (1990).

	Species name	Family	Irish Range	Continental range
Absent from Britain	*Arbutus unedo* L.	Ericaceae	W	Portugal, Spain, France and E of the Mediterranean
	Daboecia cantabrica (Huds.) K. Koch	Ericaceae	W	Azores, N Portugal, N Spain, S, W France
	Erica erigena R. Ross	Ericaceae	W	Portugal, Spain
	Erica mackaiana Bab.	Ericaceae	W	N Spain
	Minuartia recurva Schinz & Thell	Caryophyllaceae	W	N Portugal, N Spain, S France, E of the Mediterranean
	Neotinea maculata (Desf.) Stearn	Orchidaceae	W	Portugal, Spain, France and E of the Mediterranean
	Pinguicula grandiflora Lam.	Lentibulariaceae	SW	N Spain, S France,
	Saxifraga hirsuta L.	Ericaceae	SW	N Spain
	Saxifraga spathularis Brot.	Ericaceae	SW	N Portugal, N Spain
	Geomaculus maculosus Allman	Arionidae	SW	Portugal
	Lumbricus friendi Cognetti	Lumbricidae	S	Portugal, Spain, France
	Oritoniscus flavus Budde-Lund	Trichoniscidae	SE	Pyrenees
Disjunct, small populations in Britain	*Erica ciliaris* L.	Ericaceae	W, single-site, non-native[1]	Portugal, N Spain, W France
	Theba pisana Muller	Helicidae	E	SW France

[1] (Curtis 2000)

Although the term 'Lusitanian' has entered current usage, it is not an accurate one. The Roman province of Lusitania encompassed most of present day Portugal and a small part of Western Spain, as shown in Figure 1(A).

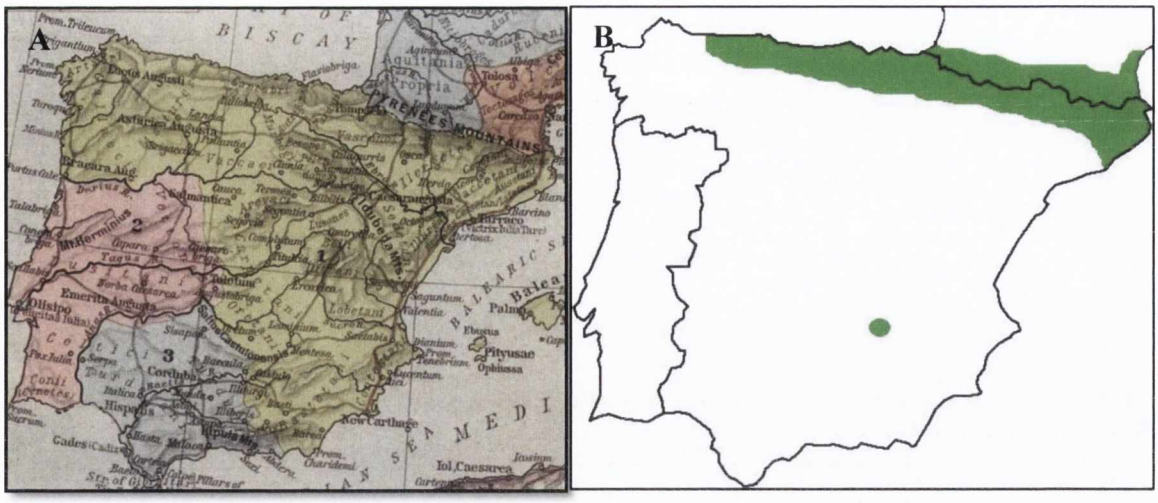


Figure 1: (A) The province of Lusitania (in pink) on a first century AD map of the Roman provinces of the Iberian peninsula (Sheperd 1911). (B) Present day borders of Portugal and Spain showing the distributions of one of the so-called 'Lusitanian' species, *Pinguicula grandiflora* in green (adapted from GBIF Portal)

Most of the species termed 'Lusitanian' have continental ranges that extend beyond these areas, and their continental range would thus be more accurately described as being Iberian. *Pinguicula grandiflora*, for example, is found across the coast of northern Spain and into southern France (Figure 1B), whilst *Minuartia recurva* can be found as far east as the Mediterranean coast of Spain.

Although the term 'Lusitanian' has maintained prominent usage in the literature (Rowe *et al.* 2006, Vialatte *et al.* 2008), 'Hiberno-Iberian' is more accurate and will be used instead. It is not a new terminology, having previously been used by Praeger (Praeger 1997).

1.1.3 Hiberno-Iberian Heathers Studied in this Thesis

As can be seen from the species list in Table 1, a large proportion of the group consist of members of the Ericaceae. Two of these, *Daboecia cantabrica* and *Erica mackaiana* were selected as the topic of this thesis.

1.2 Introduction to the Ericaceae and the species *Daboecia cantabrica* and *Erica mackaiana*

1.2.1 The family Ericaceae

The Ericaceae is one of the largest families of the flowering plants. Current taxon data maintained jointly by Kew and Missouri Botanic gardens gives the total number of species in Ericaceae as 3343, across 145 genera, with a further 433 accepted infraspecific taxon

names (The Plant List (2010). Version 1. http://www.theplantlist.org, accessed 12/06/2011). Only 22 species of Ericaceae are found in Britain and Ireland, in 12 genera (Hill *et al.* 2004).

1.2.1.1 Family description

As the family Ericaceae is so large and highly variable, a family description can be made in only very broad terms. The following, compiled from Watson (1965), Heywood *et al.* (2007), Kron *et al.* (2002), Furness (2009) and Friis *et al.* (2011), summarises the general characteristics that distinguish the Ericaceae from other flowering plants.

- **Flowers** are typically actinomorphic and hermaphrodite, sometimes occurring as solitary flowers, but more commonly as an **inflorescence** in a raceme, panicle or corymb.
- The **corolla** is generally lobed, forming a tubular, bell or urn shape, consisting of 4-5 fused petals.
- Bracts or bracteoles are present at the base of the **pedicels.**
- 8-10 **stamens** (usually double the number of petals), with appendages (typically awns or spurs) and dehiscing via short slits.
- The **ovary** is typically superior, and commonly contains 4-5 locules.
- **Fruits** are usually small loculicidal or septicidal capsules, drupes or berries.
- **Leaves** lack stipules, are usually simple and alternate, occasionally opposite or spirally arranged, often with strongly revolute margins.
- **Pollen** is usually shed as a permanent tetrahedral tetrad, however the Ericaceae are unusual in that many members of this family display variable sterility within the tetrad, thus producing also monads, dyads and triads.
- Most members of the Ericaceae are predominantly **shrubs,** some are climbers or trees, only a few are herbaceous.
- Most Ericaceae are found in nutrient-poor, usually acidic **soils**.

1.2.1.2 Ericoid mycorrhizae

One of the reasons for the success of the Ericaceae across such a diversity of habitats, many of which are nutrient deficient, is that they are capable of getting nutrition by forming mychorrizal associations with fungi in the soil (Read 1991). This symbiotic association is formed when a fungus colonises the plant root system, thereby extending the

plant's reach into the soil and aiding in the increased uptake of nutrients and water. In return the host plant supplies carbohydrates and other organic compounds to the fungus.

The formation of mycorrhizae is a phenomenon found in approximately 85% of angiosperm species (Wang and Qiu 2006), however there is a particular form of the symbiosis that is most commonly found among members of the Ericaceae and other closely related families, known as ericoid mycorrhizae (Read 1996).

1.2.1.3 Origins and phytogeography of the Ericaceae

The Ericaceae are cosmopolitan in their distribution, occurring on all continents except Antarctica, and extending from the tropics into the temperate zones, up to as far as 80°N (Stevens 1971). In tropical areas the Ericaceae are predominantly found at high altitudes (Watson 1965). Areas with the highest species diversity are the Cape region of South Africa, the Himalayas, the Andes and the South-West Pacific, including Australia (Kron and Luteyn 2005). Fossil records from the Cenozoic have shown that the Ericaceae were once much more diverse in Europe than they are currently (Friis *et al.* 2011).

The Ericaceae were thought up until recently to be of Gondwanan origin, a conclusion based largely on the fact that most of the centres of highest genetic diversity were located in former Gondwanan landmasses, however recent phylogenetic analysis has shown that the Ericaceae are in fact a Laurasian-origin family (Kron and Luteyn 2005).

The most closely related families to the Ericaceae are the Clethraceae and the Cyrillaceae (APGIII 2009).

1.2.2 *Daboecia cantabrica* (Huds.) K. Koch

1.2.2.1 Nomenclature and Classification

Daboecia cantabrica has long posed a problem to taxonomists who attempted to classify it, as evidenced by the many name changes it has undergone since its first description, the most notable of which are summarised in Table 2.

Linnaeus originally gave it the binomial *Erica daboeci* (1762) but later revised this to *Andromeda daboecia* (1767). The epithet *daboeci* was taken from the name Dabeoc, an Irish saint thought to have lived around the end of the 5th century (Nelson 1984). This epithet was lost however in further published accounts of the species as it went on to be named as *Erica cantabrica* (Tournefort 1700) and later as *Menziezia polifolia* (Jussieu 1802). The classification dilemma was finally settled by the creation of a new genus

Daboecia (Don 1834). The specific epithet *polifolia* persisted however, until it was corrected and the present name *Daboecia cantabrica* was finally instated (Koch 1869).

The designated lectotype of *D. cantabrica* is an illustration by James Petiver (Nelson 1984), shown in Figure 2.

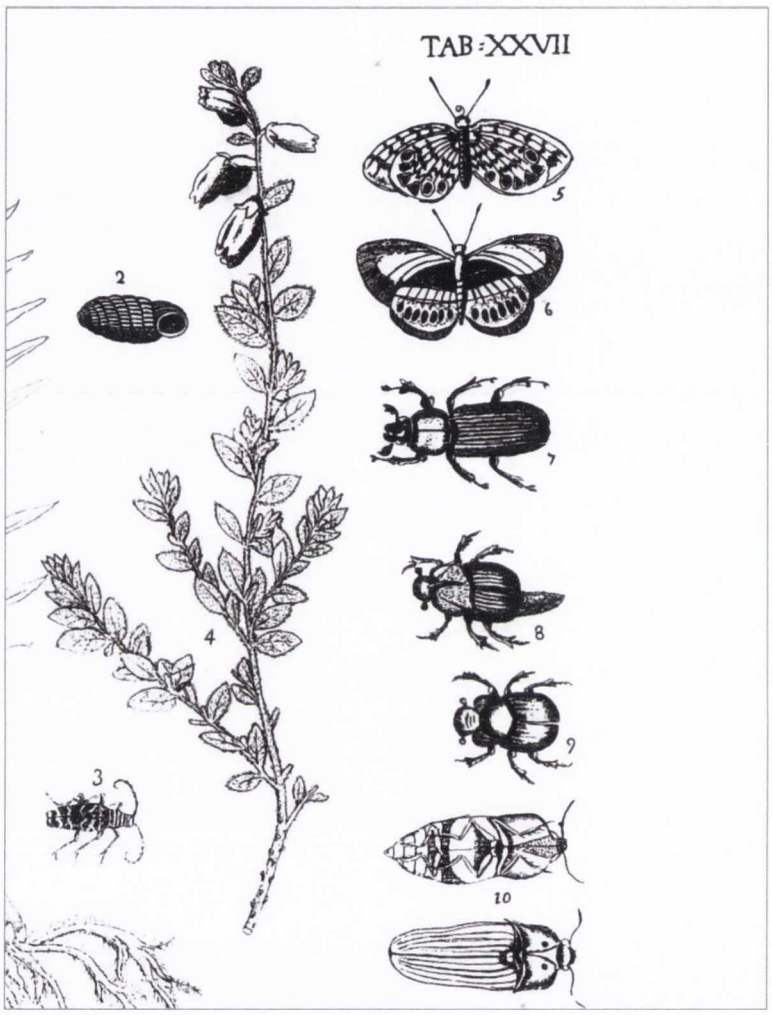


Figure 2: The designated lectotype of *D. cantabrica*, an illustration by James Petiver in *Gazophyllacii Naturae Decas Sexta*, reproduced from the copy in the type folder held in the National Botanic Gardens, Glasnevin.

Daboecia cantabrica has many morphological characteristics that differentiate it from other members of the Ericaceae and which have tended to confound efforts to accurately place it in a phylogenetic framework. The various classifications that have been produced are summarised in Table 2. Until recently general consensus placed it within the subfamily Rhododendroideae. Its distinguishing characteristics led to its inclusion within a newly created monotypic subtribe Daboecieae (Cox 1948). Molecular analysis later reorganised much of the classification of the Ericaceae. The subfamily Rhododendroideae was determined to be paraphyletic and was reorganised and amalgamated into the subfamily

Ericoideae. *Daboecia* is currently assigned to the tribe Ericeae within the subfamily Ericoideae (Kron *et al.* 2002).

Table 2: Nomenclature and placement of *D. cantabrica* within the Ericaceae, according to various authors

Early authors (no subfamilial classification)	Don (1834)
Linnaeus (1762) ***Erica daboecii*** Linnaeus (1767) ***Andromeda daboecia*** Tournefort (1700) ***Erica cantabrica*** Jussieu (1802) ***Menziezia polifolia***	Tribe Ericeae ***Daboecia polifolia*** (subtribe Andromedae) Tribe Rhodoreae Tribe Vaccinieae Tribe Pyroleae Tribe Monotropeae Tribe Epacridae
Bentham & Hooker (1876)	**Drude [36][36]**
Tribe Arbuteae Tribe Andromedeae Tribe Ericeae Tribe Rhodoreae ***Daboecia polifolia* Don.** Tribe Pyroleae	Subfamily Rhododendroideae ***Daboecia polifolia* D. Don** Subfamily Arbutoideae Subfamily Vaccinioideae Subfamily Ericoideae
Stevens (Stevens 1971)	**Kron *et al.* (2002)**
Subfamily Rhodendroideae ***Daboecia cantabrica*** (Tribe Daboecieae) Subfamily Ericoideae Subfamily Vaccinoideae Subfamily Pyroloideae Subfamily Monotropoideae Subfamily Wettstenioideae	Subfamily Enkianthoideae Subfamily Monotropoideae Subfamily Arbutoideae Subfamily Cassiopoideae Subfamily Ericoideae ***Daboecia cantabrica*** (Tribe Ericeae) Subfamily Harrimanelloideae Subfamily Styphelloideae Subfamily Vaccinoideae

1.2.2.2 Morphological description

Daboecia cantabrica is a woody branching shrub, decumbent at first at then growing erect up to 50-60cm in height. The leaves are evergreen, oval-elliptical with revolute margins. The undersides of the leaves are covered in dense white hairs; the top surfaces are dark green and sparsely covered with glandular hairs. Glandular hairs are also found on the corolla, the stems, the calyx and the capsule.

The flowers are held in terminal racemes (see Figure 3), each containing 3-12 pendent flowers. The corolla is 8-12mm long, purple and urceolate with 4 lobes. The calyx consists

of 4 dark green triangular sepals. The fruit is a septicidal ovoid capsule, with 4 valves. The seeds are reddish brown, broadly elliptical-ovate and 0.65-0.75mm in size.

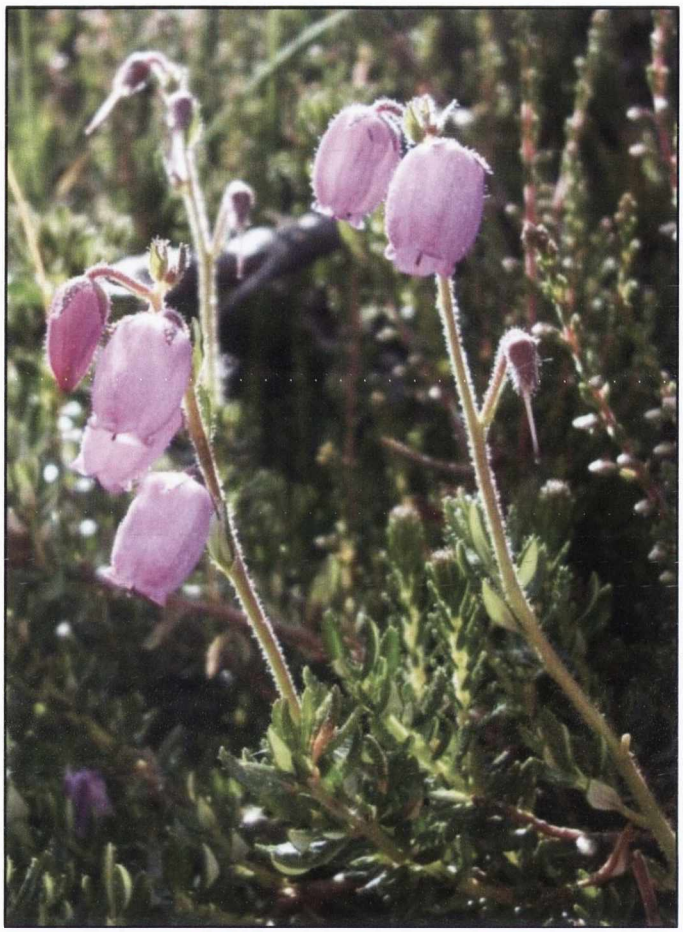


Figure 3: *D. cantabrica* growing on the cliffs at Cabo Peñas in the Asturias region of Spain.

Daboecia cantabrica shows very little morphological variation throughout its range (Woodell, 1958), with only one subordinate taxon located in the Azores. This taxon, named *D. cantabrica* subsp. *azorica*, is distinguished by a lack of glandular hairs on the corolla as well as a reduced size and compact habit (McClintock 1989). Its flower colour is more crimson than that of *D. cantabrica* (see Figure 4). It was originally classified as a separate species; *Daboecia azorica* Tutin and Warburg. The Azorean subspecies has been collected and recorded from the islands of Pico and Faial (Tutin 1953) and also grows on the island of Sao Jorge (Dias, pers. comm.)

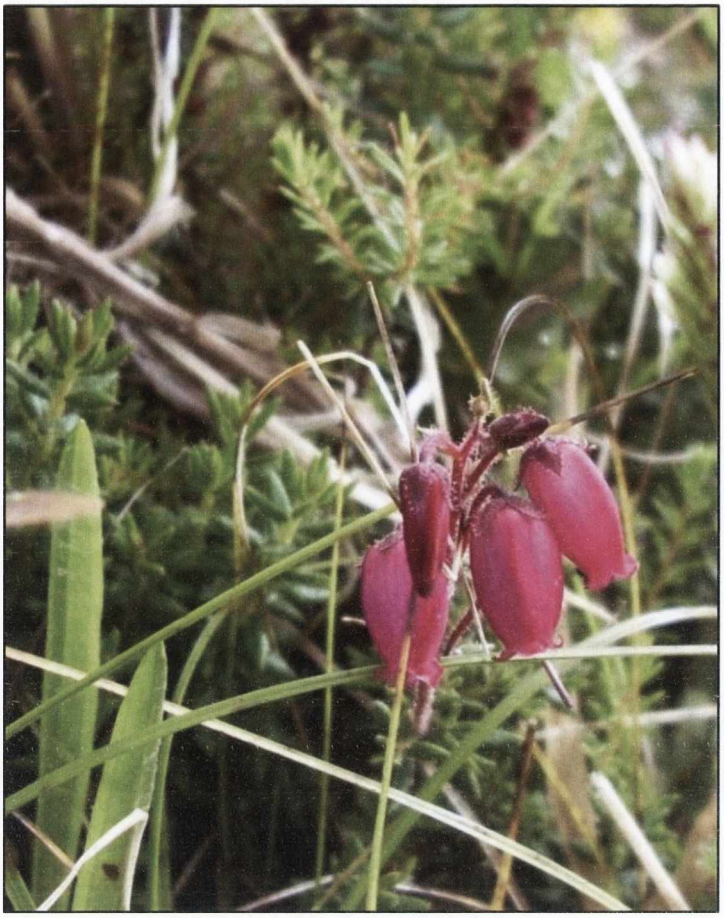


Figure 4: *D. cantabrica* subsp. *azorica* growing on the caldera on the island of Faial.

A white-flowered variety of *D. cantabrica* from Mayo was described by Mackay in 1830 (Scannell 1983), and is now grown in cultivation. The Heather Society (a British organisation dedicated to the horticulture of and research into heathers) has a record of nearly 60 cultivars of *D. cantabrica*. These vary mainly in the colour and shape of the corolla but also in leaf colour, and habit.

1.2.2.3 Distribution

The distribution of *D. cantabrica* is shown in Figure 5. In Ireland, *D. cantabrica* is confined to the counties of Mayo and Galway, in an area bounded to the north by Clew Bay and to the south by Galway Bay (Woodell 1958). To the east, its range edge has been noted to coincide with the change of substrate to carboniferous limestone (Woodell 1958).

The northernmost continental populations are in France in the region known as the Pays de la Loire, where several small populations can be found in the area around St-Cyr-des-Gats in Maine-et-Loire, as well a very small population about 100km away in the Forêt de Brissac. Two historical populations recorded in the literature from the Gironde (Laterrade 1832) and from Tarn-et-Garonne are no longer extant, the latter not having been seen since

1844 (Leblond, pers. comm.). Due to its rarity north of the Pyrenees *D. cantabrica* is a protected species in France (Le Bail 2008).

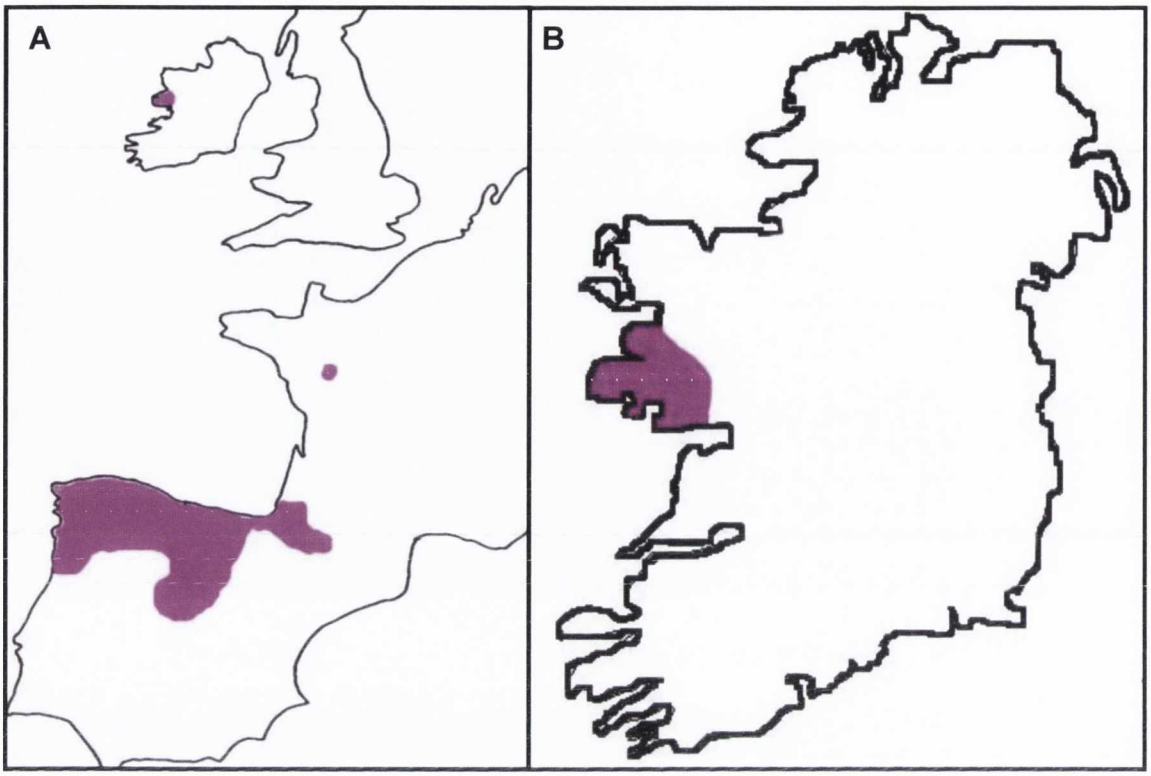


Figure 5: Distribution of *D. cantabrica* in Europe (A) and in Ireland (B). Modified from (Webb 1983).

The principal continental population begins in southern France near the Spanish border, and extends inland for approximately 120km, reaching its limit at the valleys of Ouzum and Arrens (Dupias). The species then extends southward across the border into northern Spain, along the length of the Iberian Peninsula, and down its western coast into northern Portugal. The southernmost record of *Daboecia* is in the Sierra de Gata (Franzi 1984).

1.2.3 *Erica mackaiana* Bab.

Erica mackaiana was found in Craiggamore, Connemara, County Galway in July 1835 by William McCalla (Nelson 1981). It was simultaneously discovered that same summer in the Sierra del Peral, in the Asturias by Durieu de Maisonneuve and described by Jacques Gay (Lacaita 1929). Hooker (1835) named it *Erica mackaii* after the Irish botanist James Townsend Mackay. The name was later amended to *E. mackaiana* in Mackay's Flora Hibernica (1836).

1.2.3.1 Morphology

Erica mackaiana is a dwarf shrub, around 20-30cm high, with numerous branchlets (Webb 1955). The inflorescence consists of a terminal umbel of 10-12 flowers. Flowers can vary in colour from pink to purple. The corolla is 5.5-7cm long, urceolate-tubular, with 4 lobes, the tips of which curve outwards.

Leaves are elliptical to oblong-lanceolate, arranged in whorls of four, and usually less than 5mm in length. The very short petiole is adpressed to the stem, the leaf itself spreads out. Leaves are dark green on the top surface, and white on the bottom surface, with a green midrib. Long glandular hairs are present along the leaf margins, as well as on the sepals.

The ovary is glabrous and pale green. The fruit is a subglobose capsule. Seeds are 0.4mm, elliptical-oblong and reticulate (Fagundez *et al.* 2010).

Figure 6: *E. mackaiana* growing in the Sierra de Meira, north-western Spain.

Overall there are no constant morphological differences between the Spanish and Irish populations (Webb 1955). However there do exist some variations in corolla shape, size and colour shade in the Irish populations, as well as some malformations of the stamens, stigma and styles (Nelson 1989).

One of the Irish populations (at Carna, County Galway) is distinct from all other Irish and Spanish populations in having no glandular hairs (McClintock 1983). The population at Lough Nacung, Co. Donegal is also described as having a distinctive habit, being somewhat taller and bushier than the usual form (van Doorslaer 1990a).

A double flowered variety is now in cultivation under the cultivar name 'Crawfordii'. The Heather Society has records of another nine commercial cultivars, which vary in flower colour and form.

1.2.3.2 Reproduction and hybridisation

Erica mackaiana has many morphological similarities to *E. tetralix* and *E. ciliaris* and has been shown to hybridise with both; the two hybrids produced are *Erica* x *stuartii* (with *E. tetralix*) and *E.* x *praegeri* (with *E. ciliaris*). The later occurs only in Spain, as in Ireland the population of *E. ciliaris* is restricted to a single very small population. *Erica* x *stuartii* occurs in Ireland in all areas where *E. mackaiana* is found, but in Spain it is relatively rare (Fagundez 2006).

The most notable feature of *E. mackaiana* in Ireland is that it is almost completely sterile. Mitchell & Watts (1970) proposed an ecological cause for *E. mackaiana*'s sterility in Ireland. They noted that the wet blanket-bog habitats that *E. mackaiana* currently inhabits in Ireland have only come into existence as recently as 3000 years ago, and suggested that *E. mackaiana*'s sterility has arisen due to the fact that it has survived in what is essentially a relict locality with an unsuitable habitat for this species. It has however been demonstrated (Webb 1957, McAllister 1996) that Irish material crossed with Spanish material is fertile, suggesting that its sterility is the result of a self- incompatibility caused by reduced genetic variation in the Irish populations. McAllister (1996) was able to produce seed in *E. mackaiana* by growing Irish and Spanish plants together in the same pot at Ness Botanic Gardens, Liverpool.

Webb (1951) noted that the *E.* x *stuartii* hybrids in Ireland seemed to merge gradually into the *E. tetralix* form, but remained fairly distinct from *E. mackaiana*. He attributed this to

the sterility of *E. mackaiana* primarily in its ovules, thereby negating the possibility of hybrids backcrossing with *E. mackaiana*.

1.2.3.3 Distribution

In continental Europe, *E. mackaiana* is found only within northern Spain, primarily within the province of Asturias, and eastern Galicia. (Nelson 1982). In Ireland the range is highly restricted, with five populations having been recorded to date. The largest of these is the colony discovered by McCalla at Craiggamore near Roundstone in Co. Galway (Webb 1955). A second colony is located around the shore of Upper Lough Nacung in Co. Donegal (Webb 1954). A third colony was discovered in 1874 at Carna, Co. Galway by A.G. More but after some years the colony was thought to have become extinct until its rediscovery by Webb in 1969 (McClintock 1983).

The first record of *E. mackaiana* in Co. Mayo was made in the Derry Lower townland near the town of Ballina (van Doorslaer 1990b). Most recently, a population of *E. mackaiana* as well as its hybrid *E.* x *stuartii* was discovered in South Kerry (Nelson 2005). The similarity of *E. mackaiana* to *E. tetralix*, as well as confusions with their hybrid *E.* x *stuartii* may have caused under-reporting of the occurrence of this species, which may therefore have a more widespread range.

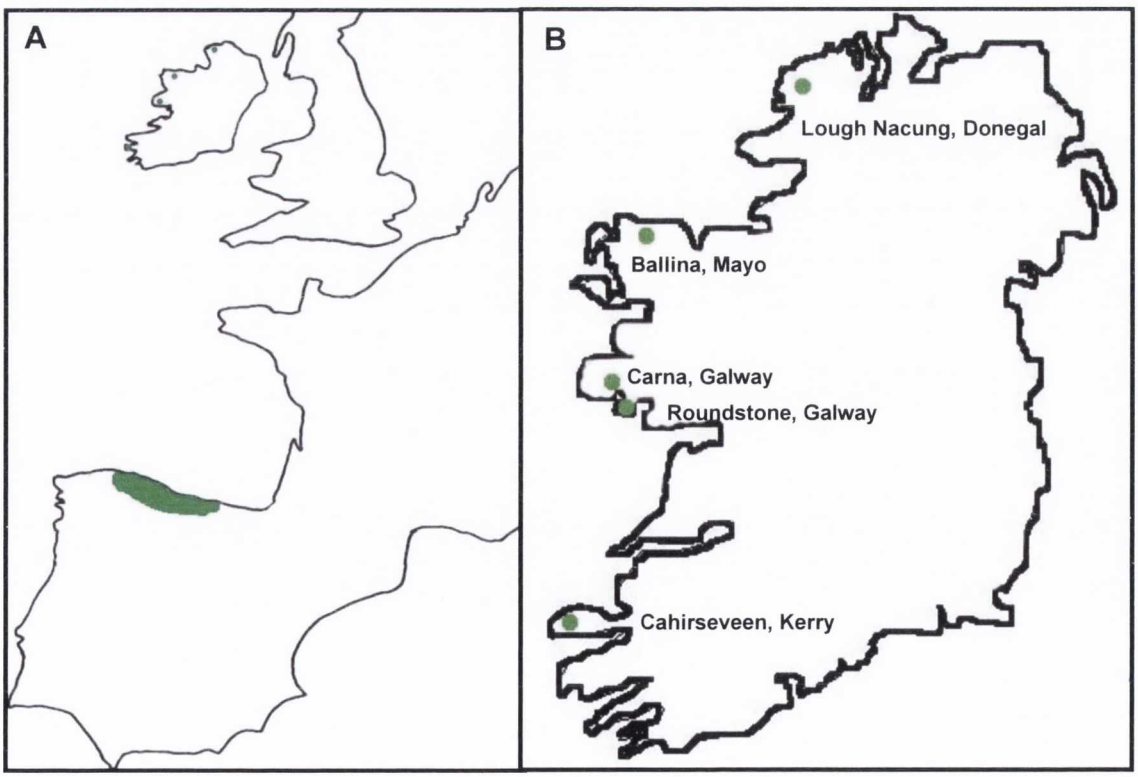


Figure 7: (A) Distribution of *E. mackaiana* in continental Europe, restricted to the north coast of Spain; (B) disjunct populations in Ireland. Modified from (Webb 1983).

1.3 Theories to explain the disjunct distribution

1.3.1 Disjunct distributions – an introduction

Disjunct ranges of plants has been a field of considerable study in all parts of the world for many years (Raven 1972). No species has a completely continuous range, as even on small scales plants will still tend to grow only in areas with suitable habitats. A range is generally considered to be disjunct if the distance between populations is larger than the species' "normal dispersal capacity" (Thorne 1972).

There are two principle factors to explain a disjunct distribution; long distance dispersal and range fragmentation.

Long distance dispersal occurs through a number of possible mechanisms associated generally with extreme weather conditions or animal migration (Normand *et al.* 2011) namely; wind (anemochory), water (hydrochory), animals (zoochory) and the intentional or unintentional action of human beings (hemerochory).

Range fragmentation refers to the process by which areas within a formerly continuous range become unsuitable for the species, usually as a result of changing climatic conditions, habitat change or biotic factors such as new predators or competitors (Hernandez-Roldan *et al.* 2011). It can also result from more drastic physical processes such as rising sea levels, or plate tectonics.

1.3.2 The fossil record

Fossil records (Jessen *et al.* 1959) show the presence of both *D. cantabrica* and *E. mackaiana* in Ireland during the Gortian interglacial, approximately 320-400 thousand years BP. The Gortian interglacial preceded the Munsterian glacial period which was the glacial period before the most recent one, known as the Midlandian. A timeline summarising these most recent glacial cycles is given in Figure 8.

There are several sites in Ireland of different ages, which are linked to the Gortian interglacial, and there is evidence to suggest that what is collectively called the Gortian is in fact a period of several successive warm stages (Vaughan *et al.* 2004).

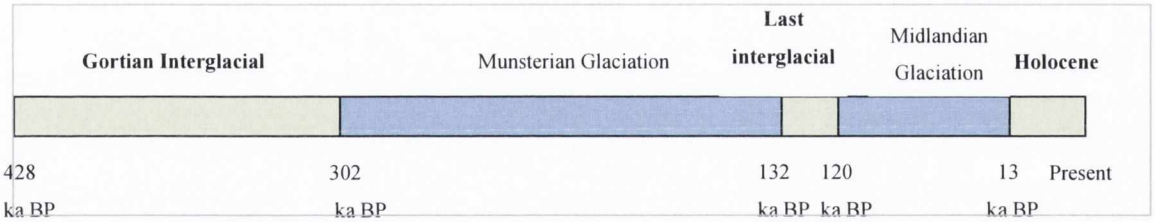


Figure 8: Timeline of the most recent interglacial and glacial periods in Ireland, times given in 1000 years before present (ka BP). Adapted from Coxon (1996).

The most reliable fossil evidence for the two species in question is based on seed fossils. Seed fossils of *D. cantabrica* were first identified in the type site at Gort (Jessen 1949), and later, along with *E. mackaiana* seeds in sites at Kilbeg, Co. Waterford (Watts and Ross 1959) and Baggotstown, Co. Limerick (Watts *et al.* 1963). *Daboecia* pollen was also found in cores from Bidart in southwest France corresponding to the Gortian interglacial (Oldfield 1968). Macrofossils of *E. mackaiana* were also found in abundance at this site (Huckerby and Oldfield 1976), which is several hundred kilometres away from current *E. mackaiana* populations in northern Spain.

The only evidence of either species having once been present further north than their current range comes from seed fossils from a peat deposit in Fugla Ness, off the north east coast of Scotland, which was correlated also to the Gortian interglacial (Birks and Ransom 1969).

Some pollen fossils from other sites throughout Ireland, Britain and France have been ascribed to *D. cantabrica* or *E. mackaiana* but for the most part they are identified as being of Ericales or Ericaceae-type without being identified to species or even genus level. This is also the case for the recently described site at Knocknacram, Co. Monaghan which could be the first example of a last interglacial site in Ireland (Vaughan *et al.* 2004).

1.3.3 Postglacial Recolonisation of Europe – an overview

During the Quaternary the ranges of temperate species expanded and contracted in response to climate change (Hewitt 1996). A pattern of climate oscillations over the last 700,000 years is characterised by cold and dry glacial periods, interspersed with warmer, wetter periods known as the interglacials (Comes and Kadereit 1998).

In Europe species ranges contracted southward and populations persisted during these cold periods in one or more of the three main refugial areas (see **Error! Reference source not ound.**); the peninsulas of Iberia, Italy and the Balkans (Taberlet *et al.* 1998). During the

warmer periods (including the current one) a process of range expansion occurred from one or multiple refugia as new areas became suitable (Waltari *et al.* 2007). Range expansion of different species occurred at different times, along different pathways and at different rates (Comes and Kadereit 1998).

The most recent glacial period (known as the Weichselian glaciation) lasted from approximately 115,000 years ago until about 10, 000 years ago (Cox and Moore 2000). The period of this glaciation which was the coldest and at which the ice sheets has reached their maximal southern extent is known as the Last Glacial maximum (LGM).

Sea levels also were affected by these climate oscillations. The volume of the ice sheet that formed in glacial periods effectively lowered sea levels. At the LGM sea levels are estimated to have been 120m lower than the current day (Peltier 2004), which would have resulted in a European coastline dramatically different to that of today, as shown in Figure 10. Ireland and Britain were not islands, but were part of the main continental landmass of Europe.

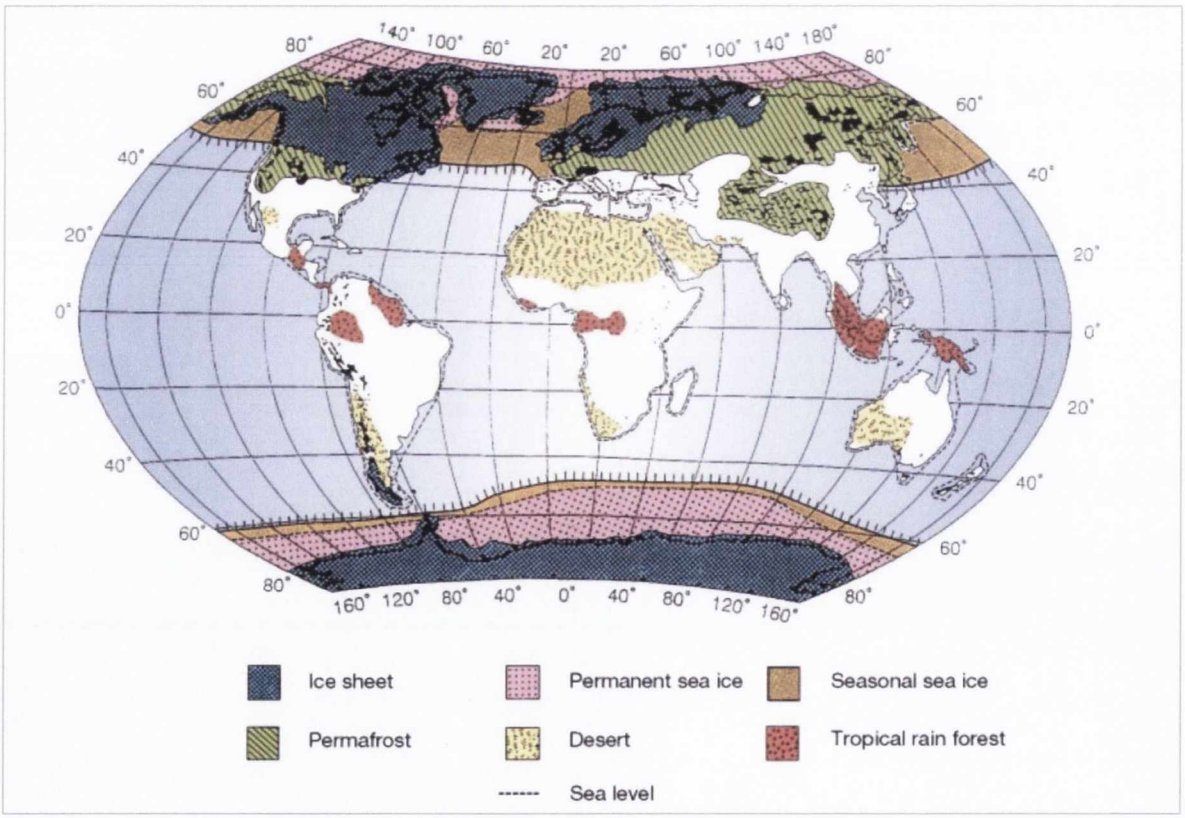


Figure 9: The maximum extent of ice and permafrost at the end of the last ice age 20,000 yr BP. The lowered sea level, large deserts and main blocks of tropical forest are indicated. From (Hewitt 2000).

The extent of the ice sheet has been revised over the years, the latest being that of Clark (2012) which shows the island of Ireland as entirely covered in ice at the (Figure 10). Earlier estimates showed parts of the west and south of Ireland remaining unglaciated throughout the LGM (see Chiverrell and Thomas 2010 for review).

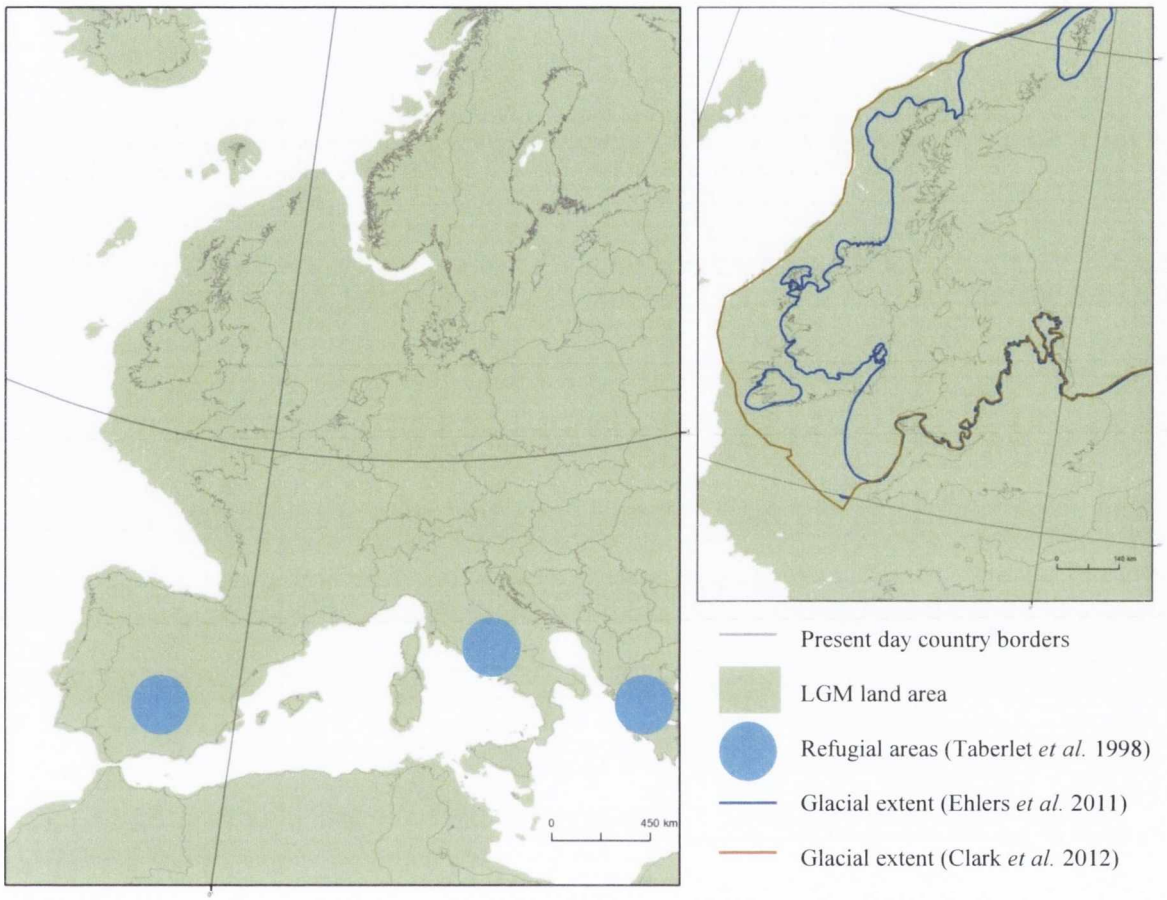


Figure 10: The land area of Europe at the LGM, location of refugia and the extent of the glacial limit according to earlier estimates (Ehlers *et al.* 2011) and more recent studies (Clark *et al.* 2012). The LGM land area is based on the ICE-5G reconstruction (Peltier 2004).

As well as the change in sea levels the cold, dry conditions at the LGM favoured extensive tundra and steppe vegetation across Eurasia, with northern hemisphere forests being displaced southward (Prentice *et al.* 2000). Analyses of pollen and macrofossil records have led to the reconstruction of vegetation biomes for different time periods in the Holocene. As shown in Figure 11, western Europe at the LGM is dominated by steppe, and there is no area of temperate deciduous forest to the south; raising the question of where the constituent taxa could have persisted at this time. It is possible that refugia for the temperate deciduous forest biome may have existed offshore at LGM, or that their characteristic taxa also persisted as components of other biomes (Prentice *et al.* 2000).

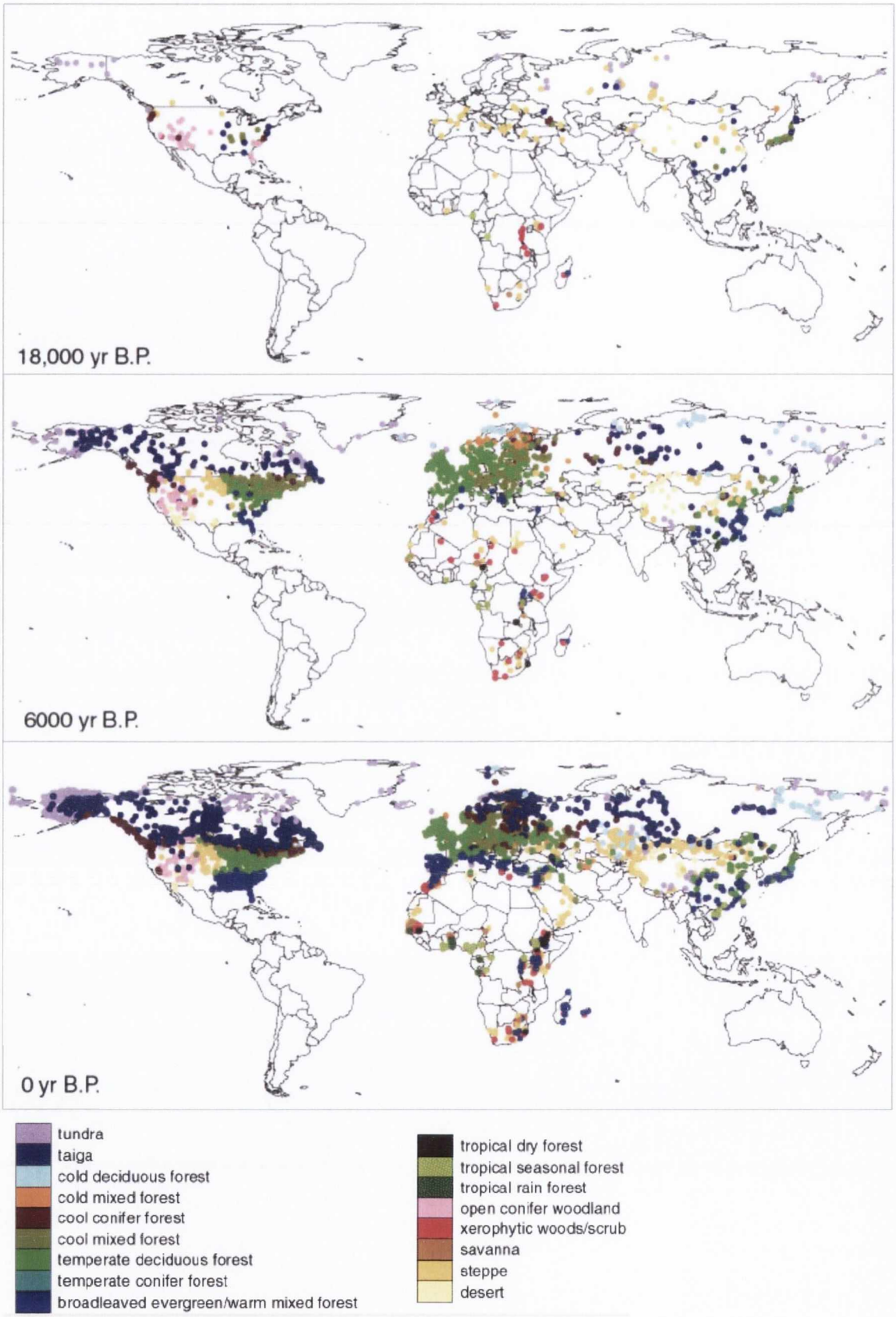


Figure 11: Biomes reconstructed from palaeoecological data for the LGM (18,000 yr BP) and mid-Holocene (6000 yr BP) and present day biomes reconstructed from surface pollen (0 yr BP). From Prentice *et al* (2000) .

1.3.4 Hypotheses for the current disjunct distribution of the study species

Forbes (1846), although he used neither the term "Hiberno-Iberian' nor 'Lusitanian', described the suite of species which gave the west and southwest of Ireland its "*botanical peculiarity*", and which were either peculiar to or abundant in the Iberian peninsula. Forbes and his contemporaries Darwin, Hennessy, Lyell and Hooker corresponded with each other on the various hypotheses that could explain the current distribution of these Irish species (DeArce 2008).

Their speculations and those that have been developed more recently can be summarised as four principal hypotheses, which have been discussed at length over the last 100 years (Reid 1913, Praeger 1939, Webb 1983, Waldren *et al.* 2006) and can be summarised as shown in Figure 12.

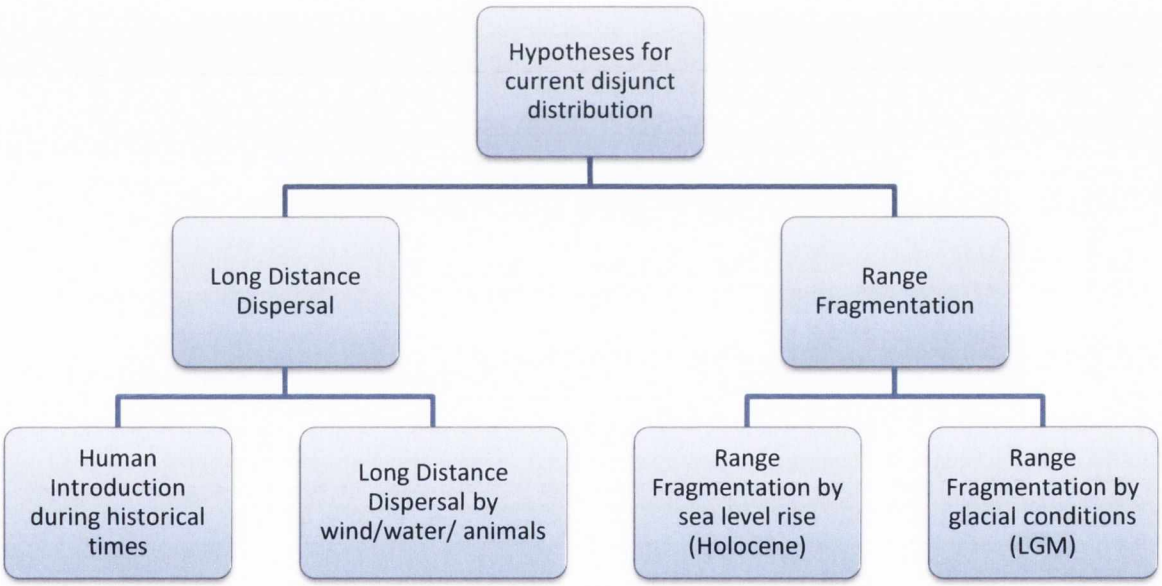


Figure 12: Possible scenarios to explain the current disjunct distribution of *D. cantabrica* and *E. mackaiana.*

1.3.4.1 Long distance dispersal scenarios

Two of the possible scenarios involve Long Distance Dispersal, one through the vectors of wind, water or animal migration, whilst the other invokes an anthropogenic introduction by humans in historical times.

Forbes (1846) discounted ocean dispersal because of the lack of an appropriate northward-flowing ocean current to explain the present-day distribution. Tests on seed buoyancy (Praeger 1913) demonstrated that it would be highly unlikely for water dispersal across such a large expanse of ocean to be possible; of the nine Hiberno-Iberian species he tested, all sank in less than an hour (Praeger 1913, 1932).

Wind dispersal for seeds not specifically designed for such also gave little support to the theory of dispersal by gales (Praeger 1939).

Reid (1913) favoured the idea that it was the chance introduction of seeds by birds or animals that provided the best explanation for the current distribution range, suggesting that birds blown off course in strong winds were the most likely carriers. However, few of the bird species that migrate from Spain and Portugal to Ireland are seed eaters, and the nearest landing places would be southern Ireland or southern Britain, where neither *D. cantabrica* nor *E. mackaiana* is found, making seed dispersal by birds an unrealistic scenario (Foss and Doyle 1990).

The hypothesis of a human introduction, whether deliberate or accidental, is based on the known historical trade links between Ireland and Iberia during the Middle Ages (Hennessey 1867, Corbet 1961). The earliest archaeological records of humans in Ireland are dated to the Mesolithic around 9000 years BP (Woodman 1985). References of human migrations from Spain to Ireland in ancient times can be found in the folklore of both populations (Madden 1863). However, the main historically verifiable contact occurred during the Middle Ages. One theory is that heather was used as packing between the wine casks that were imported into Ireland from Spain during this time (Nelson 1989).

Evidence to support a hypothesis of human introduction would have potential ramifications on the native status of these species in Ireland, as has already been shown to be the case for *E. ciliaris* (Curtis 2000).

Regardless of which vector the long distance dispersal occurred by the questions raised are the same. Are the Irish populations the result of one dispersal event or several? If long distance dispersal was possible to Ireland why did it not occur in Britain or northern France as well?

1.3.4.2 Range Fragmentation scenarios

Two additional hypotheses are based on the idea that the present day range of the study species is a decreased area which was at one time more continuous, and that the fragmentation occurred either due to the climatic changes of the LGM, or as a result of rising sea levels since the LGM.

In the first instance (range fragmentation in the LGM), it is hypothesised that what was a continuous range became fragmented during the LGM and that the disjunct northern population was able to survive the LGM in a refuge in or near current-day Ireland, establishing itself in the current areas of its Irish range as the ice retreated (Forbes 1846). These refugia may have been *in-situ* (within the boundaries of the current Irish coastline), or proximal (in areas now currently submerged).

Recent paleogeographic techniques gave indications that substantial areas off the west, south and east of Ireland were once above sea level (Edwards and Brooks 2008) and so could possibly have held putative proximal refugia. However whether or not this land area was covered by ice at the maximum glacial extent has been debated for many years. Until recently the traditional glacial extent for Ireland had parts of the West and South unglaciated at the LGM (Figure 10). However, more recently, the revised limit is much further out to sea, and completely covers the entire island of Ireland (Clark *et al.* 2012).

Phylogeographic studies on a variety of species including the natterjack toad *Bufo calamita* (Rowe *et al.* 2006), the brown trout (McKeown *et al.* 2010), Scot's Pine (Sinclair *et al.* 1998) and the brown seaweed *Fucus serratus* (Hoarau *et al.* 2007) have all supported the possibility of a western or south-western Irish refugium.

In the second instance (range fragmentation in the Holocene), it is postulated that during the LGM the distribution was restricted to southern Europe and then as it warmed, spread northwards along the coast, reaching Ireland across a land bridge before rising sea levels during the Holocene cut it off from the continent (Webb 1983). The position of Ireland and Britain on the continental shelf of Europe, and the shallower waters that now cover the hypothesised land bridge between Ireland and continental Europe, is shown in Figure 13.

Figure 13: This bathymetric chart shows the outline of the continental shelf as it extends outwards from Spain and France to encompass Ireland and Britain. Image reproduced from the General Bathymetric Chart of the Oceans (GEBCO) world map (www.gebco.net).

Figure 14 shows how over the next 10,000 years the ice sheet retreated northwards, sea levels rose and Ireland became completely separated from Britain and continental Europe at some point between 16, 000 years and 14, 000 years ago (Edwards and Brooks 2008). It is hypothesised (Praeger 1939) that at some point during this period of retreating ice, some species of both plants and animals were able to migrate fast enough from their Iberian refuge before the land bridge was cut off by the rising sea levels.

