Colonisation and diversification in invertebrates: looking within species on islands to connect pattern and process

Christiana Mara de Assis Faria

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of East Anglia, England
School of Biological Sciences

March 2015

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there-from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Thesis abstract

How species originate and how communities of species assemble are among the most intriguing questions in biology, and colonisation is a key element to understand them. Using two island scenarios and applying molecular tools, this thesis looks within species to investigate the themes of colonisation (both island colonisation and European continental recolonisation) and diversification processes in invertebrates. The aim was to address three gaps in our understanding about island colonisation, speciation and the assembly of biota. In the Canary Islands, an oceanic island system, the gaps addressed were: (i) the possibility that genomic admixture among multiple founding lineages has featured in the diversification of a very species rich coleopteran genus; and (ii) the lack of information regarding the colonisation history and dynamics of the small arthropod soil dwelling fauna. In Great Britain, a continental island system, the gap addressed was the under-explored possibility that the UK was not completely defaunated during glaciations, then recolonised from external sources, but that a more complex pattern, involving persistence within small cryptic refugia, may have featured in the history of its invertebrate soil dwelling fauna. I reveal two instances of shared mtDNA variation among weevil species from different Canarian islands for which I was able to dismiss explanations of incomplete lineage sorting and reveal a history of colonisation and speciation involving genetic admixture (first gap). I characterise Collembola evolutionary diversity within Tenerife and the distribution of lineage colonisation times, and reveal this fauna to be represented by a mosaic of very old lineages and a large number of very recently arrived lineages (second gap). Finally, I reveal signatures of survival and persistence of the Collembola fauna through the last Pleistocene glaciation in Great Britain (third gap). How these results fit into a broader evolutionary and conservation context as well as future directions are discussed.

Chapter contributions and publications

All parts of this thesis have been written by Christiana M. A. Faria, in consultation with Brent C. Emerson. Below are estimates of percentage contributions to the initial concept and development, sample collection, data collection and data analysis, as well as the publication status for each chapter.

Chapter 2: re-submitted to *Journal of Biogeography* as Faria CMA, Machado A, Amorim IR, Gage MJG, Borges AV, Emerson BC (in review), under the title: Evidence for multiple founding lineages and genetic admixture in the evolution of species within an oceanic island weevil (Coleoptera, Curculionidae) super-radiation.

- Concept and development: B Emerson 50%, C Faria 50%
- Sample collection: A Machado 100%
- Data collection and data analysis: C Faria 100%

Chapter 3: not yet published. To be submitted under authorship Faria CMA, Ramirez-Gonzalez R., Emerson BC.

- Concept: B Emerson 40%, C Faria 60%
- Sample collection: B Emerson 100%
- Data collection and analysis: Ramirez-Gonzalez 10%, C Faria 90%

Chapter 4: not yet published. To be submitted under authorship Faria CMA, Shaw P, Emerson BC.

- Concept and development: B Emerson 40%, C Faria 60%
- Sample collection: Shaw P 100 %
- Data collection and analysis: C Faria 100%

Chapter 5: not yet published. To be submitted under authorship Faria CMA, Shaw P, Emerson BC.

- Concept and development: B Emerson 30%, C Faria 70%
- Sample collection: Shaw P 100 %
- Data collection and analysis: C Faria 100%

Appendix 2 (Chapter 1): a perspective commenting on a study that investigated similar issues as those in Chapter 2. Published as Emerson and Faria (2014) Fission and fusion in island taxa – serendipity, or something to be expected? *Molecular Ecology* **23:**5132-5134.

Supervisory Information

Primary supervisor:

Dr. Brent C. Emerson
Island Ecology and Evolution Research Group,
Instituto de Productos Naturales y Agrobiología (IPNA-CSIC),
c/Astrofísico Francisco Sánchez 3,
San Cristóbal de La Laguna,
Tenerife, Canary Islands,
38206
Spain

Secondary supervisor:

Dr. Matthew J. G. Gage
School of Biological Sciences,
University of East Anglia,
Norwich Research Park,
Norwich,
Norfolk,
NR4 7TJ
United Kingdom

List of Contents

Thesis abstract………… ii

Chapter contributions and publications………… iii

Supervisory Information………… iv

List of Contents………… v

Acknowledgments………… ix

Chapter 1 – General Introduction………… 1

1.1 Background………… 1

1.2 Oceanic Island system: aids, difficulties and bias………… 3

1.2.1 The first bias………… 4

1.2.2 The second bias………… 5

1.3 Continental island systems: historical facts and gaps………… 7

1.4 References………… 12

1.5 Appendices………… 18

Chapter 2 - Speciation success on islands: evolution of the *Laparocerus tessellatus* complex (Curculionidae) in the Canary archipelago………… 28

2.1 ABSTRACT………… 29

2.2 INTRODUCTION………… 30

2.3 MATERIALS AND METHODS………… 32

2.3.1 Sampling and laboratory procedures………… 32

2.3.2 Sequence alignment, haplotype reconstruction, and sequence properties.. 35

2.3.3 Evolutionary tree and haplotype network construction………… 36

2.3.4 Testing hypothesis of monophyly………… 37

2.3.5 Analysis of linkage disequilibrium and Hardy–Weinberg equilibrium…. 37

2.3.6 Dating analysis………… 38

2.4 RESULTS………… 39

2.4.1 Mitochondrial COII gene………… 39

2.4.2 Nuclear ITS2 gene………… 43

2.4.3 Phylogenetic analysis of mtDNA COII gene sequences………… 45

2.4.4 Phylogenetic and network analyses of nuclear ITS2 sequences………… 48

2.4.5 Species boundaries in Tenerife………………………………………….. 49
2.4.6 Dating analysis……………………………………………………….. 49
2.5 DISCUSSION……………………………………………………………... 52
2.5.1 Geographic origin of the *L. tessellatus* complex………………………… 54
2.5.2 Gran Canaria species and the founding of neighbouring islands………... 55
2.5.3 Tenerife species and the founding of neighbouring islands……………... 55
2.5.4 La Palma and El Hierro–single species with multiple origins…………... 56
2.5.5 Species boundaries in Tenerife…………………………………………... 58
2.5.6 Conclusions………………………………………………………….... 59
2.6 REFERENCES 59
2.7 Appendices 65
Chapter 3 - Community level assessment of arthropod colonisation history: an example using the Collembola of Tenerife……………………………………… 68
3.1 ABSTRACT……………………………………………………………….. 69
3.2 INTRODUCTION…………………………………………………………. 70
3.3 MATERIALS AND METHODS…………………………………………….. 75
3.3.1 Sample collection and laboratory work………………………………….. 75
3.3.2 Data analysis……………………………………………………….... 77
3.3.3 Cluster assignment……………………………………………………. 77
3.3.4 Frequency distribution and cumulative frequency plots………………… 80
3.4 RESULTS…………………………………………………………….............. 80
3.4.1 PyroCleaning results……………………………………….............. 80
3.4.2 Clustering thresholds………………………………………….............. 80
3.4.3 Taxonomic assignment……………………………………….......... 81
3.4.4 Neighbour joining and distance analyses……………………….......... 85
3.4.5 Tenerife community and distributions…………………………........... 88
3.5 DISCUSSION……………………………………………………………....... 91
3.5.1 Short barcodes and diversity estimation……………………………....... 92
3.5.2 Lineages colonisation………………………………………………..... 93
3.5.3 Lineage distribution…………………………………………………....... 96

3.5.4 Diversity comparison with previous Tenerife surveys…………………....... 96
3.5.5 Limitations and future directions……………….....……………………........ 97
3.6 REFERENCES………………………………………………………….................... 98
3.7 Appendices……………………………………………………........................... 102
Chapter 4 - Quantifying genetic diversity of *Entomobrya* (Hexapoda, Collembola) lineages in the UK: have they survived through the last Pleistocene glaciation? ... 116
4.1 ABSTRACT…………………………………………………….......................... 117
4.2 INTRODUCTION……………………………………………........................... 118
4.3 MATERIAL AND METHODS…………………………………........................ 122
4.3.1 Sampling and laboratory work……………………………........................ 122
4.3.2 Sequence alignment and molecular lineage delineation............................ 124
4.3.3 Evolutionary rate estimation and dating analysis……….......................... 124
4.3.4 Geographical distribution and species diversity analysis…....................... 125
4.4 RESULTS……………………………………………………............................ 126
4.4.1 Overview of *Entomobrya* in Great Britain…………............................... 126
4.4.2 Geographical distribution and species diversity analysis.......................... 130
4.4.3 MtDNA COI rate estimation and dating analysis...................................... 132
4.5 DISCUSSION…………………………………………………........................... 133
4.5.1 Long-term persistence of *Entomobrya* in the UK……............................ 133
4.5.2 Species diversity analysis…………………………………......................... 134
4.5.3 Conclusions………………………………………………….......................... 138
4.6 REFERENCES…………………………………………………......................... 138
4.7 Appendices…………………………………………………….......................... 143
Chapter 5 - Genetic diversity and persistence of *Lepidocyrtus* (Hexapoda, Collembola) lineages through historical climate changes in the island of Great Britain………. 149
5.1 ABSTRACT………………………………………………….............................. 150
5.2 INTRODUCTION…………………………………………….............................. 151
5.3.1 Sampling and laboratory work…………………………............................. 155
5.3.2 Sequence alignment and molecular lineage delineation............................ 157
5.3.3 Molecular dating analysis………………………….................................... 158

5.3.4 Geographical distribution and haplotype diversity analysis...................... 158
5.4 RESULTS.. 160
5.4.1 Overview of *Lepidocyrtus* in Great Britain.. 160
5.4.2 Spatial distribution of lineages.. 163
5.4.3 Species diversity analysis... 164
5.4.4 Long-term persistence analysis.. 167
5.5 DISCUSSION.. 168
5.5.1 Molecular lineages and geographic distributions..................................... 168
5.5.2 Species richness analysis.. 173
5.5.3 Long-term persistence of *Lepidocyrtus* in the UK................................... 174
5.5.4 Conclusions, limitations and future work.. 176
5.6 REFERENCES... 180
5.7 Appendix.. 186
Chapter 6 – General conclusions and discussion.. 188
References... 194

Acknowledgments

I start by expressing my earnest gratitude to my supervisor Brent Emerson, for his mentoring, endless patience, kind understanding of individual conditions and great support, which allowed me to get this far. As we [his students] always said, we were very lucky to have such a brilliant (both humanist and scientifically speaking) supervisor. Thanks also go to my secondary supervisor Matt Gage, especially for providing *in situ* support while Brent and I were in separate institutions. I would also like to thank Paulo Borges (Azorean Biodiversity Group, PT) for his supervisory role in the first chapter of thesis.

I gratefully acknowledge Antonio Machado, Pedro Oromí, Peter Shaw and Brent Emerson, for providing all the samples used in this thesis. Without their help this work would not have been possible.

I thank my thesis examiners Sara Goodacre (University of Nottingham) and Richard Davies (University of East Anglia) for the thought-provoking but enjoyable discussion on my project, and for their suggestions that improved the final presentation of the thesis.

I wish to thank to the members of Brent Emerson's research group with whom I had the pleasure of overlapping, and share science and friendship: Alicia Mastretta, Conrad Gillet, James Kitson, Gerardo Hernandez, and Isabel Amorim. A special thanks to James K for patiently guiding me through lab and computer issues, and to Isabel A for her help towards the beetle's part of the thesis.

I was also very lucky to be surrounded by an incredibly supportive and wonderful set of people at the UEA Centre for Ecology, Evolution and Conservation. My PhD journey was definitely less hard, less homesick, more enjoyable, and more joyful because of you all. Thank you very much for the companionship, for sharing your knowledge and experiences, and for sharing great relaxing moments (tea-cake times, volleyball, table tennis, etc.). Thanks to Alicia Mastretta, Aly Lumley, Amanda Bretman, Amy Romans, Becca Franklin, Becky Laidlaw, Cata Gonzaléz, Cat Morrison, Claudia Frickie, Conrad

Gillett, Damian Smith, Danni Gilroy, Danni Peruffo, Dave Collins, Dave Wright, Edd Almond, Ellie Fairfield, Emma Barret, Freydís Vigfusdottir, Helena Batalha, Jacob Holland, Janet Mason, Jo Godwin, James Kitson, José Alves, Kat Bebbington, Kate Bruce, Karl Phillips, Lewis Spurgin, Liam Crowther, Lucy Friend, Matt Dickson, Mike Bolton, Mohammed Albersh, Phil Leftwich, Sarah Marburger, Sarah Yaetes, Sían Diamond, Tim Huggins, Tom Finch, Vero Mendez, Wayne Rostant, Will Nash. A special thanks to Cata for guiding with thesis formatting and beloved friendship. Special thanks to my 'Bambi-lab' (Conrad, Karl, Janet, Jacob and Mohammad) and 'Zeke Hewitt office' (Aly, Ellie, James, Matt, and Kat) fellows; you were really great and added very nice touches to my days! My special 'thank you' to Karl for helping me numerous times with English proof reading, sharing the warmth and joy of his family, being so supportive and helping through some of my challenging times, you became a great friend, many thanks. I would also like to thank the faculty at UEA for great discussions and for helping to make the CEEC such a supportive and welcoming environment.

From the time I spent in Tenerife, I would like to thank the staff of the Instituto de Productos Naturales y Agrobiología (IPNA-CSIC) for administrative help, and the members of the Island Ecology and Evolution Research Group: Heriberto López, Marta Darias, Manuel Nogales and the Arthopods Group: Núria Macías, Salva Cruz, David Hernandéz and Toni Delgado, for great companionship, support and for making my writing time more enjoyable.

I thank the UEA Research Computing Service team for helping me with the High Performance Computing Cluster. A special thanks to Paul Wright (BIO) for his prompt and generous help with my many computer issues. I also thank Scott Chamberlain (ropensci.org) and Carlos Raffi for kindly providing scripts to speed up some tiresome parts of data processing.

I am greatly thankful to the Brazilian Council for Scientific and Technological Development (CNPq) for funding my PhD abroad (scholarship and tuition fees). I thank the UEA School of Biological Sciences, the Portuguese National funding agency for Science, Research and Technology (FCT) and the Linnean Society of London (LSL),

for financing molecular work through findings granted to myself, my co-authors Paulo Borges and Isabel Amorim (FCT) and Peter Shaw (LSL), respectively. I also thank the John & Pamela Salter Charitable Trust for financial support to attend a Computational Molecular Evolution Course.

For helping me to de-stress and find the body-mind steadiness to keep on working through more difficult times, I thank to Martin Langsdon and my yoga tutors Lou Kitchener and Joss Guin (Norwich), Cristina Mantecón and Jorge Alonso (Tenerife).

I would also like to thank my beloved friends and housemates, both in Norwich and in Tenerife, for joyous companionship, support in many situations, and for bringing me to the world outside PhD: Diana, David, Miry, Manuel, Frank, Alberto A, Alberto V; the Brazilian team: Maíra, Nanda, Ju, Marcos, Dani, Rafa, Vanessa, Camila, Marcelo, Rodrigo, Edgar, Kátia, Gabriel and Ismael; and the equipo Naranja: Carlos, Sara, Nico, Javier, Adriana, Alicia, and Soren. Special thanks to Alicia for all we shared in the last year, being friends, housemates and officemates, you are unique, thanks for inspiring me with science and arts. Extra special thanks also to Soren for his help and support in the final difficult months of this PhD journey, you kept the balance in the house and were a great companion, thank you.

Lastly, I extend my heartfelt gratitude to my family, especially to my father Ildeu A Faria and my mother Maria Christina B A Faria for their endless love, strong faith and for making it possible for me to achieve this; without you I would not be anywhere near where I am today. Many thanks mummy and dad, my brother Otávio, my sisters Mariana and Morgana, my brother-in-law Wellerson, and my princepessa (god-daughter) Letícia, for keeping your support through these years, for sending your warmth and love, and for being patience with my long absence. I missed you very much. We will soon be re-united!

Chapter 1 – General Introduction

1.1 Background

How species originate and how communities of species assemble are among the most intriguing questions in biology, and colonisation is a key element to understand them. Insights on how biological communities are formed and the relative roles of factors such as immigration, establishment, speciation, and extinction, operating at broad spatiotemporal scales, have been obtained in a variety of molecular studies and field experiments (e.g. Roderick & Gillespie, 1998; Thornton et al., 2001; Emerson & Gillespie, 2008; Wilson et al., 2011). Insular systems have been particularly useful for such studies, due to their simplication, when compared to continental systems (Warren et al., 2014 and references therein). Colonisation provides the primary source for the assembly of species within a newly available area, as well as contributing species to already established assemblages. Lineages from a regional gene pool have distinct historical and ecological features, of which a subset will be involved in the colonisation of a given site (Ricklefs, 2004; Emerson & Gillespie, 2008). After arrival, lineages will pass through an environmental filter (and a biotic filter in the case of non-empty areas), with their species-specific features influencing the success of establishment and any subsequent evolutionary change (Emerson & Gillespie, 2008; HilleRisLambers et al., 2012). These evolutionary changes are affected by microevolutionary processes, including genetic drift, selection, gene flow, and mutation, that act separately or in concert to promote the rise of divergences that may follow colonisation (e.g. Grant, 1998; Roderick & Gillespie, 1998; Clegg, 2010).

Insular systems, especially oceanic islands, have great appeal over continental systems for studying evolution because of their discrete boundaries, limited size and isolation. This results in terrestrial populations that are discretely bounded with reduced gene flow between them (Emerson, 2002). Island areas and inter-island distances are easily measured, and island communities tend to be less species rich than continental communities, facilitating the cataloguing of their fauna and flora (e.g. Heaney et al., 2005). Furthermore, when the time of emergence of an island is known, studies can be placed within a temporal framework, while if islands form an archipelago, studies can

potentially be replicated (e.g. Parent et al., 2008). Thus, insular systems provide ideal theatres (both geographical and historical) to study colonisation and species diversification (Emerson 2002).

Traditionally, two types of islands are recognized, those that emerged from the ocean (oceanic islands, Darwinian islands, *de novo* islands) and those that were separated from an existing landmass, or an existing habitat (continental islands or fragment islands) (Gillespie & Roderick, 2002; Whittaker & Fernandez-Palacios, 2007). The fundamental difference between them is that *de novo* islands are completely devoid of life when they appear from the ocean, thus requiring colonisation events for the establishment of their biota, while continental islands are already populated when they become separated from the continent. Consequently, in oceanic islands, immigration will increase species number at the early stages of community assembly with rate being dependent on the distance to source areas. Over time, if isolation persists, diversification will result in new endemic species (Gillespie & Roderick, 2002). By contrast, in continental islands, ecological niches are already filled at the initial stage of their insularisation, and over time species numbers will decrease as a result of relaxation, where extinction rate exceeds colonisation (Whittaker & Fernandez-Palacios, 2007).

In this thesis, the history and dynamics of invertebrate colonisation and diversification were investigated in two systems: the Canary Islands, a *de novo* islands system that began to emerge from the ocean around 25 million years ago and its currently formed by seven major islands and several islets (Fernández-Palacios et al., 2011) (for details see Appendix 1); and the United Kingdom, a continental insular system that falls in-between the two types of islands. On the one hand, it was first fragmented from the European continent with the formation of the English Channel around 450 thousand years ago[1] (Gupta et al., 2007), thus supporting a biota upon formation. On the other hand, it was extensively covered by ice and permafrost during Pleistocene glaciations (Chiverrell & Thomas, 2010), resulting in the extirpation of its biota over much of the modern land mass. Postglacial recolonisation from the mainland would then have been

[1] Changing of sea levels following cycles of glaciations allowed Britain to be connected to Europe repeated times after the formation of the English Channel – e.g. Doggerland. It is argued that it finally became physically disconnected from the continent with the complete flooding of Doggerland around 6,500 years ago (Weninger et al., 2008; Pettitt & White, 2012)

required to fill newly opened ecological spaces (Yalden, 1982; Hewitt, 2000a; Montgomery et al., 2014), although it remains possible that some species could have survived in refugia through glaciations, thus also contributing to the re-establishment of the biota (Stewart et al., 2010) . Within these two island scenarios, this thesis looks within species to investigate the themes of colonisation (both island colonisation and European continental recolonisation), differentiation and diversification processes in invertebrates.

1.2 Oceanic Island system: aids, difficulties and bias

Typically, studying colonisation on oceanic islands is facilitated by the simplified nature of islands (as described above) in relation to continental systems, and by the easiness of determining the progression rule (Wagner & Funk, 1995), which is the most common colonisation scenario predicted and tested on island systems, also known as stepping stone colonisation. Within an archipelago, islands emerge at different times, and by contrasting the age of the islands with the phylogenetic content of a lineage, inferences of the colonization pathways and speciation can be obtained (e.g. Parent et al., 2008). The progression rule predicts that colonisation will follow the formation of the islands with direction starting from the old islands down to the young islands (Wagner & Funk, 1995). This is the predominant pattern of colonisation described for Hawaiian taxa (reviewed in Gillespie & Roderick, 2002), and it has also been observed for several Canary Island taxa (reviewed in Juan et al., 2000). Island systems also facilitate the study of the processes that lead to diversification and species formation, and phylogenetic information have helped to define a number of factors responsible for generating species richness on a given archipelago. Multiple colonization events, increased speciation through bottleneck and founder flush events, diversification *in situ* from a founding population into a number of species caused by vicariant events or by adapting to diverse environmental niches, are some of these factors (Emerson 2002). The main assumption is that "for a given species group within an archipelago, the species assemblage of an individual island within the archipelago is the result of a single colonisation event, thus, suggesting within island diversification to be the mechanism generating diversity" (Emerson 2002, pg. 952).

Despite the simplicity of predictions and assumptions provided by insular systems, the interpretation of phylogenetic information and inference of colonisation pathways can be problematic due to several issues that are likely to confound genetic signatures. These issues include back colonisation to the source (continent or another island), intra-island differentiation, recent colonisation, incomplete sampling, extinction, hybridization and incomplete lineage sorting (Juan et al., 2000; Emerson, 2002). For example, within a complex of species of the beetle *Nesotes* from the western Canary Islands, inferences of island taxa monophyly have been complicated by the apparent retention of ancestral genetic polymorphism, and an intricate pattern of colonisation has been suggested for the group (Rees et al., 2001). Moreover, a phylogenetic signature of a single colonisation episode may result from hybridization and introgression between two colonising lineages that arrived independently to an island (Emerson 2002). This is a complex issue, difficult to evaluate, and only detected in a small number of studies to date, including the Hawaiian *Lapaula* crickets (Shaw, 2002), Canary Island *Aphanathrum* beetles (Jordal et al., 2006), and mockingbirds and giant tortoises of the Galápagos archipelago (Nietlisbach et al., 2013; Garrick et al., 2014). The potential for a species within an insular setting to be the result of genomic admixture following two distinct colonisation events does not seem to be negligible. In fact, it could be a representative feature within island organisms, but the genetic signatures of this phenomenon are transient, thus complicating its detection (Emerson and Faria 2014, see Appendix). For example, Garrick et al. (2014) present strong evidence that the Galápagos giant tortoise, *Chelonoidis becki*, is the result of a double colonization event involving *C. darwini* from Santiago island. Their data reveal the two founding lineages not to be reproductively isolated as they have started to coalesce back into one. As time moves forward, coalescence of the two lineages will reach completion and the signature of genomic admixture will be eroded (Emerson and Faria, 2014).

1.2.1 The first bias

Studies of colonisation and diversification on islands are typically directed toward describing evolutionary events (such as species radiation) that follow single colonisation events, and there are only a limited number of studies describing the evolutionary consequences following multiple colonisation events. This is likely due to

the difficulty of predicting where this phenomenon is going to be found, and a potential way to circumvent this difficulty is by looking within single species, or complexes of closely related species, distributed across multiple islands in an archipelago. As the unit of branch length in a phylogenetic tree indicates the evolutionary distance, which is the product of the rate of evolution (number of substitutions per site per unit time) (Baum & Smith, 2012), this can be used to get a temporal estimate for the age of the group, which in turn will indicate the possibility of detecting this phenomenon. If very long branches (deep genetic divergence) connect species within this complex, representing millions of years since each one originated on its islands, then no signature of potential double colonisation is likely to be found as it has been overwritten. However, if very short branches (shallow genetic divergence) connect species, evolutionary time has not been long enough to erode this signature, then, double colonisation events are more likely to be found.

Chapter 2 addresses this issue by investigating the history of diversification within a species complex belonging to the most diverse genus in the Canary Islands, the weevil genus *Laparocerus,* with 128 described species that are endemic to the archipelago (Warren et al., 2014). Molecular work looking at phylogenetic relationships among many of these species (Machado, unpublished), revealed a complex of 9 closely related species – the *Laparocerus tessellatus* complex, all single island endemics distributed across 4 different islands, to be connected by shallow genetic distance (short branch lengths). This suggests a recent origin for the group, and provides an opportunity to investigate the relationships among individuals sampled from the different islands to evaluate the fit of mtDNA and nuclear sequence data to a colonisation history where species are the product of a single founding event.

1.2.2 The second bias

Studies of invertebrate colonisation on islands are also biased toward large arthropods occurring above the soil surface in the most widely investigated archipelagos. For example, in the Galápagos archipelago, studies have focused on arthropods from Insecta orders such as Lepidopetra, Hemiptera, Homoptera and Coleoptera (e.g. weevils and darkling beetles) (reviewed in Parent et al., 2008). In the Hawaii, Aranea and Insecta

(Isopoda, Odonata, Lepidoptera, Hemiptera, Orthoptera, Pscoptera and Diptera) are the main orders investigated (e.g. Gillespie & Roderick, 2002; Jordan et al., 2005; Medeiros & Gillespie, 2011; Croucher et al., 2012; Lapoint et al., 2013), while in the Canary Island studies have mainly focused on Coleoptera, Orthoptera and Aranae (e.g. Juan et al., 2000; Contreras-Díaz et al., 2007; Lopez et al., 2013; Macías-Hernández et al., 2013; Husemann et al., 2014; Opatova & Arnedo, 2014). Thus, there is a tendency towards large above-soil arthropods and against small-bodied soil-dwelling invertebrates, such as springtails and mites, for which very little information is currently available.

Even though soil species are one of the biggest components of biodiversity (accounting for approximately 25% of the 1.5 million described living species on the Earth, [Decaëns, 2010]), the invertebrate component of soil has barely been investigated in terms of the dynamics of colonisation and community assembly. In part this is probably related to the difficulty of demarcating species boundaries, and their complex taxonomy. Soil fauna has been referred to as the 'third biotic frontier' due to its enormous diversity (below-ground species richness is much higher than that found for above ground fauna and vegetation), which makes the tasks of identifying and quantifying soil fauna species virtually unmanageable. The estimated average taxonomic deficit for overall soil species is 76%, and it is higher than 90% for soil organisms smaller than 100 μm (Decaëns, 2010). Additionally, species boundaries traditionally described by morphology are being challenged by recent molecular studies that reveal high levels of cryptic species, frequently associated with very deep genetic divergences (e.g. Garrick et al., 2007; Cicconardi et al., 2010; Mortimer et al., 2012). As a result, previous assumptions about shared species between islands and continents are being refuted and the apparently wide distributions and morphological stasis of many morphospecies in different groups (e.g. Collembola, Nematodes, Acari) are being reconciled with their limited dispersal ability (Emerson et al., 2011).

Likewise, the frequently vast population densities found for below-ground fauna (e.g. up to 60 000 individuals per m^2 for Collembola [Hopkin, 1997]) complicates the tasks of demarcating community boundaries which also contributes to the bias towards large arthropods studies with more discretely delimited communities (Vamosi et al., 2009).

The much better fit between morphological species and biological species typically found for large arthropods allows a great number of inferences about patterns of colonisation to be made based on presence and absence data, however, the problematic taxonomy of soil dwelling fauna (such as springtails, mites, and earthworms) greatly constrain the acquisition of presence and absence data, thus, limiting what can be inferred about their colonisation histories.

To overcome this limitation and bias, second-generation sequencing techniques like pyrosequencing, originally developed to address similar issues of demarcating community and species boundaries in the microbial fauna (e.g. Chu et al., 2010), are emerging as a possible solution. Second-generation sequencing techniques have been applied to cryptic complex eukaryote systems, such as fungi (Rousk et al., 2010), protists (Medinger et al., 2010), marine benthic meiofauna (Fonseca et al., 2010), and arthropods (Yu et al., 2012) including soil dwelling fauna (Ramirez-Gonzalez et al., 2013), and have proved to be a useful tool to obtain information for some of these elements of biodiversity that are otherwise very difficult to study.

Chapter 3 deals with these issues of bias and limitation by employing pyrosequencing data of the mtDNA COI gene obtained from a 454-sequencing platform to investigate the history of colonisation of the Collembola community in the Canary Islands, by sampling the island of Tenerife. As very little data is currently known, the aim was to describe general broad patterns that could provide some information regarding: (i) how many Collembola lineages are there in Tenerife; (ii) where do they come from; (iii) the extent to which the Collembola community is composed of older or more recently arrived lineages. Results acquired from such an analysis may serve as a baseline for more detailed studies investigating specific details of colonisation, diversification and community structure within the Collembola community of the Canary Islands.

1.3 Continental island systems: historical facts and gaps

Although oceanic island have been the focus of the majority of island studies investigating colonisation, colonisation also concerns establishment into new areas within more continuous continental landscapes, where the interaction of geography and

climate may also lead to periods of isolation and biotic extirpation, such as the relationship between Great Britain and the European continent. At the present time, there is an apparently simple scenario because island and mainland are separated by the English Channel, a clear boundary for colonisation to Great Britain from the continent. However, through periods of geological time Great Britain was connected to the continent, and for repeated cycles it was glaciated (revised in White & Schreve, 2001). The result is a complex scenario within which to understand the dynamics of colonisation and species diversification.

In this continental island setting, climate and geography have had a much greater importance (when compared to oceanic islands) in defining the hospitable terrains where species could colonise and establish. Over the last 700 thousands of years, climatic oscillations massively changed the habitable landscape of Great Britain and European surroundings due to the repeated advance and retreat of ice sheets, directly affecting biogeographic distributions, species richness and the assembly of biota (Webb & Bartlein, 1992; Hewitt, 2003). The primary consequence of these climatic changes was that species shifted their ranges to track suitable habitats and confirmation for this came initially from fossil and pollen data (Huntley & Birks, 1983; Huntley & Thompson Webb, 1989; Bennett et al., 1991). Comprehensive analysis of pollen fossil diagrams and macrofossils demonstrated that during cold periods, as ice and permafrost covered most part of northern Europe, species went extinct over large parts of their ranges, while others dispersed to new areas or survived in refugia with suitable climate (Bennett et al., 1991 and references therein). These refugia were mainly located in southern Europe, which maintained suitable habitats due to their latitudinal position. During warm periods, ice released areas became hospitable and surviving species were able to expand their distributions and colonise new territories. These expansion and contraction range changes occurred repeatedly following the climatic oscillations and fossil data in ice cores indicated these post-glacial expansions to have been remarkably rapid for many species (e.g. Coope et al., 1977; Huntley & Birks, 1983; Bennett, 1985).

With the advent of molecular genetics, the consequence of these climatic changes on the genetic structure of species was also revealed. Paleontological studies were corroborated and advanced since it was possible to describe intraspecific geographical

structure (through the identification of lineages) and identify postglacial colonisation routes, when the location of refugia were known (Hewitt, 1996, 1999; Taberlet et al., 1998). The processes associated with range shifts (reduction in population size, fragmentation and range expansion) during the Pleistocene also impacted the genetic variability of species and this has been demonstrated for many plant and animal species (e.g Petit et al., 2003; Hewitt, 2011 and references therein). In southern Europe, as a result of the long-term isolation in refugia and varied topography, lineages are generally highly divergent, particularly if they were not the source of postglacial expansion. Likewise, due to the successive founder events following rapid postglacial colonisation, intraspecific diversity tends to decline away from refugia thus being especially low in northern Europe (e.g. Hewitt, 2004).

Concordant geographical structures found among many taxa have demonstrated how refugial genomes contributed to the re-colonization of central and northern Europe, and three general expansion routes from the southern Mediterranean refugia have been proposed (Hewitt, 1999). Contemporary biotas in northern European areas are either the result of postglacial expansion predominantly from a Balkan refugium (as exemplified by the grasshopper *Chorthippus parallellus*), or they are the result of expansion from refugia in the three Mediterranean peninsulas (as with the hedgehog *Erinaceus concolor*). The third pattern identifies distinct western and eastern lineages, suggesting expansion from both Iberian and Balkan refugia (as in the case of the brown bear *Ursus actors* [reviewed in Hewitt, 2011]). Particularly for Great Britain, the prediction that has arisen from these model studies is that, following the retreat of the British-Irish ice sheet after the last glacial maximum, fauna and flora arrived in the island exclusively from southern European refuges. Studies have supported this prediction, revealing many species that re-colonised Great Britain from Spain, such as oaks, shrews, hedgehogs and bears, while others recolonised from the Balkans, such as grasshoppers, alder, beech and newts (Hewitt, 2000).

In concert with the findings of general extinction in northern European areas and survival in southern refugia, fossil and genetic data have also provided evidence for the survival of species in unexpected areas outside the Mediterranean refugia (e.g. Kullman, 1998; Ukkonen et al., 1999; Stewart & Lister, 2001; Hänfling et al., 2002). Although

initially less recognised, possibly due to inconclusive data and their smaller impact (small size, much less number of taxa is found on them) and rarity (very improbable), it is now well recognised that non-Mediteranean refugial areas may also be considered when interpreting current genetic structures. The first genetic evidence for the existence of 'northern cryptic refugia' was identified for the pygmy shrew (*Sorex araneus* and *S. minutus*) and the bank vole (*Clethrionomys glareolus*) in central Europe (Bilton et al., 1998). Since then, a variety of studies have revealed evidence for non-Mediterranean refugia in a number of taxa (revised in Bhagwat & Willis, 2008; and Benke et al., 2009), including some species in the southern region of the British Isles, as well as in the North Sea (e.g. Bernatchez, 2001; Hänfling et al., 2002; Hoarau et al., 2007) (see Appendix 2, Tab S1).

The phylogeographic study of invertebrates, particularly those with very limited dispersal ability, has been recognised to be suitable for highlighting patterns of survival and recolonisation during Pleistocene glaciations (Nieberding et al., 2005; Garrick et al., 2007). Despite this, only a few studies to date have looked into the phylogeography of invertebrates within Europe and Great Britain (Thomaz et al., 1996; Davison, 200; Goodacre et al. 2006; Nieberding et al., 2005; McInerney et al., 2014) and to our knowledge none exist for very low dispersive taxa, such as soil dwelling fauna. In fact, Pleistocene phylogeographic studies looking at European taxa are biased towards vertebrate taxa (e.g. fishes, amphibians, birds and mammals), above ground taxa, single species, and taxa that survived the Pleistocene in southern refugia (Benke et al., 2009; also see Appendix 2, Tab S1).

The final two chapters of the thesis address these issues by investigating the recolonisation history of the Collembola fauna of Great Britain following the last maximum glacial retraction, using the mtDNA COI barcode gene. Besides looking at the patterns of diversification and lineage richness within this continental island, the main question addressed was whether Collembola recolonisation has been entirely from mainland Europe or whether there is evidence that part of the fauna persisted within Great Britain through the Last Glacial Maximum, thus contributing to the process of post-glacial recolonisation.

In summary, this thesis employs molecular tools (mtDNA and nuclear Sanger sequencing and high throughput parallel sequencing) to investigate the history of invertebrate colonisation and diversification within two scenarios, (1) an oceanic archipelago; and (2) a continental island setting where climate and geography have a long-term historical role. The aim was to address three gaps in our understanding about island colonisation, speciation and the assembly of the biota. In the oceanic islands system - the Canary Islands, the gaps addressed were: (i) the possibility that genomic admixture among multiple founding lineages has featured in the recent history of diversification of a very species rich coleopteran genus; and (ii) the lack of information regarding the colonisation history and dynamics of the small arthropod soil dwelling fauna. In the continental island system - Great Britain, the gap addressed was the under-explored possibility that the UK was not completely defaunated and recolonised from external sources, but that a more complex pattern, involving persistence within small cryptic refugia, may have featured in the history of its invertebrate soil dwelling fauna.

The timeframes involved in colonisation of the two island scenarios investigated (Canary Islands versus postglacial Great Britain) are very different when comparing the age of their formation and the time available for colonisation. The Canary Islands were available for colonisation much earlier (upon their formation, from 20 Ma for the oldest up to 1.1 Ma for the youngest islands, with the age of the islands providing upper limits for their colonisation from the continent and other islands) than the lands of Great Britain, which were greatly affected by the Quaternary glaciations due to their latitudinal position as explained above. The surface of Great Britain was still dramatically covered by ice up to 15,000 years ago (the Last Glacial Maximum extended from approximately 27,000-15,000 years, [Clark et al. 2012]), thus, fauna and flora are thought to have arrived from the mainland only recently, following the retreat of the British-Irish ice sheet after the Last Glacial Maximum. On the other hand, when taking into account the dynamic history of geological events that affected the Canary Islands (for example, the giant landslide in Tenerife that formed the Orotava Valley is estimated to have occurred around 0.72 -0.27 Ma [Watts and Masson, 1995]; the periodic volcanic activity that has occurred within the last million years in most of the Canary Islands [Carracedo et al. 1998]), which likely provoked local extinctions of

many taxa, the discrepancy between the timeframes is greatly (at least for the land affected by these events).

Despite the different time frame, the evolutionary processes operating in these two areas are expected to be similar (e.g. migration, colonisation, differentiation, adaptation, population expansion, speciation, extinction) although their frequency and spatial and temporal scale may be different. For example, large local extinctions (caused by ice or volcanisms), founder flush events, diversification in situ, historical habitat disjunctions, recent expansions are operating in both island scenarios. When taking into account time since colonisation, for the older Canary Islands, lineages have had time to colonise different niches, differentiate and adapt to different habitats. Whereas for the younger post-glacial Great Britain, time since colonisation is recent, species are found colonising different habitats but they have not had the time to specialise and differentiate yet. Thus, we are more likely to find species on the Canary Islands that are at later stages of species formation than those species that are the result of post-glacial colonisation in Great Britain. Moreover, when considering the diversity of habitats in these two areas, it is clear that the Canary Islands is naturally much richer in habitats than Great Britain, thus there are more opportunities for genetic differentiation in the oceanic islands which is reflected by its much greater diversity.

1.4 References

Baum D. & Smith S. (2012) Using Trees to Study Space, Time, and Evolutionary Diversification. *Tree Thinking: An Introduction to Phylogenetic Biology* pp. 349–388. Roberts and Company Publishers, Greenwood Village, Colorado.

Benke M., Brändle M., Albrecht C., & Wilke T. (2009) Pleistocene phylogeography and phylogenetic concordance in cold-adapted spring snails (*Bythinella* spp.). *Molecular Ecology*, **18**, 890–903.

Bennett K.D. (1985) The spread of Fagus grandifolia across eastern North America during the last 18,000 years. *Journal of Biogeography*, **12**, 147–164.

Bennett K.D., Tzedakis P.C., & Willis K.J. (1991) Quaternary Refugia of North European Trees. *Journal of Biogeography*, **18**, 103–115.

Bernatchez L. (2001) The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution; International Journal of Organic Evolution*, **55**, 351–379.

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

Bilton D.T., Mirol P.M., Mascheretti S., Fredga K., Zima J., & Searle J.B. (1998) Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society B: Biological Sciences*, **265**, 1219–1226.

Carracedo J.C., Day S.J., Guillou H., Badiola E.R., Canas J.A., & Pérez Torrado F.J. (1998) Hotspot volcanism close to a passive continental margin: The Canary Islands. *Geol. Mag*., **135**, 591–604.

Chiverrell R.C. & Thomas G.S.P. (2010) Extent and timing of the Last Glacial Maximum (LGM) in Britain and Ireland: a review. *Journal of Quaternary Science*, **25**, 535–549.

Chu H., Fierer N., Lauber C.L., Caporaso J.G., Knight R., & Grogan P. (2010) Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environmental Microbiology*, **12**, 2998–3006.

Cicconardi F., Nardi F., Emerson B.C., Frati F., & Fanciulli P.P. (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology*, **19**, 386–400.

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

Clegg S.M. (2010) Evolutionary changes following island colonization in birds: empirical insights into the roles of microevolutionary processes. *The Theory of Island Biogeography Revisited* pp. 293– 325. Princeton University Press, Princeton, NJ.

Contreras-Díaz H.G., Moya O., Oromí P., & Juan C. (2007) Evolution and diversification of the forest and hypogean ground-beetle genus *Trechus* in the Canary Islands. *Molecular Phylogenetics and Evolution*, **42**, 687–699.

Coope G.R., Pennington W., Mitchell G.F., West R.G., Morgan A.V., & Peacock J.D. (1977) Fossil Coleopteran Assemblages as Sensitive Indicators of Climatic Changes During the Devensian (Last) Cold Stage [and Discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **280**, 313–340.

Croucher P.J.P., Oxford G.S., Lam A., Mody N., & Gillespie R.G. (2012) Colonization History and Population Genetics of the Color-Polymorphic Hawaiian Happy-Face Spider *Theridion Grallator* (Araneae, Theridiidae). *Evolution*, **66**, 2815–2833.

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.

Decaëns T. (2010) Macroecological patterns in soil communities. *Global Ecology and Biogeography*, **19**, 287–302.

Emerson B.C. (2002) Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology*, **11**, 951–966.

Emerson B.C., Cicconardi F., Fanciulli P.P., & Shaw P.J.A. (2011) Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2391–2402.

Emerson B.C. & Gillespie R.G. (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology & Evolution*, **23**, 619–630.

Fernández-Palacios J.M., de Nascimento L., Otto R., Delgado J.D., García-del-Rey E., Arévalo J.R., & Whittaker R.J. (2011) A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography*, **38**, 226–246.

Fonseca V.G., Carvalho G.R., Sung W., Johnson H.F., Power D.M., Neill S.P., Packer M., Blaxter M.L., Lambshead P.J.D., Thomas W.K., & Creer S. (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications*, **1**, 1–8.

Garrick R.C., Benavides E., Russello M.A., Hyseni C., Edwards D.L., Gibbs J.P., Tapia W., Ciofi C., & Caccone A. (2014) Lineage fusion in Galápagos giant tortoises. *Molecular Ecology*, **23**, 5276–5290.

Garrick R.C., Sands C.J., Rowell D.M., Hillis D.M., & Sunnucks P. (2007) Catchments catch all: long-term population history of a giant springtail from the southeast Australian highlands - a multigene approach. *Molecular Ecology*, **16**, 1865–1882.

Gillespie R.G. & Roderick G.K. (2002) Arthropods on islands: Colonization, Speciation, and Conservation. *Annual Review of Entomology*, **47**, 595–632.

Goodacre S.L., Thomaz D., & Davies E.K. (2006) Mitochondrial gene diversity in *Cepaea* – population structure, history and positive selection. Biological Journal of the Linnean Society, 87, 167–184.

Grant P.R. (1998) *Evolution on Islands*. Oxford University Press, Oxford, UK.

Gupta S., Collier J.S., Palmer-Felgate A., & Potter G. (2007) Catastrophic flooding origin of shelf valley systems in the English Channel. *Nature*, **448**, 342–345.

Hänfling B., Hellemans B., Volckaert F. a. M., & Carvalho G.R. (2002) Late glacial history of the cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. *Molecular Ecology*, **11**, 1717–1729.

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.

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

Hewitt G.M. (2003) Ice Ages: their impact on species distributions and evolution. *Evolution of Planet Earth* (ed. by L.J. Rothschild and A.M. Lister), Academic Press.

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

Hewitt G.M. (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica*, **139**, 617–638.

HilleRisLambers J., Adler P.B., Harpole W.S., Levine J.M., & Mayfield M.M. (2012) Rethinking Community Assembly through the Lens of Coexistence Theory. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 227–248.

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

Hopkin S.P. (1997) *Biology of the Springtails*. Oxford University Press, Oxford, UK.

Huntley B. & Birks H.J.B. (1983) *An Atlas of Past and Present Pollen Maps for Europe: 0-13000 Years Ago*. Cambridge University Press, Cambridge.

Huntley B. & Thompson Webb I. (1989) Migration: Species' Response to Climatic Variations Caused by Changes in the Earth's Orbit. *Journal of Biogeography*, **16**, 5–19.

Husemann M., Deppermann J., & Hochkirch A. (2014) Multiple independent colonization of the Canary Islands by the winged grasshopper genus *Sphingonotus* Fieber, 1852. *Molecular Phylogenetics and Evolution*, **81**, 174–181.

Jordal B.H., Emerson B.C., & Hewitt G.M. (2006) Apparent "sympatric" speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology*, **15**, 2935–2947.

Jordan S., Simon C., Foote D., & Englund R.A. (2005) Phylogeographic patterns of Hawaiian Megalagrion damselflies (Odonata : Coenagrionidae) correlate with Pleistocene island boundaries. *Molecular Ecology*, **14**, 3457–3470.

Juan C., Emerson B.C., Oromí P., Hewitt G.M. (2000) Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology & Evolution*, **15**, 104–109.

Kullman L. (1998) Non-analogous tree flora in the Scandes Mountains, Sweden, during the early Holocene - macrofossil evidence of rapid geographic spread and response to palaeoclimate. *Boreas*, **27**, 153–161.

Lapoint R.T., O'Grady P.M., & Whiteman N.K. (2013) Diversification and dispersal of the Hawaiian Drosophilidae: The evolution of *Scaptomyza*. *Molecular Phylogenetics and Evolution*, **69**, 95–108.

Lopez H., Hernandez-Teixidor D., Macias-Hernandez N., Juan C., & Oromi P. (2013) A taxonomic revision and species delimitation of the genus *Purpuraria* Enderlein, 1929 (Orthoptera: Pamphagidae) using an integrative approach. *Journal of Zoological Systematics and Evolutionary Research*, **51**, 173–186.

Macías-Hernández N., Bidegaray-Batista L., Emerson B.C., Oromí P., & Arnedo M. (2013) The Imprint of Geologic History on Within-Island Diversification of Woodlouse-Hunter Spiders (Araneae, Dysderidae) in the Canary Islands. *Journal of Heredity*, **104**, 341–356.

McInerney C.E., Maurice L., Robertson A.L., Knight L.R.F.D., Arnscheidt J., Venditti C., Dooley J.S.G., Mathers T., Matthijs S., Eriksson K., Proudlove G.S., & Haenfling B. (2014) The ancient Britons: groundwater fauna survived extreme climate change over tens of millions of years across NW Europe. *Molecular Ecology*, **23**, 1153–1166.

Medeiros M.J. & Gillespie R.G. (2011) Biogeography and the evolution of flightlessness in a radiation of Hawaiian moths (Xyloryctidae: *Thyrocopa*). *Journal of Biogeography*, **38**, 101–111.

Medinger R., Nolte V., Pandey R.V., Jost S., Ottenwälder B., Schlötterer C., & Boenigk J. (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Molecular ecology*, **19**, 32–40.

Montgomery W.I., Provan J., McCabe A.M., & Yalden D.W. (2014) Origin of British and Irish mammals: disparate post-glacial colonisation and species introductions. *Quaternary Science Reviews*, **98**, 144–165.

Mortimer E., Van Vuuren B.J., Meiklejohn K.I., & Chown S.L. (2012) Phylogeography of a mite, *Halozetes fulvus*, reflects the landscape history of a young volcanic island in the sub-Antarctic. *Biological Journal of the Linnean Society*, **105**, 131–145.

Nieberding C., Libois R., Douady C.J., Morand S., & Michaux J.R. (2005) Phylogeography of a nematode (*Heligmosomoides polygyrus*) in the western Palearctic region: persistence of northern cryptic populations during ice ages? *Molecular Ecology*, **14**, 765–779.

Nietlisbach P., Wandeler P., Parker P.G., Grant P.R., Grant B.R., Keller L.F., & Hoeck P.E.A. (2013) Hybrid ancestry of an island subspecies of Galápagos mockingbird explains discordant gene trees. *Molecular Phylogenetics and Evolution*, **69**, 581–592.

Opatova V. & Arnedo M.A. (2014) Spiders on a Hot Volcanic Roof: Colonisation Pathways and Phylogeography of the Canary Islands Endemic Trap-Door Spider *Titanidiops canariensis* (Araneae, Idiopidae). *PLoS ONE*, **9**, e115078.

Parent C.E., Caccone A., & Petren K. (2008) Colonization and diversification of Galápagos terrestrial fauna: a phylogenetic and biogeographical synthesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 3347–3361.

Petit R.J., Aguinagalde I., Beaulieu J. L. de, Bittkau C., Brewer S., Cheddadi R., Ennos R., Fineschi S., Grivet D., Lascoux M., Mohanty A., Müller-Starck G., Demesure-Musch B., Palmé A., Martín J.P., Rendell S., & Vendramin G.G. (2003) Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. *Science*, **300**, 1563–1565.

Pettitt P. & White M. (2012) *The British Palaeolithic: Human Societies at the Edge of the Pleistocene World*. Routledge, London.

Ramirez-Gonzalez R., Yu D.W., Bruce C., Heavens D., Caccamo M., & Emerson B.C. (2013) PyroClean: Denoising Pyrosequences from Protein-Coding Amplicons for the Recovery of Interspecific and Intraspecific Genetic Variation. *PLoS ONE*, **8**, 1–11.

Rees D.J., Emerson B.C., Oromí P., & Hewitt G.M. (2001) Mitochondrial DNA, ecology and morphology: interpreting the phylogeography of the *Nesotes* (Coleoptera: Tenebrionidae) of Gran Canaria (Canary Islands). *Molecular Ecology*, **10**, 427–434.

Ricklefs R.E. (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, **7**, 1–15.

Roderick G.K. & Gillespie R.G. (1998) Speciation and phylogeography of Hawaiian terrestrial arthropods. *Molecular Ecology*, **7**, 519–531.

Rousk J., Bååth E., Brookes P.C., Lauber C.L., Lozupone C., Caporaso J.G., Knight R., & Fierer N. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME Journal*, **4**, 1340–1351.

Shaw K.L. (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences*, **99**, 16122–16127.

Stewart J.R. & Lister A.M. (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, **16**, 608–613.

Stewart J.R., Lister A.M., Barnes I., & Dalen L. (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 661–671.

Taberlet P., Fumagalli L., Wust-Saucy A.-G., & Cosson J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Thomaz D., Guiller A., & Clarke B. (1996) Extreme Divergence of Mitochondrial DNA within Species of Pulmonate Land Snails. Proceedings of the Royal Society of London B: Biological Sciences, 263, 363–368.

Thornton I.W.B., Cook S., Edwards J.S., Harrison R.D., Schipper C., Shanahan M., Singadan R., & Yamuna R. (2001) Colonization of an island volcano, Long Island, Papua New Guinea, and an emergent island, Motmot, in its caldera lake. VII. Overview and discussion. *Journal of Biogeography*, **28**, 1389–1408.

Ukkonen P., Lunkka J.P., Jungner H., & Donner J. (1999) New radiocarbon dates from Finnish mammoths indicating large ice-free areas in Fennoscandia during the Middle Weichselian. *Journal of Quaternary Science*, **14**, 711–714.

Vamosi S.M., Heard S.B., Vamosi J.C., & Webb C.O. (2009) Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology*, **18**, 572–592.

Wagner W.L. & Funk V.A. (1995) *Hawaiian Biogeography: Evolution on a Hot Spot Archipelago*. Smithsonian, Washington.

Warren B.H., Simberloff D., Ricklefs R.E., Aguilée R., Condamine F.L., Gravel D., Morlon H., Mouquet N., Rosindell J., Casquet J., Conti E., Cornuault J., Fernández-Palacios J.M., Hengl T., Norder S.J., Rijsdijk K.F., Sanmartín I., Strasberg D., Triantis K.A., Valente L.M., Whittaker R.J., Gillespie R.G., Emerson B.C., & Thébaud C. (2014) Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecology Letters*, **doi:10.1111/ele.12398**

Watts A.B. & Masson D.G. (1995) A giant landslide on the north flank of Tenerife, Canary Islands. Journal of Geophysical Research: Solid Earth, 100, 24487–24498

Webb T. & Bartlein P.J. (1992) Global changes during the last 3 million years: climatic controls and biotic responses. *Annual review of Ecology and Systematics*, **23**, 141–173.

Weninger B., Schulting R., Bradtmoeller M., Clare L., Collard M., Edinborough K., Hilpert J., Joeris O., Niekus M., Rohling E.J., & Wagner B. (2008) The catastrophic final flooding of Doggerland by the Storegga Slide tsunami. *Documenta Praehistorica, Vol Xxxv*, **35**, 1–24.

White M.J. & Schreve D.C. (2001) Island Britain-Peninsula Britain : palaeogeography, colonisation and the Lower Palaeolithic settlement of the British Isles. *Proceedings of the Prehistoric Society*, **66**, 1–28.

Whittaker R.J. & Fernandez-Palacios J.M. (2007) *Island Biogeography: Ecology, Evolution, and Conservation*. Oxford University Press, Oxford, UK.

Wilson J.J., Rougerie R., Schonfeld J., Janzen D.H., Hallwachs W., Hajibabaei M., Kitching I.J., Haxaire J., & Hebert P.D. (2011) When species matches are unavailable are DNA barcodes correctly assigned to higher taxa? An assessment using sphingid moths. *BMC Ecology*, **11**, 1–14.

Yalden D.W. (1982) When did the mammal fauna of the British Isles arrive? *Mammal Review*, **12**, 1–57.

Yu D.W., Ji Y., Emerson B.C., Wang X., Ye C., Yang C., & Ding Z. (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, **3**, 613–623.

1.5 Appendices

Appendix 1 - Canary Islands: brief outline on their formation and biodiversity

The Canary Islands are of particular interest both historically and biologically, and from a geological point of view they are one of the best-studied oceanic island systems (Fernández-Palacios et al., 2011), which make them an ideal template for the study of evolution. The archipelago is an island chain of seven major islands and several islets situated in the northeast Atlantic Ocean, about 500 km north from the Tropic of Cancer. The main islands are aligned from east to west: Lanzarote, Fuerteventura, Gran Canaria, Tenerife, La Gomera, La Palma and El Hierro, with Fuerteventura lying closest to the African continent (110 km), and La Palma situated the most distant (460 km). The origin of the archipelago is controversial, but there is geological support for a hotspot origin (van den Bogaard, 2013). The central and western islands are separated by deep water and have never been connected to one another, or to the continent. The most easterly islands Lanzarote, Fuerteventura and their outlying islets formed a single landmass during Pleistocene sea-lowering events. All the islands were formed in the past 25 million years (Mya), and recent K-Ar and Sr-Nd-Pb isotopic dating shows a general reduction in the age of the islands from east to west: El Hierro arose approximately 1.1 Mya, La Palma 1.7 Mya, La Gomera 12 Mya, Tenerife between 11.9 and 8.9 Mya, Gran Canaria 14.5 Mya, Lanzarote 15.5 Mya, and Fuerteventura 20 Mya (Guillou et al., 2004; van den Bogaard, 2013). These ages provide upper limits for their colonization from the continent and other islands. With the exception of La Gomera, all islands have had periodic volcanic activity after their formation, even within the last million years (Carracedo et al., 1998; Fernández-Palacios et al., 2011).

Species endemism are high in the Canary Islands for many groups of plants and invertebrates which diversified throughout the islands in a series of distinct species and subspecies: about 27% of the approximately 1000 native vascular plant species and 50% of the terrestrial invertebrate fauna (c. 6500 species) are endemic (Juan et al. 2000). Diversification within the archipelago is found among vertebrates, markedly in geckos (Nogales et al., 1998), lizards (Cox et al., 2010) and skinks (Brown & Pestano, 1998); with subspecies divergences within some birds (finches and blue tits) (Illera et al.,

2012), but the most notable examples of diversification can be found among the invertebrates (Juan et al., 2000). The species richness of the Canary Islands has in part been attributed to the great variety of habitats present within the archipelago, and this has attracted much attention from evolutionary biologists and ecologists. The islands have a subtropical climate with warm temperatures and the variation between seasons is small. Humid trade winds from northeast strongly influence the climate which, together with the drier northwest winds blowing at higher levels and the altitude of the volcanoes (e.g. 3718 m, El Teide, Tenerife), create an inversion zone and a distinct vertical stratification of the vegetation. Five vegetational strata can be distinguished: (1) from sea level up to 250 m - arid subtropical scrub; (2) from 250 to 600/800 m altitude – humid and semi-arid subtropical scrub and woods; (3) from 600 to 1200 m – humid laurel forest in the cloud belt (windward slopes); (4) from 800/1000 to 2000 m – humid to dry temperate pine forest; and (5) over 2000 m – dry subalpine scrub. Above 3000 m, a 'stone desert' stratum, with almost no vegetation, also appears in Tenerife (Morales & Pérez, 2000). The Canary Island flora and fauna has been linked to that of Madeira and, in part, to that of the Azores and the Cape Verde Islands but it also presents affinities to the Mediterranean region and, for a few groups, with other more distant regions such as East Africa, India, South America, Australia, and America. Two probable sources of colonizers are neighbouring North Africa and the Iberian Peninsula, in view of the prevailing winds and the sea currents (Juan et al., 2000).

The Canary Islands have been the focus of many phylogenetic studies with molecular techniques being employed to investigate the evolution of several groups of plant and animal species, such as *Gallotia* (lizard) (Thorpe et al., 1994), *Laurus azorica* (Canary Island laurel) (Arroyo-García et al., 2001), S*teganacaru*s (oribatid mites) (Salomone et al., 2002), *Napaeus* (land snail) (Alonso et al., 2006), bringing insights into the origins of species diversity on islands. The arthropod fauna has in particular been a focus of investigation for understanding the origins of diversity (Juan et al., 1995, 1996; Arnedo et al., 2001; Jordal & Hewitt, 2004; Emerson & Oromí, 2005) due to its high species richness within the archipelago. While many genera have been subject to higher-level phylogenetic analysis, detailed studies charting the course of evolution over shorter time periods within species are less common (but see Rees et al., 2001; Jordal et al., 2006)

References

Alonso M.R., Goodacre S.L., Emerson B.C., Ibáñez M., Hutterer R., & Groh K. (2006) Canarian land snail diversity: conflict between anatomical and molecular data on the phylogenetic placement of five new species of Napaeus (Gastropoda, Pulmonata, Enidae). *Biological Journal of the Linnean Society*, **89**, 169–187.

Arnedo M.A., Oromí P., & Ribera C. (2001) Radiation of the Spider Genus Dysdera (Araneae, Dysderidae) in the Canary Islands: Cladistic Assessment Based on Multiple Data Sets. *Cladistics*, **17**, 313–353.

Arroyo-García R., Martínez-Zapater J.M., Prieto J.A.F., & Álvarez-Arbesú R. (2001) AFLP evaluation of genetic similarity among laurel populations (Laurus L.). *Euphytica*, **122**, 155–164.

Van den Bogaard P. (2013) The origin of the Canary Island Seamount Province - New ages of old seamounts. *Scientific Reports*, **3**, 2107.

Brown R.P. & Pestano J. (1998) Phylogeography of skinks (Chalcides) in the Canary Islands inferred from mitochondrial DNA sequences. *Molecular Ecology*, **7**, 1183–1191.

Carracedo J.C., Day S.J., Guillou H., Badiola E.R., Canas J.A., & Pérez Torrado F.J. (1998) Hotspot volcanism close to a passive continental margin: The Canary Islands. *Geol. Mag.*, **135**, 591–604.

Cox S.C., Carranza S., & Brown R.P. (2010) Divergence times and colonization of the Canary Islands by Gallotia lizards. *Molecular Phylogenetics and Evolution*, **56**, 747–757.

Emerson B.C. & Oromí P. (2005) Diversification of the forest beetle genus Tarphius on the Canary Islands, and the evolutionary origins of island endemics. *Evolution; International Journal of Organic Evolution*, **59**, 586–598.

Fernández-Palacios J.M., de Nascimento L., Otto R., Delgado J.D., García-del-Rey E., Arévalo J.R., & Whittaker R.J. (2011) A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography*, **38**, 226–246.

Guillou H., Carracedo J.C., Paris R., & Pérèz Torrado F.J. (2004) Implications for the early shield-stage evolution of Tenerife from K/Ar ages and magnetic stratigraphy. *Earth and Planetary Science Letters*, **222**, 599–614.

Illera J.C., Rando J.C., Richardson D.S., & Emerson B.C. (2012) Age, origins and extinctions of the avifauna of Macaronesia: a synthesis of phylogenetic and fossil information. *Quaternary Science Reviews*, **50**, 14–22.

Jordal B.H., Emerson B.C., & Hewitt G.M. (2006) Apparent “sympatric” speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology*, **15**, 2935–2947.

Jordal B. & Hewitt G. (2004) The Origin and Radiation of Macaronesian Beetles Breeding in Euphorbia: The Relative Importance of Multiple Data Partitions and Population Sampling. *Systematic Biology*, **53**, 711–734.

Juan C., Emerson B.C., Oromí P., Hewitt G.M. (2000) Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology & Evolution*, **15**, 104–109.

Juan C., Oromi P., & Hewitt G.M. (1995) Mitochondrial DNA Phylogeny and Sequential Colonization of Canary Islands by Darkling Beetles of the Genus *Pimelia* (Tenebrionidae). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **261**, 173–180.

Juan C., Oromi P., & Hewitt G.M. (1996) Phylogeny of the genus *Hegeter* (Tenebrionidae, Coleoptera) and its colonization of the Canary Islands deduced from Cytochrome Oxidase I mitochondrial DNA sequences. *Heredity*, **76 (Pt 4)**, 392–403.

Morales G. & Pérez R. (2000) *Gran Atlas Temático de Canarias*. Editorial Interinsular Canaria, Santa Cruz de Tenerife, Tenerife.

Nogales M., López M., Jiménez-Asensio J., Larruga J.M., Hernández M., & González P. (1998) Evolution and biogeography of the genus *Tarentola* (Sauria: Gekkonidae) in the Canary Islands, inferred from mitochondrial DNA sequences. *Journal of Evolutionary Biology*, **11**, 481–494.

Rees D.J., Emerson B.C., Oromí P., & Hewitt G.M. (2001) Reconciling gene trees with organism history: the mtDNA phylogeography of three *Nesotes* species (Coleoptera: Tenebrionidae) on the western Canary Islands. *Journal of Evolutionary Biology*, **14**, 139–147.

Salomone N., Emerson B.C., Hewitt G.M., & Bernini F. (2002) Phylogenetic relationships among the Canary Island Steganacaridae (Acari, Oribatida) inferred from mitochondrial DNA sequence data. *Molecular Ecology*, **11**, 79–89.

Thorpe R.S., McGregor D.P., Cumming A.M., & Jordan W.C. (1994) DNA Evolution and Colonization Sequence of Island Lizards in Relation to Geological History: Mtdna Rflp, Cytochrome B, Cytochrome Oxidase, 12s RRNA Sequence, and Nuclear Rapd Analysis. *Evolution*, **48**, 230.

Appendix 2

Table S1 - Overview of previous studies suggesting refugia in British Isles taxa. The data used for this review were compiled from previous revisions up to 2008 (e.g. Baghawart & Willis 2008, Benke et al. 2009), and from searches in Web of Science 2008-2015. Articles were searched for the terms: "refug*" and "glacial" or "Pleistocene" and "England" or "British Isles" or "Great Britain" or "UK"

Scientific name	Group	Attribute	Evidence	British Isles refugia	References
Ursus arctos	brown bear	cold adapted	fossil	Great Britain	(Aldhouse-Green & Pettitt, 1998)
			contemporary and ancient mtDNA	Northern England	(Edwards et al., 2014)
Pinus sylvestris	tree	cod adapted	mtDNA	Ireland, Scotland	(Sinclair et al., 1998, 1999)
Salmo trutta	fresh water fish	cod adapted	mtDNA	Southern England	(Bernatchez, 2001)
Cottus gobio	fresh water fish	temperate/cold tolerant	microsatellite/mtDNA, allozyme	Southern England	(Hänfling et al., 2002)
Carex digitata	tree	temperate/cold tolerant	allozyme	Ireland, Scotland	(Tyler, 2002)
Heligmosomoides polygyrus	nematode (parasite)	temperate	mtDNA	Ireland, Scotland	(Nieberding et al., 2005)
Fucus serratus	seaweed	temperate	mtDNA	English Channel	(Coyer et al., 2003; Hoarau et al., 2007)
Palmaria palmata	seaweed	artic-cold temperate	nuclear and plastid	English Channel	(Provan et al., 2005)
Mustela erminea	stoats	thermally adapted	mtDNA	Ireland	(Martínková et al., 2007)
Tree vegetation *(Quercus, Pinus, etc.)*	tree	thermophilous	speleothem pollen	Yorkshire, Lancashire, Mendip Hills	(Caseldine et al., 2008)

Table S1 (cont)

Scientific name	Group	Attribute	Evidence	British Isles refugia	References
Rana temporaria	common frog	cold tolerant	mtDNA	Ireland	(Teacher et al., 2009)
Lepus timidus	mountain hare	cold tolerant	mtDNA, nuclear, microsatellite	Ireland	(Hughes 2009 unpub. report in Montgomery et al., 2014)
Vulpes vulpes	red fox	cold adapted	contemporary and ancient mtDNA	Britain and Ireland	(Edwards et al., 2012)
Niphargus glenniei, N. irlandicus	amphipoda	obligate groundwater	mtDNA	Ireland, Southern England	(McInerney et al., 2014)
?	shrub and tree vegetation	?	?	Southern England	(Soffer 1990 in Hänfling et al., 2002)
?	heath	?	?	Southern England	(Vincent 1990 in Hänfling et al., 2002)
?	gastropoda	?	?	Southern England	(Vincent 1990 in Hänfling et al., 2002)

? not specified

References

Aldhouse-Green S. & Pettitt P. (1998) Paviland Cave: contextualizing the "Red Lady." *Antiquity*, **72**, 756–772.

Bernatchez L. (2001) The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution; International Journal of Organic Evolution*, **55**, 351–379.

Caseldine C.J., McGarry S.F., Baker A., Hawkesworth C., & Smart P.L. (2008) Late Quaternary speleothem pollen in the British Isles. *Journal of Quaternary Science*, **23**, 193–200.

Coyer J.A., Peters A.F., Stam W.T., & Olsen J.L. (2003) Post-ice age recolonization and differentiation of *Fucus serratus* L. (Phaeophyceae; Fucaceae) populations in Northern Europe. *Molecular Ecology*, **12**, 1817–1829.

Edwards C.J., Ho S.Y.W., Barnett R., Coxon P., Bradley D.G., Lord T.C., & O'Connor T. (2014) Continuity of brown bear maternal lineages in northern England through the Last-glacial period. *Quaternary Science Reviews*, **96**, 131–139.

Edwards C.J., Soulsbury C.D., Statham M.J., Ho S.Y.W., Wall D., Dolf G., Iossa G., Baker P.J., Harris S., Sacks B.N., & Bradley D.G. (2012) Temporal genetic variation of the red fox, *Vulpes vulpes*, across western Europe and the British Isles. *Quaternary Science Reviews*, **57**, 95–104.

Hänfling B., Hellemans B., Volckaert F. a. M., & Carvalho G.R. (2002) Late glacial history of the cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. *Molecular Ecology*, **11**, 1717–1729.

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

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

McInerney C.E., Maurice L., Robertson A.L., Knight L.R.F.D., Arnscheidt J., Venditti C., Dooley J.S.G., Mathers T., Matthijs S., Eriksson K., Proudlove G.S., & Hänfling B. (2014) The ancient Britons: groundwater fauna survived extreme climate change over tens of millions of years across NW Europe. *Molecular Ecology*, **23**, 1153–1166.

Montgomery W.I., Provan J., McCabe A.M., & Yalden D.W. (2014) Origin of British and Irish mammals: disparate post-glacial colonisation and species introductions. *Quaternary Science Reviews*, **98**, 144–165.

Nieberding C., Libois R., Douady C.J., Morand S., & Michaux J.R. (2005) Phylogeography of a nematode (*Heligmosomoides polygyrus*) in the western Palearctic region: persistence of northern cryptic populations during ice ages? *Molecular Ecology*, **14**, 765–779.

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

Sinclair W.T., Morman J.D., & Ennos R.A. (1998) Multiple origins for Scots pine (*Pinus sylvestris* L.) in Scotland: evidence from mitochondrial DNA variation. *Heredity*, **80**, 233–240.

Sinclair W.T., Morman J.D., & Ennos R.A. (1999) The postglacial history of Scots pine (*Pinus sylvestris* L.) in western Europe: evidence from mitochondrial DNA variation. *Molecular Ecology*, **8**, 83–88.

Teacher A.G.F., Garner T.W.J., & Nichols R.A. (2009) European phylogeography of the common frog (*Rana temporaria*): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, **102**, 490–496.

Tyler T. (2002) Geographical distribution of allozyme variation in relation to post-glacial history in *Carex digitata*, a widespread European woodland sedge. *Journal of Biogeography*, **29**, 919–930.

MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 5132–5134

NEWS AND VIEWS

PERSPECTIVE

Fission and fusion in island taxa – serendipity, or something to be expected?

BRENT C. EMERSON* and CHRISTIANA M. A. FARIA†

**Island Ecology and Evolution Research Group, Instituto de Productos Naturales y Agrobiología (IPNA-CSIC), C/Astrofísico Francisco Sánchez 3, San Cristóbal de La Laguna, Santa Cruz de Tenerife 38206, Spain, †Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK*

A well-used metaphor for oceanic islands is that they act as 'natural laboratories' for the study of evolution. But how can islands or archipelagos be considered analogues of laboratories for understanding the evolutionary process itself? It is not necessarily the case that just because two or more related species occur on an island or archipelago, somehow, this can help us understand more about their evolutionary history. But in some cases, it can. In this issue of *Molecular Ecology*, Garrick *et al.* (2014) use population-level sampling within closely related taxa of Galapagos giant tortoises to reveal a complex demographic history of the species *Chelonoidis becki* – a species endemic to Isabela Island, and geographically restricted to Wolf Volcano. Using microsatellite genotyping and mitochondrial DNA sequencing, they provide a strong case for *C. becki* being derived from *C. darwini* from the neighbouring island of Santiago. But the interest here is that colonization did not happen only once. Garrick *et al.* (2014) reveal *C. becki* to be the product of a double colonization event, and their data reveal these two founding lineages to be now fusing back into one. Their results are compelling and add to a limited literature describing the evolutionary consequences of double colonization events. Here, we look at the broader implications of the findings of Garrick *et al.* (2014) and suggest genomic admixture among multiple founding populations may be a characteristic feature within insular taxa.

Keywords: archipelago, colonization, genomic admixture, hybridization, island, speciation

Received 8 September 2014; revised 25 September 2014; accepted 29 September 2014

Correspondence: Brent C. Emerson, Fax: +34 922260135; E-mail: bemerson@ipna.csic.es

Understanding the significance of patterns of shared genetic variation among species or populations is frequently a thorn in the side of population geneticists. Although progress is being made through the development of analytical methods, distinguishing between gene flow and incomplete lineage sorting remains a challenge (e.g. Navascués *et al.* 2014). Island archipelagos potentially offer a geographical framework that can help to distinguish between these two explanations. The geographical isolation of islands means that the colonization of new islands, or the exchange of alleles among existing island populations, is expected to involve only one or a few individuals for taxa with naturally low dispersal capacity. As a consequence of this dynamic, the sharing of ancestral genetic variation among island taxa of low dispersal ability will be limited. Within such a framework, Nietlisbach *et al.* (2013) were able to exclude an explanation of incomplete lineage sorting among mockingbirds of the Galápagos archipelago to reveal a hybrid origin between *Mimus parvulus* and *M. melanotis* for the Genovesa Island endemic subspecies *M. parvulus bauri*. In this issue of *Molecular Ecology*, Garrick *et al.* (2014) reveal the Galápagos giant tortoise species *Chelonoidis becki* (Fig. 1) to be in the process of admixture among two founding gene pools. Underlying

Fig. 1 *Chelonoidis becki*, a species of Galapagos giant tortoise endemic to Wolf Volcano on Isabela Island. Photo credit: Yale University.

© 2014 John Wiley & Sons Ltd

the apparent morphological uniformity of *C. becki*, Garrick *et al.* (2014) were able to distinguish two gene pools, derived from two separate colonization events involving *C. darwini* from a neighbouring island, that are in the process of admixing. By calibrating their molecular data, they estimate an interval of approximately 150 000 years between the two colonization events, and using simulation analyses and forward in time projections, they estimate eventual complete introgression of the two *C. becki* gene pools.

The population genetic and adaptive consequences of genomic admixture

The results of Garrick *et al.* (2014) offer an interesting insight into island colonization dynamics and their consequences. In the context of island colonization, founding events that involve only one or a few individuals will result in low genetic diversity within the founding populations, which can only be recovered over an evolutionary timescale of mutation. Genetic admixture provides a potential escape from reduced genetic variation via the production of novel genotypic combinations of alleles, coupled with new opportunities for recombination among divergent genomes that may also facilitate adaptation within novel adaptive landscapes (Mallet 2007). In the case of *C. becki*, Garrick *et al.* (2014) note that the fusion of the two gene pools has little consequence for population-level heterozygosity, as heterozygosity is already high in both gene pools. However, introgression of the two genomes does result in a substantial increase in allelic richness, which suggests an increase in genotypic variation, and thus adaptive potential (Fig. 2). Mallet (2007) has noted that admixture of divergent gene pools may facilitate speciation and

	A_1B_1	A_1B_2	A_2B_1	A_2B_2	A_1B_3	A_1B_4	A_2B_3	A_2B_4	A_3B_1	A_3B_2	A_4B_1	A_4B_2	A_3B_3	A_3B_4	A_4B_3	A_4B_4
A_1B_1	A_1A_1 B_1B_1 (1)	A_1A_1 B_1B_2 (2)	A_1A_2 B_1B_1 (3)	A_1A_2 B_1B_2 (4)	10	14	12	16	31	32	54	55	37	39	58	59
A_1B_2	A_1A_1 B_1B_2 (2)	A_1A_1 B_2B_2 (5)	A_1A_2 B_1B_2 (4)	A_1A_2 B_2B_2 (6)	11	15	13	17	32	35	55	62	38	40	64	65
A_2B_1	A_1A_2 B_1B_1 (3)	A_1A_2 B_1B_2 (4)	A_2A_2 B_1B_1 (7)	A_2A_2 B_1B_2 (8)	12	16	18	20	33	34	56	57	41	43	60	61
A_2B_2	A_1A_2 B_1B_2 (4)	A_1A_2 B_2B_2 (6)	A_2A_2 B_1B_2 (8)	A_2A_2 B_2B_2 (9)	13	17	19	21	34	36	57	63	42	44	66	67
A_1B_3	10	11	12	13	22	23	24	25	37	38	58	64	68	69	82	83
A_1B_4	14	15	16	17	23	26	25	27	39	40	59	65	69	76	83	88
A_2B_3	12	13	18	19	24	25	28	29	41	42	60	66	70	71	84	85
A_2B_4	16	17	20	21	25	27	29	30	43	44	61	67	71	77	85	89
A_3B_1	31	32	33	34	37	39	41	43	45	46	47	48	72	78	74	80
A_3B_2	32	35	34	36	38	40	42	44	46	49	48	50	73	79	75	81
A_4B_1	54	55	56	57	58	59	60	61	47	48	51	52	74	80	86	90
A_4B_2	55	62	57	63	64	65	66	67	48	50	52	53	75	81	87	91
A_3B_3	37	38	41	42	68	69	70	71	72	73	47	75	A_3A_3 B_3B_3 (92)	A_3A_3 B_3B_4 (93)	A_3A_4 B_3B_3 (94)	A_3A_4 B_3B_4 (95)
A_3B_4	39	40	43	44	69	76	71	77	78	79	80	81	A_3A_3 B_3B_4 (93)	A_3A_3 B_4B_4 (96)	A_3A_4 B_3B_4 (95)	A_3A_4 B_4B_4 (97)
A_4B_3	58	64	60	66	82	83	84	85	74	75	86	87	A_3A_4 B_3B_3 (94)	A_3A_4 B_3B_4 (95)	A_4A_4 B_3B_3 (98)	A_4A_4 B_3B_4 (99)
A_4B_4	59	65	61	67	83	88	85	89	80	81	90	91	A_3A_4 B_3B_4 (95)	A_3A_4 B_4B_4 (97)	A_4A_4 B_3B_4 (99)	A_4A_4 B_4B_4 (100)

Fig. 2 A two-locus (locus A and locus B) Punnett square of all possible genotypes resulting from admixture and random segregation and independent assortment between two gene pools G1 and G2. G1 (green) and G2 (blue) are represented by two nonshared alleles at each locus: G1 contains alleles A_1, A_2 and alleles B_1, B_2; G2 contains alleles A_3, A_4 and alleles B_3, B_4. Unique genotypes are represented numerically. A twofold increase in allele richness following admixture has little consequence for average heterozygosity, but results in a greater than ten-fold increase in genotype richness, with nine possible two-locus genotypes within each parental population, and 100 within an admixed population.

© 2014 John Wiley & Sons Ltd

adaptive radiation, and the evolutionary success of the Hawaiian silversword alliance provides one probable example. The reservoir of adaptive potential contained within the allopolyploid ancestor of the silversword alliance, resulting from the hybridization of two ecologically divergent species, has been suggested to be a factor for their evolutionary success (Barrier *et al.* 1999). In the case of *C. becki*, it seems unlikely that genomic admixture will be the source of subsequent radiation (cladogenesis). However, an increase in genetic variation, and thus adaptive potential, could facilitate anagenetic change within *C. becki*, potentially contributing to the increased phenotypic divergence within the genus *Chelonoidis*, something that has been suggested for the mockingbirds of Galápagos (Nietlisbach *et al.* 2013).

Genomic admixture and island colonization dynamics

So, to what extent are the results of Garrick *et al.* (2014) an oddity – an interesting, but potentially low frequency event, chanced upon by serendipity? Or alternatively, could genomic admixture perhaps be a significant player in the field of island biogeography? In the case of *C. becki*, mitochondrial haplotypes shared with a third species, *C. vandenberghi*, suggest that the observations of Garrick *et al.* (2014) may not be a one-off event within the history of *Chelonoidis*. Considering the broader diversity on oceanic islands, there are good reasons to expect a more than minor role for admixture among multiple founding lineages. High colonization rates among islands will push populations towards genetic homogeneity, while low colonization rates will facilitate divergence among populations on different islands and high rates of inter-island cladogenetic speciation. At intermediate rates, there are two potential scenarios: (i) when barriers to gene flow are complete, founding taxa will remain genetically distinct, facilitating inter-island cladogenetic speciation; (ii) when barriers to gene flow may not have reached completion, the potential for admixture may arise, which may either inhibit or facilitate inter-island cladogenetic speciation. The data of Garrick *et al.* (2014) clearly point to the inhibition of cladogenetic speciation, with the two colonization events from Santiago to Isabela resulting in only one species. However, if admixture results in negative fitness consequences, divergent evolution may be facilitated. Mitochondrial and nuclear DNA sequence data have been used to infer a double colonization of the Canary Island of La Palma from El Hierro by the beetle species *Aphanarthrum glabrum*. The molecular data indicate a limited period of admixture between the two founding populations, presumably with negative fitness consequences, facilitating reproductive character displacement and the completion of reproductive isolation (Jordal *et al.* 2006).

The work of Garrick *et al.* (2014), together with previous studies such as those of Nietlisbach *et al.* (2013), Jordal *et al.* (2006) and Shaw (2002), provides strong evidence for genomic admixture among independent founding populations or species within an insular setting. Serendipity may have played a role in these discoveries, but rather than being a rare event, it may just be that the phenomenon is not easy to detect. In the case of *C. becki*, if fusion were complete, and allelic richness had returned to pre-admixture levels via the joint actions of drift and selection, there may be little information content from the molecular data to indicate an origin from multiple colonizations and admixture. Similarly, if the Santiago species *C. darwini* were to go extinct, the evidence for multiple colonizations underpinning the origin of *C. becki* would disappear with it. Given (i) the favourable conditions for genomic admixture that island archipelagos can provide and (ii) the potential for recombination between parental genomes to counter founder effects, the phenomenon observed by Garrick *et al.* (2014) could be a characteristic feature of the evolutionary process on islands. Detecting it may be complicated by the aforementioned issues, but it should never-the-less be given due consideration when interpreting molecular patterns of relatedness and genetic diversity in insular settings.

References

Barrier M, Baldwin BG, Robichaux RH, Purugganan MD (1999) Interspecific hybrid ancestry of a plant adaptive radiation: allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Molecular Biology and Evolution*, **16**, 1105–1113.

Garrick RC, Benavides E, Russello MA *et al.* (2014) Lineage fusion in Galápagos giant tortoises. *Molecular Ecology*, **23**, 5276–5290.

Jordal BH, Emerson BC, Hewitt GM (2006) Apparent 'sympatric' speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology*, **15**, 2935–2947.

Mallet J (2007) Hybrid speciation. *Nature*, **446**, 279–283.

Navascués M, Legrand D, Campagne C, Cariou M-L, Depaulis F (2014) Distinguishing migration from isolation using genes with intragenic recombination: detecting introgression in the *Drosophila simulans* species complex. *BMC Evolutionary Biology*, **14**, 89.

Nietlisbach P, Wandeler P, Parker PG *et al.* (2013) Hybrid ancestry of an island subspecies of Galápagos mockingbird explains discordant gene trees. *Molecular Phylogenetic and Evolution*, **69**, 581–592.

Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 16122–16127.

B.C.E. led and C.M.A.F. contributed to the writing of the manuscript.

---

doi: 10.1111/mec.12951

© 2014 John Wiley & Sons Ltd

Chapter 2

Speciation success on islands: evolution of the *Laparocerus tessellatus* complex (Curculionidae) in the Canary archipelago

Laparocerus tessellatus - species that gives the name to the complex.
Photo by Antonio Machado

2.1 ABSTRACT

Aim To infer colonisation and speciation history for a closely related complex of nine species within the enigmatic Canary Island *Laparocerus* weevil radiation of 128 species. Using molecular dating and the spatial and temporal context that islands provide, we evaluate the possible explanations of incomplete lineage sorting and gene flow for the origin of shared variation among species from different islands.

Location Canary Islands (Gran Canaria, Tenerife, La Palma and El Hierro).

Methods Phylogenetic analyses of mitochondrial (COII) and nuclear (ITS2) sequence data and molecular dating techniques were used to infer the origin of the group in the archipelago and their history of colonisation and differentiation.

Results Gran Canaria seems to be the geographic origin of the complex, with the onset of diversification within the complex estimated to have occurred approximately 3.9 Ma. An unexpected result is that mtDNA reveals each of the single species on La Palma and El Hierro to be the product of more than one colonisation event from more than one source island. In both cases nuclear ITS2 data reveals these multiple colonisations to have been followed by genomic admixture. Complex genetic patterns within the island of Tenerife suggest that morphologically defined species are unlikely to represent biological species, but also present evidence of cryptic speciation.

Main conclusions The two gene trees present very different topologies, with a rather simple colonisation history required to explain the pattern of nuclear gene relationships, while the mtDNA gene tree implicates a much more complex history of colonisation. Explanations of incomplete lineage sorting are ruled out and a history of colonisation and speciation for the *L. tessellatus* complex involving genetic admixture is revealed.

2.2 INTRODUCTION

Charles Darwin's stop in the Galapagos Islands during his five-year voyage on the HMS Beagle revealed to him that islands are an important source of evidence for evolution. Since Darwin's time, islands have become not just a source of evidence, but theatres for investigating and understanding mechanistic explanations for evolution itself (Emerson, 2002). Research on islands continues to advance and develop our general understanding of the evolutionary process, and biologists have come to regard islands as natural laboratories for the study of evolution (e.g. Roderick & Gillespie, 1998; Parent *et al*., 2008). The evolutionary proliferation of biodiversity has progressed much farther on some islands than on others (Losos and Ricklefs, 2009), and an equally important observation is that on a given island or archipelago it is often apparent that the proliferation of life has progressed much farther within some lineages than others. For example, among the 88 genera of weevil (Coleoptera, Curculionidae) occurring naturally within the Canary Islands, there is an average of only three species per genus (Oromí *et al*., 2010). However 128 species, more than 1/3 of all the native weevil species in the Canary Islands, belong to a single genus, *Laparocerus* Shonherr, 1834 (Machado in prep.). This raises an important question, why do some lineages diversify so extensively on islands, while others do not?

The addition of molecular phylogenetic techniques to the evolutionary biologist's toolkit has seen renewed focus on the study of evolution within oceanic islands over the last two decades, enabling researchers to address why a given lineage may have speciated more extensively in some archipelagos compared to others (e.g. Amorim *et al*., 2012). However, the extent to which molecular phylogenetic analyses have gone beyond describing pattern to inferring process has been limited, with most focusing on defining the relationships among species, and inferring the timing of speciation events (e.g. Juan *et al*., 1995; Emerson *et al*., 2000a; Emerson & Oromí, 2005; Amorim *et al*., 2012). Such analyses can help to elucidate the role of geography and the relative importance of within island and between island speciation. However, by typically sampling only a few individuals within a species, phylogenetic sampling does not capture diversification processes occurring below the species level. Recent or ongoing diversification may be captured with intraspecific genetic sampling (e.g. Jordal *et al*., 2006; Spurgin *et al*., 2011), and this can be complemented by phylogeographic

sampling, for a more explicit assessment of geography and its potential role as an agent of diversification.

While phylogenetic techniques chart the histories of island biotas, the integration of population-level studies with higher order phylogenetic studies may provide greater opportunity to connect pattern and process (Losos & Ricklefs, 2009). We agree with Losos and Ricklefs (2009), that detailed population level studies can chart the course of evolution over short time periods, directly measuring the extent to which natural selection changes in strength and direction over time, but advocate a broadening of approach for population-level studies. Population-level studies that incorporate geographically explicit sampling of individuals for the reconstruction of gene genealogies can reveal the extent to which natural selection or alternative mechanisms offer explanations of evolutionary change. Molecular sampling within species and species complexes that have diverged over Quaternary timescales optimises the ratio of extant informative haplotypes to extinct uninformative haplotypes with evolutionary time, increasing the information content of DNA sequence data to connect pattern with process (Benton & Emerson, 2007). Such studies have frequently revealed substantial genetic structuring within species within islands, and evolutionary processes underpinning diversification. The repeated convergent selection for, and evolutionary origins of, cave dwelling *Palmorchestia hypogaea* amphipod populations of the Canary Island of La Palma would have been all but missed with phylogenetic sampling (Villacorta *et al*., 2008), as would have been conclusions of ancestral and derived ecological associations and niche shifts in the *Nesotes* beetles of Gran Canaria (Rees *et al*., 2001). A revealing example of the combined power of geographic sampling with both mitochondrial and nuclear genomic sampling comes from Jordal et al.'s (2006) study of the sympatric and closely related *Aphanarthrum* weevil species *A. subglabrum* and *A. glabrum* on the island of La Palma in the Canary Islands. Representative geographic sampling of these two species and other taxa from the *A. glabrum* complex on other islands, combined with the analysis of mitochondrial and nuclear loci, reveals the combined roles of geography (allopatric isolation and subsequent secondary contact) and species interactions (hybridisation and reinforcement) in driving diversification.

With 128 described species from the Canary Islands and 34 described species from the Madeira archipelago, the genus *Laparocerus* stands out as an evolutionary enigma. It is

estimated that there may be as many as 200 species distributed in the Canary and Madeira archipelagos, with single species occurring in West Morocco, and the Salvages Islands (Machado, 2011). It is the most species rich of all animal and plant genera within the Canary Islands (Arechavaleta *et al*., 2010). Species richness within the *Laparocerus* is eight times higher than that of the next most species rich weevil genus, *Acalles*, represented by 16 species (Oromí *et al*., 2010), but there is little understanding of what evolutionary processes may underpin the evolutionary success of this group. All *Laparocerus* species are flightless and most are oligophages that climb vegetation to feed upon leaves, while a few species dwell in the leaf litter or are adapted to the underground environment (Machado, 2003).

The extensive diversity within the *Laparocerus* has for many years both intrigued and intimidated biologists, but recent efforts have successfully partitioned this diversity into taxonomic units (e.g. Machado, 2006; Machado, 2009; Machado, 2012), and species complexes within this diversity have been delineated with both mtDNA and nuclear sequence data (Machado *et al*., 2008; Machado, in prep.). As a first approach to understand why *Laparocerus* has diversified so dramatically within Macaronesia, we have sampled within the well-defined *Laparocerus tessellatus* species complex across its distribution on four of the Canary Islands (Gran Canaria, Tenerife, La Palma and El Hierro). Using sequence data from mitochondrial COII and nuclear ITS2 gene regions, we investigated inter-island colonisation within the complex, and the distribution of genetic variation among species. Using the spatial and temporal context that islands provide, we evaluate the possible explanations of incomplete lineage sorting and gene flow for the origin of shared variation among species from different islands, and discuss the implications of this shared variation for the speciation process in *Laparocerus*.

2.3 METHODS

2.3.1 Sampling and laboratory procedures

The *Laparocerus tessellatus* complex comprises nine species that have been identified as monophyletic based on both mitochondrial and nuclear sequence data (Machado, unpublished data). All species are single island endemics, with five on Gran Canaria (*Laparocerus microphthalmus* Lindberg, 1950, *Laparocerus obsitus* Wollaston, 1864,

Laparocerus osorio Machado, 2012, *Laparocerus tirajana* Machado, 2012, and *Laparocerus* sp. aff. *tirajana*), two on Tenerife (*L. tessellatus* Brullé, 1839 and *L. freyi* Uyttenboogaart, 1940) and one on each of La Palma (*Laparocerus* sp. 1) and El Hierro (*Laparocerus bimbache* Machado, 2011). Samples were collected from 39 sites from 1999 to 2011 (Fig 2.1). *Laparocerus vicinus* (Gran Canaria) was sampled as outgroup, based on its close phylogenetic relationship to the *L. tessellatus* complex (Machado, unpublished data). Upon collection, samples were stored in absolute ethanol at 4°C prior to species identification (AM) and DNA extraction.

Total genomic DNA was extracted from the head and prothorax using the DNeasy 96 well Blood and Tissue Extraction Kit (QIAGEN, West Sussex, UK) following the manufacturer's instructions. After extraction, both head and prothorax were placed back in absolute ethanol with the remainder of the body and maintained at 4° C as vouchers within the collection of AM. A fragment of approximately 785 bp of the mitochondrial gene cytochromec oxidase subunit II (COII) was amplified using primers TL2-J-3038 (5'-TAATATGGCAGATTAGTGCATTGGA) (Emerson *et al*., 2000b) and TK-N 3782 (5'-GAGACCATTACTTGCTTTCAGTCATCT) (EVA-Harrison Laboratory, Cornell University, Ithaca, NY, USA). Primers M13REV-CAS5p8sFt (5'-CAGGAAACAGCTATGACCTGAACATCGACATTTYGAACGCATAT) (Ji *et al*., 2003; as modified in Regier & Shi, 2005) and CAS28sB1d (5'-TTCTTTTCCTTCSCTTAYTRATATGCTTAA) (Ji et al. 2003) were used to amplify a fragment of approximately 540 bp of the nuclear gene internal transcribed spacer 2 (ITS2).

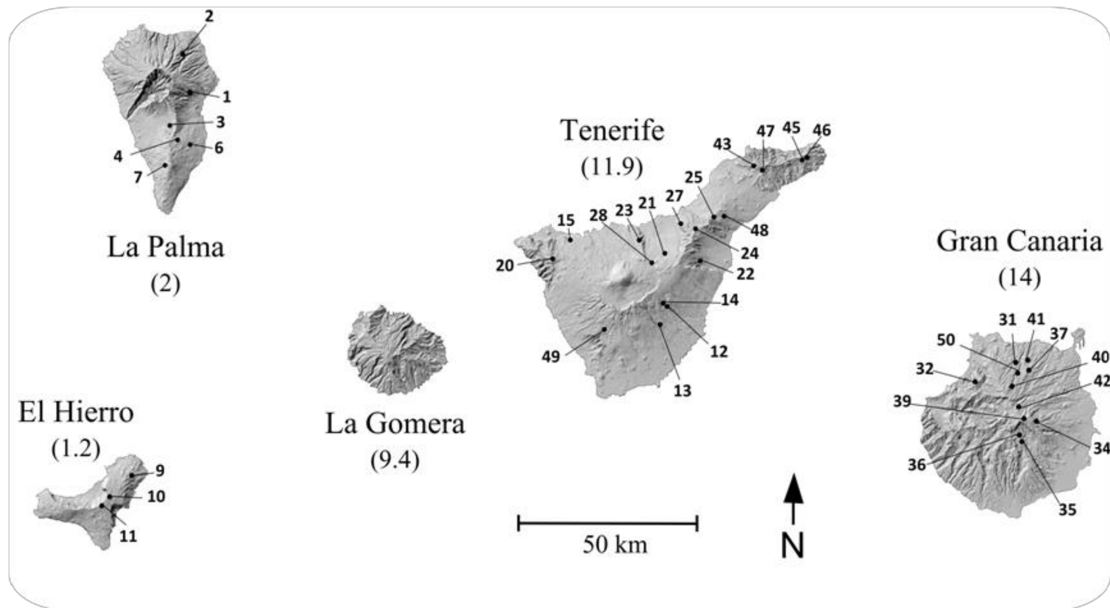


Figure 2.1. Distribution of sampling sites in the Canary Islands; a complete list of the locality names, species collected and number of individuals can be found in Table 2.1. Universal Transverse Mercator coordinates (UTM) for each site can be provided upon request. Numbers in parentheses refer to the proposed maximum estimated geological ages of the islands in millions of years.

Polymerase chain reactions (PCR) contained NH4 buffer (1x), 3.0 mMMgCl$_2$(for COII reactions and 4.0 mM MgCl$_2$ for ITS2 reactions), 0.2 mM of each dNTP, 0.4 μM of each primer and0.5 U of Taq polymerase (Bioline)in 25 μl final volume. PCR cycles were carried out using the following thermal profile for COII: 95°C for 3 min, 32 cycles at 95°C for 1 min, annealing temperature 48°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 3 min. For ITS2 the thermal profile used was: 94°C for 3 min, 34 cycles at 95°C for 40 sec, annealing temperature 50°C for 1 min, 72°C for 40 sec, and a final extension at 72°C for 2 min. Sequencing was performed in a PerkinElmer ABI3700 automated sequencer with BigDyeTerminator v3.1 Cycle Sequencing kit (applied Biosystems, California, USA). The thermal profile used for all sequencing reactions was: 96ºC for 10 sec, 50ºC for 5 sec and 60ºC for 4 min, 25 cycles. Sequences for COII were obtained with the forward primer only. The majority of COII PCR products yielded clean and unambiguous sequences with the forward primer (TL2-J-3038), but in some samples the first 48 nucleotides were not clear, requiring the use of

an internal reverse primer (LapCOIIRev1 5'-GGYATRAATCTATGATTTGTT), which binds at positions 588 to 608 at the end 3' of the aligned sequences. Sequencing was performed in both directions for the ITS2, due to heterozygosity and indel (insertion and deletion) variation.