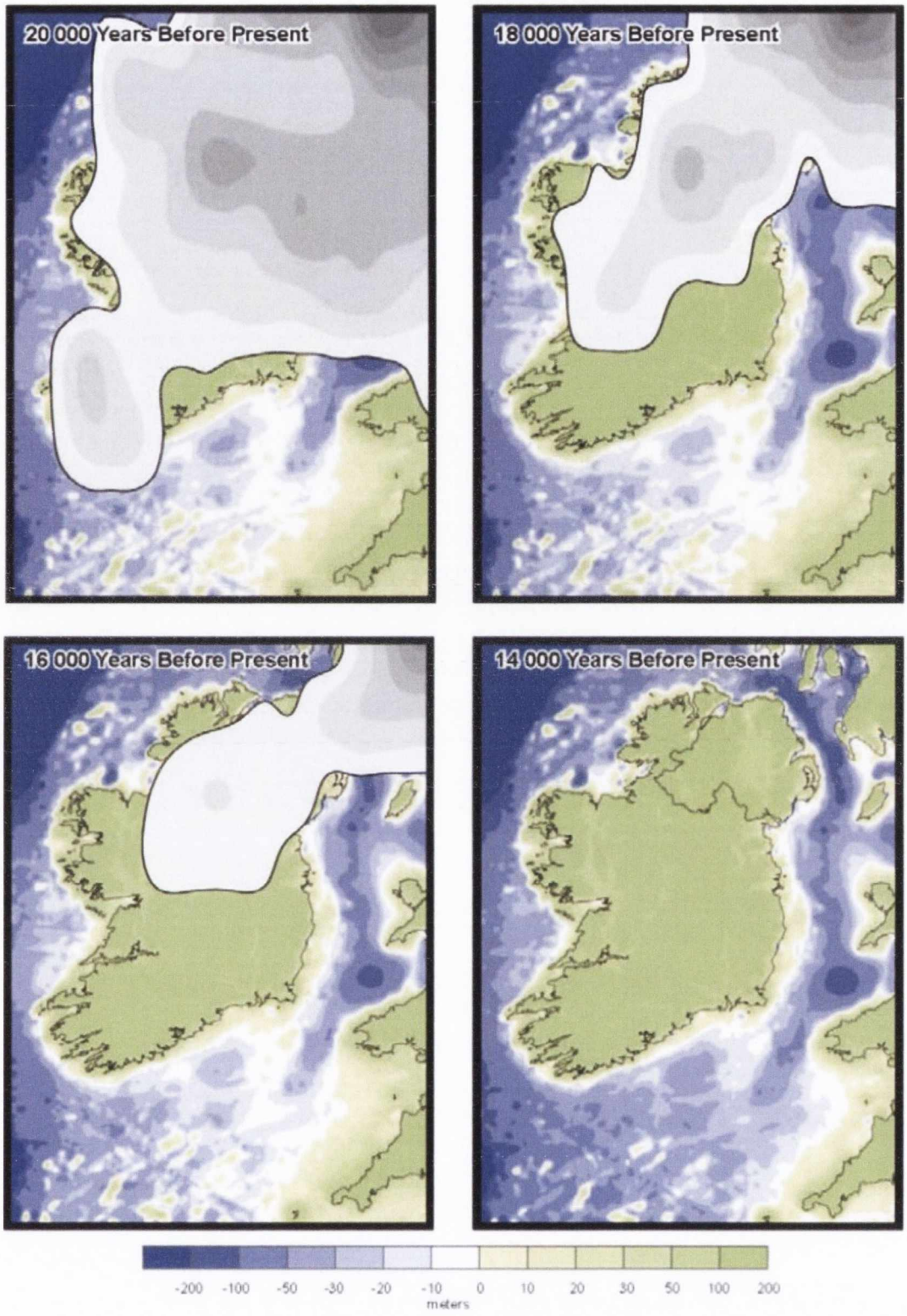


Figure 14: These paleogeographic reconstruction models show the retreat of the ice sheet and the rising sea levels that eventually isolate Ireland completely from the continent. Reproduced from (Edwards and Brooks 2008)

When considering the two range-fragmentation hypotheses the question remains – what is maintaining the disjunction? Why has there been no range-filling in the interim?

Looking at the four hypothesis overall it must be noted that one hypothesis is not necessarily mutually exclusive of another, and it is possible that the current Irish populations may represent the end product of a combination of more than one of the above scenarios, for example persistence in a glacial refuge, followed by further additions from human introduction. This is especially relevant for *E. mackaiana* with its pattern of isolated disjunct populations in Ireland which may have each had its own distinct distributional history.

It is also important to bear in mind that although both the study species share a similar range pattern as encompassed under the historical grouping of the Hiberno-Iberian flora, there are key differences between them, as outlined in Table 3. Even if the range disjunctions of two species coincide it does not mean they have had identical distribution histories (Thorne 1972), and in this case the possibility of different distributional histories is perhaps reinforced by these key differences in their current range.

Table 3: Key differences in range characteristics between *D. cantabrica* and *E. mackaiana*.

	D. cantabrica	*E. mackaiana*
Continental range	Northern Portugal, across northern coast of Spain and into France	North coast of Spain, not in Portugal or France
Isolated continental populations?	Yes (one in central Spain and two in western France)	No
Irish range	One large continuous population in west Galway and Mayo	Small disjunct populations in Kerry, Galway, Mayo and Donegal
Naturalised sites in Britain?	Yes	No

The compilation and analysis of evidence needed to either corroborate or refute any of the above hypotheses is the primary aim of this study.

1.4 Approaches and techniques selected for this study

A two-pronged approach was devised to elucidate the historical processes and events that have led to the intriguing present day disjunct distribution of *D. cantabrica* and *E. mackaiana*.

Firstly, the biogeographical aspect of the question was addressed using species distribution modelling, to investigate how the environmental and spatial components of these species'

distributions are linked, and to determine what processes or events leading to these distributions are likely or even possible given past and current environmental conditions.

Secondly, the analysis of molecular markers was carried out to reveal the current genetic relationships between individuals and populations of both species across their Irish and continental ranges, and in doing so uncover patterns or relationships which could explain the origins and history of the Irish populations.

1.4.1 Species Distribution Modelling

A species distribution model relates geographical occurrence data at known localities with the environmental and spatial information at those localities (Elith and Leathwick 2009). These types of models are also known as climate envelope models or ecological niche models, as they are based on the idea of characterising the niche of a species with the environmental data available.

1.4.1.1 Niche theory

The term "niche", first employed by Grinnell (1917) to mean the entire area in which an organism could survive if free from predators and competitors, has been utilised in several different ways over the years.

Elton's (1927) concept of the niche was more function orientated. He conceived the niche as encompassing the role of the organism, in terms of how it related to its food and to its predators; a functional unit as opposed to Grinnell's geospatial unit. The Eltonian niche concept was built upon by Gausse (1934) who formulated the competitive exclusion principle that states that no two species in a community may possess the same niche.

Hutchinson (1957) was the first to outline the niche concept mathematically, calling it an 'n-dimensional hypervolume', with every dimension representing an ecological factor, and every point within the hypervolume representing a state of the environment within which the species can survive indefinitely.

This fundamental niche of Hutchinson defines the area which a species could *potentially* inhabit if all abiotic conditions were favourable, similar to Grinnel's habitat-orientated niche. However, these abiotic factors are not the only variables that affect the ability of an organism to survive in that area. The volume of the fundamental niche is further constricted by biotic factors (such as competition, herbivory/predation and any other interaction with other species), by physical accessibility and by the capacity of a

population to adapt to environmental change (Soberón and Peterson 2005). This reduced subset of the fundamental niche is called the realised niche.

1.4.1.2 Constructing the model

A huge amount of environmental and spatial data are freely available with which to construct species distribution models, and at very high resolution. The current WorldClim datasets contain data for 19 variables, at a resolution of approximately one square kilometre. A model of each species' ecological niche can be calculated and principal components analysis (PCA) used to determine which environmental factors have the most limiting effect on the species distribution. Models can then be used to predict the potential niche of a species, which could possibly extend to regions which it currently does not occupy.

By projecting the models onto past or future climatic conditions it is also possibly to identify the potential niche of a species as it may have appeared several thousand years ago, or how it could alter under future climate change scenarios. This makes it an ideal approach to use for reconstructing the processes or events leading to the establishment of disjunct populations, as in this study.

1.4.2 Comparative DNA sequencing

To study the relationships between groups of organisms at their most fundamental level necessitates the investigation of the genetic code, in effect, the nucleotide sequences of their genome. Identifying the similarities and differences in selected regions of the genome between individuals is known as comparative DNA sequencing.

Selected gene regions are amplified by PCR (Polymerase Chain Reaction) methods and then the nucleotide sequence is determined through methods such as automated dye-terminator sequencing. Non-coding regions of DNA, or spacer regions are the most rapidly evolving, more likely to have sequence variation, and thus more useful in studies at the infraspecific level.

Due to their different modes of inheritance, nuclear and chloroplast markers have different attributes when it comes to their phylogenetic utility. Because chloroplasts are generally inherited through the maternal line (Harris and Ingram 1991), the potential for gene flow in the chloroplast genome is less than that in the nuclear genome (because pollen tends to move further than seed and recombination does not occur), and the genetic variation is more geographically structured (Cavers *et al.* 2003).

However because of this uniparental inheritance the lineage patterns revealed by cpDNA sequences will not be representative of the entire population history (Hare 2001), so in recent years the use of a combination of both chloroplast and nuclear markers has been the approach of choice for many phylogeographic studies (Beheregaray 2008).

For this study one nuclear marker and two chloroplast markers, all of which are non-coding regions, have been selected for comparative DNA sequencing.

1.4.3 Amplified Fragment Length Polymorphism (AFLPs)

Because a limited number of DNA sequences may not show a high degree of variation at the infraspecific level, another type of molecular marker, with a higher level of variation, can contribute more information about the genetic diversity of and relationships between populations. For the purposes of this study AFLPs were selected as a marker capable of yielding high levels of informative variation.

AFLPs are a powerful DNA fingerprinting technique, based on the production and amplification of fragments of total genomic DNA digested by restriction enzymes (Vos *et al.* 1995). The technique allows for the assessment of variation across the entire genome, requires no *a priori* sequence information and is highly reproducible (Meudt and Clarke 2007). Most importantly they can be used to detect genetic variation between groups that have low sequence variation, as is frequently the case in infraspecific studies (Despres *et al.* 2003), and can also help to detect clones (Hodkinson *et al.* 2002).

Preliminary AFLP analysis (Kingston and Waldren 2005) has confirmed that the Irish populations of both species can be genetically distinguished from that of the continental material, but was unable to provide conclusive support to any of the four hypotheses of distributional history. This study expanded on this previous work by including a larger sample size from a wider distribution, and carrying out more detailed analyses.

1.5 Aims of this study

This research project aimed to elucidate the historical biogeography of the two study species, *D. cantabrica* and *E. mackaiana,* through three main lines of enquiry;

- What are the environmental factors that define the current distribution of these species on the continent and in Ireland? Can we extrapolate from this how past

climate may have influenced distribution. Will future climate change have any effect on range?

- What can the lineage of DNA sequences of individuals tell us about the geographic origins of the Irish populations? What can the amount of sequence divergence tell us about the timescale of their establishment in Ireland?
- How diverse genetically are each of the two species, how is this diversity partitioned between individuals, populations and regions and what can this tell us about the relationships between continental and Irish populations of each species.

This thesis consists of five chapters. Chapters 2, 3 and 4 individually address each of the three points above. Chapter 2 describes the use of climate and soil data to carry out species distribution modelling. Chapter 3 addresses the second point, using comparative DNA sequences. Chapter 4 addresses the third point using the DNA fingerprinting technique of AFLPs. A synthesis of the findings and the inferences that can be made regarding the historical biogeography of each species are presented in Chapter 5.

CHAPTER 2 Species Distribution Modelling

2.1 Introduction

In order to elucidate the events or processes that led to the establishment of *Daboecia cantabrica* and *Erica mackaiana* in the west of Ireland we must first acknowledge the inherent spatial aspect of the question. In order to answer the "how" we must first look at 'where'.

As Webb (1983) pointed out, the question is not so much *Why do they grow in Ireland?* but rather, *Why do they not grow in Britain?*. To take that idea further we could also ask *Why do they not grow along the west coast of France?* It is precisely the disjunction that makes the distribution of interest.

The determination of whether a species establishes in a particular area is a combination of;

(1) the biotic and abiotic requirements it has for survival and reproduction and,

(2) its ability to disperse naturally into that area.

These two factors must be taken into account when considering the different scenarios under which a disjunct population distribution might have occurred; either the intervening area became uninhabitable due to a change in the biotic and abiotic environment, or its dispersal ability allowed for an event of long distance dispersal to a new area with a suitable environment (Hernandez-Roldan *et al.* 2011).

In the context of the two study species in the absence as yet of any definitive proof one way or the other, both scenarios remain in conjecture; the range could once have been more continuous than it currently is and the disjunction arose due to changing environmental factors between Ireland and the north Iberian peninsula, or the Irish populations became established in what were suitable habitats as a result of one or more long distance dispersal events, which could have been either a natural or a human-mediated event.

Determining which of these scenarios is the more likely to have occurred begins with developing a picture of the environmental factors currently defining these species ranges.

2.1.1 Climatic and edaphic requirements of study species

2.1.1.1 Lowland heaths and the oceanic climate

D. cantabrica and *E. mackaiana*, are two of a suite of species that together form a complex known as lowland heath, which in western Europe is confined to areas with an oceanic climate (Gimingham 1972). The northern limits of these lowland heaths are defined by cold temperatures and their southern limits by summer drought (Loidi *et al.* 2010).

Since Neolithic times the deforestation that accompanied the rapid expansion of farming and livestock grazing led to the formation of many lowland heaths (Crawford 2000), to the extent that today most of the current European heathlands have a secondary man-made origin (Gimingham 1972). Despite this, basic natural conditions associated with this vegetation remain the same; high rainfall, strong oceanicity and acidic soils (Loidi *et al.* 2010).

The moderating effects of the Atlantic Ocean are what characterise the distribution of the Hiberno-Iberian flora in general, this group of species being also referred to as 'Atlantic' (Webb 1983) or 'hyper-oceanic' (Crawford 2000).

2.1.1.2 Daboecia cantabrica

In Ireland *D. cantabrica* prefers areas of high rainfall, growing in areas with an annual precipitation of over 1524mm (Woodell 1958). Praeger (1939) postulated that it was this humidity requirement that constrained *D. cantabrica* and the other Hiberno-Iberian heathers from expanding eastwards across the country, rather than temperature or soil factors.

Despite its seeming preference for the milder temperatures of the oceanic climate, at higher altitudes it can be found under snow for many months of the year (Webb 1983). The plants collected for this study at Mt Torneo in the Asturian mountains were still surrounded by patches of snow in mid-July, at approximately 2000m.

In Ireland *D. cantabrica* generally occurs on thin soils over quartzite or mica schist, and on the fringe areas of peat bog, with soil pH measured at between 5.6 and 7.0 (Woodell 1958).

On the continent, pH values as low as 4.4 were recorded in the Basses Pyrenees, in France, and as high as 7.4 in some areas where *D. cantabrica* was found on calcareous soils (Woodell 1958).

Described as preferring full sun (Le Bail 2008), *D. cantabrica* is also shade tolerant and can occur in woodland and forest fringes (Woodell 1958). On the continent it prefers north-facing slopes, yet in Ireland it is more abundant on south-facing slopes (Woodell 1958).

2.1.1.3 Erica mackaiana

In Spain *E. mackaiana* grows within 35km from the coast (Nelson and Fraga 1983) in habitats characterised by mild winters and mild summers, without large temperature fluctuations. Although limited to the colder western coast of Ireland, frost resistance tests showed it to be more resistant than its current range in Ireland would indicate, suggesting that it is not frost sensitivity that limits its range in Ireland, but some other factor (Bannister and Polwart 2001).

In Spain *E. mackaiana* grows on a wide range of soil types including those on peaty acidic moorlands, granite, gneiss, schist, shales and slate, most of which are of low fertility (van Doorslaer 1990a). In all Irish habitats where *E. mackaiana* is found the soil is peaty.

The soil pH in the Spanish locations of *E. mackaiana* ranges widely from 3.5 to a high of 7.0 (Nelson and Fraga 1983), this top end being generally higher than that of its Irish habitats, which range from 3.7 to 4.8 (Webb 1955).

In Spain *E. mackaiana* occurs in a wider range of habitats than in Ireland, being found in wet flushed areas, open heathland, open coastal areas, shaded streams, and the margins of oak woodlands as well as pine and eucalyptus plantations (van Doorslaer 1990a).

The information above on the climatic and edaphic requirements of the study species are based on an assortment of personal observations and empirical measurements made by many individuals over the years and reported in the literature. At best they can provide a snapshot of conditions at the time and location at which they were measured, which in most cases did not approach nearly the complete range of either species.

However, with the recent advances in Geographical Information Systems (GIS) and the availability of high resolution climate data and soil datasets it is now possible to get a much more comprehensive picture of the abiotic factors found across a species' range, and determine what set of conditions define the area within which the species is found; in other words the development of a species distribution model (SDM).

2.1.2 Species Distribution Modelling

A species distribution model links species location data (occurrence or abundance at known sites) with the environmental conditions at those sites (Elith and Leathwick 2009). There are many different terms in use to describe what is essentially the same thing; bioclimatic envelope models, ecological niche models, or habitat suitability models. The term species distribution model will be used throughout this work, as it is the terminology currently the most widely used (Elith *et al.* 2011).

2.1.2.1 Principle of Species Distribution Modelling

The two basic inputs into a species distribution model are occurrence data and environmental data. The occurrence data should be as comprehensive as possible; comprising locations from across the entire known range of the species being studied. The environmental variables most commonly used are climate, vegetation, land use and soil type.

By extracting the environmental data at each known occurrence location, as well as the data at other locations where the species does NOT occur, a model is constructed of the potential niche for that species, i.e. the full geographic range where it has the environmental conditions that make its survival viable.

This does not however take into account the impact of biotic interactions such as competition or predation, or any barriers to dispersal into that area. It also does not take into account the potential effect of any environmental factors that were not included in the model.

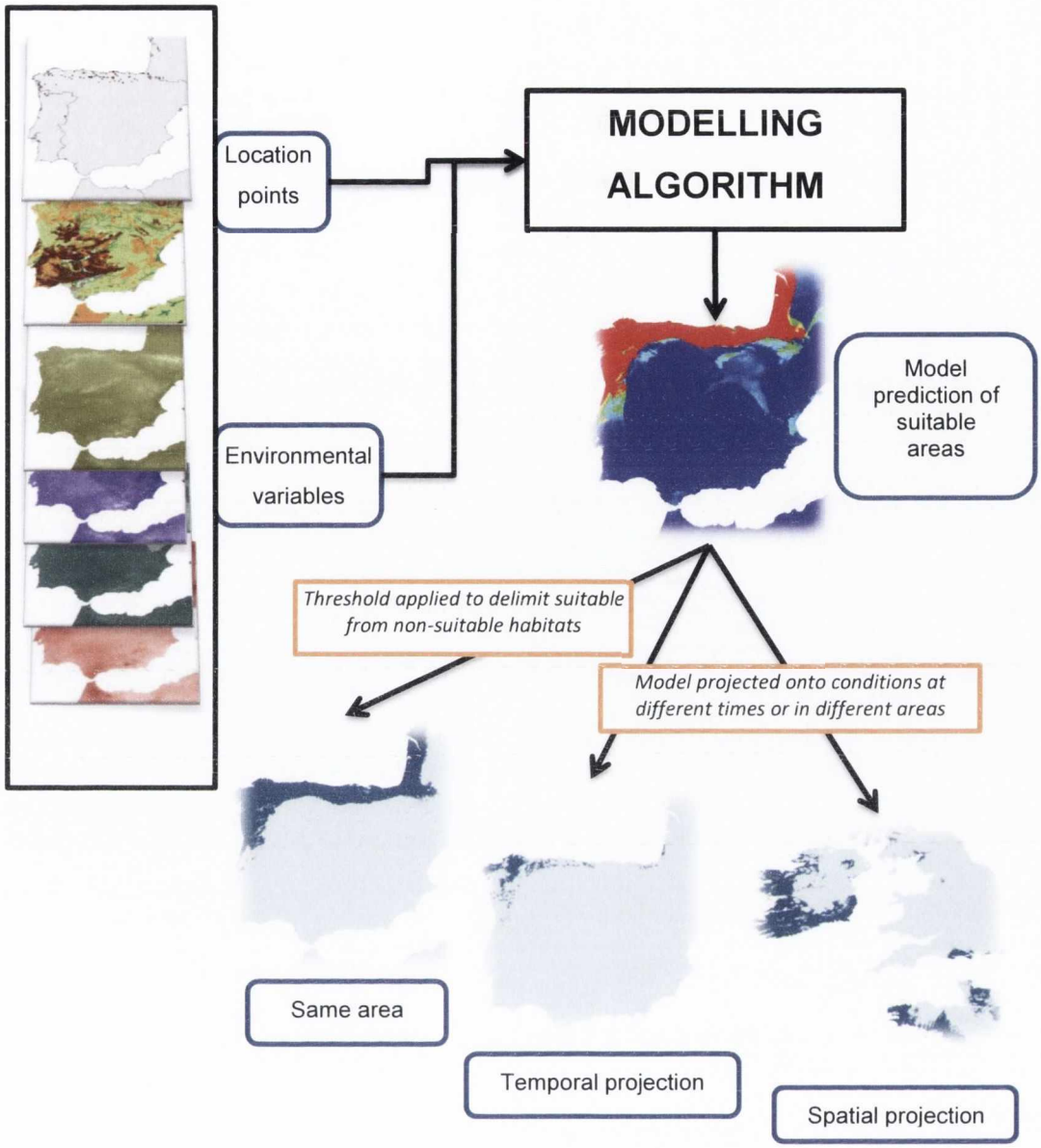


Figure 15: Diagrammatic representation of the construction of a species distribution model

2.1.2.2 Hind and Forecasting of SDMs

Many studies have applied a species distribution model to the prediction of a species distribution under climatic conditions in the past or future. For these hindcasting and forecasting applications it is essential to point out there is an assumption of niche stability. Niche stability (not to be confused with niche conservatism) refers to the similarity through time of the climatic conditions necessary for a species survival (Nogues-Bravo 2009).

2.1.2.3 Modelling techniques

Over the last 20 years a variety of modelling techniques and algorithms have been developed (Guisan and Thuiller 2005), the most extensively used being discriminant analysis (Manel *et al.* 1999), logistic regression (Brito *et al.* 1999), genetic algorithms (Stockwell and Peters 1999), artificial neural networks (Manel *et al.* 1999) and maximum entropy (Phillips and Dudik 2004). This study uses a maximum entropy technique implemented with a program called MaxEnt (Phillips *et al.* 2006), described more fully in the Methodology (Section 2.2.7).

2.1.2.4 Applications of Species Distribution Modelling

Species distribution modelling has become a widely used tool across many disciplines (Guisan and Thuiller 2005, Richards *et al.* 2007).

In the area of conservation and biodiversity species distribution models have been used to identify potential sites for the reintroduction of endangered species (Wilson *et al.* 2011, Adhikari *et al.* 2012) and the establishment of conservation reserves (Araujo *et al.* 2004). They have also been applied to the identification of areas susceptible to invasive species (Peterson and Vieglais 2001) or to disease (Machado-Machado 2012), and in the overall study of the effects of climate change on biodiversity (Pearson and Dawson 2003).

Species distribution modelling can also be employed to target field surveys in order to more rapidly discover as yet unknown populations of a rare species (Pearson *et al.* 2007, Williams *et al.* 2009). The technique has also been used to identify cryptic species (Raxworthy *et al.* 2007).

In the agricultural context species distribution models have been applied to the identification of potential areas for the cultivation of crops (Evans *et al.* 2010, Sanchez *et al.* 2010).

Species distribution modelling is not limited to terrestrial species, it has been successfully applied to marine and freshwater environments as well (Robinson *et al.* 2011) and has been implemented in studies focussing on coral (Guinan *et al.* 2009), fish (Reshetnikov and Ficetola 2011) and algae (Tyberghein *et al.* 2011).

One of the most widely used applications of species distribution models is their integration with the fields of phylogeography and historical biogeography, where testing for concordance between molecular data and the species distribution model can strengthen and

refine phylogeographic inferences (Waltari *et al.* 2007). Through setting the results of molecular phylogenies into a spatial context, species distribution models have been used in the identification of refugial areas (Flanders *et al.* 2011) and the study of the origins of disjunct populations (Hernandez-Roldan *et al.* 2011). It is in this context that species distribution modelling has been selected as a tool for the study of the distribution of *D. cantabrica* and *E. mackaiana.*

This will be the first time species distribution modelling has been applied to the question of the Hiberno-Iberian flora.

2.1.3 Aims

Construct a species distribution model for each species based on available climatic and edaphic data,

Use the model to identify the potential range of each species under current conditions,

Project the model onto past climatic conditions to identify the potential range of each species during the Last Glacial Maximum (LGM),

Investigate whether LGM glacial extents and sea-level reconstructions allow for the possibility of an Irish proximal or *in-situ* refugium,

Project the model onto future climate conditions to evaluate the effect climate change will have on the range of the two species.

2.2 Methodology

2.2.1 Occurrence data and georeferencing

A GPS reading was taken for the location of all specimens collected during fieldwork and all of these points (325 in total) were included in the dataset. Coordinates that were recorded on herbarium vouchers (TCD, D, M, K, BC) were checked for concordance with the locality description and also included in the dataset. Herbarium vouchers with precise locality names but no coordinates were georeferenced using Google Earth. Herbarium vouchers with only very vague locality names, e.g. 'Asturias', 'W. Galway', were not used.

Datasets from GBIF (Global Biodiversity Information Facility) were downloaded and assessed for their suitability, but as some very obvious mistakes in georeferencing were found they were not imported in bulk. Instead, a selection of records with precise locality descriptions were used, and the georeferencing done manually using Google Earth. The final occurrence dataset contained 514 presence points, as described in Table 1. As far as is known the points cover the entire distribution range of both species.

Table 4: Number of occurrence points used in niche modelling analysis

Species	Points from fieldwork	Georeferenced points from herbarium vouchers and GBIF database	Total
D. cantabrica	215	133	348
E. mackaiana	110	56	166
Total	325	189	514

2.2.2 Environmental data

The basis of the environmental data was a series of global temperature and precipitation datasets, known collectively as the WorldClim dataset (Hijmans *et al.* 2005). The WorldClim dataset is a comprehensive climate dataset with near-total global coverage. There is one dataset available which comprises current conditions and several others which comprise past and future climate data. The WorldClim dataset is used extensively in all climate-related modelling and GIS applications, including species distribution modelling. It is available for free download from www.worldclim.org.

The WorldClim dataset for current conditions has been compiled from records for the period 1950-2000 and has a spatial resolution of 30 arc seconds, which is approximately one square kilometre (Hijmans *et al.* 2005). The dataset is made available in the form of a raster layer for each of 19 different bioclimatic variables. The 19 bioclimatic variables used are constructed from monthly temperature and precipitation data and have been selected as those most directly related to the physiological aspects of plant growth. They represent annual trends, seasonality and extreme or limiting environmental factors.

Table 5: The 19 bioclimatic variables from the WorldClim database

Variable Name	Description (temperature values are given as °C, precipitation values are in mm).
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp)
BIO3	Isothermality (BIO2/BIO7)
BIO4	Temperature Seasonality (standard deviation of monthly means)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (coefficient of variation[1])
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

Bioclimatic variables derived from LGM climate data were also obtained from www.worldclim.org. The original LGM climate database from which they were derived is held by the Paleoclimate Modelling Intercomparison Project Phase II (http://pmip2.lsce.ipsl.fr/). The LGM layers are at a resolution of 2.5 arc minutes, and are based on the Community Climate System Model version 3 (Collins *et al.* 2006).

[1] ratio of monthly precipitation standard deviations to the annual mean.

Climate data for the Last Interglacial (Otto-Bliesner *et al.* 2006) were downloaded from www.worldclim.org at a resolution of 30 arc-seconds.

For the future projections the climate layers used had been generated with the CCCMA-CGCM31 Model (30arc seconds resolution) under two emission scenarios; SRES-A1b and SRES-B2a, for the time period of the 2080s. This dataset was downloaded directly from the General Circulation Model (GCM) data portal of the International Centre for Tropical Agriculture (CIAT) http://ccafs-climate.org/download_down.html.

The two emissions scenarios chosen, SRES-A1b and SRES-B2a, were selected from the A1 and B2 'families' of emission scenarios, respectively. The Intergovernmental Panel on Climate Change (IPCC) has developed 40 different emission scenarios, grouped under four families according to different projected economic growth, population growth and fossil fuel usage (Nakicenovic *et al.* 2000). The A1 family is the most extreme family of emissions scenarios, and B2 the most conservative (Beaumont *et al.* 2008).

To enable faster processing all climate layers were clipped to the area of Europe, including the Azorean archipelago (extent: 64°N, 31°S, -37°W, 14°E).

2.2.3 Soil Data

Soil layers were derived from the Harmonised World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC 2012). The HWSD is a 30 arc-second raster dataset comprising 21600 rows and 43200 columns, each cell being linked to harmonised soil property data which include the soil unit composition and the characterisation of various soil parameters (organic carbon, pH, soil depth, salinity, textural class and many others).

The dataset was downloaded as a raster file with an accompanying Microsoft Access database. Converting the data into the same format as the bioclimatic layers, as required for the modelling process, was a multi-step process.

Using ArcGIS 9.2 (ESRI) the raster file was converted to grid format. A database query selecting the variables of interest (Soil Unit Code and Topsoil pH) was exported and joined to the grid file using the 'Join' tool. To create a layer of just the soil unit categories, the Lookup function (using the Soil Unit Code as the lookup field) was then used to change the values of the raster cells to the soil categories. The resulting raster was then converted to integer format using the Raster Calculator, and then converted to an ASCII file. The same

process was used to create the layer of pH values, but using a Reclassify tool instead of the Integer conversion, to retain the decimalised pH values.

2.2.4 Selection of variables for analysis

In order to reduce the effect of correlation between the 19 bioclimatic variables in the model, variables that were highly correlated had to be excluded from the analysis. Using ArcMap 10, 000 points were randomly generated across Europe and the values from each underlying bioclimatic raster layer were extracted at each point. From this 19x10000-cell dataset a correlation matrix was generated using the statistical package R (http://cran.r-project.org/). Variables showing high correlations with other variables were identified and removed from the dataset. Only those variables that showed no correlations over r=0.8 were included in the model, as suggested by Tellería *et al.* (2012).

2.2.5 Description of the niche

The values of the environment variables in each raster cell that contained at least one occurrence point were extracted, using the Extraction tool in ArcGIS9.2, with the points shapefile as a mask. (The values at each occurrence point were not used because the duplication caused by within-cell multiple records would have caused an additional bias towards high-density sampling locations).

Using the statistical package R, a Principal Components Analysis was carried out on the correlation matrix of these environmental variables. The plot of the first two principal components was constructed to visualise niche differentiation between species and between Irish and Continental populations of each species. The loadings of each variable to the principal components were also calculated.

2.2.6 Construction of the SDM

The open source software package MaxEnt (Phillips and Dudik 2004) was used to carry out all the modelling predictions. Many other SDM algorithms/software packages exist, such as GARP and BIOCLIM, but MaxEnt was chosen based on its capabilities with presence-only data (rather than presence/absence data), and its performance compared to other packages (Elith *et al.* 2006). MaxEnt v3.3.3e is available at http://www.cs.princeton.edu/~schapire/maxent/.

2.2.7 Principle of MaxEnt

The MaxEnt algorithm operates on the principle of maximum entropy. The modelling process starts initially with a uniform distribution of probability (of species occurrence) over the entire area of interest. It then carries out an optimisation of the predicted distribution based on the 'background' environmental variables at each known presence point. The result is a spatial distribution that is most spread out (i.e. at maximum entropy), whilst still subject to the constraints of the known observations of the species.

To construct the models in this chapter, MaxEnt was run with the following default parameters; convergence threshold = 0.00001, default prevalence = 0.5, maximum iterations=500. The model values were generated in logistic format, which gives the species probability of presence in the range 0 (unsuitable) to 1 (optimal), and is the output format which has been shown to perform best (Phillips and Dudik 2008). For these analyses duplicate presence points within one grid cell were discarded to avoid a presence bias.

Suitable areas were defined by using a binomial threshold based on the sum of the maximum sensitivity and specificity (Liu *et al.* 2005). All of the prediction outputs were then imported into ArcMap version 9.2 for visualisation and formatting.

All output maps were projected using the Lambert Conformal Conic Projection, which is a standard projection for use across regions in the mid-latitudes (20°-60°), such as Europe.

2.2.8 Model validation

Each model was validated using the Receiver Operating Characteristic (ROC) curve. A ROC curve is obtained by plotting sensitivity (true presence rate) on the y-axis and 1-specificity (the false positive rate) on the x-axis.

The Area Under the ROC curve (AUC) provides a single measure of model performance by giving the probability that the model scores a random presence site higher than a random site from the study area (Phillips *et al.* 2009). A score of 0.5 indicates a model that cannot discriminate any better than random (Phillips *et al.* 2009) whilst models with AUC values above 0.9 are considered to be excellent (Swets 1988).

In the last decade the AUC has been accepted as the standard measure for assessing the accuracy of species distribution models although not without some controversy as to their usefulness (Lobo *et al.* 2008, Jiménez-Valverde 2012).

2.2.9 Temporal projection of the model

The SDM for each species was projected onto the same geographical area at three different time periods; the LGM (c. 21,000 years BP), the Last Interglacial (c.130,000-116,000 years BP) and 70 years into the future to the 2080s. For each of these projections the same binomial threshold of maximum sensitivity and specificity was used to define suitable areas of distribution.

Land surface area was the same at all these projections except for the LGM, when sea levels were approximately 120km higher than the present day. The CCSM model used to construct the climate layers for the LGM takes this into account and climate data are thus given for the land area as it was then.

2.3 Results

2.3.1 Current actual distribution

Table 6 lists all sampled populations of *D. cantabrica* and *E. mackaiana*, with a locality description, and an indication of the number of samples in that population that were used in the DNA sequencing and AFLP analyses described in Chapters 3 and 4. The maps in Figure 16 and Figure 17 show the locations of these populations.

A total of 26 populations of *D. cantabrica* were sampled, including two in the Azores (subsp. *azorica*) and four in Ireland. Eleven populations of *E. mackaiana* were sampled; six in Spain and five in Ireland.

Table 6: Locality description and coordinates of sampled populations of *Daboecia cantabrica* and *Erica mackaiana*, and use of samples in molecular work

COUNTRY	REGION	Pop Code	Locality	Longitude	Latitude	Description	Total samples	No. of samples used in:		
								cpDNA	ITS	AFLPs
Daboecia cantabrica **(including subsp. *azorica*)**										
France	France - Loire	Q	St-Cyr-des-Gats	-0.8628	46.5644	Side of sunflower field, under trees	30	4	29	16
		R	Brebaudet	-0.8330	46.5595	Near forest, side of road	10	4	11	8
	Southern France	S	La Rhune	-1.6277	43.3247	Side of walking track, open hillside	10	1	9	7
		T	Col d'Ibardin	-1.7069	43.3234	Edge of lake on rocky slope	10	2	10	8
		U	Pas de Roland	-1.4091	43.3074	Steep hillside along river, woodland	10	2	10	6
		V	Bidarray	-1.3636	43.2628	Rocky hillside, near road	30	1	27	19
Spain	NE Spain	A	Urkiola National Park	-2.6545	43.1004	Edge of hillside pine forest	10	2	9	5
		B	Cabo Major	-3.7951	43.4903	Cliffs, mixed with gorse and bracken	10	2	10	5
		C	Val de Aso	-3.6004	43.2272	Roadside woodland, oak	10	3	10	5
		D	Val de Aso	-3.5836	43.2063	Open hillside	10	1	8	5
		E	Cabo Penas	-5.8529	43.6558	Cliffs, open hillside	10	1	10	5
		F	San Isidro Pass	-5.4366	43.0776	Roadside	10	2	9	5
		G	Mt Torneo	-5.3909	43.0481	North side of mountain	10	3	9	4
	NW Spain/ Portugal	H	Serra de Capelada	-8.0032	43.6912	Clearing in conifer plantation	10	-	8	4
		I	Sierra de Meira	-7.2624	43.2131	Beside stream, amongst rocks	10	1	10	-
		J	Cabo Fisterra	-9.2676	42.8934	Coast, on rocky hillside	10	-	10	5
Portugal		K	Vigo	-8.7414	42.0570	Roadside edge of conifer forest	10	1	7	5
		L	Peneda Geres	-8.2873	41.7021	Roadside	10	1	9	5
		M	Peneda Geres	-8.2073	41.7367	Roadside, under Quercus	10	1	-	5
		N	Neiva Catelo	-8.7877	41.6191	Clearing in eucalypt forest	10	2	2	5
	Azores	O	Faial	-28.707	38.5804	Caldera rim, wet	10	3	10	10
		P	Pico	-28.426	38.4709	Slope of Mt Pico	12	3	12	12
Ireland	Ireland	W	Carraroe	-9.5966	53.263	Rocky outcrop	10	1	10	8
		X	Derryneen	-9.6666	53.4576	Roadside	10	1	10	5
		Y	Clifden	-10.093	53.4978	Roadside, hill	30	4	30	20
		Z	Errisbeg	-9.9799	53.3915	Open slopes	10	2	10	5
						Totals:	322 (26 pops)	48 (24 pops)	289 (25 pops)	187 (25 pops)
Erica mackaiana										
Spain	Spain	A	Cabo Penas	-5.852883	43.65575	Cliffs, open hillside	10	3	-	6
		B	Sierra de Loba	-7.945083	43.282583	Wet heathland	10	2	9	7
		C	As Pontes de GR	-7.93705	43.48125	Field edge	10	1	7	5
		D	Sierra da Capelada	-8.0032	43.6912	Clearing in conifer forest	10	1	2	5
		E	Sierra de Meira	-7.262433	43.213083	Beside stream, amongst rocks	10	2	2	7
		F	Foz	-7.2287	43.520883	Roadside in conifer forest	10	1	2	6
Ireland	Ireland	G	Roundstone 1	-9.9172	53.438233	Blanket bog	10	2	2	5
		H	Roundstone 2	-9.924033	53.441583	Blanket bog	30	4	5	17
		I	Lough Nacung	-8.152483	55.0179	Bog	26	4	4	16
		J	Knockeneden Cross	-10.07630	51.93567	Field	8	5	5	5
		K	Bellacorick, Mayo	-9.521100	54.100700	Bog, field	13	2	2	2
						Totals:	147 (11pops)	27 (11pops)	40 (10 pops)	81 (11 pops)

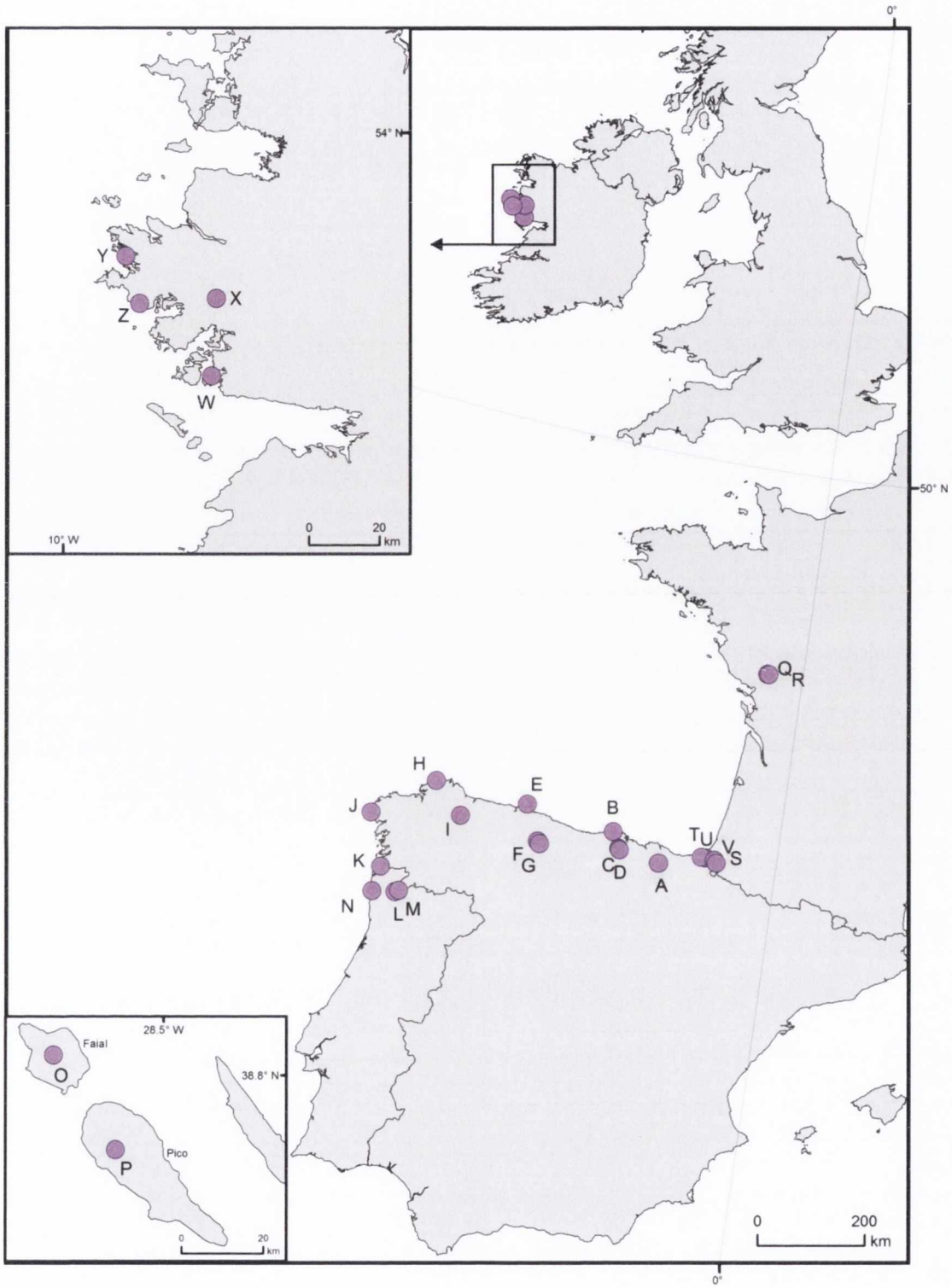


Figure 16: The 26 sampled populations (A-Z) of *D. cantabrica* in continental Europe, the Azores (inset, bottom left) and Ireland (inset, top left).

The 26 sampled populations of *D. cantabrica* were located in France, Spain, Portugal and Ireland. Two populations were sampled in the Azores, one on Faial and one on Pico. Six populations of *E. mackaiana* were sampled in Spain, and five in Ireland, their locations shown in Figure 16.

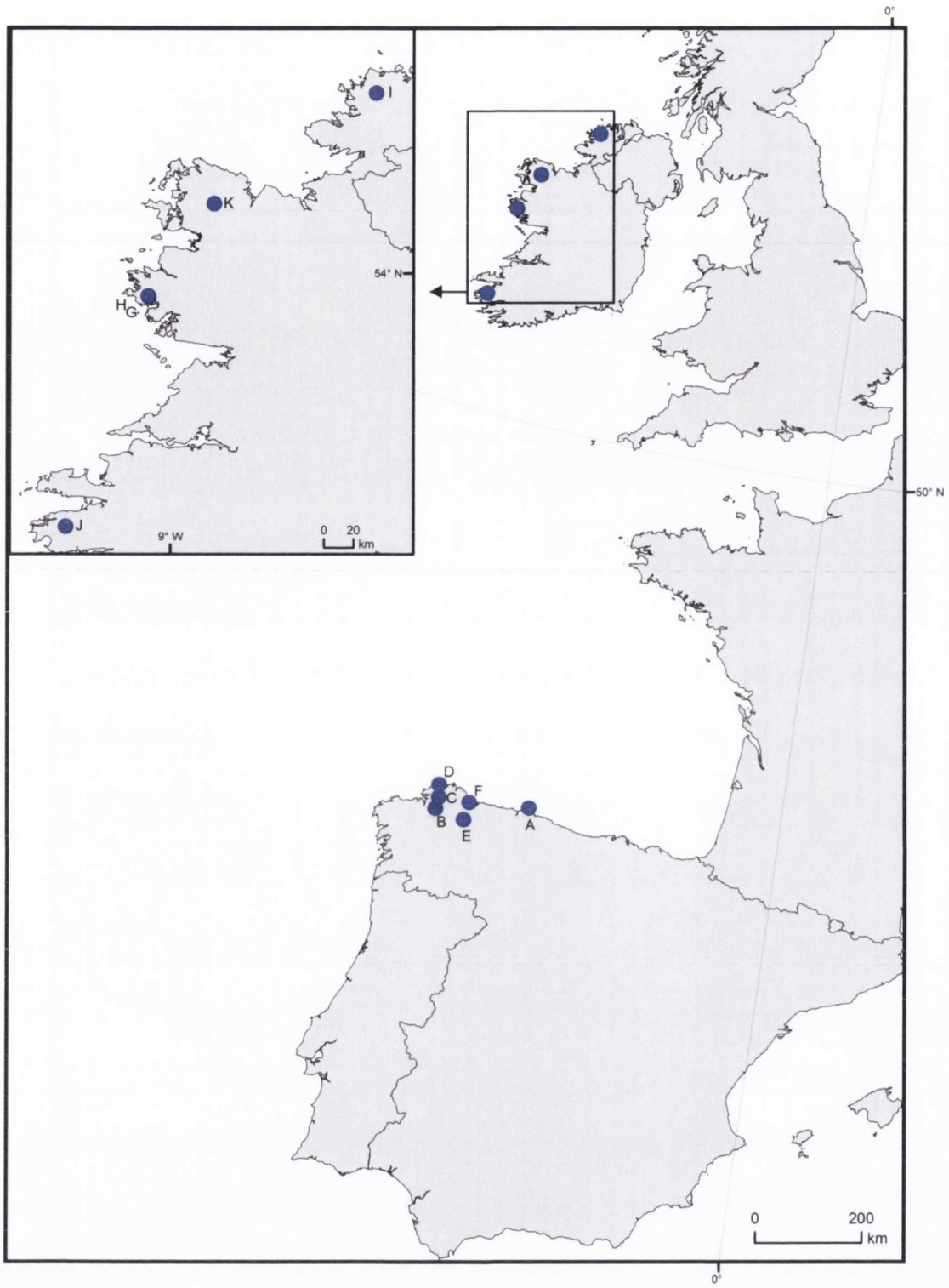


Figure 17: The 11 sampled populations of *E. mackaiana* in Spain and in western Ireland (inset).

Table 7 summarises the georeferenced occurrence points gleaned from herbarium voucher labels and from the GBIF portal; these 189 records together with 325 individual sample points from the 37 the sampled populations above together make up the 541 location points used in the species distribution modelling analysis of this chapter.

Table 7: Occurrence points of *D. cantabrica* and *E. mackaiana* from herbarium voucher specimens, either viewed in person or from GBIF portal.

	Herbarium[2]	Accession Number	Country	Locality	Latitude	Longitude
Daboecia cantabrica						
1	TCD	s.n.	Spain	Cordillera Cantabrica, c. 70km NNW of Leon on road from San Emiliano to Puerto de Ventana.	43.014929	-6.001160
2	TCD	s.n.	Spain	South side of Puerto de Piedrasluengas (Burgos)	42.921408	-4.512291
3	TCD	s.n.	France	Pas de Roland, Laxia nr Itxassou, abundant on rocks above R. Nive (Basses-Pyrenees)	43.316468	-1.403594
4	TCD	s.n.	Ireland	W. side of Ballycuirks Lough, W.Galway	53.327596	-9.166004
5	TCD	s.n.	Ireland	Near Westport, Mayo, foothills of Croagh Patrick	53.758042	-9.601574
6	TCD	s.n.	Ireland	Gentian hill, Galway (west side of hill)	53.259116	-9.110303
7	TCD	s.n.	Ireland	Oldhead, nr Louisburgh, W. Co. Mayo	53.772172	-9.776065
8	TCD	s.n.	Ireland	Grassy roadside bank 2 miles N of Oughterard	53.458226	-9.335329
9	TCD	s.n.	Ireland	Errisbeg, base of mountain on S side of big bend in road to the W of the mountain	53.391111	9.978056
10	TCD	s.n.	Ireland	Roadside, Kylemore Lough N. side	53.560833	9.871111
11	DBN	73:1981	Ireland	Cloonisle, W of Cashel, W Galway	53.434006	-9.828993
12	DBN	20:1971	Ireland	W of Lough Troskan, E of Carna, W Galway	53.329339	-9.825618
13	DBN	22:1967	Ireland	near the Cashel Hotel, Connemara	53.419371	-9.799653
14	DBN	70:1989	Ireland	Kylemore, Connemara	53.47035	-9.729482
15	DBN	220	Ireland	Leenaun, Connemara	53.59473	-9.692777
16	DBN	86:1980	Ireland	Ballinaboy, Clifden	53.460877	-10.024163
17	DBN	39:1983	Ireland	W end of Dawross peninsula, W of Letterfrack	53.562387	-10.079991
18	DBN	s.n.	Ireland	Rinvyle, Connemara	53.564994	-9.88525
19	DBN	s.n.	Ireland	Salruck Pass, Connemara	53.592695	-9.845553
20	DBN	33:1985	Ireland	Streamstown, N of Clifden, W Galway	53.526295	-10.046854
21	DBN	86:1980	Ireland	E of Carna, Galway	53.324909	-9.835406
22	DBN	1:1966	Ireland	Near Tonabrocky, 4 miles WNW of Galway	53.290303	-9.116582
23	DBN	36:1965	Ireland	by Owenwee River, W of Westport, Mayo	53.777029	-9.568172
24	DBN	36:1965	Ireland	Near north shore of Lough Mask at Srah, Mayo	53.695826	-9.361891
25	DBN	185	Ireland	West of Partry, E Mayo	53.697376	-9.288591
26	DBN	379A	Ireland	Oranhill, Galway	53.255654	-8.935826
27	DBN	s.n.	Ireland	Between Cong and Maam	53.515234	-9.42844
28	DBN	s.n.	Ireland	Killary Harbour, Co. Mayo	53.613776	-9.807006
29	DBN	115:1969	Ireland	Amongst heather in Tournakeady forest, Mayo	53.653758	-9.374785
30	DBN	27:1974	Ireland	S. side of Errislannan peninsula, S of Clifden	53.46966	-10.075809
31	DBN	18:1971	Ireland	Glen, W of Fin Lough, W Mayo	53.621503	-9.755807
32	DBN	13:1975	Ireland	Lough Truskan, hill W of Carna, Galway	53.311429	-9.790154
33	DBN	329	Ireland	Heathery bluff, W of Partry, E Mayo	53.697037	-9.294541
34	DBN	s.n.	Ireland	Edge of a bog near Carrowniskey Bridge, Co Mayo	53.696705	-9.847911
35	DBN	s.n.	Ireland	Dry bank, NW arm of Lough Corrib	53.504988	-9.527441
36	DBN	s.n.	Ireland	Rock outcrop by lake on S side of Errislannan peninsula, W Galway	53.470056	-10.065537
37	DBN	16:1993	Portugal	(2 sheets) Azores: Pico, east of main peak west of Lagoa do Caiado	38.456236	-28.267671
38	DBN	12:1994	Spain	Picos de Europa, Asturias, track to Narango de Balines at Coll above car park on way to Refugio	43.177837	-4.901944
39	DBN	41:1989	Spain	Province of Orviedo, Puerto de Ventana on road between San Emiliano and Trubia	43.058432	-6.013071
40	DBN	41:1989	Spain	Province of Orviedo,Lago Enol, near Covadonga	43.266524	-4.985278
41	DBN	41:1989	Spain	Province of Santander, Valdeprado, north of Puerto de Piedras Luengas	43.074055	-4.492545
42	DBN	41:1989	Spain	Province of Orviedo, 1 km from main Llanes-Comillas road, on road to La Borbolla, southwest of	43.382707	-4.631253
43	DBN	41:1989	Spain	Province of La Coruna east of Grana, Monte del Estado	43.192361	-8.14931
44	DBN	41:1989	Spain	Province of La Coruna, Monte Capelada, south of Cao Ortegal	43.689417	-7.935101
45	MA	168160	Spain	Villardiaz; Fonsagrada (Lugo) (Galicia)	43.120632	-7.065247
46	MA	89558	Spain	Marin (Pontevedra) (Galicia)	42.387814	-8.704062
47	MA	180956	Spain	Cerdedo (Pontevedra) (Galicia)	42.532386	-8.395268
48	MA	185303	Spain	Louorizan (Pontevedra) (Galicia)	42.402208	-8.670334
49	MA	523809	Spain	Desembocadura del R. Cabe en el R. Sil, Ferreira de Panton, Lugo (Galicia)	42.513629	-7.640207
50	MA	523808	Spain	Cerceda, La coruna (Galicia)	43.184655	-8.484906
51	MA	517196	Spain	Monfero, por la pista de Ventureira a la presa del Eume. En bordes de brezal-tojal (Galicia)	43.321971	-8.041433
52	MA	426102	Spain	Asturias, Puerto de Tarna (Asturias)	43.084851	-5.218545
53	MA	546648	Spain	Asturias, Siero, entre San Miguel del Marcenado y el cuartel Cabo Noval (Asturias)	43.38179	-5.647622
54	MA	488317	Spain	Cangas de Narcea: entre Carballo y Casas Reguera del Cabo, puente de Valsagra (Asturias)	43.184147	-6.541053
55	MA	488314	Spain	Degana: camino de Rebollar a Peranzanes (Asturias)	42.942264	-6.575827
56	MA	594099	Spain	Cudillero, Ballota (Santa Marina) (Asturias)	43.552521	-6.306031

[2] TCD=Trinity College Dublin, DBN=Dublin, MA =Real Jardin Botanico, SALA = Universidad de Salamanca, K=Kew, E= Edinburgh, BCN = Barcelona, LY=Lyon, GDA=Universidad of Grenada, ARAN=Alto de Zoroaga, SANT=Universidad de Santiago de Compostela, SEV=Universitat de Sevilla, LEB = Universidad de Leon, FCO= University of Oviedo, MAF=Universidad Complutense, GDA, COA = Universidad de Cordoba, ESP?? ABH=Universidad de Alicante