2.3.2 Sequence alignment, haplotype reconstruction, and sequence properties

All sequences were processed and ambiguous base calls manually assessed with GENEIOUS PRO 5.4.5 (Drummond *et al*., 2010). COII sequences were aligned using MAFFT 6.814 (Katoh *et al*., 2002) and ITS2 consensus sequences were aligned using FAST 1.15.7 (Bradley *et al*., 2009), as it outperformed MAFFT when dealing with ITS2 indel variation. Both alignments were then checked by eye. Haplotypes for ITS2 sequences that were heterozygous only for single nucleotide polymorphisms (SNPs) were resolved either by direct comparison to homozygous sequences or with PHASE 2.1.1 (Stephens *et al*., 2001; Stephens & Scheet, 2005). The web tool SEQPHASE (Flot, 2010) was used to create PHASE input files and to interpret PHASE output files. Haplotype determination for out of phase sequence traces, caused by alleles with indel variation, was inferred from related homozygous sequences, or with the program Indelligent (Dmitriev & Rakitov, 2008).

The total number of variable and parsimony informative sites, average and maximum pairwise genetic distances (both uncorrected and corrected) overall, within and between species were computed with MEGA 6.0 (Tamura *et al*., 2013). The entropy-based index as implemented in DAMBE 5.2.78 (Xia *et al*., 2003) was used to assess substitution saturation within the mtDNA and ITS2 genes and the RDP4 software (Martin *et al*., 2015) was applied to detect potential historical recombination events within the ITS2 nuclear gene.

RDP4 implements a range of non-parametric recombination detection methods (including BOOTSCAN, MAXCHI, CHIMAERA, 3SEQ, GENECONV, SISCAN, PHYLPRO, RDP VISRD) to analyse a given set of aligned nucleotide sequences (Martin et al., 2015). As a result, RDP4 specifies a breakdown of recombination breakpoint locations and the identities of recombinant and parental sequences. RDP4 method (Martin and Rybicki, 2000), GENECONV method (Padidam et al., 1999), MAXCHI method (Maynard Smith,

1992; Posada and Crandall, 2001) and CHIMAERA method (Posada and Crandall, 2001) are the primary exploratory recombination signal detection methods implemented in RDP4 software. BOOTSCANning method (Salminien et al., 1995; Martin et al., 2005) and SISCAN method (Gibbs et al., 2000), are designed to check the validity of the recombination results produced by the primary detection methods, but they can also be used for primary exploratory analysis in order to detect signals of recombination. All these methods were used to scan the ITS2 alignment for recombination because they treat the data in different ways (e.g. RDP, GENECONV, MAXCHI, and CHIMAERA methods only examine variable nucleotide positions in triplets of sequences sampled from the alignment, BOOTSCAN and SISCAN methods examine all variable and conserved positions, and the optimal window size varies slightly from method to method) and they all have potential problems (see RDP4 manual for details, http://web.cbio.uct.ac.za/~darren/RDP4Manual.pdf). By using a combination of approaches, there was a greater probability to detect signals of recombination, so that the consequences could be offset. Window sizes were set small enough to ensure that events involving exchanges of small tracts of sequence (<200bp) were detectable in the most divergent sequences being examined. The MAXCHI and CHIMAERA methods were set to run with a variable window size that got respectively bigger and smaller with lower and higher degrees of parental sequence divergence.

2.3.3 Evolutionary tree and haplotype network construction

Bayesian inference (BI) analyses were performed for the mtDNA and nuclear gene separately, using the parallel version of MrBayes 3.2.1 (Ronquist, 2012). Eight analyses were run each for 10 million generations using 8 MCMC (Markov chain Monte Carlo) chains, discarding 25% of samples as burn-in. The general time reversible model of sequence evolution with a gamma correction (GTR + G), with priors set to the default values, was used for both genes as recommended by Stamatakis in the RAxML 7.0.4 Manual (http://www.phylo.org/archive/news/RAxML). Trees were rooted with *L. vicinus*. The output was assessed for stationarity and convergence in TRACER v.1.5 (Rambaut & Drummond, 2009) with only estimated sample size (ESS) above 200 for all parameters being accepted. Trees were visualised in FIGTREE 1.3.1 (Rambaut, 2011). TCS v.1.21 (Clement, 2000) was employed to infer haplotype networks of the less divergent nuclear gene sequences using statistical parsimony (Templeton *et al*., 1992)

with 95% confidence limit, and the software HAPSTAR (Teacher & Griffiths, 2011) was used to draw the network.

2.3.4 Testing hypothesis of monophyly

The monophyly of species within an island was assessed comparing the log likelihoods of alternative tree topologies. Topological constraints were constructed in MRBAYES 3.2.1 (Ronquist, 2012). A hard constraint was used to force a monophyletic species group to be present in all the sampled trees and it was compared to a negative constraint, which sampled across all trees that did not contain that monophyletic group. For each constraint, four analyses of $1x10^6$ generations each were run using 8 MCMC chains and the GTR + G model, discarding 25% of samples as burn-in and checking the stationarity and convergence of runs. Bayesian consensus tree files were then used for Maximum likelihood (ML) analyses in PAMLX (Yang, 2007; Xu & Yang, 2013) with the branch lengths used as starting values for maximum likelihood iterations. The matrix of site log likelihood generated for the ML trees was used for bootstrapping (n=10 000 replicates) and compared by the one-tailed Shimodaira–Hasegawa (SH) test (Shimodaira & Hasegawa, 1999) with a correction for multiple comparisons as implemented in PAMLX (Xu & Yang, 2013). The null hypothesis considered that all topologies were equally good.

2.3.5 Analysis of linkage disequilibrium and Hardy–Weinberg equilibrium

The joint analysis of mitochondrial and nuclear markers provides the opportunity to assess the biological significance of divergent genetic lineages when those lineages are sampled in sympatry (e.g. Cicconardi *et al*., 2013). Divergent lineages can be formally evaluated for consistency with the biological species concept (BSC) (Mayr, 1942) by testing for Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) when they occur sympatrically within sampling sites. In the case of divergent genetic lineages sampled sympatrically within species of the *L. tessellatus* complex, HWE and LD were calculated using ARLEQUIN 3.5.1.2 (Excoffier & Lischer, 2010), and statistical significance was tested using 100 000 dememorization steps and 100 000 steps in Markov chain, with Bonferroni adjustments (Benjamini & Hochberg, 1995) for multiple

comparisons. The null hypothesis considered that individuals within a sampling site belong to a panmictic population.

2.3.6 Dating analysis

When trying to identify the evolutionary processes that shape species formation, distinguishing between introgression and incomplete lineage sorting is generally difficult (Toews & Brelsford, 2012). However oceanic islands offer an ideal framework to unravel these processes, mainly due to the dynamics of speciation by founder events. In an insular setting, incomplete lineage sorting is of minor consequence within species derived from a limited number of founding individuals from a different island. Thus, gene flow among species that are derived from independent colonisation events can be evaluated in a background of limited or no incomplete lineage sorting (e.g. Jordal *et al*., 2006; Nietlisbach *et al*., 2013). To further evaluate the possibility of incomplete lineage sorting we use the spatial context provided by islands and the temporal information provided by DNA sequences to identify the colonisation directions and times for genetic variation shared between species on different islands. As our interest is not the absolute timings of lineage colonisation, but the relative timings, we applied a general coleopteran COII mutation rate of 0.015 substitution/site/myr (Cicconardi *et al*., 2010) with a restricted uniform distribution interval [1.49, 1.51]. Because the *L. tessellatus* complex is comprised of closely related species that are likely to have similar biological properties and molecular rates, a strict molecular clock model was applied. The MRBAYES output trees for both the mtDNA and nuDNA gene partitions were used as starting trees for the analyses with BEAST 1.7.3 (Drummond *et al*., 2012). The estimated root age of the COII tree was used as prior information to calibrate the root height of the ITS2 tree. Analyses used a GTR+G substitution model with 4 gamma categories, a Yule tree prior, and nodes with BP support of 0.90 or higher constrained to be monophyletic. To account for the possibility of either the extinction or non-sampling of molecular lineages (Emerson, 2002), nodes representing both the earliest and the most recent possible lineage colonisation times were estimated, and the time intervals among colonising lineages compared. Input files were generated in BEAUTI 1.7.3 (Drummond *et al*., 2012), and 15 runs of 100 million generations each, sampled every 1000 generations, were performed and combined, checking sampling, mixing and

convergence to a stationary distribution with TRACER 1.5 (Rambaut & Drummond, 2009).

2.4 RESULTS

A total of 173 specimens were collected from 11 sites in Gran Canaria (*L. microphthalmus, L. obsitus, L. osorio, L. tirajana, L.* sp. aff. *tirajana*, and the outgroup *L. vicinus*), 18 in Tenerife (*L. tessellatus* and *L. freyi*), six in La Palma (*Laparocerus* sp. 1) and three in El Hierro (*L. bimbache*) islands (Fig 2.1). Site locations and number of individuals per site are detailed in Table 2.1.

2.4.1 Mitochondrial COII gene

All but one of the 173 specimens were successfully amplified and sequenced for the mitochondrial COII gene, producing an alignment of 172 sequences of 633bp. Across the ingroup, 122 polymorphic sites (of which 96 were parsimony informative) and 76 unique alleles were identified. The average pairwise p-distance was 4% across all species with a maximum of 6.9%. Within species, average pairwise p-distances ranged from 0.7% (*L. osorio*) to 3.9% (*Laparocerus* sp. 1), and between species, average pairwise p-distances ranged from 1.7% (*L. osorio* and *L. microphthalmus*) to 5.2% (*L. bimbache* and *L. tirajana*) (Table 2.2).

Table 2.1 - Details of sampling within the Canary Islands for species within the *Laparocerus tessellatus* complex. Locations coded according to Fig 2.1.

Species	Locality code	Locality	N
Laparocerus sp. 1	LP1	La Palma: Montaña de Tagoja, 1250 m	3
	LP2	La Palma: Llanada de Barlovento, 650 m	5
	LP3	La Palma: s. El Paso, 870 m	7
	LP4	La Palma: Breña Alta, Pared Vieja, 1350 m	2
	LP6	La Palma: Mazo: Venijobre, 830 m	29
	LP7	La Palma: El Paso: Mña Don Mendo, 1075 m	3
L. bimbache Machado, 2011	EH9	El Hierro: Monte Ajares, 600 m	2
	EH10	El Hierro: San Andrés: Piedras Blancas, 912 m	2
	EH11	El Hierro: Cruz de Isora, Infra Masilva, 1247 m	2
L. freyi Uyttenboogaart, 1940	TF28	Tenerife: El Portillo, 2000 m	3
	TF12	Tenerife: Cumbres de Arico, 1070 m	3
	TF13	Tenerife: Granadilla: Las Vegas, 630 m	6
	TF14	Tenerife: Arico: Contador, 1200 m	2
	TF15	Tenerife: Tanque Bajo, 500 m	1
	TF20	Tenerife: Cumbre de Bolicos, 1200 m	2
	TF21	Tenerife: s. Mña. Bermeja, 1600 m	2
	TF22	Tenerife: Güímar: Bco. del Agua, 700-800 m	9
	TF23	Tenerife: s. Icod El Alto, 1200 m	4
	TF24	Tenerife: Santa Úrsula: Bco. Bensa, 1463 m	3
	TF25	Tenerife: Tacoronte: FuenteFría, 1014 m	1
	TF27	Tenerife: Santa Úrsula: La Corujera, 600 m	3
	TF48	Tenerife: Las Raíces, 930 m	11
	TF49	Tenerife: Ifonche, 990 m	2

Table 2.1. (cont)

Species	Locality code	Locality	N
L. tessellatus Brullé, 1839	TF43	Tenerife: Pista LasYedras, 740 m	1
	TF45	Tenerife: Anaga: El Pijaral Km 4.5, 700 m	12
	TF46	Tenerife: Anaga: Chinobre, 900 m	1
	TF47	Tenerife: Anaga: Cruz del Carmen, 900 m	2
L. microphthalmus Lindberg, 1950	GC32	Gran Canaria: Tamadaba NW, 1200 m	8
L. obsitus Wollaston, 1864	GC31	Gran Canaria: Valsendero: Bco. Oscuro, 900 m	2
L. tirajana Machado, 2012	GC35	Gran Canaria: San Bartolomé, Km 1, 940 m	12
	GC36	Gran Canaria: San Bartolomé: Bco. Tirajana, 900 m	1
L. sp. aff. *tirajana*	GC34	Gran Canaria: Bco. de los Cernícalos, 1400 m	4
	GC37	Gran Canaria: Degollada de Osorio, 875 m	5
	GC39	Gran Canaria: Cumbre: Roque Redondo, 1900 m	5
L. osorio Machado, 2012	GC31	Gran Canaria: Valsendero: Bco. Oscuro, 900 m	6
	GC40	Gran Canaria: Valsendero: Bco. Cazadores, 1080 m	3
	GC41	Gran Canaria: Las Huertecillas, 650 m	1
	GC42	Gran Canaria: Bco. de la Mina, 1200 m	2
L. vicinus Lindberg, 1953	GC31	Gran Canaria: Valsendero: Bco. Oscuro, 900 m	1

Xia's index for substitution saturation produced values of 0.015 (first and second codon positions) and 0.16 (third codon position) which were significantly lower than the critical value for symmetric topologies (0.69-0.79, $P<0.001$; 0.64-0.77, $P<0.001$, respectively), suggesting that sites have reached little saturation and sequences can be reliably used for phylogenetic reconstruction.

Table 2.2 - Mean pairwise genetic distances between mitochondrial species groups (n=76 alleles). Fre = *L. freyi*, Lap = *Laparocerus* sp.1, Bim = *L. bimbache*, Tess = *L. tessellatus*, Atir = *L.* sp. aff. *tirajana*, Tir = *L. tirajana*, Mic = *L. microphthalmus*, Oso = *L. osorio*, Obs= *L. obsitus*. Average pairwise genetic distances within species are represented in diagonal. Values in diagonal inside brackets and above diagonal represent uncorrected genetic distances (p-distances). Values in diagonal outside brackets and under diagonal represent genetic corrected distances using the Tamura Nei model for nucleotide substitution.

	Fre	Lap	Bim	Tess	ATir	Tir	Mic	Oso	Obs
	Uncorrected p-distances (%)								
Fre	**4.5 (2.7)**	4.35	4.17	2.77	4.60	4.80	3.83	4.46	4.34
Lap	8.1	**8 (3.9)**	4.3	4.6	3.6	3.6	4.6	4.8	4.6
Bim	7.6	8.9	**7 (3.3)**	4.3	5.2	5.2	3.8	4.4	4.1
Tess	4.6	8.7	7.9	**5.6 (3.3)**	4.7	4.8	4.0	4.8	4.5
ATir	8.7	6.9	10.6	9.0	**2 (1.6)**	1.7	4.0	4.3	3.7
Tir	9.1	6.9	10.6	9.2	2.3	**1.4 (1.1)**	4.0	4.6	3.9
Mic	6.5	9.1	7.9	7.1	6.9	7.1	**0.9 (0.8)**	1.7	1.7
Oso	8.6	9.7	9.0	9.9	7.9	8.7	2.1	**0.8 (0.7)**	2.5
Obs	7.9	8.8	7.7	8.6	6.2	6.9	2.2	3.6	**2.7 (2.1)**
	Tamura Nei + G corrected (%)								

2.4.2 Nuclear ITS2 gene

All but two of the 173 sampled beetles were successfully amplified and sequenced for the ITS2 gene. Of the 170 beetles of the ingroup, 64 were homozygotes and 106 were heterozygotes. All but five heterozygotes were successfully resolved for indel and SNP variation. The five unphased samples were removed from the dataset. The number of SNPs per heterozygote ranged from 1 to 6. Indels ranged in length from 1-19 bp and only one individual presented more than a single indel differing between its two alleles. The ITS2 alignment consisted of 330 sequences resulting in a final alignment of 411 bp after the removal of two variable poly-A regions, and a hyper-variable region of 35 bp that could not be aligned. Across the ingroup there were 48 polymorphic sites (of which 24 were parsimony informative) and 52 unique alleles were identified. The average pairwise p-distance was 1%, with a maximum of 3.4%. Within species, average pairwise p-distance ranged from 0.3% (*L. tirajana*) to 1.4% (*L.* sp. aff. *tirajana*), and between species, average pairwise p-distance ranged from 0.4 *(Laparocerus* sp. 1 and *L. bimbache*) to 2.5 % (*L. microphthalmus* and *L. tessellatus*) (Table 2.3).

Xia's index for substitution saturation, performed on all sites for the nuclear ITS2 dataset, produced values of 0.10-0.15 which were significantly lower than the critical values for symmetric topologies (0.69-0.79, $P<0.001$), suggesting that sites have reached little saturation and sequences can be reliably used for phylogenetic reconstruction. No signal of recombination was detected for the ITS2 gene for any of the methods used. Table 2.4 describes the time taken for each method implemented in RPD4 to scan the ITS2 alignment for recombination signals and the numbers of unique events detected (none).

Table 2.3 - Mean pairwise genetic distances between nuclear species groups (n=52 alleles). Lap = *Laparocerus* sp.1, Fre = *L. freyi*, Bim = *L. bimbache*, Tess = *L. tessellatus*, Atir = *L.* sp. aff. *tirajana*, Oso = *L. osorio*, Tir = *L. tirajana*, Mic = *L. microphthalmus*, Obs= *L. obsitus*. Average pairwise genetic distances within species are represented in diagonal. Values in diagonal inside brackets and above diagonal represent uncorrected genetic distances (p- distances). Values in diagonal outside brackets and under diagonal represent genetic corrected distances using the Tamura Nei model for nucleotide substitution. n/a – no variation within *L. bimbache* (unique haplotype)

	Lap	Fre	Bim	Tess	Atir	Oso	Tir	Mic	Obs
	Uncorrected p-distance (%)								
Lap	**0.67(0.71)**	0.9	0.4	1.0	1.5	1.3	1.8	2.5	1.9
Fre	1.0	**0.95 (0.9)**	0.7	0.9	1.5	1.3	1.8	2.5	1.9
Bim	0.4	0.8	n/a	0.8	1.3	1.1	1.5	2.3	1.7
Tess	1.1	0.9	0.9	**0.95 (0.89)**	1.6	1.4	1.9	2.5	1.9
Atir	1.7	1.6	1.4	1.7	**1.52 (1.42)**	1.1	1.4	2.0	1.0
Oso	1.4	1.4	1.2	1.5	1.1	**0.68 (0.65)**	1.0	1.7	0.9
Tir	1.9	1.9	1.6	2.0	1.5	1.0	**0.26 (0.25)**	2.3	1.2
Mic	2.8	2.7	2.5	2.8	2.2	1.8	2.5	**0.32 (0.31)**	1.6
Obs	2.1	2.1	1.9	2.1	1.1	1.0	1.3	1.7	**0.81 (0.79)**
	Tamura Nei + G corrected (%)								

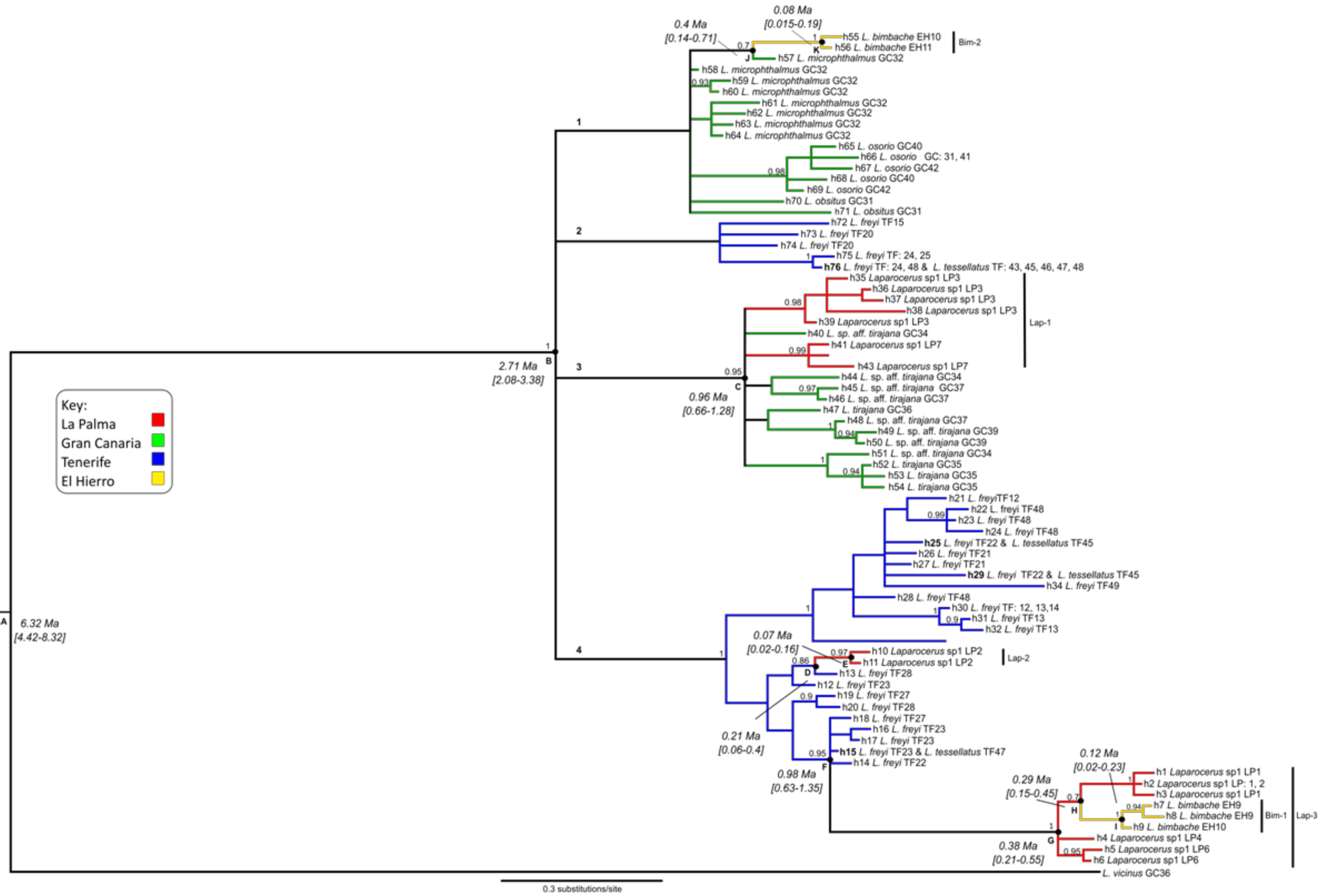
2.4.3 Phylogenetic analysis of mtDNA COII gene sequences

The bayesian tree of mtDNA sequence data reveals most of the morphologically described species within *L. tessellatus* complex to be paraphyletic or polyphyletic for this gene (Fig 2.2). Four major groups are described by the mtDNA sequence data: two poorly supported but geographically distinct clades (1 and 2) and two well supported clades (3 and 4). Within the island of Gran Canaria, *L. osorio* is the only species clearly recovered as monophyletic (Posterior Probability, PP=0.98). However, when considering only DNA sequence variation within Gran Canaria, the SH test did not reject the null hypothesis of equally good topologies for *L. microphthalmus* (p=0.457), *L.* sp. aff. *tirajana* (p=0.529) and *L. obsitus* (p=0.253) indicating they are not inconsistent with monophyly. For *L. tirajana*, SH test rejected the null hypothesis (p=0.027) and selected the negative constraint (no monophyly) as the best tree. The two species from Tenerife, *L. freyi* and *L. tessellatus* are polyphyletic and share four mtDNA haplotypes (h15, h25, h29 and h76). *Laparocerus* sp. 1 from La Palma is polyphyletic, originating from three founding mtDNA lineages (uncorrected p-distance among lineages ranges from 3.2 to 6.3%). The first of these lineages (Lap-1) comprises clade 3 with sequences from Gran Canaria (PP=1). The second (Lap-2) and third (Lap-3) La Palma lineages comprise clade 4 with sequences from Tenerife and El Hierro (PP=1).

Within clade 4, lineage Lap-2 forms a moderately supported clade with sequences from Tenerife (PP=0.86), while Lap-3 contains sequences from El Hierro and forms a highly supported clade with sequences from Tenerife (PP=1). Similar to La Palma, mtDNA sequences from the El Hierro species *L. bimbache* are polyphyletic with two lineages of independent origin (uncorrected p-distance between lineages 5.3%). The first of these (Bim-1) forms a strongly supported clade (PP=1) with sequences from La Palma while the second lineage (Bim-2) is more closely related to sequences from Gran Canaria.

Table 2.4 – Non-parametric methods (as implemented in RPD4 software) used to detect recombination signals for the ITS2 sequence data (see text for details). The time elapsed for scanning the alignment and the numbers of unique events detected are described.

Method	Time Elapsed	Unique Events (Recombination Signals)
RDP	0.30 s	0 (0)
GENECONV	0.00 s	0 (0)
BOOTSCAN	3.27 s	0 (0)
MAXCHI	0.20 s	0 (0)
SISCAN	0.34 s	0 (0)
CHIMAERA	0.00 s	0 (0)
Total	4.11 s	0 (0)



0.08 Ma
[0.015-0.19]
0.4 Ma
[0.14-0.71]
h55 L. bimbache EH10
h56 L. bimbache EH11
Bim-2
h57 L. microphthalmus GC32
h58 L. microphthalmus GC32
h59 L. microphthalmus GC32
h60 L. microphthalmus GC32
h61 L. microphthalmus GC32
h62 L. microphthalmus GC32
h63 L. microphthalmus GC32
h64 L. microphthalmus GC32
h65 L. osorio GC40
h66 L. osorio GC: 31, 41
h67 L. osorio GC42
h68 L. osorio GC40
h69 L. osorio GC42
h70 L. obsitus GC31
h71 L. obsitus GC31
h72 L. freyi TF15
h73 L. freyi TF20
h74 L. freyi TF20
h75 L. freyi TF: 24, 25
h76 L. freyi TF: 24, 48 & L. tessellatus TF: 43, 45, 46, 47, 48
h35 Laparocerus sp1 LP3
h36 Laparocerus sp1 LP3
h37 Laparocerus sp1 LP3
h38 Laparocerus sp1 LP3
h39 Laparocerus sp1 LP3
h40 L. sp. aff. tirajana GC34
h41 Laparocerus sp1 LP7
h43 Laparocerus sp1 LP7
Lap-1
h44 L. sp. aff. tirajana GC34
h45 L. sp. aff. tirajana GC37
h46 L. sp. aff. tirajana GC37
h47 L. tirajana GC36
h48 L. sp. aff. tirajana GC37
h49 L. sp. aff. tirajana GC39
h50 L. sp. aff. tirajana GC39
h51 L. sp. aff. tirajana GC34
h52 L. tirajana GC35
h53 L. tirajana GC35
h54 L. tirajana GC35
2.71 Ma
[2.08-3.38]
0.96 Ma
[0.66-1.28]
Key:
La Palma
Gran Canaria
Tenerife
El Hierro
h21 L. freyi TF12
h22 L. freyi TF48
h23 L. freyi TF48
h24 L. freyi TF48
h25 L. freyi TF22 & L. tessellatus TF45
h26 L. freyi TF21
h27 L. freyi TF21
h29 L. freyi TF22 & L. tessellatus TF45
h34 L. freyi TF49
h28 L. freyi TF48
h30 L. freyi TF: 12, 13,14
h31 L. freyi TF13
h32 L. freyi TF13
0.07 Ma
[0.02-0.16]
h10 Laparocerus sp1 LP2
h11 Laparocerus sp1 LP2
Lap-2
h13 L. freyi TF28
h12 L. freyi TF23
h19 L. freyi TF27
h20 L. freyi TF28
h18 L. freyi TF27
h16 L. freyi TF23
h17 L. freyi TF23
h15 L. freyi TF23 & L. tessellatus TF47
h14 L. freyi TF22
0.21 Ma
[0.06-0.4]
0.98 Ma
[0.63-1.35]
6.32 Ma
[4.42-8.32]
0.12 Ma
[0.02-0.23]
0.29 Ma
[0.15-0.45]
h1 Laparocerus sp1 LP1
h2 Laparocerus sp1 LP: 1, 2
h3 Laparocerus sp1 LP1
h7 L. bimbache EH9
h8 L. bimbache EH9
h9 L. bimbache EH10
Bim-1
Lap-3
h4 Laparocerus sp1 LP4
h5 Laparocerus sp1 LP6
h6 Laparocerus sp1 LP6
0.38 Ma
[0.21-0.55]
L. vicinus GC36
0.3 substitutions/site

Figure 2.2. Bayesian phylogenetic tree of the mtDNA COII haplotypes (633 bp) inferred using the GTR+G model of sequence evolution showing relationships and estimated divergence times of the Canarian Islands *L. tessellatus* complex. The tree is rooted with *L. vicinus*. Bayesian posterior probabilities are shown above nodes. Italic numbers in some selected nodes indicate the estimated age of the divergence event of that node and the 95% highest posterior density intervals obtained with a fixed substitution rate (see Materials and Methods). Letters and numbers immediately to the right of species names correspond to island codes and sampling localities, see Table 2.1. Shared haplotypes are highlighted in bold. A full version of the tree with all sampled individuals is provided in Fig. S2.1.

2.4.4 Phylogenetic and network analyses of nuclear ITS2 sequences

Within the bayesian ITS2 tree the earliest branching events are comprised solely of DNA sequences from Gran Canaria, with the remaining sequences from Gran Canaria forming a well supported clade (clade C, PP=1) with sequences from the remaining islands (Fig. 2.3). Only *L. microphthalmus* and *L. tirajana* from Gran Canaria are recovered as monophyletic. Although *L. osorio* is not monophyletic, the nuclear data is not inconsistent with its monophyly when only Gran Canarian species are considered (SH test, p=0.202). However, Shimodaira-test rejected the hypothesis of equally good topologies for *L. obsitus* and *L.* sp. aff. *tirajana* (p=0.043 and p=0.029 respectively), and in both cases the negative constraint (no monophyly) was selected as the best tree. The Tenerife species *L. freyi* and *L. tessellatus* are polyphyletic and share six haplotypes (h6, h20, h21, h24, h29, and h28). For a more resolved understanding of haplotype relationships within the clade comprising sequences from all islands (clade C), they were used for a network analysis, with closely related sequences from *L. osorio* (h48, h49 and h50) included as an outgroup to provide temporal information regarding derived and ancestral haplotypes (e.g. Zarza *et al*., 2008). A single network with no reticulations among haplotypes was recovered (Fig. 2.4) with haplotype h20 from Tenerife identified as the most recent common ancestor (MRCA) haplotype within the clade. Haplotype h20 would appear to be either unsampled, or extinct on Gran Canaria,

given that *L*. sp. aff. *tirajana* haplotypes (h42 and h43) are directly derived from it. Two haplotypes are shared between Tenerife and La Palma (h2 and h6), with one of these (h2) also being shared with El Hierro (Fig.2.4).

2.4.5 Species boundaries in Tenerife

Mitochondrial divergent lineages occurring sympatrically were found at two sampling sites in Tenerife (Fig. 2.2, locations TF45 and TF48) and the possibility of cryptic speciation was evaluated by assessing co-segregation of alleles (e.g. Cicconardi *et al*., 2013). The genotypes sampled at these sites are listed in Table 2.4. For all 11 individuals sampled at locality TF45, mitochondrial lineages correspond to different nuclear groupings, while at locality TF48, individuals from divergent mitochondrial lineages share the same nuclear grouping. At site TF45, both null hypotheses of linkage equilibrium and HWE were rejected ($p\text{-adj}<0.001$), while at site TF48 none of the null hypotheses could be rejected ($p>0.05$).

2.4.6 Dating analysis

The BEAST analysis of mtDNA COII sequences yielded an estimate of approximately 6.32 Ma [95% HPD: 4.42-8.32 Ma] for the divergence of the *L. tessellatus* complex from its sister lineage *L. vicinus*, with initial diversification within the complex estimated at approximately 2.71 Ma [HPD 2.08-3.38]. These two absolute age estimates are tentative, as they are reliant on a general coleopteran mtDNA rate, and thus should be viewed as suggestive. For the remaining nodes age estimates are used for the relative comparison of the earliest and most recent possible times for inter-island DNA sequence colonisation events (Fig. 2.2 and Table 2.5). La Palma is estimated to have been colonised by a Gran Canaria lineage (node C) approximately 0.96 Ma [0.66-1.28], and by two Tenerife lineages, one colonising between 0.21 Ma [0.06-0.4] (node D) and 0.07 Ma [0.02-0.16] (node E), and the other between 0.98 Ma [0.63-1.35] (node F) and 0.38 Ma [0.21-0.55] (node G). El Hierro is estimated to have been colonised by a La Palma lineage between 0.29 Ma [0.15-0.45] (node H) and 0.12 Ma [0.02-0.23] (node I) and by a lineage from Gran Canaria between 0.4 Ma [0.14-0.71] (node J) and 0.08 Ma [0.015-0.19] (node K). Nodes D, H and J are not well supported but their ages were

calculated in order to obtain an estimate of the earliest possible time for a colonisation event.

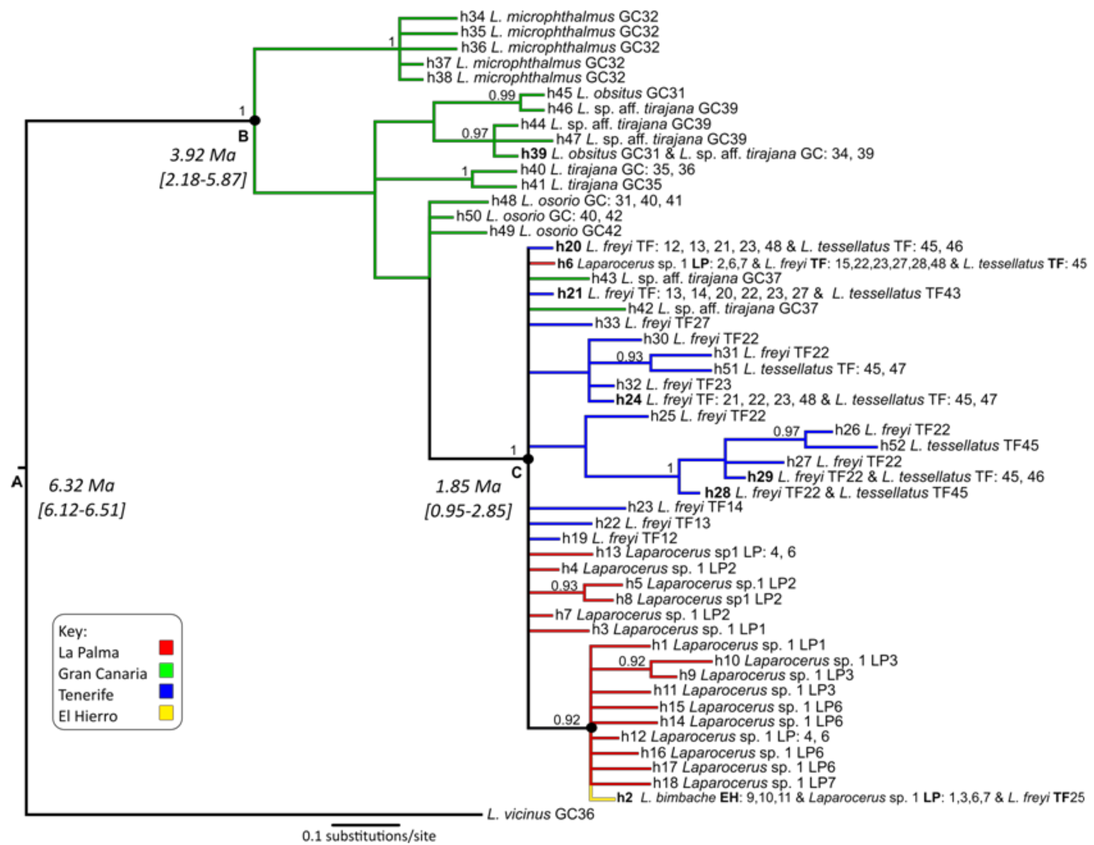


Figure 2.3. Bayesian phylogenetic tree inferred from the ITS2 nuclear gene (411 bp) using the GTR+G model of sequence evolution showing relationships and estimated divergence times of the Canarian Islands *L. tessellatus* complex. The tree is rooted with *L. vicinus*. Bayesian posterior probabilities are shown above nodes. Italic numbers in some selected nodes indicate the estimated age of the divergence event of that node and the 95% highest posterior density intervals obtained with a constrained root (see Materials and Methods). Letters and numbers immediately to the right of species names correspond to location codes, see Table 2.1. Shared haplotypes are highlighted in bold. A full version of the tree with all sampled individuals is provided in Fig.S2.2.

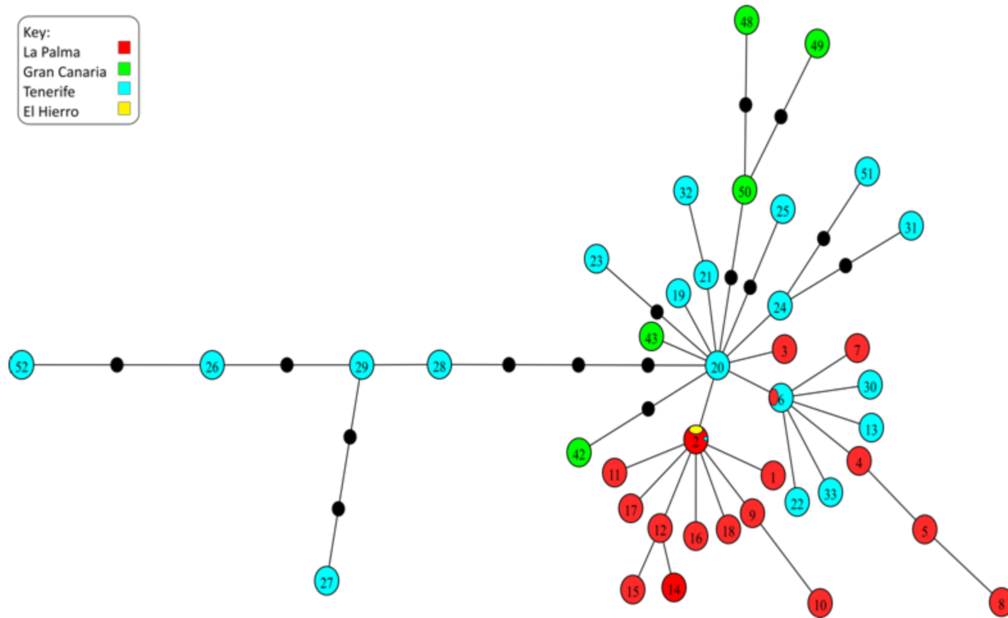


Figure 2.4. Haplotype network corresponding to the ITS2 sequence variation in the less divergent group of the *L. tessellatus* complex (clade C in Fig 2.3, see text for more details). Colours correspond to islands, and differently coloured segments within circles represent haplotype sharing across islands. The Gran Canarian outgroup *L. osorio* is represented by green haplotypes 48, 49 and 50. Lines represent a mutational step; black circles represent missing or unsampled haplotypes.

Using the root age of the mtDNA gene tree (Fig. 2.2, 6.32 Ma) as a prior to calibrate the ITS2 tree root height (Fig. 2.3), ages were estimated for the MRCA of the *L. tessellatus* complex and nodes representing sequence divergence events involving more than one island (Fig. 2.3 and Table 2.6). The time of MRCA of the *L. tessellatus* complex was estimated to be 3.92 Ma [HPD 2.17-5.87]. Tenerife, La Palma and El Hierro nuclear sequences are estimated to have diverged from Gran Canaria sequences (node C) at approximately 1.85 Ma [HPD 0.95-2.85].

Table 2.4 - Genotypes of individuals collected sympatrically in two localities in Tenerife (TF45 – El Pirajal and TF48 – Las Raíces), which represent divergent lineages (according to Figs. 2.2 and 2.3). MtDNA alleles are preceded by letter C and nuclear alleles are preceded by letters IT.

TF45 (El Pirajal)				TF48 (Las Raíces)			
Individual	COII lineage	Genotype		Individual	COII lineage	Genotype	
150	Clade 2	C76	IT20/29	165	Clade 2	C76	IT24/24
153		C76	IT51/51	167		C76	IT06/06
154		C76	IT51/51	169		C76	IT24/24
155		C76	IT20/52	172		C76	IT24/24
156		C76	IT20/52	174		C76	IT24/24
158		C76	IT51/51	175		C76	IT24/24
148	Clade 4	C25	IT06/24	166	Clade 4	C22	IT20/24
151		C29	IT20/28	171		C22	IT24/24
152		C29	IT29/29	173		C23	IT24/24
147		C33	IT06/06	170		C24	IT20/24
149		C33	IT06/06	168		C28	IT24/24

2.5 DISCUSSION

We investigated the history of diversification within the *L. tessellatus* species complex in the Canary Islands using a combination of sequence data from one mitochondrial and one nuclear gene. The two gene trees present very different topologies, with a rather simple colonisation history required to explain the pattern of nuclear gene relationships, while the mtDNA gene tree implicates a much more complex history of colonisation. Using the geographic context of the islands themselves, and relative temporal information from the gene trees, we were able to identify the geographic origin of the complex, and dismiss explanations of incomplete lineage sorting to reveal a history of colonisation and speciation involving genetic admixture.

Table 2.5 - Estimated relative times in million years (Ma) of *L. tessellatus* complex mitochondrial lineages expressed as mean values with 95% highest posterior density (HPD) intervals. ET = earliest possible time; MRT = most recent possible time. Island codes: EH- El Hierro, GC-Gran Canaria, LP– La Palma, TF-Tenerife. La Palma lineages= Lap-2 and Lap-3; El Hierro lineages = Bim-1 and Bim-2.

Node	Description	Mean value (Ma)	95% HPD
A	root of the tree	6.32	4.45-8.36
B	Ingroup	2.71	2.08-3.38
C	Divergence of LP sequences from GC sequences	0.96	0.66-1.28
D	ET for colonisation of LP from TF sequences (Lap-2)	0.21	0.06-0.4
E	MRT for colonisation of LP from TF sequences (Lap-2)	0.07	0.02-0.16
F	ET for colonisation of LP from TF sequences (Lap-3)	0.98	0.63-1.35
G	MRT for colonisation of LP from TF sequences (Lap-3)	0.38	0.21-0.55
H	ET for colonisation of EH from LP sequences (Bim-1)	0.29	0.15-0.45
I	MRT for colonisation of EH from LP sequences (Bim-1)	0.12	0.02-0.23
J	ET for colonisation of EH from GC sequences (Bim-2)	0.4	0.14-0.71
K	MRT for colonisation of EH from GC sequences (Bim-2)	0.08	0.015-0.19

Table 2.6 - Estimated relative times in million years (Ma) of *L. tessellatus* complex ITS2 lineages expressed as mean values with 95% highest posterior (HPD) density intervals. Island codes = EH-El Hierro, GC-Gran Canaria, LP-La Palma, TF-Tenerife.

Node	Description	Mean value (Ma)	95% HPD
A	root of the tree	6.32	6.12-6.51
B	Ingroup	3.92	2.18-5.87
C	Divergence of TF, LP and EH sequences from GC sequences	1.85	0.95-2.85

2.5.1 Geographic origin of the *L. tessellatus* complex

The nuclear sequence data indicates the geographic origin of the complex to be Gran Canaria and, assuming a rate of molecular evolution not dissimilar from a general coleopteran rate, initial diversification within the complex would have started around 3.92 Ma. The earliest branching events within the ITS2 tree are uniquely composed of individuals from Gran Canaria, with sequences from all other islands restricted to a single well defined clade that also includes DNA sequences from Gran Canaria (clade C, Fig. 2.3). The complex relationships within the clade composed of sequences from all islands are best understood when viewed as a network (Fig. 2.4). All nuclear ITS2 DNA sequence variation sampled on the islands of Tenerife, La Palma and El Hierro is derived from a single ancestral sequence that is identified as haplotype h20 from Tenerife. Given the inferred Gran Canaria origin for the complex, h20 must be either unsampled or extinct on Gran Canaria, but the derived haplotypes h42 and h43, sampled from *L.* sp. aff. *tirajana*, implicate this taxon in the colonisation of Tenerife, La Palma and El Hierro. Relationships among mtDNA sequences provide little evidence that can be used to infer the geographic origin of the group. However, the *L. tessellatus* complex is nested within a well supported monophyletic clade comprised of an additional nine mitochondrial lineages, of which all are almost exclusively comprised of species from Gran Canaria (Machado, unpublished data), which supports a Gran Canarian origin for the complex.

2.5.2 Gran Canaria species and the founding of neighbouring islands

When considering Gran Canarian species alone (i.e. ignoring all other islands), monophyly is either supported, or not rejected, for both molecular markers, for two species, *L. osorio* and *L. microphthalmus*. *Laparocerus osorio* is monophyletic for mtDNA, while *L. microphthalmus* is monophyletic for the nuclear marker, and for both species a hypothesis of monophyly cannot be rejected (Shimodaira Hasegawa test) for the second molecular marker. The other species are monophyletic or not inconsistent with monophyly for only one marker. *Laparocerus tirajana* is monophyletic for the ITS2 gene but monophyly is rejected for mitochondrial data (Shimodaira Hasegawa test). *Laparocerus obsitus* and *L.* sp. aff. *tirajana* are not inconsistent with a hypothesis of monophyly for the mitochondrial gene but they are for the nuclear marker. When taking into account species from all islands, hypotheses of monophyly are rejected for *L. microphthalmus, L. obsitus, L. tirajana* and *L.* sp. aff. *tirajana*. The molecular data identify colonisation events to neighbouring islands involving two of the five Gran Canarian species. Mitochondrial DNA sequences sampled from *L. bimbache* (El Hierro) are recovered within *L. microphthalmus* (Gran Canaria), while mtDNA sequences from *Laparocerus* sp. 1 (La Palma) are recovered within *L. tirajana* and *L.* sp. aff. *tirajana* (Gran Canaria). All nuclear variation from all species occurring outside Gran Canaria is recovered within *L.* sp. aff. *tirajana* (Gran Canaria).

2.5.3 Tenerife species and the founding of neighbouring islands

Most DNA sequences sampled within *L. tessellatus* were also found in *L. freyi*. Of the five mitochondrial and eight nuclear sequences sampled from *L. tessellatus*, four and six are shared with *L. freyi,* suggesting that *L. tessellatus* may be of recent origin from an *L. freyi* like ancestor. Both markers present an absence of monophyly for genetic variation within Tenerife and identify colonisation events from Tenerife to both La Palma and El Hierro. Mitochondrial DNA sequences sampled from both *Laparocerus* sp. 1 (La Palma) and *L. bimbache* (El Hierro) are recovered within a clade of sequences from Tenerife (Fig. 2.2) and nuclear haplotypes from *Laparocerus* sp. 1 (La Palma) and *L. bimbache* (El Hierro) are shared with Tenerife (Fig. 2.4).

2.5.4 La Palma and El Hierro–single species with multiple origins

The nuclear sequence data indicates both of the younger islands of La Palma and El Hierro to be derived from a single founding event and island. ITS2 sequences from La Palma are consistent with Tenerife as a source, while those of El Hierro are derived from either La Palma or Tenerife (Figs. 2.3 & 2.4). In contrast, mitochondrial relationships are in sharp disagreement with this simple colonisation scenario. *Laparocerus* sp. 1 (La Palma) is composed of multiple mitochondrial lineages, derived from either Tenerife (*L. freyi* and *L. tessellatus*) or Gran Canaria (*L. tirajana* and *L.* sp. aff. *tirajana*) (Fig. 2.2). *Laparocerus bimbache* (El Hierro) is composed of two mitochondrial lineages, one derived from *Laparocerus* sp. 1 (La Palma), while the other is derived from *L. microphthalmus* (Gran Canaria) (Fig. 2.2). This pattern of mixed ancestry suggests either: (i) species origin on La Palma and El Hierro involving genetic admixture from multiple founding species, or (ii) incomplete lineage sorting (Funk & Omland, 2003; Toews & Brelsford, 2012). Distinguishing between these two processes is typically a challenge, but the dynamics of speciation by founder events between islands enables us to exclude lineage sorting by evaluating three expectations from incomplete lineage sorting. We deal with each of these three expectations below in turn.

Expectation 1: All sequence variation shared among islands was present in the ancestral gene pool. This would require colonisation events between islands involving large numbers of founding individuals, such that a substantial amount of the standing ancestral genetic variation would be transferred. This is not consistent with the colonisation dynamics of flightless beetles (Ikeda *et al*., 2012; Vogler & Timmermans, 2012). This would then need to be followed by non-random extinction (or sampling) of genetic variation within both the source and founded islands. As an example, the following scenario would be necessary to explain patterns of genetic diversity on La Palma if *Laparocerus* sp. 1 were the result of a single colonisation event from Gran Canaria. All sequence variation within lineage 4 (Fig. 2.2) would have been present within the ancestral population of Gran Canaria. Additionally, all sequence variation within lineage 3 (Fig. 2.2) would have been present within Gran Canaria. Reciprocal and extensive non-random mtDNA lineage extinction on both islands would then be required to explain the absence of Tenerife sequences within

lineage 3, and Gran Canaria sequences in lineage 4. As we sampled across the range of each species, it is equally unlikely that such a large number of sequences are present, but were not sampled.

Expectation 2: If the species from La Palma and El Hierro are each the product of a single colonisation event, colonisation times of all molecular lineages shared between these and other islands should coincide. The relative time interval for the colonisation of a lineage to an island ranges from the lowest to the highest bound of the two time intervals (earliest and most recent possible) measured for each lineage, e.g. La Palma was colonised from one Tenerife lineage between 0.02-0.4 Ma (minimum and maximum HPD values for nodes D and E, Fig. 2.2). There is substantial overlap in the time intervals estimated for different lineages colonising La Palma and El Hierro, and this is a consequence of the wide 95% posterior intervals estimated by the Bayesian MCMC method when dealing with shallow genetic divergence (Brown & Yang, 2010). Despite this difficulty, results clearly reject the hypothesis of a single colonisation for La Palma. The mitochondrial lineage Lap-1 on La Palma involved a colonisation event (node C, between 0.66 and 1.28 Ma) that does not coincide with lineage Lap-2 (node D and E, between 0.02 and 0.4 Ma) (Fig. 2.2). Thus temporal information alone indicates that La Palma was colonised at least two times.

Expectation 3: A signature of incomplete lineage sorting for nuclear gene is expected to be more exaggerated than for the mitochondrial gene, due to its much bigger effective population size (Ballard & Whitlock, 2004). In contrast to this expectation, the signature of potential incomplete lineage sorting among islands presented by the nuclear gene (Fig. 2.4) is much less than expected if the patterns of mtDNA relatedness are due to incomplete lineage sorting. If we consider the shared mtDNA variation observed between Gran Canaria and both La Palma and El Hierro to be the result of incomplete lineage sorting, its absence within the more slowly evolving nuclear gene, with its much larger effective population size becomes difficult to explain.

It should be noted that many other studies have found mtDNA gene trees to be less resolved than nuclear gene trees and while many of them have indicated retention of ancestral polymorphism as the possible explanation, several others were able to

explain this difference through the acts of genomic admixture (e.g. Shaw et al., 2000; Peters et al., 2007; Roos et al., 2011; Choleva et al., 2014; Dong et al., 2014). In fact, a search in the literature for mito-nuclear discordance reveals the majority (if not all) of the studies favour hybridization. In the few cases where incomplete lineage sorting was clearly demonstrated (Rheindt et al., 2009; Wilson et al., 2011; Tang et al., 2012; Pagès et al., 2013), our prediction of a more exaggerated signature of incomplete lineage sorting for the nuclear gene was met. For example, Rheindt et al. (2009) investigated species-level paraphyly and polyphyly in the montane Neotropical Elaenia flycatchers using a combination of mitochondrial and nuclear genes and population genetic methods. They demonstrated the pattern of polyphyly to be based on an interplay of three different factors: "(i) faulty taxonomy […]; (ii) a late Pleistocene hybridization event that resulted in two morphologically and ecologically distinct species sharing extremely similar mitochondrial DNA but distinct nuclear DNA profiles; and (iii) incomplete lineage sorting in a nuclear marker that results in a polyphyletic placement of species that are otherwise well-differentiated in mitochondrial DNA, morphology and ecology" (Rheindt et al., 2009, pg 143).

2.5.5 Species boundaries in Tenerife

An interesting genetic pattern within Tenerife is the finding of divergent mtDNA lineages occurring in sympatry in two locations. While in one location (site TF48), the two mtDNA lineages show no evidence that they might represent biological species, in the other location (site TF45), there is a strong nuclear signature of limited gene flow among individuals from these two mtDNA lineages. This complex pattern of sympatric coexistence may be explained by differences in environmental backgrounds that might facilitate contact and gene exchange between lineages in one site but enhance isolation and prevent gene flow in the other. It is not an unusual pattern and it has been reported for other species such as the Towhee birds *Pipilo maculatus* and *P. ocai* in southern Mexico, which hybridize extensively in several locations, but coexist in sympatry with no evidence of gene flow in other locations (Sibley & Sibley, 1964). Thus, while molecular data does not support the division of the *L. tessellatus* complex in Tenerife into *L. tessellatus* and *L. freyi,* it does suggest that there to be more than one species, potentially related to the two mtDNA lineages within the islands. However, more sampling is required to specifically address this issue.

2.5.6 Conclusions

Species that are the product of genetic admixture among multiple colonising lineages within island archipelagos, as demonstrated here, have rarely been reported for animal taxa. However, this is perhaps not to be unexpected, because island archipelagos present a geographic matrix where both geographic isolation and secondary contact are a function of colonisation dynamics and frequency in space and time. High colonisation rates among islands will push populations toward panmixia, while low colonisation rates will facilitate divergence. At intermediate rates, where genomic divergence is promoted, but where barriers to gene flow may not have reached completion, the potential for admixture may arise. Indeed, genomic admixture has been invoked to explain discordant mito-nuclear tree topologies for a subspecies of the Galápagos mockingbird genus *Mimus* (Nietlisbach *et al*., 2013). Our findings, like those of Nietlisbach *et al*. (2013) raise an interesting question – to what extent might admixture be a driver of diversification itself? The potential importance of genetic admixture as a driver of speciation is well recognised (e.g. Shaw, 2002; Mallet, 2007; Schwenk *et al*., 2008). In the context of island colonisation, founding events that involve only one or a few individuals will result in low genetic diversity within the founding populations, which can only be recovered over an evolutionary time-scale of mutation. Genetic admixture provides a potential escape from reduced genetic variance via recombination among divergent genomes that may also facilitate adaptation within novel adaptive landscapes (Mallett, 2007). Further work is needed to address this issue, and new techniques such as reduced genome sequencing and genotyping by sequencing (e.g. Mastretta-Yanes *et al*., 2014) should prove very useful in this respect.

2.6 REFERENCES

Amorim, I.R., Emerson, B.C., Borges, P.A.V. & Wayne, R.K. (2012) Phylogeography and molecular phylogeny of Macaronesian island *Tarphius* (Coleoptera: Zopheridae): why are there so few species in the Azores? *Journal of Biogeography*, **39**, 1583–1595.

Arechavaleta, M., Rodrígues, S., Zurita, N. & García, A. (2010) *Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres. 2009*. Gobierno de Canarias.

Ballard, J.W.O. & Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729-744.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodological*, **57**, 289–300.

Benton, M.J. & Emerson, B.C. (2007) How did life become so diverse? The dynamics of diversification according to the fossil record and molecular phylogenetics. *Palaeontology*, **50**, 1-19.

Bradley, R.K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I. & Pachter, L. (2009) Fast Statistical Alignment. *Plos Computational Biology*, **5**

Brown, R.P. & Yang, Z. (2010) Bayesian dating of shallow phylogenies with a relaxed clock. *Systematic Biology*, **2**, 191-131.

Choleva L., Musilova Z., Kohoutova-Sediva A., Paces J., Rab P., & Janko K. (2014) Distinguishing between Incomplete Lineage Sorting and Genomic Introgressions: Complete Fixation of Allospecific Mitochondrial DNA in a Sexually Reproducing Fish (*Cobitis*; Teleostei), despite Clonal Reproduction of Hybrids. *PLoS ONE*, **9**: e806.

Cicconardi, F., Fanciulli, P.P. & Emerson, B.C. (2013) Collembola, the biological species concept and the underestimation of global species richness. *Molecular Ecology*, **22**, 5382-5396.

Cicconardi, F., Nardi, F., Emerson, B.C., Frati, F. & Fanciulli, P.P. (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology*, **19**, 386-400.

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

Dmitriev, D.A. & Rakitov, R.A. (2008) Decoding of Superimposed Traces Produced by Direct Sequencing of Heterozygous Indels. *Plos Computational Biology*, **4.**

Dong F., Zou F.-S., Lei F.-M., Liang W., Li S.-H., & Yang X.-J. (2014) Testing hypotheses of mitochondrial gene-tree paraphyly: unravelling mitochondrial capture of the Streak-breasted Scimitar Babbler (Pomatorhinus ruficollis) by the Taiwan Scimitar Babbler (*Pomatorhinus musicus*). *Molecular Ecology*, **23**, 5855–5867.

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*, **29**, 1969-73.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. & Wilson, A. (2010) Geneious v5.3. In: *Available from* http://www.geneious.com

Emerson, B.C. (2002) Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology*, **11**, 951-966.

Emerson, B.C. & Oromí, P. (2005) Diversification of the forest beetle genus *Tarphius* on the Canary Islands, and the evolutionary origins of island endemics. *Evolution*, **59**, 586-598.

Emerson, B.C., Oromí, P. & Hewitt, G.M. (2000a) Interpreting colonisation of the *Calathus* (Coleoptera: Carabidae) on the Canary Islands and Madeira through the application of the parametric bootstrap. *Evolution*, **267**, 2199-2205.

Emerson, B.C., Oromi, P. & Hewitt, G.M. (2000b) Colonization and diversification of the species *Brachyderes rugatus* (Coleoptera) on the Canary Islands: Evidence from mitochondrial DNA COII gene sequences. *Evolution*, **54**, 911-923.

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

Flot, J.F. (2010) SEQPHASE: a web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources*, **10**, 162-166.

Funk, D.J. & Omland, K.E. (2003) Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics*, **34**, 397-423.

Gibbs M.J., Armstrong J.S., & Gibbs A.J. (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics (Oxford, England)*, **16**, 573–582.

Ikeda, H., Nishikawa, M. & Sota, T. (2012) Loss of flight promotes beetle diversification. *Nature Communications*, **3**, 648.

Ji, Y.J., Zhang, D.X. & He, L.J. (2003) Evolutionary conservation and versatility of a new set of primers for amplifying the ribosomal internal transcribed spacer regions in insects and other invertebrates. *Molecular Ecology Notes*, **3**, 581-585.

Jordal, B.H., Emerson, B.C. & Hewitt, G.M. (2006) Apparent 'sympatric' speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology*, **15**, 2935-2947.

Juan, C., Oromi, P. & Hewitt, G.M. (1995) Mitochondrial-DNA phylogeny and sequential colonization of Canary Islands by darkling beetles of the genus *Pimelia* (Tenebrionidae). *Proceedings of the Royal Society of London Series B-Biological Sciences*, **261**, 173-180.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059-3066.

Losos, J.B. & Ricklefs, R.E. (2009) Adaptation and diversification on islands. *Nature*, **457**, 830-836.

Machado, A. (2003) Sobre el método de colectar *Laparocerus* Schönherr, 1834 y el reconocimiento de sus marcas en las hojas (Coleoptera, Curculionidae). *Vierae*, **31**, 407-420.

Machado, A. (2006) The type material of the species of *Laparocerus* Schonherr, 1834 (Coleoptera, Curculionidae, Entiminae). *Journal of Natural History*, **40**, 2001-2055.

Machado, A. (2009) New *Laparocerus* Schoenherr, 1834 from La Palma, Canary Islands (Coleoptera, Curculionidae, Entiminae). *Graellsia*, **65**, 183-224.

Machado, A. (2011) Los *Laparocerus* Schoenherr, 1834 de las Canarias orientales, islas Salvajes y Marruecos (Coleoptera, Curculionidae, Entiminae). *Graellsia*, **67**, 205-241.

Machado, A. (2012) Nuevos taxones de *Laparocerus* Schoenherr, 1834 de la isla de Gran Canaria (Coleoptera, Curculionidae, Entiminae). *Boletín de la Sociedad Entomológica Aragonesa*, **51**, 5-46.

Machado, A. (in prep.) Monograph of the *Laparocerus*.

Machado, A., López, M., Almeida, T. & Hernández, M. (2008) Mitochondrial DNA phylogenetic analysis of the genus *Laperocerus* (Coleoptera, Curculionidae, Entiminae). I. The Madeiran clade. *Zoology Scripta*, **37**, 415-427.

Mallet, J. (2007) Hybrid speciation. *Nature* **446**, 279–283.

Martin D.P., Murrell B., Golden M., Khoosal A., & Muhire B. (2015) RPD4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, **1**, vev003.

Martin D.P., Posada D., Crandall K.A., & Williamson C. (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS research and human retroviruses*, **21**, 98–102.

Martin D. & Rybicki E. (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics, 16, 562–563. Maynard Smith J. (1992) Analyzing the mosaic structure of genes. Journal of Molecular Evolution, 34, 126–129.

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T.H., Piñero, D. & Emerson, B.C. (2014) RAD sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Molecular Ecology Resources, **doi: 10.1111/1755-0998.1229**

Mayr, E. (1942) *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. Columbia University Press, New York.

Nietlisbach, P., Wandeler, P., Parker, P.G., Grant, P.R., Grant, B.R., Keller, L.F. & Hoeck, P.E.A. (2013) Hybrid ancestry of an island subspecies of Galápagos mockingbird explains discordant gene trees. *Molecular Phylogenetics and Evolution*, **69**, 581–592.

Oromí, P., Cruz, S. & Báez, M. (2010) Psocoptera, Mallophaga, Anoplura, Hemiptera, Thysanoptera, Coleoptera, Strepsiptera, Siphonaptera. *Lista de especies silvestres de Canarias*. *Hongos, plantas y animales terrestres*. *2009* (ed. by M. Arechavaleta, S. Rodrígues, N. Zurita and A. García), pp. 230-302. Gobierno de Canarias, Santa Cruz de Tenerife.

Padidam M., Sawyer S., & Fauquet C.M. (1999) Possible emergence of new geminiviruses by frequent recombination. Virology, **265**, 218–225.

Pagès M., Bazin E., Galan M., Chaval Y., Claude J., Herbreteau V., Michaux J., Piry S., Morand S., & Cosson J.-F. (2013) Cytonuclear discordance among Southeast Asian black rats (Rattus rattus complex). Molecular Ecology, **22**, 1019–1034.

Parent, C.E., Caccone, A. & Petren, K. (2008) Colonization and diversification of Galápagos terrestrial fauna: a phylogenetic and biogeographical synthesis. Philosophical Transactions of the Royal Society B: Biological Sciences, **363**, 3347–3361.

Peters Y.Z. (2007) Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (Anas spp.). Evolution; international journal of organic evolution, 61, 1992–2006.

Posada D. & Crandall K.A. (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proceedings of the National Academy of Sciences of the United States of America, **98**, 13757–13762.

Rambaut, A. (2011) FigTree v1.3.1, Available from http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut, A. & Drummond, S.J. (2009) *Tracer v1.5*. *Available from* http://beast.bio.ed.ac.uk/Tracer.

Rees, D.J., Emerson, B.C., Oromí, P. & Hewitt, G.M. (2001) Mitochondrial DNA, ecology and morphology: interpreting the phylogeography of the *Nesotes* (Coleoptera: Tenebrionidae) of Gran Canaria (Canary Islands). *Molecular Ecology*, **10**, 427-434.

Regier, J.C. & Shi, D. (2005) Increased yield of PCR product from degenerate primers with nondegenerate, nonhomologous 5 ' tails. *Biotechniques*, **38**, 34-+.

Rheindt F.E., Christidis L., & Norman J.A. (2009) Genetic introgression, incomplete lineage sorting and faulty taxonomy create multiple cases of polyphyly in a montane clade of tyrant-flycatchers (Elaenia, Tyrannidae). Zoologica Scripta, **38**, 143–153.

Roderick, G.K. & Gillespie, R.G. (1998) Speciation and phylogeography of Hawaiian terrestrial arthropods. Molecular Ecology, **7**, 519-531.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard, J. P. Huelsenbeck (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. . *Systematic Biology*, **61**, 539-542.

Roos C., Zinner D., Kubatko L.S., Schwarz C., Yang M., Meyer D., Nash S.D., Xing J., Batzer M.A., Brameier M., Leendertz F.H., Ziegler T., Perwitasari-Farajallah D., Nadler T., Walter L., & Osterholz M. (2011) Nuclear versus mitochondrial DNA: evidence for hybridization in colobine monkeys. BMC Evolutionary Biology, **11**, 77.

Salminen M.O., Carr J.K., Burke D.S., & McCutchan F.E. (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS research and human retroviruses, **11**, 1423–1425.

Schwenk, K., Brede, N. & Streit, B. (2008) Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2805–2811.

Shaw, K.L. (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences*, **99**, 16122-16127.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, **16**, 1114–1116.

Sibley, C.G. & Sibley, F.C. (1964) Hybridization in the red-eyed towhees of Mexico: The populations of the southeastern plateau region. *Auk*, **81**, 479-504.

Spurgin, L.G., Van Oosterhout, C., Illera, J.C., Bridgett, S., Gharbi, K., Emerson, B.C. & Richardson, D.S. (2011) Gene conversion rapidly generates major histobompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, **20,** 5213-5225**.**

Stephens, M. & Scheet, P. (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, **76**, 449-462.

Stephens, M., Smith, N.J. & Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978-989.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Tang Q.-Y., Liu S.-Q., Yu D., Liu H.-Z., & Danley P.D. (2012) Mitochondrial capture and incomplete lineage sorting in the diversification of balitorine loaches (Cypriniformes, Balitoridae) revealed by mitochondrial and nuclear genes. *Zoologica Scripta*, **41**, 233–247.

Teacher, A.G.F. & Griffiths, D.J. (2011) HapStar: automated haplotype network layout and visualization. *Molecular Ecology Resources*, **11**, 151-153.

Templeton, A.R., Crandall, K.A. & Sing, C.F. (1992) A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA-sequence data. 3. Cladogram estimation. *Genetics*, **132**, 619-633.

Toews, D.P.L. & Brelsford, A. (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907-3930.

Villacorta, C., Jaume, D., Oromí, P. & Juan, C. (2008) Under the volcano: phylogeography and evlution of the cave-dwelling *Palmarchestia hypogaea* (Amphipoda, Crustacea) at La Palma (Canary Islands). *BMC Biology*, **6:7.**

Vogler, A.P. & Timmermans, M.J. (2012) Speciation: don't fly and diversify? *Current Biology*, **22**, R284-R286.

Xia, X.H., Xie, Z., Salemi, M., Chen, L. & Wang, Y. (2003) An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, **26**, 1-7.

Xu, B. & Yang, Z. (2013) PAMLX: a graphical user interface for PAML. *Molecular Biology and Evolution*, **30**, 2723-2724.

Wilson R.E., Eaton M.D., Sonsthagen S.A., Peters J.L., Johnson K.P., Simarra B., & McCracken K.G. (2011) Speciation, subspecies divergence, and paraphyly in the cinnamon teal and blue-winged teal. *The Condor*, **113**, 747–761.

Yang, Z. (2007) PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586-1591.

Zarza, E., Reynoso, V.H. & Emerson, B.C. (2008) Diversification in the northern neotropics: mitochondrial and nuclear DNA phylogeography of the iguana *Ctenosaura pectinata* and related species. *Molecular Ecology*, **17**, 3259-3275.

2.7 Appendices (electronic version)

Figure S2.1

Bayesian phylogenetic tree of the complete mtDNA COII data set (172 sequences, 633 bp) inferred using the GTR+G model of sequence evolution showing relationships within the Canarian Islands *L. tessellatus* complex. The tree is rooted with *L. vicinus*. Bayesian posterior probabilities are shown below nodes. Letters and numbers immediately to the right of species names correspond to location codes, see Table 2.1.

Figure S2.2

Bayesian phylogenetic tree of the complete ITS2 nuclear gene dataset (330 sequences, 411 bp) inferred using the GTR+G model of sequence evolution showing relationships within the Canarian Islands *L. tessellatus* complex. The tree is rooted with *L. vicinus*. Bayesian posterior probabilities are shown below nodes. Letters and numbers immediately to the right of species names correspond to location codes, see Table 2.1.

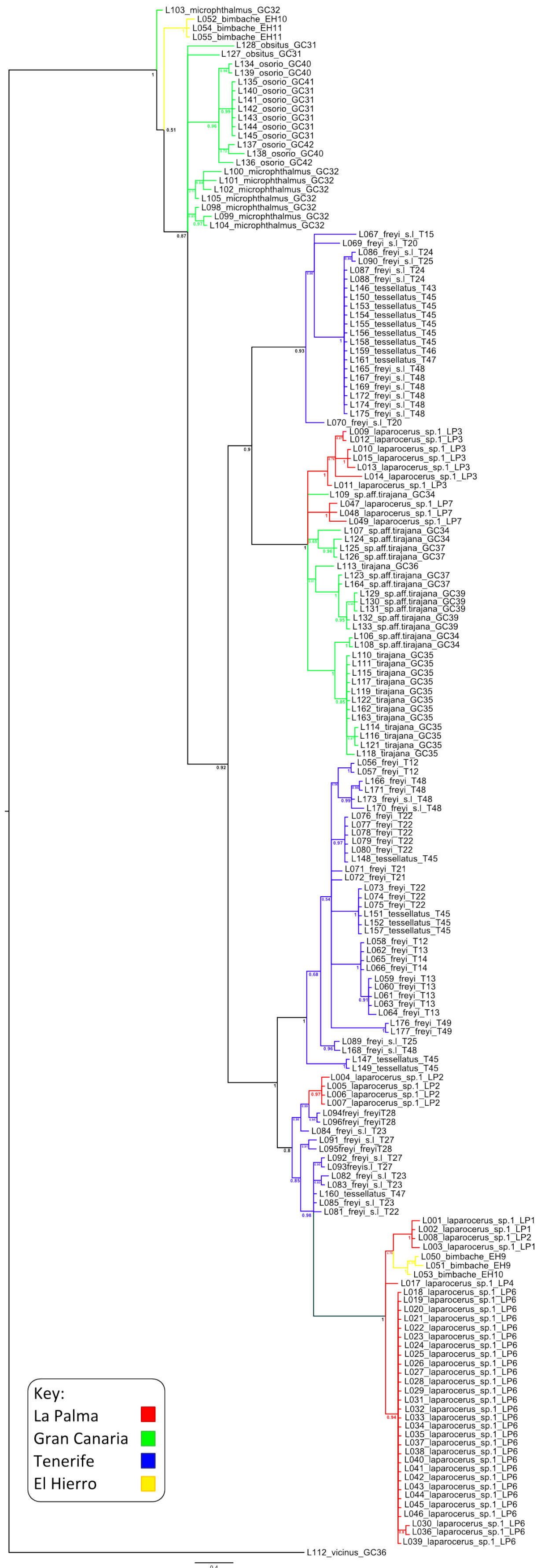


Figure S2.1

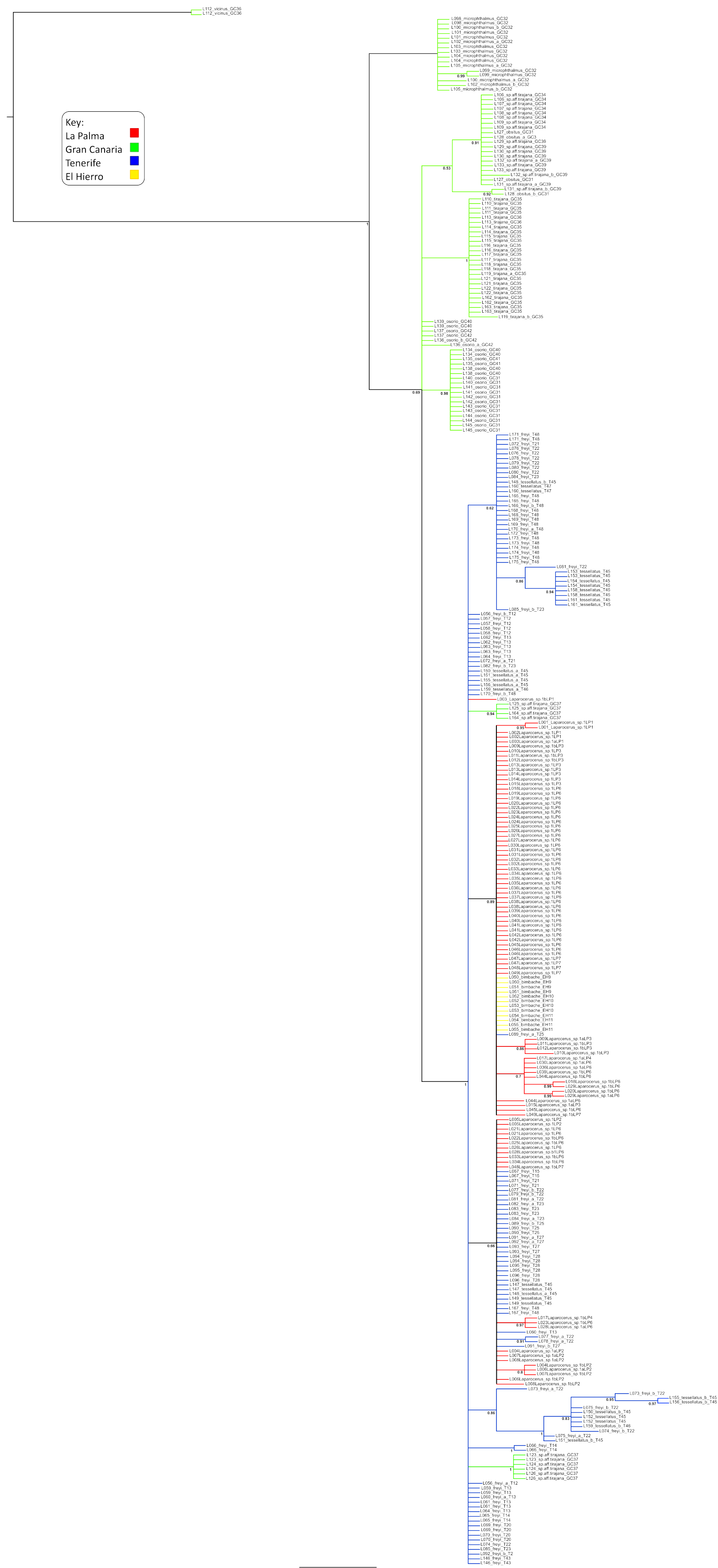
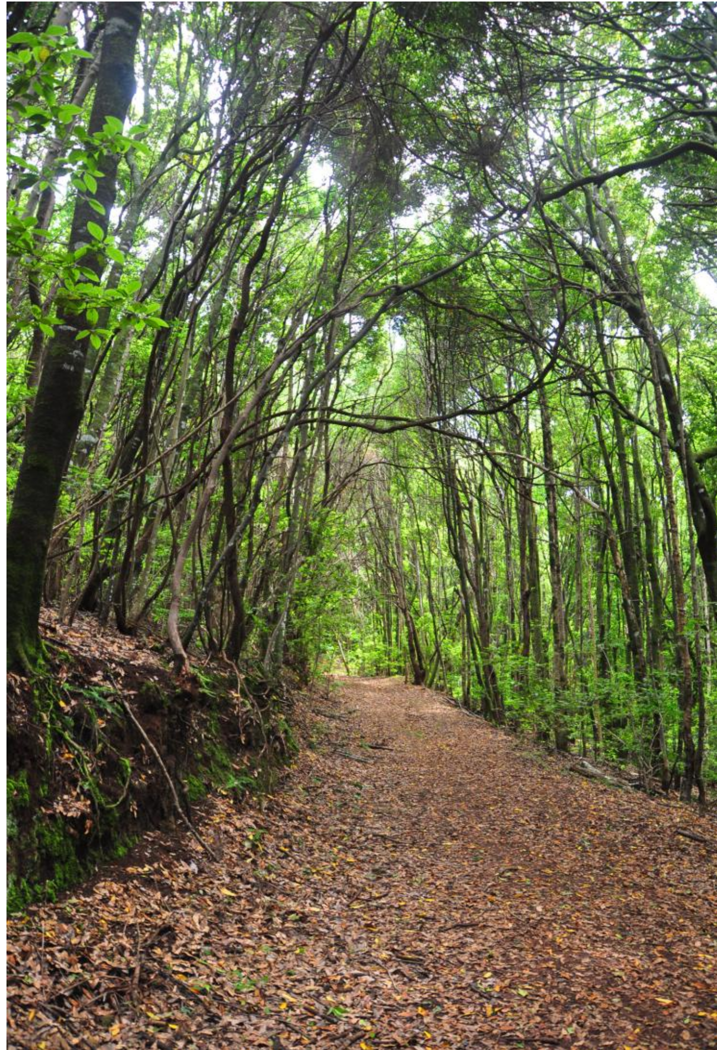


Figure S2.2

Chapter 3

Community level assessment of arthropod colonisation history: an example using the Collembola of Tenerife



Sampling site in the laurel forest of Anaga, northeast of Tenerife. Photo by Conrad Gillett

3.1 ABSTRACT

Oceanic islands offer an ideal temporal and spatial framework to study community processes, in particular the dynamics of colonisation within mainland-island models over time. However, only a few studies to date have examined this dynamic using the historical information provided by molecular data for a large number of species. Collembola is an ancient group of small, wingless invertebrates, extremely abundant in soil and leaf litter, and widely distributed throughout the world. Due to their high density, virtual ubiquity and limited dispersal abilities, they represent a good model system to study evolutionary processes. In this study, Collembola were sampled from Tenerife to understand the process of island colonisation for small soil dwelling invertebrate fauna, an understudied component of biodiversity despite its significant representation. Specifically, a high throughput parallel sequencing approach with the barcode gene COI was used to quantify species richness, estimate the genetic relatedness of island and mainland taxa, and infer the distribution of lineage colonisation times. In total, 117 Collembola lineages are described for Tenerife community with an average of 15 lineages sampled per site. The majority of lineages are restricted to a single site, and only a few lineages are widely distributed. The plot of frequency distributions of the observed genetic distances reveals a clear bimodal distribution of pairwise distances between the Tenerife OTUs and their closest matches outside the archipelago (BOLD database). These results indicate the Collembola of Tenerife to be a mosaic of taxa that are either genetically very closely related to non-Canary Island species, or taxa that are genetically very divergent from non-Canary Island species, with a scarcity of lineages of intermediate relatedness. The broad geographic distribution, and apparent genetic uniformity of many of non-Canary Island taxa that are genetically very close to Canary Island taxa reveals a probable origin by human introduction for some of these species.

3.2 INTRODUCTION

Community composition at any point in time results from the interplay between immigration, speciation and extinction. It is constrained by the evolutionary history and availability of species in the regional species pool (Ricklefs, 2004), the dispersal abilities of these species that allow them to reach a new area (e.g. an island) (Lomolino et al., 2010), and their ability to tolerate abiotic (environmental filtering) and biotic (interactions among species) conditions once they arrived in the newly colonised area (HilleRisLambers et al., 2012). There has been increasing interest in understanding community assembly and the processes that drive it at multiple spatio-temporal scales (e.g. Parent & Crespi, 2006; Emerson & Gillespie, 2008; Papadopoulou et al., 2011), and this has been facilitated by the use of molecular data together with developments in computing power and informatics tools. Studies of phylogenetic community ecology have greatly illuminated our knowledge about the formation and structuring of biological communities and we now have a better understanding of the processes driving community assembly, maintenance and changes over time (Cavender-Bares et al., 2009; Vamosi et al., 2009; Vellend et al., 2014). This is fundamental in a time of rapid global change and pervasive anthropogenic threats to biological communities worldwide.

The theory of island biogeography (MacArthur and Wilson, 1967) provides the most influential general framework to study biological communities, specifying that the rates of interisland colonization and extinction are responsible for the generation and maintenance of an island biota (Hengeveld, 2002). The theory was later expanded to incorporate within-island speciation as another important process in generating diversity on islands at particular spatial and temporal scales (Losos & Schluter, 2000). In the early stages of community formation, immigration is thought to be the dominant process, and the dispersive, adaptive and competitive quality of colonists will determine their establishment success in the new colonized area. Island biogeography theory (IBT) predicts that the size of an island and its distance to a source area are important variables because they determine the amount of empty niches available to colonize, and the immigration probability, which will vary negatively with distance. Over time, niches are filled, migrant alleles are less likely to become fixed in the island population and

speciation can become a more important process (Emerson & Gillespie, 2008). Despite efforts to place phylogenetic analyses of communities within deeper temporal scales, studies have mostly focused on ecological times (examining recent dispersal, colonization, establishment) rather than evolutionary times where speciation features. Several recent examples have included a multi-taxa approach to assess genetic diversity of local assemblages, including tenebrionid beetle communities sampled across continental islands of the Aegean archipelago (Papadopoulou et al., 2011), and a large-scale DNA sequencing approach to study the genotypic diversity of the entire aquatic communities of beetles across Europe (Baselga et al., 2013). While these studies considered evolutionary time taking age of the islands and age of clades into account, they have addressed very different temporal scales.

Understanding the processes underlying community assemblage (immigration, speciation and extinction) depends on the ability to measure their rates and variation over time, and this can be facilitated by the application of molecular data and use of appropriate model systems. Oceanic islands offer an ideal temporal and spatial system to study community processes, in particular the dynamics of colonisation within mainland-island models over time, because of their discrete nature and often known geological history. However, only a few studies have examined this dynamic using the historical information provided by molecular data for a large number of species. For example, Ricklefs and Bermingham (2001) examined the connectivity between island and mainland bird species by estimating genetic distances between populations and analysing the distribution of colonisation times of extant island lineages (Ricklefs & Bermingham, 2001). They plotted the cumulative curve of lineages of land birds in the Lesser Antilles as a function of increasing relative age of colonisation. This curve was expected to increase exponentially towards an equilibrium point, when colonisation and extinction are constant over time and homogeneous over lineages (Ricklefs & Bermingham, 2004). However, they found it did not reach equilibrium for the Lesser Antilles birds and presented a striking change in slope at a mitochondrial DNA genetic distance of approximately 2% (Fig 3.1). An explanation put forward for this result was the speciation threshold model (Johnson et al. 2000), according to which the genetic divergence among recent colonists is slowed by migration from the source, and species accumulate at a low genetic distance prior to a speciation threshold. Once this threshold

is achieved, source and island populations diverge more rapidly generating a heterogeneous distribution of apparent ages (genetic distance values) of island taxa. The shape of this distribution is expected to be a negative exponential decay (Fig 3.2) with a high density prior to the speciation threshold and a sparser distribution over a wider range of divergence values above that threshold (Fig 3, Johnson et al., 2000). This approach and expectations (cumulative curve of lineages with an abrupt change of slope and negative exponential decay of island taxa ages) are useful because they give insights about the dynamics of colonisation over time and the build-up of island biota. However, they are only useful for very dispersive taxa, and we lack data or models for other groups, especially low dispersive taxa such as flightless groups.

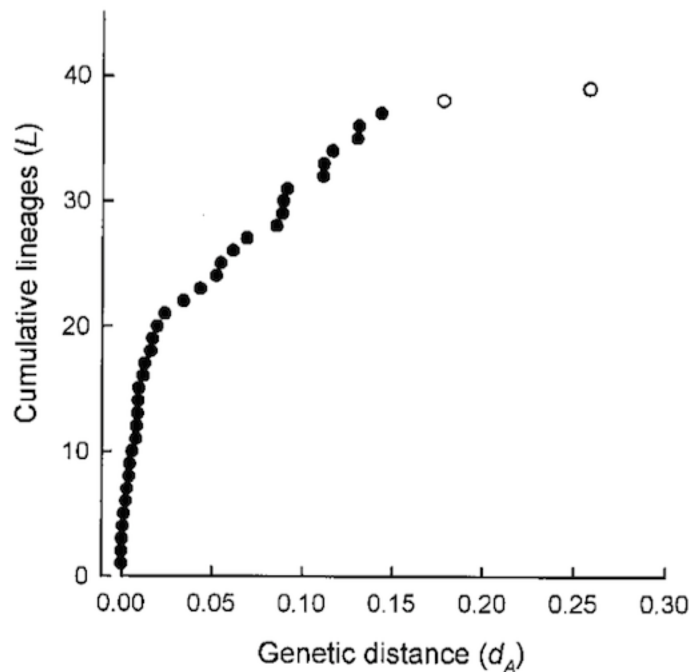


Figure 3.1. Cumulative curve of lineages of land birds in the Lesser Antilles as a function of increasing relative age of colonisation (genetic distance, d_A). Contrary to expectations, the curve did not reach equilibrium (and it presented a striking change in slope at a mitochondrial DNA genetic distance of approximately 2% (reproduced from Rickelfs and Bermingham [2001]).

Collembola is an ancient group of flightless invertebrates whose families evolved more than 200 million years ago (Hopkin, 1997). They are small, wingless animals, extremely

abundant in soil and leaf litter, and widely distributed throughout the world, including areas of extreme climates such as deserts and polar regions (Hopkin, 1997). Their key role for ecosystem functioning lies in influencing soil structure and decomposition through feeding on fungal hyphae and releasing faecal pellets, which become part of the black humus (Hopkin, 1997). Due to their high density, virtual ubiquity and limited dispersal abilities, they represent a good model system to study evolutionary processes. Although one of the weaknesses of the group is the difficulty to assign taxonomy (the "taxonomic impediment"), molecular studies addressing springtails as model organisms demonstrate their ability to inform about evolutionary history even at very fine geographical scales (e.g Garrick et al., 2007; Cicconardi et al., 2010). High levels of genetic differentiation consistent with cryptic diversity has been recently documented for many collembolan genera and species around the world, e.g. *Lepidocyrtus* from the North-Western Mediterranean basin and Panama (Cicconardi et al., 2010, 2013), *Megalothorax* from France (Schneider et al., 2011), *Tomocerus* from China (Zhang et al., 2014), *Bilobella aurantiaca* from France (Porco et al., 2012a), *Frisea frisea* from Antartica (Torricelli et al., 2010), *Parasitoma notabilis* from many regions in Europe and Canada (Porco et al., 2012b). These studies have also repeatedly revealed deep divergences among lineages within species defined by classical taxonomy, indicating that the broad distributions of many species recognised by traditional morphological taxonomy in fact represent cryptic species diversity.

The aim of this chapter is to use Collembola sampled from the island of Tenerife to understand the process of island colonisation for small invertebrates from the soil fauna. Specifically, a molecular approach is used to estimate the genetic relatedness of island and mainland taxa, and infer the distribution of lineage colonisation times. In the Canary Islands, Collembola diversity has been mainly documented based on morphological taxonomy and although most of the species are thought to be native, the possibility that some species could be the result of introductions has been suggested (Arechavaleta et al., 2010) and recently reported. Ramirez-Gonzalez et al. (2013) found three Tenerife sequences assigned to *Parasitoma notabilis* to be identical to sequences sampled from Europe, North Africa and Australia. Attempts to solve the possible origins (whether native or introduced) of the species become more complex due to the fact that many species have broad geographic ranges that extend beyond the archipelago

(Emerson et al., 2011) and by the fact that cryptic diversity is extensive within Collembola (e.g Porco et al., 2012a).

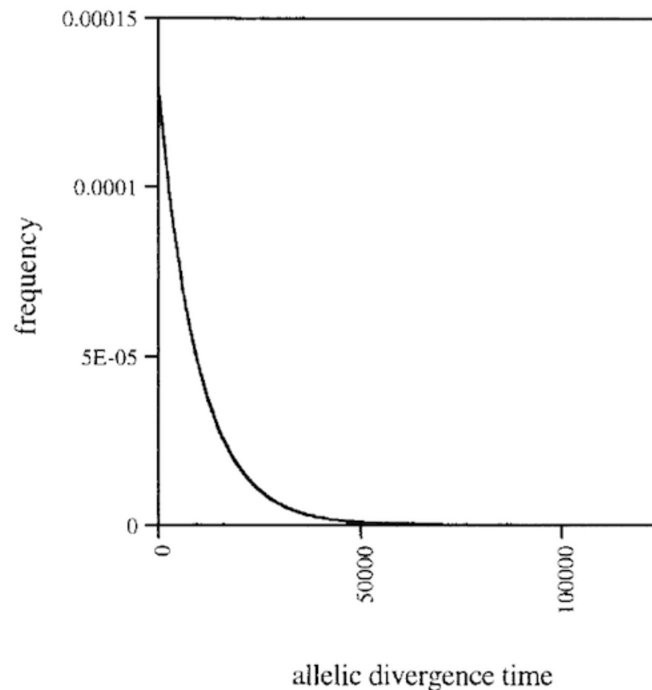


Figure 3.2. Mean allelic divergence times of alleles across species on the same island. This distribution is a negative exponential decay with ahigh density prior to the speciation threshold and a sparser distribution over a wider range of divergence values above that threshold (reproduced from Johnson et al. [2000]).

To characterise Collembola evolutionary diversity in Tenerife and infer the history of colonisation of the island, we analysed sequence data (220 bp of the barcode gene COI obtained from a 454 sequencing platform) from 2500 individuals collected at 25 sampling sites from across the island. Next generation sequencing (NGS) allows large-scale DNA sequencing of entire species assemblages and this provides a way to quantify species and genetic diversity collectively for a given community (Baselga et al., 2013). The barcode gene COI is used as the amplicon for high throughput parallel (HTP) sequencing because of its distinct advantages for the assessment of intra and

interspecific analyses of mesofaunal taxa compared to commonly used markers for amplicon HTP sequencing, such as 16S for bacteria (e.g. Chu et al., 2010), rRNA for fungi (e.g. Rousk et al., 2010), protists (e.g. Medinger et al., 2010) and marine mesofaunal elements (e.g. Fonseca et al., 2010). MtDNA COI is a single copy gene, has highly conserved regions spanning variable regions which facilitates cross-taxon amplification, it maximises the capture sample of taxonomic diversity for pyrosequencing within a focal group, and has faster evolutionary substitution rate compared to the nuclear rRNA genes (Emerson et al., 2011). Furthermore, it has no indels (facilitating the identification of non-target sequences) and it is well represented on public databases (Ramirez-Gonzalez et al., 2013). At the time of this study, there were 50,294 Collembola specimens with barcodes in the BOLD Identification System (http://www.boldsystems.org) from different parts of the world. This represents a valuable tool to assign taxonomy for the 454 sequences from Tenerife and to understand the dynamics of colonisation and community assembly of islands and archipelagos.

A previous pilot study analysed sequence data from two of the 25 sampling sites of this study to develop an algorithm to separate pyrosequencing noise from true sequence diversity (Ramirez-Gonzalez et al., 2013). The study revealed the potential for pyrosequencing data to successfully recover both interspecific and intraspecific sequence diversity. Furthermore, the comparison of Canary Island sequence data to the BOLD database revealed it to be mainly comprised of two classes of sequences, across the 24 divergent lineages found. Seven sequences (30%) were genetically identical or nearly identical to sequences sampled outside the archipelago, while 13 sequences (54%) did not have any close match outside the archipelago (between 89-92% similarity). Although a rather small dataset, it suggested a bimodal distribution for the frequency of allelic divergences across species on a single island. Here we extend the sampling of Ramirez-Gonzalez et al. (2013) to a dataset of 25 sites for a broader analysis of the Collembola community assembly within the Canary Islands to assess whether this bimodal distribution could represent a general pattern for the group.

3.3 MATERIALS AND METHODS

3.3.1 Sample collection and laboratory work

Soil samples were collected at 25 sampling sites from across Tenerife (Fig 3.2) and placed in Tullgren funnels to extract Collembola, as described in Ramirez-Gonzalez et al. (2013). For each site, one hundred Collembola were randomly sampled and combined for DNA extraction, followed by amplification of 307 bp mtDNA COI barcode region using primers 454ColFol-for and 454Col307-rev, with 454 adaptors A and B at the 5' ends of the respective primers. For each of the 25 pools, PCRs were performed in triplicate; with each replicate having a different multiplex identifier MID-tagged A adaptor (only the forward direction was pyrosequenced). MIDs allow different samples to be sequenced together on a single 454 plate and then separated bioinformatically for downstream analysis. DNA extraction, PCR and sequencing conditions are described in detail in Ramirez-Gonzales et al. (2013). PCR products were purified, normalized and equimolar pooled, then sequenced on two 1/2 plates of the Roche 454 GSFLX pyrosequencing platform within The Genome Analysis Centre, Norwich Research Park, Norwich, United Kingdom.

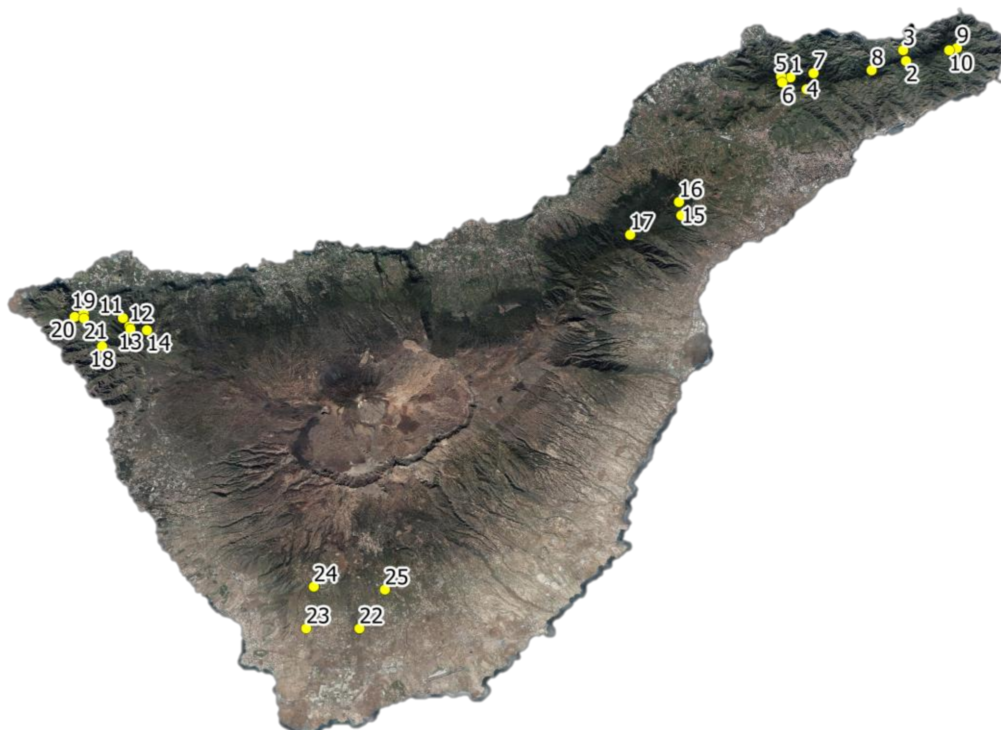


Figure 3.3. Distribution of sampling sites in the island of Tenerife; a complete list with geographic coordinates and number of lineages sampled per localities can be found in Table 3.1.