	Herbarium[2]	Accession Number	Country	Locality	Latitude	Longitude
57	MA	611329	Spain	Branillin, Sierra del Cueto Negro (Asturias)	42.979843	-5.768082
58	MA	89544	Spain	Reinosa (Santander) (Cantabria)	43.021186	-4.137562
59	MA	89546	Spain	Suances (Santander) (Cantabria)	43.424803	-4.043129
60	MA	89548	Spain	Monte Corona, Cabezon de la Sal (Santander) (Cantabria)	43.312041	-4.233119
61	MA	89551	Spain	De la sierra de Rozas, valle de Valdeporres (Santander) (Cantabria)	43.046976	-3.7502
62	MA	89550	Spain	Del Cueto, Castro Urdiales (Santander) (Cantabria)	43.378165	-3.23053
63	MA	437500	Spain	Cantabria: Vega de Liebana, subiendo al puerto de San Glorio, Penas de la Horcada (Cantabria)	43.058962	-4.664011
64	MA	544531	Spain	Prov. Oviedo. NW part of Sierra de los Vientos, 1.5 km NW of Munas de Abajo. (Cantabria)	43.489086	-6.427624
65	MA	653596	Spain	Santander, Bezana (Cantabria)	42.958261	-3.836379
66	MA	724824	Spain	Soba, Puerto de los Tornos (Cantabria)	43.136408	-3.455702
67	MA	751132	Spain	Cantabria, Heras (Cantabria)	43.394508	-3.760942
68	MA	719040	Portugal	Azores, isla Pico, Cabeco das Cabras a Ponta do Pico	38.476018	-28.423641
69	MA	89564	Portugal	Ponte de Barca, entre Travassos e Povoa.	41.803102	-8.421569
70	MA	755675	Spain	Leon, Marana, Mampodre (Castilla & Leon)	43.031034	-5.199403
71	MA	695387	Spain	Palencia, 7 km de Villalba de Guardo, laguna de la isla (Castilla & Leon)	42.697782	-4.906707
72	MA	643861	Spain	Burgos, cercanias de Huidobro, melojar sobre suelo arenoso (Castilla & Leon)	42.766799	-3.692835
73	MA	467362	Spain	Burgos, Basconcillos del Tozo (Castilla & Leon)	42.711197	-3.994241
74	MA	89535	Spain	Castille: Montagnes de Pancorvo (Castilla & Leon)	42.75089	-3.388364
75	MA	317884	Spain	Valdegovia (VI): Vallegrull. (Logrono, Pais Vasco, Navarra)	42.878389	-3.251618
76	MA	249093	Spain	Guipuzcoa: Cestona, monte de Santa Engracia (Logrono, Pais Vasco, Navarra)	43.238635	-2.26316
77	MA	301091	Spain	Pto de Barazar (Vizcaya) (Logrono, Pais Vasco, Navarra)	43.064448	-2.708559
78	MA	89553	Spain	Sierra Urbasa (Navarra). Pau. (Logrono, Pais Vasco, Navarra)	42.831658	-2.160526
79	MA	614529	Spain	Alava: Asparenna-Zalduendo, sierra de Urkilla, Zamarraundi(Logrono, Pais Vasco, Navarra)	42.913002	-2.316988
80	MA	173677	Ireland	on roadside bank above Streamstown bay, 2 miles north of Clifden, West Co. Galway	53.527561	-10.056251
81	MA	301093	Ireland	lieu-dit Claggan, le long de la route L101, 4km SW de Cornamona, pres de la partie W de Lough	53.525832	-9.505096
82	MA	768540	Portugal	Minho, pr. Figueiredo	41.859444	8.404722
83	MA	488319	Spain	Asturias, Ibias: puerto del Conio, entre Moal y Centenales km 32-33	43.03316239	-6.72911405
84	MA	329743	Spain	Puerto de las Senales (Leon)	43.077352	-5.260248
85	K	H/1503/94 - 29	Spain	Roadside 2 km S.E. of Espinama, Picos de Europa, Santander Prov.	43.113942	-4.771071
86	K	H/1503/94 - 30	Spain	Province Santander, rocky roadside bank between Pechon and Unquera, above Ria Tinamayor	43.379015	-4.50496
87	K	s.n.	France	Pyrenees: Roncesvalles, near head of the pass	43.007592	-1.317786
88	K	s.n.	France	Towards the Spanish frontier from St Jean Pied de Port	43.152748	-1.234332
89	K	s.n.	Spain	Santander province; below Puerto des Alisas, SW of Santander	43.301767	-3.651959
90	K	s.n.	Spain	Province Santander, East side of mouth of Tina Mayor estuary, 2km north east of Unquera	43.384682	-4.509045
91	E	112/94 - 6	Spain	Dist. Leon: Below (North of) Puerto de Ponton. Cordillera Cantabrica, near Picos de Europe	43.112778	-5.015
92	E	E00079124	Spain	Prov. Navarra: Burguete; Valcalos.	43.033311	-1.322617
93	E	112/94 - 1	France	Inter Itsatsou et le pas de Roland prope thermas Cambo. Pyr. Occid.	43.323609	-1.397366
94	E	112/94 - 27	Ireland	nr, Maam Cross, Co. Galway	53.458701	-9.542411
95	E	112/94 - 62	Ireland	Morisk, Co. Mayo	53.777779	-9.641153
96	E	112/94 - 61	Ireland	Letterfrack, Galway	53.552979	-9.951362
97	E	112/94 - 56	Ireland	Ballynahinch, Galway	53.452376	-9.86731
98	E	112/94 - 32	Ireland	Along south shore of Lough Corrib, NW of Oughterard, Connemara	53.46485	-9.36572
99	E	112/94 - 30	Ireland	Inverin, Jarconnaught, Co. Galway	53.242953	-9.457307
100	E	112/94 - 34	Ireland	Lough Fee near Kylemore, Connemara, W. Galway	53.575342	-9.813077
101	E	112/94 - 36	Ireland	nr Cregduff Lough, W. Galway	53.388676	-9.92667
102	E	112/94 - 51	Ireland	On the high ground at the back of Renvyle House with the other heathers	53.608318	-9.996568
103	BCN	59950	Spain	Sotoa l'Hotel de la Corza Blanca. Brañavieja (Alto Campoo)	43.018917	-4.21357
104	BCN	59952	Spain	Pinar de Puebla de Lillo. Cord. Cantabria. Léon	43.020563	-5.276985
105	BCN	59954	Spain	S. Vicente de la Barquera	43.385227	-4.400386
106	BCN	59955	Spain	Peña Labra, sota la Corza Blanca	43.042515	-4.386136
107	BCN	59956	France	Pyr Atlantiques; Entre Arette la Pierre St Martin I Osse-en-Aspe	43.036943	-0.691395
108	BCN	59957	Spain	Pueble de Lillo (Léon)	43.02053064	-5.31705401
109	BCN	59958	Spain	Navarra: Valle de Irati entre Olaldea I Arive	42.935964	-1.278376
110	BCN	59959	Spain	Asturias, Siero, entre San Miguel del Marcenado y el cuartel Cabo Noval	43.41841637	-5.76697329
111	BCN	59961	Spain	Brañillín, Puerto de Pajares (Asturias)	43.00438300	-5.76057352
112	BCN	59962	Spain	Araba, Puerto de Azaceta	42.72146129	-2.51145420
113	BCN	59964	Spain	Guitiriz (Lugo)	43.178389	-7.913751
114	BCN	59968	Spain	Lourizán (Pontevedra)	42.401906	-8.670685
115	BCN	59970	Spain	Asturias, Siero, entre San Miguel del Marcenado y el cuartel Cabo Noval	43.41841637	-5.76697329
116	BCN	59971	Spain	A Coruña: Rois	42.72224273	-8.75572385
117	BCN	59972	Spain	Leon; Cofiñal	43.04216034	-5.27121795
118	BCN	59978	France	Basses Pyrénées: Vallée du Saison, entre Licq et Saint Engrâce	43.044129	-0.882141
119	BCN	22245	Spain	Puerto del Escudo, hacia Santander	43.049606	-3.860047
120	BCN	38487	Spain	Cantabria: Heras	43.33362702	-5.09702951
121	BCN	44013	Spain	Oviedo; Caravia c. Mirador del Fito	43.4394493	-5.19409831
122	LY	Prêt 01	France	Bedous, Pyrenées Occidentales	42.997917	-0.59376
123	LY	Prêt 02	Spain	Obarenes bois	42.685212	-3.131847
124	LY	Prêt 11	France	Pas de Roland, près Cambo, Basses Pyrénees	43.314148	-1.39908
125	LY	Prêt 12	Spain	Fontarrabie (Hondarribia)	43.363534	-1.816883
126	LY	Prêt 24	France	Pic de Listo, The-Valée d'Ossau, Basses Pyrenées	42.996929	-0.389619

	Herbarium[2]	Accession Number	Country	Locality	Latitude	Longitude
127	LY	Prêt 32	France	Maine-et-Loire, Foret de Brissac, près de l'ancien étang de Monthayer	47.333717	-0.457847
128	LY	Prêt 53	Spain	Bentanzos, Galicia	43.280959	-8.211219
129	LY	Prêt 66	France	Eaux Bonnes, Basses Pyrénées	42.976363	-0.388999
130	LY	Prêt 70	Spain	Montes en Obarenes, Pancorbo	42.635541	-3.114394
131	LY	Prêt 77	France	Lescun (B-Pyr)	42.936427	-0.634614
132	LY	Prêt 82	Spain	Soncillo, Burgos	42.967773	-3.786873
133	LY	Prêt 83	Spain	Biscaye, Viscaya, Urberuaga	43.331473	-2.682496
134	LY	Prêt 84	Spain	Oviedo, S Esteban in collibus maritimis	43.341982	-5.816532
Erica mackaiana						
1	SALA	67684-1	Spain	Saja	43.14449	-4.284419
2	SALA	102872-1	Spain	Siero, Bimenes, Campa El Calero, proximidades a El Plano	43.33001	-5.593888
3	MA	598852-1	Spain	Abaño, San Vicente de la Barquera	43.36115	-4.384678
4	GDA	51626-1-1	Spain	Asturias, Concejo de Caravia, subida al Mirador del Fito, Duyos, próx. Cuetu L´Abeyera	43.45883	-5.204777
5	MA	680886-1	Spain	Bárcena Mayor	43.14486	-4.191592
6	ARAN	23618-1	Spain	Bustriguado. Bustriguado ; Sierra del Escudo de Cabuerniga	43.2942	-4.350093
7	SANT	35445	Spain	C: A Capela, Vilar de Mouros, Serra de Forgoselo	43.4632	-8.079104
8	SANT	42691	Spain	C: Monfero; Serra da Loba	43.32404	-8.054459
9	SEV	42037-1	Spain	Cadavedo. Luarca	43.5434	-6.387609
10	LEB	2767-1	Spain	Canero. Argumosa	43.54233	-6.462756
11	FCO	25709-1	Spain	Cangas de Onís; Margolles, La Granda	43.40753	-5.117613
12	SALA	111281-1	Spain	Caravia, c. Mirador del Fito	43.4521	-5.189668
13	MAF	158091-1	Spain	Cariño, Serras da Capelada	43.72949	-7.884328
14	FCO	29374-1	Spain	Carreño; Próximo a Tamón	43.53566	-5.826513
15	MA	680884-1	Spain	Comillas	43.38287	-4.294132
16	SANT	12076	Spain	Cuadramón (Lugo)	43.48357	-7.499627
17	MA	594105-1	Spain	Cudillero, Ballota, ermita de San Roque	43.54193	-6.321879
18	FCO	11573-1	Spain	Cudillero; explanada de Oviñana.	43.59	-6.24
19	MA	680894-1	Spain	Divisoria entre Los Corrales de Buelna y San Felices (entre "La Garmia" y el "Acebal del Iso").	43.25633	-4.026989
20	FCO	28368-1	Spain	El Franco; Andina	43.48641	-6.828032
21	SEV	111487-1	Spain	Entre Pechon et Pesues, à l'est d'Unquera	43.38236	-4.490482
22	MA	642592-1	Spain	Grado	43.39873	-6.070468
23	FCO	29241-1	Spain	Grado; A 1 km de San Martín de Gurullés en dirección a Grado	43.16953	-6.103043
24	FCO	16861-1	Spain	Llanes; Niembro	43.44098	-4.848605
25	MA	680890-1	Spain	Los Tojos (sobre Soja).	43.15573	-4.255115
26	GDA	23192-1-1	Spain	Lugo, c. Mondoñedo.	43.42523	-7.361202
27	SANT	50001	Spain	Lugo: A Pastoriza	43.30135	-7.348079
28	SANT	50002	Spain	Lugo: A Pontenova	43.34652	-7.195309
29	FCO	25712-1	Spain	Melide; Mellid	42.91653	-8.007449
30	MA	680893-1	Spain	Molledo (ladera del Pico Jano).	43.12195	-4.032683
31	SANT	12075	Spain	Momán (Lugo)	43.33488	-7.858677
32	FCO	2970-1	Spain	Morcín; laderas del Monsacro, Sierra del Aramo.	43.26885	-5.884322
33	FCO	14948-1	Spain	Navia; proximidades de la playa de La Vega, Puerto Vega.	43.5584	-6.676712
34	GDA	52671-1-1	Spain	Oviedo, Caravia, c. Mirador del Fito	43.36334	-5.789415
35	COA	21353-1	Spain	Parque Natural de Urkiola	43.10215	-2.641404
36	MA	680883-1	Spain	Pechón (Val de San Vicente)	43.38975	-4.496968
37	MNHN	P00420300	Spain	Peñaflor [Asturies, près d'Oviedo]	43.40259	-6.042766
38	MA	680885-1	Spain	Pendueles	43.3992	-4.636457
39	COA	21352-1	Spain	Playa Cueva. Luarca	43.55279	-6.470928
40	SALA	110436-1	Spain	Puente Viesgo, Alto de Capia	43.304	-4.012608
41	SANT	12631	Spain	Puerto de la Xesta, NE of Villalba	43.38857	-7.419406
42	MNHN	P00392384	Spain	Raices, près Avilés (Oviedo)	43.56962	-5.951949
43	ESP003	NC064347	Spain	San Esteban, Pravia, province of Asturias	43.55638	-6.085299
44	SANT	12175	Spain	San Román (Lugo)	42.88995	-7.570404
45	MA	680887-1	Spain	San Sebastián de Garabandal (Rionansa)	43.20259	-4.422502
46	ARAN	23619-1	Spain	San Vicente de la Barquera. San Vicente de la Barquera	43.38493	-4.403776
47	MA	680892-1	Spain	Santillana del Mar, falda del cerro Huervo, entre Santillana y Ubiarco	43.39257	-4.113932
48	ABH	44473-1	Spain	Siero; Bimenes, Campa El Calero	43.33012	-5.571846
49	MA	680888-1	Spain	Sobre Bárcena Mayor (Los Tojos), cueva de los Albarqueros.	43.14305	-4.192215
50	MA	621894-1	Spain	Taramundi, Ouría	43.40879	-7.058079
51	MA	680889-1	Spain	Ucieda (Ruente).	43.26088	-4.26963
52	FCO	24250-1	Spain	Valdés; Busto	43.5581	-6.466588
53	FCO	11487-1	Spain	Valdés; Cadavedo	43.54651	-6.387535
54	FCO	11800-1	Spain	Valdés; camino vecinal a Argumoso, Canero.	43.54208	-6.463299
55	FCO	14684-1	Spain	Valdés; Leiriella.	43.46676	-6.538015
56	SEV	117114-1	Spain	Viveiro, Chavin, Monte Castelo	43.61128	-7.612713

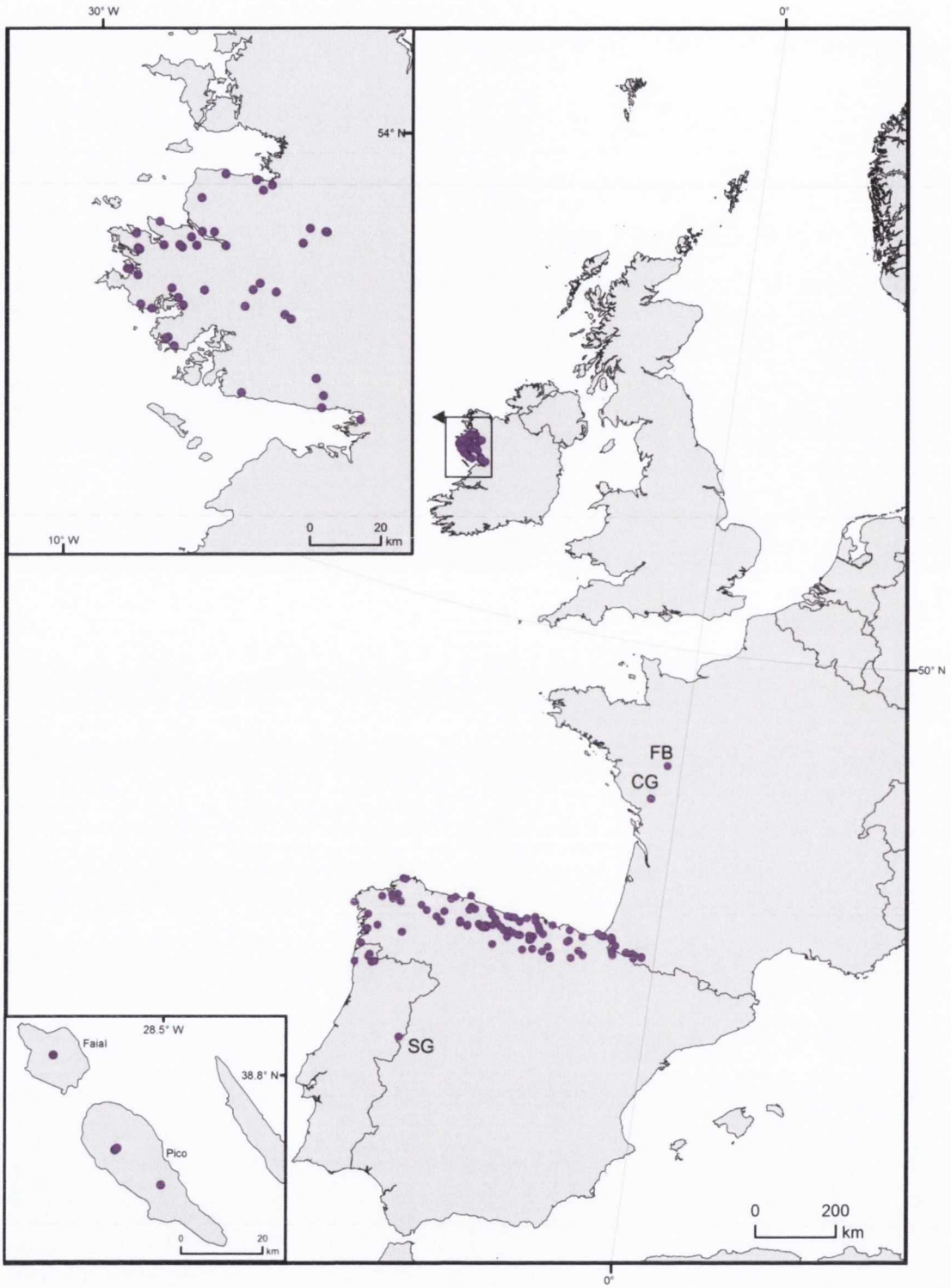


Figure 18: Occurrence points from the current distribution range of *D. cantabrica,* as given in Table 7. Isolated populations identified as follows; CG = St Cyr-des-Gats, Vendee, FB = Forêt de Brissac, Maine-et-Loire, SG=Sierra de Gata, Extremadura.

Figure 18 and Figure 19 show the distribution of the 541 presence records used in these analyses. All are current, the previously recorded localities in Moissac (Tarn-et-Garonne)

and Gensac (Gironde) and which have not been seen since the 18th century were not included in the dataset.

D. cantabrica's distribution in Ireland, as shown in Figure 18, is clearly limited to the peninsula bounded north and south by Clew bay and Galway Bay, respectively, and comprised mainly of County Galway, and part of southwest County Mayo. In Europe the bulk of the *D. cantabrica* records are along the northern Spanish coast, and in the border area between France and Spain.

Three disjunct populations exist, two in the Pays de la Loire; in the départements of Maine-et-Loire and Vendée which are 350km and 440km away from the nearest southern populations, respectively, and in the Spanish province of Extremadura very close to the Portuguese border, 130km away from its nearest population to the north-west of Portugal.

Figure 19 shows that on the continent *E. mackaiana* exists in a fairly continuous band along the northern Spanish coastline. The easternmost limit of the bulk of the localities is approximately 200km away from the French border, staying to the west of Santander, however there is one locality 200km further east (in Urkiola National Park), recorded from a specimen lodged at Cordoba Botanic Gardens (COA, #21353-1).

In Ireland it grows in a far more disjunct pattern of isolated populations in the counties of Kerry, Galway, Mayo and Donegal.

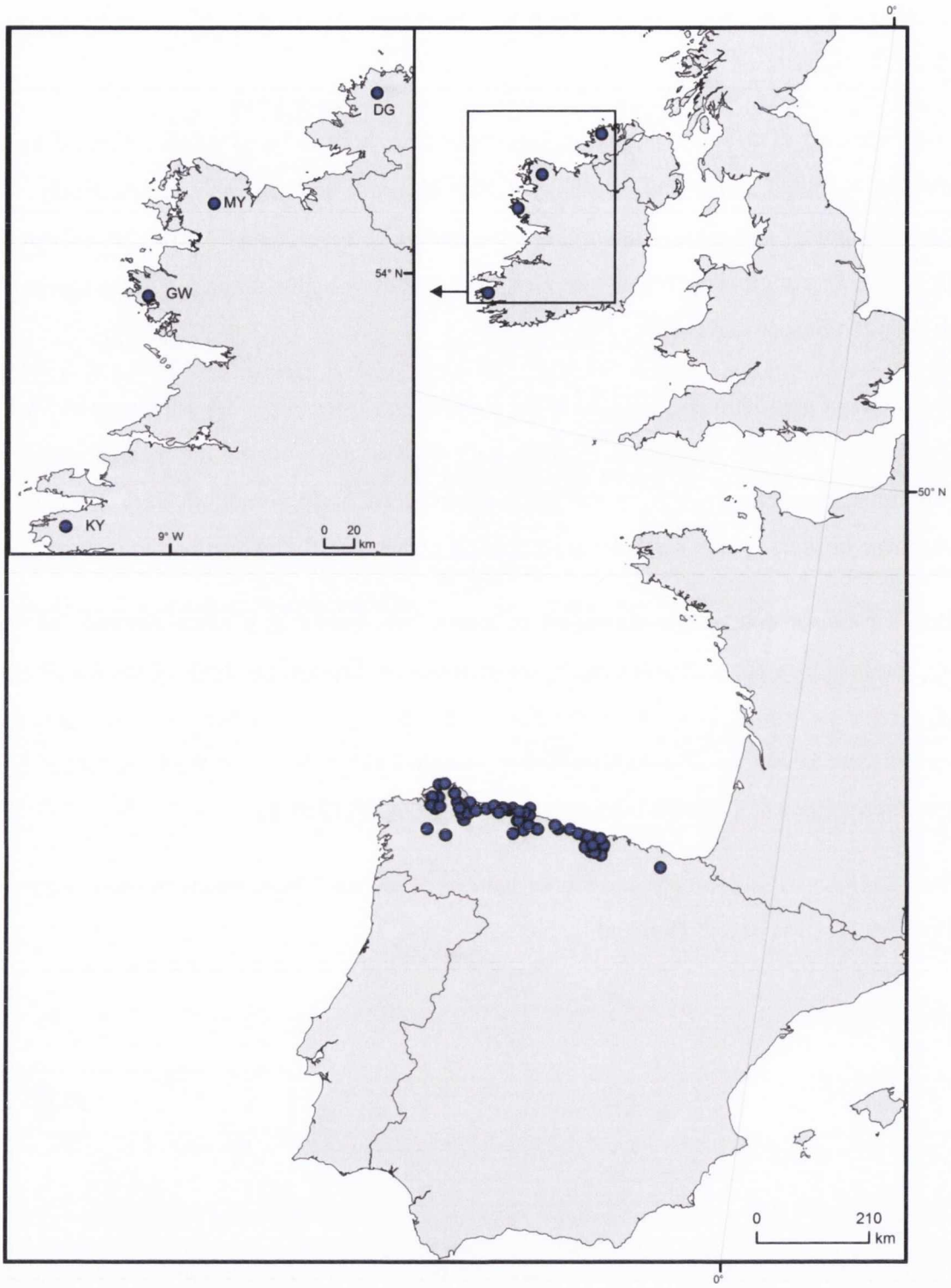


Figure 19: Occurrence points from the current distribution range of *E. mackaiana* in Europe. Populations of *E. mackaiana* in Ireland identified as follows; GW=Roundstone bog, Co. Galway, DG = Lough Nacung, Co. Donegal, MY = Bellacorick, Co. Mayo, KR = Cahirseveen, Co. Kerry.

2.3.2 Bioclimatic ranges for study species

Table 8: Upper and lower limits for the 19 bioclimatic variables at *D. cantabrica* and *E. mackaiana* population localities

Variable	*D. cantabrica*		*E. mackaiana*	
	MIN	MAX	MIN	MAX
1 Annual Mean Temperature	4.3	14.9	8.2	14.4
2 Mean Diurnal Range	5.8	10.5	5.9	10.0
3 Isothermality (BIO2/BIO7)	3.4	4.4	3.7	4.4
4 Temperature Seasonality	3.2	6.0	3.1	4.7
5 Max Temperature of Warmest Month	17.6	27.4	16.4	24.9
6 Min Temperature of Coldest Month	-5.6	7.7	0.0	6.3
7 Temperature Annual Range	14.8	26.3	14.5	23.5
8 Mean Temperature of Wettest Quarter	-0.7	12.7	4.8	12.8
9 Mean Temperature of Driest Quarter	9.2	20.2	9.1	19.2
10 Mean Temp. of Warmest Quarter	11.3	20.3	13.2	19.5
11 Mean Temp. of Coldest Quarter	-1.6	10.2	4.1	9.7
12 Annual Precipitation	673	1481	788	1481
13 Precipitation of Wettest Month	72	226	106	172
14 Precipitation of Driest Month	10	70	29	82
15 Precipitation Seasonality	13	53	20	41
16 Precipitation of Wettest Quarter	204	600	274	499
17 Precipitation of Driest Quarter	65	259	113	255
18 Precipitation of Warmest Quarter	69	296	130	327
19 Precipitation of Coldest Quarter	186	596	204	472

Overall the range of conditions in which in which *D. cantabrica* is found is wider than that for of *E. mackaiana.* Both species had mean annual temperature ranges which did not exceed 15°C, with *D. cantabrica* tolerating temperatures as low as -5.6°C and as high as 27.4°C, whilst *E. mackaiana* had a slightly more moderate range of between 0°C and 24.9°C.

In terms of precipitation both species were found in locations with annual rainfall reaching a maximum of 1481mm. Even the driest months had precipitation levels of 10-70mm for *D. cantabrica*, and 29-82mm for *E. mackaiana.*

2.3.3 Selection of environmental variables for SDM

A correlation matrix for the 19 bioclimatic variables across the area of Europe showed many variables highly correlated to each other. Table 2 shows the correlation matrix using Pearson's correlation statistic.

Table 9: Correlation values of bioclimatic layers. Correlation values above r=0.8 are highlighted in red.

	BIO1	BIO2	BIO3	BIO4	BIO5	BIO6	BIO7	BIO8	BIO9	BIO10	BIO11	BIO12	BIO13	BIO14	BIO15	BIO16	BIO17	BIO18	BIO19
BIO1	1.000																		
BIO2	0.622	1.000																	
BIO3	0.561	0.555	1.000																
BIO4	0.217	0.661	-0.227	1.000															
BIO5	0.902	0.863	0.466	0.577	1.000														
BIO6	0.891	0.263	0.583	-0.228	0.623	1.000													
BIO7	0.446	0.894	0.132	0.919	0.784	0.003	1.000												
BIO8	0.520	0.355	0.196	0.299	0.497	0.385	0.330	1.000											
BIO9	0.769	0.605	0.481	0.202	0.763	0.654	0.456	0.101	1.000										
BIO10	0.963	0.747	0.445	0.469	0.977	0.743	0.659	0.537	0.766	1.000									
BIO11	0.953	0.437	0.634	-0.081	0.750	0.978	0.182	0.421	0.744	0.842	1.000								
BIO12	-0.679	-0.668	-0.437	-0.430	-0.743	-0.466	-0.579	-0.551	-0.421	-0.728	-0.556	1.000							
BIO13	-0.484	-0.574	-0.356	-0.420	-0.573	-0.277	-0.513	-0.488	-0.223	-0.543	-0.353	0.950	1.000						
BIO14	-0.833	-0.700	-0.439	-0.393	-0.874	-0.643	-0.607	-0.465	-0.703	-0.871	-0.744	0.814	0.613	1.000					
BIO15	0.607	0.339	0.190	0.121	0.577	0.556	0.296	0.090	0.688	0.606	0.617	-0.173	0.104	-0.667	1.00				
BIO16	-0.499	-0.597	-0.371	-0.434	-0.592	-0.282	-0.533	-0.513	-0.233	-0.561	-0.364	0.956	0.996	0.623	0.099	1.00			
BIO17	-0.829	-0.678	-0.441	-0.374	-0.859	-0.652	-0.581	-0.482	-0.661	-0.860	-0.744	0.848	0.657	0.990	-0.638	0.664	1.000		
BIO18	-0.852	-0.728	-0.547	-0.333	-0.878	-0.681	-0.582	-0.376	-0.754	-0.871	-0.779	0.841	0.663	0.961	-0.599	0.672	0.963	1.000	
BIO19	-0.430	-0.545	-0.245	-0.477	-0.539	-0.197	-0.533	-0.573	-0.141	-0.510	-0.281	0.913	0.955	0.571	0.138	0.963	0.608	0.577	1.000

After discarding highly correlated variables (r>0.8), eight variables remained (Table 10). These eight variables were used then for the following PCA analysis of the niche, and for the development of the species distribution model.

Table 10: The eight bioclimatic variables used to construct the species distribution model

BIO1 Annual Mean Temperature	BIO2 Mean Diurnal Temperature Range	BIO3 Isothermality	BIO4 Temperature Seasonality
BIO8 Mean Temperature of Wettest Quarter	BIO9 Mean Temperature of Driest Quarter	BIO12 Annual Precipitation	BIO15 Precipitation Seasonality

2.3.4 Visualising the climatic envelope

A principal components analysis was carried out on the eight climatic variables across all locality points for each species. The factor loadings of each of the variables is shown in Table 11.

Table 11: Factor loadings from PCA of bioclimatic variables for *D. cantabrica* and *E. mackaiana*

D. cantabrica								
Importance of components:								
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
Standard deviation	1.7445929	1.602529	0.9815324	0.9311469	0.6519779	0.3311514	0.1271604	0.0833661
Proportion of Variance	0.3804505	0.3210124	0.1204257	0.1083793	0.0531344	0.0137077	0.0020212	0.0008687
Cumulative Proportion	0.3804505	0.7014629	0.8218887	0.930268	0.9834024	0.99711	0.9991313	1
Loadings:								
BIO1: Annual Mean Temp.	-0.348	0.473		-0.103	-0.211	0.457	-0.492	0.386
BIO2: Mean Diurnal Temp. Range	0.475	0.228	0.395	0.126		-0.292	0.137	0.662
BIO3: Isotherm-ality	-0.164	-0.193	0.886	-0.267		0.13		-0.232
BIO4: Temp. seasonality	0.499	0.266		0.195	-0.182	-0.155	-0.586	-0.489
BIO8: Mean Temp. of Wettest Qtr.	-0.376	0.373		-0.417		-0.73		
BIO9: Mean Temp. of Driest Qtr.		0.608			-0.204	0.284	0.619	-0.334
BIO12: Annual Precipitation	-0.362	-0.217		0.568	-0.676	-0.198		
BIO15: Precipitation seasonality	-0.319	0.246	0.222	0.602	0.638	-0.108		
E. mackaiana								
Importance of components:								
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
Standard deviation	2.0203501	1.2586966	1.1772447	0.920297	0.2310421	0.1816529	0.0864542	0.0846425
Proportion of Variance	0.5102268	0.1980396	0.1732381	0.1058683	0.0066726	0.0041247	0.0009343	0.0008955
Cumulative Proportion	0.5102268	0.7082664	0.8815046	0.9873729	0.9940454	0.9981702	0.9991045	1
Loadings:								
BIO1: Annual Mean Temp.	-0.434		0.386		-0.137	-0.487	0.57	0.267
BIO2: Mean Diurnal Temp. Range	-0.321	0.45	-0.417	0.125		0.268	0.513	-0.405
BIO3: Isotherm-ality		0.624	0.191	0.617	-0.102	-0.169	-0.288	0.263
BIO4: Temp. seasonality	-0.31		-0.635	-0.229	-0.196		-0.196	0.609
BIO8: Mean Temp. of Wettest Qtr.	-0.4	-0.295	0.325	0.221	-0.275	0.703		0.152
BIO9: Mean Temp. of Driest Qtr.	-0.482			-0.201	-0.302	-0.294	-0.514	-0.531
BIO12: Annual Precipitation	0.457	0.177		-0.245	-0.821		0.12	
BIO15: Precipitation seasonality	-0.102	0.533	0.345	-0.628	0.293	0.279		0.141

Figure 20 (A and B) shows the scatterplots of the first two principal components for *D. cantabrica* and *E. mackaiana*, accounting for 70% and 71% of the total variability, respectively.

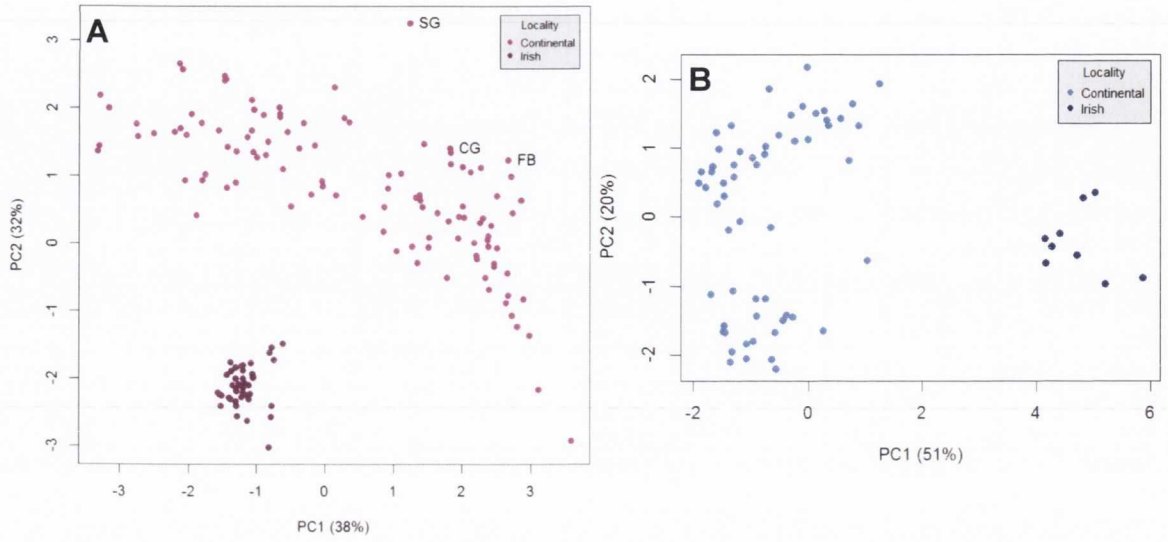


Figure 20: Principal Components Analysis of climatic variables throughout the current range of (A) *D. cantabrica* and (B) *E. mackaiana*.

In both species there is a clear difference in climate between the Irish localities and the Continental ones.

For *D. cantabrica* the variables that contribute the most towards PC1 are Mean Diurnal Temperature Range and Temperature Seasonality. The variables that contribute the most towards PC2 are Annual Mean Temperature and Mean Temperature of Driest Quarter. The two outliers in the bottom right corner of the plot correspond to two sites of high altitude in the Cantabrian mountains. The three geographically isolated populations are identified in the plot (SC, CG and FB), and their geographic location can be seen in Figure 17.

A different pattern was found for *E. mackaiana*, with precipitation variables making a bigger contribution than for *D. cantabrica*. PC1 is mainly made up of the Mean Temperature of Driest Quarter and Annual Precipitation, and PC2 of Isothermality and Precipitation Seasonality.

The spread of *D. cantabrica* localities in relation to individual bioclimatic variables was also plotted (Figure 21).

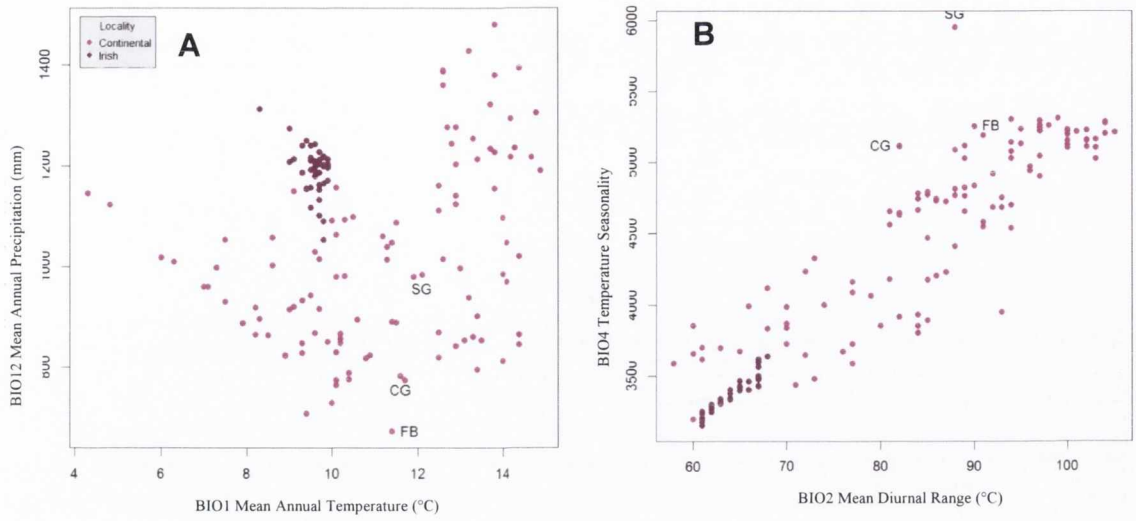


Figure 21: Plots of climatic variables at *D. cantabrica* locations

Figure 21 shows the Sierra de Gata location (SG) nested within the main spread of Annual Temperature and Precipitation, but very much an outlier when looked at across the two most important variables in PC1; Mean Diurnal Range and Temperature Seasonality.

E. mackaiana shows the Irish-Continental disjunction moderately well when viewed across Annual Precipitation and Annual Temperature (Figure 22A), however the disjunction is far more apparent when the Mean Temperature of Driest Quarter is substituted for Mean Annual Temperature (Figure 22B).

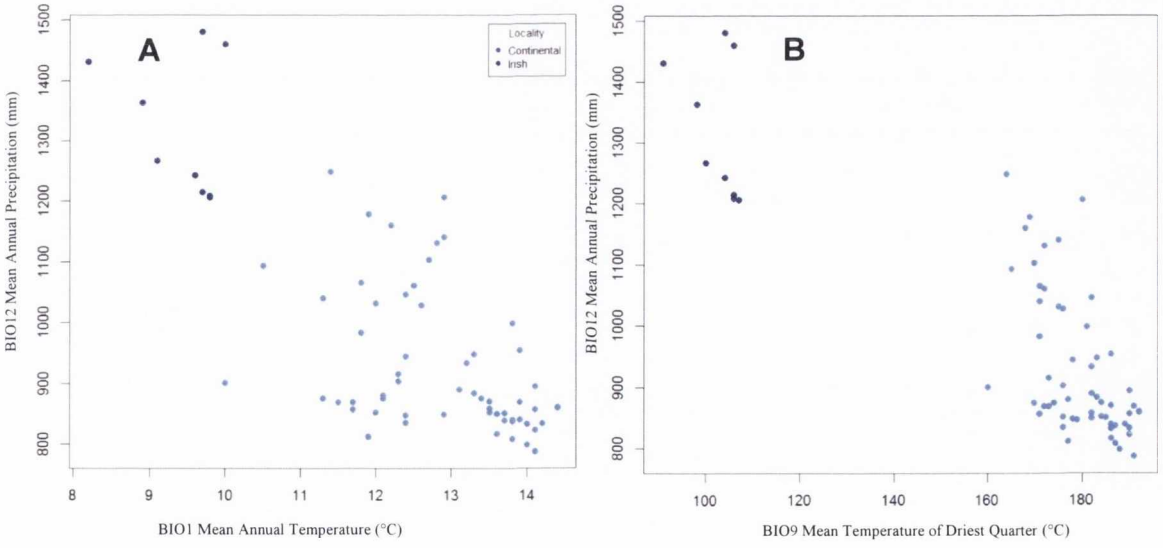


Figure 22: Plots of climatic variables at *E. mackaiana* locations

2.3.5 Model validation

The Receiver Operating Characteristics for the models run on both species (with all ten variables) are shown in Figure 23.

Both models had AUC values of over 0.9 (0.981 for *D. cantabrica*, 0.993 for *E. mackaiana*), indicating excellent performance in terms of correctly predicting presence sites as suitable areas.

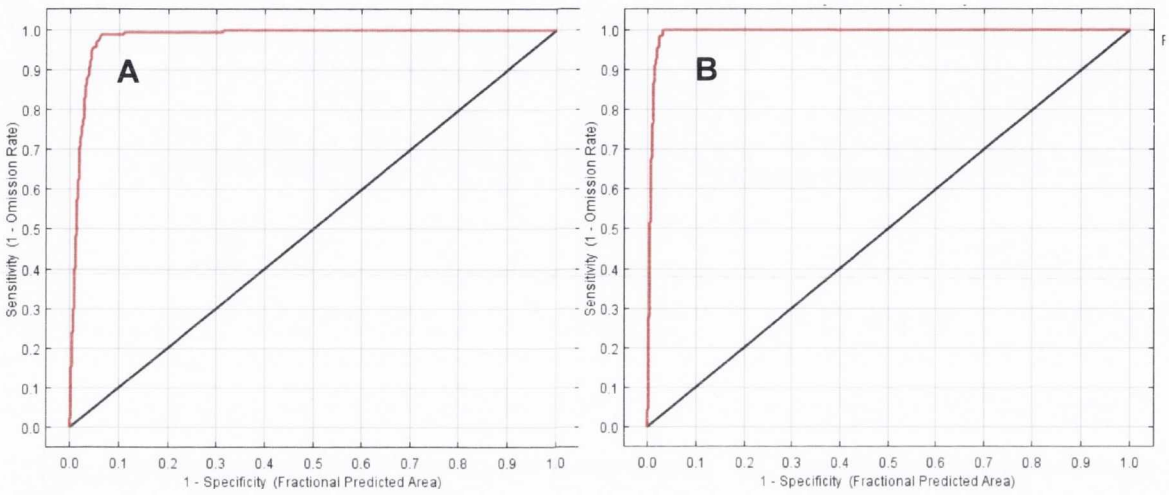


Figure 23: Receiver Operating Curves (ROC) for the MaxEnt models run for *D. cantabrica* (A) and *E. mackaiana* (B).

The results of the MaxEnt jacknife operation are shown in Figure 24 (*D. cantabrica*) and Figure 25 (*E. mackaiana*). For both species variables BIO4 and BIO9 (temperature seasonality and the mean temperature of driest quarter) were clearly the two variables contributing most to the model (showing the highest gain when used individually). For *D. cantabrica* the decrease in gain was most apparent when these two variables were each excluded, however for *E. mackaiana* it was the exclusion of BIO15 (precipitation seasonality) that decreased the gain most of all.

For *E. mackaiana* the two soil variables were the variables that contributed least to the model, with soil category (hwsd_sc) less than pH (hwsd_ph). For *D. cantabrica* soil category contributed the least but pH was more important than three of the climate variables (BIO1: annual mean temperature, BIO2: mean diurnal range and BIO15: precipitation seasonality).

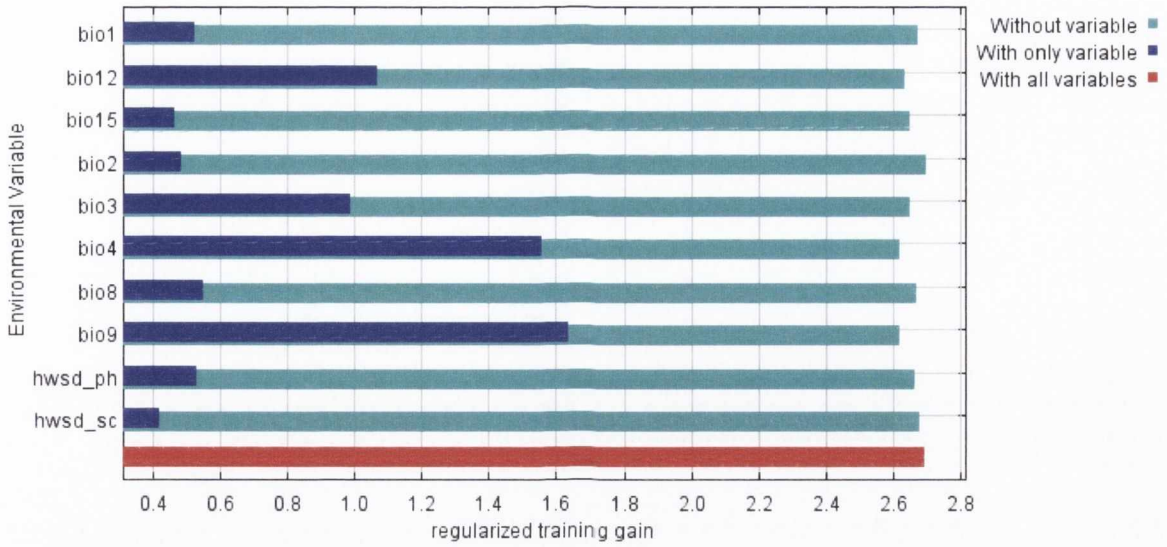


Figure 24: Jacknife results of regularised training gain for *D. cantabrica* SDM

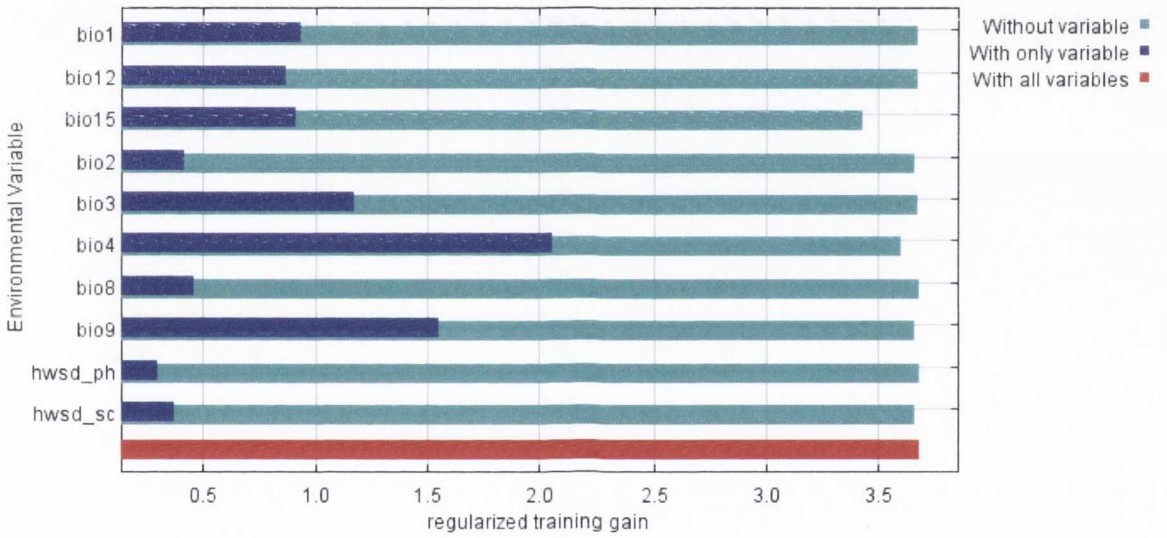


Figure 25: Jacknife results of regularised training gain for the *E. mackaiana* SDM

2.3.6 Modelling Results

The following section outlines the results of applying a MaxEnt species distribution model to the two study species to predict potentially suitable areas for their growth. For all map figures the dark shaded areas are those regions that the model predicts as suitable according to the threshold set (maximum training and specificity) as described in the Methodology (page 40).

2.3.6.1 Potential ranges under current conditions

Figure 26 and Figure 27 summarise the results of the initial runs of the SDM for both species. Each figure shows three separate runs, one with the two soil layers, one with the eight climate layers and one with all ten layers together.

The model predictions clearly showed that using soil data alone without any climate inputs resulted in huge over-predictions of suitable habitats. The models run using climate alone and climate and soil data together showed extremely similar results, and much better reflected the actual ranges of the two species.

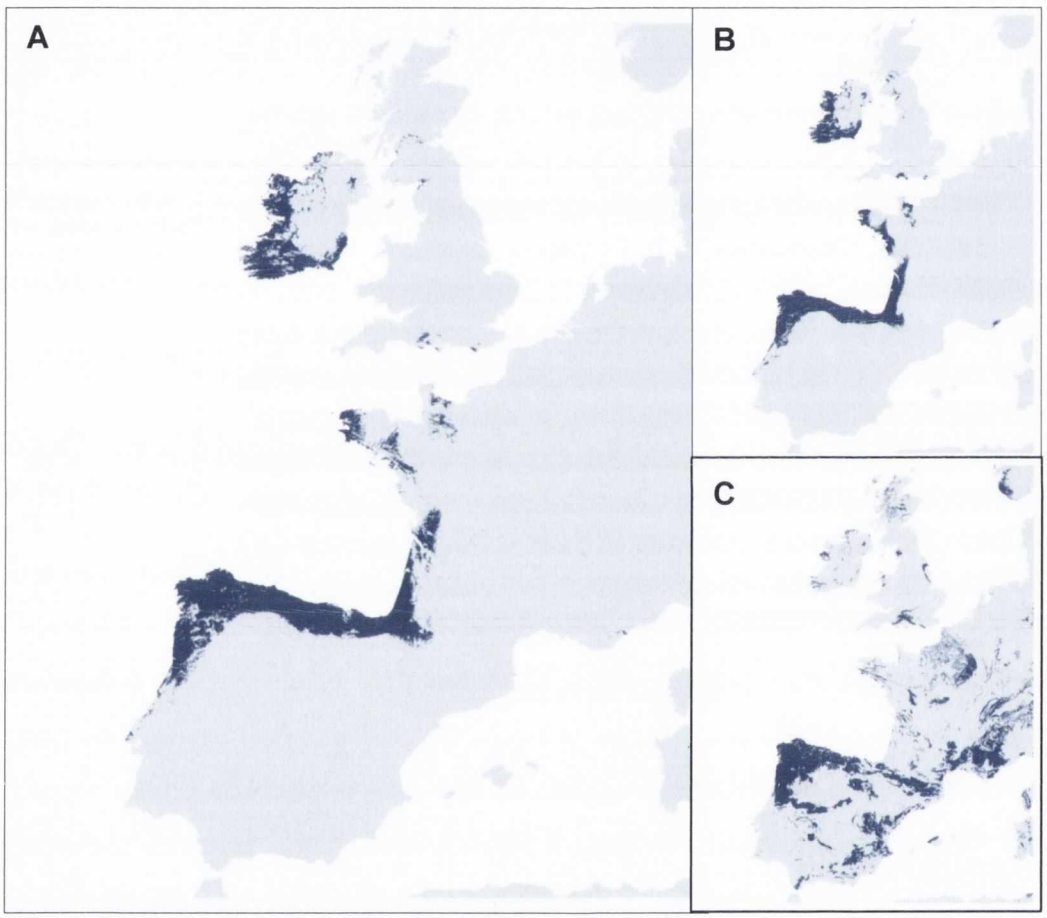


Figure 26: Potential range of *D. cantabrica* under current climatic conditions, based on Bioclimatic (B) and Soil (C) inputs separately, and combined (A).

For *D. cantabrica* the model predicts a much wider distribution in Ireland than it currently holds, including almost all the southern half of the country (Figure 26). *D. cantabrica*'s potential continental range shows a discontinuous corridor of suitable conditions from the area of the current main population up the coast of France to Brittany, as well as a small area on the northern coast of France around Normandy. The main area of discontinuity on the French coast was around the departéments of Loire Atlantique and Vendée.

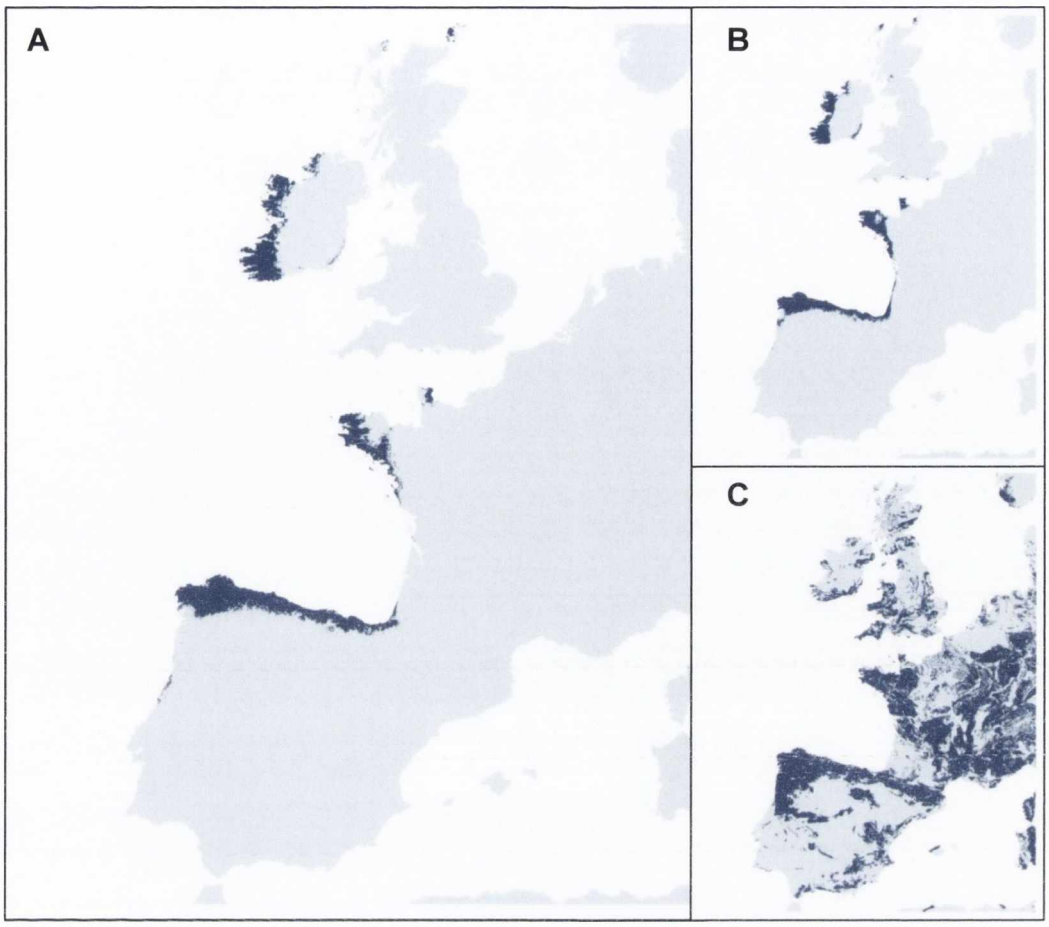


Figure 27: Potential range of *E. mackaiana* under current climatic conditions, based on bioclimatic (B) and soil (C) layers separately, and combined (A)

In Ireland *E. mackaiana*'s potential range is still very clearly restricted to the west coast, although in a more continuous band than the current disjunct populations. The potential continental range shows, in addition to the current actual range, suitable areas in the north-west of France (Brittany and Pas-de-Calais).

None of the three outlying continental populations (FB – Forêt de Brissac, SG – Sierra de Gata, CG – St-Cyr-des-Gats) are found in areas predicted by the model as suitable (Figure 28), although the St-Cyr-des-Gats population is proximal to a suitable area.

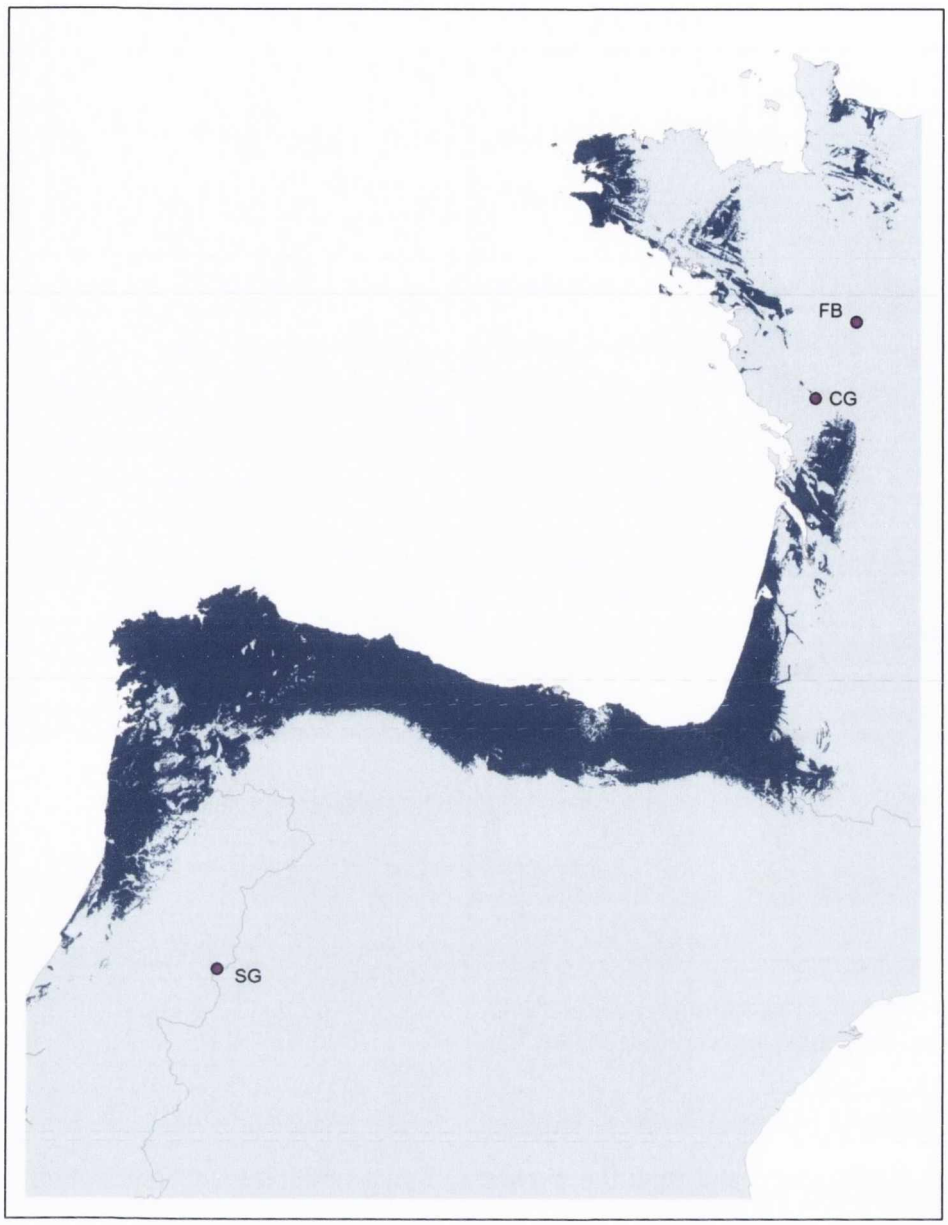


Figure 28: Outlying populations of *D. cantabrica* in continental Europe overlaid on the prediction map.

To investigate how the inclusion of these three populations affected the prediction across the entire range of the species, another model was run on a dataset without these populations. The result was a smaller area predicted as suitable, although the overall range predicted remains relatively similar (Figure 29A). The elimination of the three localities as presence points in the model decreased the predicted range overall by 16 %, mostly in the region of north-western France as well as in northern Portugal (see Figure 29B). For the rest of the predictions the models were run including these 3 populations.

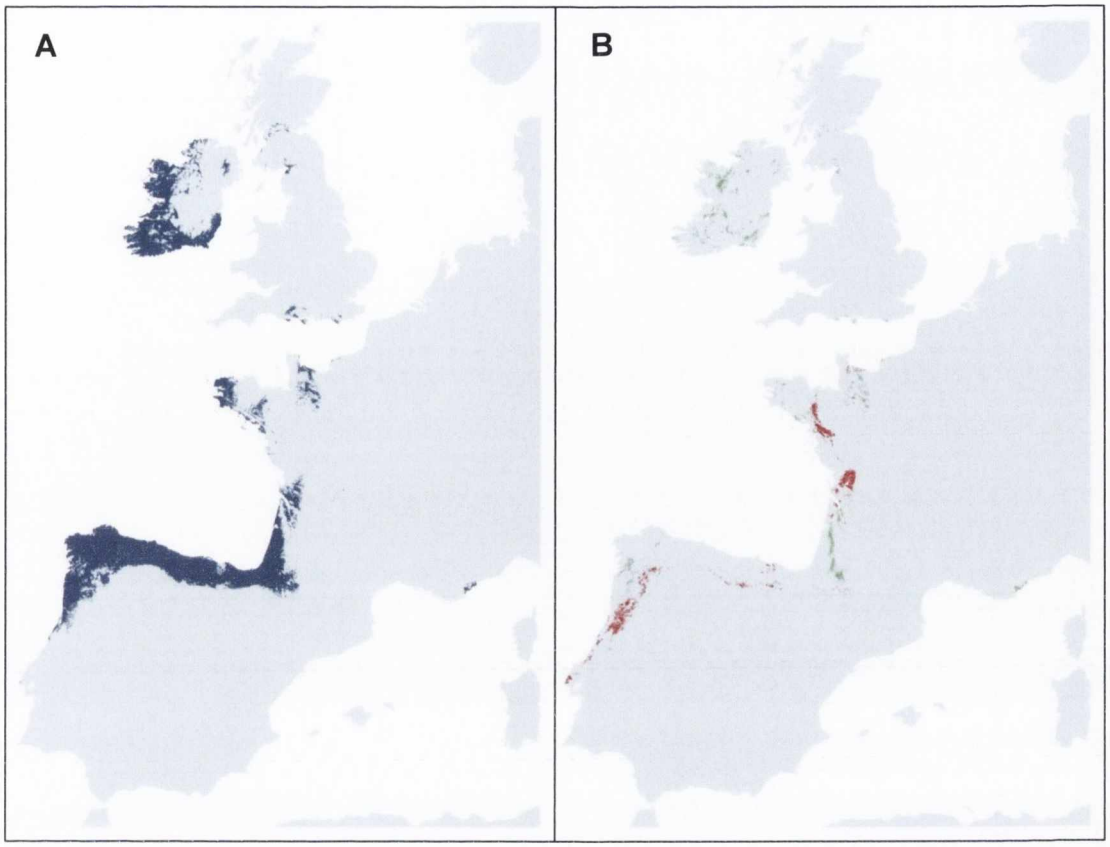


Figure 29: (A) Potential range of *D. cantabrica* modelled with the exclusion of outlying continental populations. (B) decrease (green) and increase (red) in predicted suitable areas when outliers are excluded.

The model also demonstrated the majority of the island of Britain is unsuitable for *D. cantabrica*, and does not predict suitability of areas even where naturalised populations are known to occur, with the exception of a small area of the south coast, as shown in Figure 30. These naturalised British populations were not included in the model.

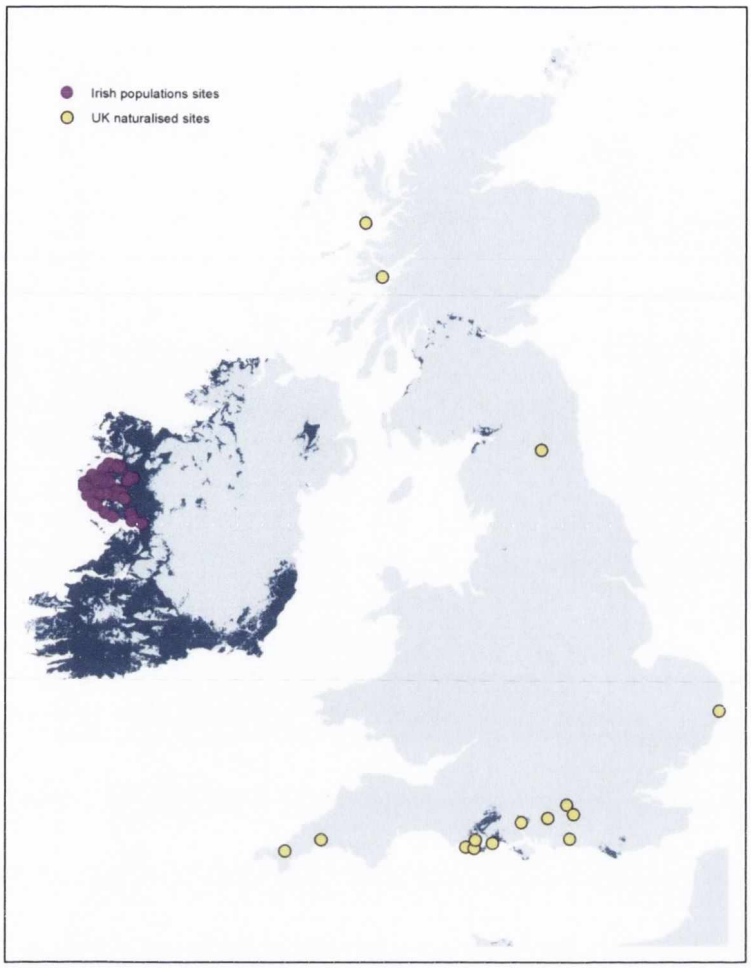


Figure 30: Naturalised sites (yellow) in the United Kingdom, and current Irish population (purple) overlaid on the predicted potential range of *D. cantabrica*.

A further series of models were run to compare the difference in range predictions if only the main continental populations were used as presence points, excluding the outliers and excluding Ireland (in other words assuming the Irish locations represented an introduced, non-native population).

For *D. cantabrica* the predicted range in Ireland was dramatically reduced when the model was run using continental locations alone as input (see Figure 31). Almost the entire west coast was then predicted as being unsuitable, except for small areas in coastal Kerry. The only area of note which was still predicted as being suitable was in Waterford.

On the continent the main difference between the two predictions was the reduction in the width and continuity of the predicted range along the southwest coast of France.

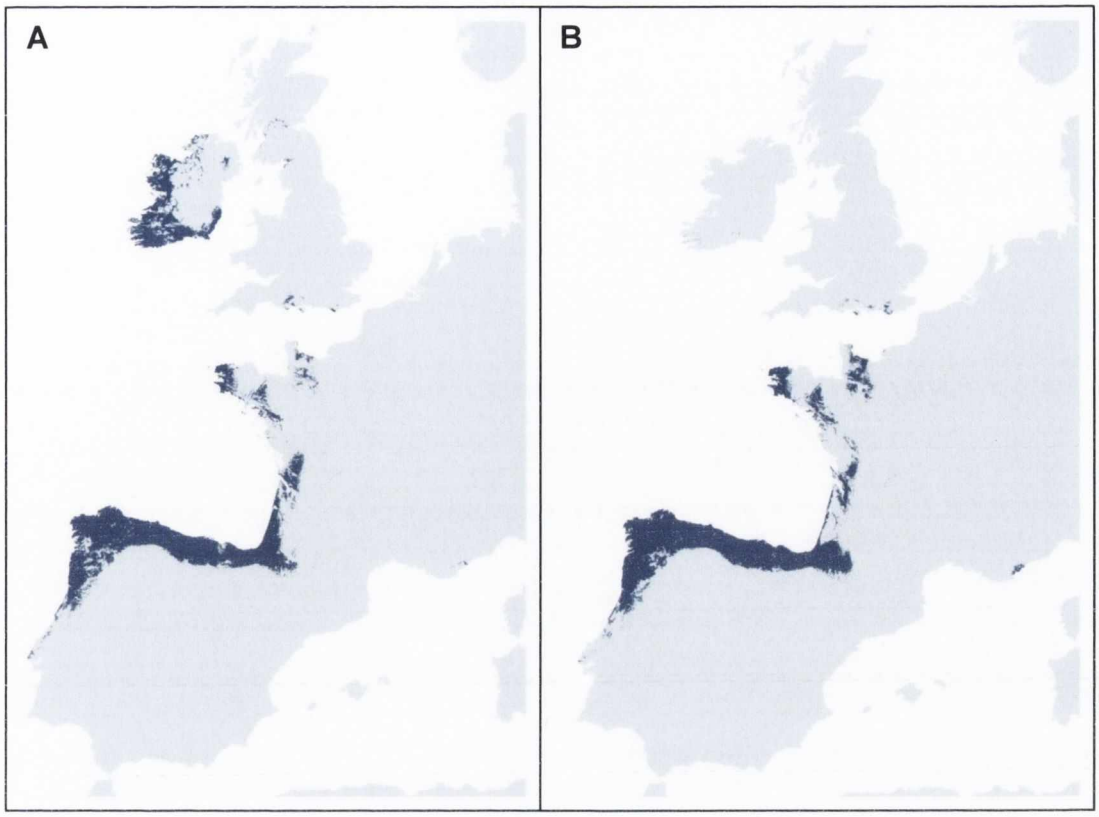


Figure 31: Predicted range of *D. cantabrica* based on all known localities (A) and only main continental range localities (B).

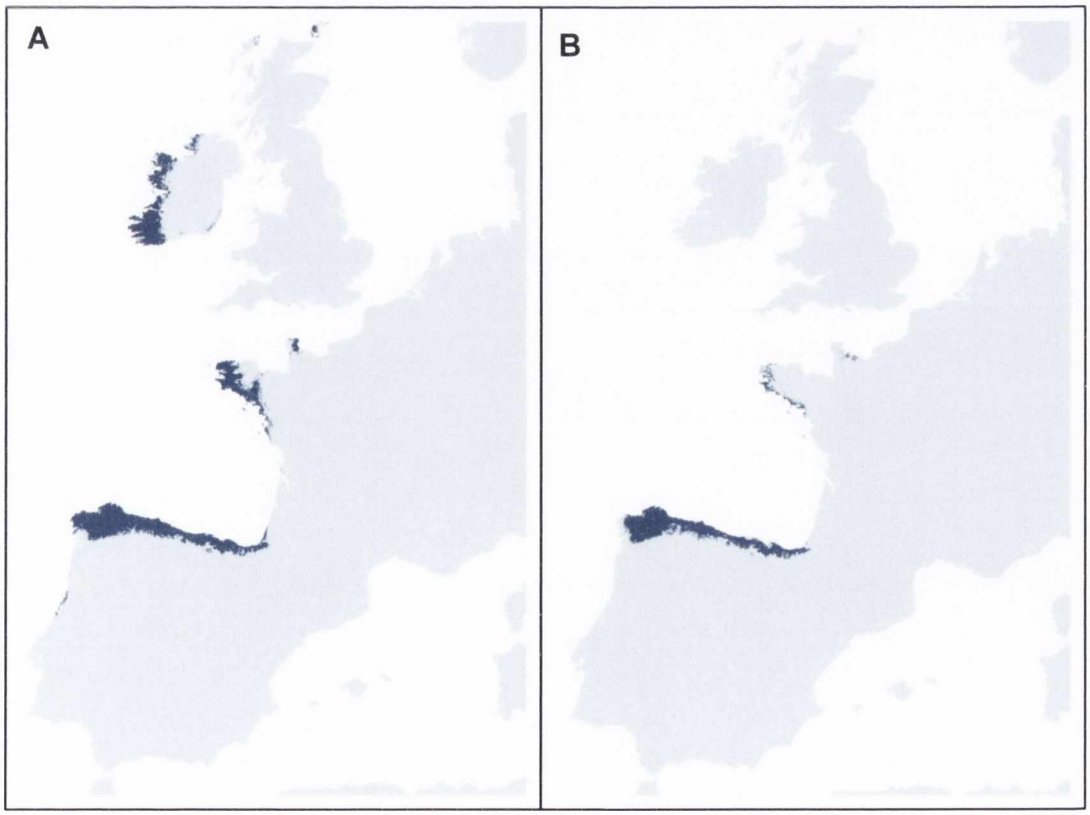


Figure 32: Predicted range of *E. mackaiana* based all known localities (A) and only main continental range localities (B).

For *E. mackaiana* an even more extreme change was found when Irish localities were excluded as presence points; the model made no predictions of suitable habitats in either Ireland or Britain, and the continental range was clearly reduced, most noticeably in the north west of the French coast (Figure 32).

For both species the continental range predicted by the model most closely resembled the actual continental range when the Irish presence data was excluded from the model.

2.3.6.2 Hindcasting to the Last Glacial Maximum

The models run for all the past and future projections included all points, and were based on climate layers alone, as soil data were not available for these times. The source of the LGM climate data is the Paleoclimate Modelling Intercomparison Project Phase II (http://pmip2.lsce.ipsl.fr/), based on the Community Climate System Model version 3 (Collins *et al.* 2006).

Projecting the MaxEnt models constructed based on current day conditions onto the landscape of the LGM yielded the following maps of suitable areas. Figure 33 describes the potential distribution areas at the Last Glacial Maximum (around 21, 000 years ago).

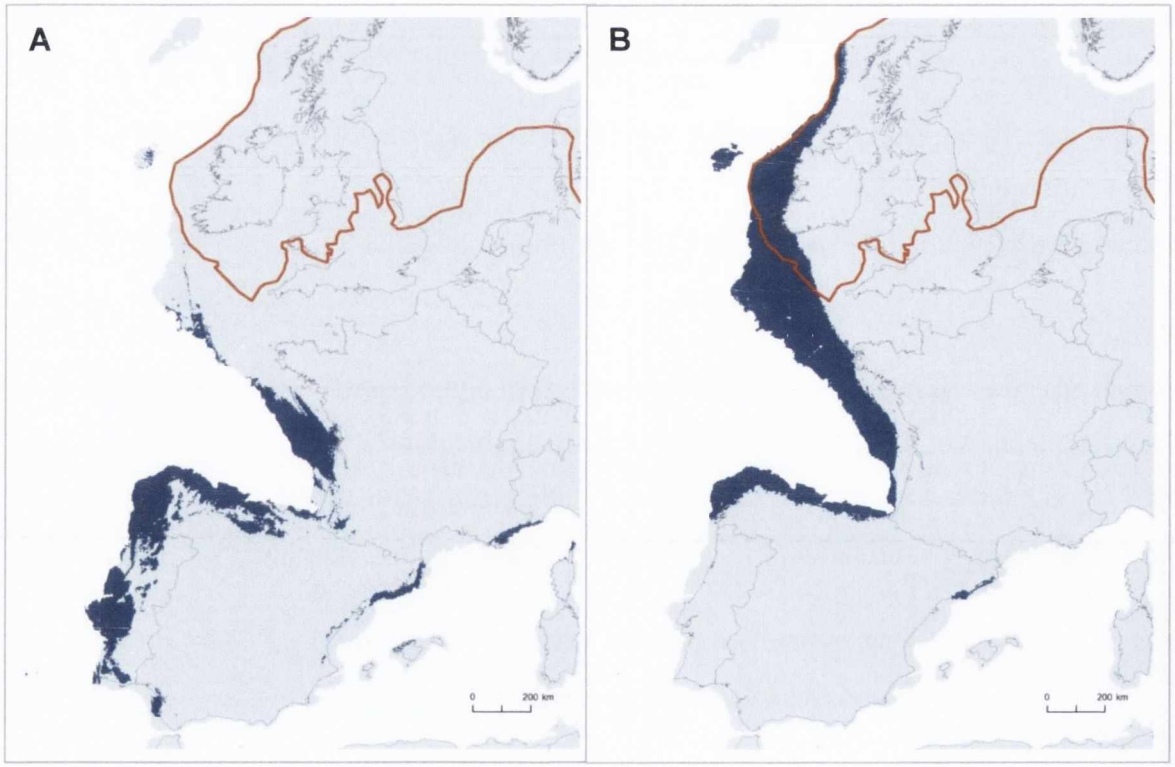


Figure 33: Predicted potential range of (A) *D. cantabrica* and (B) *E. mackaiana* during the LGM. The shaded areas delimit the LGM coastline, and current day coastlines and country borders are overlaid for reference. The glacial extent (orange) is from Clarke *et al.* (2012).

The predicted range was very different between the two species. For *D. cantabrica* the predicted range extended much further south than the current distribution, as far as the south-western tip of the Iberian peninsula. To the north, the predicted range at the LGM included much of the then coastline offshore what is now France, including a small area to the south of Ireland. However no areas in Ireland or Britain were predicted as suitable.

For *E. mackaiana* the predicted range at the LGM was a clear continuous coastal strip stretching from the present day distribution in northern Spain all the way up the Atlantic coastline to include the areas west of Ireland, and as far north as the area off the north coast of Scotland.

In order to examine the case for potential glacial refugia in or near Ireland for both species, this area was looked at in more detail (Figure 34).

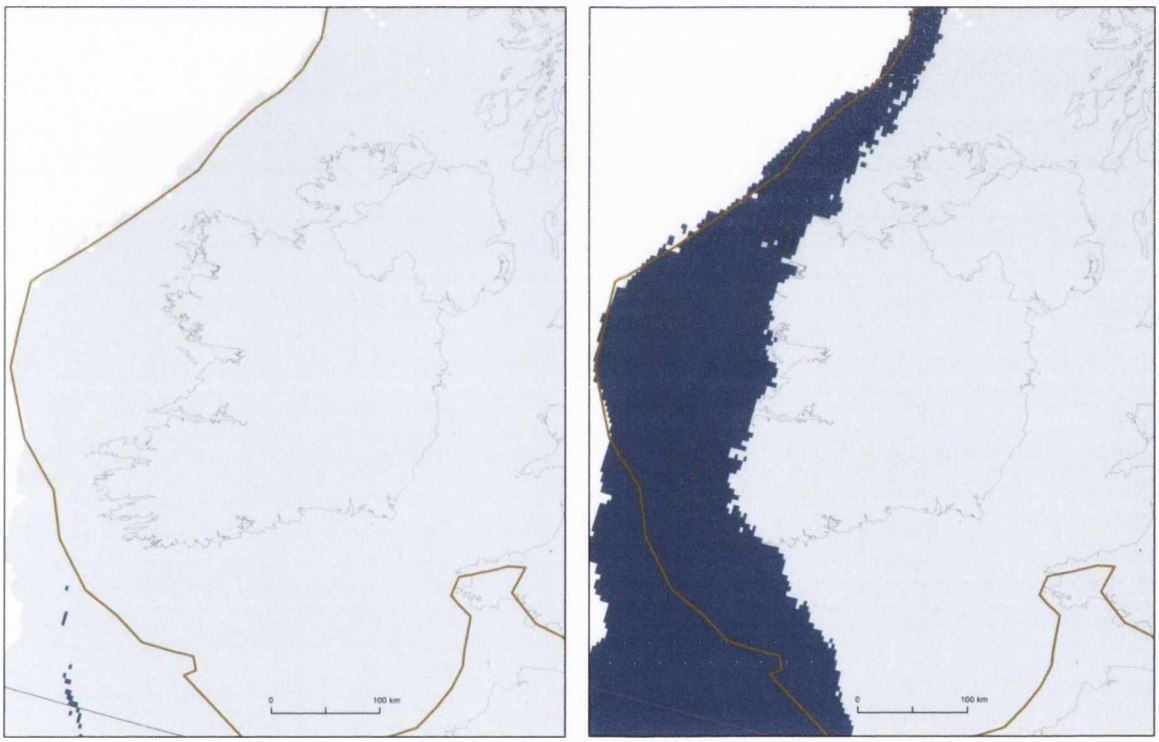


Figure 34: Close up of potential distributions of (A) *D. cantabrica* and (B) *E. mackaiana* in and around Ireland at LGM, glacial extent (orange line) and current coastlines also shown.

The glacial extent of Clark *et al* (2012) precludes any possibility of an *in-situ* refugium in any part of the current island of Ireland, as it extends well beyond the present-day coastlines. The nearest significant area of suitable conditions for *D. cantabrica* is to the south-west, off present-day Kerry. For *E. mackaiana*, a great deal more suitable area was predicted near and in parts of the west coast of Ireland (Figure 34B).

For *E. mackaiana* there would appear to be a much larger area of suitable conditions unglaciated, stretching across a large part of Kerry to off the current coast of Donegal. Current locations of *E. mackaiana* which were unglaciated and predicted as suitable are the Kerry, Galway and Mayo populations, although all three occur in extremely close proximity to the glacial extent. The location of the current Donegal population was completely glaciated.

2.3.6.3 Projections to the Last Interglacial

The potential range predicted by the model at the time of the last interglacial period (120, 000 years ago) is shown in Figure 35.

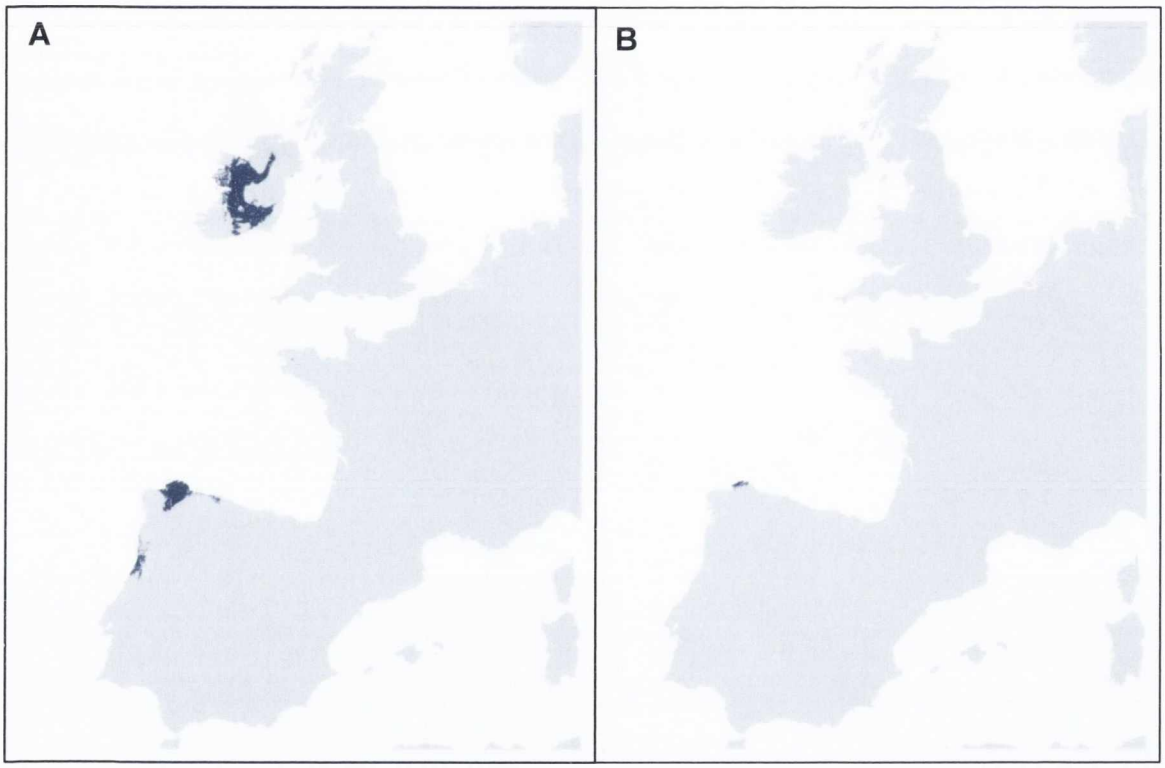


Figure 35: Range projections for *D. cantabrica* (A) and *E. mackaiana* (B) at the Last Interglacial.

D. cantabrica shows a much reduced potential range compared to today, mostly in central Ireland, the Galician coast, and the northwest coast of Portugal. The continental areas show an overlap with the current range but the Irish predicted range is very different.

The prediction showed a small area of suitability for *E. mackaiana* during the Last Interglacial in the north-eastern corner of the Iberian peninsula.

Unfortunately there exists no fossil evidence reliably dated to the Last Interglacial in Ireland to confirm whether these predictions are accurate. However two sites from the Gortian Interglacial (i.e. the penultimate interglacial) are overlaid on the projection and show a close concordance with the predicted suitable area (Figure 36).

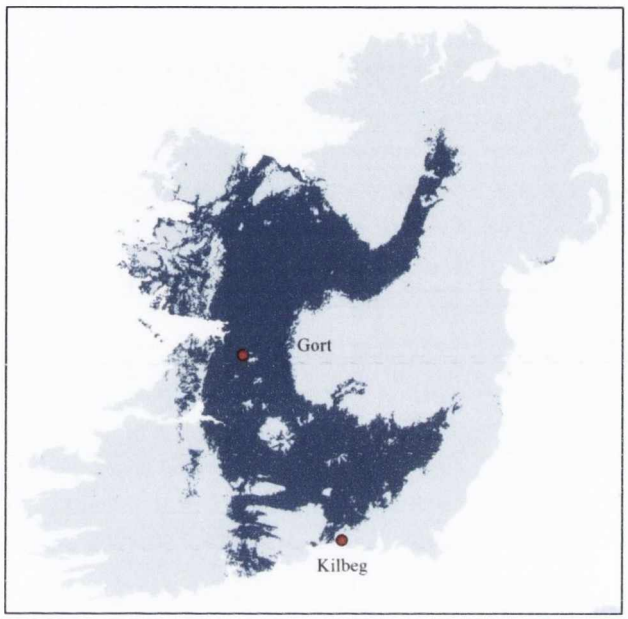


Figure 36: Darker areas correspond to the predicted potential range of *D. cantabrica* during the Last Interglacial (120,000 years ago). Two sites containing Gortian interglacial fossil evidence of *D. cantabrica* are shown in red.

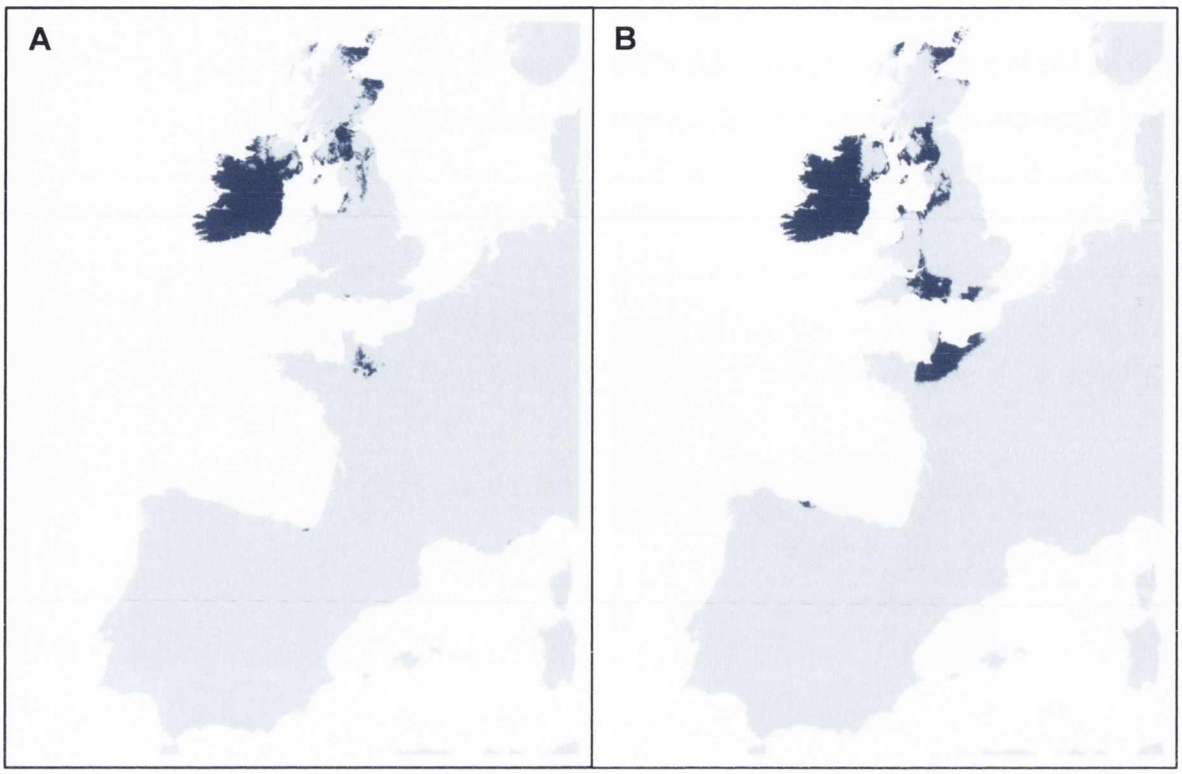
2.3.6.4 Forecasting the effects of future climate change

Although not relevant to the origins of the study species, it was considered important to also look at how future climate change could potentially alter the distribution of *D. cantabrica* and *E. mackaiana* in continental Europe and in Ireland.

Figure 37 shows the predicted potential range of the D. cantabrica and E. mackaiana approximately 70 years from now, under two different emission scenarios; SRES-A1b (the more drastic scenario), and SRES-B2a (the more conservative scenario).

Both emissions scenarios show similar patterns for the two species; a drastic reduction in their current continental range and an increase in Irish range. Under SRES-A1b there are also substantial areas in the north of Britain predicted as becoming suitable for the growth of both of species.

Climate change scenario: SRES-A1b



Climate change scenario: SRES-B2a

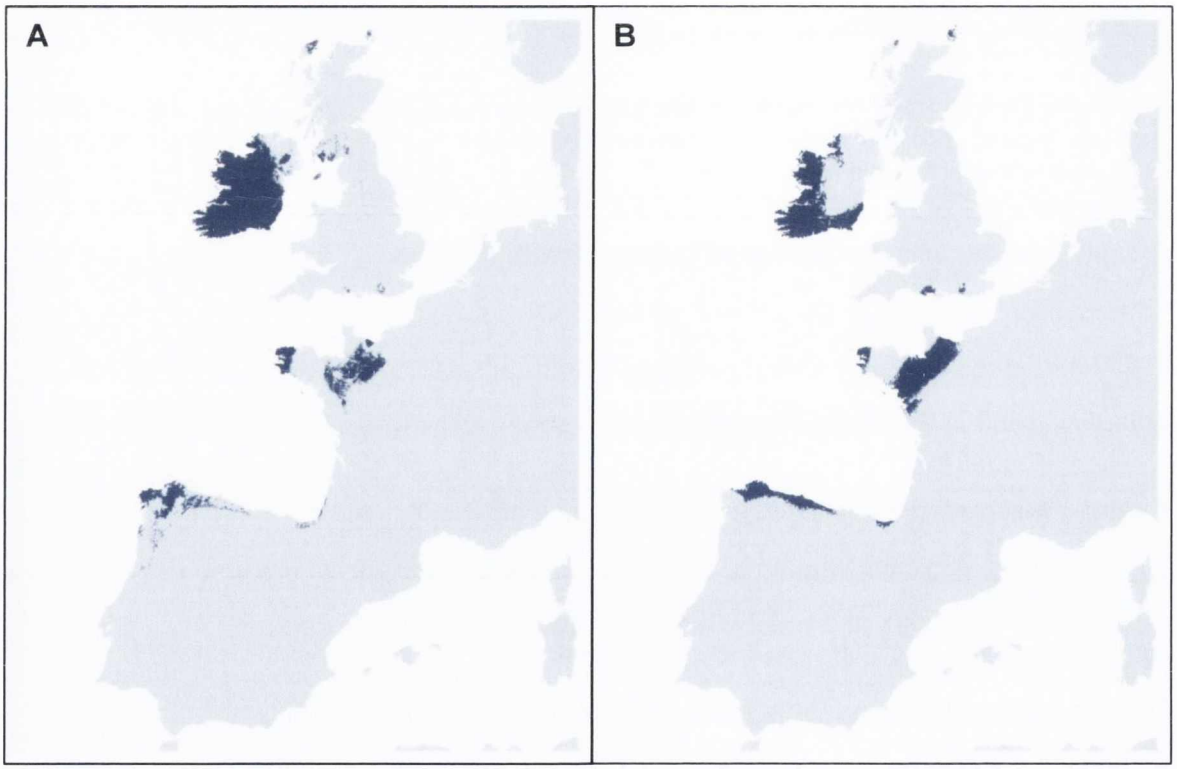


Figure 37: Darker areas correspond to the predicted potential range of (A) *D. cantabrica* and (B) *E. mackaiana* projected into the future to the decade of the 2080s, under two different emissions scenarios; SRES-A1b and SRES-B2a.

2.4 Discussion

2.4.1 Insights into climatic ranges tolerated by the two species

The results of the PCA clearly demonstrate that despite the similarities in climate which may exist between the continental and Irish locations of the study species there is in fact a quantifiable difference in the climatic envelope between the two areas.

It was also shown that the climatic ranges tolerated by *D. cantabrica* are greater than those tolerated by *E. mackaiana*, as evidenced by its larger geographic range.

2.4.2 Continental range over-predictions

For both species the models predicted as suitable areas in which they are currently not found. The extent of this over-prediction varied depending on whether the model was based only on the continental location points, or on the Irish ones as well. I begin by discussing the implications of an over-prediction when based solely on the continental locations, and then move on to discussing the implications of the even greater over-prediction when using Irish locations.

When based only on the continental points the model over-predicted eastwards for *E. mackaiana*, approximately 200 km beyond its more easterly location point. For *D. cantabrica* the model over-predicted northwards up the west coast of France.

This indicates that both species are not completely filling the range potentially available to them due to (i) certain limiting factor(s) not accounted for in the model, or (ii) an inability to disperse effectively from their current range.

Interestingly the model run only with soil data, whilst extensively over-predicting did suggest a possible barrier to explain the restriction of *D. cantabrica*'s French range to the southern corner of the French coast. The model showed a large region of unsuitable habitat just to the north of the current range edge near Biarritz.

This area corresponds to what is known as 'les sables des Landes', or 'les landes de Gascogne' ('landes' means heathland or moor in French, and Landes is also the name of the French Département covering this area).

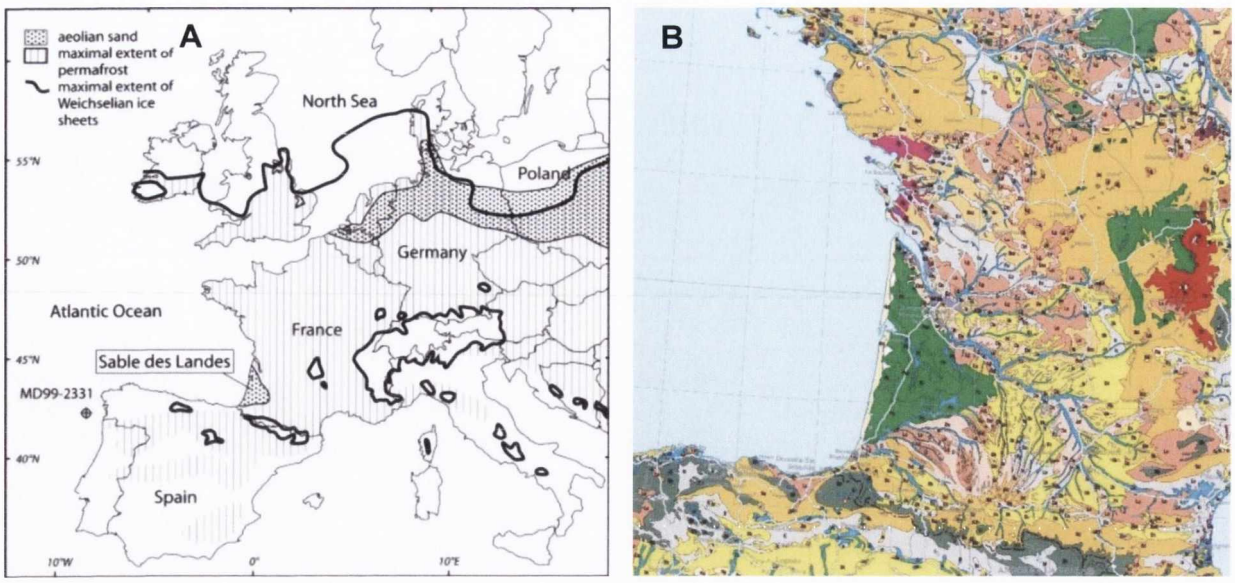


Figure 38: (A) Les sables de Landes, reproduced from (Bertran *et al.* 2009). (B) Close-up of the region showing soil types, reproduced from Jones *et al.* (2005).

This region was formed by aeolian coversand in the late glacial (Bertran *et al.* 2009). There is possibly some factor associated with this highly distinct region that is acting as a barrier to dispersal or establishment further up the west coast of France.

I return now to discussing the even greater over-prediction made when Irish locations are included in the model. The over-prediction is not unexpected, since the earlier PCA analysis made it clear that the Irish locations experienced climatic conditions outside of the ranges of the continental conditions, and it would therefore be logical for the predicted continental range to be expanded.

However, similar patterns of over-prediction of native range have been demonstrated in other species when location data of populations known to be introductions are included in the distribution model (Beerling *et al.* 1995, Fitzpatrick *et al.* 2007). Therefore the results in this study do raise the question of whether the Irish locations should be regarded as an 'introduced' population, in contrast to the 'native' population of the continent. Human introduction is already postulated as a possible origin of the Irish populations, and these results would appear to give some weight to the idea.

2.4.3 Irish range predictions

When based purely on continental locations the model failed to predict any of the areas currently inhabited by either species in Ireland. In fact it predicted a complete absence of *E. mackaiana* from Ireland, and only a small area of Waterford as being suitable for *D. cantabrica*. This supports the findings revealed by the PCA; that the niche described by

the Irish locations is substantially different to the niche described by the continental locations.

The Irish populations of both species therefore demonstrate a succesful adaptation to the Irish climate. Adaptation to local growing conditions across broad latitudinal ranges has been studied in other species (Scots pine: Eriksson *et al.* 1980, aspen: Hall *et al.* 2007, poplar: Keller *et al.* 2011).

When the model was run using all locations, including the Irish ones, there was then a considerable over-prediction within Ireland for both species. The entire western third of the country was predicted as suitable, as well as large areas of the south for *D. cantabrica*. Wexford was also predicted as suitable for both species.

This indicates that the current Irish range of *D. cantabrica* and *E. mackaiana* is constrained either by an unknown factor not included in the model, or an inability to disperse.

For *E. mackaiana* the obvious reason for its lack of range expanson in Ireland is the fact that it does not set seed, and vegetative propagation does not allow for dispersal over any great distance. Figure 32 shows that a model based only on continental points does not predict any suitable habitat in Ireland. The lack of seed set in *E. mackaiana* may thus be a result of climatic factors.

Similarly to the continent, even though it did not show in the combined model, the soil factors alone indicate a possible reason why *D. cantabrica* is restricted to its current position, supporting earlier statements (Woodell 1958) attributing its eastern limits to limestone barriers.

Landuse is another factor, which although not included in the model could also be an important factor in limiting range expansion for both these species.

2.4.4 Range in Britain

Models based on only on continental location points do not predict the presence of either *D. cantabrica* or *E. mackaiana* anywhere in Britain, echoing the current situation of an absence of any known native populations of either species. When Irish locations are added to the models of both species, a few isolated areas are predicted as suitable. In the case of *D. cantabrica* some of these predicted areas correspond with a few of the naturalised populations that are found in Britain.

If we consider the Irish populations as an introduction then the natural absence of these species from Britain is easily attributable to the lack of suitable habitats (based on a purely 'continental-type' climatic niche). The Irish populations, despite showing an ability to establish in and adapt to a different type of climate, apparently lack the dispersal capabilities required to extend further into Ireland, or cross the Irish Sea to Britain.

2.4.5 Irish refugium

The range predictions to LGM show clearly that both species could have survived in areas further to the north than they are currently distributed. The modelling results for the LGM hindcasting, particularly for *E. mackaiana*, suggest that the range of both species was at one time less disjunct than it is now, and that some populations extended as far north as the unglaciated areas near the current Irish west coast.

It is important to note however that the LGM projections do not include the soil variables in the model, which may be very important in restricting range and acting as a barrier to dispersal.

The scenario of an proximal or in-situ refuge to explain the origin of the Irish populations hinges primarily on delimiting the extent of the the British-Irish ice sheet. Recent work by Clark *et al* (2012) proposes a much larger area of glaciation and effectively discounts hitherto existing possibilities of unglaciated areas in the west and southwest of Ireland. Based on this, the hypothesis of an Irish refugium for these and indeed for any other species is ruled out.

2.5 Long distance dispersal in postglacial times

If considering natural long distance dispersal events as a possibility for the establishment for the current Irish populations there is presumption of stochasticity and therefore many other suitable locations in Ireland in which these dispersals could have occurred to. Even if only considering a strictly 'continental' niche, there are locations in the north and west of France that are suitable (in terms of both climate and soil) which could potentially support the establishment of new populations.

Many of the natural dispersal mechanisms which explain long distance dispersal events have been deemed highly unlikely; there are few migratory birds to Ireland from Iberia (Reid 1913) and those that do are generally not seed eaters (Foss and Doyle 1988). Ocean currents do not flow in the right direction (Forbes 1846), and even if they did the seeds would not remain floating for long enough (Praeger 1913), nor are they adapted for wind

dispersal (Praeger 1939). Therefore we can assume that any dispersal event that occurred was just as likely to establish in any of the other areas of Ireland, Britain or northwest France predicted as suitable by the models. The fact that for both species (and for other Hiberno-Iberian species) there is a pattern of preference for the west coast suggest that there is a less stochastic mechanism operating, such as the possibility of human introduction and subsequent naturalisation.

2.5.1 Climate change consequences

In terms of the Irish range of both species, neither would appear to be under any threat from the effects of climate change. For both species their current locations continue to be predicted as suitable under both a conservative and an extreme climate change scenario.

The situation for the continental ranges however is more worrying, as for both species there are results ndicating a significant reduction in range in the future. There may already be evidence of such effects; a recent study of seed production in *D. cantabrica* at high altitudes has demonstrated significant reductions in the years with hot summers and low precipitation (Vera 2011).

CHAPTER 3 Comparative DNA Sequencing and Haplotype Analysis

3.1 Introduction

Prior to the advent of molecular biology techniques systematists relied on comparative morphology, physiology and other measurable phenotypic characteristics to make phylogenetic inferences on the relationships between groups of organisms.

Variation in phenotype, however, is a combined result not only of genetic variation but also of environmental influences (Beebee and Rowe 2004). The ability of an organism to express different phenotypes depending on the biotic or abiotic environment is known as phenotypic plasticity (Agrawal 2001).

In order to infer relationships between groups of organisms, only those traits which are genetically transmitted are informative (Avise 1994). Because of phenotypic plasticity, the measurement of morphological variation alone is not an accurate reflection of genetic variation as it can result in an overestimation (Freeland 2005).

The direct quantification of genetic variation has only become possible due to the advances in molecular biology that enabled the development and use of molecular markers (Beebee and Rowe 2004).

3.1.1 Molecular markers

Molecular markers are sections of the genome selected for analysis on the basis that they contain readily identifiable alternative sequences, known as alleles (Beebee and Rowe 2004). The number of alleles and their repartition among different individuals is then investigated to inform on genetic variability and relationships within a species.

3.1.2 Comparative DNA sequencing

The nucleotide sequence contains within it the record of evolutionary relationships to the genomes of other individuals (Avise 1994). Comparative DNA sequencing, where the actual nucleotide sequence along a particular length of DNA is compared between individuals is one way to examine this evolutionary history.

3.1.2.1 Nuclear, mitochondrial and chloroplast DNA

Plants have three genomes, the nuclear genome, the mitochondrial genome and the chloroplast genome. These three genomes differ in terms of their size and mode of inheritance (Petit *et al.* 2005).

The plant nuclear genome consists of linear DNA organised in the form of discrete chromosomes, varying in number from n=2 to as high as n=600 (Heslop-Harrison and Schwarzacher 2011).

All angiosperm chloroplast genomes are circular molecules, varying in size in a very narrow range of between 135kb and 160kb (Palmer and Stein 1986).

In angiosperms chloroplast and mitochondrial DNA is typically (but not exclusively) maternally inherited (Harris and Ingram 1991).

For phylogenetic analysis in plants the mitochondrial genome has historically been used much less frequently, as it is highly rearranged and has a slower rate of nucleotide substitution than the chloroplast or nuclear genomes (Bakker *et al.* 2006, Seberg and Petersen 2006). However, some recent studies have demonstrated successful applications of mitochondrial DNA analysis in phylogenetic analysis (e.g. Sherwood 2008, Kyoda and Setoguchi 2010).

Early phylogenetic studies used almost exclusively chloroplast markers, until it became evident that phylogenetic inference was affected by the use of uniparentally inherited loci (Alvarez and Wendel 2003), and nuclear markers became commonly used in conjunction with the plastid markers.

When selecting markers for phylogenetic analysis it is important to select those with rates of evolution appropriate for the level of analysis (Soltis and Soltis 1992). Markers with very low rates of evolution will not be suitable for infraspecific analysis, as they are unlikely to show enough variation. Likewise markers with very high mutation rates may not be suitable for inferring phylogenies at the genus or family level because of homoplasy (Soltis and Soltis 1992).

Chloroplast genomes have approximately half the rate of evolution of nuclear genomes (Wolfe *et al.* 1987), however the rate of evolution at different areas within the chloroplast genome can vary widely. Non-coding regions of the chloroplast genome have been shown

to have much higher rates of evolution and are often used in phylogeographic analysis at low taxonomic levels (Diekmann *et al.* 2012).

This study focussed on infraspecific variation and therefore markers were selected from non-coding regions of both chloroplast and nuclear DNA, which had high rates of mutation. Markers that had already been used in the family Ericaceae were also considered when selecting markers. The selected markers are outlined in the Methodology (page xx).

3.1.3 Haplotypes and networks

The term haplotype refers to the unique combination of nucleotides within one or more sequences of DNA that contain polymorphisms. Individuals of the same haplotype will have identical bases at all of the polymorphic sites, as well as across the portions of the sequence that do not vary. A sequence that differs at even a single nucleotide will constitute a different haplotype.

Whilst the genetic relationships between different species are often represented by bifurcating trees, the genetic relationships between haplotypes within a species are more appropriately represented by multifurcating networks (Schaal *et al.* 2003). This is because ancestral haplotypes often coexist with current ones, and recombination produce reticulate relationships between haplotypes (Posada and Crandall 2001).

3.1.4 Molecular dating

The idea that molecular evolution occurs at an approximately uniform rate over time (known as the molecular clock hypothesis) was formulated in the early 1960s (Kumar 2005). At this time the molecules in question were proteins, and it was not until PCR methods became widespread in the 1970s that the molecular clock hypothesis became applied to DNA sequences.

Since that time sequence analysis and fossil calibration have determined that molecular evolution rates vary significantly between different species, especially when life histories are different, for example generation time (Lanfear *et al.* 2010). Whilst this makes it difficult to assess divergence times for groups of plants that have no fossil calibration there do exist in the literature instances of using published rates, based on similarity in life history or phylogenetic relatedness.

For non-coding chloroplast sequences an average rate of 5.0 x 10^{-10} substitutions/site/year, with a standard deviation 10^{-10}, has been proposed as a rate that would cover the variation across species (Palmer 1991). This rate has been used in phylogenetic studies in various families (Huttunen *et al.* 2008, Aigoin *et al.* 2009), including the Ericaceae (Désamor *et al.* 2011).

For ITS a review of rates across different genera and families revealed a range of rates differing by at least an order of magnitude (Kay *et al.* 2006). There were striking differences in the rates of ITS evolution between life history categories; higher rates being recorded in annuals compared to perennials and in herbaceous species compared to woody species. The rate for woody perennials ranged from 0.38 × 10^{-9} to 7.83 × 10^{-9} substitutions/site/year (mean = 2.15 × 10^{-9} subs/site/yr) (Kay *et al.* 2006).

3.1.5 Hybrid investigation

Hybrid organisms are those that result from the interbreeding of two separate species, and is now known to be widespread in many different taxonomic, groups including over 23, 000 species of plants (Freeland 2005).

E. x *stuartii* was first described as a hybrid in 1902 (Linton 1902) although the parentage presumed at that time was *E. mackaiana* x *E. meditteranea.* Before that it had been described variously as either a different form or a subspecies of *E. mackaiana* (Nelson 1995). It was not until 1977 that an accidental reversion in cultivation identified its correct parentage as being of *E. mackaiana* and *E. tetralix* (McClintock 1979).

Since a few samples of *E.* x *stuartii* were also collected as part of this study, the opportunity was taken to confirm their parentage by DNA sequencing.

3.1.6 DNA sequence analysis and infraspecific phylogeography

DNA sequencing has become one of the standard methodologies in plant phylogeography and has been used to investigate the genetic relationships of present-day populations of various species in order to elucidate the historical mechanisms that have led to their current distribution.

Two ways in which sequence data can be used to identify whether disjunct populations are the result of vicariance or of dispersal events are, firstly, with the use of a molecular clock

and, secondly, through the examination of the branching order of haplotype genealogies (Freeland 2005).

A molecular clock can be used to determine timing since the divergence of two haplotypes. We can then see if this timescale corresponds to (or precludes) a known or hypothesised vicariance or dispersal event. If investigating whether dispersal was by human means for example, the estimate of divergence times would have to fall within the timescale of known human travel between the disjunction. Molecular clock methods have in some cases identified evidence of human introduction of populations once thought to be native. The reverse is also true, molecular clocks have been used in studies that revealed the native status of disjunct populations previously thought to be introduced (Martin-Bravo *et al.* 2009).

By analysing the branching order of the gene genealogies (haplotypes) we can compare the evolutionary relationships of populations to their geographical distributions, and in this way determine relative importance of past dispersal versus vicariance events (Freeland 2005).

DNA sequencing has been used in the identification of potential refugial areas, in the tracking of routes of postglacial colonisation (Petit *et al.* 2002, Petit *et al.* 2003, Beatty and Provan 2011), and in the dating of range fragmentations (Schneeweiss and Schonswetter 2010).

DNA sequencing analysis has been used extensively to elucidate the floral and faunal postglacial colonisation of western Europe, and many of these studies have specifically investigated the origins of the Irish flora and fauna, identifying evidence for a variety of mechanisms including post or late-glacial colonisation from continental Europe (oaks: Kelleher *et al.* 2004, stoats: Martinkova *et al.* 2007), persistence in glacial refugia (Scots pine: Sinclair *et al.* 1998, red seaweed: Provan *et al.* 2005) and human introduction (pygym shrews: Mascheretti *et al.* 2003).

DNA sequence analysis has not yet been applied in phylogeographic analysis of any other members of the Hiberno-Iberian flora specifically, but sequence analysis (of ITS) has been utilised in a study on the natterjack toad, which occurs in southwestern Ireland and Iberia, and also in two disjunct regions of England (the northwest coast and the southeast), as well as in much of north-central Europe (Rowe *et al.* 2006). In this case the authors concluded that the toad survived north of the Iberian peninsula throughout the LGM, followed local

population expansions in the early postglacial into Britain and Ireland, with the Irish populations being most closely related to those in northern Britain.

Whilst sequencing offers the best way of establishing the genealogy of a gene lineage, the often limited amounts of variation found at the infraspecific level mean that studies are often carried out in conjunction with other forms of molecular markers such as microsatellites or Amplified Fragment Length Polymorphism (AFLPs). Whilst the focus of this chapter is on sequence analysis, the following chapter focuses on AFLPs.

3.1.7 Aims

The aims of this chapter are to;

1. Identify haplotypes for the cpDNA and rDNA (ITS) markers in individuals from across the range of each of the study species;
2. Examine the geographical distribution of these haplotypes;
3. Estimate a time-scale for divergence of haplotypes based on the amount of variation between them.
4. To confirm the parentage of *Erica* x *stuartii* as being of *E. tetralix* x *E. mackaiana*.

3.2 Methodology

3.2.1 Sampling

Table 12 outlines the number of samples sequenced from each of the populations sampled, for the different markers. The location map of these populations can be found in Figure 16 and Figure 17 on pages 43 and 44.

Table 12: Location of populations and number of samples used for DNA sequencing of the ITS and cpDNA markers.

COUNTRY	REGION	Pop Code	Locality	Total samples	No. of samples used in:	
					cpDNA	ITS
Daboecia cantabrica (including subsp. *azorica*)						
France	France - Loire	Q	St-Cyr-des-Gats	30	4	29
		R	Brebaudet	10	4	11
	Southern France	S	La Rhune	10	1	9
		T	Col d'Ibardin	10	2	10
		U	Pas de Roland	10	2	10
		V	Bidarray	30	1	27
Spain	NE Spain	A	Urkiola National Park	10	2	9
		B	Cabo Major	10	2	10
		C	Val de Aso	10	3	10
		D	Val de Aso	10	1	8
		E	Cabo Penas	10	1	10
		F	San Isidro Pass	10	2	9
		G	Mt Torneo	10	3	9
	NW Spain/ Portugal	H	Serra de Capelada	10	-	8
		I	Sierra de Meira	10	1	10
		J	Cabo Fisterra	10	-	10
Portugal		K	Vigo	10	1	7
		L	Peneda Geres	10	1	9
		M	Peneda Geres	10	1	-
		N	Neiva Catelo	10	2	2
	Azores	O	Faial	10	3	10
		P	Pico	12	3	12
Ireland	Ireland	W	Carraroe	10	1	10
		X	Derryneen	10	1	10
		Y	Clifden	30	4	30
		Z	Errisbeg	10	2	10
			Totals:	322 (26 pops)	48 (24 pops)	289 (25 pops)
Erica mackaiana						
Spain	Spain	A	Cabo Penas	10	3	-
		B	Sierra de Loba	10	2	9
		C	As Pontes de GR	10	1	7
		D	Sierra da Capelada	10	1	2
		E	Sierra de Meira	10	2	2
		F	Foz	10	1	2
Ireland	Ireland	G	Roundstone 1	10	2	2
		H	Roundstone 2	30	4	5
		I	Lough Nacung	26	4	4
		J	Knockeneden Cross	8	5	5
		K	Bellacorick, Mayo	13	2	2
			Totals:	147 (11pops)	27 (11pops)	40 (10 pops)

3.2.2 DNA Extraction

Leaf tissue samples were collected together with herbarium vouchers from the localities described previously in Chapter 2. From each sampled plant one or two branchlets were

stripped of their leaves, and the leaves placed immediately in silica gel, as recommended by Chase & Hills (1991).

Approximately 15-20mg of leaf tissue was used from each silica-dried sample for the DNA extraction. Leaf tissue was placed in a 2ml Eppendorf tube with a carbide bead and ground using a mixer mill for 2 minutes. Total DNA extraction was carried out using DNeasy Plant Mini kits (Qiagen), and the accompanying buffers, according to the manufacturer's protocol, namely;

Cell Lysis: 400µl of Buffer AP1 and 4µl of RNase stock solution were added to each sample, vortexed vigorously then incubated for 10min at 65C, mixing 3 times by inversion during incubation.

Precipitation: 130µl of Buffer AP2 was added to each sample, vortexed and incubated for 5min on ice, then centrifuged at 14000rpm for 5min.

Lysate filtration: the supernatant was then poured into a spin column and centrifuged at 14000rpm for 2min. The flow-through was transferred to a 2ml Eppendorf tube, taking care not to disturb the pellet.

Binding: 67µl (or1.5x volume of flow-through) of Buffer AP3was then added and mixed in, then the solution was added in two lots to a spin column, spun at 8000rpm for a minute each time, and the flow-through discarded.

Washing: 500µl of Buffer AW was added to the spin column, centrifuged at 8000rpm for 1 minute and the flow-through discarded. This step was then repeated.

Elution: 100µl of Buffer AE was then added to the spin column, inserted in a fresh Eppendorf as the final collection tube, then incubated for 5min at room temperature and then centrifuged at 8000rpm for 1min. This step was repeated, with a further 100µl of AE, eluting into the same collection tube.

3.2.3 Amplification of markers

3.2.3.1 Markers selected for DNA sequence analysis

The three markers selected for comparative sequencing analysis were the two chloroplast markers; the *trnH-psbA* spacer and the *trnL-F* spacer, and the Internal Transcribed Spacer (ITS).

trnH-psbA spacer

trnH and *psbA* are two chloroplast genes found on the Large Single Copy (LSC) region of the chloroplast genome, adjacent to Inverted Repeat A (IRA). The *psbA* gene codes for the D1 polypeptide of the photosystem II reaction centre complex. The *trnH* gene codes for tRNA-GUG, which is the code for Valine. The non-coding space between these two genes is known as the *trnH-psbA* spacer.

The *trnH-psbA* spacer is a commonly used locus in phylogenetic studies, and has also been suggested as an alternative barcode to *matK* and *rbcL* (Whitlock *et al.* 2010). The primers used to amplify the spacer region for this study were *psbA* (Sang *et al.* 1997) and *trnH* (Tate and Simpson 2003).

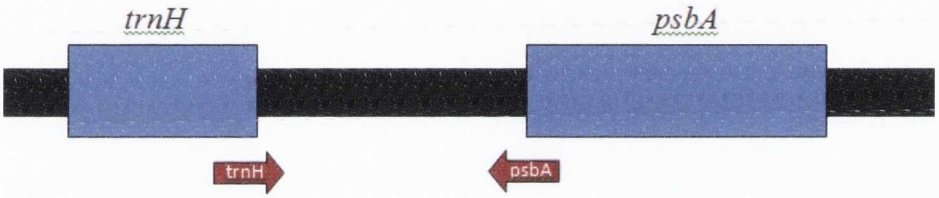


Figure 39: The *trnH-psbA* intergenic spacer and associated primers

trnL-trnF Intergenic Spacer

Two other adjacent genes on the chloroplast LSC are *trnL* and *trnF*, which code for tRNA-UAA (start codon) and tRNA-GAA (glutamic acid) respectively. The *trnL* gene is divided by a Group I intron known as the *trnL* intron. Between *trnL* and *trnF* lies a non-coding region known as the trnL-trnF intergenic spacer.

The primers used to amplify this spacer were the universal primers e and f developed by Taberlet (1991).