3.3.2 Data analysis

Sequences generated from the Roche 454 GSFLX platform were processed using the PyroClean denoising algorithm (Ramirez-Gonzalez et al., 2013). PyroClean involves five steps for the removal of non-target sequences (sequencing error, PCR error, pseudogenes, chimeric sequences, or contamination) and the generation of a PyroCleaned output alignment. In the first step, the raw 454 library data was split by MID and converted to fasta format. The forward primer was trimmed from the fasta file and reads were collapsed to unique sequences, with the number of reads appended to each sequence name. In step 2, an informative consensus reference sequence was constructed to detect and correct read errors in homopolymer regions (e.g. insertion and deletions - indels) that are quantifiable as frameshift events. This reference sequence was then used in step 3 as anchor points for denoising reads generated from pyrosequencing. Reads were analysed individually by identifying and correcting indels as well as insertion deletion compensation events that are associated with homopolymers. In step 4 reads divergent from the consensus reference sequence (with nucleotide variation attributable to PCR error, sequencing error, numts, or non-target sequence) were filtered out and the filter threshold used was 1 permissible mismatch between a read and the consensus reference sequence. Low frequency sequences such as singletons were excluded a priori. The alignment was then trimmed to exclude remaining 3' homopolymer error and sequences that differed by only an ambiguity code from another sequence were merged into a single sequence without the ambiguity code. Finally, in step 5, reads presenting nucleotide variation consistent with the consensus reference sequence, but still attributable to PCR or sequencing error, or numts, were removed. The PyroCleaned output alignment obtained was then processed to obtain clusters of sequences presumably representing Collembola species and to remove clusters of sequences that cannot be reliably attributed to Collembola clusters (remaining non-target sequences that persisted in the alignment).

3.3.3 Cluster assignment

DNACLUST (Ghodsi et al., 2011) was used for clustering the large number of sequences of the PyroCleaned output alignment down to OTUs (Operational Taxonomic

Unit, “OTU picking”). Here lineages are defined operationally as a cluster of similar sequences. To define the similarity threshold for clustering, the alignment was subjected to DNACLUST analyses using a range of similarities values (from 85% to 99% similarity, with increments of 1%) and the number of clusters obtained at each value was quantitatively compared in relation to the rate of increase in the number of OTUs produced by increasing the similarity level. The threshold that generated the maximum number of clusters with the smallest rate of increase (i.e. minimum number of singleton clusters) was chosen as the similarity value for defining lineages. Then, to assign taxonomy and accept (or not) an OTU as a Collembola lineage, the most frequent sequence of each cluster (i.e. the one with the highest number of reads) was submitted to the BOLD Identification System (http://www.boldsystems.org), and using bioinformatic tools implemented within this system, a distance analysis was performed. By default, BOLD uses the Kimura 2 parameter distance model to calculate pairwise distances between the searched sequence and all the other sequences on the database, and returns the percentage of similarity for the top 99 closest matches found. The maximum similarity hit to a non-Canarian Collembola and the number of matches to both Collembola and non-Collembola sequences within the 99 top matches were computed and compared. The limited taxonomic resolution of the first 220 bases of the barcode gene and the disproportionately higher number of non-Collembola sequences on the BOLD database when compared to Collembola resulted in many PyroCleaned sequences matching to non-collembolan taxa on the BOLD database in Ramirez-Gonzalez et al. (2013). To take this into account, we took two approaches to define a threshold for accepting a cluster to be a Collembola species. First, we assessed the taxonomic resolution of the first 220 bases of the barcode gene by submitting 20 randomly chosen Collembola sequences downloaded from GenBank and 27 Tenerife Sanger sequences derived from Collembola (provided by BCE) both in their full (~600 bp) and short (220 bp) lengths to BOLD and compared how many matches to a Collembola each sequence length retrieved. Second, we evaluated the impact of the much greater representation of Insecta sequences within the BOLD database (2,257,609 Insecta specimens with barcodes as opposed to 50,294 Collembola specimens) by computing the number of matches to a collembolan sequence that is retrieved when submitting a random Insecta sequence to the BOLD system. Fifty randomly chosen Insecta sequences from the dataset in Liu et al. (2013) were downloaded from GenBank,

trimmed down to the first 220 bp and submitted to BOLD. These two approaches provided us with the cut off point, i.e. the minimum number of matches to Collembola-taxa for accepting an OTU as a true Collembola. To further assign taxonomy and estimate the number of lineages in Tenerife, all genuine Collembola clusters (one representative sequence per cluster again chosen as the one with the highest number of reads) were subjected to a neighbour-joining analysis with the 27 Tenerife Sanger Sequences using p-distances in MEGA6 (Tamura et al., 2013) and the uncorrected genetic distances between all pair of sequences in the alignment was computed. Tenerife Sanger sequences that were closely related to any barcode cluster were collapsed to a single cluster.

Table 3.1 – Sampling sites within the island of Tenerife. Locations coded according to Fig 3.3.

Locality Code	Latitude	Longitude	N lineages
1	28° 32.161	016° 17.660	17
2	28° 32.977	016° 12.080	19
3	28° 33.508	016° 12.220	17
4	28° 31.579	016° 16.878	15
5	28° 32.211	016° 18.111	11
6	28°31.913	016°18.054	13
7	28° 32.374	016° 16.544	13
8	28° 32.519	016° 13.742	17
9	28° 33.613	016° 09.625	12
10	28° 33.507	016° 09.998	12
11	28° 20.288	016° 49.865	19
12	28° 19.876	016° 49.554	15
13	28° 19.748	016° 49.506	14
14	28° 19.687	016° 48.708	13
15	28° 25.356	016° 22.921	14
16	28° 26.014	016° 23.036	21
17	28° 24.402	016° 25.390	14
18	28° 18.891	016° 50.871	19
19	28° 20.514	016° 51.762	18
20	28° 20.343	016° 52.198	23
21	28° 20.270	016° 51.746	26
22	28° 04.966	016° 38.454	12
23	28° 04.969	016° 41.024	11
24	28° 07.031	016° 40.656	13
25	28° 06.884	016° 37.225	12

3.3.4 Frequency distribution and cumulative frequency plots

Once genuine Collembola short barcode clusters were separated from non-Collembola clusters (by checking the similarity hit and number of matches to Collembola sequences returned when queried on BOLD), and OTUs were verified (by the joint neighbour-joining analysis with the taxonomically referenced Sanger sequences), a frequency distribution of pairwise genetic distances between every lineage and its closest sequence outside the archipelago (which is the first match to a non-Tenerife Collembola retrieved by BOLD) was computed, and a cumulative frequency plot of the number of lineages with increasing genetic distance was constructed.

3.4 RESULTS

3.4.1 PyroCleaning results

After data processing with the Roche 454 software package, amplicon PCRs for the three MIDs within each pool (25 sites) yielded broadly similar unique raw pyrosequence counts with a few exceptions (Table S3.1). Unique raw reads had a mean maximum length of 537 bp and were on average 348 bases long. The consensus reference sequence for Collembola (step 2) was built using 1556 collembolan COI gene sequences that were available from EMBL/GenBank. In step 3, the PyroClean algorithm was implemented to correct homopolymer error and singletons were excluded. The alignment was then trimmed to the first 220 nucleotides and filtered for 1 mismatch to the consensus reference (step 4). A second filter was performed in step 5 and the average number of unique sequences within each pool was 42 (ranging from 28 to 76). None of the sequences were found to be of chimeric origin. The final PyroCleaned output alignment consisted of 3,400 unique sequences.

3.4.2 Clustering thresholds

The rate of increase in the number of clusters obtained when running DNACLUST at a range of similarities was small when increasing the threshold from 85% up to 95% (with an average of 10 clusters added for each threshold increase), medium, when rising

from 96% to 98%, and very big when rising from 99% to 99.9% (Table 3.2). We chose 95% similarity for clustering, as this is the highest similarity value before the number of lineages greatly increased. The rate of increase in the number of clusters was smaller than in the previous two thresholds (additional 10 as opposed to 17 and 14 clusters added at similarities 94% and 93%, respectively), indicating this to be a suitable similarity threshold for clustering. The 3,400 sequences in the PyroCleaned output alignment were then clustered at 95% similarity and reduced down to a total of 182 clusters that are at least 5% divergent from each other.

3.4.3 Taxonomic assignment

Seventeen out of the 20 randomly chosen Collembola Sanger sequences submitted to BOLD matched only to Collembola taxa within the top 99 matches when searched with their full length (Table 3.3). The three remaining sequences matched mostly to Collembola but presented 1 to 3 matches (out of the top 99) to an Insecta sequence. When reduced to the first 220 bp, only two sequences matched only to Collembola sequences. The remaining 18 sequences presented at least a few, but usually many matches to non-Collembola taxa, especially to Insecta. The majority of the Tenerife Sanger sequences, when reduced to 220 bp, had very low frequency matches to another Collembola sequences on BOLD (<10% out of the top 99) and they mostly presented low similarity hits (<94%) (Table S3.2). Surprisingly, two of these sequences (Col213 and Col226) matched to not a single Collembola on BOLD. When queried in its full length (692 bp), Col226 matched only to 4 Collembola, and its top similarity hit was 79.8%, the lowest hit found among all sequences searched on BOLD for this study. The full length of Col213 was only 282 bp long, and it matched to a single Collembola sequence, with top similarity hit 91.3%. These tests indicated that true Collembola sequences may exhibit low matches to Collembola sequences on BOLD, and apparent higher matching to Insecta sequences, and that, although this is most dramatic at reduced sequence lengths, it is also a feature of full sequence lengths. When submitting randomly chosen Insecta sequences to BOLD, also reduced to the first 220 nucleotides, the majority of them matched only to Insecta sequences on the database (Table 3.4). Five out of the 50 Insecta sequences presented Collembola matches in very low

frequencies (maximum number of Collembola matches retrieved was 5 out of the 99 top matches).

Based on the above findings, and in the absence of high similarity matches, we set a threshold of at least 6 Collembola matches within the top 99 BOLD matches for a cluster to be considered to be a species of Collembola. With this threshold, a total of 129 out of the 182 clusters submitted to BOLD were considered to represent a species of Collembola (Table S3.3), with 29 of these having an identical or nearly identical (≥99% match) sequences on the BOLD database. The remaining 100 clusters presented a variety of low to high levels of genetic similarity (98% to 83%) to the most closely related Collembola sequence outside the Canary Islands (Table S3.3).

Table 3.2 - Number of clusters produced when clustering the PyroCleaned output alignment (n=3,400 sequences) at increasing similarity values, and the number of additional clusters added (rate of increase) at each threshold.

Similarity threshold	N of clusters	Rate of Increase
0.85	82	-
0.86	93	11
0.87	103	10
0.88	110	7
0.89	117	7
0.9	124	7
0.91	133	9
0.92	141	8
0.93	155	14
0.94	172	17
0.95	182	10
0.96	221	39
0.97	271	50
0.98	359	88
0.99	567	208
0.995	803	236
0.999	1056	253

The 53 clusters that matched to none or to a very few Collembola sequences were removed from the dataset. Highest matching hits for presumably non-Collembola clusters ranged from 89% to 98% similarity to another sequence on the database and the identities of those with high matches (≥96% similar) were mostly Lepidoptera, Diptera and Coleoptera (Table S3.4). Among the clusters with low similarity hits (90 to 92%), the highest hit for 13 clusters was not to an Insecta sequence but to the following taxa: Amphipoda (5), Actinopterygii (2), Anura (1), Decapoda (1), Euonychophora (2), Rodentia (1), Passeriformes (1). It should be noted that, based on the analyses described above, some of these may actually represent Collembola. This type II error (rejection of Collembola lineages as Collembola) should be offset by a lower type I error (acceptance of non-Collembola lineages as Collembola).

Table 3.3 Number of matches to Collembola, Insecta or other taxa within the top 99 matches returned by the BOLD System Identification database when querying with randomly chosen full length Collembola Sanger sequences, and shorter lengths of these.

Query Sequence	Full length (~600 bp)			Short length (220 bp)		
	Collembola	Insecta	Other taxa	Collembola	Insecta	Other taxa
1	99	0	0	25	73	1
2	99	0	0	4	95	0
3	99	0	0	93	6	0
4	99	0	0	44	52	3
5	99	0	0	4	95	0
6	99	0	0	96	3	0
7	99	0	0	99	0	0
8	99	0	0	1	98	0
9	99	0	0	29	65	5
10	96	3	0	2	97	0
11	99	0	0	25	49	25
12	99	0	0	98	1	0
13	99	0	0	26	47	26
14	97	2	0	36	63	0
15	99	0	0	63	10	26
16	99	0	0	48	51	0
17	99	0	0	99	0	0
18	99	0	0	53	40	6
19	99	0	0	26	68	5
20	98	1	0	2	97	0

Table 3.4 - Number of matches to Insecta, Collembola or other taxa within the top 99 matches returned by BOLD System Identification database when querying with randomly chosen Insecta sequences reduced to 220 bp.

Query Sequence		Number of matches to:		
		Insecta	Collembola	Other taxa*
1	JQ344393	94	**5**	0
2	JQ344935	99	0	0
3	JQ344793	99	0	0
4	JQ344582	99	0	0
5	JQ344597	99	0	0
6	JQ344719	99	0	0
7	JQ344720	99	0	0
8	JQ344696	99	0	0
9	JQ344697	99	0	0
10	JQ344468	99	0	0
11	JQ344529	99	0	0
12	JQ344857	99	0	0
13	JQ344965	96	**2**	1
14	JQ344967	99	0	0
15	JQ344828	98	0	1
16	JQ345029	99	0	0
17	JQ344361	98	0	1
18	JQ344842	82	0	17
19	JQ344545	99	0	0
20	JQ344788	99	0	0
21	JQ344413	99	0	0
22	JQ344508	99	0	0
23	JQ344768	97	0	2
24	JQ344777	97	**1**	1
25	JQ344784	99	0	0
26	JQ344788	99	0	0
27	JQ344880	99	0	0
28	JQ344949	97	0	2
29	JQ344975	99	0	0
30	JQ345012	99	0	0
31	JQ344413	98	0	1
32	JQ344437	99	0	0
33	JQ344470	99	0	0
34	JQ344536	99	0	0
35	JQ344564	99	0	0
36	JQ344660	99	0	0
37	JQ344676	99	0	0
38	JQ344782	97	0	2

Table 3.4 (cont.)

Query Sequence		Number of matches to: Insecta	Collembola	Other taxa*
39	JQ344869	99	0	0
40	JQ344879	99	0	0
41	JQ344888	99	0	0
42	JQ344890	99	0	0
43	JQ344917	98	**1**	0
44	JQ344939	99	0	0
45	JQ344956	99	0	0
46	JQ345001	96	**3**	0
47	JQ344524	99	0	0
48	JQ344527	99	0	0
49	JQ344577	99	0	0
50	JQ344694	99	0	0

*These taxa include Arachnida, Malacostraca, Maxillopoda, Decapoda, Onychophoridae, Amphibia and Mammalia

3.4.4 Neighbour joining and distance analyses

The NJ and uncorrected pairwise distance analyses with the 129 presumed Collembola clusters (the sequence with the highest number of reads selected per each cluster) plus the 27 Tenerife Sanger sequences revealed 14 pairs of OTUs to be less than 5% divergent from each other. Although they mostly represent comparisons between Sanger sequences and closely related short barcode clusters (n=10 pairs), they were also found in comparisons between pairs of short barcode clusters (n=4 pairs) indicating DNACLUST failed to group them into single clusters in these instances. The 14 pairs were each collapsed to a single cluster. Furthermore, the NJ analysis revealed a well-supported group of clusters (99 bootstrap) assigned to Neanuridae comprised of 15 clusters of low frequency of reads (maximum number of reads was on average 21) closely related to a single cluster of high frequency of reads (maximum number of reads = 660) revealing an unexpected pattern of presumed numts. These numts were removed from the dataset and the Neanuridae group was reduced to the single cluster of highest frequency of reads. The final NJ dendrogram comprises 117 lineages of which eight are Sanger sequences lineages (Fig 3.4). Uncorrected pairwise divergence comparisons among OTUs ranged from 5 to 37% with a mean of 21%.

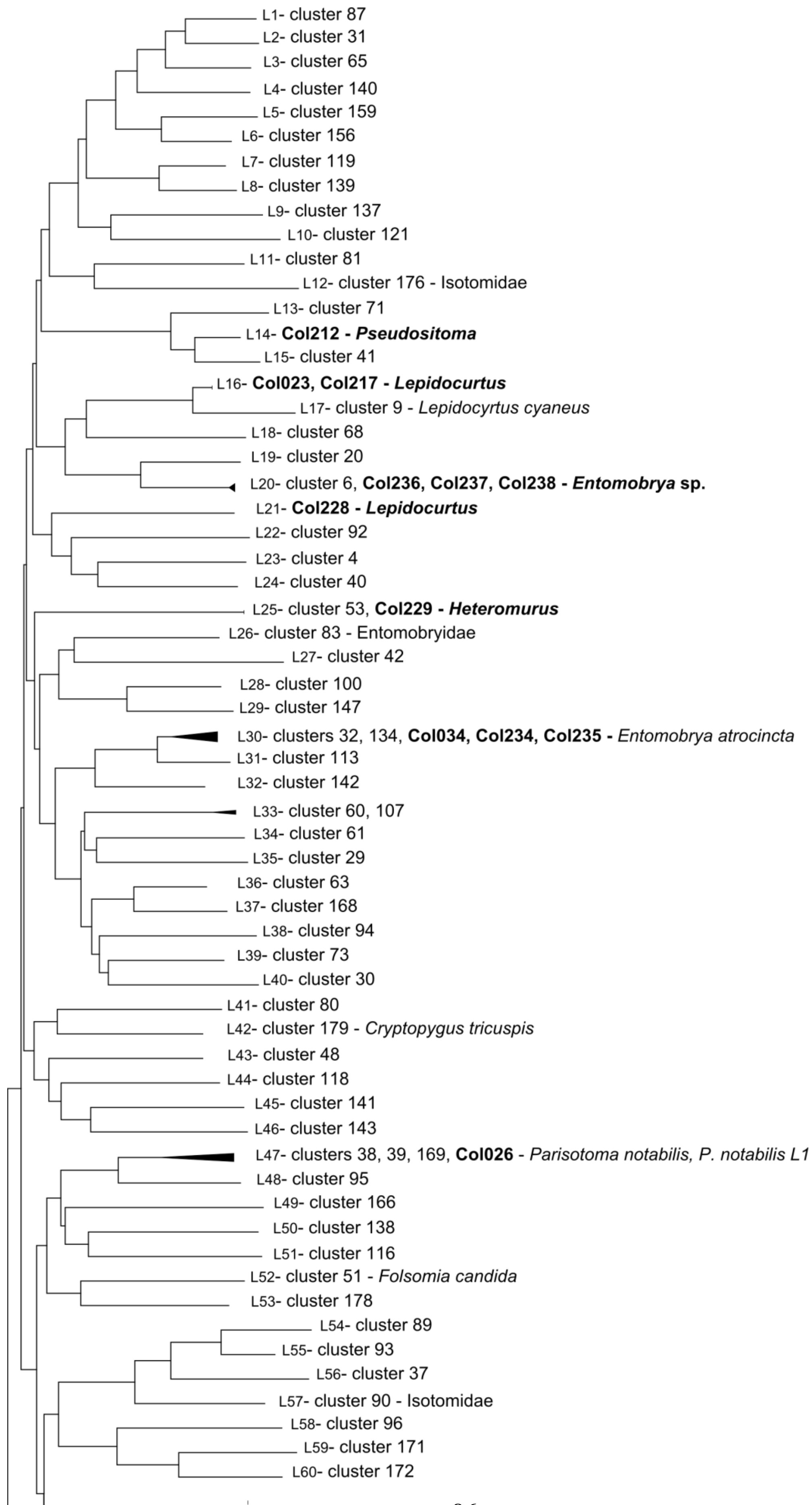

L1- cluster 87
L2- cluster 31
L3- cluster 65
L4- cluster 140
L5- cluster 159
L6- cluster 156
L7- cluster 119
L8- cluster 139
L9- cluster 137
L10- cluster 121
L11- cluster 81
L12- cluster 176 - Isotomidae
L13- cluster 71
L14- **Col212 - *Pseudositoma***
L15- cluster 41
L16- **Col023, Col217 - *Lepidocurtus***
L17- cluster 9 - *Lepidocyrtus cyaneus*
L18- cluster 68
L19- cluster 20
L20- cluster 6, **Col236, Col237, Col238 - *Entomobrya* sp.**
L21- **Col228 - *Lepidocurtus***
L22- cluster 92
L23- cluster 4
L24- cluster 40
L25- cluster 53, **Col229 - *Heteromurus***
L26- cluster 83 - Entomobryidae
L27- cluster 42
L28- cluster 100
L29- cluster 147
L30- clusters 32, 134, **Col034, Col234, Col235 -** *Entomobrya atrocincta*
L31- cluster 113
L32- cluster 142
L33- cluster 60, 107
L34- cluster 61
L35- cluster 29
L36- cluster 63
L37- cluster 168
L38- cluster 94
L39- cluster 73
L40- cluster 30
L41- cluster 80
L42- cluster 179 - *Cryptopygus tricuspis*
L43- cluster 48
L44- cluster 118
L45- cluster 141
L46- cluster 143
L47- clusters 38, 39, 169, **Col026** - *Parisotoma notabilis, P. notabilis L1*
L48- cluster 95
L49- cluster 166
L50- cluster 138
L51- cluster 116
L52- cluster 51 - *Folsomia candida*
L53- cluster 178
L54- cluster 89
L55- cluster 93
L56- cluster 37
L57- cluster 90 - Isotomidae
L58- cluster 96
L59- cluster 171
L60- cluster 172

Fig 3.4 (cont.)

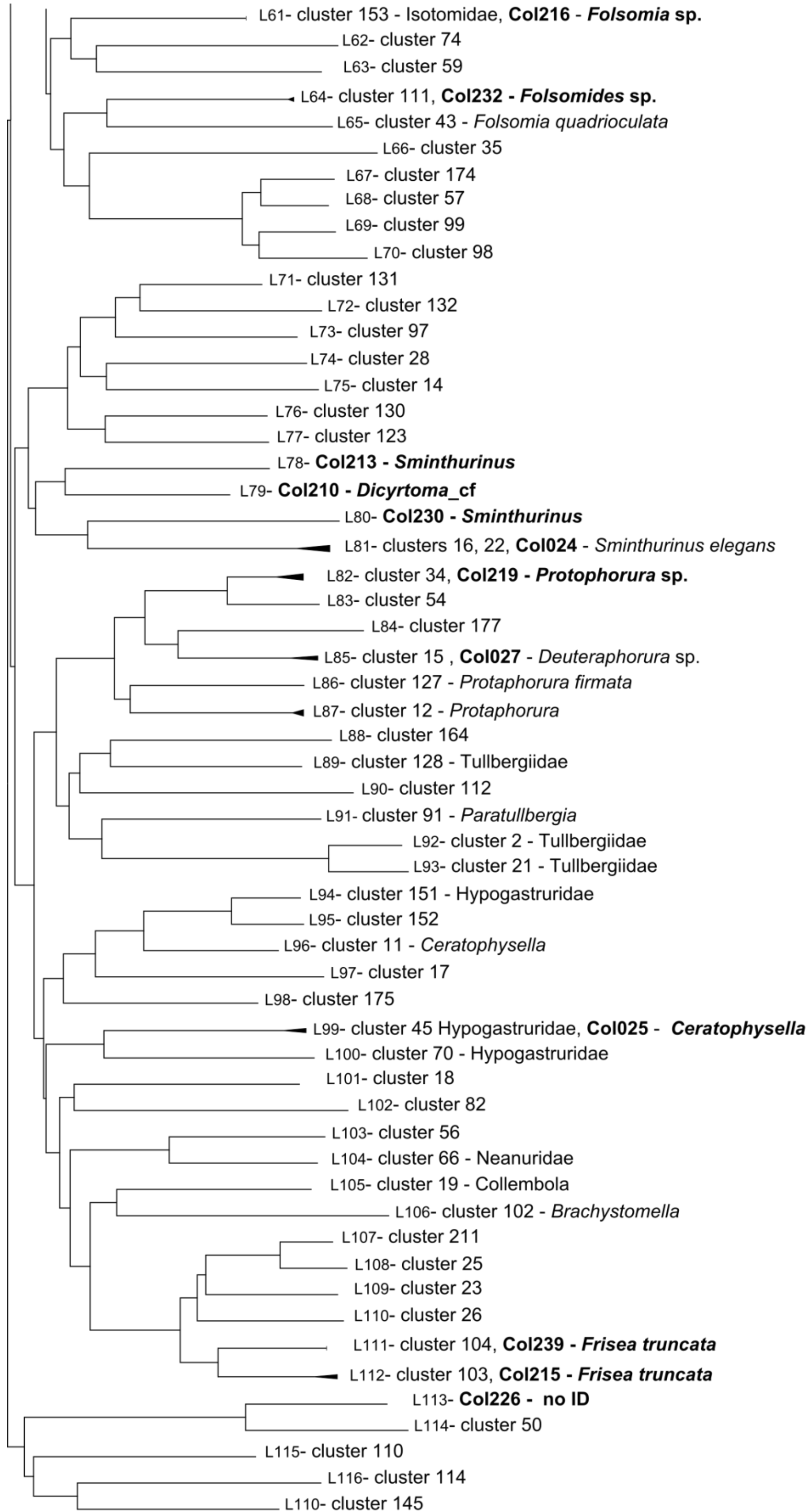


Figure 3.4. Neighbour joining tree of PyroCleaned sequences derived from 25 sampling sites in the island of Tenerife, and 27 taxonomically identified Sanger sequences sampled from Tenerife (in bold). Clusters were assessed for taxonomic identity against the BOLD Identification Database, and matches higher than 99% are reported. Taxonomy assignment for all lineages can be found in Table S3.3.

3.4.5 Tenerife community and distributions

Of the 117 Collembola lineages, based on sequence identity to the Tenerife Sanger sequences and ≥99% similarity to a sequence within the BOLD database, the following genera were identified: *Lepidocyrtus, Ceratophysella, Brachystomella, Deuteraphorura, Dicyroma, Heteromurus, Sminthurinus, Entomobrya, Parasitoma, Folsomia, Frisea, Pseudosomita, Neanura, Paratullbergia, Protaphorura, Cryptopygus, Folsomides*. The geographic distribution of the 29 lineages with ≥ 99% similarity to BOLD sequences revealed the presence of related sequences in relatively close continents (Europe and North Africa) as well as much further away (Canada, USA, South Africa, Australia and New Zealand) (Table S3.5). Across the sampling sites, the total number of lineages per site ranged from 11 to 26, with a mean of 15.8 (Table 3.1). The average number of sites at which a lineage occurs is 3.6, with the majority of lineages restricted to a single site, and only a few lineages being relatively widely distributed (Fig 3.5).

The plot of frequency distributions of the observed genetic distances (Fig 3.6) reveals a clear bimodal distribution of pairwise distances between the Tenerife OTUs (n=115, because two Sanger Sequence had no match to a Collembola on the database) and their closest matches outside the archipelago (BOLD database). The bimodal distribution has a peak at 0 per cent divergence (representing 20% of all sequence comparisons), which

decreases to reach 4 per cent divergence (with only 1 sequence comparison). A second peak appears at 8 percent divergence (representing 15% of all sequence comparisons) that is slightly reduced up to 10 percent divergence (summing up frequencies at 9 and 10 percent divergences it represents 24% of all sequence comparisons), after which the number of sequences reduced (Fig 3.6). This distribution indicates two main classes of data for Tenerife Collembola sequences, they are either very closely related or very distantly related to Collembola sequences sampled from outside the archipelago.

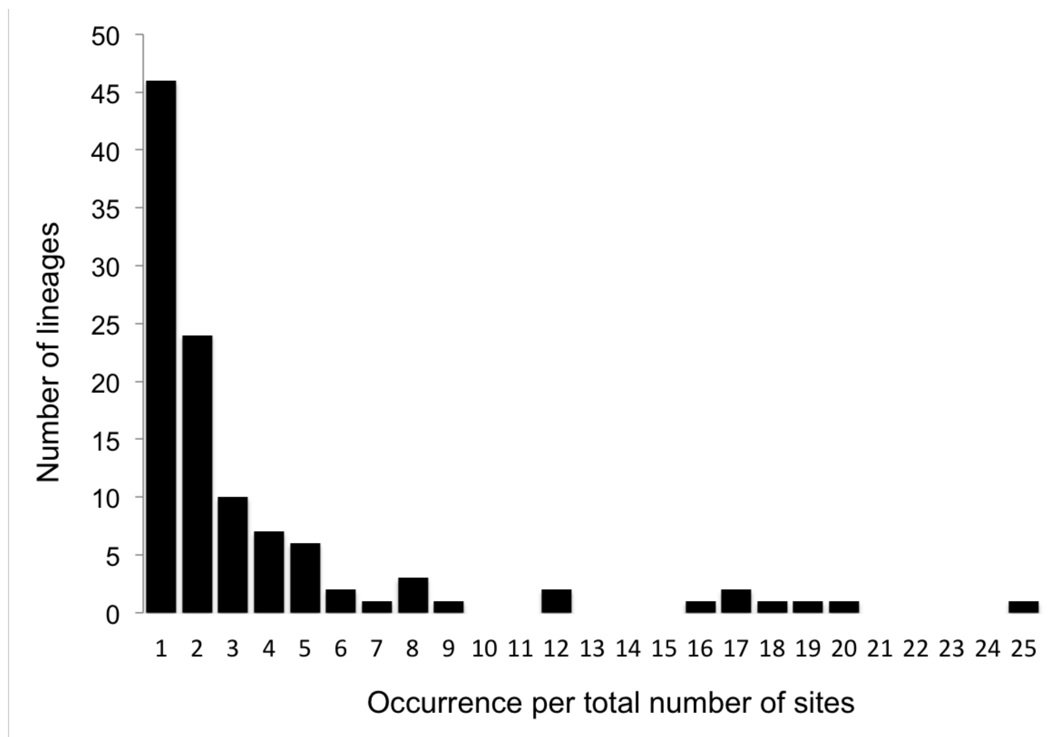


Figure 3.5. Right-skewed distribution of Collembola lineages across 25 sampling sites in the island of Tenerife. The majority of lineages occur in only a single or a few sites, with only a few lineages being comparatively widely distributed.

The lineage accumulation curve with increasing genetic distance reveals two main changes in the slope of the curve indicating changes in the rate at which lineages are being accumulated. The first change occurs between 2 and 4 % divergence when

lineages accumulate at a much slower pace compared to the previous divergence values (0 and 2%). Then, lineages start to accumulate increasingly faster up to 10% divergence, after which another change in slope shows a reduction in the number of lineages being accumulated (Fig 3.7). This curve also indicates Tenerife Collembola to be composed of two main groups, one group of very closely related sequences and another group of more distantly related sequences to Collembola sequences sampled from outside the archipelago.

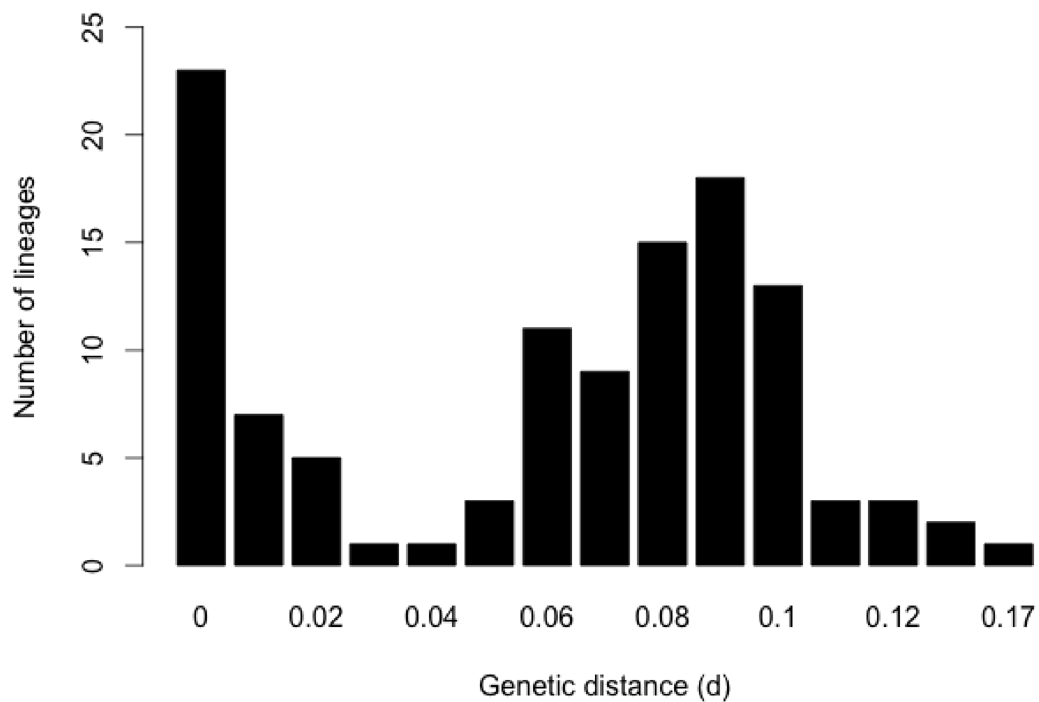


Figure 3.6. Bimodal distribution of pairwise genetic distances between Tenerife Collembola lineages and their closest matches outside the archipelago (BOLD database).

A comparison between figures 3.5 and 3.6 reveals that all widespread lineages (those that occur in more than 20 sites) are closely related to their mainland counterparts (similarity >99%). Whereas most of the unique/single site lineages are more diverged from their mainland counterparts than widespread lineages (63% of the unique site lineages are less than 92% similar to their mainland counterparts). However, at least 12% of these single site lineages are as closely related to their mainland counterparts as the widespread lineages are (being >99% similar to their mainland counterparts).

3.5 DISCUSSION

We investigated the colonisation history of the Collembola fauna of the Canary Islands by sampling Collembola DNA sequences from Tenerife, and comparing their relatedness to a database of sequences sampled from outside the Canary Islands. Results indicate the Collembola of Tenerife to be a mosaic of taxa that are genetically very closely related to non-Canary Island species, and taxa that are genetically very divergent from non-Canary Island species, with a scarcity of lineages of intermediate relatedness. The broad geographic distribution, and apparent genetic uniformity of many of non-Canary Island taxa that are genetically very close to Canary Island taxa reveals a probable origin by human introduction for some of these species.

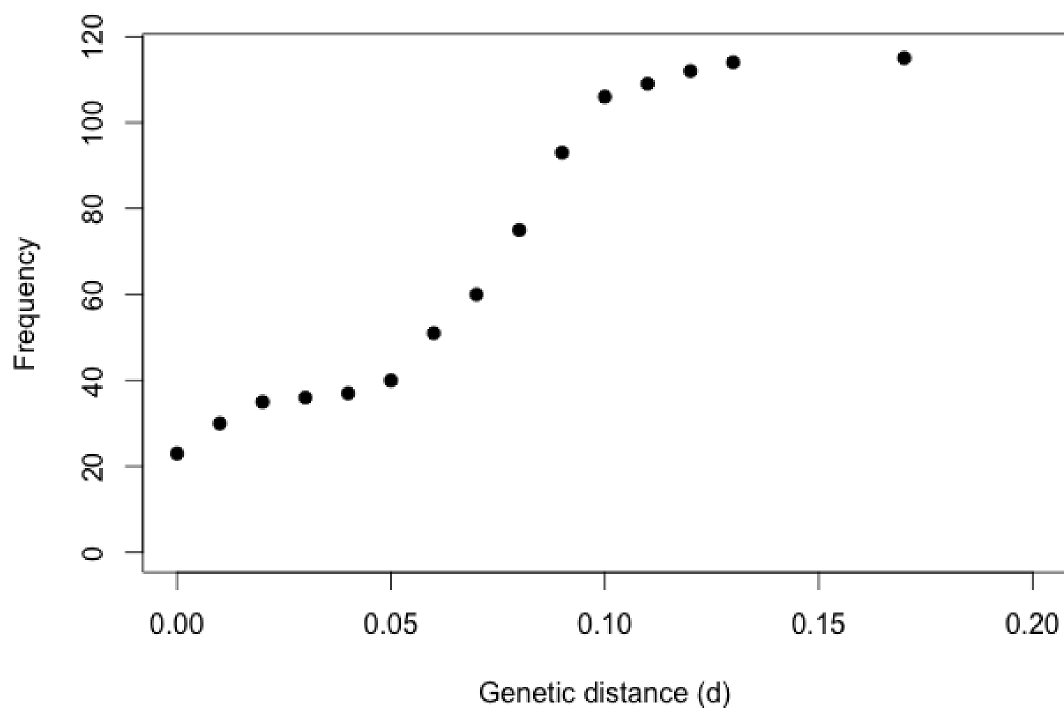


Figure 3.7. Cumulative frequency graph of Tenerife Colembolla lineages with increasing relative colonisation time (pairwise genetic distance, d, between Tenerife lineages and their closest matches outside the archipelago - BOLD database).

3.5.1 Short barcodes and diversity estimation

Short DNA barcode sequences from the PyroCleaned sequencing output alignment were delineated with a 95% similarity threshold criterion, as it represented an optimal threshold for the number of clusters produced. This similarity threshold implies a

minimum divergence limit of 5% to delineate lineages/OTUs, which is consistent with previous Collembola studies that applied coalescent tree models to identify lineages based on coalescent units - branches below which tree shape shifts from a species to a population pattern (Cicconardi et al., 2010, 2013; Zhang et al., 2014). For example, using the GMYC model (Pons et al., 2006) to define OTUs, Cicconardi et al. (2010, 2013) provided evidence that lineages as low as 5% divergence are consistent with biological species within *Lepidocyrtus* morphospecies sampled from the northwestern Mediterranean basin and from Panama. Similarly, Zhang et al. (2014) reported a minimum interspecific distances of 4% among species within the *Tomocerus nigrus* complex and only 2% among species within *Tomocerus ocreatus* complex, both sampled from China. These small minimum divergences contrast with those reported in other Collembola studies that typically applied higher divergence thresholds to define OTUs (e.g. 14% in Porco et al., 2014). However, these studies based their choices of threshold on the estimate of barcode gaps (Meyer & Paulay, 2005) between intraspecific and interspecific distances. More importantly, they did not consider coalescent units to define species, likely resulting in an underestimation of diversity as closely related species are lumped together.

After filtering the data for clusters that were not consistent with Collembola sequences and removing presumed numts, presumed Collembola clusters were assessed against the BOLD database and reference Sanger sequences to assign taxonomy. This facilitated the identification of some taxa to family, genus or species levels. The mean interlineage distance detected in this study (26.8%) is in accordance with those observed in prior studies on Collembola communities, e.g. mean distance of 28.7% among 97 lineages of the Churchill collembolan community (Porco et al., 2014), although we sampled a wider range of divergences (5-53% in Tenerife compared to 14-44% in Churchill).

It should be noted that the number of lineages found here are based on the best model provided by previous studies that indicate minimum divergence threshold as low as 5% to be consistent with biological species within Lepidocyrtus morphospecies sampled elsewhere (Cicconardi et al. 2010, 2013). Results from Chapter 2 demonstrate the difficulties to assign mtDNA lineages to species, for example, indicating that a single species can be formed by admixture of two divergent lineages of independent origin

(e.g. Laparocerus bimbache is originated from two founding mtDNA lineages that are 5.3% divergent). Although this fact could point to the possibility that some of the Collembola lineages could in fact represent a single biological species (indicating we have overestimated lineages diversity), the nature of Collembola data is quite distinct from the coleopteran data evaluated in Chapter 2 in terms of intraspecific and interspecific variation. For example, high levels of cryptic diversity have been documented within many collembolan species around the world, and low minimum interspecific distances (<5%) have been found among species, for example, of the Tomocerus nigrus complex (4%) and Tomocerus ocreatus complex (2%) (Zhang et al. 2014). Thus, these findings together with the fact that we were not able to assess cryptic diversity with our data, point to the possibility that our results could be an underestimation (more than overestimation) of diversity as we could be lumping together closely related species.

3.5.2 Lineages colonisation

The bimodal distribution of the pairwise distances between Tenerife lineages and their closest match outside the archipelago is consistent with two main cohorts of colonisation in the history of community assemblage in the island (Fig 3.6). Taking genetic distances as a proxy for colonisation time, it would seen clear that there is an older colonisation cohort, which seems to have been followed by a period of less frequent immigration events or decreased ability to establish. A more recent colonisation cohort indicates that niches were either not saturated, or perhaps became available after local extinctions. The cumulative lineage curve also shows the existence of two periods of increased lineage accumulation indicating two main cohorts of colonisation, an older and a more recent one (Fig 3.7). These figures differ considerably from the negative exponential decay predicted for the distribution of allelic divergence time across species on a single island under the speciation model (Johnson et al. 2000, Figs 3 & 6) and from the lineage accumulation curve with a single change of slope that fit the speciation-divergence model observed for land birds in the Lesser Antilles (Ricklefs and Bermingham 2004, Fig 1). While different distribution plots and cumulative curves would not be unexpected for Collembola, based on their much lower dispersal capability when compared to high dispersal taxa, such as birds, the effect of

this would might be expected to affect more the rate of decay or rate of accumulation (slopes) of the graphs, rather than their shapes. To understand the shape of these curves, we need to consider island biogeography theory, Collembola life history, and recent human impact.

With regard to the older distribution of Collembola divergences, two possibilities can be suggested to explain this. The first is that these taxa do not have a genetically close relative outside of the Canary Islands, due to them being the product of a relatively old colonisation event, as has been demonstrated for many insect groups, e.g. the tenebrionid genera *Hegeter* and *Pimelia* inhabiting Gran Canaria are estimated to result from a colonisation event occurred at least 8 and 9 Ma, respectively (Juan et al., 1995, 1996). The second potential explanation is the limited resolution of the BOLD database, such that some of these lineages may also occur outside the Canary Islands, but have not been sampled. While such a limitation of the BOLD database is real and unavoidable, if we take into account the range of relatedness among lineages of Collembola (e.g. Cicconardi et al. 2010, 2013), then lineages of intermediate relatedness would also be expected to have been sampled. We cannot exclude the possibility that some divergent lineages may have closely related lineages outside the Canary Islands. However, a bimodal distribution for the divergences among Collembola species has also been demonstrated in a broad taxonomic assessment of cryptic diversity within Collembola using publicly available sequence data (Emerson et al., 2011). The frequency plot of the pairwise genetic distances observed among 866 public sequences belonging to 105 species of Collembola for the mtCOI DNA gene, also revealed two peaks, one comprised of very distantly related species and another of very closely related species (Fig 4, Emerson et al., 2011), suggesting that a bimodal distribution is likely to be a feature (more than an artefact) for the frequency of genetic relatedness among Collembola species.

With regard to the distribution of relatively young lineages, both natural and non-natural introductions would seem to be the most plausible explanation. Springtails lack an effective way to avoid desiccation during dispersal but several studies have collectively showed that some species can survive both on and in seawater for long periods (e.g. more than 15 days, as reported in Moore, 2002), and survival could be extended by rafting or in floating debris (Stevens et al., 2006). Passive dispersal has been related to

natural introductions of Collembola in several regions, for example, in the Galápagos archipelago, transport of various terrestrial arthropods (including springtails) on the sea surface between islands has been recorded (Peck, 1994). Wind has also been linked to natural introductions for a set of taxa in the Southern Hemisphere (Greenslade et al., 1999; Muñoz et al., 2004). Therefore, similar natural passive dispersal could also account for Tenerife lineages that are closely related to sequences from near or relatively near sources. However, reliable estimates of dispersal and establishment for Collembola are lacking. In contrast, molecular data indicates that dispersal over timescales measured in millennia is very limited (e.g. Stevens et al., 2006; Garrick et al., 2007; Cicconardi et al., 2010, 2013)

Human mediated introductions are more likely to explain Tenerife lineages that are genetically identical or nearly identical to sequences sampled in remote continents. Passive dispersal of Collembola through human activity has been implicated in the sharing of lineages across distant continents, e.g. *Cerathophysella denticulata* lineage L3 found in Canada, South Africa, Australia and New Zealand (Porco et al., 2012a), and such introductions are considered to have mainly been accidental, e.g. in soil used as ship ballast (Vázquez & Simberloff, 2001). It has also been demonstrated that these unintentional introductions have resulted in similar genetic structure between populations of species occurring in different continents, e.g. European and North American populations of *Parasitoma notabilis, Neanura muscorum, Orchesella cincta* and *O. villos*a, suggesting massive and multiple introductions (Porco et al., 2013). Our data set presents similar examples of faunal exchange. For example, four lineages sampled on Tenerife have also been sampled in Canada, and another two Tenerife lineages have also been sampled in Australia. Genetic identity, or near identity across vast geographic distances is consistent with passive transport through human activities.

3.5.3 Lineage distribution

The distribution of lineages sampled in Tenerife with the majority of lineages restricted to a single site, and only a few lineages being widely distributed across the 25 sites in Tenerife (Fig 3.5), fits classic expectations for species-range-size distributions (Gaston, 1996). Within most natural taxonomic assemblages, the geographic ranges of species

generally present a right-skewed distribution, as the majority of species tend to be rare, and only a few species tend to be common (Gaston, 1996). The identity of the three most common lineages (which occur in > 20 out of the 25 sites sampled in Tenerife) reveals these to be presumed recently introduced species that belong to common widespread taxa, the genera *Entomobrya* and *Parasitoma* and the family Hypogasturidae. Their relatedness to sequences outside the archipelago indicate they could be the result of natural introduction but also human introduction, since the three of them are identical or nearly identical to sequences sampled both in Europe (France and Moldova) and in further continents (USA, Canada and Australia, respectively). This approach provides a minimum quantitative estimate of probable human introduction, i.e. we can categorise species being a probable or not human introduction. Also, from a conservation point of view, given that geographic range size and risk of extinction are often negatively correlated (Gaston, 1994), one of the consequences of this strong right skew to species-range-distribution is that a disproportionate number of species need to be screened if one has to identify species with a high probability of loss (Gaston, 1996).

3.5.4 Diversity comparison with previous Tenerife surveys

Since 2001, information on species diversity from published data, reports and taxonomists personal communications have been collected and made public as a list of species by the Biodiversity database of the Canary Islands (Arechavaleta et al., 2010). In its last edition from 2009, it reported the presence of 80 Collembola species in Tenerife among which 19 are endemic. Our dataset reveals the presence of 117 lineages as a result of sampling a limited volume of soil from 25 point localities within the island of Tenerife (Fig 3.1). Despite the fact that most sequences could not be assigned to taxonomy because they did not have any close match to sequences on the BOLD database, the taxonomic identity of the 29 clusters with high matches (≥99%) on BOLD, together with the identity of the Sanger sequences, added several taxa to the currently known diversity of Collembola in the island: three families (Dicyrtomidae, Katiannidae and Tullbergiidae), six genera (*Cryptopygus, Deuteraphorura*, *Friesea, Neanura, Micranurida, Paratullbergia*) and nine species *(Entomobrya atrocincta, Folsomia quadrioculata, Lepidocyrtus cyaneus and* L. curvicollis*, Sminthurinus niger* and *S. elegans*, *Friesea truncata*, *Protaphorura fimata* and *Cryptopygus tricuspis)* are newly

recorded for Tenerife based on this study. It also confirmed the presence of genera and species listed previously: *Brachystomella*, *Pseudosomita*, and *Folsomides, Ceratophysella gibbosa*, *Heteromurus nitidus*, *Parasitoma notabilis* (also *P. notabilis* L1 and L3), and *Folsomia candida*. Although we have not directly assessed cryptic speciation, neither have we been able to identify most of our clusters, we are reporting a remarkable increase in the number of OTUs for Tenerife compared to previous reports based solely on morphology. These results confirm the utility of short 454 mini barcodes for rapid assessment of the biodiversity of poorly known groups such as soil invertebrates and support previous suggestions that the species diversity in Collembola has been severely underestimated (Cicconardi et al. 2010).