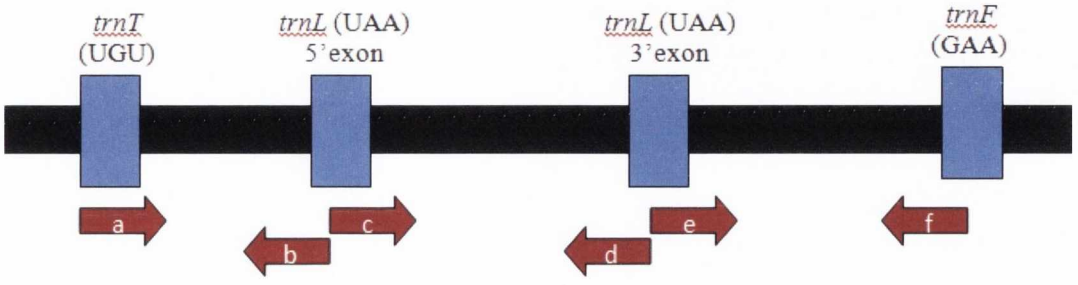


Figure 40: The *trnL-trnF* gene regions, and Taberlet's universal primers a-f used to amplify different regions.

The Internal Transcribed Spacer (ITS)

The Internal Transcribed Spacer (ITS) refers to two non-coding regions (ITS-1 and ITS-2) situated in between coding regions of DNA that code for the components of ribosomal RNA (rRNA). In plants the four coding regions that code for the various components of rRNA are 18S, 5.8S, 28S and 5S. ITS1 is situated between 18S and 5.8S, and ITS 2 is located between 5.8S and 28S (Figure 41).

Typically the entire ITS sequence of both ITS1 and ITS2 and the coding region of 5.8S between them is sequenced as a whole. The 5.8S region in the middle of the sequence is much more highly conserved than the adjacent ITS regions.

The primers used to amplify the ITS for this study were AB101 and AB102 (Sun *et al.* 1994).

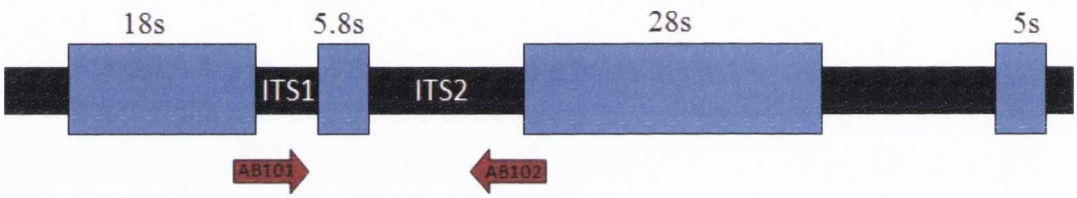


Figure 41: The Internal Transcribed Spacer (ITS) and the location of the primers AB101 and AB102

ITS is one of the most widely used markers in plant phylogenetic and phylogeographic analysis, due primarily to its high copy number, fast rate of evolution and its biparental inheritance (Calonje *et al.* 2009). Although it is most commonly used at the interspecific and intergeneric level, its use at the infraspecific level has become more common in recent years and its utility has been demonstrated across a wide variety of species and genera (Holderegger and Abbott 2003, Barker *et al.* 2005, Kawase *et al.* 2007, Huttunen *et al.* 2008, Singh *et al.* 2008, Beatty and Provan 2011, Douglas *et al.* 2011, Queiroz *et al.* 2011, Hedren *et al.* 2012).

Markers were amplified using Polymerase Chain Reaction (PCR). PCR reactions were carried out in a total reaction volume of 25µl, comprised of 15.625µl of ultrapure water, 5µl of 5x Colourless GoTaq Flexi Buffer (at 5 units μl^{-1}, Promega), 0.5µl of 10mM dNTPs, 0.25µl of each primer at 20pmol μl^{-1}, 2µl of 25mM $MgCl_2$, 0.125µl of GoTaq DNA

Polymerase (Promega) and 1.25µl of total DNA (approximately 100ngµl^{-1}) as the template.

PCR reactions were carried out with the thermocycler T-Professional (Biometra) with the following program: an initial denaturation at 94°C for 90 seconds followed by 28 cycles of denaturation at 94°C for 30 seconds, annealing at 52-62 °C for 1 minute (see Table 13 for primer-specific annealing temperatures) and extension at 72°C for 1 minute. A final extension phase was run at 72°C for 7 minutes, before storage at 4°C.

3.2.3.2 Primers

The primer pairs used for amplification of each region are described in Table 13.

Table 13: Primer pairs and annealing temperatures for the three markers sequenced

Marker	Primer	Sequence	Annealing temp (°C)	Reference
ITS	AB101: AB102:	5'-ACGAATTCATGGTCCGGTGAAGTGTTC-3' 5'-TAGAATTCCCCGGTTCGCTCGCCGTTA-3'	62	(Douzery *et al.* 1999)
trnL-trnF spacer	e: f:	5'-GGTTCAAGTCCCTCTATC-3' 5'-ATTTGAACTGGTGACACGAG-3'	62	(Taberlet *et al.* 1991).
trnH-psbA spacer	trnH: psbA:	5'-CGCGCATGGTGGATTCACAATCC-3' 5'-GTWATGCAYGAACGTAATGCTC-3'	52	(Sang *et al.* 1997) (Tate and Simpson 2003)

3.2.3.3 PCR purification

Purification of the PCR products to eliminate excess primer was carried out using one of two methods.

For some samples the JETquick PCR purification spin column kit (Genomed) was used, according to the manufacturer's protocol. For the remainder of the samples a cleanup protocol using Exonuclease and Shrimp Alkaline Phosphatase was used, commonly called ExoSAP.

For the ExoSAP purification the total volume of each reaction was 10µl, comprised of 5µl of PCR product, 0.3µl of Exonuclease (20U/µl), 2µl of Shrimp Alkaline Phosphatase (1U/µl) and 2.7µl of water. The reaction was carried out with a thermocycler programme of 37°C for 30minutes, followed by 82°C for 20 minutes. Purified PCR products were stored at 4°C until use (or at -20°C for long-term storage).

3.2.3.4 Cycle Sequencing

Purified PCR products then underwent a cycle sequencing reaction using a reaction mix of 5.5µl ultrapure water, 1µl of BigDye Terminator version 3.1 (Applied Biosystems), 1.5µl of accompanying buffer (5x, Applied Biosystems), 0.5µl of primer (20pmol/µl) and 1.5µl purified PCR product as the template (total reaction volume of 10µl). Two reactions were performed for each sample, one using the forward primer and the other using the reverse primer.

The cycle sequencing thermocycler program was as follows; an initial denaturation step at 96°C for 1 minute followed by 28 cycles of denaturation at 96°C for 10 seconds, annealing at 50°C for 15 seconds and extension at 76°C for 4 minutes. The cycle sequencing products were then stored at 4°C.

The sequencing reaction products were then cleaned using the Applied Biosystems Big Dye X-Terminator™ purification protocol. 45µl of SAM™ solution and 10µl of X-Terminator™ solution were added to each cycle sequencing reaction. The plate was then vortexed at maximum speed for 30minutes, and then centrifuged to eliminate air bubbles in the samples before loading into the sequencer.

3.2.3.5 Sequencing

Automated sequencing was carried out using the Applied Biosystems 3130xl Genetic Analyser.

3.2.3.6 Sequence Assembly and Alignment

The chromatograms from the forward and reverse sequencing reactions were assembled and aligned with the commercially available software programme Geneious Pro™ v5.4 (Drummond *et al.* 2011). All chromatograms were manually checked for ambiguities and polymorphisms, and amended as necessary.

The global alignment was carried out using the MUSCLE (Multiple Sequence Comparison by Log-Expectation) algorithm (Edgar 2004), implemented within Geneious, using default settings and a maximum of 6 iterations.

The final alignments of the two cpDNA markers were concatenated into a single alignment.

3.2.4 Haplotype networks and mapping

Alignments were exported from Geneious in PHYLIP format. Chloroplast haplotypes were identified and haplotype networks constructed using the open source software program TCS (Clement *et al.* 2000).

TCS collapses sequences into haplotypes and then an absolute distance matrix is constructed. This absolute distance matrix is constructed by comparing the characters for each sequence and recording the raw number of changes between the sequences. TCS then uses statistical parsimony to create a haplotype network.

When collapsing sequences to haplotypes, TCS encounters problems when the sequences *only* differ at missing or ambiguous character. Because the ITS marker of *D. cantabrica* had many ambiguous characters TCS was not used and the ITS haplotypes were therefore identified and networked manually.

Haplotype piecharts were constructed in MS Excel 2010. Background maps for the haplotype mapping were assembled and formatted in ArcMap 9.2 (ESRI).

3.2.5 Phylogenetic trees

A phylogenetic analysis was carried out on the ITS haplotypes of *D. cantabrica* and *E. mackaiana*, using additional ITS sequences from other species within the Ericaceae. These additional sequences were downloaded from Genebank (Table 14).

Table 14: Genebank accession numbers of Ericaceae sequences used in phylogenetic analysis.

Species	**Accession number**
Erica ciliaris	HQ858977
Erica cinerea	HQ858978
Erica tetralix	HQ859295
Erica vagans	HQ859319
Calluna vulgaris	HM854157
Erica arborea	HQ858911
Erica australis	HQ858926
Erica carnea	HQ858963
Rhododendron ponticum	X97415
Andromeda polifolia	AF358872
Arbutus unedo	AF091953

The Genebank sequences were aligned with the *D. cantabrica* and *E. mackaiana* haplotype sequences using Geneious, and the trimmed alignment exported as a Nexus file.

Maximum parsimony (MP) and maximum likelihood (ML) analyses were carried out using PAUP (Swofford 2003).

Maximum parsimony (MP) analysis was conducted using a heuristic search with random stepwise addition (1000 replicates) and TBR branch-swapping. A total of 1775 trees were retained, and from these a strict consensus tree and a 50% majority rule tree were generated.

Prior to carrying out the maximum likelihood analysis, the programme ModelTest Version 3.7 (Posada and Crandal 1998) was used to test alternative models of evolution for the sequences. The model selected (according to the Akaike Information Criterion) was the GRT+G model. Maximum likelihood analysis was conducted using the selected GTR+G model parameters, with a heuristic search and SPR branch-swapping.

3.2.6 Molecular clock test

To test whether a molecular clock approach could be used to estimate divergence times of haplotypes, a likelihood ratio test was carried out in PAUP. The likelihood ratio test assesses the likelihood of competing hypotheses, in this case, the null hypothesis is that there is a molecular clock operating (and therefore a constant rate of substitution), whereas the alternative hypothesis is that there is no molecular clock (and lineages varying at different rates). The likelihood values of the tree topologies under both hypotheses were compared using a chi-square test.

3.3 Results

Variation was found in at least one base position for all three markers analysed for both species. Overall the ITS marker revealed more haplotypes than either of the chloroplast markers.

3.3.1 Chloroplast haplotype distribution

In both *D. cantabrica* and *E. mackaiana*, and for both chloroplast markers there was variation at least one site. The majority of the variation was in the form of SNPs, with a few indels, the longest being 5 nucleotides.

3.3.1.1 cpDNA haplotype analysis of D. cantabrica

There were 25 populations sampled, with sample sizes ranging from n=1 to n= 5. The total number of samples was 48.

The concatenated multiple sequence alignment of the *trnH-psbA* and *trnL-F* loci was 802 bp in length. The TCS analysis grouped the 48 samples into 13 haplotypes (H1-H13). The sequence differences between each haplotype are summarised in Table 15.

In total there were 22 variable sites. 20 of which were SNPs, and 2 were multi-nucleotide indels. Of the SNPs, 9 were single indels and the other 11 were substitutions.

Table 15: The 13 cpDNA haplotypes of *D. cantabrica*, with polymorphic sites identified by base position across the concatenated alignment of the *trn-psbA* and *trnL-F markers*

bp / Haplotypes	47	74	82	105	125	225	237-241	244	249	250	268	291	528	529	530	540	541	542	551	635-641	722
H1	A	C	-	-	C	C	-----	A	G	A	C	G	T	-	A	-	T	T	A	ATATCA	G
H2	A	C	-	-	C	T	-----	A	G	A	C	G	T	-	A	-	T	T	A	ATATCA	G
H3	A	C	-	-	C	C	-----	A	G	A	C	G	T	A	A	-	T	T	A	ATATCA	G
H4	A	C	-	-	C	C	-----	A	G	A	C	G	T	-	A	T	T	T	A	ATATCA	G
H5	G	C	-	-	C	C	-----	A	G	A	C	G	T	-	A	-	T	T	A	ATATCA	G
H6	A	T	-	-	C	C	-----	A	G	A	C	G	T	-	A	-	T	T	A	ATATCA	G
H7	A	C	-	-	C	C	-----	A	T	A	C	G	T	-	A	-	T	T	T	ATATCA	G
H8	A	C	-	-	C	C	-----	A	G	A	C	A	T	-	A	-	-	T	T	-----	G
H9	A	C	-	-	C	C	-----	A	G	A	C	A	T	-	A	-	-	-	T	-----	G
H10	A	C	A	-	T	C	-----	G	G	G	A	A	-	-	-	-	T	T	T	ATATCA	G
H11	A	C	A	-	T	C	AAAAA	G	G	G	A	A	-	-	A	-	T	T	T	ATATCA	A
H12	A	C	A	-	T	C	AAAAA	G	G	G	A	A	-	A	A	-	T	T	T	ATATCA	G
H13	A	C	A	T	T	C	AAAAA	G	G	G	A	A	-	A	A	-	T	T	T	ATATCA	G

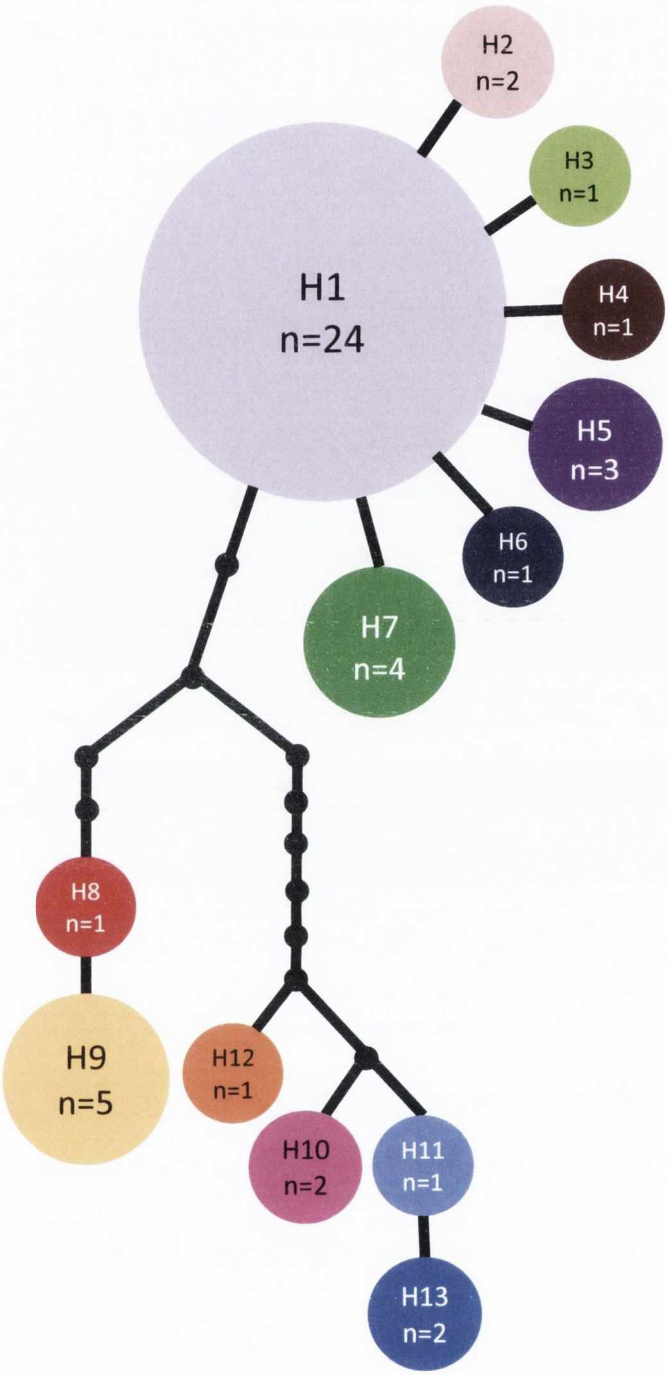


Figure 42: Network of cpDNA haplotypes of *D. cantabrica.* Each line represents a change at one nucleotide, black dots represent missing haplotypes.

The relationship between haplotypes is depicted in the network in Figure 42. The most common haplotype was H1 (n=24). The 12 other haplotypes ranged in frequency from n=1 to n=5. The 5 haplotypes H2 – H7 differed from H1 at a single nucleotide. H8 and H9 were on a more distant branch, which was sister to another branch containing H10-H13.

The distribution of these 13 haplotypes across the 24 populations sampled is shown in Figure 43.

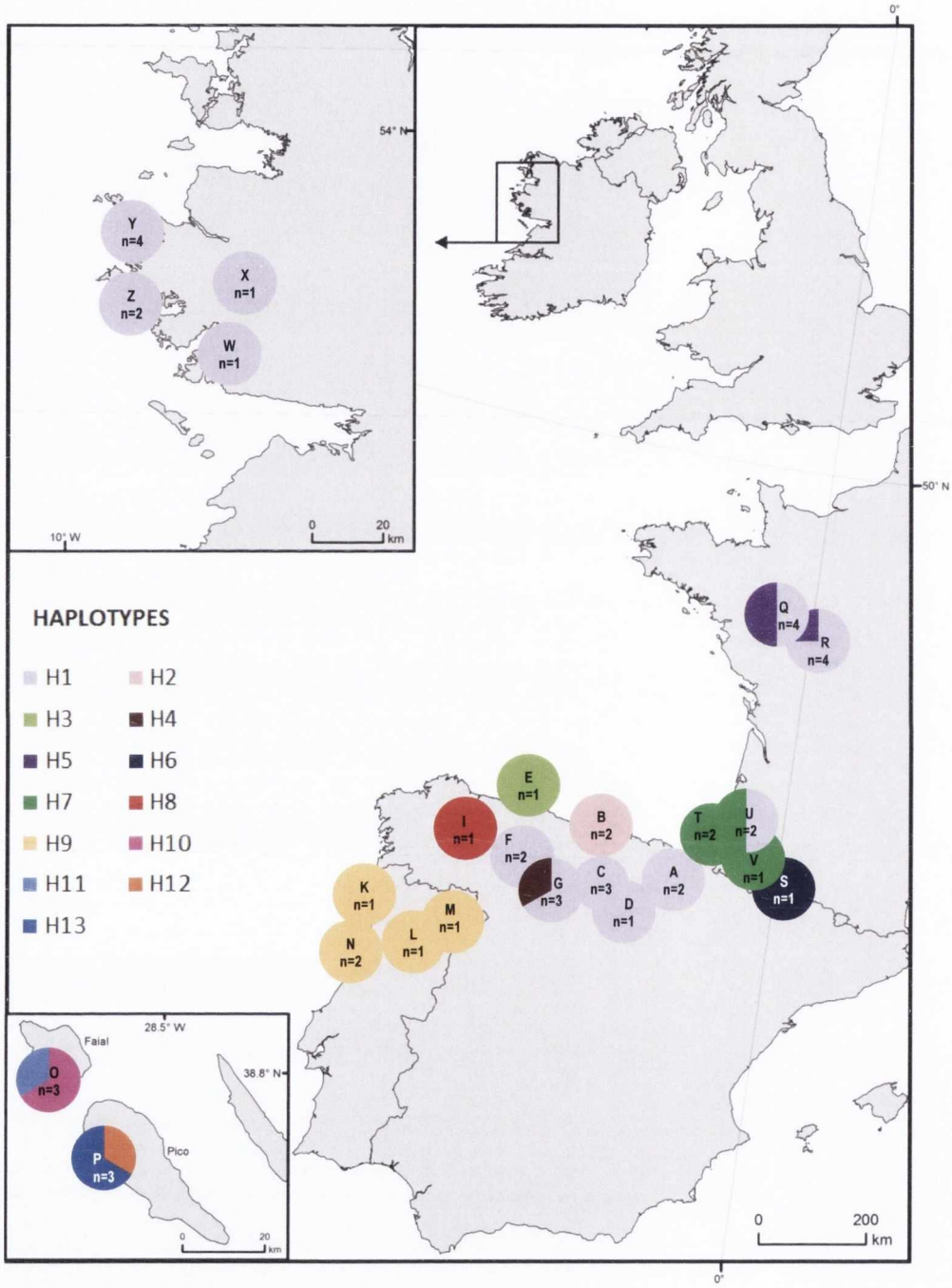


Figure 43: Distribution of cpDNA haplotypes of D. *cantabrica*. Each circle is a population, with colours indicating the proportion of samples belonging to which haplotype, and n=total number of samples in population.

H1, the most common haplotype, is found in 12 of the populations, being found across northern Spain, southern France, the Loire and Ireland. H1 was however *not* found in the populations in the northwest of Spain, nor in Portugal, nor was it present in the two populations in the Azores.

The Azores populations contained four haplotypes unique to them, not found in any other population on the continent, nor in Ireland.

Two haplotypes (H12 and H13) were found on the island of Pico, and the other two (H10 and H11) were found on the island of Faial. The network shows that H13 originates from H11, but that H12 was an earlier haplotype originating from an ancestor of H10, H11 and H13.

On the network the 4 Azorean haplotypes (H10-H13) are found on the branch most distant from H1. Sister to this branch are the haplotypes H8 and H9, which the map indicates are present only in north-west Spain and Portugal.

The four Irish populations contained only the H1 haplotype, there were no haplotypes unique to Ireland. The two populations in the the Loire contained the H1 haplotype as well as a unique haplotype (H5). Five haplotypes: H2, H3, H4, H6 and H8 were each found in only one population; H6 in southern France and the others in Spain.

3.3.1.2 cpDNA haplotype analysis of E. mackaiana

There were 11 populations sampled, with sample sizes ranging from n=1 to n= 5. The total number of samples was 27.

The concatenated multiple sequence alignment of the *trnH-psbA* and *trnL-F* loci was 737 bp in length. The TCS analysis grouped the 27 samples into 4 haplotypes (H1-H4). The sequence differences between each haplotype are summarised in

Table 16

In total there were three variable sites; one indel and two substitutions. The first two polymorphisms were in the *trnH-psbA* marker, and the third was from the *trnL-F* marker.

Table 16: The four cpDNA haplotypes of *E. mackaiana,* with polymorphic sites identified by base position across the concatenated alignment of the *trn-psbA* and *trnL-F* markers

bp / Haplotypes	164	185	387
H1	-	T	T
H2	A	T	T
H3	-	C	T
H4	-	T	G

The network of these haplotypes is shown in Figure 44.

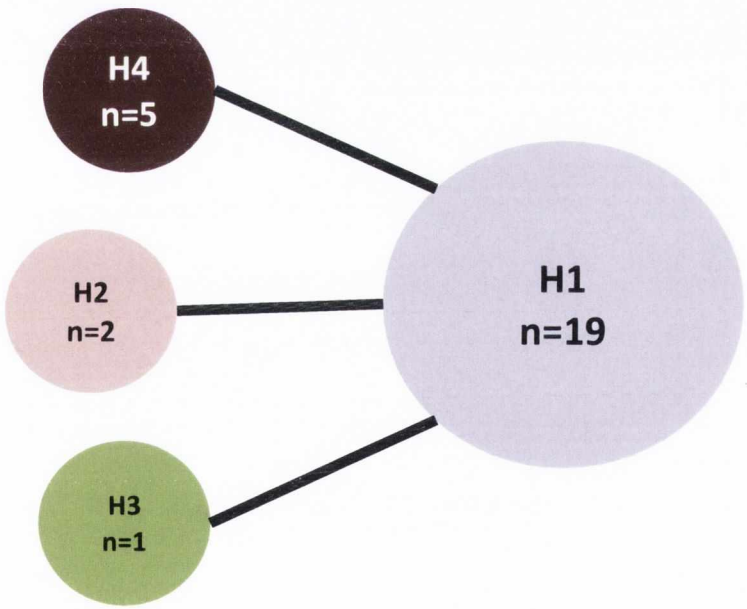


Figure 44: Haplotype network of cpDNA haplotypes of *E. mackaiana*

H1 was the most frequent haplotype, occurring in 19 of the 27 samples. H1, H2 and H3 all differed from H1 at a single base position. H3 was found in only one individual and H2 in two individuals. The distribution of these four haplotypes across the 11 populations sampled is shown in Figure 45.

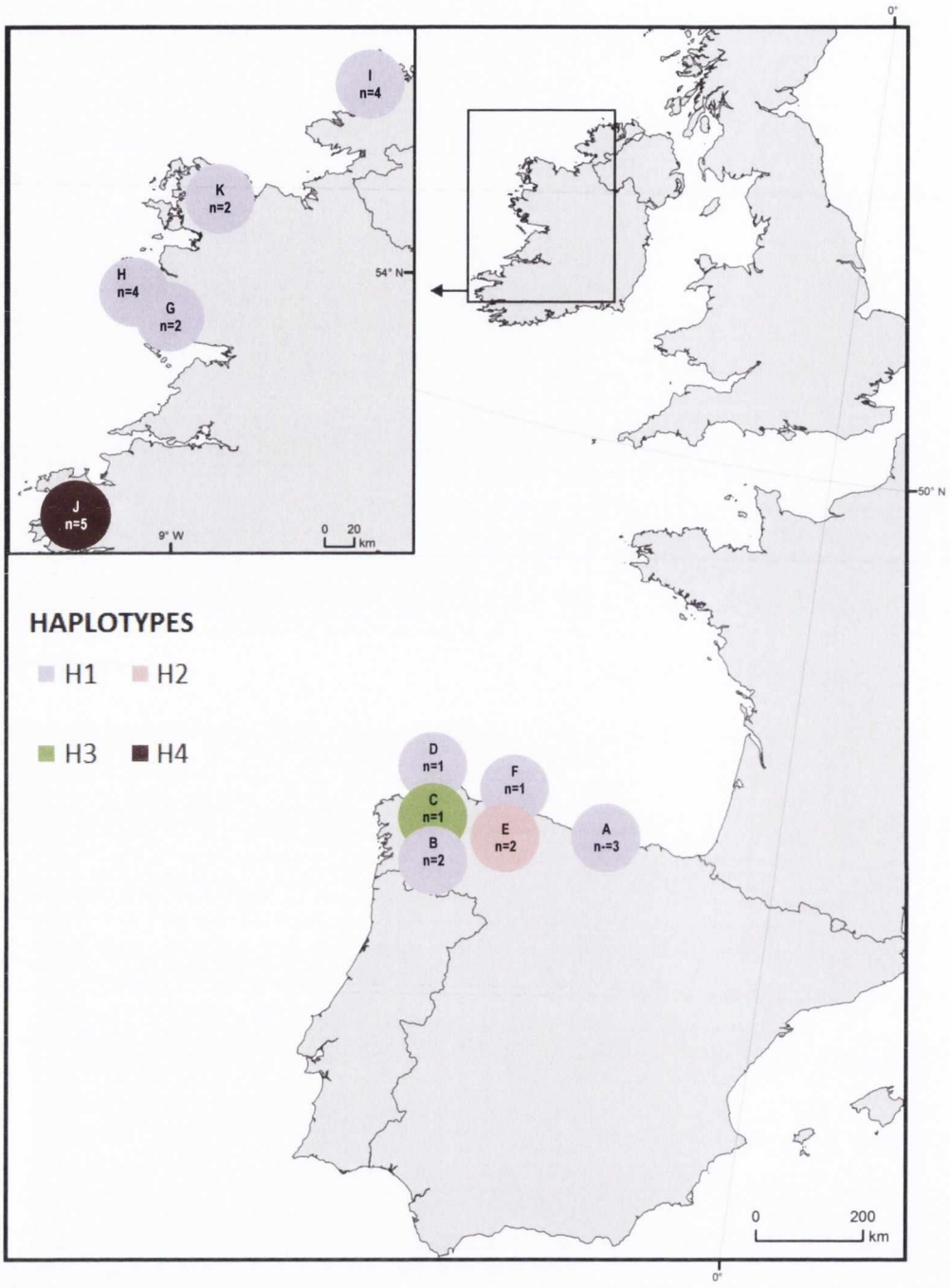


Figure 45: Distribution of cpDNA haplotypes of *E. mackaiana*. Each circle is a population, with colours indicating the proportion of samples belonging to which haplotype, and n=total number of samples in population

H1 is the most widely distributed haplotype, found in four of the six continental populations and four of the five Irish populations. The four samples from the Kerry population were all H4, which is unique to Ireland. H2 and H3 were found in the Sierra de Meira (Populations E) and As Pontes de Garcia (Population C) respectively, in Galicia.

3.3.2 Nuclear ribosomal DNA (ITS) haplotype distribution

3.3.2.1 Nuclear DNA haplotypes of D. cantabrica

289 sequences from 25 populations were analysed. The trimmed alignment was 802bp long. None of the polymorphic sites fell within the 5.8S conserved region.

There was variation at 11 base positions in the ITS sequence. Five nucleotide sites in particular exhibited a high degree of polymorphism both between and within samples. Because of the quantity of these within-sample polymorphisms (ambiguous codes Y, M, W, K), it was problematic to infer haplotypes with TCS. Instead, PAUP was used to find which of the 11 polymorphic characters were phylogenetically informative. From six phylogenetically informative characters identified by PAUP 26 haplotypes were identified manually (Table 17).

Table 17: ITS haplotypes in *D. cantabrica*, based on the 6 phylogenetically informative characters identified by PAUP.

bp	9	176	348	630	643	578
H1	C	Y	A	C	W	-
H2	C	C	M	C	A	-
H3	C	C	A	C	A	-
H4	C	Y	M	C	W	-
H5	C	C	C	C	A	-
H6	Y	C	M	Y	A	-
H7	T	C	A	Y	A	-
H8	T	C	A	T	A	-
H9	C	C	M	C	W	-
H10	C	Y	A	C	A	-
H11	C	T	A	C	T	-
H12	C	Y	M	C	A	-
H13	Y	C	A	Y	A	-
H14	C	C	A	C	W	-
H15	C	T	A	C	W	-
H16	Y	C	A	C	A	-
H17	C	Y	M	C	T	-
H18	C	Y	A	C	T	-
H19	T	C	M	Y	A	-
H20	T	C	A	M	A	-
H21	Y	C	M	C	A	-
H22	C	Y	A	C	W	C
H23	C	C	M	C	A	C
H24	C	C	A	C	A	C
H25	C	Y	M	C	W	C
H26	C	Y	A	C	A	C

Base codes
C = Cytodine
T = Thymidine
G = Guanine
A = Adenosine

Ambiguity codes
Y = C or T
M = A or C
W = A or T
K = G or T

Trees constructed using maximum parsimony are shown in Figure 46.

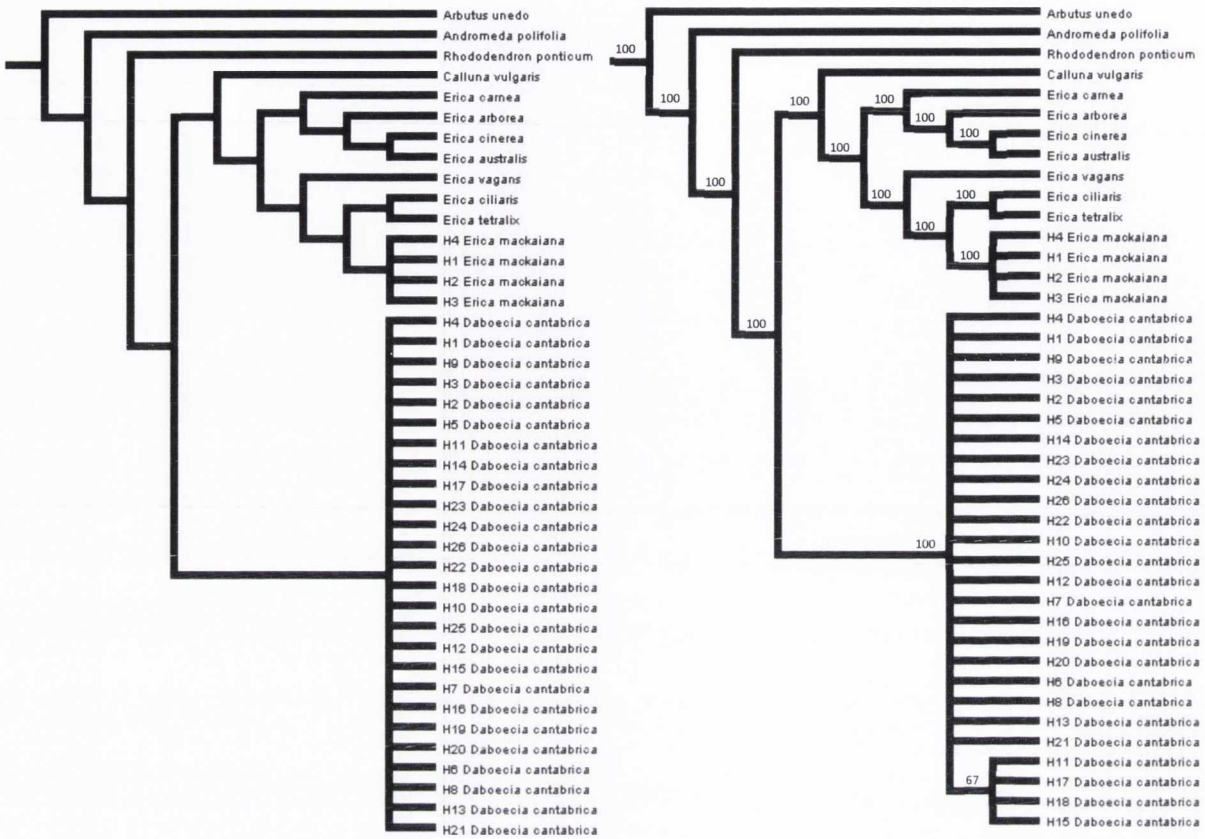


Figure 46: (A) Strict consensus tree of 1775 trees. (B) Maximum parsimony tree with 50% majority rule. The numbers at the nodes are the percentages of trees with that node supported.

The maximum parsimony trees showed the relative position of *D. cantabrica* and *E. mackaiana* relative to other European Ericaceae. *E. mackaiana* is most closely related to *E. ciliaris* and *E. tetralix*. *D. cantabrica* is shown as being a sister group to the Erica species, and to *Calluna vulgaris*. However, maximum parsimony failed to adequately resolve the tree at the haplotype level for either *D. cantabrica* or *E. mackaiana*.

There was a subgrouping of the H11, H17, H10 and H15 haplotypes of *D. cantabrica* present in the majority rule consensus tree, which was supported in 67% of the trees. None of these four haplotypes were found in Ireland.

The two trees that yielded the likelihood values are shown in Figure 47.

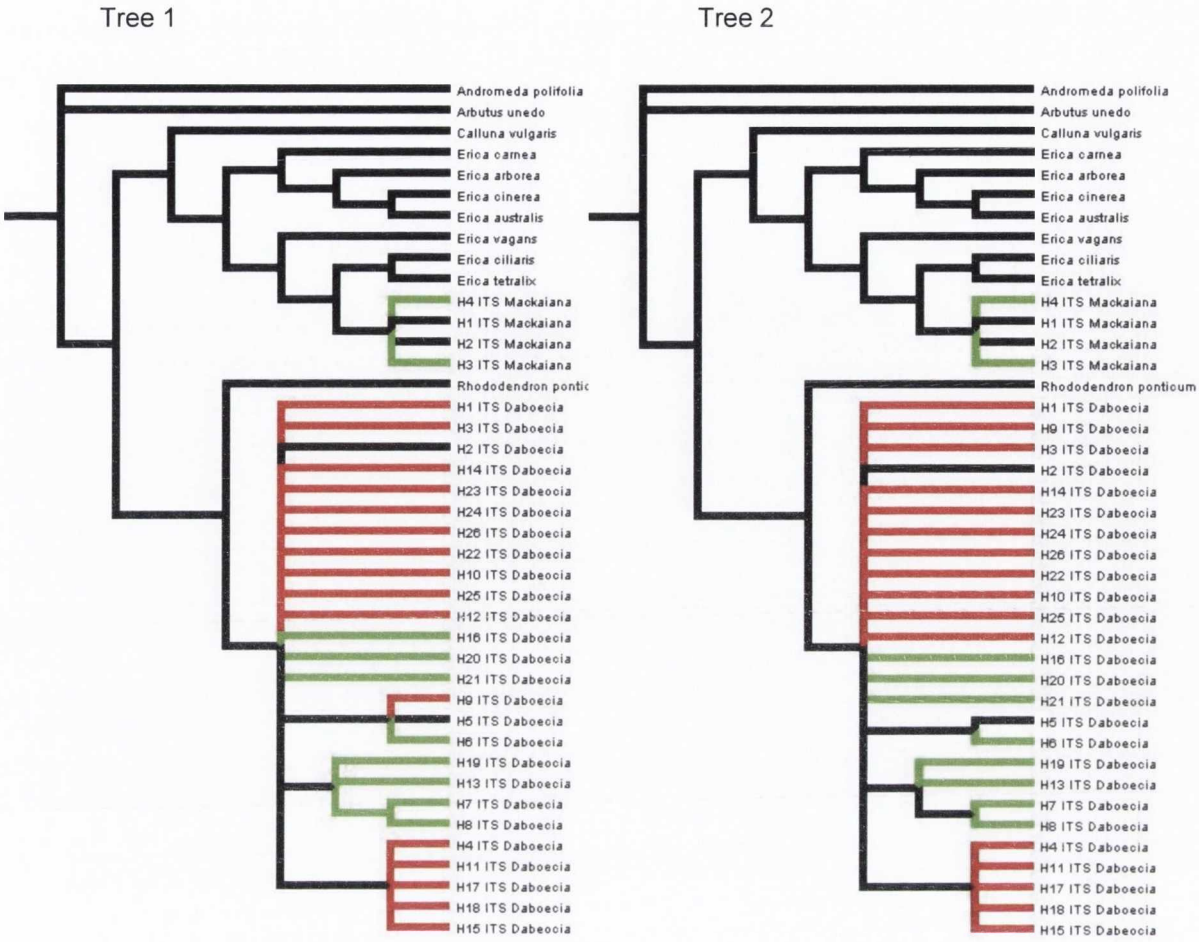


Figure 47: Maximum likelihood trees of haplotypes of *D. cantabrica* and *E. mackaiana* and other Ericaceae species in Europe. Haplotypes of *D. cantabrica* and *E. mackaiana* unique to Ireland are indicated in green, and those not found in Ireland are indicated in red.

Two trees of equally high likelihood were generated, both with log likelihoods of 2883.5. The two trees differ only in the placement of haplotype H9 of *D. cantabrica*. In Tree 1 it is unresolved among other *D. cantabrica* haplotypes, and in Tree 2 it is in a cluster with H5 and H6. Two other subgroups of *D. cantabrica* haplotypes are indicated on both trees: one containing four haplotypes unique to Ireland (H19, H13, H7 and H8), the other containing five haplotypes unique to continental Europe (H4, H11, H17, H18 and H15).

The construction of a haplotype network was done manually and is depicted in Figure 49. Each haplotype was represented as a circle proportionate to the number of samples in that haplotype overall (ranging from n=1 to n=54). Two haplotypes were linked if they differed at a single nucleotide, including a change from a monomorphism to a polymorphism and vice

versa. Figure 48 shows an example of part of the network, with the sequences and the changes shown.

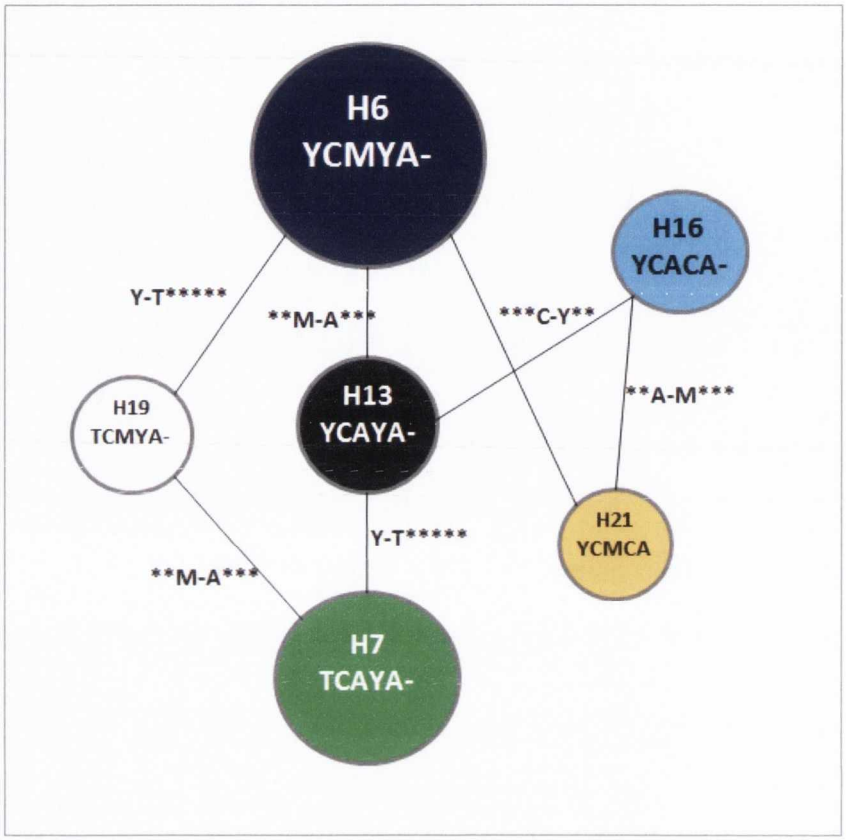


Figure 48: Detail of a portion of the *D. cantabrica* ITS haplotype network; filled circles represent haplotypes, link lines represent a change at one nucleotide position, described beside it, unchanging nucleotides are represented by a *.

The sequences and the changes were omitted from the full network (Figure 49) for clarity. Figure 50 shows the distribution of these 26 haplotypes across Europe.

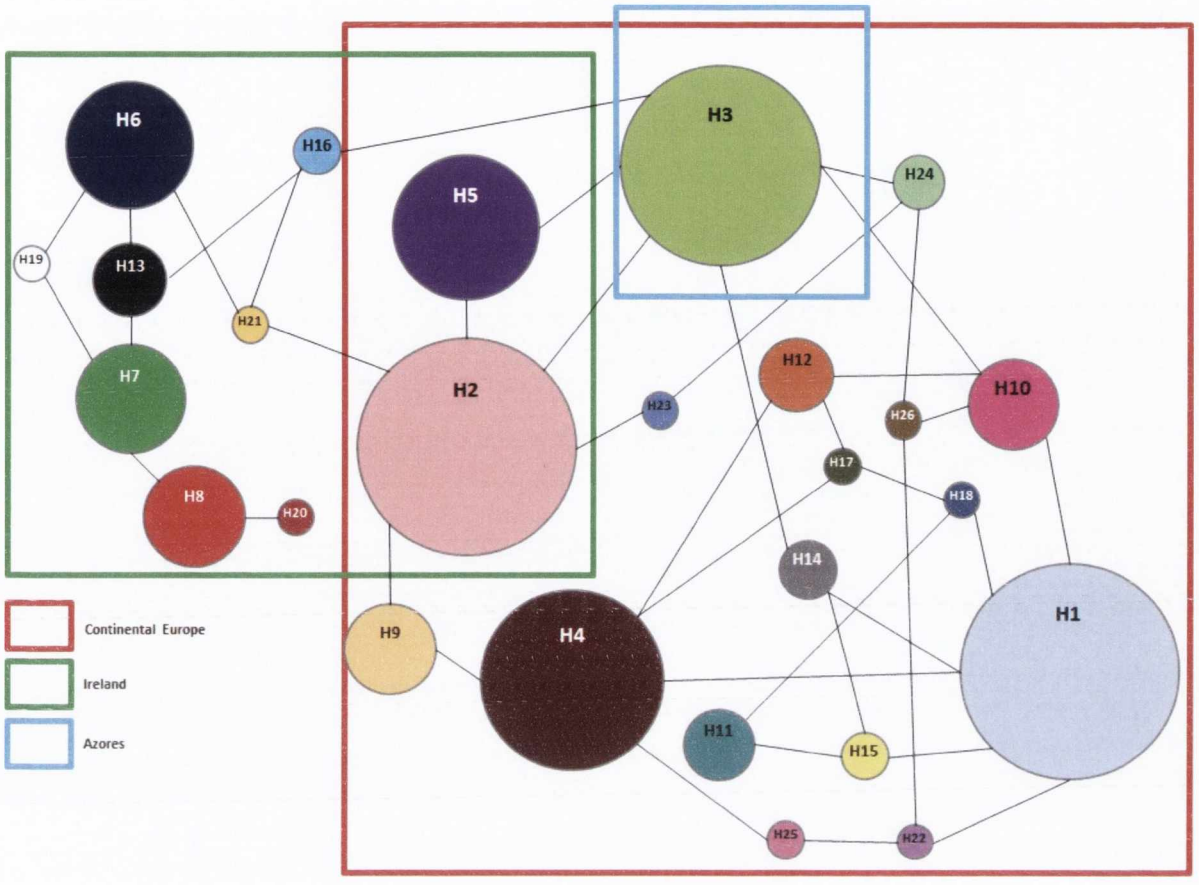


Figure 49: Network of the 26 ITS haplotypes in *D. cantabrica*. Each circle represents a haplotype, the size of which is roughly proportional to the haplotype frequency. Each link line corresponds to a single nucleotide difference, including a change from a monomorphic to a polymorphic state. The rectangles represent the groups of haplotypes found in continental Europe (□), the Azores (□) and Ireland (□).

The most common haplotypes were H1, H2 and H3 which together represented 55% of the samples. Of these three haplotypes, all three occurred in continental Europe, but only H2 was present in Ireland, and only H3 in the Azores.

The next most frequent haplotypes were H4, H5, H6, and H7 which were all found in at least 10 samples. Of these H6 and H7 were unique to Ireland, H4 was unique to parts of continental Europe and H5 was found in both Ireland and the continent.

The only two haplotypes that were found in both Ireland and the continent (overlap of red and green rectangles in Figure 49) were H2 and H5.

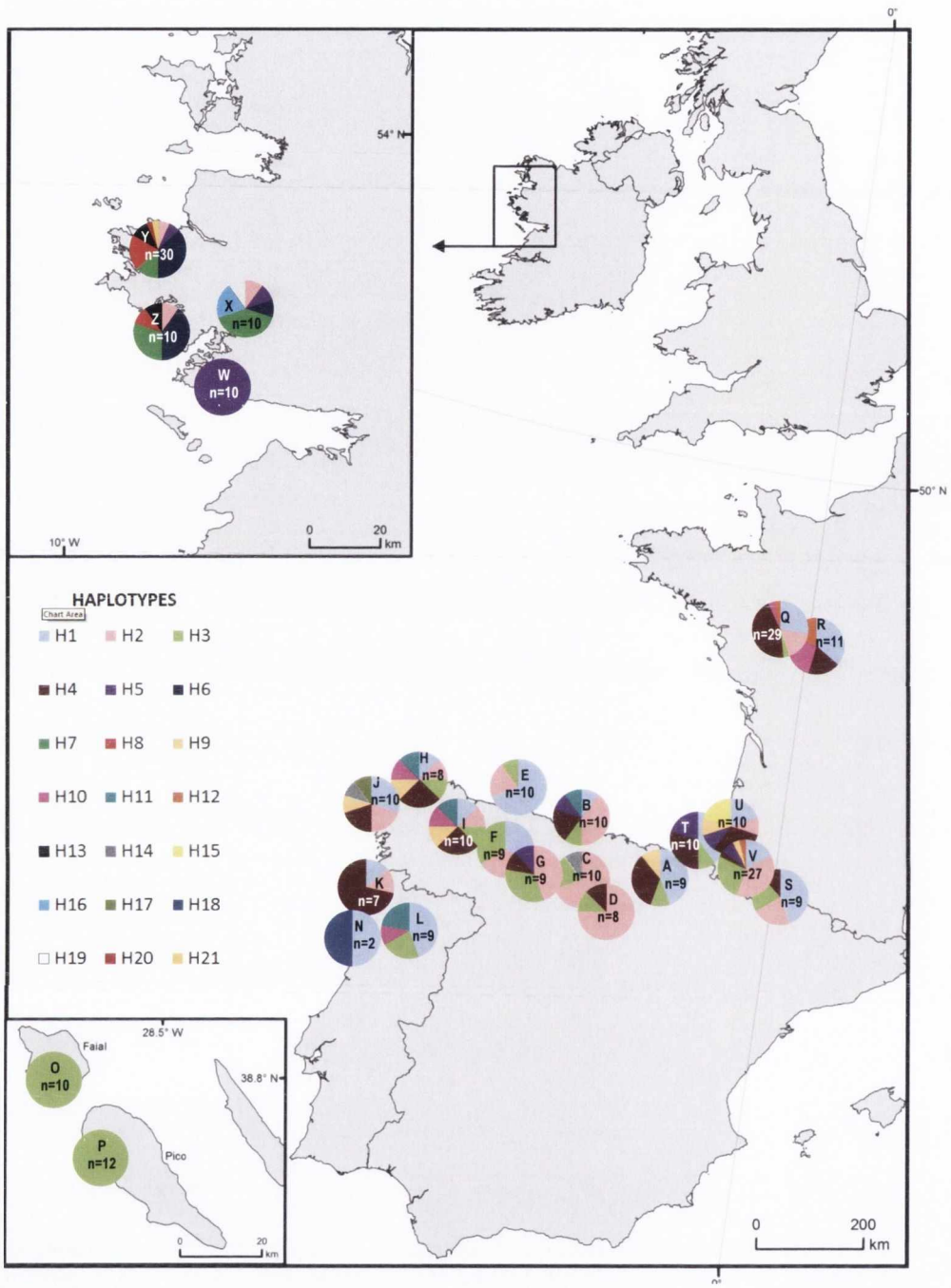


Figure 50: Haplotype distribution of ITS in *D. cantabrica*. Each circle is a population, with colours indicating the proportion of samples belonging to which haplotype, and n=total number of samples in population.

Seven haplotypes (H15-H21) were each found in one population only; H15 in southern France, H16, H19, H20 and H21 in Ireland, and H16 and H17 in NW Spain. There were no haplotypes unique to the populations in the Loire.

3.3.2.2 Nuclear DNA haplotypes of Erica mackaiana

Four haplotypes were found for the ITS marker in *E. mackaiana*, encompassing variation at three base positions. Two of these were single nucleotide indels, the other was a polymorphism (C/T) in two haplotypes and a C in the other two haplotypes (summary in Table 18).

Table 18: Table of ITS haplotypes for *E. mackaiana*

bp / Haplotypes	8	318	754
H1	A	C	-
H2	-	C	-
H3	A	Y	-
H4	A	C	A

Maximum parsimony (Figure 46) and maximum likelihood (Figure 47) analysis of the *E. mackaiana* haplotypes yielded identical results– there is no resolution of the tree at the haplotype level for *E. mackaiana*. All trees did however show *E. tetralix* and *E. cinerea* as the closest of the other Ericaceae species to *E. mackaiana*.

A network of the *E. mackaiana* chloroplast haplotypes is shown below in Figure 51 and Figure 52 shows the distribution of these haplotypes in Spain and Ireland.

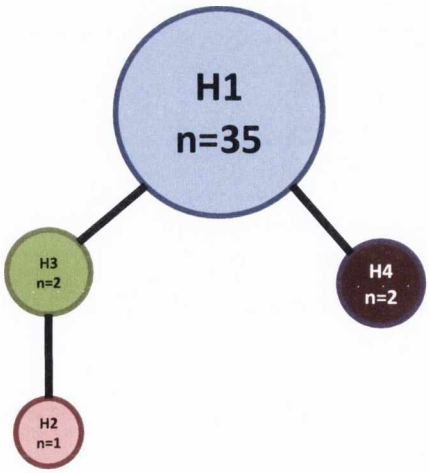


Figure 51: Haplotype network of ITS haplotypes of *E. mackaiana*

Haplotype H1 was the most common haplotype, found in the majority of populations in both Spain and Ireland.

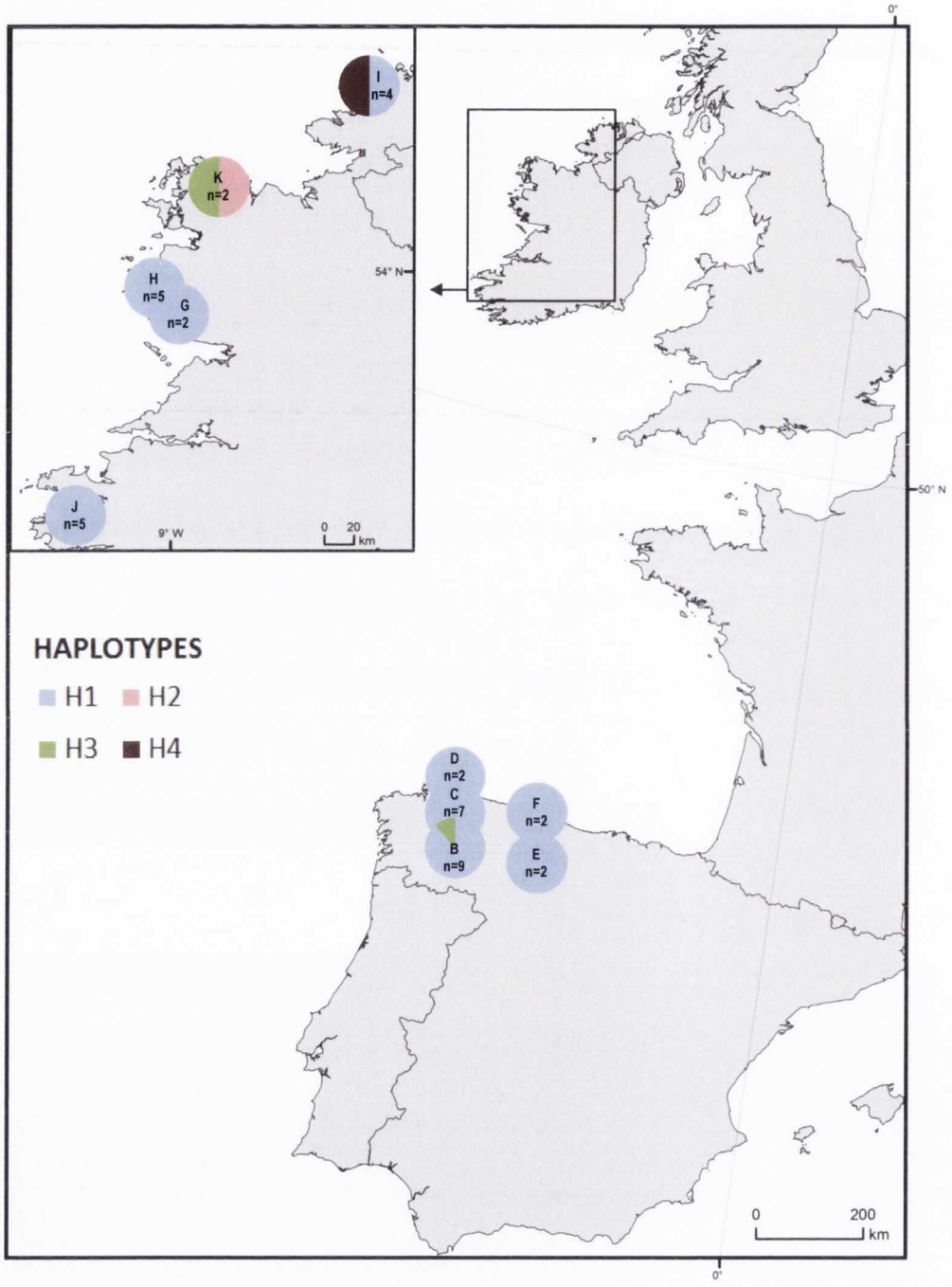


Figure 52: map of ITS haplotypes in *E. mackaiana*

H3 and H4 each differed from H1 at one site, and H2 was different from H3 at another site. Two unique haplotypes to Ireland were identified; Haplotype H4, found only in the Donegal population and H2, found only in the Mayo population.

The Mayo population therefore contained two haplotypes (H2 and H3) not identified in any other population in Ireland, H3 being also found in the Sierra de Loba population of Spain.

3.3.3 Molecular clock test

The likelihood scores yielded by the Likelihood Ratio Test were -2925.8 when a molecular clock was enforced, and -2893.5 when it was not. A chi-square test using 39 degrees of freedom (n-2) determined that the null hypothesis could be rejected ($p<0.05$) and therefore that the data does not fit a clock model.

3.3.4 Genetic evidence for the parentage of hybrid *E. x stuartii*

One sample of *E.* x *stuartii* (from the Kerry population) was also sequenced, as well as a sample of *E. tetralix,* which was confirmed as definitely *E. tetralix* by comparison to a Genbank *E. tetralix* ITS sequence (Accession no. AY520806).

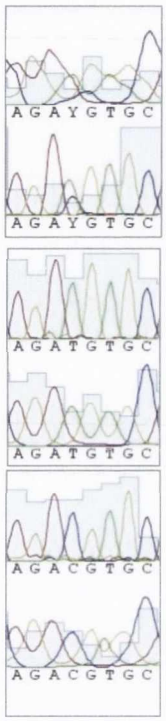
When the assembled sequences from the *E.* x *stuartii* sample was aligned to those of *E. mackaiana* and *E. tetralix* the relationships between them were easily identifiable. For both chloroplast markers the *E.* x *stuartii* sequence was identical to that of *E. tetralix*.