3.5.5 Limitations and future directions

Technical restrictions of 454 next-generation sequencers (see below) and limitations of the BOLD database are potential limitations within this study. Our barcode sequences of 220 bp length retrieved a greatly reduced number of matches to Collembola sequences from the public database when compared to the long Sanger sequences, but they were still able to capture the main taxa identities, supporting the idea that a small portion of DNA barcode gene, a mini-barcode (Meusnier et al., 2008), can also provide species-level resolution (James et al., 2010). Furthermore, by using a conservative approach to accept clusters as presumed Collembola species, we were able to exclude probable non-Collembola sequences.

NGS technologies have revolutionised the field of DNA sequencing by enabling rapid and inexpensive sequencing of large amount of data. As demonstrated here, the 454 pyrosequencing and data handling with the PyroClean algorithm efficiently produced a large number of sequences facilitating the study of Collembola community diversity in Tenerife. However, a significant drawback of many NGS technologies is the short length of the reads produced and, specifically to 454 pyrosequencing, errors associated with homopolymers. Developments in computational tools are helping to deal with these types of errors, e.g. the PyroClean algorithm applied here, and NGS technology is advancing at an unprecedented speed (with focus in increasing read-lengths). One potentially promising direction is the bulk de novo mitogenome assembly from pooled

total DNA, which has been demonstrated to efficiently and economically produce a large number of complete or near complete mitogenome DNA sequences from numerous samples, without the need of PCR amplification or any enrichment (Gillett et al., 2014). As well as producing long sequences, this approach also removes PCR bias, which should improve the detection of species within a pool of DNA samples. Obtaining longer barcode sequences would result in better matches to BOLD sequences, thus enhancing taxonomic assignment, while the ability to obtain whole or partial mtDNA genomes would also permit broader taxonomic assessments with other gene regions stored within other public databases such as GenBank.

3.6 REFERENCES

Arechavaleta M., Rodríquez S., Zurita N., & García A. (2010) *Lista de especies silvestres de Canarias*. *Hongos, plantas y animales terrestres*. *2009*. Santa Cruz de Tenerife: Gobierno de Canarias, Santa Cruz de Tenerife, Tenerife.

Baselga A., Fujisawa T., Crampton-Platt A., Bergsten J., Foster P.G., Monaghan M.T., & Vogler A.P. (2013) Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nature Communications*, **4**, 1892.

Cavender-Bares J., Kozak K.H., Fine P.V.A., & Kembel S.W. (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters*, **12**, 693–715.

Chu H., Fierer N., Lauber C.L., Caporaso J.G., Knight R., & Grogan P. (2010) Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environmental Microbiology*, **12**, 2998–3006.

Cicconardi F., Fanciulli P.P., & Emerson B.C. (2013) Collembola, the biological species concept and the underestimation of global species richness. *Molecular Ecology*, **22**, 5382–5396.

Cicconardi F., Nardi F., Emerson B.C., Frati F., & Fanciulli P.P. (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology*, **19**, 386–400.

Emerson B.C., Cicconardi F., Fanciulli P.P., & Shaw P.J.A. (2011) Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2391–2402.

Emerson B.C. & Gillespie R.G. (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology & Evolution*, **23**, 619–630.

Fonseca V.G., Carvalho G.R., Sung W., Johnson H.F., Power D.M., Neill S.P., Packer M., Blaxter M.L., Lambshead P.J.D., Thomas W.K., & Creer S. (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications*, **1**, 98.

Garrick R.C., Sands C.J., Rowell D.M., Hillis D.M., & Sunnucks P. (2007) Catchments catch all: long-term population history of a giant springtail from the southeast

Australian highlands - a multigene approach. *Molecular Ecology*, **16**, 1865–1882.

Gaston K.J. (1994) What is rarity? *Rarity* pp. 1–21. Springer Netherlands, Chapman & Hall, London.

Gaston K.J. (1996) Species-range-size distributions: patterns, mechanisms and implications. *Trends in Ecology & Evolution*, **11**, 197–201.

Ghodsi M., Liu B., & Pop M. (2011) DNACLUST: accurate and efficient clustering of phylogenetic marker genes. *BMC Bioinformatics*, **12**, 271.

Gillett C.P.D.T., Crampton-Platt A., Timmermans M.J.T.N., Jordal B.H., Emerson B.C., & Vogler A.P. (2014) Bulk De Novo Mitogenome Assembly from Pooled Total DNA Elucidates the Phylogeny of Weevils (Coleoptera: Curculionoidea). *Molecular Biology and Evolution*, **31**, 2223–2237.

Greenslade P., Farrow R.A., & Smith J.M.B. (1999) Long distance migration of insects to a subantarctic island. *Journal of Biogeography*, **26**, 1161–1167.

Hengeveld R. (2002) MacArthur, R.H. and E.O. Wilson (1967, reprinted 2001). The Theory of Island Biogeography. *Acta Biotheoretica*, **50**, 133–136.

HilleRisLambers J., Adler P.B., Harpole W.S., Levine J.M., & Mayfield M.M. (2012) Rethinking Community Assembly through the Lens of Coexistence Theory. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 227–248.

Hopkin S.P. (1997) *Biology of the Springtails*. Oxford University Press, Oxford, UK.

James S.W., Porco D., Decaëns T., Richard B., Rougerie R., & Erséus C. (2010) DNA barcoding reveals cryptic diversity in *Lumbricus terrestris* L., 1758 (Clitellata): resurrection of L. herculeus (Savigny, 1826). *PloS One*, **5**, e15629.

Johnson K.P., Adler F.R., & Cherry J.L. (2000) Genetic and Phylogenetic Consequences of Island Biogeography. *Evolution*, **54**, 387–396.

Juan C., Oromi P., & Hewitt G.M. (1995) Mitochondrial DNA Phylogeny and Sequential Colonization of Canary Islands by Darkling Beetles of the Genus *Pimelia* (Tenebrionidae). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **261**, 173–180.

Juan C., Oromi P., & Hewitt G.M. (1996) Phylogeny of the genus *Hegeter* (Tenebrionidae, Coleoptera) and its colonization of the Canary Islands deduced from Cytochrome Oxidase I mitochondrial DNA sequences. *Heredity*, **76 (Pt 4)**, 392–403.

Liu S., Li Y., Lu J., Su X., Tang M., Zhang R., Zhou L., Zhou C., Yang Q., Ji Y., Yu D.W., & Zhou X. (2013) SOAPBarcode: revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons. *Methods in Ecology and Evolution*, **4**, 1142–1150.

Lomolino M.V., Brown J.H., & Sax D.F. (2010) Reticulations and Reintegration of "a Biogeography of the Species." *The Theory of Island Biogeography Revisited* pp. 13–51. Princeton University Press, New Jersey.

Losos J.B. & Schluter D. (2000) Analysis of an evolutionary species–area relationship. *Nature*, **408**, 847–850.

Medinger R., Nolte V., Pandey R.V., Jost S., Ottenwälder B., Schlötterer C., & Boenigk J. (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Molecular ecology*, **19**, 32–40.

Meusnier I., Singer G.A., Landry J.-F., Hickey D.A., Hebert P.D., & Hajibabaei M. (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**, 214.

Meyer C.P. & Paulay G. (2005) DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biol*, **3**, e422.

Moore P.D. (2002) Biogeography: Springboards for springtails. *Nature*, **418**, 381–381.

Muñoz J., Felicísimo A.M., Cabezas F., Burgaz A.R., & Martínez I. (2004) Wind as a long-distance dispersal vehicle in the Southern Hemisphere. *Science*, **304**, 1144–1147.

Papadopoulou A., Anastasiou I., Spagopoulou F., Stalimerou M., Terzopoulou S., Legakis A., & Vogler A.P. (2011) Testing the Species–Genetic Diversity Correlation in the Aegean Archipelago: Toward a Haplotype-Based Macroecology? *The American Naturalist*, **178**, 241–255.

Parent C.E. & Crespi B.J. (2006) Sequential Colonization and Diversification of Galápagos Endemic Land Snail Genus *Bulimulus* (Gastropoda, Stylommatophora). *Evolution*, **60**, 2311–2328.

Peck S.B. (1994) Sea-Surface (Pleuston) Transport of Insects between Islands in the Galápagos Archipelago, Ecuador. *Annals of the Entomological Society of America*, **87**, 576–582.

Pons J., Barraclough T., Gomez-Zurita J., Cardoso A., Duran D., Hazell S., Kamoun S., Sumlin W., & Vogler A. (2006) Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Systematic Biology*, **55**, 595–609.

Porco D., Bedos A., Greenslade P., Janion C., Skarżyński D., Stevens M.I., Jansen van Vuuren B., & Deharveng L. (2012a) Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. *Invertebrate Systematics*, **26**, 470-477.

Porco D., Decaëns T., Deharveng L., James S.W., Skarżyński D., Erséus C., Butt K.R., Richard B., & Hebert P.D.N. (2013) Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. *Biological Invasions*, **15**, 899–910.

Porco D., Potapov M., Bedos A., Busmachiu G., Weiner W.M., Hamra-Kroua S., & Deharveng L. (2012b) Cryptic Diversity in the Ubiquist Species Parisotoma notabilis (Collembola, Isotomidae): A Long-Used Chimeric Species? *PLoS ONE*, **7**, e46056.

Porco D., Skarżyński D., Decaëns T., Hebert P.D.N., & Deharveng L. (2014) Barcoding the Collembola of Churchill: a molecular taxonomic reassessment of species diversity in a sub-Arctic area. *Molecular Ecology Resources*, **14**, 249–261.

Ramirez-Gonzalez R., Yu D.W., Bruce C., Heavens D., Caccamo M., & Emerson B.C. (2013) PyroClean: Denoising Pyrosequences from Protein-Coding Amplicons for the Recovery of Interspecific and Intraspecific Genetic Variation. *PLoS ONE*, **8**, e57615.

Ricklefs R.E. (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, **7**, 1–15.

Ricklefs R.E. & Bermingham E. (2001) Nonequilibrium diversity dynamics of the Lesser Antillean avifauna. *Science*, **294**, 1522–1524.

Ricklefs R.E. & Bermingham E. (2004) Application of Johnson et al.'s speciation threshold model to apparent colonization times of island biotas. *Evolution*, **58**, 1664–1673.

Rousk J., Bååth E., Brookes P.C., Lauber C.L., Lozupone C., Caporaso J.G., Knight R., & Fierer N. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME Journal*, **4**, 1340–1351.

Schneider C., Cruaud C., & D'Haese C.A. (2011) Unexpected diversity in *Neelipleona* revealed by molecular phylogeny approach (Hexapoda, Collembola). *Soil Organisms*, **83**, 383–398.
Stevens M.I., Greenslade P., Hogg I.D., & Sunnucks P. (2006) Southern Hemisphere Springtails: Could Any Have Survived Glaciation of Antarctica? *Molecular Biology and Evolution*, **23**, 874–882.
Tamura K., Stecher G., Peterson D., Filipski A., & Kumar S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.
Torricelli G., Carapelli A., Convey P., Nardi F., Boore J.L., & Frati F. (2010) High divergence across the whole mitochondrial genome in the "pan-Antarctic" springtail *Friesea grisea*: Evidence for cryptic species? *Gene*, **449**, 30–40.
Vamosi S.M., Heard S.B., Vamosi J.C., & Webb C.O. (2009) Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology*, **18**, 572–592.
Vázquez D.P. & Simberloff D. (2001) Taxonomic Selectivity in Surviving Introduced Insects in the United States. *Biotic Homogenization* (ed. by J.L. Lockwood and M.L. McKinney), pp. 103–124. Kluwer Academic/Plenum Publishers, New York.
Vellend M., Lajoie G., Bourret A., Múrria C., Kembel S.W., & Garant D. (2014) Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Molecular Ecology*, **23**, 2890–2901.
Zhang F., Yu D., Luo Y., Ho S.Y.W., Wang B., & Zhu C. (2014) Cryptic diversity, diversification and vicariance in two species complexes of *Tomocerus* (Collembola, Tomoceridae) from China. *Zoologica Scripta*, **43**, 393–404.

3.7 Appendices

Table S3.1 - PyroClean results for mtDNA COI amplicons generated from community samples of Collembola from 25 sites on the island of Tenerife. Each site has been amplified in triplicate (indicated by the MID code). Total count of raw reads and read

counts after steps 3, 4 and 5 (described in the text) of the PyroClean process are presented.

Sites	Code	Raw reads	step 3	step 4	step 5
	COI01.MID1	7206	4059	144	25
1	COI01.MID2	6744	3769	130	24
	COI01.MID3	6829	3855	140	35
	COI01.MID4	5524	3155	64	33
2	COI01.MID5	6629	3930	83	37
	COI01.MID6	5359	3220	78	34
	COI01.MID7	7942	4743	187	22
3	COI01.MID8	8122	4963	170	20
	COI01.MID9	6736	4068	139	19
	COI01.MID10	5505	3357	125	26
4	COI01.MID11	8588	5234	211	31
	COI01.MID12	8522	5474	200	30
	COI01.MID13	11260	6642	211	32
5	COI01.MID14	10020	5988	208	36
	COI01.MID15	10847	6441	200	30
	COI02.MID1	5228	5176	121	36
6	COI02.MID2	5049	4976	111	33
	COI02.MID3	4773	4710	125	35
	COI02.MID4	5759	5700	114	26
7	COI02.MID5	4937	4880	115	28
	COI02.MID6	4777	4711	112	25
	COI02.MID7	5781	5720	150	30
8	COI02.MID8	4354	4325	125	30
	COI02.MID9	7415	7360	169	35
	COI02.MID10	852	837	43	19
9	COI02.MID11	5040	4986	129	36
	COI02.MID12	7990	7898	175	38
	COI02.MID13	8822	8721	243	56
10	COI02.MID14	6932	6864	210	54
	COI02.MID15	9799	9685	242	57
	COI02.MID16	7171	7023	124	53
11	COI02.MID17	5553	5456	115	42
	COI02.MID18	7460	7316	140	49

Tab S3.1. (cont)

Sites	Code	Raw reads	step 3	step 4	step 5
12	COI02.MID19	6816	6758	220	48

	COI02.MID20	7776	7711	220	54
	COI02.MID21	4767	4743	184	48
	COI02.MID22	1923	1906	99	28
13	COI02.MID23	3566	3517	137	36
	COI02.MID24	3280	3243	134	39
	COI02.MID25	14931	14782	229	52
14	COI02.MID26	7486	7422	197	49
	COI02.MID27	7405	7325	169	47
	COI02.MID28	6813	6730	165	32
15	COI02.MID29	5129	5087	112	32
	COI02.MID30	7704	7628	175	32
	COI03.MID1	4564	4538	129	45
16	COI03.MID2	5589	5530	168	42
	COI03.MID3	5032	5002	159	38
	COI03.MID4	7844	7815	241	50
17	COI03.MID5	6274	6251	233	52
	COI03.MID6	5148	4895	189	47
	COI03.MID7	5539	5509	195	44
18	COI03.MID8	6214	6185	213	43
	COI03.MID9	7465	7428	268	58
	COI03.MID10	1322	1322	79	31
19	COI03.MID11	4907	4891	191	63
	COI03.MID12	10769	10738	270	71
	COI03.MID13	10713	10671	278	73
20	COI03.MID14	6988	6966	267	56
	COI03.MID15	12281	12243	289	69
	COI03.MID16	4033	4011	149	37
21	COI03.MID17	4431	4418	172	43
	COI03.MID18	3614	3596	156	44
	COI03.MID19	7153	7129	229	58
22	COI03.MID20	9284	9239	238	55
	COI03.MID21	5734	5725	204	47
	COI03.MID22	8678	8628	199	50
23	COI03.MID23	7832	7805	196	49
	COI03.MID24	8709	8672	214	54
	COI03.MID25	10999	10937	178	52
24	COI03.MID26	4990	4961	119	33
	COI03.MID27	8406	8356	146	43
	COI03.MID28	8356	8295	275	76
25	COI03.MID29	4067	4053	173	70
	COI03.MID30	5771	5717	231	82

Table S3.2 - Taxonomy of Collembola Tenerife Sanger sequences (n=28, length=220 bp) submitted to the BOLD Identification System (http://www.boldsystems.org). Col counts = number of matches to a collembolan sequence within the top 99 matches returned by BOLD.

		BOLD taxonomy				
Sanger seq	Col counts	Order	Family	Genus	Species	Similarity hit (%)
Col023	15	Entomobryomorpha	Entomobryidae	Lepidocyrtus	*cyaneus*	100
Col024	1	Symphypleona	Katiannidae	Sminthurinus	*elegans*	100
Col025	6	Poduromorpha	Hypogastruridae			100
Col026	99	Entomobryomorpha	Isotomidae	Parisotoma	*notabilis L1*	100
Col027	3	Poduromorpha	Onychiuridae	Deuteraphorura	*sp*.	100
Col034	94	Entomobryomorpha	Entomobryidae	Entomobrya	*atrocincta*	100
Col210	3	Poduromorpha	Onychiuridae			94.83
Col211	13	Poduromorpha	Neanuridae	Friesea		89.37
Col212	83	Entomobryomorpha	Isotomidae			90.29
Col213	0					na
Col215	32	Poduromorpha	Hypogastruridae			91.58
Col216	58	Entomobryomorpha	Isotomidae			99.07
Col217	35	Entomobryomorpha	Entomobryidae	Lepidocyrtus	*cyaneus*	100
Col218	71	Collembola	Poduromorpha	Onychiuridae	Protaphorura	99.54
Col219	5	Poduromorpha	Onychiuridae	Orthonychiurus		98.1
Col221	11	Poduromorpha	Onychiuridae	Protaphorura		100
Col226	0					na
Col228	8	Entomobryomorpha				92.24
Col229	34	Entomobryomorpha	Entomobryidae			90.83

Tab S3.2. (cont)

Sanger seq	Col counts	BOLD taxonomy				
		Order	Family	Genus	Species	Similarity hit (%)
Col230	4	Symphypleona	Bourletiellidae			89.58
Col232	42	Entomobryomorpha	Isotomidae			92.59
Col234	91	Entomobryomorpha	Entomobryidae			100
Col235	92	Entomobryomorpha	Entomobryidae			100
Col236	38	Collembola				92.73
Col237	38	Collembola				92.73
Col238	53	Collembola				92.38

Table S3.3 - Taxonomy assignment of Collembola lineages (n=129) obtained after clustering the PyroCleaned output alignment of the Tenerife 454 sequencing data at 95% similarity threshold (see text for details). Sequences were submitted to the BOLD Identification System (http://www.boldsystems.org), and both taxonomy and similarity hit for the first sequence match within the top 99 matches returned by BOLD are shown. Taxonomy is accepted only for clusters with similarity hit ≥ 99%, highlighted in bold. N seq = number of unique sequences per cluster. Col counts = number of matches to a collembolan sequence within the top 99 matches.

			BOLD taxonomy				
Cluster	N seq	Col counts	Order	Family	Genus	Species	Similarity hit (%)
2	21	4	**Poduromorpha**	**Tullbergiidae**			**99.04**
4	52	30	Entomobryomorpha	Tomoceridae	*Tomocerus*		90.36
6	82	20	Entomobryomorpha	Isotomidae			91.18
9	90	11	**Entomobryomorpha**	**Entomobryidae**	***Lepidocyrtus***	***cyaneus***	**100**
11	3	10	**Poduromorpha**	**Hypogastruridae**	***Ceratophysella***		**100**
12	64	21	**Poduromorpha**	**Protaphorura**			**99.05**
14	17	18	Symphypleona	Sminthurididae			97.88
15	7	3	**Poduromorpha**	**Onychiuridae**	***Deuteraphorura***	***sp.***	**100**
16	44	1	**Symphypleona**	**Katiannidae**	***Sminthurinus***	***elegans***	**99.03**
17	8	6	Poduromorpha	Neanuridae	*Neanura*	*muscorum*	89.53
18	2	9	Entomobryomorpha	Isotomidae	*Cryptopygus*	*tricuspis*	92.5
19	3	11	**(Collembola)**				**100**
20	1	78	(Collembola)				92.62
21	7	3	**Poduromorpha**	**Tullbergiidae**			**100**
22	7	2	**Symphypleona**	**Katiannidae**	***Sminthurinus***	***elegans***	**99.5**
23	14	32	Poduromorpha	Neanuridae			91.38
25	3	14	Poduromorpha	Neanuridae	*Friesea*		87.75
26	7	17	Entomobryomorpha				89.74

Table S3.3. (cont)

Cluster	N seq	Col counts	BOLD taxonomy				
			Order	Family	Genus	Species	Similarity hit (%)
28	44	24	(Collembola)				89.44
29	3	15	Entomobryomorpha	Entomobryidae	*Entomobrya*	marginata	91.3
30	10	24	Entomobryomorpha	Entomobryidae			91.57
31	9	44	Entomobryomorpha	Entomobryidae			90.36
32	260	93	**Entomobryomorpha**	**Entomobryidae**	***Entomobrya***	***atrocincta***	**100**
34	34	5	Poduromorpha	Onychiuridae	*Orthonychiurus*		98.09
35	5	8	Neelipleona	Neelidae		*sp. DPCOL68266*	98.11
37	55	23	Poduromorpha	Hypogastruridae	*Hypogastrura*		94.44
38	204	99	**Entomobryomorpha**	**Isotomidae**	***Parisotoma***	***notabilis L1***	**100**
39	151	99	**Entomobryomorpha**	**Isotomidae**	***Parasitoma***	**notabilis**	**100**
40	15	19	Entomobryomorpha	Entomobryidae	*Entomobrya*	*gisini*	91.67
41	130	40	Entomobryomorpha	Oncopoduridae			90.98
42	4	35	Symphypleona	Dicyrtomidae			92.65
43	11	62	**Entomobryomorpha**	**Isotomidae**	***Folsomia***	***quadrioculata/ penicula***	**100**
45	140	6	**Poduromorpha**	**Hypogastruridae**			**100**
48	13	11	Poduromorpha	Neanuridae	*Neanura*	*muscorum*	94.19
50	6	12	Entomobryomorpha				83.33
51	15	36	**Entomobryomorpha**	**Isotomidae**	***Folsomia***	***candida***	**99.52**
53	3	27	Entomobryomorpha				91.18
54	6	8	Poduromorpha	Onychiuridae	*Orthonychiurus*		94.2
56	4	3	**Poduromorpha**	**Neanuridae**	***Micranurida***		**100**
57	3	23	Entomobryomorpha	Isotomidae			92.11
59	10	8	Entomobryomorpha	Entomobryidae	*Entomobrya*		88.17

Table S3.3. (cont)

Cluster	N seq	Col counts	BOLD taxonomy				
			Order	Family	Genus	Species	Similarity hit (%)
60	29	9	Entomobryomorpha	Entomobryidae		*sp. DPCOL12495*	92.22
61	11	42	(Collembola)				91.06
63	43	83	Entomobryomorpha	Isotomidae			92.27
65	2	50	Entomobryomorpha				90.36
66	2	17	**Poduromorpha**	**Neanuridae**			**99.52**
68	1	28	Entomobryomorpha	Entomobryidae			92.25
70	9	13	**Poduromorpha**	**Hypogastruridae**			**100**
71	5	26	Entomobryomorpha				90.4
73	9	99	(Collembola)				94.66
74	25	50	(Collembola)				87.5
75	5	60	Poduromorpha	Neanuridae	*Neanura*	*muscorum*	97.67
76	13	99	Poduromorpha	Neanuridae	*Neanura*	*muscorum*	95.38
77	12	90	Poduromorpha	Neanuridae	*Neanura*	*muscorum*	94.83
78	36	68	**Poduromorpha**	**Neanuridae**	***Neanura***	***muscorum***	**99.52**
79	6	69	Poduromorpha	Neanuridae	*Neanura*	muscorum	94.39
80	61	18	Entomobryomorpha	Entomobryidae	*Entomobrya*		94.17
81	4	18	Entomobryomorpha	Entomobryidae			95.18
82	49	94	**Poduromorpha**	**Neanuridae**	***Neanura***	***muscorum***	**100**
83	34	11	**Entomobryomorpha**	**Entomobryidae**			**100**
84	9	99	Poduromorpha	Neanuridae			96.77
85	7	96	Poduromorpha	Neanuridae			95.16
86	51	64	Poduromorpha	Neanuridae	*Neanura*	*muscorum*	96.55
87	23	91	Entomobryomorpha	Entomobryidae			90.36
88	9	99	Poduromorpha	Neanuridae	*Neanura*	*muscorum*	96.55

Table S3.3. (cont)

Cluster	N seq	Col counts	BOLD taxonomy				
			Order	Family	Genus	Species	Similarity hit (%)
89	6	19	Entomobryomorpha	Isotomidae	*Parisotoma*	notabilis	98.49
90	3	7	**Entomobryomorpha**	**Isotomidae**			**100**
91	25	22	**Poduromorpha**	**Tullbergiidae**	***Paratullbergia***		**100**
92	6	61	Entomobryomorpha	Tomoceridae	*Tomocerus*		90.36
93	1	21	Entomobryomorpha	Isotomidae	*Parisotoma*	*notabilis*	95.5
94	2	86	(Collembola)				90.15
95	1	93	Entomobryomorpha	Isotomidae	*Parisotoma*	*notabilis L1*	93.72
96	13	18	Entomobryomorpha	Tomoceridae	*Tomocerus*		93.98
97	40	7	Entomobryomorpha	Isotomidae			92.27
98	3	23	Entomobryomorpha	Isotomidae			91.18
99	4	13	Entomobryomorpha	Isotomidae			92.65
100	7	91	Entomobryomorpha				91.97
102	1	12	**Poduromorpha**	**Brachystomellidae**	***Brachystomella***		**99.45**
103	12	17	(Collembola)				89.74
104	5	8	Poduromorpha	Onychiuridae			93.48
107	11	9	Poduromorpha	Onychiuridae			91.3
110	2	7	Entomobryomorpha	Isotomidae			91.23
111	11	39	Entomobryomorpha	Isotomidae			91.78
112	19	8	Entomobryomorpha	Entomobryidae	*Entomobrya*	*nivalis*	88.89
113	1	70	(Collembola)				97.57
114	7	18	**Neelipleona**	**Neelidae**			**100**
116	21	12	Poduromorpha	Onychiuridae			91.3
118	11	34	Entomobryomorpha	Entomobryidae			91.54
119	20	73	(Collembola)				93.94

Table S3.3. (cont)

Cluster	N seq	Col counts	BOLD taxonomy				
			Order	Family	Genus	Species	Similarity hit (%)
121	7	70	Entomobryomorpha	Entomobryidae			90.65
123	13	7	Poduromorpha	Hypogastruridae	*Xenylla*	*humicola*	92.39
124	6	56	Poduromorpha	Neanuridae	*Neanura*	muscorum	96.55
127	17	34	**Poduromorpha**	**Onychiuridae**	***Protaphorura***	***fimata***	**100**
128	6	10	**Poduromorpha**	**Tullbergiidae**			**100**
130	16	6	Poduromorpha	Hypogastruridae	*Xenylla*	*humicola*	92.05
131	6	11	Entomobryomorpha	Isotomidae	*Isotoma*	marionensis	92.5
132	6	7	Entomobryomorpha	Isotomidae	*Isotoma*		88.62
134	1	82	Entomobryomorpha	Entomobryidae			98.36
137	8	59	Entomobryomorpha	Bombinatoridae	*Bombina*	*maxima*	90
138	1	21	Entomobryomorpha	Tomoceridae	*Pogonognathellus*	*flavescens*	92.5
139	3	79	Collembola				91.82
140	3	98	Entomobryomorpha	Isotomidae	*Cryptopygus*	*caecus*	90
141	7	12	Entomobryomorpha				92.31
142	1	7	Collembola				97.44
143	7	56	Entomobryomorpha	Entomobryidae	*Entomobrya*	*nivalis*	91.23
145	1	7	Poduromorpha	Neanuridae	*Neanura*	*muscorum*	87.37
147	1	12	Entomobryomorpha	Entomobryidae			*94.33*
151	12	36	**Poduromorpha**	**Hypogastruridae**			**100**
152	5	10	Poduromorpha	Hypogastruridae			93.81
153	4	27	**Entomobryomorpha**	**Isotomidae**			**99.02**
154	6	72	Poduromorpha	Neanuridae			97.79
155	4	68	Poduromorpha	Neanuridae	*Neanura*	muscorum	97.67
156	3	26	Entomobryomorpha	Isotomidae	*Desoria*	*tshernovi*	92.54

Table S3.3. (cont)

Cluster	N seq	Col counts	BOLD taxonomy				
			Order	Family	Genus	Species	Similarity hit (%)
159	4	33	Entomobryomorpha	Isotomidae	*Cryptopygus*	tricuspis	91.67
160	1	94	**Poduromorpha**	**Neanuridae**	**Neanura**	***muscorum***	**100**
161	3	97	Poduromorpha	Neanuridae	*Neanura*	*muscorum*	97.08
162	1	83	Poduromorpha	Neanuridae	*Neanura*	*muscorum*	98.54
164	1	8	Entomobryomorpha	Entomobryidae	*Entomobrya*	*nivalis*	91.23
166	9	31	Entomobryomorpha				90.7
168	14	32	Entomobryomorpha	Entomobryidae			93.98
169	1	99	Isotomidae	Parisotoma	*notabilis*		97.7
171	15	14	Entomobryomorpha	Isotomidae			89.15
172	7	36	Entomobryomorpha	Isotomidae	*Folsomia*	*quadrioculata*	91.5
174	6	18	Entomobryomorpha	Isotomidae			92.86
175	3	94	Poduromorpha	Hypogastruridae	*Ceratophysella*		98.57
176	7	31	**Entomobryomorpha**	**Isotomidae**			**99.05**
177	1	33	Collembola				87.88
178	2	9	Poduromorpha	Onychiuridae		*sp. DPCOL95864*	94.38
179	1	18	**Entomobryomorpha**	**Isotomidae**	***Cryptopygus***	***tricuspis***	**100**

Table S3.4 - Taxonomy assignment of non-Collembola lineages (n=53) obtained after clustering the PyroCleaned output alignment of the Tenerife 454 sequencing data at 95% similarity threshold (see text for details). Sequences were submitted to the BOLD Identification System (http://www.boldsystems.org), and both taxonomy and similarity hit for the first sequence match within the top 99 matches returned by BOLD are shown. N seq = number of unique sequences per cluster. Col counts = number of matches to a collembolan sequence within the top 99 matches. Table S3.5 - Geographic distribution of the most related sequence (in BOLD database) to the 29 Collembola lineages sampled in Tenerife.

			BOLD taxonomy				
Cluster	n seq	Col counts	Order	Family	Genus	Species	Similarity hit (%)
1	10	3	Diptera	Xyloryctidae	*Scieropepla*	*sp. ANIC1*	92.22
3	4	1	Lepidoptera	Lycaenidae	*Phengaris*	teleius	97.47
5	56	5	Lepidoptera	Noctuidae	*Agrotis*	*infusa*	96.3
7	8	2	Hemiptera	Coreidae	*Gralliclava*	*horrens*	93.83
8	3	4	Diptera	Agromyzidae			93.69
10	4	1	Lepidoptera	Oecophoridae	*Sympoecila*	callisceptra	92.86
13	1	0	Actinopterygii	Perciformes	*Apogonidae*		90.48
24	5	5	Poduromorpha	Hypogastruridae	*Hypogastrura*		92.22
27	8	2	Diptera				93.59
33	31	0	Lepidoptera	Nymphalidae	*Napeogenes*	*sylphis rindgei*	95.56
36	15	0	Lepidoptera	Hepialidae	*Phassus*		97.53
44	48	0	Diplostraca	Daphniidae	*Ceriodaphnia*		92.31
46	5	3	Lepidoptera	Geometridae	*Epyaxa*	*subidaria*	94.87
47	11	3	Lepidoptera	Oecophoridae	*Palimmeces*	*sp. ANIC19*	93.33
49	15	0	Lepidoptera	Nymphalidae	*Oleria*	*onega janarilla*	88.89
52	7	2	Coleoptera	Chrysomelidae	*Cassida*	*murraea*	92.31
55	7	5	Lepidoptera	Noctuidae			91.03

Tab S3.4. (cont)

Cluster	n seq	Col counts	BOLD taxonomy				
			Order	Family	Genus	Species	Similarity hit (%)
58	9	0	Lepidoptera	Crambidae	*Patissa*	*tinctalis*	94.25
62	3	0	Lepidoptera	Hepialidae	*Phassus*		97.53
64	3	1	Lepidoptera	Lycaenidae	*Phengaris*	teleius	97.47
67	3	1	Lepidoptera	Oecophoridae	*Sympoecila*	callisceptra	92.86
69	15	3	Passeriformes	Prunellidae	*Prunella*	*modularis*	92.31
72	31	0	Diptera	Culicidae	*Anopheles*	*crucians*	97.7
101	12	5	Poduromorpha	Hypogastruridae	*Ceratophysella*	*gibbosa*	95.24
105	23	1	Coleoptera	Scarabaeidae	*Brachysternus*		91.18
106	22	0	Lepidoptera	Choreutidae	*Tebenna*	*balsamorrhizella*	92.22
108	3	5	Plecoptera	Perlodidae	*Isoperla*		92.86
109	4	2	Diptera				91.67
115	28	3	Lepidoptera	Hesperiidae	*Polytremis*	*pellucida*	93.83
117	46	4	Amphipoda	Acanthogammaridae	*Eulimnogammarus*	*vittatus*	90.72
120	11	0	Diptera	Culicidae	*Aedes*	*albopictus*	96.55
122	10	4	Lepidoptera	Ampeliscidae			94.12
125	22	2	Lepidoptera	Noctuidae			93.59
126	32	0	Lepidoptera	Nolidae	*Meganola*	*albula*	97.62
129	8	0	Euonychophora	Peripatidae	*Epiperipatus*		92.31
133	4	3	Amphipoda	Ischyroceridae			92.16
135	1	3	Lepidoptera	Geometridae	*Epyaxa*	*subidaria*	93.59
136	3	2	Diptera	Stratiomyidae			92.59
144	3	3	Lepidoptera	Tortricidae	*Ancylis*		90.12
146	3	2	Trichoptera	Uenoidae	*Neophylax*	*splendens*	93.59

Tab S3.4. (cont)

Cluster	n seq	Col counts	BOLD taxonomy				
			Order	Family	Genus	Species	Similarity hit (%)
148	2	5	Poduromorpha	Onychiuridae	*Orthonychiurus*		94.29
149	2	0	Perciformes	Sparidae	*Pagrus*	*auriga*	90.12
150	5	4	Odonata	Macromiidae	*Macromia*	*illinoiensis*	89.29
157	3	0	Hymenoptera	Braconidae	*Notiospathius*	*sp. AZR-2011*	89.29
158	2	0	Euonychophora	Peripatidae	*Epiperipatus*		91.03
163	5	4	Coleoptera	Curculionidae	*Falsanchonus*	*emeishanicus*	92.31
165	3	2	Lepidoptera	Gracillariidae	*Phyllonorycter*	*lucetiella*	94.81
167	6	3	Poduromorpha	Hypogastruridae			91.9
170	1	0	Oecophoridae	Philobota	*scitula*		96.15
173	3	5	Coleoptera	Curculionidae	*Acalyptus*	*carpini*	90.12
180	5	0	Diptera	Simuliidae	*Simulium*	*ornatum*	94.05
181	1	1	Symphypleona	Sminthuridae			93.69
182	6	2	Lepidoptera	Xyloryctidae	*Cryptophasa*	*phaeochtha*	94.87

Table S3.5 - Geographic distribution of the most related sequence (in BOLD database) to the Collembola lineages with similarity hit ≥ 99% sampled in Tenerife.

Cluster	Similarity (%)	Geographic distribution of related sequence
2	99.04	France, South Africa
9	100	not provided
11	100	Croatia
12	99.05	France, Argelia
15	100	not provided
16	99.03	not provided
19	100	Finland
21	100	South Africa
32	100	Moldova, France, USA, South Africa
38	100	France
43	100	France, Italy
45	100	Australia, France
51	99.52	Canada, Germany, France, South Africa, Cameron
56	100	France
66	99.52	France
70	100	France, Corsica
82	100	France, South Africa, Canada, Poland, Moldova
83	100	France, Australia, South Africa
90	100	France
91	100	France
102	99.45	Australia, South Africa
114	100	France, South Africa
127	100	Norway, France
128	100	France
151	100	South Africa
153	99.02	France
175	98.57	Australia, New Zealand
176	99.05	South Africa
179	100	New Zealand, South Africa

Chapter 4

Quantifying genetic diversity of *Entomobrya* (Hexapoda, Collembola) lineages in the UK: have they survived through the last Pleistocene glaciation?

Entomobrya albocincta (top left), *E. nicoleti* (bottom left), *E. nivalis* (right), UK.

Photos from http://www.stevehopkin.co.uk/collembolagallery/

4.1 ABSTRACT

Great Britain represents a complex continental island system within which to understand the dynamics of colonisation and species diversification due to its geographic and climatic history. Through periods of geological time, this island was connected to the continent, and for repeated cycles it was glaciated, resulting in the extirpation of its biota over much of the modern land mass. While postglacial recolonisation from the mainland must have been required to fill newly opened ecological spaces, it remains possible that some species could have survived in local refugia during glacial periods, thus also contributing to the re-establishment of biota. Sedentary soil dwelling invertebrates present unique features that may allow them to survive through periods of dramatic climatic change at the regional scale, and have demonstrated remarkable genetic signatures of local persistence through such periods. Despite this, only a few studies to date have looked into the phylogeography of invertebrates within Great Britain. Using the mtDNA COI barcode gene, this study investigated the impact of the last Pleistocene glaciation on the diversity of Collembola within Great Britain, and evaluated signatures of long-term survival. In total, thirteen *Entomobrya* lineages were identified for Great Britain with an overall mean divergence of 20%. Within lineages, overall mean divergences ranged from 0.1 to 4%. The investigation of geographical variation in genetic diversity revealed that overall lineage richness and intra lineage richness are equivalent in glaciated and unglaciated areas. Non-random geographic patterns of genetic variation (i.e. a geographically localised range for a monophyletic group) were found, indicating the existence of genetic variation that evolved within the UK. Using a conservative mtDNA COI evolutionary substitution rate of 0.0504 substitutions/site/myr, estimated dates of the onset of these *in situ* diversification events indicate they have been present for extended periods of time (21,000 and 45,000 years), implying survival through historical climatic and environmental changes. Similar investigations for other soil dwelling fauna, particularly for more species rich groups, may provide stronger evidence for species diversity differences between glaciated and unglaciated areas and further support for endemism and ancient fauna in the UK.

4.2 INTRODUCTION

The diversity and distribution of a particular present-day biota is the result of its combined biogeographic response to both historical and recent processes. While environmental and physiological conditions, particularly the ability to disperse, play a major role in determining a species range (Blackburn & Gaston, 2003), it is undeniable that historical factors have also greatly shaped contemporary patterns of biodiversity, sometimes even in a greater extent than contemporary factors such as climate regimes (Stevens, 2006). In other words, species richness at different spatial and temporal scales is constrained not only by environmental factors but also by evolutionary and ecological history. Thus, the current geographical distribution of species reflects their ability to adapt to local environmental conditions, interact with other species and successfully reproduce in a given location. It also reflects their capacity to successfully colonise areas once an appropriate niche becomes available, as well shift latitudinal and altitudinal range in response to climate change over historical time frames (Brown et al., 1996; Hewitt, 2000; Strona et al., 2012; Wisz et al., 2013).

A prime example of historical climate change affecting the distribution of species is the repeated cycles of global periods of cooling (with increased advance of ice sheets) followed by global periods of warming (when ice sheets retracted and melted with resulting rise of the sea levels) that occurred during the Quaternary (Webb & Bartlein, 1992) . These climate oscillations have significantly affected the distribution of species worldwide particularly causing repeated cycles of expansion and contractions of their distributions, which led to the structured distribution of genetic diversity observed in populations today (e.g. Hewitt, 2004). In the northern hemisphere, during the Pleistocene glaciations (from 2.59-0.01 Ma), as high latitudes were covered by ice and permafrost, northern populations either went extinct or retreated their distributions southward, surviving in refugia - locations that provided suitable habitats for the long-term persistence of populations, over several glacial-interglacial cycles, representing a reservoir of evolutionary history (Tzedakis et al., 2013). As the temperature rose, during the interglacials, southern populations expanded northwards enlarging their distributional ranges, and mixed, or not, to various degrees where they met (Hewitt, 1996; Taberlet et al., 1998). In Europe, these refugia were mainly located in the Iberian, Italian and Balkan peninsulas and evidence to support surviving in these refugia with

subsequent shift of distribution has been clearly demonstrated using DNA relationships for a variety of European species, including the meadow grasshopper-*Chorthippus parallelus* (Hewitt, 1999), the hedgehog *Erinaceus concolor* (Seddon et al., 2001), and the brown bear *Ursus arctos* (Sommer & Benecke, 2005) among others (reviewed in Hewitt, 2011).

Due to their northern geographical latitudinal range, Britain and Ireland had large areas repeatedly covered by glaciers and permafrost throughout the Pleistocene glaciations (Chiverrell & Thomas, 2010), thus becoming inhospitable and restricting the persistence of terrestrial fauna. As these islands were probably isolated from the continent during the interglacials, especially with the formation of the English Channel approximately 0.45 Ma (Gupta et al., 2007), colonisation of terrestrial taxa from continental Europe was also largely prevented. These factors had biogeographical and ecological impacts for the fauna in Britain and Ireland and consequently, these islands have historically been considered of reduced diversity, limited endemism, and with the majority of the current fauna relatively recent and mainly derived from continental Europe (Yalden, 1999; Searle, 2008; Montgomery et al., 2014). Colonisation from Europe is thought to have occurred during a short period after ice retreat and before rising sea levels submerged the land bridge between them (Yalden, 1982; Hewitt, 2004). These considerations, however, have been challenged by evidence of several species that present high levels of diversity, local endemism or persistence through glaciations in some areas of Ireland (Teacher et al., 2009), the English Channel (Coyer et al., 2003; Provan et al., 2005; Hoarau et al., 2007) and more recently in the UK (Edwards et al., 2012; McInerney et al., 2014). For example, in a study investigating the biogeography of the marine red seaweed *Palmaria palmata* along its distribution across Europe and North America coasts, the highest levels of haplotype and nucleotide diversity were found in the English Channel and several haplotypes from this region were not found anywhere else along its distribution revealing a Pleistocene marine glacial refugium in the English Channel (Provan et al., 2005). A recent study provides evidence of an ancient groundwater fauna endemic to Britain and Ireland, of which two amphipod species from the genus *Niphargus* were found to have survived the entire Pleistocene in refugia and also persisted for at least 19.5 million years (McInerney et al., 2014). In the case of groundwater dwelling organisms, persistence through climate changes may have been facilitated by the relative stability of aquatic environments, due to the slower

change in temperature and chemistry of the groundwater compared to the surface water (Gibert et al., 1994; McInerney et al., 2014). However, with the exception of red foxes there has been no other clear evidence for terrestrial fauna persisting the Pleistocene glaciations in the UK. In the case of the red fox *Vulpes vulpes,* it is an ecologically adaptable species whose cytb haplotypes have been found in England during the Pleistocene, also in the Holocene, as well as in modern samples (Edwards et al., 2012).

The possibility of a surface biota surviving in refugia within the UK has been suggested from radiocarbon dates of fossil and vertebrate remains found in several caves sites (Stewart & Lister, 2001; Sommer et al., 2008). Red deer *Cervus elaphus* remains from Ossom's cave, dated to 14263 cal. yr BP (Meiri et al., 2013), and from Kent's Cave, estimated to be pre-LGM (between ca. 60,000 and 25,000 cal.yr BP) based on associated dates (Higham et al., 2011; Meiri et al., 2013), are suggested to imply a possible existence of red deer surviving in refugium in England during the LGM (Stewart & Lister, 2001; Meiri et al., 2013). Although analysis of ancient and contemporary red deer DNA do not support such existence (it shows instead red deer disappeared from central and northern regions and were mainly restricted to European southern refugia during glaciations, and expanded from the Iberian refugium to England and other Central and Northern Europe countries at the end of the LGM, Meiri et al., 2013), persistence of certain species in southern areas is not to be unexpected, as the glaciers did not reach the most southern parts of the UK. Latest estimates indicate that, during the last glacial maximum, the British ice sheet achieved its maximum extent at different times (between 27 and 15 ka) in different sectors, and at maximal extent it spread to the south and east incorporating ice grounds as far as Wales, the Lake District, and Kerry (Clark et al., 2012). Thus, ice free habitats in different places at different times may have been available in some parts of Britain, especially in the South, which could have facilitated refugia and long-term persistence for certain species. If this had happened, the survival of populations in southern British refugial areas would have allowed their lineages to diverge and accumulate genetic differences and with the rising of temperature and retraction of the ice sheet, they could have expanded and colonised northern areas of the UK in a similar way as it has been demonstrated for northern and southern European species (Hewitt 1996, 1999). Genetic signatures of these processes are likely to be seen in the geographical structuring of alleles and in the variation of haplotype richness between unglaciated and glaciated areas. However, these signatures

may be obscured depending on the life history traits of the species, i.e., those with great dispersal abilities and high levels of gene flow.

Sedentary invertebrates, in particular soil invertebrates like Collembola, present unique features that allow them to survive harsh conditions and they have exceptional ability to capture and retain geographic signatures of historical long-acting processes such as climatic cycles (e.g Garrick et al., 2007, Cicconardi et al., 2010, Cicconardi et al., 2013). They are very small, requiring little habitat (i.e. free ice habitat) to live, wingless, with highly restricted dispersal capability and thus reduced gene flow, and are probably the most abundant hexapods on Earth, reaching densities of up to 60 000 individuals per m^2 (Hopkin, 1997). This high abundance level possibly confers on them an increased probability, relative to the majority of arthropod species, of viable fragmented populations persisting and surviving during periods of climatic and/or ecological change (Emerson et al., 2011). Previous molecular studies have demonstrated Collembola to present ancient phylogeographical patterns, which are less likely to be obscured or overwritten by local extinctions induced by the Quaternary glaciations, and also indicate lineages to have persisted through severe climatic conditions such as the Pleistocene glaciations in the Antarctic continent and in the southeast Australian highlands (Garrick et al., 2007; Stevens et al., 2007; Ávila-Jiménez & Coulson, 2011) and also the Messinian Salinity crisis in the Mediterranean basin (Cicconardi et al., 2010). Thus, they offer an ideal opportunity to investigate faunal persistence and survival in possible refugial areas within the UK. Currently, 322 Collembola species are listed for the UK, of which only one has been clearly identified as an endemic species (Shaw et al., 2013). This chapter focuses on a common and widespread Collembolan genus, *Entomobrya*, looking at the geographical distribution of lineages and haplotype diversity across glaciated and unglaciated areas of the UK. It is hypothesised that southern regions of the UK that have remained ice free have higher haplotype diversity than northern areas as a consequence of the extent of coverage by the ice sheet. To evaluate the relative importance of southern areas of the UK in determining the current distribution of genetic diversity in this continental island, we have sequenced the mtDNA COI gene of 722 specimens of *Entomobrya* sampled from across Britain and Wales. Specifically we ask (i) Are there lineages that survived the Pleistocene glaciations? (ii) Are there non-random patterns of geographical distribution of lineages (e.g. geographically localized distribution of monophyletic clades)? (iii) Are there

signatures of greater lineage diversity in the South (or unglaciated areas) compared to the North (or glaciated areas)? (iv) Are there signatures of greater intra-lineage genetic diversity in the South compared to the North?

4.3 MATERIAL AND METHODS

4.3.1 Sampling and laboratory work

Samples were collected from leaf litter and tree surfaces from 98 sites distributed across Britain and Wales from 2011 to 2012 (Fig 4.1). Animals were extracted with Tullgren funnels from litter samples and by vacuum from tree surfaces. Samples were individually placed in 96-well PCR plates and stored into absolute ethanol at 4°C prior to identification by PS. Identification to morphospecies followed Hopkin (2007) and at least 1 photo per morphospecies per site was taken to help posterior identification checks.

A total of 722 individuals were individually extracted for their DNA using the DNeasy 96 well Blood and Tissue Extraction Kit (QIAGEN, West Sussex, UK). Manufacturer's instructions were used with the following modifications aimed to preserve the exoskeleton of each animal. The digestion step was performed directly within the PCR plates where samples had been stored after removing ethanol, to avoid touching the samples. A reduced volume (of 100 μL) of proteinase K + Buffer ATL working solution was added per sample to fit the smaller volume of each well of the PCR plate (200 μL) compared to that of the collection tubes of the DNeasy 96 plate kit (1.500 μL). Samples were incubated for just one hour to avoid lysing the skeleton. After incubation, the lysate was transferred to the DNeasy 96 collection tubes and the PCR plates with remaining exoskeletons were refilled with absolute ethanol and stored as vouchers at 4°C.

Each sample was amplified for the 658 bp barcode with primers ColFol-for and ColFol-rev which are a modification of the primer LCO1490 (Ramirez-Gonzalez et al., 2013). Polymerase chain reactions (PCR) contained NH4 buffer (1x), 3.0 mM $MgCl_2$, 2.5 mM of each dNTP, 0.4 μM of each primer and 0.5 U of Taq polymerase (Bioline) in 25 μL final volume. PCR cycles were carried out using the following thermal profile: 95°C for

2 min, 40 cycles of 95°C for 1 min; 52°C for 45 s, 72°C for 1 min; and finally 72°C for 5 min. PCR products were cleaned with ExoRap protocol, normalized and sent to Eurofins for sequencing with the reverse primer.

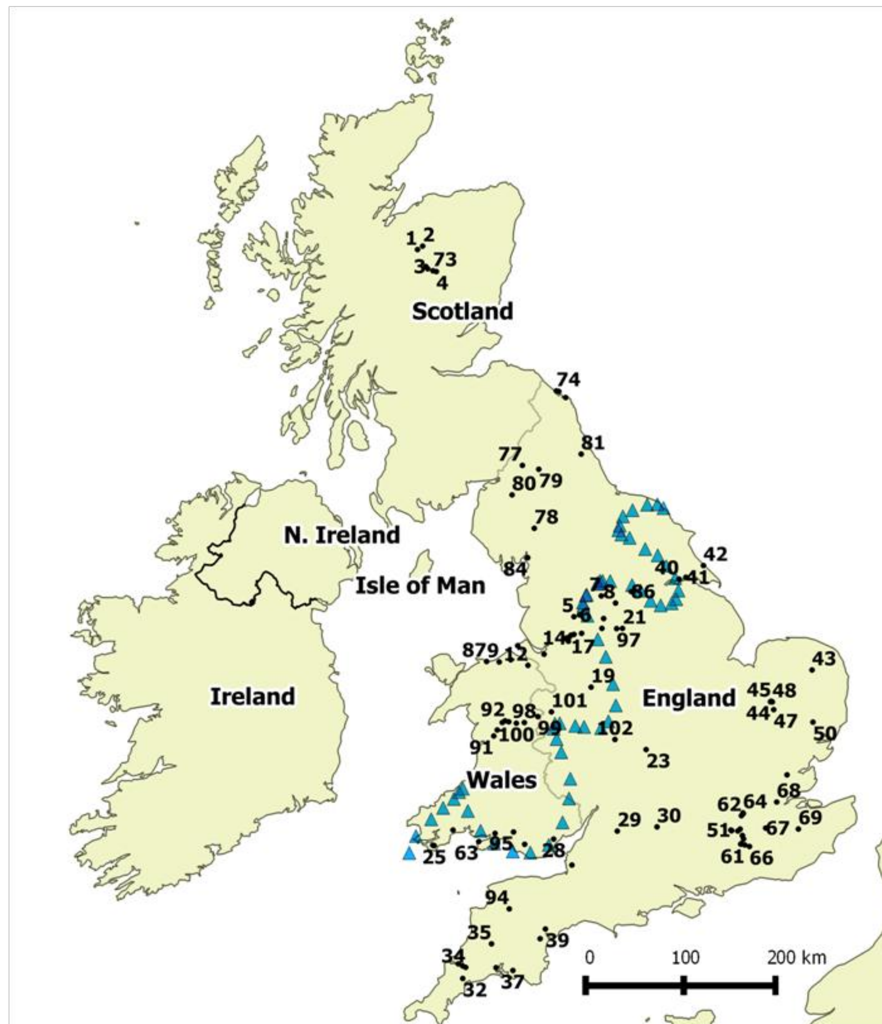


Figure 4.1. Distribution of sampling sites in Great Britain; a complete list of geographic coordinates and number of individuals sampled per locality can be found in Table S4.1. Blue triangles indicate the maximum extent of the British-Irish ice sheet during the last Pleistocene glacial period.

4.3.2 Sequence alignment and molecular lineage delineation

MtDNA COI sequences were processed and ambiguous base calls manually assessed with Geneious Pro 5.6.4 (http://www.geneious.com, Kearse et al., 2012) and aligned using mafft 6.814 (Katoh et al., 2002). Haplotypes were collapsed using DNACollapser (Villesen, 2007). The total number of variable and parsimony informative sites, and uncorrected pairwise genetic distances (overall mean, minimum and maximum) both within and among lineages (the most frequent sequence with the minimum number of differences was selected to represent each lineage) were computed with MEGA6 (Tamura et al., 2013). Sequences were subjected to a neighbour-joining analysis (NJ) using p-distances in MEGA6 (Tamura et al., 2013), and their uncorrected genetic pairwise distances were also computed. Molecular lineages, here defined operationally as a cluster of similar sequences, were delineated using as a threshold the minimum pairwise p-distance found among 19 *Lepidocyrtus* (Collembola, Entomobryidae) mtDNA COI lineages collected sympatrically in six sampling sites in Panama in a previous study by Cicconardi et al (2013). In this study, the mtDNA COII gene was used and lineages, delineated based on coalescent units, diverging from as low as 5% were shown to be consistent with biological species. MtDNA COI sequences for the same sympatric lineages were obtained from Emerson et al. (2011). After identifying clusters, one sequence per lineage was submitted to the BOLD Identification System (http://www.boldsystems.org) to verify geographical distribution of identical or nearly identical sequences sampled from outside the UK.

4.3.3 Evolutionary rate estimation and dating analysis

In the absence of fossil or geological points to estimate divergence times, general molecular clocks are frequently applied (Cicconardi et al., 2010). For the mtDNA COII gene, rates ranging from 1.2% to 4.96% pairwise divergence per million years (Ma), have been proposed for arthropods (Papadopoulou et al., 2010; Ho & Lo, 2013, and references therein). From a comparative analysis of the mtDNA COII rate variation among 20 arthropod groups, Cicconardi et al. (2010) found the genus *Lepidocyrtus* to fall within medium-high rates and indicated a mean rate of 0.0245 substitutions/site/Ma to be a conservative (fast) estimate for the COII. This rate was used to estimate the evolutionary rate of the mtDNA COI for the same group. To do that, eight *Lepidocyrtus*

COII sequences that occur sympatrically in Panama, obtained from Cicconardi et al. (2013), and their corresponding COI sequences (Emerson et al. 2011), were used to run two independent BEAST analyses (Drummond et al., 2012) setting the rate of 0.0245 substitutions/site/Ma to both. The model of nucleotide substitution for each gene was assessed in MEGA6 (Tamura et al., 2013) and the Yule speciation prior was used. The root height of each gene tree was calculated in five independent runs of 10^8 MCMC generations, sampling every 1000 generations. Convergence among the independent runs, and expected sampling sizes (ESS) for the posterior distribution and parameters were evaluated using Tracer v1.6 (Rambaut et al., 2014). The two root ages obtained were quantitatively compared to calculate the rate of substitution for the COI gene. This rate was then applied to another BEAST analysis to estimate the time of the most recent common ancestor (MRCA) for nodes of interest in the *Entomobrya* NJ tree. The most appropriate substitution model for the *Entomobrya* dataset was inferred using MEGA6 (Tamura et al., 2013). An uncorrelated lognormal relaxed clock (Drummond et al., 2006) and a Yule speciation prior were used with the same settings for the MCMC.

4.3.4 Geographical distribution and species diversity analysis

To assess geographical patterns in the distribution of genetic diversity, the localities of occurrence of each haplotype were mapped using QGIS (QGIS Development Team, 2014), then visually inspected for non-random patterns of geographical distribution, i.e., monophyletic clades of mtDNA sequences limited to geographically closely located sites. To evaluate genetic diversity differences between glaciated and unglaciated area, abundance matrices with lineages (OTUs) as rows and localities as columns were created and imported to EcoSim v7.72 (Gotelli & Entsminger, 2000). Analyses were also performed for genetic diversity within each lineage, in which haplotypes were used as rows in the abundance matrices. Localities were grouped into two biogeographical areas: glaciated and unglaciated and they were defined according to the maximum limits of the British-Irish ice sheet (BIIS) during the last Pleistocene glacial period reported in Clark et al. (2012, see Fig. 5). This is the most updated reconstruction of the extent and times of BIIS, based on different lines of evidences such as moraines, lateral meltwater channels, and subglacial melt. The ice sheet was around 0.72 million km^2 in area and with a probable volume of just below 800,000 km^3 (Clark et al., 2012). Although it reached its maximum extent at different times in different sectors (see Fig. 15 in Clark

et al 2012), its overall extent, independent of the time, was used to classify each locality as occurring in a glaciated or unglaciated area (Fig 4.1). To test whether genetic diversity (total lineage richness and haplotype richness within each lineage) is equivalent between these two areas, controlling for differences in sampling and abundances, a rarefaction algorithm as implemented in EcoSim v7.72 (Gotelli & Entsminger, 2000), in which a specified number of individuals are randomly drawn from a community sample, was used to rarefy the larger community down to the abundance level of the smaller. The total abundance of the smaller community being compared was used as the single abundance threshold for each analysis. Simulations were run using the "Species Richness" as the species diversity index. The random number seed was set to 10 and simulations were run for 1000 iterations. Diversity curves and their 95% confidence intervals were inspected to check whether assemblages were significantly different from one another. The mean and variance of diversity generated in each simulation were used to formally test the hypothesis of equivalent species diversity for glaciated and unglaciated communities by checking whether the observed diversity of the smaller community fell within the 95% confidence interval of the simulations for the rarefied community.

4.4 RESULTS

4.4.1 Overview of *Entomobrya* in Great Britain

A total of 722 specimens were sampled from the genus *Entomobrya* from across 90 sampling sites (Tab S4.1, Fig 4.1), which were assigned to 6 morphospecies (*Entomobrya albocincta, E. intermedia, E. marginata, E. multifasciata, E. nicoleti* and *E. nivalis*). Two *Katianna* sp. sequences were sampled as an outgroup. Five morphospecies were found in high numbers and they occur in many locations while only a small number of *E. marginata* individuals were found in a few sites (Tab 4.2). This morphospecies is known to be rare and scarce in the UK contrasting with the common and widespread status of the other five morphospecies. Sequences from the mitochondrial gene COI were obtained from 659 specimens from across the 90 sampling sites. The remaining 63 specimens either failed to amplify or produce sequences of good quality traces. The final COI alignment, with 659 sequences, was 561 bp long and yielded 173 unique haplotypes. Across the ingroup (657 sequences),

there were 217 polymorphic sites of which 200 were parsimony informative. Among the nineteen *Lepidocyrtus* COI lineages collected sympatrically in six Panama localities (see methods for details), the minimum pairwise p-distance value found was 4% divergence (Tab 4.1), thus 96% similarity was the threshold used for clustering sequences down to molecular lineages. In total, thirteen OTUs were identified using DNA barcoding of which three were singletons (Fig 4.2). The minimum, maximum and overall mean divergence among lineages was 7%, 20% and 16% respectively. Within lineages, overall mean divergences ranged from 0.1 to 4 and maximum divergences ranged from 0.2 to 4% (Tab 4.2). Sequences assigned to the morphospecies *E. albocincta* corresponded to a single COI lineage, while the other five morphospecies were composed of two or three divergent mtDNA lineages (Tab 4.2, Fig 4.2). The number of haplotypes per lineage ranged from 7 to 44 for the six more abundant lineages and 1 to 2 for the remaining seven lineages (Tab 4.2). The geographic distribution of the nine lineages with >99% similarity to BOLD sequences revealed the presence of related sequences mainly in Canada and France, but also in Moldova, Norway and South Africa (Tab 4.2).

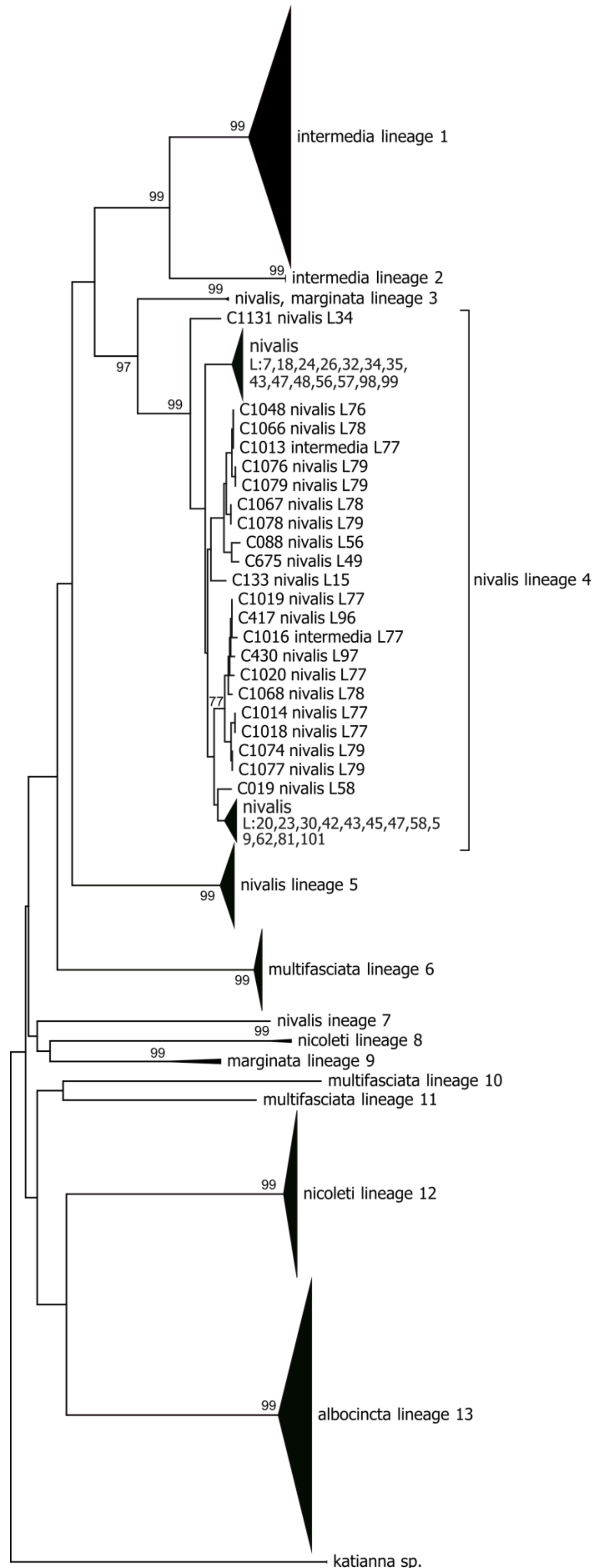
Table 4.1 - Minimum, maximum and overall mean uncorrected p-distances found among *Lepidocyrtus* COI lineages collected sympatrically from six localities in Panama (see text for details).

Localities	N sympatric lineages	Min p-dis	Max p-dis	Overall p-dist
L.b1b-SanFelix	3	0.15	0.17	0.11
L.b2b-ElValle	3	0.05	0.17	0.08
L.ve-Fortuna	3	0.21	0.19	0.13
L.ve-P.I.L.A.	3	0.04	0.11	0.07
L.ve-Darien	4	0.04	0.20	0.17
L.ve-SanFelix	3	0.12	0.22	0.15

Table 4.2. Summary data for each lineage, reporting the number of individuals (n ind) and of mtDNA COI haplotypes (n hap) found, the number of sites each lineage was sampled (n sites), the number of sequences sampled from single localities (n single), the number of localities sampled in glaciated (n glac) and unglaciated (n unglac) areas, the mean and maximum observed intraspecific p-distances, and BOLD information on the locality of occurrence of the identical or near identical sequence found in this database.

	morphospecies	n ind	n hap	n sites	n single	n glac	n unglac	Mean/max intra-lineage p-dist	BOLD most frequent sequence
Lineage 1	*intermedia**	165	22	47	11	21	26	2.1/4	Canada
Lineage 2	*intermedia*	5	2	5	1	3	2	0.1/0.2	Canada
Lineage 3	*nivalis/marginata*	2	2	2	2	1	1	0.2/0.2	-
Lineage 4	*nivalis**	94	38	33	29	12	21	1.3/3.2	France, Canada
Lineage 5	*nivalis**	54	16	24	12	6	18	0.4/1.6	Norway
Lineage 6	*mutifasciata*	52	7	14	4	8	6	0.1/0.5	France, Canada, S. Africa
Lineage 7	*nivalis*	1	1	1	1	0	1	-	-
Lineage 8	*nicoleti*	2	2	1	0	1	0	1.4/1.4	-
Lineage 9	*marginata*	3	2	2	2	0	2	2.4/3.6	-
Lineage 10	*multifasciata*	1	1	1	1	0	1	-	France, Canada
Lineage 11	*multifasciata*	1	1	1	1	0	1	-	France, Canada
Lineage 12	*nicoleti**	106	44	33	37	16	17	0.3/1.6	-
Lineage 13	*albocincta*	172	37	55	27	15	40	1.2/2.9	France, Canada, Moldova

*Indicates lineages mostly composed of one morphospecies but individuals from different morphospecies are also found within these lineages



99
intermedia lineage 1
99
99
intermedia lineage 2
99
nivalis, marginata lineage 3
C1131 nivalis L34
nivalis
L:7,18,24,26,32,34,35,
43,47,48,56,57,98,99
97
99
C1048 nivalis L76
C1066 nivalis L78
C1013 intermedia L77
C1076 nivalis L79
C1079 nivalis L79
C1067 nivalis L78
C1078 nivalis L79
C088 nivalis L56
C675 nivalis L49
C133 nivalis L15
nivalis lineage 4
C1019 nivalis L77
C417 nivalis L96
C1016 intermedia L77
C430 nivalis L97
C1020 nivalis L77
C1068 nivalis L78
77
C1014 nivalis L77
C1018 nivalis L77
C1074 nivalis L79
C1077 nivalis L79
C019 nivalis L58
nivalis
L:20,23,30,42,43,45,47,58,5
9,62,81,101
99
nivalis lineage 5
99
multifasciata lineage 6
nivalis ineage 7
99
nicoleti lineage 8
99
marginata lineage 9
multifasciata lineage 10
multifasciata lineage 11
99
nicoleti lineage 12
99
albocincta lineage 13
katianna sp.
0.02

Figure 4.2. Neighbour joining tree of 659 *Entomobrya* mtCOI Sanger sequences derived from 90 sampling sites across Great Britain (see Fig 4.1). A threshold of 96% similarity was used to cluster sequences into molecular lineages (see methods for details). *Katianna* sp. was sampled as an outgroup. Numbers immediately to the right of morphospecies names correspond to sampling localities. Further lineage details can be found in Table 4.2

4.4.2 Geographical distribution and species diversity analysis

To geographically illustrate the ranges of individual lineages, they were mapped according to their sampling localities (Figs 4.4, S4.1 to S4.3), and checked for non-random patterns of geographical distribution of genetic variation. Lineages 1, 4, 12 and 13 are the most abundant (n individuals > 90) and widespread (n localities > 30) lineages with ranges extending across the geographic extent of sampling sites (Figs 4.4 and S4.1). Lineages 5 and 6 have medium abundances (n individuals >50) and are also relatively well distributed (Fig S4.2). The remaining lineages are represented by a very small number of individuals and occur in a single or very few localities (Fig S4.3). Among the six more abundant lineages, only lineage 4 presented a pattern of non-random geographical structuring of genetic variation (Fig 4.2). Within this lineage, clear structuring of genetic variation was found in two monophyletic clusters of DNA sequences, NivG1 and NivG2, comprised of ten individuals (7 haplotypes) and 6 individuals (3 haplotypes), respectively (bootstrap support: NivG1 = 0.91; NivG2 = 0.67) (Figs 4.2, 4.3 and 4.4). Both clades are found in only a few geographically proximate sites in the North, Northwest and Yorkshire regions.

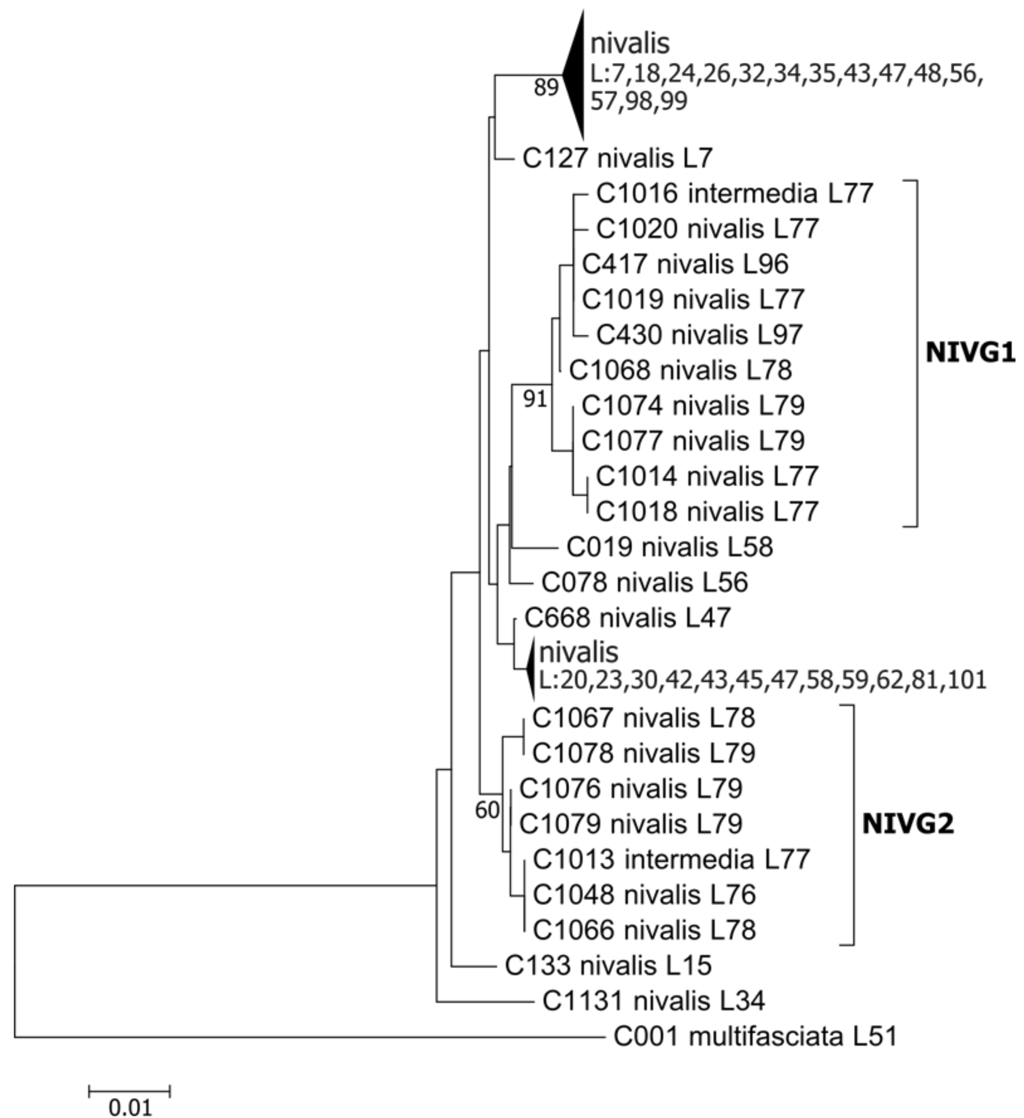


Figure 4.3. Neighour joining tree of mtCOI *Entomobrya* lineage 4 sequences derived from 33 sampling sites across Great Britain. Highlighted are two monophyletic groups NIVG1 and NIVG comprised of 9 haplotypes that were only found in six proximate sampling sites (see Fig 4.4)

After delimiting the maximum extent of the BIIS ice sheet onto Great Britain (Fig. 4.1), sampling localities were classified as being located in the glaciated or unglaciated area and their incidence data were pooled together to test the predictions that lineage diversity and intra-lineage diversity are greater in unglaciated than glaciated areas.

Overall lineage richness and intra-lineage haplotype richness (haplotypes within each of the six more abundant lineages) were compared between these two areas, after controlling for sampling and abundance differences in Ecosim. In total, 43 sampling sites were located in areas covered by the ice sheet, while 56 were located in areas that have not been covered by ice during glaciations. The investigation of geographical variation in *Entomobrya* genetic diversity revealed that overall lineage richness is equivalent in glaciated and unglaciated areas (Tab 4.3). After rarefying the unglaciated community data down to the abundance level of the glaciated community, it was found that for 1000 random samples of 264 individuals, there was an average of 11 lineages with a variance of 0.9 (Tab 4.3). The confidence interval indicates that 95% of the times that a random sample of 264 individuals is drawn from the unglaciated community, it is expected to find between 9-12 different lineages. The observed diversity of the glaciated community (n=9 lineages) is inside this confidence interval, indicating these communities are equivalent in diversity. Looking within lineages, observed haplotype richness of the smaller communities fell within the simulated 95% confidence interval for all but one lineage. Thus, haplotype richness did not differ between glaciated and unglaciated communities for all lineages, except lineage 12 for which haplotype richness was greater in glaciated than unglaciated areas (Tab 4.3).

4.4.3 MtDNA COI rate estimation and dating analysis

A dataset of both COII and COI sequences for eight *Lepiocyrtus* lineages sampled in Panama (Cicconardi et al., 2013) were used to estimate the rate of substitution of the COI gene. Tamura 3-parameter plus Gamma (T92+G) was selected as the best model of molecular evolution for the COII data, and Tamura-Nei plus Gamma (TN93+G) as the best model for the COI data. Using a conservative *Lepidocyrtus* COII evolutionary rate of 0.0245 substitutions/site/Ma (Cicconardi et al., 2010), BEAST estimated the root age of the eight Panama *Lepidocyrtus* COII sequences to be 35 Ma (HPD=12.8-60 Ma), and for the COI sequences, this age was 72 Ma (HPD=24-142 Ma). Therefore, the COI gene is revealed to be approximately twice faster than COII, and a rate of 0.0504 substitutions/site/Ma is used for subsequent COI dating analysis.

The best model of molecular evolution for the two monophyletic groups detected within *E. nivalis* lineage 4 (NivG1 and NivG2) was T92. Adopting the closest model in BEAST

(TN93) and the estimated COI substitution rate (0.0504 substitutions/site/myr), the age of the MRCA for NivG1 is 45,000 years ago (HPD= 10,000-85,000) and for NivG2 is 20,100 years ago (HPD=3,380-49,800). By using this conservative (fast) substitution rate, the ages of lineages are more likely to be underestimates than overestimates (Cicconardi et al. 2013). Thus, it is possible to hypothesize that these *E. nivalis* sequence variations must have survived through the last Pleistocene glaciation, whose last glacial ice sheet subsumed most of Britain from 27,000 and 15,000 years ago (Clark et al., 2009), persisted through the different climatic changes that followed it, and evolved independently with limited, if any, gene flow among neighbouring sites.

4.5 DISCUSSION

Through the analysis of mtDNA COI DNA sequence data, this study presents new information on the spatial pattern of genetic diversity within the springtail genus *Entomobry*a across the UK. We have investigated the impact of the last Pleistocene glaciation on lineage richness and haplotype richness between glaciated and unglaciated areas and identified signatures of long-term persistence of *Entomobrya* through the last Pleistocene glaciation.

4.5.1 Long-term persistence of *Entomobrya* in the UK

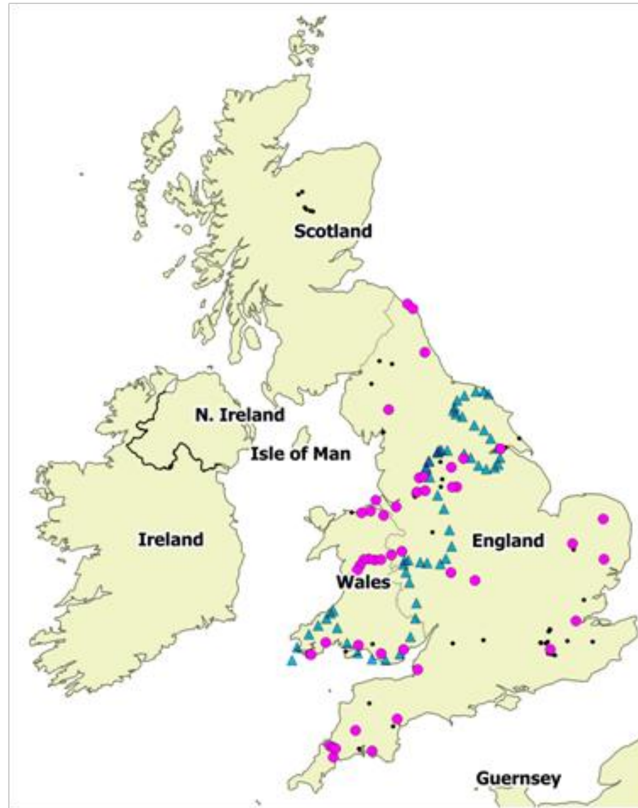
The Pleistocene is regarded as the most influential climatic period that shaped contemporary biogeographical patterns in Europe (Taberlet et al., 1998; Hewitt, 2004). In Britain and Ireland, glaciations had a major impact on the terrestrial fauna, and both genetic and paleontological lines of evidence suggest a reduction in species richness, little endemism and a recent origin for the fauna that currently inhabit both regions (Yalden, 1999; Hewitt, 2000; Searle, 2008). This fauna is thought to have arrived from continental Europe during the late part of the Pleistocene/Holocene (Yalden, 1982; Hewitt, 2004), consequently endemic biota is uncommon and limited to a few species, such as the Celtic woodlouse (*Metatrichoniscoides celticus*), subterranean amphipod (*Niphargus glenniei, N. irlandicus*), Dorset tineid moth (*Eudarcia richardsoni*), Lundy Cabbage flea beetle (*Psylliodes luridipennis*) and Lundy weevil (*Ceutorhynchus contractus pallipes*) (Proudlove et al., 2003; Barnard, 2011; McInerney et al., 2014), that have probably only been present for a few thousands of years. Our data are

consistent with this scenario, but also reveal evidence for population persistence of *Entomobrya* within Great Britain, estimated to extend back at least some 45,000 years. Six geographically proximate sampling sites are represented exclusively by 9 haplotypes that comprise two monophyletic groups, not sampled within any other site. These six sites (Fig 4.4) span both glaciated and unglaciated terrains of Great Britain, with most sequence variation concentrated within the sampling sites extending well into the glaciated northern areas. The estimated age of 45,000 years for the onset of genetic differentiation within the oldest of these two clades indicates persistence through the abrupt climatic and geological changes that have occurred over the last Pleistocene glaciation, at least during the Middle, Late Pleniglacial, Lateglacial and Younger Dryas, in the North and North West of Britain, which was covered by the ice sheet. One possible explanation is survival of Collembola in ice-free habitats within the glaciated area. Although the idea of ice-free areas within glaciated northern regions may seem unlikely, it is important to consider scale. The small size of Collembola, their potentially high population densities of up to 60,000 individuals per m^2, and their soil dwelling habit means that the persistence of populations across time scales measured in tens of thousands of years may be accommodated by relatively small areas of ice-free habitat.

4.5.2 Species diversity analysis

The extent of the BIIS in Great Britain reached as south and east as Wales, the Lake District, and Kerry during the last glaciation (Clark et al., 2012). If soil-dwelling micro-fauna survived within these southern regions that remained ice free, they would be expected to exhibit higher species richness and higher genetic diversity within species, compared to glaciated areas. Our analysis of species diversity, considering divergent genetic lineages to approximate biological species, does not support this hypothesis and indicates north and south communities to have equivalent lineage richness. This is perhaps not surprising due to the small number of *Entomobrya* mtDNA COI lineages found across the UK, potentially limiting statistical power. However, within lineages, genetic diversity also did not differ between glaciated and unglaciated areas for all but one lineage (lineage 12), which in contrast to our expectation presented greater DNA sequence diversity in the North.

a)



b)

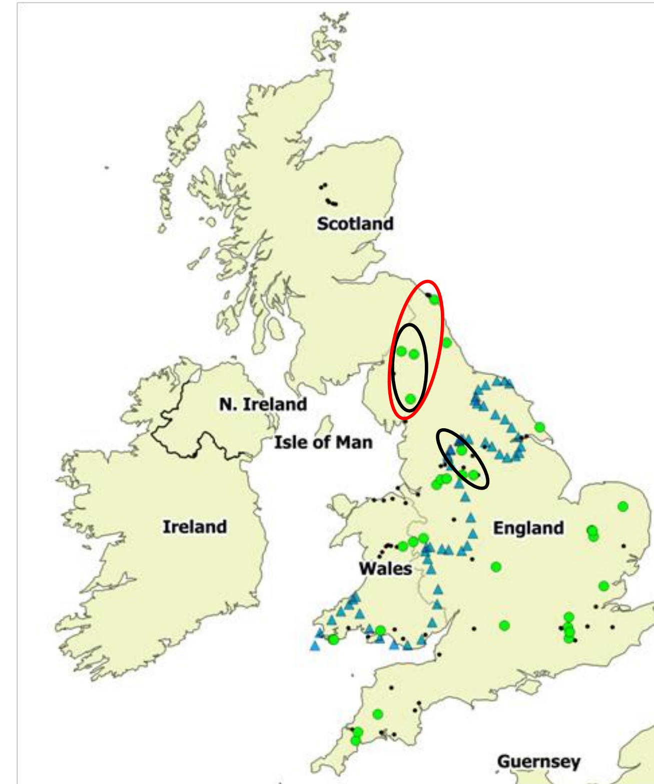


Figure 4.4 Geographical distributions of a) lineage 1 – pink dots, and b) lineage 4 – green dots, according to their sampling localities. Black and red ellipses in b) indicate geographically localized ranges of two monophyletic groups (NIVG1 and NIVG2 respectively) found in lineage 4 (see text for details). Black dots indicate remaining sampling sites. Blue triangles indicate the maximum extent of the British-Irish ice sheet during the last Pleistocene glacial period.

Table 4.3. Summary of Ecosim data entry and diversity results (mean and 95% confidence interval – CI) for the overall lineage richness and for each of the most abundant lineages. For all simulated pairs, the larger community was rarefied down to the abundance level of the smaller community being compared and the observed diversity (n lin/hap) of the smaller community was tested to see if it fell within the estimated 95% confidence interval of the larger community. N ind = number of individuals, N lin/hap = observed number of lineages or of haplotypes, glac = community in the glaciated area, unglac = community in the unglaciated area.

	All lineages		lineage 1		lineage 4		lineage 5		lineage 6		lineage 12		lineage 13	
	glac	unglac	glac	unglac	glac	unglac	glac	unglac	glac	unglac	glac	unglac	glac	unglac
N ind	264	394	75	87	30	63	14	40	20	32	58	50	31	141
N lin/hap	9	12	14	16	15	26	6	12	3	6	27	20	6	33
Mean	-	11	-	14.7	-	16	-	6.3	-	4.4	23.7	-	-	10
95% CI	-	9 to 12	-	13 to 16	-	13 to 20	-	4 to 9	-	2 to 6	21 to 26	-	-	6 to 14

To our knowledge, this is the first study to compare genetic diversity of extant communities from glaciated and unglaciated areas across the UK. Other studies have looked at the consequences of the ice ages on the genetic variation of different species but they have frequently taken a broader spatial scale. One example is provided by McInerney et al (2014), who investigated genetic variation and patterns of geographical distribution of *Niphargus* species across its entire European distribution (Spain, British Isles, West Europe, central Europe, Italy, Balkans, Ponto-Danubian, Caucasus), and found species richness to be highest in south-east Europe compared to the north-west. Although Great Britain is considered in this study, samples sites were mainly concentrated in unglaciated areas and species richness was not compared between glaciated and unglaciated areas within the UK. They also found a very ancient groundwater fauna in NW Europe, with lineages surviving in Britain and Ireland for approximately 20 millions years, thus persisting throughout the multiple glaciations of the Quaternary (McInerney et al., 2014). Our analysis considered much smaller spatial (Great Britain only) and temporal scales (45 Ka for *Entomobrya* vs. 19.5 Ma for *Niphargus*), and although signatures of higher lineage diversity in the southern areas of England as opposed to the northern areas was not found (probably due to the low number of species), signatures of higher haplotype richness in the northern areas were detected for one lineage (lineage 12), as well as signatures of haplotypes surviving and persisting in ice-free habitat of glaciated areas at least during the middle and late stages of the last glaciation in the UK.

This finding of *Entomobrya* lineages persisting through the last glaciation is congruent with those of other studies that demonstrate Collembola surviving extreme climatic disturbances, such as glaciations in Australia and in the Antarctic, and the Messinian Salinity Crisis in the Mediterranean basin (Garrick et al., 2007; Stevens et al., 2007; Cicconardi et al. 2010; Ávila-Jiménez & Coulson, 2011). For example, in Tallaganda, southeast Australia, analysis of *Acanthanura* sp. n. mtDNA COI and nuclear haplotypes found marked population structure and deep phylogeographical breaks dating back to the Early-Mid Pliocene (5.0-3.5Ma), which are explained by survival through Pleistocene or earlier climatic cycles in moist forest refuges (Garrick et al 2007). Deep mtDNA COII divergences and strong fine scale population structure were also found for *Lepidocyrtus* lineages in the Mediterranean basin, and their ages of divergence indicate that at least 36 distinct lineages were already established by the time of the Messinian

salinity crisis (~5.5 Ma), implying survival through the abrupt climatic and ecological changes that occurred in the region (Cicconardi et al. 2010). Together, these studies demonstrate the marked resilience of small, abundant Collembola fauna to historical climate cycles and their utility to investigate the impacts of such changes on the current genetic structuring of the soil diversity.

4.5.3 Conclusions

Our study provides evidence of endemic genetic variation and persistence of Collembola lineages in the UK through extreme climatic changes that characterized the Pleistocene. Non-random geographic patterns of genetic variation, revealed by the geographically localized range of monophyletic groups, indicate the existence of genetic variation that evolved within the UK. Estimated dates of the onset of these *in situ* diversification events indicate they have been present for extended periods (21,000 and 45,000 years) of time, thus implying survival through historical climatic and environmental changes. Similar investigations for other soil dwelling fauna, particularly for more species rich groups, may provide stronger evidence for species diversity differences between glaciated and unglaciated areas and further support for endemism and ancient fauna in the UK. This is a fertile ground for research that, coupled with a detailed sampling in the continental Europe, may shed light into our knowledge about the sources of genetic variation and phylogeographical consequences of Pleistocene glaciations for terrestrial invertebrate fauna in the UK.

4.6 REFERENCES

Ávila-Jiménez M.L. & Coulson S.J. (2011) A Holarctic Biogeographical Analysis of the Collembola (Arthropoda, Hexapoda) Unravels Recent Post-Glacial Colonization Patterns. *Insects*, **2**, 273–296.

Barnard P.C. (2011) *The Royal Entomological Society Book of British Insects*. Wiley-Blackwell, Hoboken, NJ.

Blackburn T.M. & Gaston K.J. (2003) *Macroecology: Concepts and Consequences: 43rd Symposium of the British Ecological Society*. Blackwell Science Ltd, Oxford, UK.

Brown J.H., Stevens G.C., & Kaufman D.M. (1996) The geographic range: Size, Shape, Boundaries, and Internal Structure. *Annual Review of Ecology and Systematics*, **27**, 597–623.

Chiverrell R.C. & Thomas G.S.P. (2010) Extent and timing of the Last Glacial Maximum (LGM) in Britain and Ireland: a review. *Journal of Quaternary Science*, **25**, 535–549.

Cicconardi F., Fanciulli P.P., & Emerson B.C. (2013) Collembola, the biological species concept and the underestimation of global species richness. *Molecular Ecology*, **22**, 5382–5396.

Cicconardi F., Nardi F., Emerson B.C., Frati F., & Fanciulli P.P. (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology*, **19**, 386–400.

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

Clark P.U., Dyke A.S., Shakun J.D., Carlson A.E., Clark J., Wohlfarth B., Mitrovica J.X., Hostetler S.W., & McCabe A.M. (2009) The Last Glacial Maximum. *Science*, **325**, 710–714.

Coyer J.A., Peters A.F., Stam W.T., & Olsen J.L. (2003) Post-ice age recolonization and differentiation of *Fucus serratus* L. (Phaeophyceae; Fucaceae) populations in Northern Europe. *Molecular Ecology*, **12**, 1817–1829.

Drummond A.J., Ho S.Y.W., Phillips M.J., & Rambaut A. (2006) Relaxed Phylogenetics and Dating with Confidence. *PLoS Biology*, **4**, 0699–0710.

Drummond A.J., Suchard M.A., Xie D., & Rambaut A. (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Edwards C.J., Soulsbury C.D., Statham M.J., Ho S.Y.W., Wall D., Dolf G., Iossa G., Baker P.J., Harris S., Sacks B.N., & Bradley D.G. (2012) Temporal genetic variation of the red fox, *Vulpes vulpes*, across western Europe and the British Isles. *Quaternary Science Reviews*, **57**, 95–104.

Emerson B.C., Cicconardi F., Fanciulli P.P., & Shaw P.J.A. (2011) Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2391–2402.

Garrick R.C., Sands C.J., Rowell D.M., Hillis D.M., & Sunnucks P. (2007) Catchments catch all: long-term population history of a giant springtail from the southeast Australian highlands — a multigene approach. *Molecular Ecology*, **16**, 1865–1882.

Gibert J., Danielopol D., & Stanford J.A. (1994) *Groundwater Ecology*. Academic Press, San Diego, California.

Gotelli N.J. & Entsminger G.L. (2000) *EcoSim: Null models software for ecology. Version 5.0*. Acquired Intelligence Inc. & Kesey-Bear, Available from http://homepages.together.net/~gentsmin/ecosim.htm..

Gupta S., Collier J.S., Palmer-Felgate A., & Potter G. (2007) Catastrophic flooding origin of shelf valley systems in the English Channel. *Nature*, **448**, 342–345.

Hewitt G.M. (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.

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

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

Hewitt G.M. (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica*, **139**, 617–638.

Higham T., Compton T., Stringer C., Jacobi R., Shapiro B., Trinkaus E., Chandler B., Gröning F., Collins C., Hillson S., O'Higgins P., FitzGerald C., & Fagan M. (2011) The earliest evidence for anatomically modern humans in northwestern Europe. *Nature*, **479**, 521–524.

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

Hopkin S.P. (1997) *Biology of the Springtails*. Oxford University Press, Oxford, UK.

Hopkin S.P. (2007) *A Key to the Collembola (springtails) of Britain and Ireland*. Field Studies Council-AIDGAP, Shrewsbury, UK.

Ho S.Y.W. & Lo N. (2013) The insect molecular clock. *Australian Journal of Entomology*, **52**, 101–105.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., & Drummond A. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

McInerney C.E., Maurice L., Robertson A.L., Knight L.R.F.D., Arnscheidt J., Venditti C., Dooley J.S.G., Mathers T., Matthijs S., Eriksson K., Proudlove G.S., & Hänfling B. (2014) The ancient Britons: groundwater fauna survived extreme climate change over tens of millions of years across NW Europe. *Molecular Ecology*, **23**, 1153–1166.

Meiri M., Lister A.M., Higham T.F.G., Stewart J.R., Straus L.G., Obermaier H., González Morales M.R., Marín-Arroyo A.B., & Barnes I. (2013) Late-glacial recolonization and phylogeography of European red deer (*Cervus elaphus* L.). *Molecular Ecology*, **22**, 4711–4722.

Montgomery W.I., Provan J., McCabe A.M., & Yalden D.W. (2014) Origin of British and Irish mammals: disparate post-glacial colonisation and species introductions. *Quaternary Science Reviews*, **98**, 144–165.

Papadopoulou A., Anastasiou I., & Vogler A.P. (2010) Revisiting the Insect Mitochondrial Molecular Clock: The Mid-Aegean Trench Calibration. *Molecular Biology and Evolution*, **27**, 1659–1672.

Proudlove G.S., Wood P.J., Harding P.T., Horne D.J., Gledhill T., & Knight L.R.F.D. (2003) A review of the status and distribution of the subterranean aquatic Crustacea of Britain and Ireland. *Cave and Karst Science*, **30**, 53–74.

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

QGIS Development Team (2014) *QGIS Geographic Information System*. Open Source Geospatial Foundation Project. Available from http://qgis.osgeo.org.

Rambaut A., Suchard M.A., Xie D., & Drummond A.J. (2014) *Tracer v1.6*. Available from http://beast.bio.ed.ac.uk/Tracer.

Ramirez-Gonzalez R., Yu D.W., Bruce C., Heavens D., Caccamo M., & Emerson B.C. (2013) PyroClean: Denoising Pyrosequences from Protein-Coding Amplicons

for the Recovery of Interspecific and Intraspecific Genetic Variation. *PLoS ONE*, **8**, 1–11.

Searle J.B. (2008) The colonization of Ireland by mammals. *The Irish Naturalists' Journal*, **29**, 109–115.

Seddon J.M., Santucci F., Reeve N.J., & Hewitt G.M. (2001) DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor*: Pleistocene refugia, postglacial expansion and colonization routes. *Molecular Ecology*, **10**, 2187–2198.

Shaw P., Faria C., Emerson B., & others (2013) Updating taxonomic biogeography in the light of new methods–examples from Collembola. *Soil Organisms*, **85**, 161–170.

Sommer R.S. & Benecke N. (2005) The recolonization of Europe by brown bears *Ursus arctos* Linnaeus, 1758 after the Last Glacial Maximum. *Mammal Review*, **35**, 156–164.

Sommer R.S., Zachos F.E., Street M., Jöris O., Skog A., & Benecke N. (2008) Late Quaternary distribution dynamics and phylogeography of the red deer (*Cervus elaphus*) in Europe. *Quaternary Science Reviews*, **27**, 714–733.