When the ITS sequence was compared to ITS from *E. mackaiana* and *E. tetralix* the hybridisation was clear. The difference between ITS in *E. mackaiana* and *E. tetralix* consists of 15 substitutions and one indel. The *E.* x *stuartii* was polymorphic at each site of difference between *E. mackaiana* and *E. tetralix*, these results are summarised in Table 19.

Table 19: Differences between ITS sequence in *E. mackaiana* and *E. tetralix*, and corresponding heterozygosity in the hybrid *E. x stuartii*

bp / Species	89	111	119	191	208	241	292	308	310	504	505	507	584	599	650	720
E. mackaiana	A	C	C	T	A	C	G	C	T	C	G	C	C	G	T	T
E. x *stuartii*	A/C	C/G	C/T	T/C	A/G	C/T	G/A	C/T	T/C	C/T	G/A	C/-	C/T	G/A	T/C	T/A
E. tetralix	G	G	T	C	G	T	A	T	C	T	A	-	T	A	C	A

In *E.* x *stuartii* the substitutions in the parent sequence show up either as two different bases on the same strand, or as different bases on each strand, see Figure 53 for an example (C in *E. mackaiana* and T in *E. tetralix*).



E. x *stuartii* (double peak C/T)

E. tetralix (T)

E. mackaiana (C)

Figure 53: Substitution showing as a polymorphism in the *E.* x *stuartii* ITS sequence

The single indel difference between *E. mackaiana* and *E. tetralix* sequences causes a discrepancy in sequence length between the strands in *E.* x *stuartii*. The resulting sequence 'stutter' is very clear in the *E.* x *stuartii* chromatogram, as shown in Figure 54. The sequence before the indel in each direction is clear but after the indel the rest of the sequence is out of phase due to the length discrepancy.

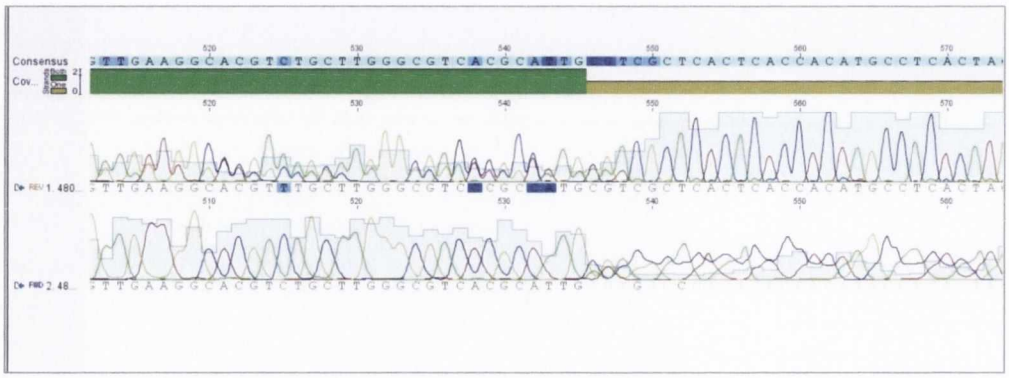


Figure 54: Chromatogram of *E.* x *stuartii* ITS showing the stutter caused by the indel.

3.4 Discussion

3.4.1 General findings

The number of haplotypes identified was greater for the ITS sequences (30 haplotypes) than when using the chloroplast sequences (17 haplotypes). This is in line with the known fact that nuclear DNA has substitution rates around double those of chloroplast DNA (Wolfe *et al.* 1987). It may also have been influenced by the fact that sample sizes for the chloroplast analysis were smaller than for the ITS analysis.

There was also a difference between species in the number of haplotypes identified, with more haplotypes identified in *D. cantabrica* than in *E. mackaiana*. This is most likely due to the fact that fewer populations of *E. mackaiana* were sampled (especially on the continent), and sample sizes were smaller. These facts should be borne in mind when interpreting the results.

The phylogenetic analyses using maximum likelihood and maximum parsimony produced trees that did not have fully resolved topologies at the haplotype level, therefore it was difficult to draw any useful conclusions from them. The *E. mackaiana* haplotypes were not resolved by either of the analyses, whereas there was some clustering of the *D. cantabrica* haplotypes, which is discussed below.

The likelihood ratio test showed that the sequence data for the ITS was not clock-like and therefore it was not possible to draw conclusions of the timing of haplotype divergence using published rates of substitution. Therefore most of the conclusions are based on interpreting the patterns of the distribution of haplotypes, in particular haplotypes unique to certain areas, and how haplotypes are related to one another, as shown by the haplotype networks.

3.4.2 *Daboecia cantabrica*

Both markers showed a similar pattern of one or two very frequent haplotypes, showing linkages to the other, less frequent haplotypes.

Overall the greatest numbers of haplotypes were located across the main continental range of *D. cantabrica*, from northern Portugal to the south of France. Each of the three other regions where *D. cantabrica* is found (Azores, the Loire, Ireland) had fewer haplotypes, and each is discussed in turn.

3.4.2.1 Azores

The ancestral link between the Azorean subspecies *D. cantabrica* subsp. *azorica* and the continent is clearly shown by the cpDNA haplotype network. The four Azorean haplotypes are sister to another branch of two haplotypes (H8 and H9) found only in the NW of Spain and in Portugal. This region corresponds to the part of *D. cantabrica*'s range which is geographically closest to the Azores.

Although both the islands of Faial and Pico, where the sampled populations are found, are less than 1 million years old, the oldest of the Azores islands are approximately 8 million years old (Carine and Schaefer 2010).

Sao Jorge is approximately 5.5 million years old (Rumeu *et al.* 2011), and there are records of *D. cantabrica* subsp. *azorica* on Sao Jorge (Dias, pers comm), although no plants were found during the collection trip for this study. It is therefore possible that the species established on Sao Jorge first, before then spreading to Faial and Pico.

It is also possible that *D. cantabrica* was once established on the older islands to the west (Sao Miguel and Santa Maria) but has since become extinct there.

The branching pattern of the 4 Azorean haplotypes (H10-H13) indicates the possibility of flow between the islands in both directions. The common ancestor of these 4 haplotypes could have become established either on Pico or another island (for example Sao Jorge). From this the H12 haplotype then evolved on Pico, whilst another branch led to the establishment of H10 and H13 on Faial. H11, derived from H13 then became established back on Pico.

Whilst the chloroplast markers show a clear divergence in haplotypes between the Azores and continental Europe, the ITS sequences show less variation within the Azores, with the existence of a single haplotype (monomorphic at all the variable loci) across the two populations sampled on Faial and Pico. This same haplotype is a common one across the continental range.

Taken together, the pattern of the chloroplast and ITS haplotypes suggest that a single founding event was responsible for the establishment of *D. cantabrica* in the Azores. The interpretation of these results is however based on a very small sample size, and further work is required before drawing more concrete conclusions.

3.4.2.2 France - Loire

The populations in the Loire show only two cpDNA haplotypes, one of which (H5) is unique to that region. Six ITS haplotypes were present, none of which are unique to the Loire.

The unique cpDNA haplotype (H5) is a single basepair different to the prevailing haplotype (H1) found across the main continental range. Since ITS has faster rates of evolution than chloroplast DNA it would be more likely that a unique haplotype would be found in the ITS marker than in the cpDNA. Therefore another possibility is that H5 represents a haplotype present but unsampled in Spain or the south of France, which has more recently dispersed somehow to the Loire, and is therefore not unique after all.

The ITS marker shows a higher frequency of H4 than in the rest of continental Europe, but otherwise the Loire populations have a haplotype profile similar to that of the majority of the continental range. The fact that there are six haplotypes present in the Loire populations could indicate a history of multiple founding events from the main continental range.

3.4.2.3 Ireland

Irish populations of *D. cantabrica* showed a single cpDNA haplotype (H1), which is the most common haplotype across the continental range.

A very different pattern emerges, however, from the ITS haplotype network, with 8 different haplotypes unique to Ireland. Eight haplotypes not found in continental Europe are found in Ireland, only 2 haplotypes are shared between the two regions (H2 and H5). On the continent H2 is widely distributed and very common but H5 is restricted to the region of southern France and the eastern half of northern Spain.

What the network of ITS haplotypes indicates overall is that the H2 haplotype of the continent became established in Ireland and then led to the evolution of six more haplotypes, unique to Ireland and not found on the continent. This is not compatible with a scenario of recent human introduction but instead suggests one or more far earlier founding events. The maximum likelihood trees show clustering of unique Irish haplotypes as well as another cluster of unique Spanish haplotypes, which reinforces the conclusion that at least some of the Irish haplotypes have had time to evolve separately from the Spanish ones.

3.4.3 *Erica mackaiana*

Overall the variability across the ITS and chloroplast DNA markers in *E. mackaiana* was much lower than that in *D. cantabrica*.

For all three markers the most common haplotype occurring in Ireland was also the most common haplotype occurring overall. However, three unique haplotypes were found in the Irish populations of *E. mackaiana*; two in the ITS marker (Mayo and Donegal populations) and one in the chloroplast marker (Kerry population).

The three haplotypes unique to Ireland are also unique to individual populations, which would indicate that gene flow within Ireland is restricted, not surprising considering *E. mackaiana's* almost total sterility in Ireland. This would also mean that it is unlikely that any Irish population gave rise to another, but instead is suggestive of independent founding events for each population.

3.4.4 Genetic evidence for *Erica* x *stuartii* parentage

Despite not having a direct bearing on the phylogeographic analysis of *E. mackaiana*, the opportunity was taken to apply the DNA sequencing techniques described in this chapter to investigating the molecular evidence for the hybrid *E.* x *stuartii*. The parentage of this hybrid has been presumed *E. mackaiana* x *E. tetralix* through morphological characteristics and crosses made in cultivation.

The chloroplast and nuclear sequences of the hybrid *E.* x *stuartii* show a clear relationship when compared to the corresponding sequences of its presumed parents *E. mackaiana* and *E. tetralix*.

The ITS sequence in *E.* x *stuartii* shows the clear inheritance of DNA from both parents. The *E.* x *stuartii* shows polymorphism at each site of base substitution in the parent sequences, and at the indel site.

Moreover, *E.* x *stuartii* shares identical chloroplast sequences with *E. tetralix* at both markers tested, giving strong indication of *E. tetralix* as the maternal parent.

Whilst the unidirectional nature of this hybridisation (*E. mackaiana* pollen fertilising *E. tetralix*) has been surmised (from *E. mackaiana*'s reduced pollen production but complete inability to set seed) this is the first time that molecular evidence has confirmed it. However further sequencing with more samples from the other populations in Ireland is necessary before it can be concluded that the hybridisation is consistently unidirectional in this manner.

CHAPTER 4 Genetic Diversity and Differentiation

4.1 Introduction

The previous chapter dealt with the phylogeny of populations, based on lineages of sequence variation in three regions of non-coding DNA. At the broad geographic scale this type of molecular marker system, with its low level of variation is well suited to the elucidation of the phylogeography of the populations across their entire range.

In order to look more closely at the relationships between the populations at smaller geographic levels, to investigate the levels of genetic diversity both within and between them, and to determine patterns of gene flow and population differentiation, a much more variable marker system is generally required (Ouborg *et al.* 1999).

This chapter describes the use of Amplified Fragment Length Polymorphism (AFLPs) to closely examine population diversity and structure in *Daboecia cantabrica* and *Erica mackaiana*, within Ireland and across their European ranges.

4.1.1 Principle of fingerprinting techniques and AFLPs

A fingerprint technique refers to any technique capable of generating a unique and reproducible genetic profile or 'fingerprint' of an individual (Beebee and Rowe 2004). The most commonly used fingerprinting techniques are those based on the amplification capability of the Polymerase Chain Reaction (PCR) (McGregor *et al.* 2000), and include Random Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Simple Sequence Repeats (SSRs, also known as microsatellites) and Inter Simple Sequence Repeats (ISSRs).

AFLPs are based on the selective amplification of restriction fragments from the digestion of total genomic DNA (Vos *et al.* 1995). The selection of which subset of fragments are amplified is determined by the primers used during each of two amplification rounds, which are complementary to sequences at the restriction site in combination with a ligated adaptor sequence (Figure 55).

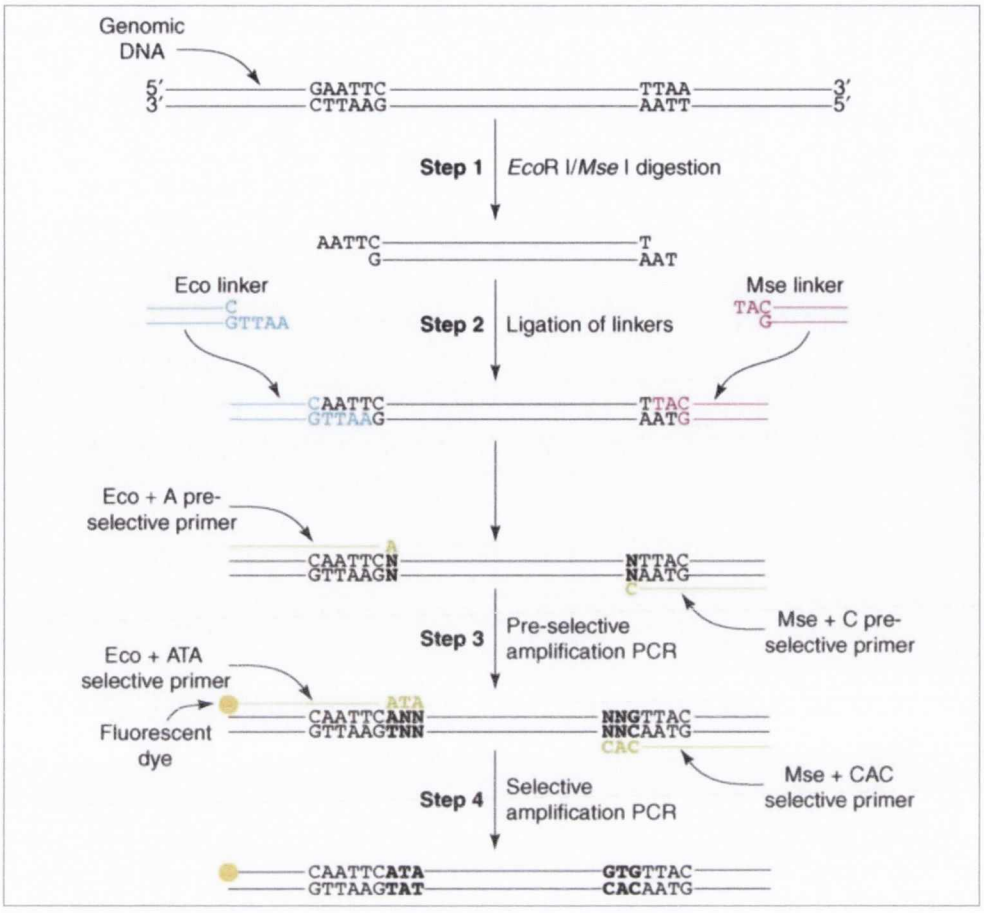


Figure 55: The four-step process of generating AFLPs, reproduced from Meudt & Clark (2007).

The polymorphism that is detected by AFLP arises from the variation in nucleotide sequence either in the restriction site, or in the sequence complementary to the adaptors and primers (Matthes *et al.* 1998). This variation occurs across the entire genome in both coding and non-coding regions.

4.1.1.1 AFLP pros and cons

One of the main advantages of using AFLPs as a molecular marker is that they are generally highly polymorphic across a large number of loci, and do not require any prior sequence information (Rafalski *et al.* 1996). Review studies of the technique have also demonstrated that it is both highly sensitive and reproducible (Meudt and Clarke 2007).

AFLPs are what is known as a dominant marker, since they record only presence or absence at any one locus and cannot distinguish between individuals that are heterozygous or homozygous for the presence of a particular fragment. Compared to the other most commonly

used dominant markers (RAPDs and ISSRs), AFLPs are considered the most reliable for showing population substructure as they generate more reproducible banding patterns (Beebee and Rowe 2004).

Despite these advantages, AFLPs have certain challenges that need to be taken into account. The process involves many steps, making it both labour-intensive and time consuming (Meudt and Clarke 2007). As with other random amplification techniques there also exists the potential for problems arising due to the amplification of homoplasious fragments, i.e. fragments of the same length but which are not of the same sequence, or that have evolved independently and have the same sequence (Caballero *et al.* 2008).

4.1.1.2 AFLP use in phylogeographic studies

AFLPs are commonly used markers in plant population genetics studies and have been applied to many intraspecific phylogeographic analyses to investigate the historical processes resulting in current distribution patterns.

In other species which, like the Hiberno-Iberian heathers, demonstrate a wide range disjunction, AFLPs have been used to investigate hypotheses of origin based on survival in glacial refugia (e.g. Schönswetter *et al.* 2007), vicariance (e.g. Kropf *et al.* 2008, Schneeweiss and Schonswetter 2010, Treier and Muller-Scharer 2011), and long distance dispersal events (e.g. Lihova *et al.* 2009).

In other studies the analysis of AFLPs has also been used to explore the possibility of human introduction of disjunct populations (e.g. Gaudeul 2006, Duffy *et al.* 2009).

4.1.2 Principles of Population Genetics

4.1.2.1 Genetic diversity

Genetic diversity is the amount of heritable variation found within a biological entity such as an individual, a population or a species (Freeland 2005). At its most fundamental level, genetic diversity within an individual refers to the alleles possessed by that individual at any one variable locus, or any type of DNA polymorphism among homologous loci. At the population level genetic diversity describes the extent of variation among individuals or how the different alleles present within the population are distributed across individuals.

The analyses described in this chapter are focussed on investigating the genetic diversity within populations, and at a broader geographical scale, among populations within and among regions.

Patterns of genetic diversity can be indicative of a population's evolutionary history. If a population is recently established from a single or few individuals that population will typically have lower levels of genetic diversity than the source population(s) from which they are derived. This is known as the founder effect (Mayr 1942).

4.1.2.2 Genetic distance

Genetic distance refers to a quantifiable extent of genetic difference between individuals or populations (Nei 1987). Genetic distance measures can be informative of a population's history in a variety of ways.

By measuring the genetic distances between populations we can establish which populations are most closely related to each other genetically, even if geographically they are not close to each other (Palmer *et al.* 2006). The pattern of genetic distance between a population and those geographically close to and far away from it can reveal historical processes such as routes of postglacial colonisation, possible long distance dispersal events, or relict survival.

4.1.2.3 Population differentiation

When two populations are isolated, the two processes of mutation and genetic drift lead over time to genetic differentiation in the populations (Holsinger and Weir 2009). At its most extreme genetic drift can in some cases lead to the extinction of alleles and loss of polymorphism in a population, so that a locus becomes fixed for a single allele (Lowe *et al.* 2004).

In the case of Irish populations showing unique genetic differentiation there would be greater argument for affording them conservation protection.

4.1.3 Previous investigations

AFLPs were previously used in a study (Kingston & Waldren, 2005) of these two species as well as *Erica erigena* R. Ross, another member of the Irish Ericaceae showing a disjunct European range. This study revealed evidence of a founder effect occurring in the Irish populations of both species, and the authors concluded that the Irish populations had been

separated from the continental populations for a considerable period of time, most likely arriving by postglacial colonisation along the coast of France. The current study will expand upon this earlier work, by including a larger number of populations across a wider geographic range and using further statistical analysis.

4.1.4 Aims

The approaches taken in this chapter aimed to:

- assess genetic diversity between and within the Irish populations and the continental populations and examine for evidence of a founder effect,
- analyse genetic distances between Irish and continental populations to determine which continental populations are most closely related to the Irish ones,
- determine whether genetic distance and geographic distance are correlated,
- identify the level of population differentiation for both species.

4.2 Methodology

4.2.1 Sampling

Sampling was carried out as described in Chapter 2. DNA samples from 25 populations of *D. cantabrica* and 11 populations of *E. mackaiana* were included for analysis. The names and localities of these populations, and the number of samples from each population that was used in the AFLP analysis is found in Table 6 (page 42), and their geographic locations are shown below in Figure 56.

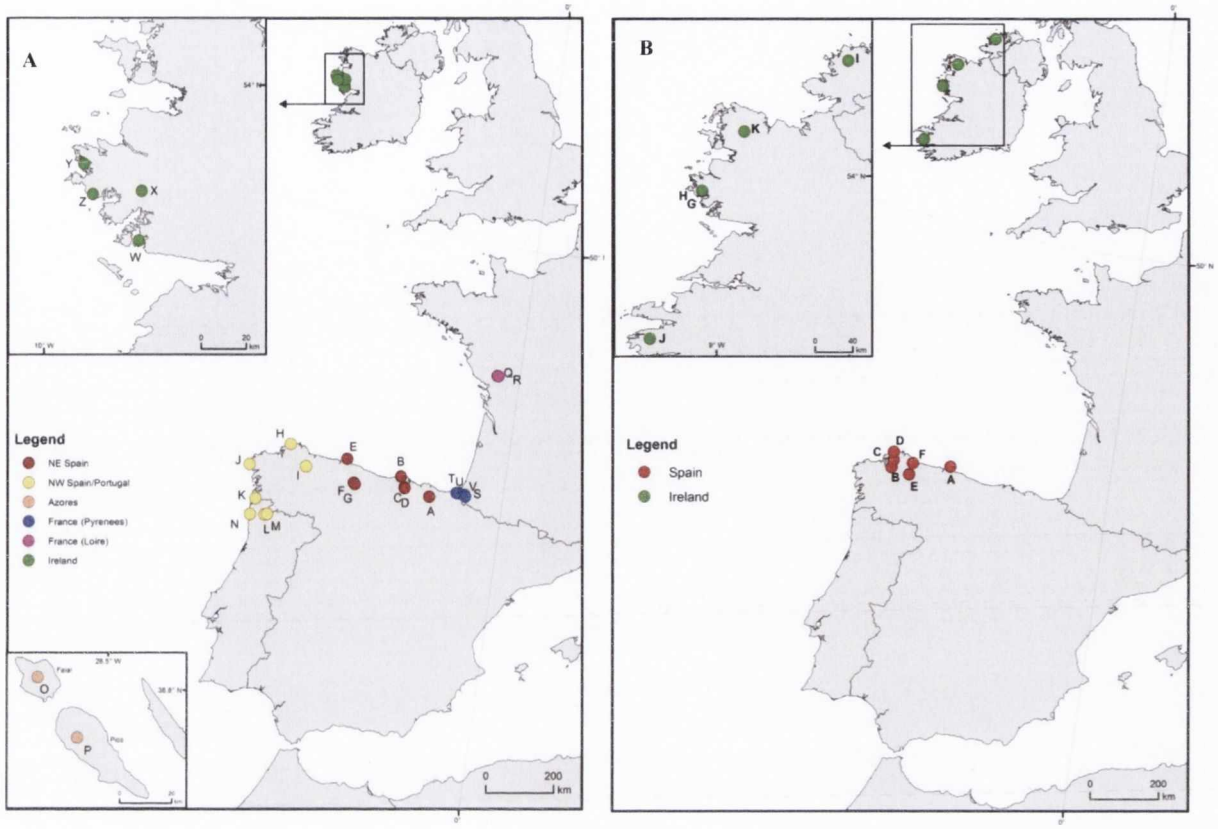


Figure 56: Population locations of AFLP samples for *D. cantabrica* (A) and *E. mackaiana* (B), with regional groupings indicated by colours; 6 regions for *D. cantabrica* and 2 regions for *E. mackaiana*.

4.2.2 AFLP Protocol

The AFLP methodology was based on the principles and techniques of Vos *et al.* (1995), and generally followed the protocol described in the AFLP® Plant Mapping protocol (Applied Biosystems) with some modifications in reaction volumes.

4.2.2.1 DNA extraction and quantification

DNA extractions were carried out using the Qiagen Plant Mini Kits as described in Chapter 2. Before beginning the AFLP protocol, the DNA concentrations of a subset of the samples were measured using a spectrophotometer (UVmini-1240). DNA concentrations were all found to be between 0.01 and 0.03 μg/μl, which is above the minimum concentration recommended by the ABI protocol (0.009 μg/μl). DNA concentrations were not standardised.

4.2.2.2 Restriction-ligation

The restriction enzymes used for the digestion of the genomic DNA were EcoRI and MseI (New England Biolabs), and the ligase enzyme used was T4DNA ligase (Invitrogen).

Reaction volumes were halved from those in the ABI protocol, in order to economise on reagents. The total reaction volume of 5.5µl comprised 2.75µl of total DNA and 2.75µl of mastermix. The makeup of the mastermix per sample was as follows: 0.5µl of each set of adaptor pairs (100U/µl), 0.55µl 10x T4 DNA Ligase buffer with ATP, 0.55µl 0.5M NaCl, 0.275µl 1mg/ml BSA, 0.5 units MseI, 2.5 units EcoRI, 6700 cohesive end units T4 DNA Ligase and 0.115µl water.

Samples were incubated for two hours at 37°C in a thermal cycler. The total volume of the restriction/ligation product was then diluted with 94.5µl of water.

4.2.2.3 Pre-selective amplification

Pre-selective amplification was achieved with the AB Pre-selective Primers that complement the EcoRI and MseI adaptors with an additional nucleotide i.e. MseI+C (5' GAT GAG TCC TGA GTA AC 3') and EcoRI+A (5' GAC TGC GTA CCA ATT CA 3').

The primers, adaptors and PCR mastermix (Core Mix) used were all part of the Applied Biosystems Ligation and Pre-selection Amplification Kit.

Each pre-selective amplification reaction totalled 5µl, comprising the following: 1µl diluted restriction/ligation product, 3.75µl Core Mix and 0.25µl Pre-selective Primers. Reaction samples were initially denatured for 2 minutes at 72°C, followed by 20 cycles of: 94°C for 20 seconds, 56°C for 30 seconds and 72°C for 2 minutes, ending with 30 minutes at 60°C, and then paused at 4°C.

1µl of the pre-selective amplification product was diluted with 19µl of water before storage at 4°C. The remaining 4ul was run on a 1.2% agarose gel to check that amplification had occurred successfully.

4.2.2.4 Selective amplification

In the second round of PCR, three separate selective amplifications were carried out on the diluted pre-selective amplification products, each using a different labelled fluorescent primer.

The choice of these primer combinations was based on those that were used in a previous study of these species (Kingston and Waldren 2005). However, due to financial constraints the same primer combinations were used for both *D. cantabrica* as well *E. mackaiana*, whereas in the previous study the *E. mackaiana* primer combination had been slightly different.

Table 20: Comparison of selective primer combinations used in this study, and previously by Kingston and Waldren (2005).

	This study		Kingston and Waldren (2005)	
	MseI primer	EcoRI primer	MseI primer	EcoRI primer
D. cantabrica	CTC	ACC-FAM	CTC	ACC-FAM
	CAG	ACT-JOE	CAG	ACT-JOE
	CTC	AGC-NED	CTC	AGC-NED
E . mackaiana	CTC	ACC-FAM	CAG	ACA-FAM
	CAG	ACT-JOE	CAC	AAC-JOE
	CTC	AGC-NED	CAG	AAG-NED

Each of the three selective amplification reactions were carried out separately in 5µl reaction volumes comprising the following: 0.75µl of the diluted pre-selective amplification product, 3.75µl of Core Mix, 0.25µl of the MseI primer and 0.25µl of the EcoRI labelled primer.

Reactions were initially denatured at 94°C for two minutes, followed by 10 touchdown cycles of 94°C for 20 seconds, 66°C for 30 seconds (reduced by 1°C per cycle), 72°C for two minutes, and then 20 cycles of 94°C for 20 seconds, 56°C for 30 seconds, 72°C for two minutes, with a final hold at 60°C for 30 minutes, and then holding at 4°C.

4.2.3 Fragment analysis

For each sample the three selective amplifications products were combined, and 1ul of the mixture loaded for fragment analysis on an Applied Biosystems Genetic Analyzer 3100. Each sample also contained 0.2µl of ROX 500HD size standard. A 23-second capillary injection time was used.

4.2.4 Quality control

To check reproducibility, each 96-well plate of samples contained a within-plate replicate as well as a between-plate replicate. In each case replicates refer to samples from the same DNA extraction which have undergone separate restriction-ligation and amplification reactions – either on the same plate, or on another plate Figure 57.

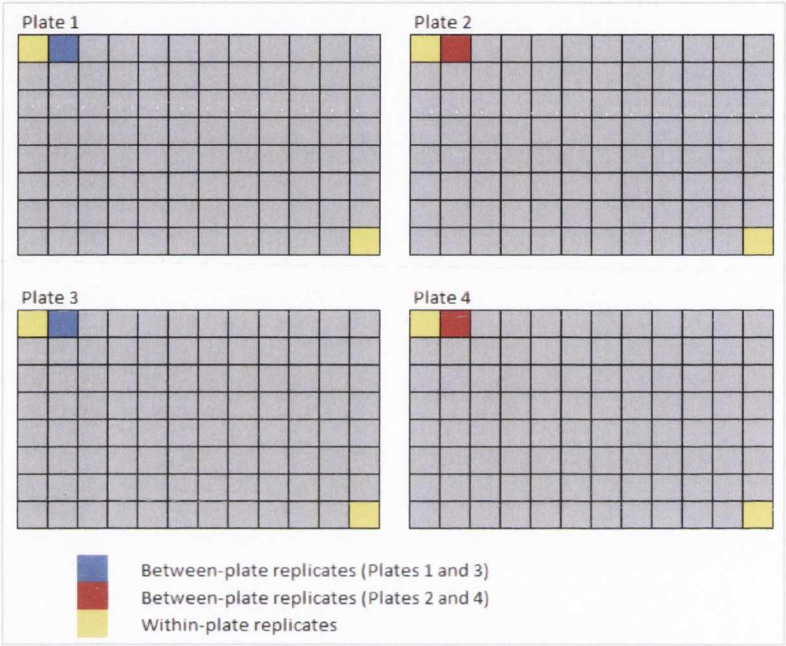


Figure 57: Replicate plan for AFLP plates

4.2.5 Scoring

Automated peak analysis was carried out with GeneMapper 4.0 using default settings, a bin width of 1.0bp, minimum peak height of 50, minimum fragment length of 50bp and maximum fragment length of 500bp.

This initial dataset was then revised based on manual scoring to identify genuine peaks that had not been called and to detect and exclude spurious peaks or 'noise'.

4.2.6 Dataset compilation

Due to the high number of polymorphic loci generated, smaller fragments were excluded from the dataset of each species to minimise the risk of homoplasy effects. For the *E. mackaiana* dataset, all loci less than 75bp long were excluded. For *D. cantabrica*, because of an even greater number of loci, the minimum length was raised to 100bp. Any locus that did not

contain a peak of at least 200 in height was also excluded. The final number of loci remaining in the datasets was 977 *for D. cantabrica* and 600 for *E. mackaiana*.

For a few samples, one primer pair set of the three failed. Not wanting to exclude the samples from the dataset, they were included and the failed primer-set loci were coded as missing data in the data analysis.

4.2.7 Software Packages Used in Data Analysis

4.2.7.1 GenAlEx (Genetic Analysis In Excel)

GenAlEx (Peakall and Smouse 2006) is an open source software package that operates as an add-in to Microsoft Excel. GenAlEx is capable of carrying out a wide range of genetic analyses with codominant, haploid and binary data. In this study GenAlEx was used to calculate the majority of the genetic diversity and genetic distance measures, as well as carry out the Mantel tests and AMOVAs.

The version of GenAlEx used for this study was V6.41, available for download from: www.anu.edu.au/BoZo/GenAlEx/

4.2.7.2 PAUP (Phylogenetic Analysis Using Parsimony)*

PAUP* (Swofford 2003) is a commercially available software package that enables the construction and analysis of phylogenetic trees. In this study, PAUP* was used for the construction of the UPGMA and Neighbour-Joining trees. The version of PAUP* used for this study was PAUP*4.0 for Windows (Command Line interface).

4.2.7.3 Mesquite

Mesquite (Maddison and Maddison 2011) is an open source software package for phylogenetic analysis. For this study it was used primarily to edit the trees produced by PAUP*.

4.2.7.4 STRUCTURE

STRUCTURE (Pritchard *et al.* 2000) is an open source program which uses a Bayesian algorithm to infer population structure among individuals based on multi-locus genotype data. The version of STRUCTURE used in this analysis was STRUCTURE 2.2 (with graphical front end), downloaded from http://pritch.bsd.uchicago.edu/structure.html

4.2.7.5 distruct

The open source package *distruct* (Rosenberg 2004) was used to convert the STRUCTURE output files into a graphical representation. *distruct* (Version 1.1) is command-line driven and was downloaded from http://www.stanford.edu/group/rosenberglab/distructDownload.html.

4.2.8 Analysis

An analysis of each species' binary dataset was carried out to elucidate the genetic diversity both within and between populations and to quantify the genetic distance between populations. A variety of clustering methods were used to visualise these distances on trees and dendrograms, and Mantel tests were carried out to determine whether there were correlations between genetic and geographic distances between populations. Population differentiation and structuring were investigated using Analysis of Molecular variance (AMOVA) and the Bayesian clustering program STRUCTURE.

4.2.8.1 Genetic diversity

The genetic diversity within a population (i.e. between individuals of that population) can be expressed in a number of ways, depending on the nature of the data available (codominant/dominant, haploid/diploid). Those which can be applied to AFLP data (i.e. dominant, multi-locus, binary data) are as follows;

Proportion of polymorphic loci (PPL)

One of the simplest ways to express genetic diversity in a population is to calculate the proportion of polymorphic loci. This is the number of polymorphic loci divided by the total number of loci (both polymorphic and monomorphic).

$$\mathbf{PPL} = \frac{\mathbf{n_p}}{\mathbf{n_{total}}}$$

where,

n_p = number of polymorphic loci

n_{total} = total number of loci

This is more suitable to codominant markers since with dominant markers the estimate produced is biased below the real number (Peakall and Smouse 2006).

Nei's Gene Diversity statistic (H_e)

Another measure of diversity is derived from the frequency of alleles at each locus, and represents the probability that at a single locus any two alleles chosen at random from the population are different to each other. When averaged over all loci this is known as the expected heterozygosity (H_e), or Nei's gene diversity (Nei 1973).

For one locus,

$$\mathbf{h_j = 1 - p^2 - q^2}$$

where,

h_j = heterozygosity per locus

$p + q$ = frequencies of alleles

L= total number of loci

Average for multiple loci,

$$\mathbf{H_e = \frac{\sum_j^L h_j}{L}}$$

Nei's Unbiased Gene Diversity Statistic (UH_e)

The gene diversity measure above is biased in favour of larger sample sizes. Another measure, Nei's unbiased gene diversity is given below, and is used when sample sizes among the populations vary. It multiplies the original Nei's gene diversity statistic by a derivative of the sample size of the population; $\frac{2n}{2n-1}$ (Nei 1987).

$$\mathbf{UH_e = \frac{\sum_j^L h_j}{L} \times \frac{2n}{2n-1}}$$

where,

h_j = heterozygosity per locus

L= total number of loci

n = sample size

4.2.8.2 Genetic distance

Nei's genetic distance and unbiased genetic distance

The standard Nei's genetic distance is based on the calculation of Nei's genetic identity first. The genetic identity and genetic distance between 2 populations, x and y, are calculated as follows;

$$\text{Nei GD} = -\ln(\text{Nei ID})$$

$$\text{NEI ID} = \frac{J_{xy}}{\sqrt{(J_x J_y)}}$$

where,

p_{ix}, p_{iy} = frequency of the *i*th allele in population x and population y.

k = number of loci

$$J_{xy} = \sum_{i=1}^{k} p_{ix}p_{iy}, \qquad J_x = \sum_{i=1}^{k} p_{ix}^2, \qquad J_y = \sum_{i=1}^{k} p_{iy}^2$$

Similarly to Nei's gene diversity measures, Nei's genetic distance also has an Unbiased estimate to take into account different sample sizes. Nei's Unbiased Genetic Distance is calculated the same way as above, with the only difference being that unbiased estimates of J_x and J_y (UJ_x, UJ_y) are used, calculated as follows;

$$UJ_x = \frac{2n\left(\sum_{i=1}^{k} p_{ix}^2 - 1\right)}{2n-1} \qquad UJ_y = \frac{2n\left(\sum_{i=1}^{k} p_{iy}^2 - 1\right)}{2n-1}$$

Tree Construction

To more easily visualise the genetic distances they were presented in the form of trees or dendrograms, using UPGMA (Unweighted Pair Group Arithmetic Means) and NJ (Neighbour-Joining) methods.

Both methods work essentially by grouping together the most similar populations or individuals, based on the pairwise distance matrix calculated. The UPGMA method (Michener and Sokal 1957) begins by finding the two most similar populations or individuals in the distance matrix, joining them at a node, then creating a new distance matrix based on

the mean of those grouped distances to the rest of the populations. The algorithm then finds the next most similar populations and joins them to the tree, and continues on in this way. The UPGMA method assumes that a constant rate of evolution has occurred in the different branches, and a root is estimated as part of the algorithm.

The NJ method (Saitou and Nei 1986) operates in a similar to UPGMA, the main difference being that it subtracts the distance to all other nodes from the pairwise distance. Also important is the fact that the NJ method produces an unrooted tree where the rates of evolution are free to vary in different lineages.

Principal Coordinates Analysis (PCoA)

PCoA is a multivariate technique that finds and plots the major patterns within a multivariate dataset, in this case it was used to find the patterns of variation in genetic distances between populations, in particularly to identify groupings of populations based on genetic distance. Each successive axis that the PCoA finds explains proportionately less of the variation, so in the case of distinct grouping existing in the dataset they would typically be evident on the first 2 or 3 axes.

Mantel Tests

A Mantel test is a test for correlation between two matrices, most frequently used in phylogeographic studies to test for correlation between geographic and genetic distances between populations (Beebee and Rowe 2004). The distance matrix used was that containing the Unbiased Nei genetic distance values, explained in Section 0. The Mantel Tests in this study were run in GenAlEx, and carried out across all populations of each species as well as populations within selected regions. Geographic distances were calculated in GenAlEx using a modification of the Haversine formula (Peakall and Smouse 2012).

4.2.8.3 Population Differentiation

AMOVA and Phi (Φ) statistics

AMOVA (Analysis of Molecular VAriance) is based on comparing multilocus profiles between individuals (Excoffier *et al.* 1992). The genetic variance is divided into a between-

and a within-population component by an analysis of a matrix of squared differences between profiles.

The statistics upon which the AMOVA is based are Wright's F-statistics (Wright 1965), all of which are derived from F, the inbreeding coefficient of a population. The three F-statistics relate to the amount of heterozygosity at various levels of population structure;

- F_{IT} : the inbreeding coefficient of individuals relative to the total population
- F_{IS} : the inbreeding coefficient of individuals relative to the subpopulation
- F_{ST}: a measure of the diversity of randomly chosen alleles within a subpopulation, relative to the total population i.e. a measure of the differentiation of that sup-population.

When a dominant marker in binary form is used (such as AFLPs), an analogue of the F-statistics is used instead. These are known as the Phi-statistics. Phi_{PT} is the equivalent of F_{ST}.

Missing data are problematic for pairwise distance-based analyses such as AMOVA and Mantel tests. Because some samples had missing data the setting to interpolate missing data was selected. GenAlEx interpolates the missing individual-by-individual pairwise distance by inserting an average genetic distance for each population level pairwise contrast.

Bayesian Clustering analysis with STRUCTURE

The raw binary dataset of AFLP data for each species was then submitted to Bayesian clustering analysis to look for population structuring. There are several methods of Bayesian clustering analysis available (Banks and Eichert 2000, Dawson and Belkhir 2001, Corander *et al.* 2008) but as STRUCTURE is the most widely used it was the method selected for this study.

To convert the binary dataset into a form usable by STRUCTURE, the data for each sample was repeated, so that it was in two rows. The Recessive Alleles function was then selected, and a row of 0s given to indicate the recessive state at each locus. The ancestry model selected was Admixture, allowing for an individual to have some fraction of its genome drawn from any of the populations.

The K values (the user-defined number of clusters) was set from K=1 to K=8 and 3 replicates were run for each K to test the stability of the results. The length of the burn-in period was set to 50,000 and the MCMC chains were run for a further 50,000 replicates.

The results of the STRUCTURE analysis were translated to graphical form using the program *distruct*.

The best value of K was selected using two plotting methods; the L(K) method described in the STRUCTURE user manual, and the ΔK method described by Evanno *et al.* (2005).

The L(K) method entails plotting the means of the estimated posterior log probability of the data over the run replicates (denoted as LnP(D) in the STRUCTURE output file) for each K-value. The true value of K (i.e. the true number of populations) is most often indicated by the point on the plot where the L(K) reaches a plateau.

The ΔK method also uses the log probability values. It is an ad hoc statistic calculated in the following way;

(i) $L'(K) = L(K)-L(K-1)$

(ii) $L''(K) = L'(K+1)-L'(K)$

(iii) $\Delta K = \frac{|L''(K)|}{s[L(K)]}$

In the 3rd step the second order rate of change of the log probability (ie $|L''(K)|$) is divided by the standard deviation of the log probability between replicate runs ($s[L(K)]$).

When ΔK is plotted against K, a clear peak in the graph indicates the true value of K.

4.3 Results

4.3.1 Dataset Overall and Quality Control

A summary of the final sample sizes and number of loci in each dataset is given in Table 21. One within-plate replicate failed. 14 samples had approximately a third of the loci coded as missing data due to one of the primer pair sets failing. One sample failed consistently across all three primer pair combinations and was therefore eliminated from the dataset completely.

Table 21: Final dataset summary statistics

		D. cantabrica	*E. mackaiana*
Number of samples		190	82
Number of populations		25	11
Number of samples with some missing data		14	0
Number of loci per primer pair:	***FAM***	*345*	*190*
	JO	*322*	*179*
	NED	*410*	*231*
	Total	1077	600

4.3.2 Genetic diversity

The three measures of the genetic diversity within the populations all showed different but comparable results (Figure 55). In general the population with the higher PPL values also had the higher UH_e and H_e values. The UH_e values for each population were slightly higher than the H_e values.

Overall the Nei statistics ranged from approximately 0.07 to 0.2, whilst all but two of the populations had PPL values over 0.3, the highest being 0.78.

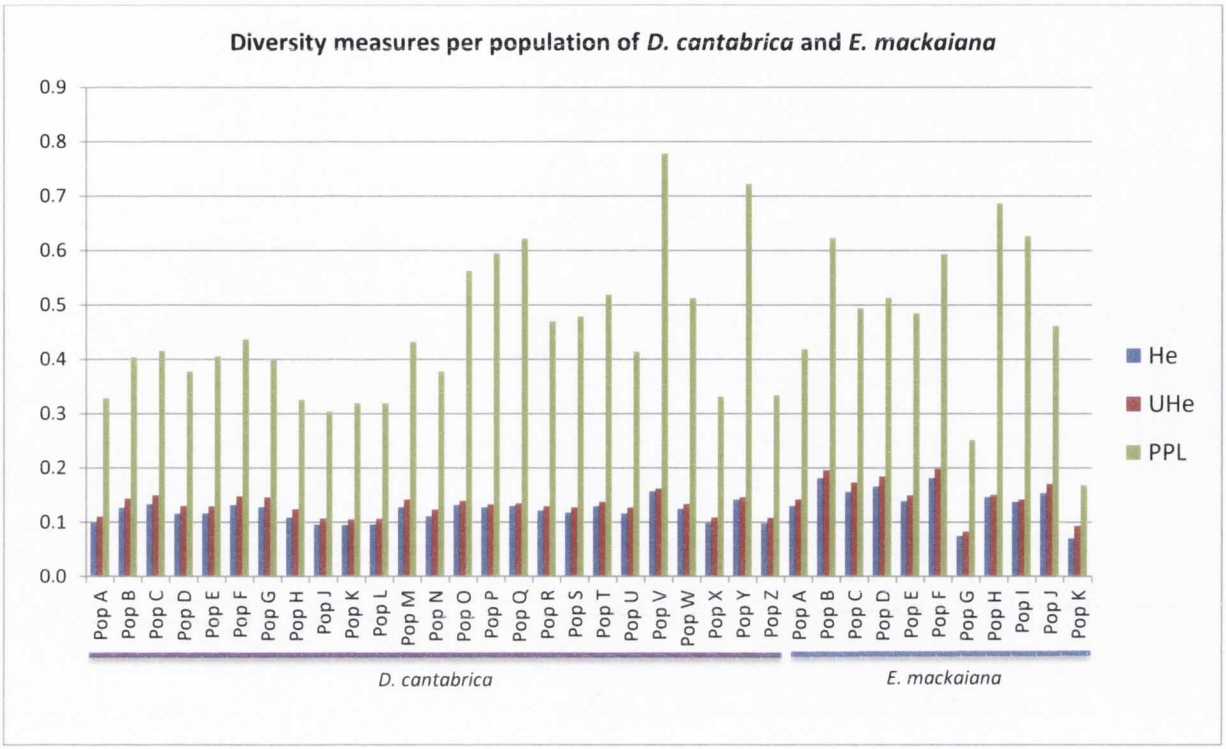


Figure 58: Genetic diversity measures for populations of *D. cantabrica* and *E. mackaiana*. PPL=Proportion of Polymorphic Loci, H_e=Nei's gene diversity, UH_e=Nei's unbiased gene diversity.

The average UH_e across all populations of *D. cantabrica* was significantly lower (p=0.039) than the average across all populations of *E. mackaiana*. The genetic diversity results for the individual populations of *D. cantabrica* and *E. mackaiana* are outlined in Table 22 and Table 24, respectively.

Correlations between sample size and genetic diversity values showed a much higher correlation existing for the PPL measure than for the Nei statistics, as shown in Figure 59. In the *D. cantabrica* dataset the correlation was 0.93 between PPL and sample size, and 0.47 between UH_e and sample size. Similarly, for *E. mackaiana* there was a high correlation (r=0.70) between sample size and PPL, and a much lower correlation (r=0.08) between sample size and UHe.

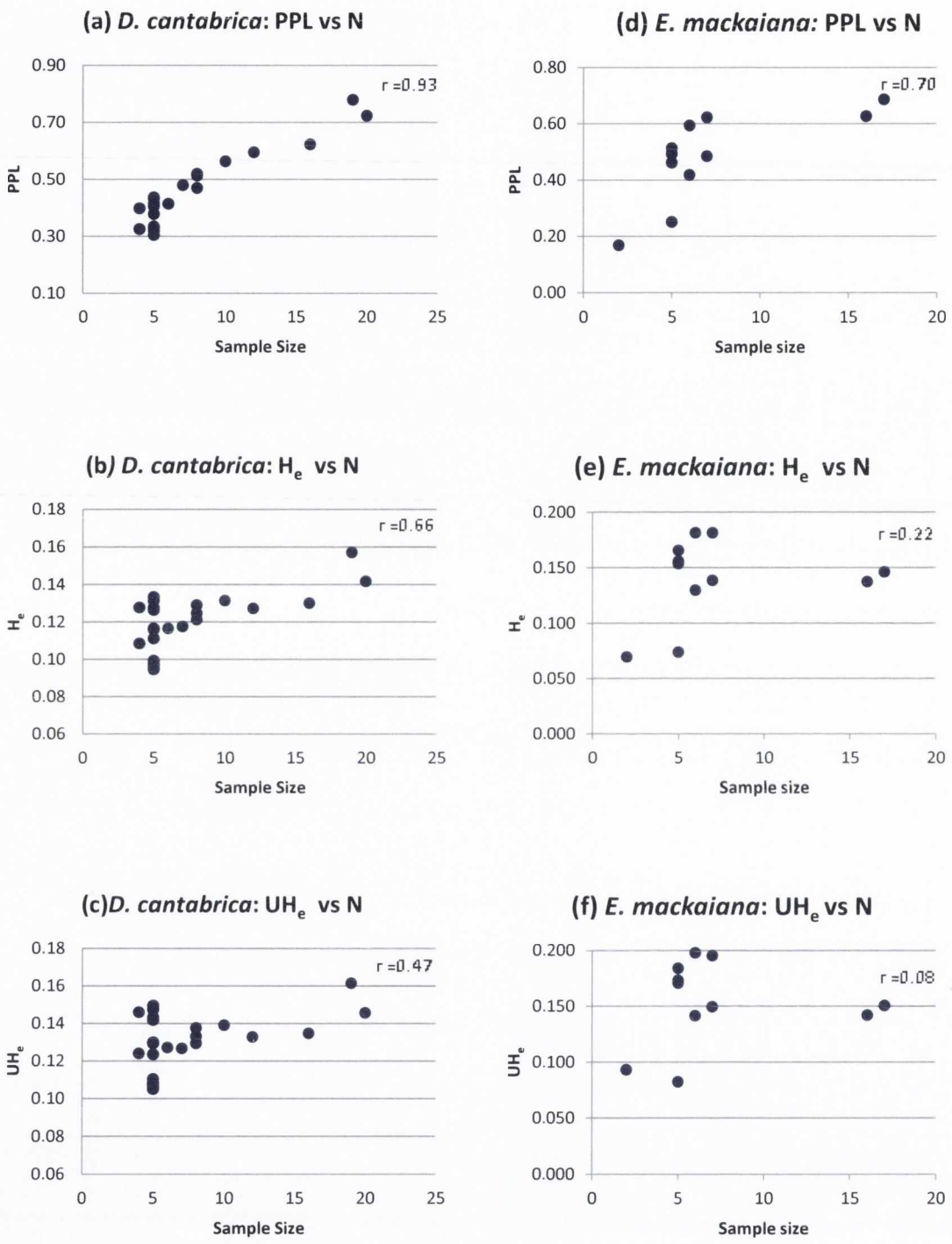


Figure 59: Diversity statistics and their correlation to sample size for *D. cantabrica* (a)-(c) and *E. mackaiana* (d)-(f).

4.3.2.1 Genetic diversity in populations of D. cantabrica

The genetic diversity values for the 25 populations of *D. cantabrica* are presented in Table 22. The table also shows the populations grouped according to six regions. These regional groupings include northern Spain (all populations from the French border to the Asturias), northwest Spain and Portugal (populations in Galicia and Portugal), the French Loire region, southern France (the populations on the French side of the Spanish-French border), the Azores (the island of Faial and Pico) and Ireland.

The population with the highest genetic diversity across all three parameters (PPL, H_e and UH_e) was Pop V (Bidarray) in the southern France region. The lowest gene diversities were found in Pop J, K and L in NW Spain.

In Ireland the population with the highest genetic diversity was Pop Y (Clifden). Pop Z (Errisbeg) had the lowest genetic diversity across all three parameters.

Table 22: Genetic diversity in 25 *D. cantabrica* populations

Region	**Population**	**N**	**Diversity Measures**		
			H_e	**UH_e**	**PPL**
N SPAIN	Pop A Urkiola National Park	5	0.099	0.110	0.33
	Pop B Cabo Major	5	0.126	0.143	0.40
	Pop C Val de Aso	5	0.133	0.149	0.42
	Pop D Cabo Penas, Asturias	5	0.116	0.130	0.38
	Pop E San Isidro Pass	5	0.116	0.129	0.41
	Pop F Mt Torneo	5	0.131	0.147	0.44
	Pop G Sierra de Capellada	4	0.128	0.146	0.40
NW SPAIN & PORTUGAL	Pop H Sierra de Meira	4	0.108	0.124	0.33
	Pop J Cabo Fisterra	5	0.095	0.107	0.30
	Pop K Vigo	5	0.095	0.105	0.32

Region	Population	N	Diversity Measures		
			H_e	UH_e	PPL
	Pop L Peneda-Gerês (1)	5	0.095	0.106	0.32
	Pop M Peneda-Gerês (2)	5	0.128	0.142	0.43
	Pop N Neiva Catelo	5	0.111	0.123	0.38
AZORES	Pop O Faial	10	0.131	0.139	0.56
	Pop P Pico	12	0.127	0.133	0.59
FRANCE (LOIRE)	Pop Q St-Cyr-des-Gats, Vendée	16	0.130	0.135	0.62
	Pop R Brebaudet, Vendée	8	0.121	0.129	0.47
FRANCE (SOUTH)	Pop S La Rhune	7	0.118	0.127	0.48
	Pop T Col d'Ibardin	8	0.129	0.138	0.52
	Pop U Pas de Roland	6	0.116	0.127	0.41
	Pop V Bidarray	19	0.157	0.161	0.78
IRELAND	Pop W Carraroe, Co. Galway	8	0.125	0.133	0.51
	Pop X Derryneen, Co Galway	5	0.098	0.109	0.33
	Pop Y Clifden	20	0.142	0.146	0.72
	Pop Z Errisbeg	5	0.097	0.108	0.33
TOTAL	***Min***	***4***	***0.095***	***0.105***	***0.30***
	Max	***20***	***0.157***	***0.161***	***0.78***

The diversity statistics were then averaged across populations within each region, as shown in

Table 23. A slightly different pattern emerged for the Nei statistics compared to the PPL.

Southern France and the Azores both showed the highest genetic diversity, whether using PPL or the Nei statistics. When using the Nei statistics as the measure of genetic diversity Ireland showed diversity between that of northern Spain and northwest Spain, but when using PPL the

Irish genetic diversity was higher than both of those regions. The France (Loire) region had a higher genetic diversity than Ireland or either of the Spanish regions, both for PPL and the Nei measures.

When the continental regions were agglomerated into an overall region 'continental Europe' (

Table 23), and compared to the Azores and Ireland regions, the Azores had the highest genetic diversity across all three measures. The lowest genetic diversity was found in Ireland, if measured with the Nei statistics, or in continental Europe if measured with PPL.

Table 23: *D. cantabrica* genetic diversity across regions

Region		No. of Pops	H_e	UH_e	PPL
6 regions	NE Spain	7	0.121	0.136	0.395
	NW Spain/Portugal	6	0.105	0.118	0.346
	Azores	2	0.129	0.136	0.579
	France (Loire)	2	0.126	0.132	0.546
	France (South)	4	0.130	0.138	0.548
	Ireland	4	0.115	0.124	0.475
3 regions	Continental Europe	19	0.119	0.130	0.428
	Azores	2	0.129	0.136	0.579
	Ireland	4	0.115	0.124	0.475

4.3.2.2 Genetic diversity in populations of E. mackaiana

Table 24 shows how the populations with the highest or lowest diversities varied depending on the measure used. The highest diversity was found in Pop F when using the Nei measures of H_e and UH_e, but in Pop H (Roundstone2) when using PPL. The lowest diversity was found in the Irish Pop K (Mayo) when using H_e or PPL, but in Pop G (Roundstone1) when using UH_e.

Table 24: Genetic diversity in 11 populations of *E. mackaiana*

Region	Population	N	Diversity Measures		
			H_e	UH_e	PPL
SPAIN	Pop A Cabo Penas	6	0.130	0.141	0.42
	Pop B Sierra da Loba	7	0.181	0.195	0.62
	Pop C Garcia Rodriguez	5	0.156	0.173	0.49
	Pop D Sierra de Capelada	5	0.166	0.184	0.51
	Pop E Sierra da Meira	7	0.139	0.149	0.49
	Pop F Foz	6	0.181	0.198	0.59
IRELAND	Pop G Roundstone1	5	0.074	0.082	0.25
	Pop H Roundstone2	17	0.146	0.151	0.69
	Pop I Donegal	16	0.137	0.142	0.63
	Pop J Kerry	5	0.154	0.171	0.46
	Pop K Mayo	2	0.070	0.093	0.17
TOTAL	MIN	2	0.070	0.082	0.17
	MAX	17	0.181	0.198	0.69

When the genetic diversity figures were averaged across the populations in the two regions Ireland and Spain (Table 25), Ireland showed consistently lower genetic diversity than Spain for all 3 parameters.

Table 25: Genetic diversity for *E. mackaiana* populations aggregated across regions

REGION	No. of Pops	H_e	UH_e	PPL
Spain	6	0.159	0.174	0.521
Ireland	5	0.116	0.128	0.439

4.3.3 Genetic Distance

4.3.3.1 Genetic distances between populations of D. cantabrica

Table 7 shows the pairwise population matrix of genetic distances between the 25 *D. cantabrica* populations, calculated using the Nei Unbiased Genetic Distance statistic.

The smallest genetic distance was 0.007, between Pop O (Faial) and Pop P (Pico) in the Azores. The smallest distance after that was between Pop Q and Pop R, the two populations in the Loire. Pop T in southern France had an equally low genetic distance (0.009) from Pop V (also in southern France) from that of Pop D (in NE Spain).

The two highest genetic distances were both associated with Pop X in Ireland, which had a distance of 0.069 to Pop H and 0.067 to Pop K, both of which are Spanish populations. Three of the four Irish populations were genetically most similar to populations in southern France, the other was genetically closest to Pop D in Spain. The Azores populations were genetically most similar to Pop V (Bidarray) in southern France, although they also showed small genetic distances to populations across Spain and Portugal.

Both of the Loire populations were most similar to Pop T in southern France.

Table 26: Pairwise Population Matrix of Nei Unbiased Genetic Distance for 25 populations of *D. cantabrica*

PopA	PopB	PopC	PopD	PopE	PopF	PopG	PopH	PopJ	PopK	PopL	PopM	PopN	PopO	PopP	PopQ	PopR	PopS	PopT	PopU	PopV	PopW	PopX	PopY	PopZ	
0.000																									PopA
0.040	0.000																								PopB
0.022	0.048	0.000																							PopC
0.012	0.040	0.019	0.000																						PopD
0.040	0.017	0.050	0.038	0.000																					PopE
0.020	0.044	0.021	0.020	0.044	0.000																				PopF
0.051	0.028	0.054	0.042	0.028	0.052	0.000																			PopG
0.031	0.057	0.037	0.027	0.056	0.036	0.060	0.000																		PopH
0.026	0.050	0.034	0.019	0.049	0.026	0.052	0.023	0.000																	PopJ
0.031	0.054	0.032	0.024	0.049	0.030	0.057	0.025	0.014	0.000																PopK
0.026	0.051	0.033	0.018	0.046	0.029	0.055	0.022	0.010	0.011	0.000															PopL
0.050	0.028	0.057	0.046	0.023	0.056	0.032	0.053	0.041	0.040	0.039	0.000														PopM
0.025	0.039	0.028	0.019	0.033	0.029	0.045	0.022	0.013	0.012	0.011	0.030	0.000													PopN
0.037	0.048	0.044	0.037	0.045	0.042	0.050	0.056	0.049	0.046	0.040	0.053	0.039	0.000												PopO
0.032	0.041	0.042	0.032	0.038	0.038	0.050	0.054	0.042	0.043	0.037	0.047	0.033	0.007	0.000											PopP
0.027	0.023	0.032	0.023	0.021	0.030	0.029	0.039	0.032	0.035	0.029	0.029	0.023	0.039	0.031	0.000										PopQ
0.015	0.030	0.022	0.016	0.030	0.022	0.040	0.034	0.027	0.033	0.026	0.039	0.023	0.039	0.032	0.008	0.000									PopR
0.020	0.036	0.024	0.016	0.032	0.021	0.042	0.037	0.031	0.033	0.026	0.046	0.022	0.036	0.032	0.021	0.017	0.000								PopS
0.017	0.030	0.024	0.009	0.022	0.022	0.035	0.034	0.028	0.030	0.024	0.037	0.020	0.033	0.029	0.015	0.013	0.011	0.000							PopT
0.039	0.028	0.049	0.040	0.023	0.047	0.035	0.055	0.054	0.059	0.056	0.040	0.040	0.056	0.048	0.026	0.030	0.029	0.025	0.000						PopU
0.021	0.021	0.028	0.021	0.018	0.027	0.026	0.039	0.033	0.036	0.032	0.032	0.024	0.032	0.027	0.017	0.017	0.016	0.009	0.016	0.000					PopV
0.031	0.040	0.038	0.025	0.039	0.032	0.038	0.041	0.037	0.039	0.037	0.049	0.027	0.052	0.048	0.031	0.029	0.024	0.024	0.040	0.025	0.000				PopW
0.049	0.031	0.062	0.051	0.031	0.057	0.038	0.069	0.061	0.067	0.061	0.044	0.049	0.059	0.053	0.041	0.044	0.043	0.037	0.036	0.026	0.024	0.000			PopX
0.029	0.024	0.039	0.027	0.021	0.036	0.029	0.045	0.039	0.042	0.037	0.035	0.028	0.039	0.033	0.025	0.026	0.023	0.020	0.032	0.016	0.012	0.013	0.000		PopY
0.035	0.058	0.037	0.026	0.059	0.037	0.057	0.042	0.037	0.038	0.038	0.064	0.034	0.060	0.055	0.046	0.040	0.035	0.032	0.061	0.038	0.014	0.047	0.026	0.000	PopZ

Using the distance matrix in Table 26, a UPGMA dendogram was constructed from these distances (Figure 60). The Irish populations form a cluster, as do the two Azorean populations and the two populations in the Loire. The clustering patter for the other regions is less clear, with populations from across Spain clustering with populations from southern France and from Portugal in different groups.

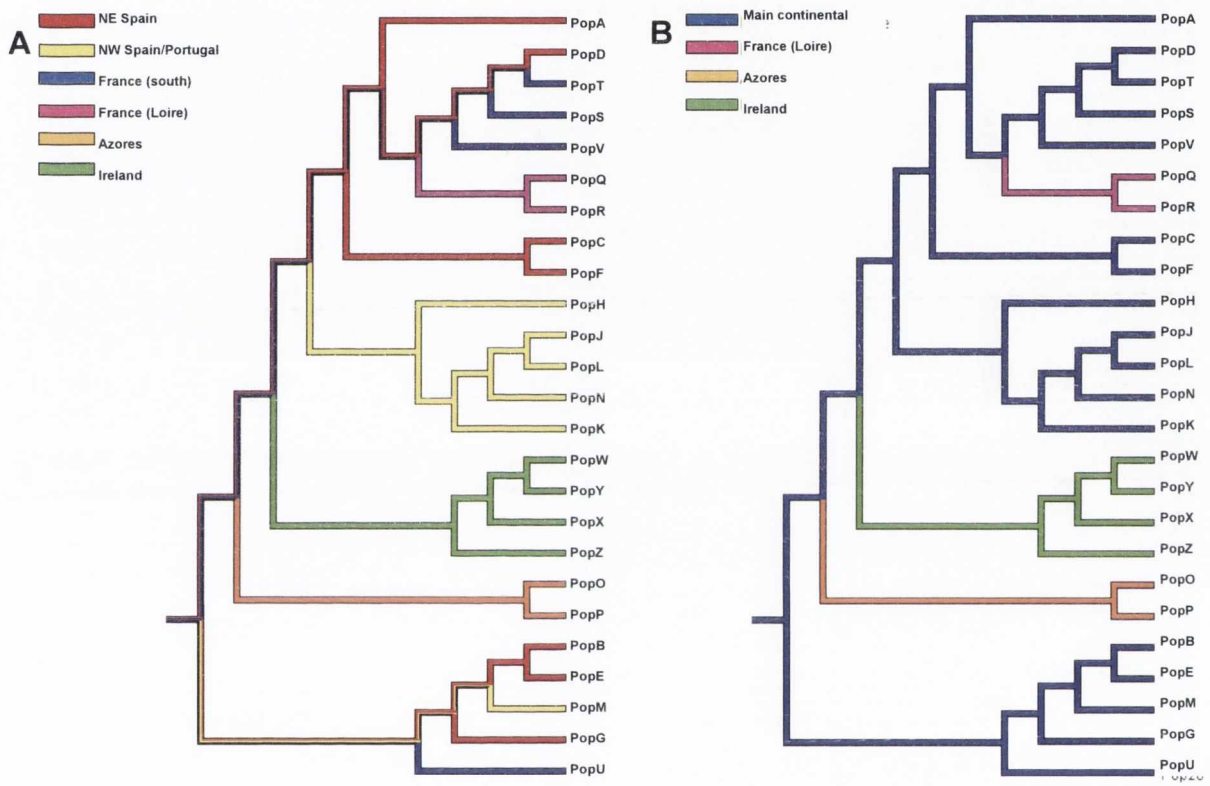


Figure 60: (A) UPGMA dendogram of genetic distance between populations of *D. cantabrica*. Colours indicate region. In (B) the same UPGMA dendogram had the NE Spain, NW Spain/Portugal and southern France regions all coloured blue.

A NJ tree was constructed for the 187 individuals across the 25 populations (). Only the Azores populations retained their clustering. The Irish individuals did not form a single cluster, but two distinct ones. Towards the top of the tree the Irish cluster contained mostly individuals from PopX and PopY, while the cluster further down contained primarily individuals from PopW and PopZ.

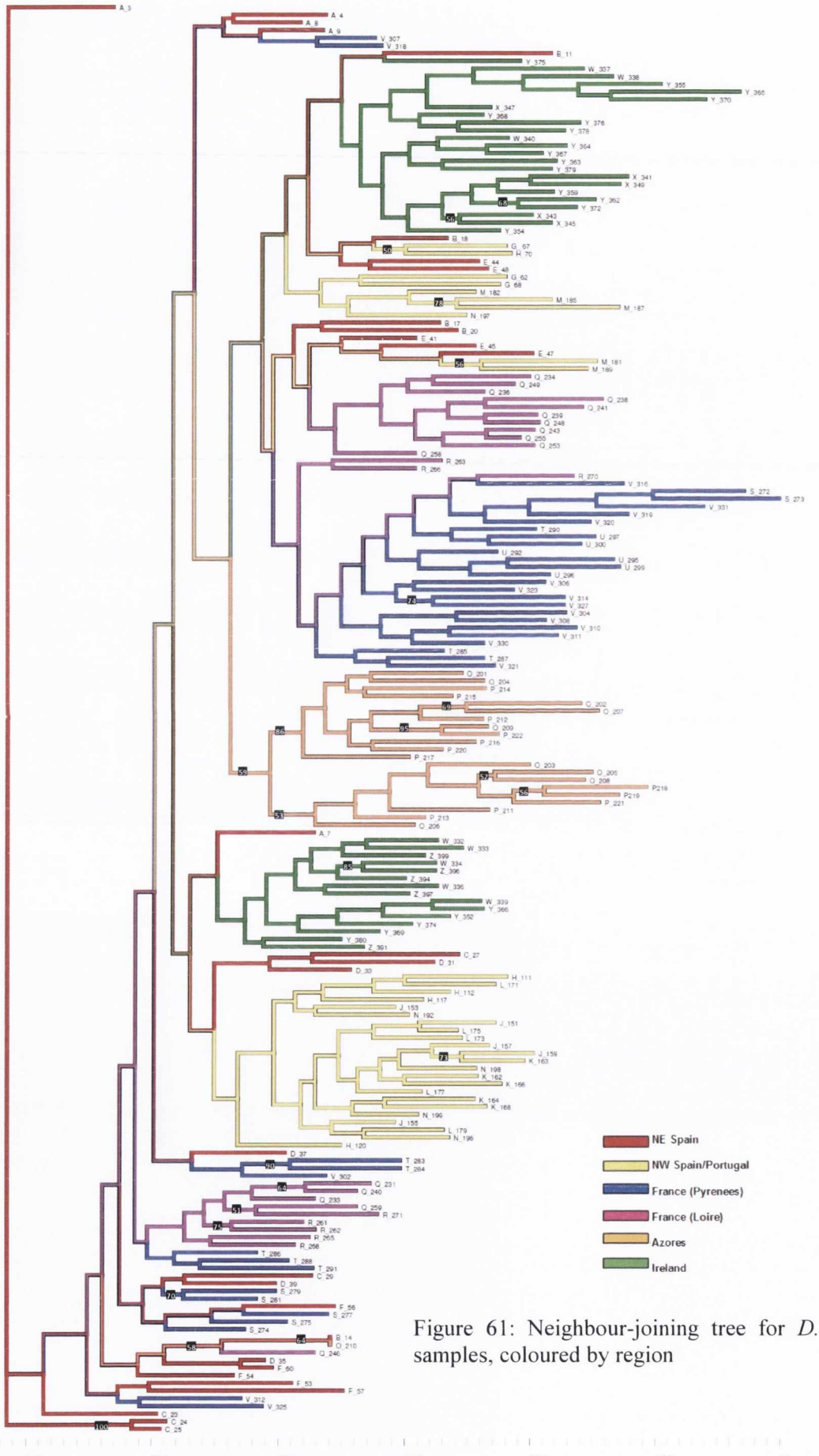


Figure 61: Neighbour-joining tree for *D. cantabrica* samples, coloured by region

The majority of the Loire individuals clustered together in a branch close to that containing mostly individuals from southern France. A few more samples from the Loire were located in a lower part of the tree, again on a branch with some individuals from southern France.

The unbiased Nei's genetic distance measures between populations (as used in the UPGMA tree) were also used to carry out a PCoA.

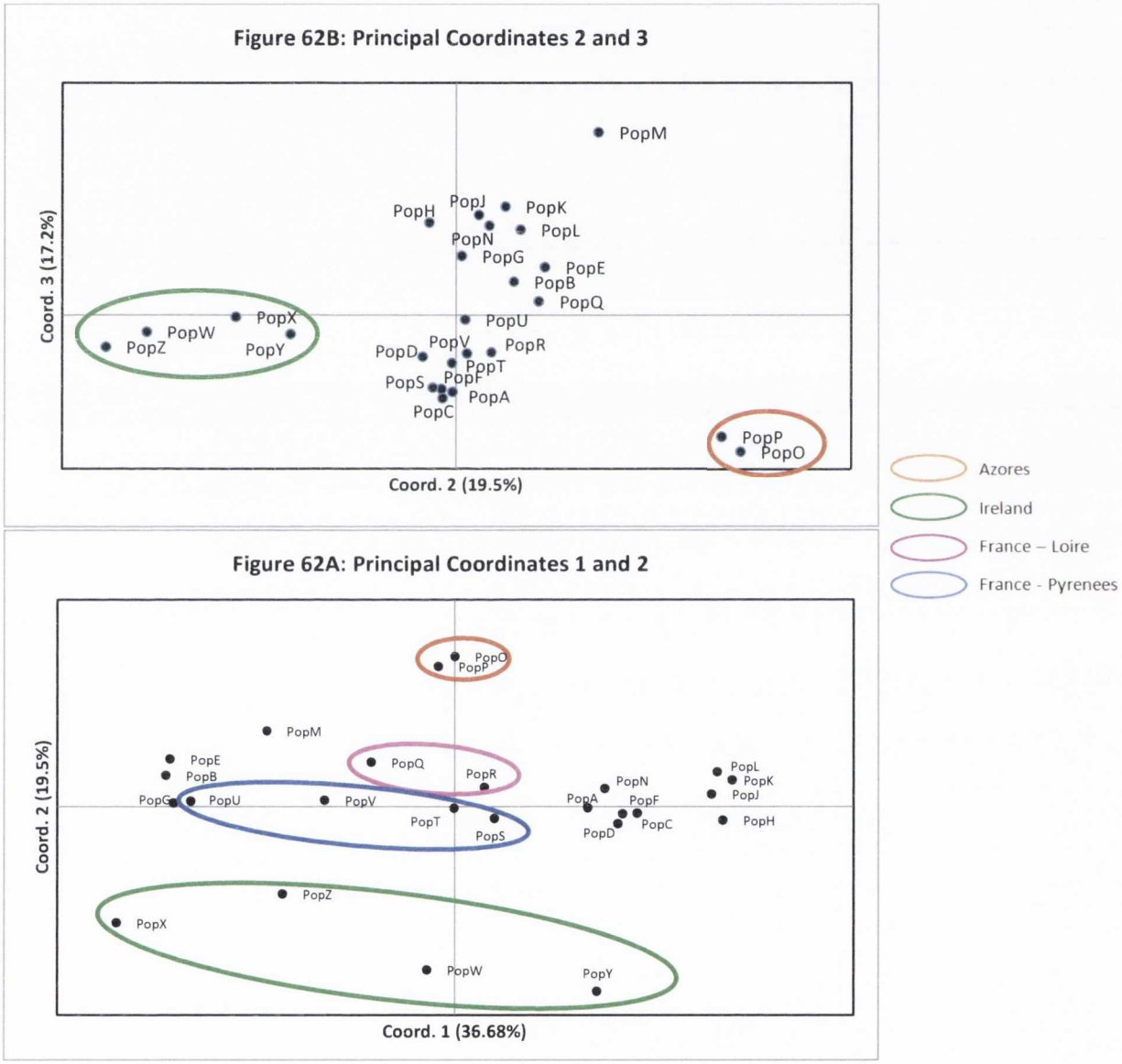


Figure 62: Principal Coordinates Analysis of the genetic distances between populations of *D. cantabrica*, based on Nei's Unbiased Genetic Distances. 61A shows Coordinates 1 and 2 (58% of the total variation), 61B shows Coordinates 2 and 3 (36.7% of the total variation). The circled groups indicate populations belonging to a particular region; Ireland (green), Azores (orange), France - Loire (pink), southern France (blue).

The first three axes account for 73% of the variation. The variation along the first two axes is depicted in Figure 6a. The second axis clearly separates out the Irish populations, as well as the Azorean populations. In Figure 6b the spread along the third axis is clear for the continental populations, whilst the Irish and Azorean populations form their own distinct clusters on either side of the Continental cluster.

The result of a Mantel test to examine the relationship between geographic distance and genetic distance of populations is shown in Figure 63. The correlation is statistically significant (p=0.01) but very small (R^2=0.1161). The large disjunction on the geographic distance axis is due to the long distance between the Azores and the rest of the populations.

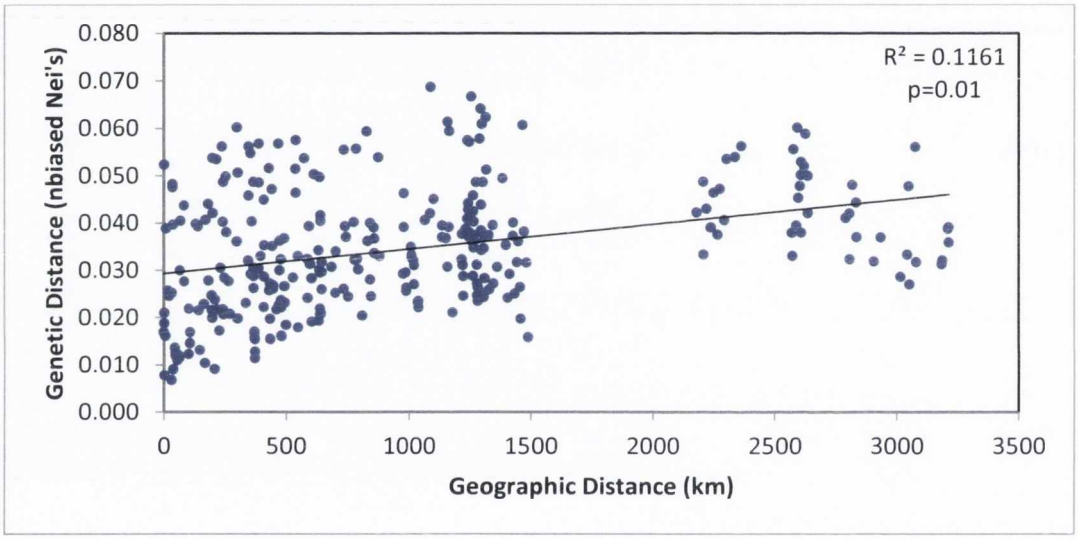


Figure 63: Mantel Test on *D. cantabrica* Nei's Unbiased Genetic Distance vs Geographical Distance

To eliminate the interference of this distance, another Mantel Test was done, omitting the Azorean populations, and working just with the Irish and Continental populations (see Figure 64). The correlation becomes weaker but is still significant with the exclusion of the Azorean populations from the Mantel Test (R^2=0.082, p=0.01).

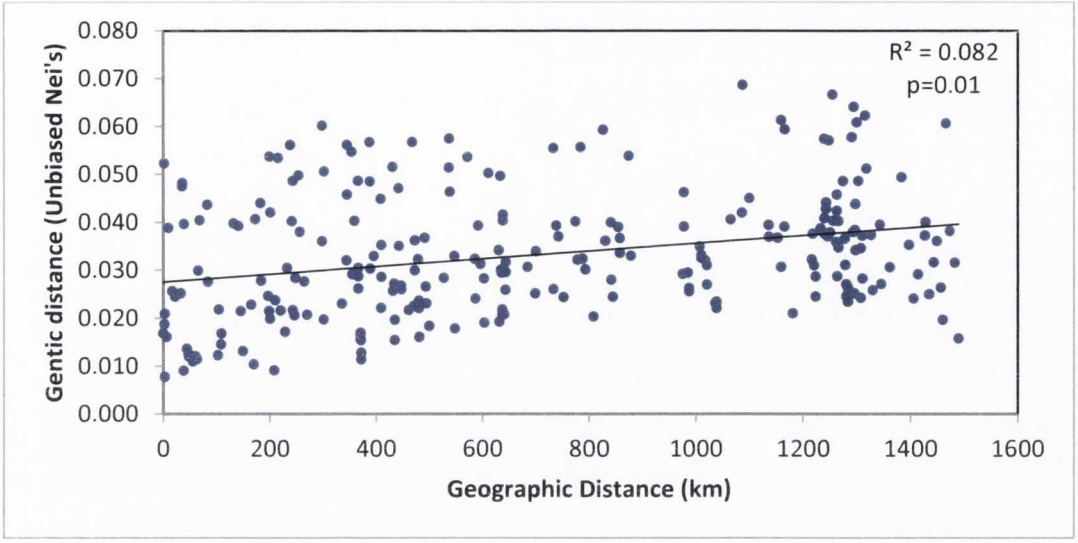


Figure 64: Mantel Test of the correlation between genetic and geographic distances of *D. cantabrica* populations (excluding the Azores).

4.3.3.2 Genetic distance between populations of E. mackaiana

A matrix of genetic distances between populations of *E. mackaiana* was generated using Nei's unbiased genetic distance statistic. The smallest genetic distance overall was between PopG and PopH which were the two populations in Galway. In the Spanish populations the smallest genetic distance was 0.023, between PopB and PopF.

The largest genetic differences were found between the Irish populations, the highest distance being between PopJ and PopK. Among the Spanish populations the highest value was 0.047, the distance between PopC and each of PopA, PopE and PopF.

Table 27: Nei's unbiased genetic differences between 11 populations of *E. mackaiana*

	PopA	PopB	PopC	PopD	PopE	PopF	PopG	PopH	PopI	PopJ	PopK
PopA	0.000										
PopB	0.029	0.000									
PopC	0.047	0.033	0.000								
PopD	0.038	0.029	0.040	0.000							
PopE	0.041	0.031	0.047	0.045	0.000						
PopF	0.026	0.023	0.047	0.034	0.037	0.000					
PopG	0.074	0.075	0.079	0.069	0.084	0.074	0.000				
PopH	0.061	0.060	0.068	0.058	0.062	0.058	0.017	0.000			
PopI	0.070	0.055	0.063	0.065	0.067	0.064	0.105	0.082	0.000		
PopJ	0.071	0.070	0.077	0.076	0.080	0.071	0.119	0.101	0.096	0.000	
PopK	0.110	0.104	0.117	0.108	0.114	0.090	0.137	0.124	0.131	0.143	0.000

THE UPGMA dendogram of the populations (Figure 65) shows the Irish populations cluster together, with the most closely related Spanish populated being PopC (Garcia Rodriguez).

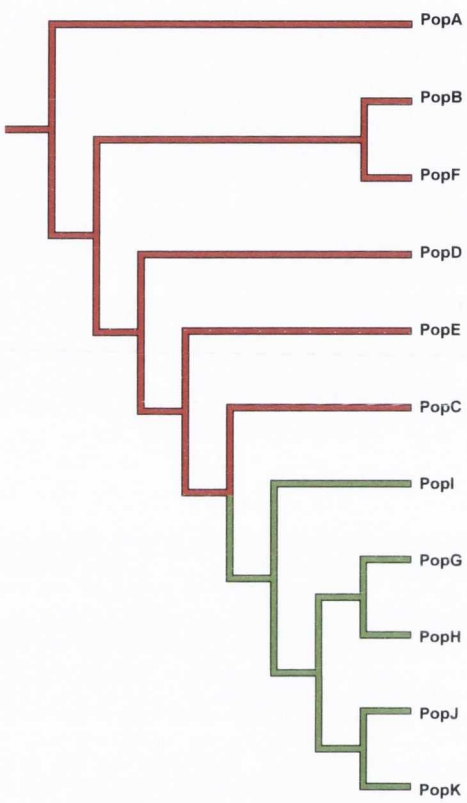


Figure 65: UPGMA dendogram of the 11 *E. mackaiana* populations. Spanish populations are indicated in red and Irish ones in green.

A neighbour joining tree was constructed of all individuals in the 11 populations, and is shown in Figure 66. Overall the Irish individuals fall into four main clusters corresponding reasonably well to their geographical populations the nodes of which are indicated on the tree;

- 1. both individuals from PopK (Mayo)
- 2. PopG and PopH (Galway)
- 3. PopI (Donegal)
- 4. Mostly PopJ, with a single individual from PopH and one from PopF

The pattern of the tree shows a two way split, with the Mayo and Galway populations on one side and the Kerry and Donegal populations on the other.

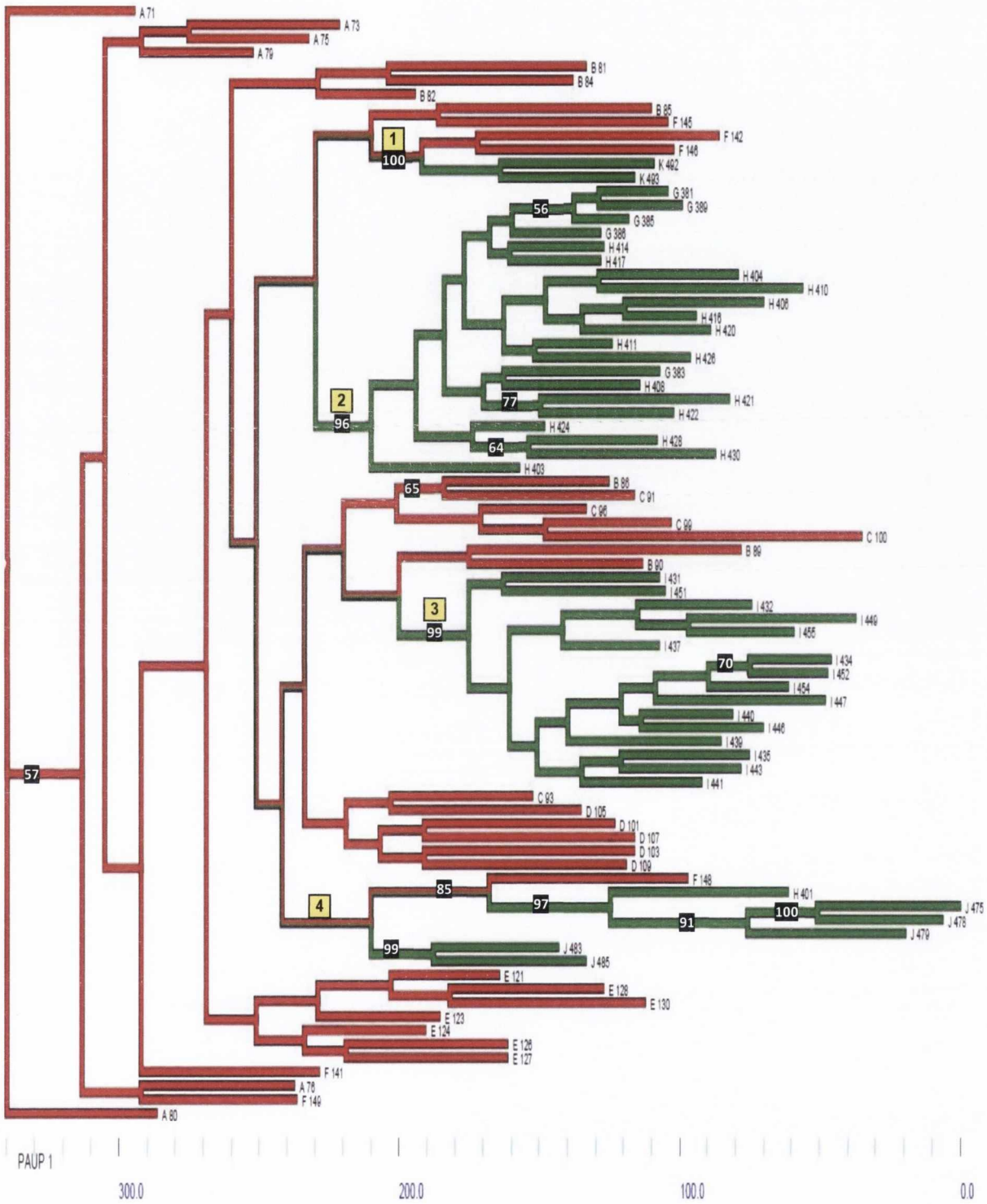


Figure 66: Neighbour-joining tree of *E. mackaiana* individuals. Irish samples are coloured green and Spanish samples are coloured red. Sample Number and Population are given for each individual

The PCoA (Figure 67) of the distance matrix between populations showed the Spanish populations clustering closely together with the Irish populations being separate in different directions. The one Irish population to cluster closely with the Spanish populations was PopI (Donegal). Among the Spanish populations PopF shows the most variation from the rest.

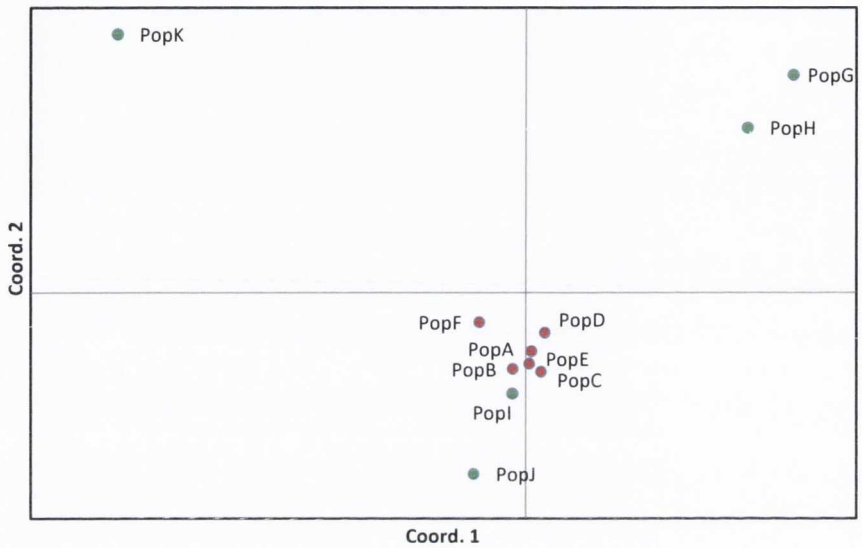


Figure 67: PCoA of *E. mackaiana* populations, Irish populations in green, Spanish populations in red.

The two Galway populations group together, but are clearly distinct from the main cluster along both axes. The Kerry (PopJ) and Donegal (PopI) populations are closely associated with the Spanish populations, being only slightly separated from them along the Coordinate 2 axis. The Mayo population is clearly separated from the main cluster along both axes.

The Mantel test showed a low level of correlation (R^2=0.0612, p=0.01) between genetic distance and geographic distance for the *E. mackaiana* populations (Figure 68).

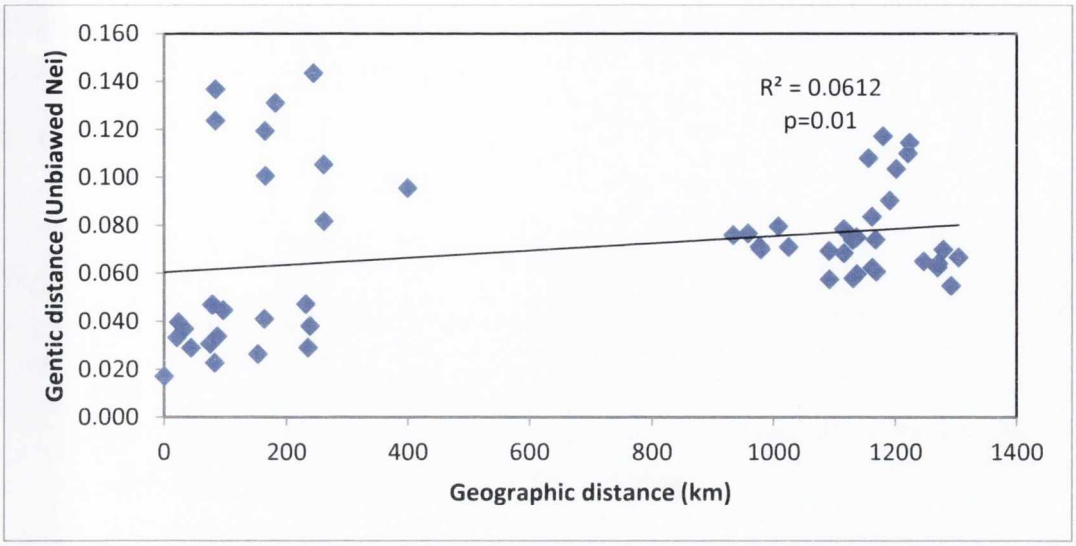


Figure 68: Mantel test for *E. mackaiana* populations

4.3.4 Population Differentiation

AMOVA was used to estimate the level of population differentiation for each species. For both species, when regional grouping were omitted, the AMOVA showed significant differentiation of populations; much more so in *E. mackaiana* (Phi_{PT}=0.201) than in *D. cantabrica* (Phi_{PT}=0.101).

4.3.4.1 Differentiation of D. cantabrica populations

AMOVA was carried out to estimate the level of population structuring in *D. cantabrica*, the results of which are summarised in Table 28. The Phi_{PT} value of 0.101 indicated a moderate differentiation between the populations of *D. cantabrica*, with a p value of 0.01.

Table 28: AMOVA of 25 *D. cantabrica* populations with no regional groupings

Source	df	SS	MS	Est. Var.	%
Between Pops	24	5116.28	213.18	13.12	10.13%
Within Pops	162	18868.65	116.47	116.47	89.87%
Total	186	23984.93		129.60	100%
Stat		Value		P	
Phi_{PT}		0.101		0.010	

Two further AMOVAs were carried out to investigate whether regional groupings also showed differentiation in the dataset. Two different regional groupings were made, one with six regions and one with 4 regions. The six regions are those shown in Figure 56; the Loire, S France, E Spain, W Spain, Azores and Ireland.

For the four-region grouping the continuous continental range (S France, NE Spain, NW Spain) was merged into one region, the other three regions remaining the same.

Table 29: AMOVA of 25 *D. cantabrica* populations with six regional groupings

Source	df	SS	MS	Est. Var.	%
Between Regions	5	2141.45	428.29	8.70	8.03%
Between Pops	19	2604.51	137.08	6.61	6.10%
Within Pops	162	15071.29	93.03	93.03	85.86%
Total	186	19817.25		108.35	100%
Stat		Value		P	
Phi_{RT}		0.080		0.010	
Phi_{PR}		0.066		0.010	
Phi_{PT}		0.141		0.010	

Table 30: AMOVA of 25 *D. cantabrica* populations with four regional groupings

Source	df	SS	MS	Est. Var.	%
Between Regions	3	1498.24	499.41	7.63	6.94%
Between Pops	21	3247.71	154.65	9.17	8.35%
Within Pops	162	15071.29	93.03	93.03	84.71%
Total	186	19817.25		109.83	100%
Stat		Value		P	
Phi_{RT}		0.069		0.010	
Phi_{PR}		0.090		0.010	
Phi_{PT}		0.153		0.010	

In both cases the within-population variation remained high, accounting for around 85% of the total variation. When the populations were grouped into six regions (Table 29) the variation between regions was higher than that between populations (PhiRT=0.080, p=0.01). However

when the three contiguous continental regions were grouped as one (Table 30), the variation between populations became greater than the variation between regions (PhiRT=0.069, p=0.01).

With both the regional groupings, the Phi_{PT} value was much higher than for when there was no regional grouping.

The output of the STRUCTURE clustering analysis was then examined at each value of K. The results of the L(K) and ΔK methods for identifying the true K are shown in Figure 69.

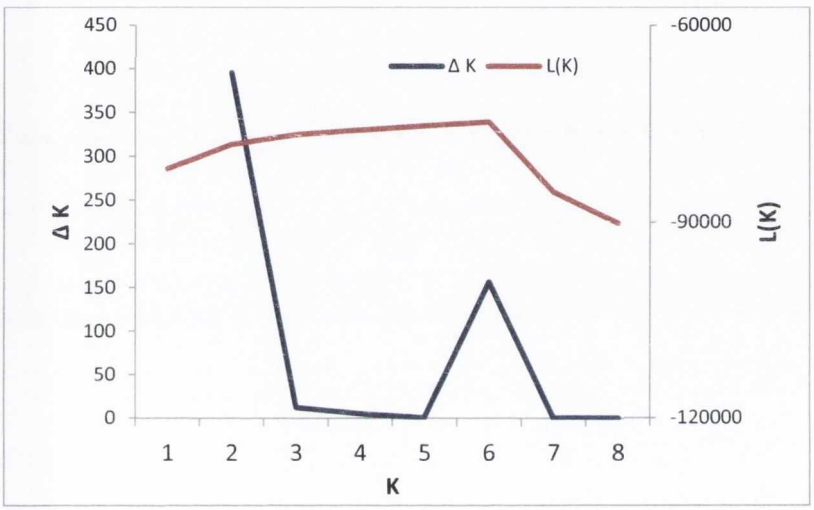


Figure 69: L(K) and Δ K plots for *D. cantabrica*

The mean L(K) increased with increasing values of K, with a plateau appearing to form at K=2 and then a fall at K=6. The ΔK vs. K plot showed peaks at K=2 and K=6. Because both of these values of K appeared to be important the STRUCTURE analysis at both values is described in Figure 70.

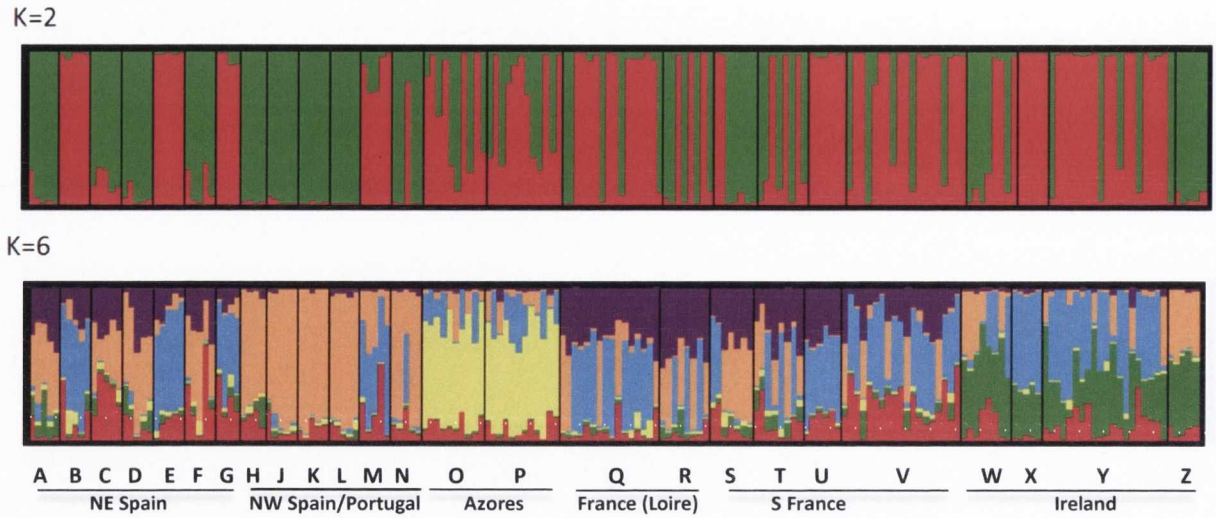


Figure 70: Population structure for *D. cantabrica* at K=2 and K=6. Individuals are grouped into their sample populations (Pop1-Pop25). Colours represent hypothetical clusters (two colours for K=2, six colours for K=6). Each individual is represented by a vertical bar, coloured proportionately according to the cluster it is assigned to.

There was no discernible pattern evident from the K=2 clustering output, but at K=6, several features were noted.

The two Azorean populations show significant differentiation from the others. The Irish populations also show clear differentiation from the rest of the populations. The populations of northwest Spain show some differentiation from the rest of the continental populations including those in the Loire.

Whilst the two Loire populations do not show any clear differentiation it is obvious they are more closely allied to the southern France and northern Spain areas, than to the northwest Spain/Portugal populations.

4.3.4.2 Differentiation of E. mackaiana populations

The populations of *E. mackaiana* showed significant population differentiation (Phi_{PT}=0.201, p=0.01), with 20% of the total variation being accounted for by variation between populations
(

Table 31).

Table 31: AMOVA of 25 *E. mackaiana* populations with no regional groupings

Source	df	SS	MS	Est. Var.	%
Between Pops	10	1682.32	168.23	15.22	20.13%
Within Pops	70	4225.68	60.37	60.37	79.87%
Total	80	5908.00		75.58	100%
Stat		**Value**		**P**	
Phi_{PT}		0.201		0.010	

The populations were then grouped into regions (Ireland and Spain) to see whether there was also differentiation between regions (Table 32).

When the Irish and continental regional groupings were taken into account the AMOVA results showed very little differentiation between regions (Phi_{RT}=0.028, p=0.01), with variation remaining moderate between populations (Phi_{PT}=0.211, p=0.01).

Table 32: AMOVA of 25 *E. mackaiana* populations with two regional groupings

Source	df	SS	MS	Est. Var.	%
Between Regions	1	276.56	276.56	2.14	2.80%
Between Pops	9	1405.76	156.20	14.01	18.30%
Within Pops	70	4225.68	60.37	60.37	78.89%
Total	80	5908.00		76.52	100%
Stat		**Value**		**P**	
Phi_{RT}		0.028		0.010	
Phi_{PR}		0.188		0.010	
Phi_{PT}		0.211		0.010	

The output of the STRUCTURE clustering analysis was then examined at each value of K. The results of the L(K) and ΔK methods for identifying the true K are shown in Figure 71.

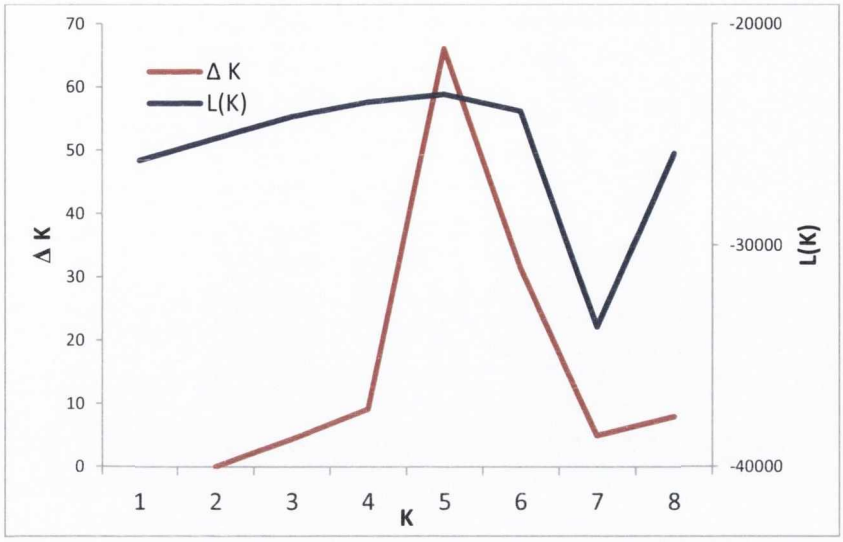


Figure 71: ΔK and L(K) plots for *E. mackaiana*

The L(K) plot shows a plateau at the K=5 point, and the ΔK plot shows a clear spike at this value of K as well. The STRUCTURE output for K=5 was therefore selected for the examination of population differentiation in the *E. mackaiana* populations (Figure 72).

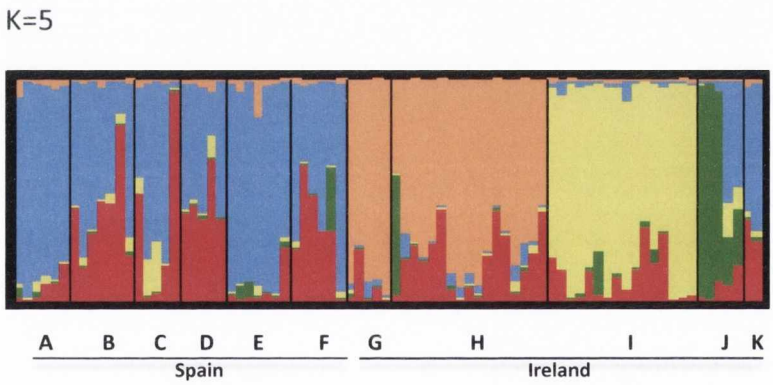


Figure 72: STRUCTURE clustering of *E. mackaiana* at K=5. Colours represent the five hypothetical clusters. Each individual is represented by a vertical bar, coloured proportionately according to the cluster it is assigned to

The output shows a clear differentiation between the Irish (PopG-K) and Spanish (PopA-F) populations, although PopJ and PopK (Kerry and Mayo) show some affinity to the Spanish populations. The two Galway populations (PopG and PopH) are clearly differentiated from anything else, as is the Donegal population (PopI). Among the Spanish populations there is less differentiation between populations. Within the Kerry population there is evidence of

some individuals but not others showing clearly assignment to a particular cluster (green) which is rarely found anywhere else.

4.4 Discussion

The focus of this chapter was on investigating the genetic diversity and genetic differentiation of, and the genetic distances between, populations of *D. cantabrica* and *E. mackaiana*.

A significant limitation of this section of the study was the under-sampling of certain populations due to time and financial constraints. The results of genetic diversity in Figure 58 and Figure 59 indicate a correlation between higher sample sizes and higher genetic diversity. The following discussion is therefore undertaken bearing in mind that differences in genetic diversity could be affected by sample size variation.

4.4.1 *D. cantabrica* discussion

4.4.1.1 Overall

Genetic diversity within populations overall is considered to be moderately low, with genetic diversity statistics ranging from 0.105 to 0.161. These figures are, however, consistent with the genetic diversity found in an earlier study of *D. cantabrica* (Kingston and Waldren 2005), which showed values of between 0.11 and 0.14 for the six populations studied.

Whilst the genetic diversity values are low they are not substantially lower than the genetic diversity of continental and Scottish populations of the widespread common heather *Calluna vulgaris* (L.) Hull, for which genetic diversity values from 0.087 to 0.232 were recorded (Mahy *et al.* 1999). Interestingly this study also showed the highest rates of genetic diversity of *C. vulgaris* occurring in the area of the Spanish-French border, which is consistent with the findings for *D. cantabrica* in this study.

Genetic distances between populations ranged from 0.007 to 0.069 using the Nei unbiased genetic distance statistic. The Mantel test showed that there was a positive correlation between geographic and genetic distance, but it was very weak (R^2=0.082 after exclusion of the oceanic effect of the Azorean populations). There were also instances of genetic similarity between geographically distant populations. Overall there would appear not to be any strong indication of an effect of isolation by distance.

Genetic differentiation between populations (Phi_{PT}=0.141) and between regions (Phi_{RT}=0.069) was significant and these values indicate a moderate level of genetic differentiation between populations and between regions (Hartl and Clark 2007). This differentiation was clearly visualised from the STRUCTURE analysis, which showed Ireland and the Azores as each being distinct from the continental populations, and which also showed a differentiation of the NW Spain & Portugal populations from the other populations on the continent.

4.4.1.2 Azores

The two Azorean populations sampled, on the islands of Faial and Pico both showed similar levels of diversity (Faial=0.139, Pico=0.133). These figures lie around the middle of the range of genetic diversities measured in the other populations across Europe.

The PCoA of genetic distances between populations showed a clear separation from the continental and Irish populations. However an inspection of the distance matrix did reveal instances of continental populations being more genetically similar to one of the Azorean populations than to another continental population.

The SRUCTURE analysis showed a clear differentiation of the Azorean populations from the other populations at K=6, and the AMOVA showed moderate differentiation between regions, one of which contained the two Azorean populations.

All three measures are concordant with a substantial length of time since isolation from the continental populations, and the subsequent genetic drift that has led to a morphologically distinct form.

Whether or not this morphologically distinct form constitutes a different species or a subspecies of *D. cantabrica* has been a matter of controversy (Nelson 1996). The fact that hybrids between the Azorean form and the 'normal' *D. cantabrica* grown in cultivation are fully fertile and backcross readily resulted in its current classification as a subspecies (McClintock 1989).

The genetic evidence discovered in this study would support this classification over that of a different species. The STRUCTURE analysis failed to clearly delimit the Azorean populations from the rest of Europe at a hypothetical cluster number of 2 (which might have been expected

had the dataset contained two separate species). Also, in both the UPGMA and the NJ tree, the Azorean populations cluster within the tree, rather than on a separate branch.

Therefore, whilst it is fair to say that Azorean populations are in many ways genetically distinct from the rest of the European populations, there is no evidence to suggest that this distinction is of sufficient magnitude to warrant recognition as a separate species. This study would therefore support maintaining its current status as *D. cantabrica* subsp. *azorica*, as proposed by McClintock (1989).

4.4.1.3 France - Loire

Genetic diversity, as measured by Nei's unbiased gene diversity statistic, for the two populations in the Loire were 0.135 and 0.129, which fall approximately midway in the total range of genetic diversity across all the populations (0.105 to 0.161).

The most genetically similar population to either of the Loire populations was Pop T (the Col d'Ibardin), in southern France. Another very small genetic distance was recorded with Pop A (Urkiola), approximately 100km away over the Spanish border.

This similarity to the southern France area is reflected in both the UPGMA and NJ trees as well. The PCoA shows the two Loire populations located within the broad continental range, not distinct from it, as the Azores and Irish populations are.

The STRUCTURE analysis does not show a clear differentiation of the Loire populations from the rest of the continental populations but it is apparent that they are much more similar to those of southern France and northern Spain than to those populations found further away in the northwest of Spain and in Portugal.

Because there is no clear genetic differentiation of these populations it is unlikely that they have been isolated for a substantial amount of time, and it is more likely that they have become established in their current locations relatively recently. Whether or not they were established naturally through long distance dispersal or through human introduction cannot be demonstrated by these genetic analyses.

4.4.1.4 Ireland

The genetic diversity of the Irish populations of *D. cantabrica* ranged from 0.109 to 0.146. Pop Y (Clifden) had the highest genetic diversity of the Irish populations, and the third highest overall (equal to Pop G in northern Spain).

Two other populations in Ireland, Pop X (Derryneen) and Pop Z (Errisbeg), showed some of the lowest genetic diversity values across all the populations.

The genetic distance measures revealed that three out of the four Irish populations were genetically most similar to populations in southern France, the other being most similar to a population in northern Spain (Pop D Cabo Penas in the Asturias).

The PCoA of genetic distance showed a clear separation of the Irish populations from the continental populations. The Irish populations do not exhibit the close clustering to each other that the Azorean populations do, showing considerable distance from each other along the first axis (PC1), although much less on the second and third axes. Moreover, the Irish populations are slightly nearer to the continental populations than the Azorean ones are. The genetic similarity to the populations of southern France rather than other regions of the continent is also evident from the PCoA.

The UPGMA tree showed the Irish populations clustering together as one group within the tree, but the NJ tree of individual genetic distances showed the Irish individuals clustered into two distinct groups. The NJ tree illustrates that although at the level of population genetic distance, Irish populations have a similarity to the southern French populations, that relationship is not replicated when genetic distances between individuals are examined.

The STRUCTURE analysis showed a clear differentiation of the Irish populations from the other populations of *D. cantabrica* on the continent and the Azores. Within the Irish populations there was some evidence of sub-structuring, as Pop Z (Errisbeg) showed a slight differentiation from the other three populations.

The genetic differentiation of the Irish populations indicates a substantial time period of isolation of these populations from the continental populations. This in itself makes it unlikely that the Irish populations were established solely as the result of human introduction.

The genetic diversity of two of the four Irish populations is very low in comparison to those on the continent, but another shows quite high diversity, and the four populations show substantial genetic distances from each other, although less so than from the continental populations as a whole. If the Irish populations as a group had a much lower genetic diversity than the continent that would indicate the existence of single founder event. However, the pattern found in this study is more indicative of multiple founder events.

The differentiation between the eastern and western parts of Spain (as shown in the STRUCTURE output in Figure 70) suggests that during the last glacial period, *D. cantabrica* persisted in two separate areas in the Iberian peninsula. This east-west pattern of differentiation has been found in other phylogeographic studies of other species (Schmitt 2007).

The genetic distance analysis revealed similarities between the Irish populations and those in southern France. Two scenarios are therefore possible. The range of *D. cantabrica* could have at one time extended up the west coast of France and into Ireland when sea-levels were lower, and the range became disjunct when sea-levels rose. The Irish populations would in this case be more similar to these southern French populations because they would have been geographically the closest ones. Alternatively, the range has always been disjunct but natural long distance dispersal events led to the establishment of populations in Ireland that had their origins in populations in southern France.

Fossil evidence of a more continuous range of *D. cantabrica* is what is required to substantiate the first scenario. In the absence of that, either hypothesis remains viable.

4.4.2 *E. mackaiana* discussion

The Irish populations of *E. mackaiana* all showed differences in the three genetic measure.

The overall genetic diversity of the *E. mackaiana* populations in Ireland was low when compared to the Spanish populations. Genetic diversity values ranged from 0.074 to 0.154 in Ireland, and from 0.141-0.198 in Spain. This in itself is indicative of a founder effect, suggesting that the Irish populations were established from a very small number of individuals.

The AMOVA showed that there was very little genetic differentiation between Spain and Ireland was low, with the majority of variation partitioned between individuals within populations or between populations. This would suggest that not a long time has passed since the Irish populations became genetically isolated from the Spanish ones. However, it must also be taken into account that due to its sterility in Ireland *E. mackaiana* reproduces mainly vegetatively and therefore low genetic diversity could also be intensified as a result of the lack of recombination.

Examining the relationship of individual populations of *E. mackaiana* in Ireland to each other and to the Spanish populations reveals patterns not easily explained.

A previous study of genetic diversity in *E. mackaiana* (Kingston and Waldren 2005) found that Irish samples did not cluster together in a UPGMA tree of individuals; instead Donegal clustered with some Spanish populations, and Galway with other Spanish populations. This study replicates that pattern, also showing the Donegal and Galway populations in separate clusters of the NJ tree. On this tree it was also shown that generally the Mayo individuals cluster with the Galway individuals and the Kerry individuals with the Donegal individuals.

This pattern corresponds very roughly to the genetic distance between populations as shown with the PCoA. In the PCoA the Donegal and Kerry populations are shown as being closer to each other than to Galway or Mayo, and were also close to the Spanish populations; the Galway and Mayo populations were shown in the PCoA as very distant both to each other and to the Spanish populations.

The STRUCTURE analysis revealed that the Mayo population clustered more strongly to the Spanish populations rather than the Irish ones. The Mayo population also had the second lowest genetic diversity (UH_e=0.093). However the Mayo population also had a very small sample size (n=2). Therefore, the inference that the Mayo population is a result of a recent introduction, whilst in concordance with the genetic evidence, is by no means conclusive.

The STRUCTURE analysis revealed the possibility of admixture in the Kerry population, with some individuals showing assignment to a cluster found almost nowhere else. This highlights the possibility (previously encountered by Kingston & Waldren (2005)) that certain individuals whilst morphologically almost indistinct from *E. mackaiana*, are in fact the hybrid

E. x *stuartii*. The Kerry population has been documented as being very densely populated with hybrids and with *E. tetralix*, with one study showing that out of 86 individuals sampled only 2 were morphologically true *E. mackaiana* (Nelson 2005). It is therefore possible that some of the Kerry individuals in this study have some level of *E. tetralix* admixture. The remaining *E. mackaiana* individuals in the Kerry sample nevertheless show a greater similarity to the Spanish populations than to any of the Irish ones, which again suggests a more recent establishment than the Galway or Donegal populations. The Kerry population is the mostly recently discovered in Ireland (2005), though this does not preclude its existence here for much longer.

In summary, whilst the genetic composition of the Irish populations of *E. mackaiana* are confused somewhat by the possibility of hybrid contamination in the samples, and small sample sizes, they would appear to have established independently, the Mayo and Kerry populations are likely the most recent to have become established, whilst the Donegal and Galway populations show indications of a more lengthy separation from the Spanish populations.

CHAPTER 5 General Discussion

This chapter will begin with a very brief summary of what each of the previous chapters dealt with and what the overall findings were. The next section will synthesise in more detail how the findings for each species contributed to establishing a clearer picture of the processes and events that led to the establishing of the Irish populations of *D. cantabrica* and *E. mackaiana* at such a great disjunction from the populations in the Iberian peninsula.

5.1 Summary

Chapter 1 was an introductory chapter in which the taxonomy, morphology and ecology of *D. cantabrica* and *E. mackaiana* was described, as well as their disjunct distribution which mirrors that of a few other floral and faunal species, and which has been described in this study as 'Hiberno-Iberian'. The different scenarios for how this range disjunction may have arisen were described, within the context of the glacial and post-glacial landscape of Western Europe.

Three principal lines of enquiry were then pursued in the next three chapters, each of which aimed to contribute to the elucidation of which scenario or scenarios are the most plausible in explaining the distribution of *D. cantabrica* and *E. mackaiana* in Europe, with a particular emphasis on how and when their establishment in Ireland occurred.

5.1.1 Chapter 2

In Chapter 2 the location data of extant populations of both species across their Irish and continental ranges were linked to the bioclimatic factors and soil characteristics at these locations. From this a species distribution model was constructed and used to predict the full extent of the potential range of each species under current conditions, conditions at the Last Glacial Maximum and at the Last Interglacial, and conditions 70 years into the future.

The principal findings of this chapter were:

- the Irish populations occupy a range of climatic conditions which are substantially different from those occupied by the continental populations

- the potential range of each species under current climatic conditions is more extensive than the actual range in both the continent and in Ireland. This suggests that other factors are limiting range expansion, such as soils, vegetation and land use.
- hindcasting the model to the LGM shows the potential range of both species occurring in a continuous region along the Atlantic coast to the west coast of Ireland
- glacial extents and predicted potential ranges show that there were proximal refugial areas where both species could possibly have either survived the last glaciation, or become established early in the late glacial.
- forecasting to the future shows that even under conservative climate change scenarios, the range of suitable conditions on the continent will drastically diminish, whilst those in Ireland however remain largely unchanged.

5.1.2 Chapter 3

Chapter 3 dealt with the lineage of DNA sequences of two chloroplast and one nuclear region of the genome, all of which were non-coding and therefore selectively neutral. From these sequences, haplotypes were developed. Haplotype networks were constructed to demonstrate the relationships between haplotypes, and the distribution of these haplotypes across the geographical range of each species was mapped.

The principal findings of this chapter were:

- across the two species 17 chloroplast haplotypes and 30 nuclear (ITS) haplotypes were found, consistent with the higher rate of substitution in nuclear DNA.
- for *D. cantabrica* 13 chloroplast haplotypes and 26 nuclear (ITS) haplotypes were identified. Six of the ITS haplotypes were unique to Ireland.
- the number of haplotypes identified for *E. mackaiana* was much lower, consisting of 4 chloroplast and 4 ITS haplotypes. One chloroplast haplotype and two ITS haplotypes were unique to Ireland.

5.1.3 Chapter 4

In Chapter 4 AFLPs were used to determine the extent of genetic diversity and genetic differentiation in continental and Irish populations of both species. The AFLP data were also

used to determine genetic distances between individuals and between populations, and these distances were depicted on UPGMA and Neighbour-Joining trees.

The principal findings of this chapter were:

- Genetic differentiation between the Irish and continental populations was evident for both species
- Genetic diversity of the Irish populations was comparable to that of the continent in the case of *D. cantabrica* and somewhat lower for *E. mackaiana*
- There was a small yet significant correlation between genetic distance and geographic distance between populations.

5.2 Synthesis of findings

5.2.1 Synthesis of findings for *D. cantabrica*

The continental range of *D. cantabrica* showed some evidence of genetic differentiation among regions, with the northwestern part of the Iberian peninsula demonstrating differentiation from the main continental range consisting of the rest of Spain and the south of France. The Loire populations also showed no significant differentiation from the main continental range.

The Azores and Ireland however both showed clear genetic differentiation from the continental populations. Further findings of the two molecular chapters give strong indications that the Irish populations, as well as those of the Azores, have been separate from the current continental populations for a significant period of time.

Genetic diversity in the Irish populations was low, but at a similar level to the continental populations, and not indicative of a recent founding event. This finding was consistent with earlier work done with AFLPs.

Although DNA sequencing revealed no unique chloroplast haplotypes for Ireland, it did show that the Irish populations have two haplotypes which are commonly found on the continent,

and from these a further six appear to have been generated, which are not found anywhere except Ireland. This is not compatible with a recent introduction, but suggests one or more far earlier founding events, of at least 59,000 years, and possibly longer.

The two ITS haplotypes shared between Ireland and the continent are common across the continental range so this did not give any indications as to likely areas of provenance of the Irish populations. However, the AFLP analysis revealed close genetic distances between Irish populations and certain populations in the southern France region.

Populations in Ireland show both high and low genetic diversity, overall the picture is not one of a greatly reduced diversity compared to the continent, which makes it unlikely that a single founder event led to the establishment of the Irish populations.

The pattern of chloroplast haplotypes showed clear evidence of the provenance of the Azorean populations, showing that the lineage can be traced back to the northwest area of the Iberian peninsula. Although the Azorean populations show a clear genetic differentiation from the continental populations, distance measures showed that in terms of genetic distance they were not substantially more distant genetically than other populations within the continent. The current taxonomic status of the Azorean populations as a subspecies of *D. cantabrica* was deemed more accurate than its proposed categorisation as a separate species.

The Loire populations were shown to not be genetically differentiated from the main continental range. One unique chloroplast haplotype and no unique ITS haplotypes were identified in this region. Genetic distance analysis showed the Loire populations had close genetic relationships with populations in southern France.

The species distribution modelling showed that the Loire populations lie either outside or on the border of areas predicted as suitable. Taken together the molecular and climate modelling results indicate that the Loire populations are much more recently established than the Irish or Azorean populations.

Chapter 2 also showed that Irish populations of *D. cantabrica* occupy a niche very different to that of the continent, thereby showing that a significant process of adaptation has taken place which has allowed them to establish so successfully in Ireland.

Hindcasting of the species distribution model to the LGM shows possible corridor of suitable climatic conditions up the Atlantic coastline and across the landbridge to areas just offshore of their current range, although this does not take into account possible soil limiting factors at that time.

Evidence of their existence in Ireland (but not in their current location) prior to the LGM is know from fossil records dated to the Gortian Interglacial. Hindcasting of the species distribution model to the Last Interglacial corresponds well to two known locations of these fossils.

Taken altogether the results indicate two possible scenarios are likely for the Irish populations. The first possibility is that the current Irish populations became established from postglacial colonisation along the Atlantic coast of France and across the landbridge with intervening populations since becoming extinct. The current populations in the Loire are most likely not remnants of this colonisation process, but are more recently established from populations in the south of France.

The second possibility is that the range of *D. cantabrica* has been disjunct since before the LGM or during late glacial times, and that the current Irish populations were established by long distance dispersal events from the Iberian peninsula to unglaciated areas off the current Irish coastline.

Either way, the combined evidence supports the fact that they are native to Ireland, and are not recent human-mediated introductions.

5.2.2 Synthesis of findings for *E. mackaiana*

The molecular analysis of the *E. mackaiana* populations was complicated by smaller sample sizes and fewer populations than were analysed for *D. cantabrica*. Therefore the results are less conclusive and should be considered with the methodological limitations described in Section 5.4 borne in mind.

The DNA sequencing and haplotype analysis identified one chloroplast haplotype (in Kerry), and two ITS haplotypes (Mayo and Donegal) which are unique to Ireland, and were not identified in any continental population. This in itself suggests different origins for the Irish populations and minimal if non-existent gene flow between populations.

AFLP analysis revealed a possible founder effect in the Irish populations, evident in their lower genetic diversity compared to continental populations. Overall there was less of a differentiation between the continental and the Irish populations than was evident for *D. cantabrica*, which suggests that *E. mackaiana* in Ireland has experienced a more recent separation from the continent. However, it must be borne in mind that their sterility in Ireland would also contribute to a lower genetic diversity, due to a lack of recombination.

The inference that the different populations in Ireland arose from different founding events was also supported by the Neighbour-joining tree based on genetic distance, with the Irish individuals did not cluster together, but instead were roughly formed into two groups, one with the Donegal and Kerry populations, the other with the Galway and Mayo populations.

The STRUCTURE analysis of population differentiation showed signs that the Mayo and Kerry populations contained individuals more closely associated to Spanish individuals than did the Donegal and Galway populations, which suggests that they are more recently separated from the continent.

The findings of Chapter 2 were very similar to those patterns described for *D. cantabrica*; the Irish *E. mackaiana* have demonstrated an adaptation to an environment very different to that in their continental range, although the fact that *E. mackaiana* is almost completely sterile in Ireland factor leads one to question whether the adaptation is all that successful.

Similarly to *D. cantabrica* the species distribution model predicted a potential range under current conditions that is more extensive than that currently occupied. Hindcasting to the LGM shows possible corridor of suitable climatic conditions up the Atlantic coastline and across the land bridge to areas just offshore of their current range.

Overall the patterns revealed are not conclusive, but it would seem most likely that *E. mackaiana* became established in Ireland in postglacial times. Whilst the possibility of human introduction (particularly for the Kerry and Mayo populations) is not entirely ruled out, the three unique haplotypes identified strongly suggests a much earlier establishment in Ireland than the last few hundred years.

5.2.3 Synthesis of findings for *E.* x *stuartii*

A hybrid plant, fitting the morphological description of *E.* x *stuartii* had chloroplast and ITS sequences compared to those of putative parents *E. tetralix* and *E. mackaiana.*

In the hybrid individual sequenced, *E. mackaiana* was the maternal parent. Whether this directionality of the hybridisation is consistent for all *E.* x *stuartii* can only be determined by greater sampling in this population and across the other populations in Ireland.

5.3 European postglacial colonisation and the Iberian peninsula

The Iberian Peninsula is one of the three main Pleistocene glacial refugia in the European subcontinent (Hewitt 2000). Phylogeographic studies on a wide variety of plants and animals have demonstrated that the Iberian peninsula was an important source for western and northern European colonisation after the ice ages (Gomez and Lunt 2007).

Ireland's location on the far-western fringe of Europe, and its insularity from the continent due to postglacial sea-level rise has largely determined its floristic profile, which is essentially depauperate compared to continental western Europe. It has been estimated that one third of the Holocene tree genera in north-western Europe did not succeed in reaching Ireland during this northward colonisation since the LGM (Mitchell 2006).

Whilst the majority of the Irish biota reached Ireland through Britain, there is evidence of a growing number of species whose genetic relationships indicate that a more direct route of colonisation directly from southern Europe, and specifically the Iberian peninsula. Haplotype analysis of white oaks in Ireland for example showed no evidence of a route through Britain, but instead suggested a southern Irish entry point directly from Iberia (Kelleher *et al.* 2004).

In animals the same pattern has also been found for a variety of species including Irish hares (Hamill *et al.* 2006), the natterjack toad (Rowe *et al.* 2006) and *Cepaea nemoralis* snails (Davison 2000), all of which revealed high genetic divergence between British and Irish populations and provide support for separate routes of colonisation into Britain and Ireland during postglacial times.

It is therefore very likely that both of the species in this study also entered Ireland not through Britain, but as a direct process of colonisation from the Iberian peninsula, and their absence

from Britain in the present-day is due to their never having been established there at all since the last ice age.

There is now growing evidence that many species existed in separate 'refugia within refugia' i.e. disjunct refugia areas within the Iberian peninsula (Gomez and Lunt 2007). One of the predominating patterns has been a strong genetic differentiation into eastern and western lineages (Schmitt 2007). Whilst the population structure did not allow firm conclusions to be drawn regarding *E. mackaiana*, an east-west differentiation was shown in the AFLP analysis of the Iberian populations of *D. cantabrica*, and Irish populations showed a closer affinity to the eastern Iberian populations than to the western ones.

Ongoing efforts to date and delimit the extent of the last British-Irish ice sheet have now effectively ruled out the possibility of an in-situ or proximal glacial refuge in Ireland. Other phylogeographic studies which have implicated an in-situ glacial refuge as the source location of current Irish populations will have to be revisited in the light of up-to-date knowledge on the glacial extent.

5.4 Implications

The fact that there is strong evidence for their status as Irish natives, as well as the fact that climate change scenarios indicate a significant impact on their range in continental Europe, means that both these species should be afforded some legal protection here in Ireland, which currently they are not.

There are two mechanisms by which plants can be given legal protection in Ireland; firstly under the national Flora Protection Order 1999, or secondly, under EU legislation as part of Annex II or IV of the EU Habitats Directive. Annex II lists species of community interest whose conservation requires the designation of special areas of conservation. Annex IV lists species of community interest in need of strict protection.

Currently the only specific recognition of either species is *E. mackaiana*'s listing as a 'Rare' species (rather than threatened or vulnerable) in the Irish Red Data book (Curtis and McGough 1988). Those individuals or populations which occur in Special Areas of Conservation (SACs) are protected, but this does not protect the entire range of either species in Ireland.

Work is underway to update the red data lists, and preliminary consultations have proposed the inclusion of *D. cantabrica* on the list, as well as the conversion of *E. mackaiana*'s status from Rare to Vulnerable. The findings of this study support the inclusion of both species on the red data lists, as well their protection under either the Flora Protection Order or the EU Habitats Directive.

5.5 Methodological limitations

5.5.1 Sampling

In Chapters 3 and 4 results of the molecular analysis must be taken in context of the extent of sampling. Efforts were made to cover as much of the continental range as possible, however the extremely large distances involved, the costs, and the limited time frame of one field trip meant that coverage was not as extensive as it could have been.

This is especially true for of *E. mackaiana*, which is definitely under-sampled in the eastern part of its continental range. Therefore care should be taken when considering such things as certain haplotypes being unique to Ireland, which may simply have not been sampled in the continental populations.

In the Azores, although *D. cantabrica* is known to grown on Sao Jorge, no populations could be located due to time limitations and extremely bad weather.

The Forêt de Brissac population in the Loire also could not be located from directions provided by the Conservatoire Botanique, and follow-up requests for the provision of DNA samples were unsuccessful.

No samples were taken from the British naturalised populations of *D. cantabrica*, as these are assumed to have been established from plants in cultivation, and therefore uninformative to the study. However, this should perhaps have been verified within the context of this project.

5.5.2 Sample sizes

Molecular work is inherently expensive, and the cost of reagents (in particular for the AFLPs), as well as the per-sample cost of sequencing and fragment analysis meant that only a subset of the total number of samples actually collected were analysed.

This problem was compounded by the fact that the AFLP protocol proved extremely challenging to carry out successfully, with a great deal of time and a large amount of reagents being used up in failed runs.

By the time they started working there was severe financial restriction, on reagents and the cost of fragment analysis, and therefore the number of samples finally analysed were not as many as had been envisaged.

In many cases populations were represented by less than 5 individual samples; which can impact greatly on measures such as gene diversity or the number of haplotypes within a population or within a region. This therefore should be taken into account when considering the results of both AFLPs and the haplotype analysis.

As far as was possible, with the financial constraints that existed, molecular analysis was carried out for every population sampled, even though the number of individual samples within each populations was sometimes lower than five.

5.5.3 Molecular dating methods

The molecular dating method used in the study was no more than a very rough approximation, using published general rates of substitution for chloroplast and nuclear DNA. Not only are the lineages of the sequences uncalibrated by fossil or geographical evidence, but it is also not known if the substitution rates are constant on all branches.

5.6 Future Research

The gaps in the fossil record are a clear missing puzzle piece in the story of the Hiberno-Iberian flora. Ericaceous pollen is known for being difficult to identify to species level and many published studies have not attempted to do so. It would be worthwhile to re-examine some of these fossil samples with the current Scanning Electron Microscopy (SEM)

techniques now available, to strengthen the fossil record of *D. cantabrica* and *E. mackaiana*, and in doing so build up a clearer picture of past distributions of both species.

A detailed study of the reproductive biology of both species is important in terms of their conservation in Ireland. In particular it would be valuable to have a clearer picture of their pollination strategies, seed dispersal capabilities and vegetative reproduction. The mechanisms underlying the sterility of *E. mackaiana* in Ireland will also be informative to conservation strategies. It would also be of interest to examine more closely the directionality of the *E.* x *stuartii* hybrid, and investigate the frequency and direction of back-crossing with either of the parental species.

5.7 Concluding remarks

Taken in total the results of the study indicate that both species have been present in Ireland for a considerable time period, which precludes the possibility of human introduction, although the status of the Kerry and Mayo *E. mackaiana* populations is less clear. More fossil evidence and more detailed molecular dating of DNA sequence divergence would be needed to narrow down the timescale of establishment in Ireland for both of these species.

However, since they almost certainly do not have their origins in anthropogenic introduction indicates that they are native to Ireland, and therefore an important component of the Irish flora. The fact that their continental range appears to be threatened by climate change, and their ability to disperse is constrained by soils means that Ireland could become a stronghold for both species in Europe.

The protection of these two species in Ireland is therefore essential. Neither species is currently afforded any legal conservation protection. It is therefore recommended that both be given protected conservation status under either the Irish Floral Protection Order, or the EU Habitats Directive.

This study has elucidated some of the dynamics underlying the distribution of these two species of the Irish flora, as well as contributed to the wider body of knowledge on the postglacial floral recolonisation of Western Europe.

References

Adhikari, D., S. K. Barik, and K. Upadhaya. 2012. Habitat distribution modelling for reintroduction of *Ilex khasiana* Purk., a critically endangered tree species of northeastern India. Ecological Engineering **40**:37-43.

Agrawal, A. A. 2001. Ecology - Phenotypic plasticity in the interactions and evolution of species. Science **294**:321-326.

Aigoin, D. A., N. Devos, S. Huttunen, M. S. Ignatov, J. M. Gonzalez-Mancebo, and A. Vanderpoorten. 2009. And if Engler was not completely wrong ? Evidence for multiple evolutionary origins in the moss flora of Macaronesia. Evolution **63**:3248-3257.

Alvarez, I. and J. F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution **29**:417-434.

APGIII. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society **161**:105-121.

Araujo, M. B., M. Cabeza, W. Thuiller, L. Hannah, and P. H. Williams. 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. Global Change Biology **10**:1618-1626.

Avise, J. C. 1994. Molecular markers, natural history and evolution. Chapman & Hall, New York.

Bakker, F. T., F. Breman, and V. Merckx. 2006. DNA sequence evolution in fast evolving mitochondrial DNA nad1 exons in Geraniaceae and Plantaginaceae. Taxon **55**:887-896.

Banks, M. A. and W. Eichert. 2000. WHICHRUN (version 3.2): A computer program for population assignment of individuals based on multilocus genotype data. Journal of Heredity **91**:87-89.

Bannister, P. and A. Polwart. 2001. The frost resistance of ericoid heath plants in the British Isles in relation to their biogeography. Journal of Biogeography **28**:589-596.

Barker, N. P., I. von Senger, S. Howis, C. Zachariades, and B. S. Ripley. 2005. Plant phylogeography based on rDNA ITS sequence data: two examples from the Asteraceae. Pages 217-244 *in* F. T. Bakker, L. W. Chatrou, B. Gravendeel, and P. B. Pelser, editors. Plant Species-Level Systematics: New Perspectives on Pattern & Process. A R G Gantner Verlag K G, Koenigstein.

Beatty, G. E. and J. Provan. 2011. Phylogeographic analysis of North American populations of the parasitic herbaceous plant *Monotropa hypopitys* L. reveals a complex history of range expansion from multiple late glacial refugia. Journal of Biogeography **38**:1585-1599.

Beaumont, L. J., L. Hughes, and A. J. Pitman. 2008. Why is the choice of future climate scenarios for species distribution modelling important? Ecology Letters **11**:1135-1146.

Beebee, T. J. C. and G. Rowe. 2004. An Introduction to molecular ecology. Oxford University press, New York.

Beerling, D. J., B. Huntley, and J. P. Bailey. 1995. Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surfaces. Journal of Vegetation Science **6**:269-282.

Beheregaray, L. B. 2008. Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. Molecular Ecology **17**:3754-3774.

Bentham, G. and J. D. Hooker. 1876. Genera plantarum: ad exemplaria imprimis in herberiis Kewensibus servata definita. A. Black, London.

Bertran, P., G. Allenet, T. Ge, F. Naughton, P. Poirier, and M. F. S. Goni. 2009. Coversand and Pleistocene palaeosols in the Landes region, southwestern France. Journal of Quaternary Science **24**:259-269.

Birks, H. J. B. and M. E. Ransom. 1969. An interglacial peat at Fugla Ness, Shetland. New Phytologist **68**:777-796.

Brito, J. C., E. G. Crespo, and O. S. Paulo. 1999. Modelling wildlife distributions: Logistic multiple regression vs overlap analysis. Ecography **22**:251-260.

Caballero, A., H. Quesada, and E. Rolan-Alvarez. 2008. Impact of Amplified Fragment Length Polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. Genetics **179**:539-554.

Calonje, M., S. Martin-Bravo, C. Dobes, W. Gong, I. Jordon-Thaden, C. Kiefer, M. Kiefer, J. Paule, R. Schmickl, and M. A. Koch. 2009. Non-coding nuclear DNA markers in phylogenetic reconstruction. Plant Systematics and Evolution **282**:257-280.

Carine, M. A. and H. Schaefer. 2010. The Azores diversity enigma: why are there so few Azorean endemic flowering plants and why are they so widespread? Journal of Biogeography **37**:77-89.

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

Chase, M. W. and H. H. Hills. 1991. Silica gel: an ideal material for field preservation of leaf samples for DNA studies. Taxon **40**:215-220.

Chiverrell, R. C. and G. S. P. Thomas. 2010. Extent and timing of the Last Glacial Maximum (LGM) in Britain and Ireland: a review. Journal of Quaternary Science **25**:535-549.

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

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology **9**:1657-1659.

Collins, W. D., C. M. Bit, M. L. Blackmon, G. B. Bonan, C. S. Bretherton, J. A. Carton, P. Chang, S. C. Doney, J. J. Hack, T. B. Henderson, J. T. Kiehl, W. G. Large, D. S. McKenna, B. D. Santer, and R. D. Smith. 2006. The community climate system model (CCSM3) Journal of Climate **19**:2122–2143.

Comes, H. P. and J. W. Kadereit. 1998. The effect of Quaternary climatic changes on plant distribution and evolution. Trends in Plant Science **3**:432-438.

Corander, J., J. Siren, and E. Arjas. 2008. Bayesian spatial modeling of genetic population structure. Computational Statistics **23**:111-129.

Corbet, G. B. 1961. Origin of British insular races of small mammals and of Lusitanian fauna. Nature **191**:1037-1040.

Cox, C. B. and P. D. Moore. 2000. Biogeography, an ecological and evolutionary approach. 6th edition. Blackwell Science, Oxford.

Cox, H. T. 1948. Studies in the comparative anatomy of the Ericales 1. Ericaceae - subfamily Rhododendroideae. American Midland Naturalist **39**:220-245.

Coxon, P. 1996. The Gortian temperate stage. Quaternary Science Reviews **15**:425-436.

Crawford, R. M. M. 2000. Ecological hazards of oceanic environments. New Phytologist **147**:257-281.

Curtis, T. 2000. The change in the legal status of *Erica ciliaris* L. Dorset heath in Ireland. Yearbook of the Heather Society:89-92.

Curtis, T. G. F. and H. N. McGough. 1988. The Irish Red Data Book 1. Vascular Plants. Wildlife Service Ireland, Dublin.

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.

Dawson, K. J. and K. Belkhir. 2001. A Bayesian approach to the identification of panmictic populations and the assignment of individuals. Genetical Research **78**:59-77.

DeArce, M. 2008. Darwin's Irish correspondence. Biology and Environment-Proceedings of the Royal Irish Academy **108B**:43-56.

Désamor, A., B. Laenen, N. Devos, M. Popp, J. M. González-Mancebo, M. A. Carine, and A. Vanderpoorte. 2011. Out of Africa: north-westwards Pleistocene expansions of the heather *Erica arborea*. Journal of Biogeography **38**:164-176.

Despres, L., L. Gielly, W. Redoutet, and P. Taberlet. 2003. Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. Molecular Phylogenetics and Evolution **27**:185-196.

Diekmann, K., T. R. Hodkinson, and S. Barth. 2012. New chloroplast microsatellite markers suitable for assessing genetic diversity of *Lolium perenne* and other related grass species. Annals of Botany.

Don, D. 1834. An attempt at a new arrangement of the Ericaceae. Edinburgh New Philosophical Journal **17**:150-160.

Douglas, N. A., W. A. Wall, Q. Y. Xiang, W. A. Hoffmann, T. R. Wentworth, J. B. Gray, and M. G. Hohmann. 2011. Recent vicariance and the origin of the rare, edaphically specialized Sandhills lily, *Lilium pyrophilum* (Liliaceae): evidence from phylogenetic and coalescent analyses. Molecular Ecology **20**:2901-2915.

Douzery, E. J. P., A. M. Pridgeon, P. Kores, H. P. Linder, H. Kurzweil, and M. W. Chase. 1999. Molecular phylogenetics of *Diseae* (Orchidaceae): A contribution from nuclear ribosomal ITS sequences. American Journal of Botany **86**:887-899.

Drude. 1897. Ericaceae.*in* A. Engler and K. Prantl, editors. Die Natürlichen Pflanzenfamilien nebst ihren Gattungen und wichtigeren Arten, insbesondere den Nutzpflanzen, unter Mitwirkung zahlreicher hervorragender Fachgelehrten begründet. W. Engelmann, Leipzig.

Drummond, A. J., B. Ashton, S. Buxton, M. Cheung, A. Cooper, C. Duran, M. Field, J. Heled, M. Kearse, S. Markowitz, R. Moir, S. Stones-Havas, S. Sturrock, T. Thierer, and A. Wilson. 2011. Geneious Pro™. Page Available from http://www.geneious.com/.

Duffy, K. J., G. Scopece, S. Cozzolino, M. F. Fay, R. J. Smith, and J. C. Stout. 2009. Ecology and genetic diversity of the dense-flowered orchid, *Neotinea maculata*, at the centre and edge of its range. Annals of Botany **104**:507-516.

Dupias, G. 1988. L'extension de la flore Atlantique au versant nord des Pyrénées. Pages 529-534 *in* I. d. E. Altoaragoneses, editor. Homenaje a Pedro Montserrat.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**:1792-1797.

Edwards, R. and A. Brooks. 2008. The island of Ireland; drowning the myth of an Irish landbridge, Mind the Gap - The Postglacial Colonization Conference 2006. Special Supplement to The Irish Naturalists' Journal. **10**:19-34.

Ehlers, J., P. L. Gibbard, and P. D. Hughes, editors. 2011. Quaternary glaciations - extent and chronology; a closer look. Elsevier.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Elith, J. and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology Evolution and Systematics **40**:677-697.

Elith, J., S. J. Phillips, T. Hastie, M. Dudik, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions **17**:43-57.

Elton, C. S. 1927. Animal Ecology. Sidgwick and Jackson, London.

Eriksson, G., S. Andersson, V. Eiche, J. Ifver, and A. Persson. 1980. Severity index and transfer effects on survival and volume production of *Pinus sylvestris* in Northern Sweden. Studia Forestalia Suecica **156**:1-31.

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

Evans, J. M., R. J. Fletcher, and J. Alavalapati. 2010. Using species distribution models to identify suitable areas for biofuel feedstock production. Global Change Biology Bioenergy **2**:63-78.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data. Genetics **131**:479-491.

Fagundez, J. 2006. Two wild hybrids of *Erica* L. (Ericaceae) from northwest Spain. Botanica Complutensis **30**:131-135.

Fagundez, J., R. Juan, I. Fernandez, J. Pastor, and J. Izco. 2010. Systematic relevance of seed coat anatomy in the European heathers (Ericeae, Ericaceae). Plant Systematics and Evolution **284**:65-76.

FAO/IIASA/ISRIC/ISSCAS/JRC. 2012. Harmonised World Soil Database Version 1.2. FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Fitzpatrick, M. C., J. F. Weltzin, N. J. Sanders, and R. R. Dunn. 2007. The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? Global Ecology and Biogeography **16**:24-33.

Flanders, J., L. Wei, S. J. Rossiter, and S. Y. Zhang. 2011. Identifying the effects of the Pleistocene on the greater horseshoe bat, *Rhinolophus ferrumequinum*, in East Asia using ecological niche modelling and phylogenetic analyses. Journal of Biogeography **38**:439-452.

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. Great Britain Geological Survey Memoir **1**:336–432.

Foss, P. J. and G. J. Doyle. 1988. Why has *Erica erigena* (the Irish heather) such a markedly disjunct European distribution? Plants Today **1**:161-168.

Foss, P. J. and G. J. Doyle. 1990. The history of *Erica erigena* R. Ross, an Irish plant with a disjunct European distribution. Pages 1-16 Journal of Quaternary Science.

Franzi, V. A. 1984. Algunas plantas interesantes de la Sierra de Gata (Cáceres). Studia Botanica **3**:301-304.

Freeland, J. R. 2005. Molecular ecology. John Wiley and Sons.
Friis, E. M., P. R. Crane, and K. R. Pederson. 2011. Early flowers and angiosperm evolution. Cambridge University Press, Cambridge.
Furness, C. 2009. Pollen evolution and development in Ericaceae, with particular reference to pseudomonads and variable pollen sterility in Styphelioideae. International Journal of Plant Sciences **170**:476-494.
Gaudeul, M. 2006. Disjunct distribution of *Hypericum nummularium* L. (Hypericaceae): molecular data suggest bidirectional colonization from a single refugium rather than survival in distinct refugia. Biological Journal of the Linnean Society **87**:437-447.
Gause, G. F. 1934. The struggle for existence. Williams & Wilkins, Baltimore.
Gimingham, C. H. 1972. Ecology of heathlands. Chapman & Hall, London.
Gomez, A. and D. H. Lunt. 2007. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula.
Grinnell, J. 1917. The niche-relationships of the California thrasher. The Auk **34**:427-433.
Guinan, J., C. Brown, M. F. J. Dolan, and A. J. Grehan. 2009. Ecological niche modelling of the distribution of cold-water coral habitat using underwater remote sensing data. Ecological Informatics **4**:83-92.
Guisan, A. and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters **8**:993-1009.
Hall, D., V. Luquez, V. M. Garcia, K. R. St Onge, S. Jansson, and P. K. Ingvarsson. 2007. Adaptive population differentiation in phenology across a latitudinal gradient in European Aspen (*Populus tremula* L.): a comparison of neutral markers, candidate genes and phenotypic traits. Evolution **61**:2849-2860.
Hamill, R. M., D. Doyle, and E. J. Duke. 2006. Spatial patterns of genetic diversity across European subspecies of the mountain hare, *Lepus timidus* L. Heredity **97**:355-365.
Hare, M. P. 2001. Prospects for nuclear gene phylogeography. Trends in Ecology & Evolution **16**:700-706.
Harris, S. A. and R. Ingram. 1991. Chloroplast DNA and biosystematics - the effects of intraspecific diversity and plastid transmission. Taxon **40**:393-412.
Hartl, D. L. and A. G. Clark. 2007. Principles of population genetics. 4 edition. Sinauer Associates, Inc., Sunderland, Massachusetts.
Hedren, M., S. Nordstrom, and D. Stahlberg. 2012. Geographical variation and systematics of the tetraploid marsh orchid *Dactylorhiza majalis* subsp. *sphagnicola* (Orchidaceae) and closely related taxa. Botanical Journal of the Linnean Society **168**:174-193.
Hennessey, H. 1867. On the origin of the South European plants found growing in the West and South of Ireland. Proceedings of the Royal Irish Academy **10**:66-70.
Hernandez-Roldan, J. L., C. Murria, H. Romo, G. Talavera, E. Zakharov, P. D. N. Hebert, and R. Vila. 2011. Tracing the origin of disjunct distributions: a case of biogeographical convergence in *Pyrgus* butterflies. Journal of Biogeography **38**:2006-2020.
Heslop-Harrison, J. S. and T. Schwarzacher. 2011. Organisation of the plant genome in chromosomes. The Plant Journal **66**:18-33.
Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature **405**:907-913.
Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society **58**:247-276.
Heywood, V. H., R. K. Brummit, A. Culham, and O. Seberg. 2007. Flowering plant families of the world. Firefly Books Ltd., Ontario, Canada.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965-1978.

Hill, M. O., C. D. Preston, and D. B. Roy. 2004. PLANTATT - Attributes of British and Irish plants: status, size, life history, geography and habitats. Centre for Hydrology and Ecology.

Hoarau, G., J. A. Coyer, J. H. Veldsink, W. T. Stam, and J. L. Olsen. 2007. Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*. Molecular Ecology **16**:3606-3616.

Hodkinson, T. R., M. W. Chase, and S. A. Renvoize. 2002. Characterization of a genetic resource collection for *Miscanthus* (Saccharinae, Andropogoneae, Poaceae) using AFLP and ISSR PCR. Annals of Botany **89**:627-636.

Holderegger, R. and R. J. Abbott. 2003. Phylogeography of the Arctic-Alpine *Saxifraga oppositifolia* (Saxifragaceae) and some related taxa based on cpDNA and its sequence variation. American Journal of Botany **90**:931-936.

Holsinger, K. E. and B. S. Weir. 2009. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). Nature Reviews Genetics **10**:639-650.

Hooker, W. J. 1835. Another heath found in Ireland. Companion to the Botanical Magazine **1**:158.

Huckerby, E. and F. Oldfield. 1976. The Quaternary vegetational history of the French Pays Basque. II. Plant macrofossils and additional pollen-analytical data. New Phytologist **77**:499-526.

Hutchinson, G. E. 1957. Concluding remarks: Cold Spring Harbor symposium on quantitative biology **22**:415-427.

Huttunen, S., L. Hedenas, M. S. Ignatov, N. Devos, and A. Vanderpoorten. 2008. Origin and evolution of the northern hemisphere disjunction in the moss genus *Homalothecium* (Brachytheciaceae). American Journal of Botany **95**:720-730.

Jessen, K. 1949. Studies in the late Quaternary deposits and flora history of Ireland. Proceedings of the Royal Irish Academy. Section B: Biological, Geological, and Chemical Science **52**:85-290.

Jessen, K., S. T. Anderson, and A. Farrington. 1959. The interglacial deposit near Gort, Co. Galway, Ireland. Proceedings of the Royal Irish Academy. Section B: Biological, Geological, and Chemical Science **60**:3-77.

Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Global Ecology and Biogeography **21**:498-507.

Jones, A., L. Montanarella, and R. Jones. 2005. Soil Atlas of Europe. Office for Official Publications of the European Communities, Luxembourg.

Jussieu, A. L. d. 1802. Mémoire sur la plante nomée par les botanistes *Erica daboecia* et sur la nécessite de la rapporter à un autre genre et à une autre famille. Annales du Muséum d'Histoire Naturelle **1**:52-56.

Kawase, D., T. Yumoto, K. Hayashi, and K. Sato. 2007. Molecular phylogenetic analysis of the infraspecific taxa of *Erigeron thunbergii* A. Gray distributed in ultramafic rock sites. Plant Species Biology **22**:107-115.

Kay, K. M., J. B. Whittall, and S. A. Hodges. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. Bmc Evolutionary Biology **6**.

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

Keller, S. R., R. Y. Soolanayakanahally, R. D. Guy, S. N. Silim, M. S. Olson, and P. Tiffin. 2011. Climate-driven local adaptation of ecophysiology and phenology in balsam poplar, *Populus balsamifera* L. (Salicaceae). American Journal of Botany **98**:99-108.

Kingston, N. and S. Waldren. 2005. The historical biogeography of the 'Lusitanian' Ericaceae in the Irish flora. Pages 203-215 *in* F. T. Bakker, L. W. Chatrou, B. Gravendeel, and P. B. Pelser, editors. Plant Species-Level Systematics: New Perspectives on Pattern & Process. A.R.G Gantner Verlag, Ruggell.

Koch, K. 1869. Dendrologie; Bäume, Sträucher und Halbsträucher, welche in Mittel- und Nord- Europa in Freien kultiviert. Erlangen F. Enke.

Kron, K. A., W. S. Judd, P. F. Stevens, D. M. Crayn, A. A. Anderberg, P. A. Gadek, C. J. Quinn, and J. L. Luteyn. 2002. Phylogenetic classification of Ericaceae: molecular and morphological evidence. Botanical Review **68**:335-423.

Kron, K. A. and J. L. Luteyn. 2005. Origins and biogeographic patterns in Ericacae: New insights from recent phylogenetic analyses. Pages 479-500 *in* Symposium on Plant Diversity and Complexity Patterns - Local, Regional and Global Dimensions. . The Royal Danish Academy of Sciences and Letters, Copenhagen.

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

Kumar, S. 2005. Molecular clocks: four decades of evolution. Nature Reviews Genetics **6**:654-662.

Kyoda, S. and H. Setoguchi. 2010. Phylogeography of Cycas revoluta Thunb. (Cycadaceae) on the Ryukyu Islands: very low genetic diversity and geographical structure. Plant Systematics and Evolution **288**:177-189.

Lacaita, C. C. 1929. Duriaei iter Asturicum botanicum. Journal of Botany **67**:256-258, 310.

Lambeck, K. 1995. Late Devensian and Holocene shorelines of the British Isles and North Sea from models of glacio-hydro-isostatic rebound. Journal of the Geological Society, London **152**:437-448.

Lanfear, R., J. J. Welch, and L. Bromham. 2010. Watching the clock: Studying variation in rates of molecular evolution between species. Trends in Ecology & Evolution **25**:495-503.

Laterrade, J. F. 1832. Additions à la flore Bordelaise. Actes de la Société Linnéenne **25**:3-20.

Le Bail, J. 2008. Plan de conservation directeur en faveur de la bruyère de Saint-Daboec (Daboecia cantabrica (Hudson) C. Koch) en région Pays de la Loire. Conservatoire Botanique National de Brest.

Lihova, J., T. Carlsen, C. Brochmann, and K. Marhold. 2009. Contrasting phylogeographies inferred for the two alpine sister species *Cardamine resedifolia* and *C. alpina* (Brassicaceae). Journal of Biogeography **36**:104-120.

Linnaeus, C. v. 1762. Species plantarum: exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas. Impensis Direct. Laurentii Salvii, Holmiae.

Linnaeus, C. v. 1767. Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Typis Ioannis Thomae, Vindobonae.

Linton, E. F. 1902. *Erica stuarti*, novo hybr. The annals of Scottish natural history **11**:176-177.

Liu, C. R., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography **28**:385-393.
Lobo, J. M., A. Jimenez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography **17**:145-151.
Loidi, J., I. Biurrun, J. Antonio Campos, I. Garcia-Mijangos, and M. Herrera. 2010. A biogeographical analysis of the European Atlantic lowland heathlands. Journal of Vegetation Science **21**:832-842.
Lowe, A., S. Harris, and P. Ashton. 2004. Ecological genetics: design, analysis and application. Blackwell Science Ltd, Oxford.
Machado-Machado, E. A. 2012. Empirical mapping of suitability to dengue fever in Mexico using species distribution modeling. Applied Geography **33**:82-93.
Mackay, J. T. 1836. Flora Hibernica: comprising the flowering plants, ferns, Characeæ, Musci, Hepaticæ, Lichenes and Algæ of Ireland. William Curry Jun and Company, Dublin.
Madden, R. R. 1863. On the migrations from Spain to Ireland. Proceedings of the Royal Irish Academy **8**:372-393.
Maddison, W. P. and D. R. Maddison. 2011. Mesquite: a modular system for evolutionary analysis Version 2.75.
Mahy, G., R. A. Ennos, and A. L. Jacquemart. 1999. Allozyme variation and genetic structure of Calluna vulgaris (heather) populations in Scotland: the effect of postglacial recolonization. Heredity **82**:654-660.
Manel, S., J. M. Dias, S. T. Buckton, and S. J. Ormerod. 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. Journal of Applied Ecology **36**:734-747.
Martin-Bravo, S., P. Vargas, and M. Luceno. 2009. Is *Oligomeris* (Resedaceae) indigenous to North America? Molecular evidence for a natural colonization from the Old World. American Journal of Botany **96**:507-518.
Martinkova, N., R. A. McDonald, and J. B. Searle. 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., M. B. Rogatcheva, I. Gunduz, K. Fredga, and J. B. Searle. 2003. How did pygmy shrews colonize Ireland? Clues from a phylogenetic analysis of mitochondrial cytochrome b sequences. Proceedings of the Royal Society of London Series B-Biological Sciences **270**:1593-1599.
Matthes, M. C., A. Daly, and K. J. Edwards. 1998. Amplified Fragment Length Polymorphism (AFLP). Pages 183–192 *in* A. Karp, P. G. D. S. Isaac, and D. S. Ingram, editors. Molecular tools for screening biodiversity. Chapman & Hall, London.
Mayr, E. 1942. Systematics and the Origin of Species. Columbia University Press, New York.
McAllister, H. 1996. Reproduction in *Erica mackaiana*. Yearbook of the Heather Society:43-46.
McClintock, D. 1979. The status of, and correct name for, *Erica* 'Stuartii'. Watsonia **12**:249-252.
McClintock, D. 1983. The Carna colony of *Erica mackaiana*, a new variety. Irish Naturalists' Journal **21**:85-86.
McClintock, D. 1989. The heathers of Europe and adjacent areas. Botanical Journal of the Linnean Society **101**:279-289.

McGregor, C. E., C. A. Lambert, M. M. Greyling, J. H. Louw, and L. Warnich. 2000. A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. Euphytica **113**:135-144.

McKeown, N. J., R. A. Hynes, R. A. Duguid, A. Ferguson, and P. A. Prodohl. 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.

Meudt, H. M. and A. C. Clarke. 2007. Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends in Plant Science **12**:106-117.

Michener, C. D. and R. R. Sokal. 1957. A quantitative approach to a problem in classification. Evolution **11**:130-162.

Mitchell, F. J. G. 2006. Where did Ireland's trees come from? Biology and Environment **106B**:251-259.

Mitchell, G. F. and W. A. Watts. 1970. The History of the Ericaceae in Ireland during the Quaternary Epoch. Pages 13-21 *in* D. Walker and R. G. West, editors. Studies in the Vegetational History of the British Isles; essays in honour of harry Godwin. Cambridge University Press.

Nakicenovic, N., J. Alcamo, G. Davis, B. de Vries, J. Fenhann, S. Gaffin, K. Gregory, A. Grubler, T. Y. Jung, T. Kram, E. L. La Rovere, L. Michaelis, S. Mori, T. Morita, W. Pepper, H. M. Pitcher, L. Price, K. Riahi, A. Roehrl, H.-H. Rogner, A. Sankovski, M. Schlesinger, P. Shukla, S. J. Smith, R. Swart, S. van Rooijen, N. Victor, and Z. Dadi. 2000. Special Report on Emissions Scenarios : a special report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America **70**:3321-3323.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nelson, E. C. 1981. William McCalla: A Second 'Panegyric' for an Irish Phycologist. Irish Naturalists' Journal **20**:275-283.

Nelson, E. C. 1982. Historical records of the Irish Ericaceae *Arbutus unedo*, *Daboecia cantabrica*, *Erica erigena*, *Erica mackaiana* and *Ledum palustre*. Irish Naturalists' Journal **20**:364-369.

Nelson, E. C. 1984. Dabeoc - A saint and his heather. Year Book of the Heather Society **3**:41-46.

Nelson, E. C. 1989. Heathers in Ireland. Botanical Journal of the Linnean Society **101**:269-277.

Nelson, E. C. 1995. *Erica* x *stuartii*: the authorship reconsidered. Watsonia **2520**:275-278.

Nelson, E. C. 1996. Hybrids in St. Dabeoc's heath, *Daboecia* (Ericaceae): some nomenclatural adjustments. New Plantsman **3**:84-85.

Nelson, E. C. 2005. *Erica mackaiana* Bab. and *Erica x stuartii* (Macfarl.) Mast. (Ericaceae): Two heathers new to South Kerry (v.c. H1), Ireland. Watsonia **25**:414-417.

Nelson, E. C. and M. I. Fraga. 1983. Studies in *Erica mackaiana* Bab. II: Distribution in northern Spain. Glasra **7**:25-33.

Nogues-Bravo, D. 2009. Predicting the past distribution of species climatic niches. Global Ecology and Biogeography **18**:521-531.

Normand, S., R. E. Ricklefs, F. Skov, J. Bladt, O. Tackenberg, and J. C. Svenning. 2011. Postglacial migration supplements climate in determining plant species ranges in Europe. Proceedings of the Royal Society B-Biological Sciences **278**:3644-3653.

Oldfield, F. 1968. Quaternary vegetational history of French Pays Basque I. Stratigraphy and pollen analysis. New Phytologist **67**:677-731.

Otto-Bliesner, B. L., S. J. Marshall, J. T. Overpeck, G. H. Miller, A. Hu, and C. L. I. P. members. 2006. Simulating Arctic climate warmth and icefield retreat in the Last Interglaciation. Science **311**:1751-1753.

Ouborg, N. J., Y. Piquot, and J. M. van Groenendael. 1999. Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology **87**:551-568.

Palmer, J. D. 1991. Plastid chromosomes; structure and evolution. Pages 5-53 *in* L. Bogorad and I. K. Vasil, editors. The Molecular Biology of Plastids. Academic Press, San Diego.

Palmer, J. D. and D. B. Stein. 1986. Conservation of chloroplast genome structure among vascular plants. Current Genetics **10**:823-833.

Palmer, M. A., D. A. Falk, and J. B. Zedler. 2006. Ecological theory and the restoration of populations and communities.*in* D. A. Falk, M. A. Palmer, and J. B. Zedler, editors. Foundations of restoration ecology. Island Press, Washington D.C.

Peakall, R. and P. E. Smouse. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**:288-295.

Peakall, R. and P. E. Smouse. 2012. GenAlEx 6.5: Appendix 1 - Methods and Statistics. Article 1471-8278.

Pearson, R. G. and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography **12**:361-371.

Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography **34**:102-117.

Peltier, W. R. 2004. Global glacial isostasy and the surface of the ice-age earth: The ice-5G (VM2) model and grace. Annual Review of Earth and Planetary Sciences **32**:111-149.

Peterson, A. T. and D. A. Vieglais. 2001. Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem. Bioscience **51**:363-371.

Petit, R. J., I. Aguinagalde, J.-L. d. Beaulieu, C. Bittkau, B. Simon, R. Cheddadi, R. Ennos, S. Fineschi, D. Grivet, M. Lascoux, A. Mohanty, G. Mülller-Starck, B. Demesure-Musch, A. Palmé, J. P. Martín, S. Rendell, and G. G. Vendramin. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. Science **300**:1563-1565.

Petit, R. J., S. Brewer, S. Bordacs, K. Burg, R. Cheddadi, E. Coart, J. Cottrell, U. M. Csaikl, B. van Dam, J. D. Deans, S. Espinel, S. Fineschi, R. Finkeldey, I. Glaz, P. G. Goicoechea, J. S. Jensen, A. O. Konig, A. J. Lowe, S. F. Madsen, G. Matyas, R. C. Munro, F. Popescu, D. Slade, H. Tabbener, S. G. M. de Vries, B. Ziegenhagen, J. L. de Beaulieu, and A. Kremer. 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.

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

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231-259.

Phillips, S. J. and M. Dudik. 2004. A maximum entropy approach to species distribution modeling. Pages 655-662 *in* Twenty-First International Conference on Machine Learning. AMC Press.

Phillips, S. J. and M. Dudik. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography **31**:161-175.

Phillips, S. J., M. Dudik, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications **19**:181-197.

Posada, D. and K. A. Crandal. 1998. ModelTest: testing the model of DNA substitution. Bioinformatics **14**:817-818.

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

Praeger, R. L. 1913. On the buoyancy of the seeds of some Britannic plants. The Scientific Proceedings of the Royal Dublin Society **14**:13-62.

Praeger, R. L. 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.

Praeger, R. L. 1939. The Hooker Lecture: The relations of the flora and fauna of Ireland to those of other countries. Proceedings of the Linnean Society of London **151**:192-213.

Praeger, R. L. 1997. The way that I went: an Irishman in Ireland. Collins Press, Dublin.

Prentice, I. C., D. Jolly, and B. participants. 2000. Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. Journal of Biogeography **27**:507-519.

Preston, C. D., D. A. Pearman, and T. D. Dines. 2002. New atlas of the British and Irish flora. Oxford University Press, Oxford.

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

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

Queiroz, C. D., F. R. D. Batista, and L. O. de Oliveira. 2011. Evolution of the 5.8S nrDNA gene and internal transcribed spacers in *Carapichea ipecacuanha* (Rubiaceae) within a phylogeographic context. Molecular Phylogenetics and Evolution **59**:293-302.

Rafalski, J. A., J. M. Vogel, M. Morgante, W. Powell, C. Andre, and S. V. Tingey. 1996. Generating and using DNA markers in plants. Pages 75-134 *in* B. Birren and E. Lai, editors. Analysis of Non-Mammalian Genomes – A Practical Guide. Academic Press, New York.

Raven, P. H. 1972. Plant species disjunctions - summary. Annals of the Missouri Botanical Garden **59**:234-246.

Raxworthy, C. J., C. M. Ingram, N. Rabibisoa, and R. G. Pearson. 2007. Applications of ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. Systematic Biology **56**:907-923.

Read, D. J. 1991. Mycorrhizas in ecosystems. Cellular and Molecular Life Sciences **47**:376-391.

Read, D. J. 1996. The structure and function of the ericoid mycorrhizal root. Annals of Botany **77**:365-374.

Reid, C. 1913. The relations of the present plant population of the British Isles to the glacial period. Journal of Ecology **1**:42-46.

Reshetnikov, A. N. and G. F. Ficetola. 2011. Potential range of the invasive fish rotan (Perccottus glenii) in the Holarctic. Biological Invasions **13**:2967-2980.
Reynolds, S. 2002. A checklist of alien plants in Ireland. National Botanic Gardens, Dublin.
Rich, T. C. G. 2005. Distribution of the Irish whitebeam, *Sorbus hibernica* E. F. Warb. (Rosaceae). Watsonia **25**:369-380.
Richards, C. L., B. C. Carstens, and L. L. Knowles. 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. Journal of Biogeography **34**:1833-1845.
Robinson, L. M., J. Elith, A. J. Hobday, R. G. Pearson, B. E. Kendall, H. P. Possingham, and A. J. Richardson. 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Global Ecology and Biogeography **20**:789-802.
Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes **4**:137-138.
Rowe, G., D. J. Harris, and T. J. C. Beebee. 2006. Lusitania revisited: a phylogeographic analysis of the natterjack toad *Bufo calamita* across its entire biogeographical range. Molecular Phylogenetics and Evolution **39**:335-346.
Rumeu, B., J. Caujape-Castells, J. Luis Blanco-Pastor, R. Jaen-Molina, M. Nogales, R. B. Elias, and P. Vargas. 2011. The colonization history of *Juniperus brevifolia* (Cupressaceae) in the Azores Islands. Plos One **6**.
Saitou, N. and M. Nei. 1986. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Japanese Journal of Genetics **61**:611-611.
Sanchez, A. C., P. E. Osborne, and N. Haq. 2010. Identifying the global potential for baobab tree cultivation using ecological niche modelling. Agroforestry Systems **80**:191-201.
Sang, T., D. J. Crawford, and T. F. Stuessy. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). American Journal of Botany **84**:1120-1136.
Scannell, M. J. P. 1983. *Erica erigena* R. Ross and *Daboecia cantabrica* (Huds.) C. Koch. - some further notes on the historical records. Irish Naturalists' Journal **21**:29-30.
Schaal, B. A., J. F. Gaskin, and A. L. Caicedo. 2003. Phylogeography, haplotype trees, and invasive plant species. Journal of Heredity **94**:197-204.
Schmitt, T. 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. Frontiers in Zoology **4**.
Schneeweiss, G. M. and P. Schonswetter. 2010. The wide but disjunct range of the European mountain plant *Androsace lactea* L. (Primulaceae) reflects Late Pleistocene range fragmentation and post-glacial distributional stasis. Journal of Biogeography **37**:2016-2025.
Schönswetter, P., J. Suda, M. Popp, H. Weiss-Schneeweiss, and C. Brochmann. 2007. Circumpolar phylogeography of *Juncus biglumis* (Juncaceae) inferred from AFLP fingerprints, cpDNA sequences, nuclear DNA content and chromosome numbers. Molecular Phylogenetics and Evolution **42**:92-103.
Seberg, O. and G. Petersen. 2006. Mitochondrial DNA Sequences in plant phylogenetics and evolution: symposium at XVII International Botanical Congress, Vienna, Austria. Taxon **55**:833-835.
Sheperd, W. R. 1911. Historical Atlas. Henry Holt and Company.
Sherwood, A. R. 2008. Phylogeography of *Asparagopsis taxiformis* (Bonnemaisoniales, Rhodophyta) in the Hawaiian Islands: two mtDNA markers support three separate introductions. Phycologia **47**:79-88.

Sinclair, W. T., J. D. Morman, and R. A. Ennos. 1998. Multiple origins for Scots pine (*Pinus sylvestris* L.) in Scotland: evidence from mitochondrial DNA variation. Heredity **80**:233-240.

Singh, A., R. M. Devarumath, S. RamaRao, V. P. Singh, and S. N. Raina. 2008. Assessment of genetic diversity, and phylogenetic relationships based on ribosomal DNA repeat unit length variation and Internal Transcribed Spacer (ITS) sequences in chickpea (*Cicer arietinum*) cultivars and its wild species. Genetic Resources and Crop Evolution **55**:65-79.

Soberón, J. and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiversity Informatics **2**:1-10.

Soltis, D. E. and P. S. Soltis. 1992. Choosing an approach and an appropriate gene for phylogenetic analysis. Pages 117-150 *in* D. E. Soltis, P. S. Soltis, and J. J. Doyle, editors. Molecular Systematics of Plants. Chapman and Hall, New York.

Stevens, P. F. 1971. Classification of Ericaceae - subfamilies and tribes. Botanical Journal of the Linnean Society **64**:1-&.

Stockwell, D. R. B. and D. P. Peters. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographic Information Systems **13**:143-158.

Sun, Y., D. Z. Skinner, G. H. Liang, and S. H. Hulbert. 1994. Phylogenetic analysis of *Sorghum* and related taxa using Internal Transcribed Spacers of nuclear ribosmal DNA. Theoretical and Applied Genetics **89**:26-32.

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. Science **240**:1285-1293.

Swofford, D. L. 2003. PAUP* Phylogenetic Analysis Using Parsimony (* and other methods) Version 4. Sinauer Associates Inc., Sunderland, Massachusetts.

Taberlet, P., L. Fumagalli, A.-G. West-Saucy, and J.-F. Cossons. 1998. Comparative Phylogeography and postglacial colonisation routes in Europe. Molecular Ecology **7**:453-464.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of 3 noncoding regions of chloroplast DNA. Plant Molecular Biology **17**:1105-1109.

Tate, J. A. and B. B. Simpson. 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. Systematic Botany **28**:723-737.

Tellería, J. L., T. Santos, P. Refoyo, and J. Muñoz. 2012. Use of ring recoveries to predict habitat suitability in small passerines. Diversity and Distributions:1-9.

Thorne, R. F. 1972. Major disjunctions in the geographic ranges of seed plants. Quarterly Review of Biology **47**:365-411.

Tournefort, J. P. 1700. Institutiones Rei Herbariæ. Parisiis: E Typographia Regia, Paris.

Treier, U. A. and H. Muller-Scharer. 2011. Differential effects of historical migration, glaciations and human impact on the genetic structure and diversity of the mountain pasture weed *Veratrum album* L. Journal of Biogeography **38**:1776-1791.

Tutin, T. G. 1953. The vegetation of the Azores. Journal of Ecology **41**:53-61.

Tyberghein, L., H. Verbruggen, K. Pauly, C. Troupin, F. Mineur, and O. De Clerck. 2011. Bio-ORACLE : a global environmental dataset for marine species distribution modelling. Global Ecology and Biogeography **21**:272–281.

van Doorslaer, L. 1990a. The ecology of *Erica mackaiana* Bab. with reference to its conservation in Connemara, Ireland. University College Galway, Galway.

van Doorslaer, L. 1990b. *Erica mackaiana* Bab. in Mayo, a new county record. Irish Naturalists' Journal **23**:268-270.

Vaughan, A. P. M., L. A. Dowling, F. J. G. Mitchell, S. E. Lauritzen, A. M. McCabe, and P. Coxon. 2004. Depositional and post-depositional history of warm stage deposits at Knocknacran, Co. Monaghan, Ireland: implications for preservation of Irish last interglacial deposits. Journal of Quaternary Science **19**:577-590.

Vera, M. L. 2011. Heathlands in the Cantabrian mountains (north Spain) under climate schange scenario. 12th Heathland Workshop, 12-18 June 2011, University of Oviedo, Spain, (Poster presentation).

Vialatte, A., A. Guiller, A. Bellido, and L. Madec. 2008. Phylogeography and historical demography of the Lusitanian snail *Elona quimperiana* reveal survival in unexpected separate glacial refugia. Bmc Evolutionary Biology **8**:339.

Vincent, P. 1990. The Biogeography of the British Isles. Routledge, London.

Vos, P., R. Hogers, M. BLeeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research **23**:4407-4414.

Waldren, S., P. Coxon, and N. Kingston. 2006. Phytogeography of Ireland, past and present. Pages 31-39 *in* S. Leach, C. Page, Y. Peytoureau, and M. Sanford, editors. Botanical Links in the Atlantic Arc. Botanical Society of the British Isles, London.

Waltari, E., R. J. Hijmans, A. T. Peterson, A. S. Nyari, S. L. Perkins, and R. P. Guralnick. 2007. Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. Plos One **2**.

Wang, B. and Y. L. Qiu. 2006. Phylogenetic distribution and evolution of mycorrhizas in land plants. Mycorrhiza **16**:299-363.

Watson, L. 1965. The taxonomic significance of certain anatomical variations among Ericaceae. Journal of the Linnean Society of London, Botany **59**:111-125.

Watts, W. A. and R. Ross. 1959. Interglacial deposits at Kilbeg and Newtown, Co. Waterford. Proceedings of the Royal Irish Academy. Section B: Biological, Geological, and Chemical Science **60**:79-134.

Watts, W. A., B. W. Sparks, and J. H. Dickson. 1963. Interglacial deposits at Baggotstown, near Bruff, Co. Limerick. Proceedings of the Royal Irish Academy. Section B: Biological, Geological, and Chemical Science **63**:167-189.

Webb, D. A. 1951. Hybrid plants in Ireland Irish Naturalists' Journal **10**:201-205.

Webb, D. A. 1954. Notes on four Irish heaths. Irish Naturalists' Journal **11**:186-192.

Webb, D. A. 1955. *Erica mackaiana* Bab. (*E. mackayi* Hook. ined.). Journal of Ecology **43**:319-330.

Webb, D. A. 1957. Description of exhibit of *Erica mackaiana* Bab. Proceedings of the BSBI **3**:105.

Webb, D. A. 1983. The flora of Ireland in its European context. Journal of Life Sciences Royal Dublin Society **4**:143-160.

Whitlock, B. A., A. M. Hale, and P. A. Groff. 2010. Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. Plos One **5**.

Williams, J. N., C. W. Seo, J. Thorne, J. K. Nelson, S. Erwin, J. M. O'Brien, and M. W. Schwartz. 2009. Using species distribution models to predict new occurrences for rare plants. Diversity and Distributions **15**:565-576.

Wilson, C. D., D. Roberts, and N. Reid. 2011. Applying species distribution modelling to identify areas of high conservation value for endangered species: A case study using *Margaritifera margaritifera* (L.). Biological Conservation **144**:821-829.

Wolfe, K. H., W. H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences of the United States of America **84**:9054-9058.

Woodell, S. R. J. 1958. *Daboecia cantabrica* K. Koch (*D. polifolia* D. Don; *Menziesia polifolia* Juss.). Journal of Ecology **46**:205-216.

Woodman, P. C. 1985. Prehistoric settlement and environment. Pages 251-278 *in* K. J. Edwards and W. P. Warren, editors. The Quaternary History of Ireland. Academic Press, London.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution **19**:395-420.