Stevens M.I., Frati F., McGaughran A., Spinsanti G., & Hogg I.D. (2007) Phylogeographic structure suggests multiple glacial refugia in northern Victoria Land for the endemic Antarctic springtail *Desoria klovstadi* (Collembola, Isotomidae). *Zoologica Scripta*, **36**, 201–212.

Stevens R.D. (2006) Historical processes enhance patterns of diversity along latitudinal gradients. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2283–2289.

Stewart J.R. & Lister A.M. (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, **16**, 608–613.

Strona G., Galli P., Montano S., Seveso D., & Fattorini S. (2012) Global-Scale Relationships between Colonization Ability and Range Size in Marine and Freshwater Fish. *PLoS ONE*, **7**, 1–8.

Taberlet P., Fumagalli L., Wust-Saucy A.-G., & Cosson J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Tamura K., Stecher G., Peterson D., Filipski A., & Kumar S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Teacher A.G.F., Garner T.W.J., & Nichols R.A. (2009) European phylogeography of the common frog (*Rana temporaria*): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, **102**, 490–496.

Tzedakis P.C., Emerson B.C., & Hewitt G.M. (2013) Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, **28**, 696–704.

Villesen P. (2007) FaBox: an online toolbox for fasta sequences. *Molecular Ecology Notes*, **7**, 965–968.

Webb T. & Bartlein P.J. (1992) Global changes during the last 3 million years: climatic controls and biotic responses. *Annual review of Ecology and Systematics*, **23**,141–173.

Wisz M.S., Pottier J., Kissling W.D., Pellissier L., Lenoir J., Damgaard C.F., Dormann C.F., Forchhammer M.C., Grytnes J.-A., Guisan A., Heikkinen R.K., Høye T.T., Kühn I., Luoto M., Maiorano L., Nilsson M.-C., Normand S., Öckinger E., Schmidt N.M., Termansen M., Timmermann A., Wardle D.A., Aastrup P., & Svenning J.-C. (2013) The role of biotic interactions in shaping distributions and

realised assemblages of species: implications for species distribution modelling. *Biological Reviews of the Cambridge Philosophical Society*, **88**, 15–30.
Yalden D. (1999) *The History of British Mammals*. T & A D Poyser, London.
Yalden D.W. (1982) When did the mammal fauna of the British Isles arrive? *Mammal Review*, **12**, 1–57.

4.7 Appendices

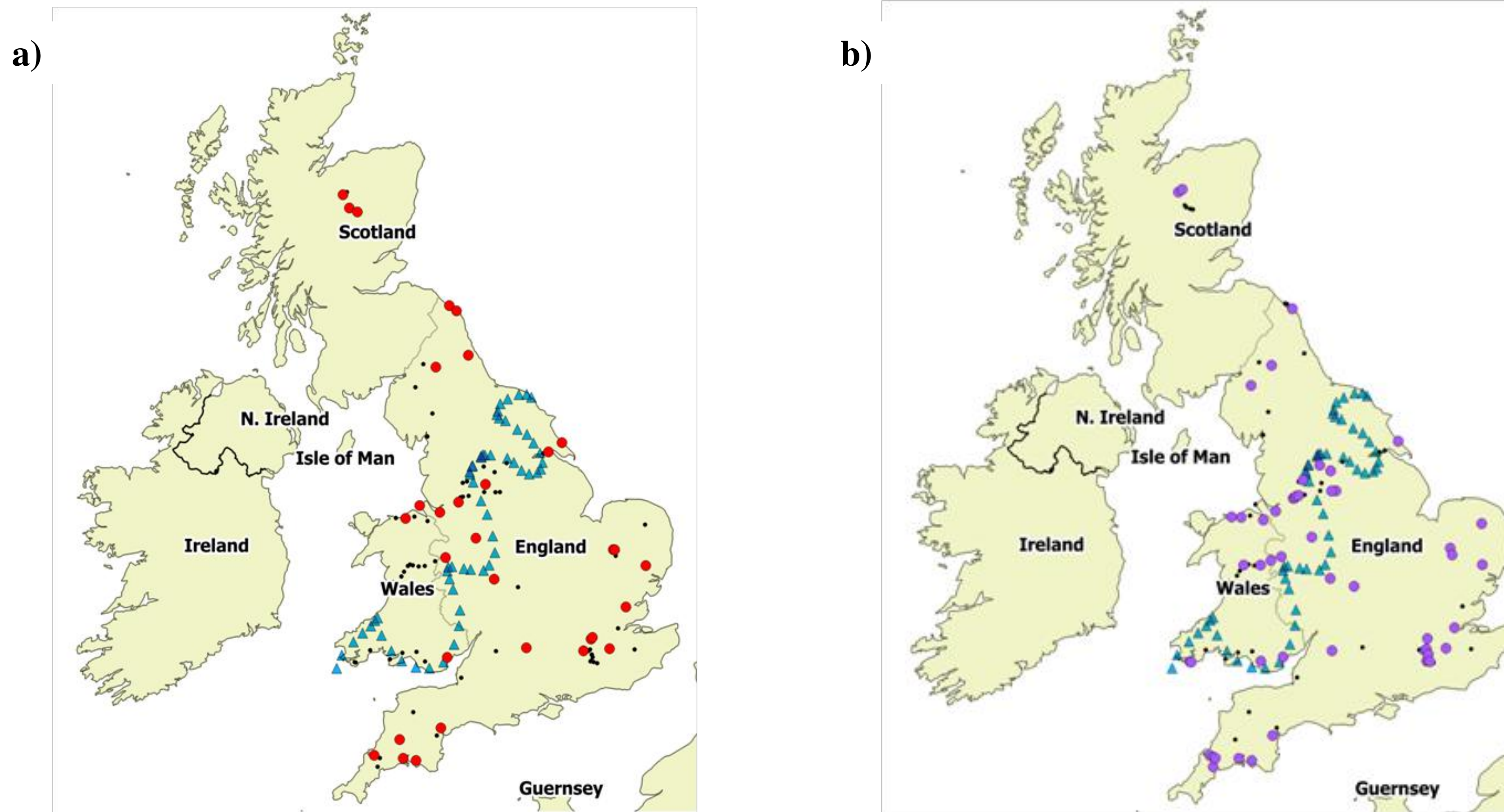
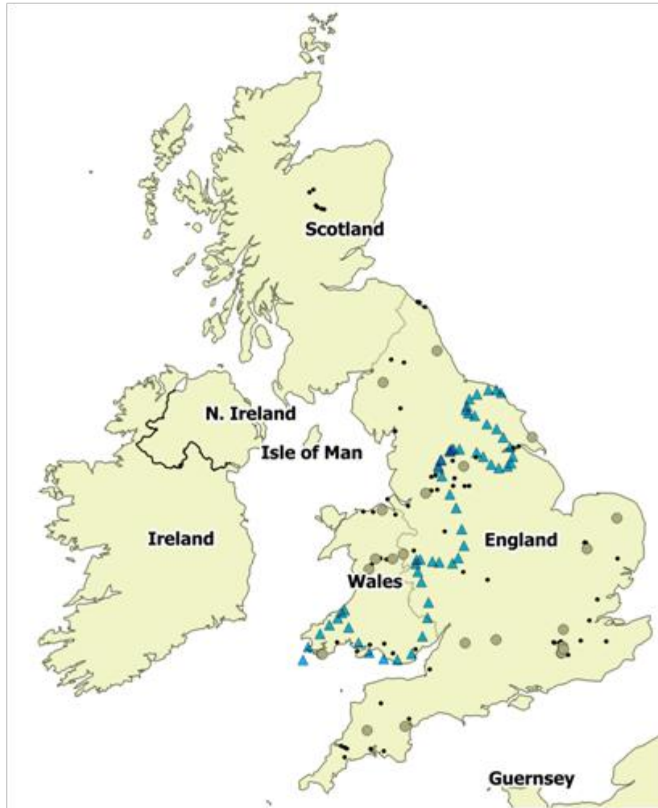


Figure S4.1 Geographical distributions of a) *Entomobrya* lineage 12 – red dots, and b) *Entomobrya* lineage 13 – purple dots, according to their sampling localities. Small black dots indicate remaining sampling sites. Blue triangles indicate the maximum extent of the British-Irish ice sheet during the last Pleistocene glacial period.

a)



b)

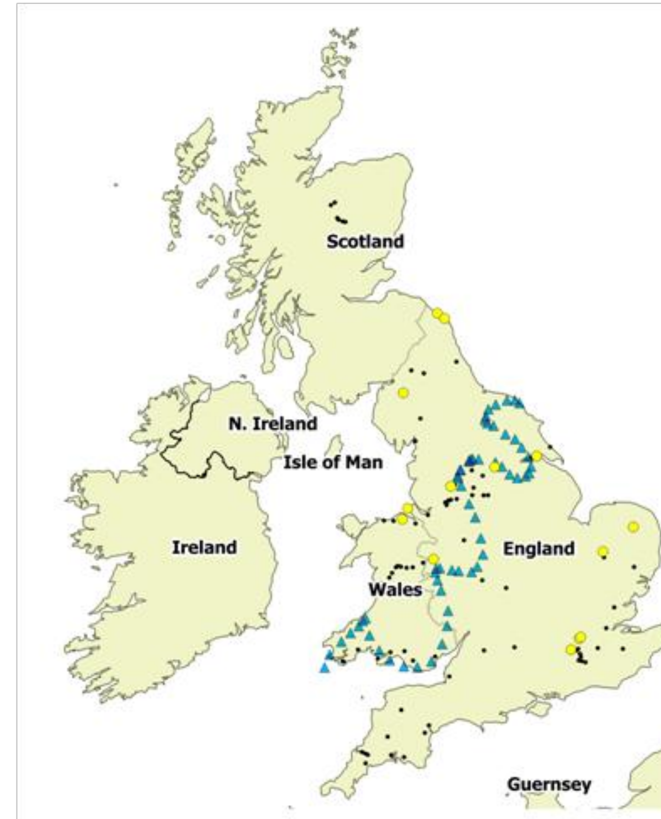


Figure S4.2 Geographical distributions of a) lineage 5 – grey dots, and b) lineage 6 – yellow dots, according to their sampling localities. Small black dots indicate remaining sampling sites. Blue triangles indicate the maximum extent of the British-Irish ice sheet during the last Pleistocene glacial period.

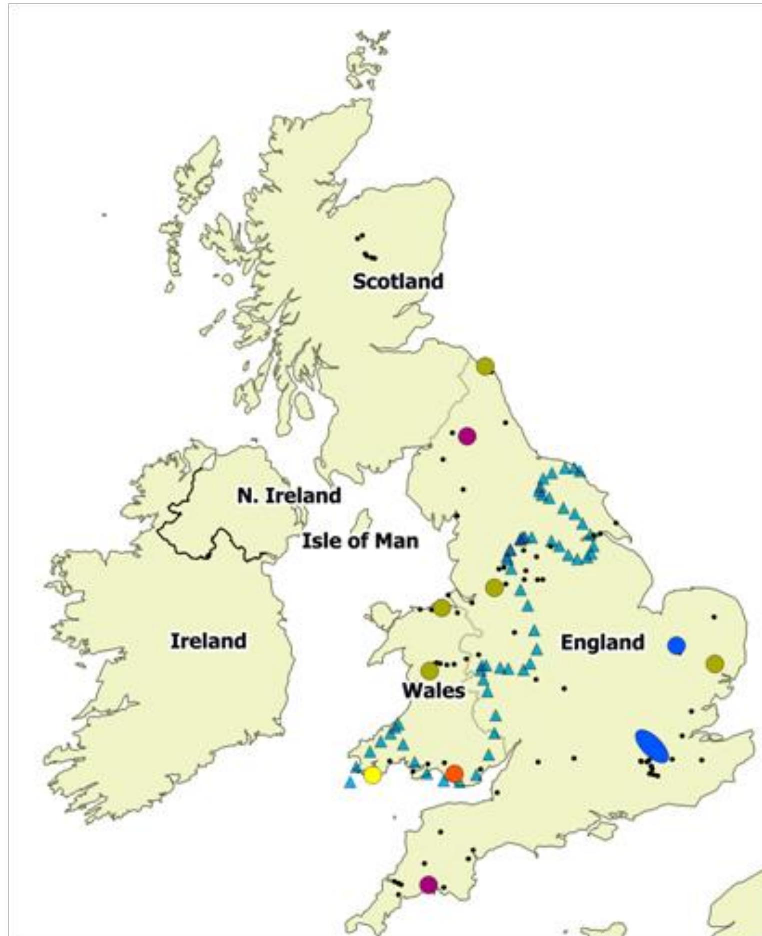


Figure S4.3. Geographical distributions of *Entomobrya* mtDNA COI lineages according to their sampling localities. Lineage 2 – green dots, lineage 3- dark purple dots, lineages 7 and 10 – dark blue ellipse, lineage 8 - orange dot, lineage 9 – dark blue dot and ellipse, lineage 11- yellow dot. Black dots indicate remaining sampling sites. Blue triangles indicate the maximum extent of the British-Irish ice sheet during the last Pleistocene glacial period.

Table S4.1. Sampling sites within the island of Great Britain, reporting the geographic coordinates and number of *Entomobrya* individuals collected per locality. Locations coded according to Fig 4.1.

Locality	Latitude	Longitude	N individuals
1	57.226940	-3.744723	3
2	57.250828	-3.646657	1
3	57.056782	-3.650115	2
4	56.992771	-3.484993	3
5	53.567368	-2.232038	11
6	53.574394	-2.140705	7
7	53.711956	-1.745219	4
8	53.621056	-1.544229	8
9	53.267548	-3.517447	12
10	53.270206	-3.332264	9
11	53.390812	-3.188403	9
12	53.191521	-3.080297	3
13	53.269020	-2.800899	7
14	53.383728	-2.403826	2
15	53.348923	-2.387444	5
16	53.383045	-2.369771	8
17	53.401390	-2.331261	3
18	53.403618	-2.281280	2
19	52.886627	-2.164448	22
20	53.400555	-2.160005	2
21	53.374790	-1.511842	3
22	51.768250	-4.620054	10
23	52.215801	-1.503167	15
24	51.653542	-4.957234	10
25	51.647945	-4.937279	17
26	51.678989	-3.993226	19
27	51.534397	-3.574383	19
28	51.539238	-3.128327	4
29	51.511406	-2.158805	17
30	51.482948	-1.557258	15
31	50.512501	-4.822230	9
32	50.369995	-4.786108	9
33	50.487782	-4.762217	5
34	50.470009	-4.720558	5
35	50.654724	-4.290828	5
36	50.426395	-4.275012	4
37	50.376942	-4.034168	5
38	50.633892	-3.565275	3
39	50.716389	-3.465000	3
40	53.723221	-0.479807	9

Table S4.1. (cont)

Locality	Latitude	Longitude	N individuals
41	53.727322	-0.369791	4
42	53.800648	-0.057395	14
43	52.623665	1.246741	36
44	52.419460	0.515290	2
45	52.423801	0.527952	4
47	52.341007	0.537557	11
48	52.414165	0.542738	17
49	51.718281	0.523698	1
50	52.146744	1.081943	11
51	51.321110	-0.475282	11
52	51.313606	-0.466388	3
53	51.316383	-0.464724	1
54	51.299717	-0.374997	11
55	51.173615	-0.371669	2
56	51.176941	-0.360549	23
57	51.312778	-0.335275	2
58	51.248577	-0.320651	14
59	51.248592	-0.320650	3
60	51.215530	-0.314491	1
61	51.157887	-0.314006	10
62	51.433052	-0.276102	13
63	51.626289	-4.256871	2
64	51.456116	-0.244719	0
65	51.448338	-0.242504	20
66	51.135246	-0.249651	0
67	51.273663	0.042656	2
68	51.490414	0.285921	3
69	51.200829	0.518717	0
72	57.032429	-3.625677	1
73	57.005009	-3.541238	0
74	55.684959	-1.838683	6
75	55.675194	-1.799497	1
76	55.606403	-1.705754	10
77	55.054352	-2.617450	8
78	54.451942	-2.606670	4
79	54.991104	-2.363755	11
80	54.798607	-2.868619	5
81	55.056610	-1.632874	8
83	54.190464	-2.797168	0
84	54.191376	-2.795045	0
85	53.498901	-1.774891	6
86	53.698200	-1.264826	7

Table S4.1. (cont)

Locality	Latitude	Longitude	N individuals
87	53.291092	-3.712894	8
88	52.697304	-3.519311	1
89	52.699707	-3.621308	2
90	52.638248	-3.716803	11
91	52.587330	-3.784479	5
92	52.711727	-3.575345	0
93	51.269722	-2.919990	1
94	50.955917	-3.952013	1
95	51.665211	-3.712372	0
96	53.411362	-1.830386	1
97	53.383255	-1.602855	12
98	52.666122	-3.282838	11
99	52.700470	-3.060518	18
100	52.672260	-3.407084	3
101	52.722786	-2.839915	40
102	52.362823	-1.945560	13

Chapter 5

Genetic diversity and persistence of *Lepidocyrtus* (Hexapoda, Collembola) lineages through historical climate changes in the island of Great Britain

Lepidocyrtus curvicollis (Entomobryidae), UK. Photo from
http://www.stevehopkin.co.uk/collembolagallery/

5.1 ABSTRACT

Recent work on British soil dwelling fauna has revealed evidence for endemic genetic variation indicating the survival of springtail lineages (genus *Entomobrya*) through the abrupt climatic and geological changes that occurred over the last Pleistocene glaciation in Great Britain. To investigate support for this evidence and the generality of these findings for the British Collembola fauna, we further sampled within the more species rich genus *Lepidocyrtus* with the mtDNA COI gene across Great Britain to (i) assess lineage richness between glaciated and unglaciated communities, (ii) evaluate signatures of disjunct distributions and structuring of genetic variation, and (iii) estimate the ages of the onset of diversification within geographically localised lineages. Using 96% similarity threshold to cluster sequences into molecular lineages, a total of 22 *Lepidocyrtus* OTUs were identified, with an overall mean divergence of 14%. Lineages presented widespread ranges (n=3), disjunctive distributions (n=3), or were geographically restricted to a few proximate sites (n=7). Lineage richness was significantly different between glaciated and unglaciated communities, with lower than expected lineage richness found for the glaciated community. Within widespread lineages, haplotype richness did not differ between these two communities. Using a conservative mtDNA COI rate of 0.0504 substitutions/site/Ma, estimated ages for lineages with genetic variation geographically structured ranged from 112,800 to 326,400 years ago, which substantially predates the Last Glacial Maximum. Theses results corroborate previous signatures of differentiation and persistence of Collembola lineages in Great Britain and are also congruent with other studies showing soil dwelling fauna surviving harsh climatic conditions, possibly due to their particular life history traits, which confer on them abilities to resist extreme conditions. Placing these findings into the European context, Collembola extend the list of species showing evidence of survival in high latitude refugia, confirming a more mosaic model of persistence during Pleistocene glaciations, which involves not only southern peninsulas but also northern refugia areas.

5.2 INTRODUCTION

Understanding species distributions across the Earth has been of general interest to ecologists, evolutionary biologists and biogeographers for centuries (Ricklefs, 2004 and references therein), who have faced the challenging tasks of characterizing general spatial patterns of biodiversity, and determining the main factors that describe them. Recent years have seen considerable advances in methodological and analytical tools (e.g. Elith & Leathwick, 2009; Jetz & Fine, 2012; Fonseca et al., 2014) that allow exploration of global patterns in biodiversity distribution, including hotspots of diversity, change with spatial scales such as species-area relationships, and the influence of gradients such as latitudinal variation in species richness (Gaston, 2000, 2009). A combination of historical (e.g. speciation, extinction, dispersal, continental drift, and glaciation) and contemporary (e.g. climate, topographic relief, biotic interactions, ecological tolerances) factors have been found to explain these patterns (e.g. Hewitt, 1996; Hawkins & Porter, 2003; Whittaker et al., 2007). Within this area of interest the influence of historical climate change on shaping the evolution and distribution of diversity has received a lot of interest within the European landscape, primarily through phylogeographic analysis (e.g. Hewitt, 2000; Wilson & Veraguth, 2010; Davison et al., 2011).

Due to the repeated Pleistocene glaciations and climatic fluctuations, many European temperate and boreal taxa were forced to contract their distributions into low-latitudes, persisting in three major southern refugial areas within the Mediterranean peninsulas, the Iberian, Italian and Balkan peninsulas (Bennett et al., 1991; Taberlet et al., 1998; Hewitt, 2004). These southern populations then expanded their ranges towards mid and high-latitudes when conditions improved during interglacial periods (e.g. Bennett et al., 1991; Hewitt, 2004) . More recently, phylogeographic and paleoecological research have also begun to provide support for a model where certain species persisted in mid- to high-latitude refugia during glaciations and contributed to the recolonization of Central and Northern Europe when glaciers receded (Provan & Bennett, 2008; Hofreiter & Stewart, 2009; Stewart et al., 2010). This has been demonstrated for a variety of taxa, including invertebrates (e.g. Nieberding et al., 2005; Benke et al., 2009), amphibians (e.g. Teacher et al., 2009), reptiles (e.g. Ursenbacher et al., 2006), mammals (e.g. Deffontaine et al., 2005; Valdiosera et al., 2007) and plants (e.g Petit et al., 2003).

Consequently, there is now an increasing recognition of a potentially more mosaic model of persistence, involving not only southern but also central and northern refugial areas.

Surviving in more northern refugia would require species to exhibit traits that are expected to promote persistence in these regions such as cold tolerance, small body sizes, short generation time, generalist habitat preference, northerly present-day distributions (Bhagwat & Willis, 2008; Hopper, 2009). There are a number of reasons these traits are thought to be important: (i) cold tolerance implies a number of physiological adaptations to survive in cooler climates such as desiccation tolerance (Marshall & Coetzee, 2000); (ii) small body size implies a higher metabolic rate which enables survival in cold climates (Li & Wang, 2005); (iii) short generation time would also be advantageous as it allows fast reproduction and dispersal in harsh environments (Clark et al., 2001); (iv) habitat-generalist species are more tolerant to a varied range of environmental conditions than habitat-sensitive species are (Cooper & Gessaman, 2004); (v) present-day northerly distributions also presume species that are inherently capable of surviving in cooler climates (Bhagwat & Willis, 2008). High mobility is also thought to confer an advantage for vertebrates when surviving in refugia surrounded by a permafrost landscape conferring the ability to disperse into suitable areas as climate changes (Andreev, 1999). These geographical, ecological and life-history traits have been investigated in a variety of European woody plants and vertebrates to test their correlation with survival in northern refugia during full glacial times (Bhagwat & Willis, 2008). Wind-dispersed, habitat generalist trees with the ability to reproduce vegetatively, and habitat generalist mammals with present-day northerly distributions have been clearly demonstrated to encompass the traits that allowed them to persist in northern refugia (Bhagwat & Willis, 2008).

While the analysis of Bhagwat and Willis (2008) looked at a number of woody plants and mammals, and traits were based on these groups, we are not aware of similar analyses for invertebrates, a very diverse group among which many species naturally hold many of these traits (e.g. limited desiccation tolerance and reduced body size of Antarctic microarthropods adapted to harsh environmental conditions [McGaughran et al., 2010]). Among invertebrates, soil dwelling species present the potential to survive through extreme environmental conditions in northern refugia as they have small body

sizes, are typically ubiquitous, and are often present in high population densities (which potentially confers a degree of local demographic stability, i.e. maintenance of viable populations through time at small spatial scales [Garrick et al., 2008; Emerson et al., 2011]) . They also present low vagility, which, contrary to the model for vertebrates, has been suggested to correlate with survival in small refugial areas (Moritz et al., 2001; Hugall et al., 2002), suggesting their suitability for recovering phylogeographic signatures of persistence through long-acting climatic cycles. While for large vertebrates such as deer, fox, alce (Bhagwat and Willis, 2008 and references therein) high mobility would be useful to help encounter suitable remaining habitat across the landscape, this does not apply to soil dwelling species. Their ubiquity implies that rather than needing to disperse to find suitable environmental conditions, they would already be present at possible refugial sites. Soil organisms may of course undergo passive long distance dispersal (e.g. wind dispersal), but results from molecular work suggest this to be of limited importance. Molecular data reveal that soil dwelling Collembola may not move even short distances of only tens of kilometers over timescales exceeding the Pleistocene (e.g. Garrick et al., 2007; Cicconardi et al., 2010). Phylogeographic analyses of some soil fauna (e.g. springtails, mites and roundworms) have frequently revealed deep evolutionary divergences among lineages, dating back millions of years, which implies lineages to have persisted through multiple glacial cycles and other potential vicariant events (Nieberding et al., 2005; Garrick et al., 2007; Stevens et al., 2007; McGaughran et al., 2008, 2010; Cicconardi et al., 2010; Mortimer et al., 2012). As a consequence, remarkable population structuring among geographically proximate sites can be found within soil dwelling taxa. For example, a fine scale phylogeographic analysis of the long-term population history of the giant springtail *Acanthanura* sp. from the Australian highlands found marked population structure, deep molecular divergences and abrupt mtDNA COI breaks among six *Acanthanura* sp. n., with the oldest break possibly dating back to the Early–Mid Pliocene (5.0–3.5 Myr), suggesting lineages to have survived through Pleistocene climatic cycles in montane temperate forests refuges in southeastern Australia (Garrick et al., 2007). In the North-Western Mediterranean basin, deep mtDNA COII divergences and remarkable geographic structure were also found for *Lepidocyrtus* lineages, and their ages of divergence indicate many of them were already established by the time of the Messinian salinity crisis (~5.5 Ma), implying survival through the abrupt climatic and ecological changes that occurred in the region (Cicconardi et al. 2010).

Other soil taxa also provide examples of probable persistence of taxa in geographic areas subjected to harsh environmental conditions during the glacial periods. A phylogeographic study of the nematode *Heligmosomoides polygyrus* in the western Paleartic region revealed a highly differentiated northern clade in Ireland and Denmark, with its time of differentiation, estimated between 2.02±0.21 and 1.46±0.19Myr, suggesting it may have survived the Quaternary ice ages in refugia in the southern parts of the British Isles and Denmark (Nieberding et al., 2005). In a broad-scale molecular analysis of ameronothroid mites from the maritime and sub-Antartic regions, deep divergence times were found among groups (e.g. mean dates ranging from 6.15 to 9.47 Myr among the genera *Podacarus, Alaskozetes* and *Halozetes*) that significantly predate the Pleistocene, thus suggesting their survival in refugia within the Antarctic Peninsula, which remained heavily glaciated since 12 Myr until the end of the Last Glacial Maximum (Mortimer et al., 2011).

Although at a smaller temporal scale (thousands instead of million of years), recent work on British soil dwelling fauna has also revealed evidence for survival and persistence of springtail lineages of the genus *Entomobrya* through the abrupt climatic and geological changes that occurred over the last Pleistocene glaciation in Great Britain. This is indicated by the occurrence of two monophyletic groups of mtDNA sequences with geographically localized distributions in discrete areas in the North and North West regions of England. Together with the estimated age of 45,000 years for the onset of genetic differentiation within the oldest of these two monophyletic groups, which suggest endemic genetic variation diversifying *in situ*. Surprisingly, sampling sites for both groups were mainly concentrated in the northern glaciated areas of the Great Britain, indicating persistence at least through the last Pleistocene glaciation in ice-free habitats that may have existed within these areas. While this scenario suggests evidence for long-term persistence, it was only found within one out of the 13 *Entomobrya* lineages currently sampled across the UK (Faria et al., Chapter 4). If this scenario were real, we would expect to find signatures of persistence in other Collembola species.

Here we assess the generality of the finding of chapter 4 by further sampling within the genus *Lepidocyrtus* across Great Britain. Like the genus *Entomobrya*, *Lepidocyrtus* is

also commonly found and widespread in this continental island. If Collembola lineages survived the last Pleistocene glaciation in Great Britain, the following genetic signatures are expected to be found: (i) lower lineage richness and intra-lineage genetic richness in the northern glaciated areas compared to the southern unglaciated areas. This pattern is generally referred to as 'southern richness and northern purity,' derived from the fact that northern areas are expected to be recolonized from subsets of the genetic diversity present in southern refugial populations (Hewitt, 1996); (ii) non-random spatial patterns of genetic variation structuring (i.e. monophyletic group of sequences with geographically localized distribution in discrete areas in the UK) pointing to endemic variation that has diversified within and are limited to areas of long-term persistence; (iii) divergence times for the onset of this endemic genetic variation that predate the Last Glacial Maximum-LGM; (iv) disjunct distributions of lineages, suggesting lineages with previously widespread geographic distributions that became patchy. Such patterns are expected when a widespread taxa becomes restricted to a refugial area, but fails to achieve a broader distribution as climate and conditions improve. Such a pattern is mediated by the leading-edge effect – pioneers population from related taxa rapidly expand their distribution and recolonize ice released areas. Once the empty space has been filled by these pioneers populations, it is more difficult for related taxa to colonise and establish (Hewitt, 2004). We evaluate evidence for these expectations by sampling the mtDNA COI gene for 428 specimens of *Lepidocyrtus* sampled from across Great Britain.

5.3 MATERIAL AND METHODS

5.3.1 Sampling and laboratory work

Between 2011 to 2012, as part of a bigger project to estimate Collembola diversity in the UK, more than 1000 individuals were collected from 98 sites distributed across Britain and Wales (Fig 5.1). Among them, a total of 428 individuals belonging to the genus *Lepidocyrtus* were sampled from 68 of the 98 sampling sites (Table S5.1). The remaining individuals belong to the genus *Enotmobrya*, which were analysed in the previous chapter. Animals were collected from leaf litter (extracted with Tullgren funnels) and from tree surfaces (extracted by using a vacuum). Table S5.1, Supporting Information, lists morphospecies, their collection sites, number of individuals per sites

and national grid information. Upon collection, samples were individually placed in 96-well PCR plates and stored into absolute ethanol at 4°C prior to identification by PS. Identification to morphospecies followed Hopkin (2007) and at least 1 photo per morphospecies per site was taken to help posterior identification checks.

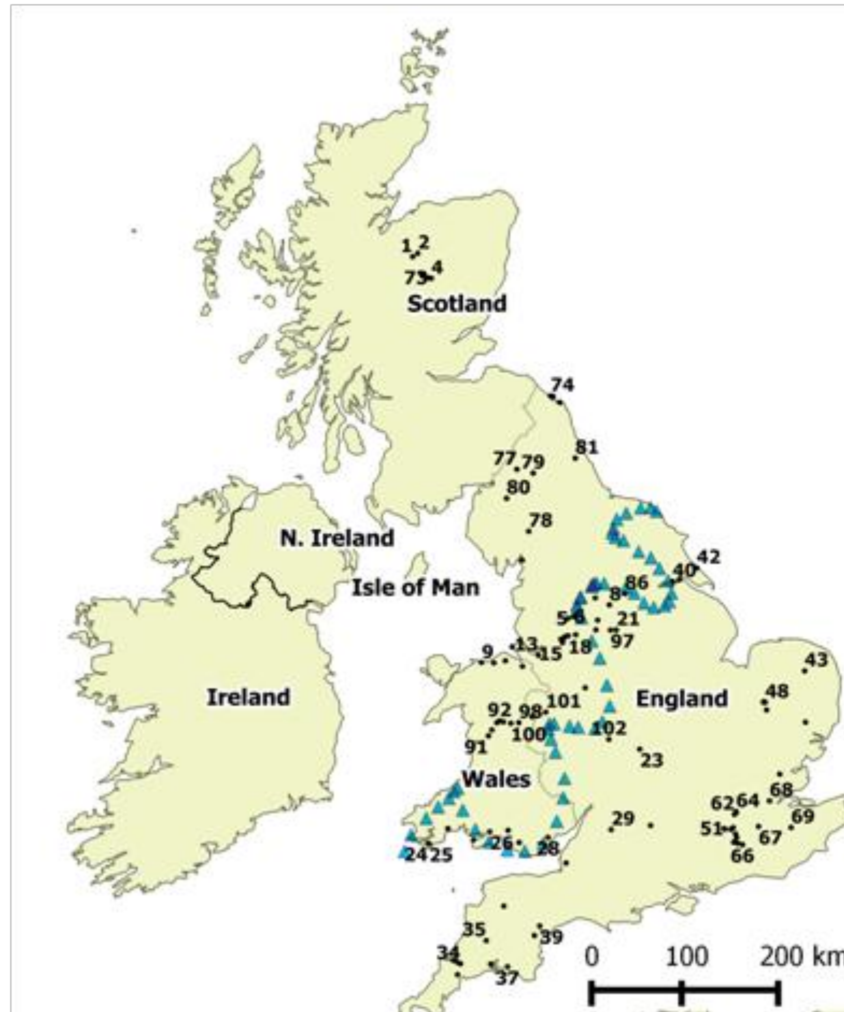


Figure 5.1. Distribution of sampling sites in Great Britain (a total of 98 sites were sampled – black dots, and *Lepidocyrtus* was sampled from 68 – labelled sites); a complete list of geographic coordinates and number of individuals sampled per locality can be found in Table S5.1. Blue triangles indicate the maximum extent of the British-Irish ice sheet during the last Pleistocene glacial period.

DNA was isolated from individual specimens using the DNeasy 96 well Blood and Tissue Extraction Kit (QIAGEN, West Sussex, UK). Manufacturer's instructions were used with modifications aimed to preserve the exoskeleton of each animal (for details see methods in chapter 4). After incubation, the PCR plates with remaining exoskeletons were refilled with absolute ethanol and stored as vouchers at 4°C.

The primers used to amplify a fragment of the cytochrome c oxidase subunit I (COI) of 658 base pairs (bp) were ColFol-for (5'-TTTCAACAAATCATAARGAYATYGG-3') and ColFol-rev (5'-TAAACTTCNGGRTGNCCAAAAAATCA-3'). ColFol-for is a modification of primer LCO1490 (Folmer et al., 1994) aimed to improve matching to Collembola and to have full degeneracy across the last three 3' codons (Ramirez-Gonzalez et al., 2013). ColFol-rev is a modification of primer HCO2198 (Folmer et al., 1994) also aimed to improve matching to Collembola (Ramirez-Gonzalez et al., 2013). Polymerase chain reactions (PCR) contained NH4 buffer (1x), 3.0 mM $MgCl_2$, 2.5 mM of each dNTP, 0.4 µM of each primer and 0.5 U of Taq polymerase (Bioline) in 25 µL final volume. PCR cycles were carried out using the following thermal profile: 95°C for 2 min, 40 cycles of 95°C for 1 min; 52°C for 45 s, 72°C for 1 min; and finally 72°C for 5 min. PCR products were cleaned with ExoRap protocol, normalized and sent to Eurofins for sequencing with the reverse primer.

5.3.2 Sequence alignment and molecular lineage delineation

MtDNA COI sequences were processed and ambiguous base calls manually assessed with Geneious Pro 5.6.4 (http://www.geneious.com, Kearse et al., 2012) and aligned using mafft 6.814 (Katoh et al., 2002). Haplotypes were collapsed using DNACollapser (Villesen, 2007). The total number of variable and parsimony informative sites, and uncorrected pairwise genetic distances (overall mean, minimum and maximum) both within and among lineages (the most frequent sequence with the minimum number of differences was selected to represent each lineage) were computed with MEGA6 (Tamura et al., 2013). Sequences were subjected to a neighbour-joining analysis (NJ) using p-distances in MEGA6 (Tamura et al., 2013), and their uncorrected genetic pairwise distances were also computed. Molecular lineages, here defined operationally as a cluster of similar sequences (Operational taxonomic units, OTUs), were delineated using as a threshold the minimum pairwise p-distance of 4% found among 19

Lepidocyrtus mtDNA COI lineages collected sympatrically in six sampling sites in Panama in a previous study by Cicconardi et al (2013) (for details see methods chapter 4). After identifying clusters, one sequence per lineage was submitted to the BOLD Identification System (http://www.boldsystems.org) to test for identical or nearly identical sequences sampled from outside the UK.

5.3.3 Molecular dating analysis

In order to obtain a temporal framework for the onset of diversification within geographically localized monophyletic groups found in the *Lepidocyrtus* NJ tree, node ages were estimated using Beast v. 1.7.4 (Drummond et al., 2012) under a relaxed clock method that assumes a lognormal distribution of rates. The mtDNA COI rate of 0.0504 substitutions/site/Ma, obtained after quantitatively comparing the root ages of eight Panamian *Lepidocyrtus* COII sequences and their corresponding COI sequences (for details see chapter 4), was used following a normal prior distribution. Best-fit substitution models were selected independently for each monophyletic group using Bayesian information criterion (BIC) in MEGA6 (Tamura et al., 2013). The tree prior followed the Yule process speciation model. Five independent Markov Chain Monte Carlo searches were run for 10^8 generations, sampling every 1000 generations. Stationarity, convergence, and expected sampling sizes (ESS) for the posterior distribution and parameters were evaluated using Tracer v1.6 (Rambaut et al., 2014).

5.3.4 Geographical distribution and haplotype diversity analysis

To assess geographical patterns in the distribution of genetic diversity, the localities of occurrence of each lineage haplotype were mapped using QGIS (QGIS Development Team, 2014), then visually inspected for non-random patterns of geographical distribution, i.e., disjunction distributions and monophyletic clades of mtDNA sequences limited to geographically closely located sites. Arbitrary cut-off points (based on the mean and maximum number of sites and number of individuals sampled per lineage) were used to describe spatial patterns of distribution and abundance levels of lineages. Spatial distributions were classified as widespread = lineage with haplotypes sampled from more than 40 sites with ranges widely distributed across the geographic extent of sampling sites; spread = lineage with haplotypes sampled from across 20 to 39

sampling sites with ranges relatively well distributed across most of the sampling sites; localised = lineage with haplotypes restricted to a few closely located sampling sites; disjunctly distributed = linage sampled from a few sampling sites located in opposite areas (e.g. east and west) of the geographic extent of the sampling sites. Abundance levels were classified as low abundance = 2 to 10 individuals, medium abundance = 11 to 30 individuals, high abundance = 31 to 50 individuals, very high abundance = above 51 individuals. To evaluate differences in genetic diversity between glaciated and unglaciated areas, abundance matrices with lineages (OTUs) as rows and localities as columns were created and imported to EcoSim v7.72 (Gotelli & Entsminger, 2000). Analyses were also performed for genetic diversity within each lineage, in which haplotypes were used as rows in the abundance matrices. Localities were grouped into two biogeographical areas: glaciated and unglaciated and they were defined according to the maximum limits of the British-Irish ice sheet (BIIS) during the last Pleistocene glacial period reported in Clark et al. (2012). To test whether genetic diversity (total lineage richness and haplotype richness within each lineage is equivalent between these two areas, controlling for differences in sampling and abundances, a rarefaction algorithm as implemented in EcoSim v7.72 (Gotelli & Entsminger, 2000), was used to rarefy the larger community down to the abundance level of the smaller. The total abundance of the smaller community being compared was used as the single abundance threshold for each analysis. Simulations were run using the "Species Richness" as the species diversity index. The random number seed was set to 10 and simulations were run for 1000 iterations. Diversity curves and their 95% confidence intervals were inspected to check whether assemblages were significantly different from one another. The mean and variance of diversity generated in each simulation were used to formally test the hypothesis of equivalent species diversity for glaciated and unglaciated communities by checking whether the observed diversity of the smaller community felt within the 95% confidence interval of the simulations for the rarefied community.

5.4 RESULTS

5.4.1 Overview of *Lepidocyrtus* in Great Britain

A total of 428 specimens were sampled from the genus *Lepidocyrtus* from across 68 sampling sites (Tab S5.1, Fig 5.1), which were assigned to 4 common morphospecies (*Lepidocyrtus curvicollis, L. cyaneus, L. lanuginosus, L. lignorum*), two unknown morphospecies (*Lepidocyrtus* sp. 1 cf. orange, *Lepidocyrtus* sp. 2) plus one specimen of each of *L. ruber* and *L. violaceus*. One *Katianna* sp. sequence was sampled as an outgroup. Sequences from the mitochondrial gene COI were obtained from 391 individuals, while the remaining 37 specimens either failed to amplify or produce sequences of good quality traces. The final COI alignment, with 392 sequences, was 556 bp long and yielded 92 unique haplotypes. Across the ingroup, there were 240 polymorphic sites of which 225 were parsimony informative. Using 96% similarity as the threshold for clustering sequences down to molecular lineages, a total of 22 OTUs were identified of which five were singletons (Fig 5.2). The minimum, maximum and overall mean divergence among lineages was 4.5%, 23.9% and 14.4% respectively. Within lineages, overall mean divergences ranged from 0.02% to 2.36% and maximum divergences ranged from 0.18% to 4.68% (Tab 5.1). Sequences assigned to the morphospecies *L. cyaneus* and *L. lanuginosus* corresponded to nine and six divergent mtDNA COI lineages respectively, while those assigned to *L. lignorum* and *L. curvicollis* corresponded to three and two divergent COI lineages respectively (Tab 5.1, Fig 5.2). Seven out of the 22 lineages are a mixture of morphospecies but they are mainly represented by one morphospecies (e.g. most individuals of lineage 3 are *L. lignorum*), apart from lineage 1 which is the most abundant (n=189 specimens) and widespread (n=41 sites) lineage with very low sequence variation (only 9 haplotypes), equally represented by two morphospecies, *L. lanuginosus* and *L. lignorum* (n=83 and n=82 respectively), plus a few specimens of *L. curvicollis*, *L. cyaneus* and *L. ruber*. The number of haplotypes per lineage ranged from 1 to 14 with an average of 4 haplotypes per lineage (Tab 5.1). The geographic distribution of the 13 lineages with >99% similarity to BOLD sequences revealed the presence of related sequences mainly in Canada and France, but also in Germany, Italy, Poland, USA, Australia and Tasmania (Tab 5.1).

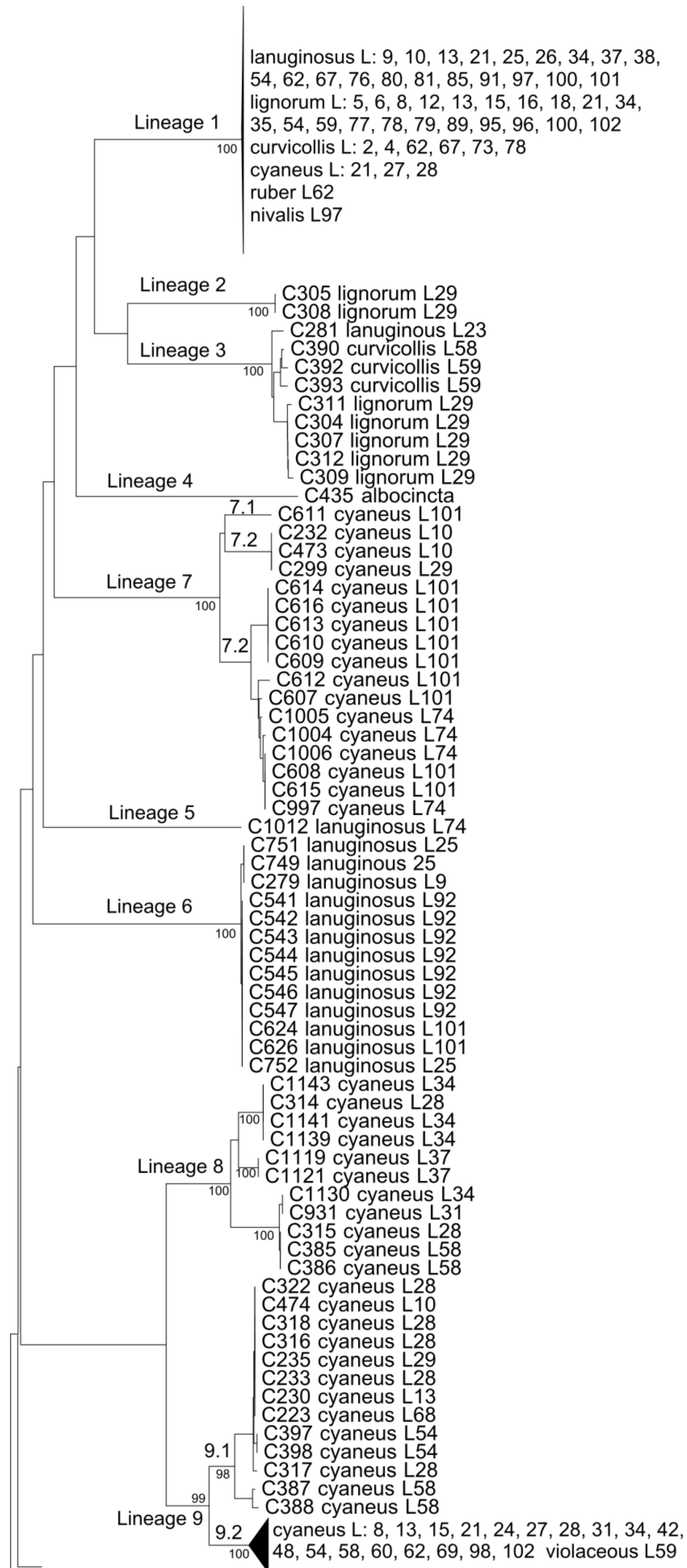

lanuginosus L: 9, 10, 13, 21, 25, 26, 34, 37, 38,
54, 62, 67, 76, 80, 81, 85, 91, 97, 100, 101
lignorum L: 5, 6, 8, 12, 13, 15, 16, 18, 21, 34,
35, 54, 59, 77, 78, 79, 89, 95, 96, 100, 102
curvicollis L: 2, 4, 62, 67, 73, 78
cyaneus L: 21, 27, 28
ruber L62
nivalis L97
Lineage 1
100
Lineage 2
C305 lignorum L29
C308 lignorum L29
100
C281 lanuginous L23
Lineage 3
C390 curvicollis L58
C392 curvicollis L59
C393 curvicollis L59
100
C311 lignorum L29
C304 lignorum L29
C307 lignorum L29
C312 lignorum L29
C309 lignorum L29
Lineage 4
C435 albocincta
7.1
C611 cyaneus L101
7.2
C232 cyaneus L10
C473 cyaneus L10
C299 cyaneus L29
Lineage 7
100
C614 cyaneus L101
C616 cyaneus L101
C613 cyaneus L101
C610 cyaneus L101
7.2
C609 cyaneus L101
C612 cyaneus L101
C607 cyaneus L101
C1005 cyaneus L74
C1004 cyaneus L74
C1006 cyaneus L74
C608 cyaneus L101
C615 cyaneus L101
C997 cyaneus L74
Lineage 5
C1012 lanuginosus L74
C751 lanuginosus L25
C749 lanuginous 25
C279 lanuginosus L9
Lineage 6
C541 lanuginosus L92
C542 lanuginosus L92
100
C543 lanuginosus L92
C544 lanuginosus L92
C545 lanuginosus L92
C546 lanuginosus L92
C547 lanuginosus L92
C624 lanuginosus L101
C626 lanuginosus L101
C752 lanuginosus L25
C1143 cyaneus L34
C314 cyaneus L28
100
C1141 cyaneus L34
C1139 cyaneus L34
Lineage 8
C1119 cyaneus L37
100
C1121 cyaneus L37
100
C1130 cyaneus L34
C931 cyaneus L31
100
C315 cyaneus L28
C385 cyaneus L58
C386 cyaneus L58
C322 cyaneus L28
C474 cyaneus L10
C318 cyaneus L28
C316 cyaneus L28
C235 cyaneus L29
C233 cyaneus L28
C230 cyaneus L13
C223 cyaneus L68
C397 cyaneus L54
9.1
C398 cyaneus L54
98
C317 cyaneus L28
99
C387 cyaneus L58
C388 cyaneus L58
Lineage 9
9.2
cyaneus L: 8, 13, 15, 21, 24, 27, 28, 31, 34, 42,
48, 54, 58, 60, 62, 69, 98, 102 violaceous L59
100

Fig 5.2 (cont)

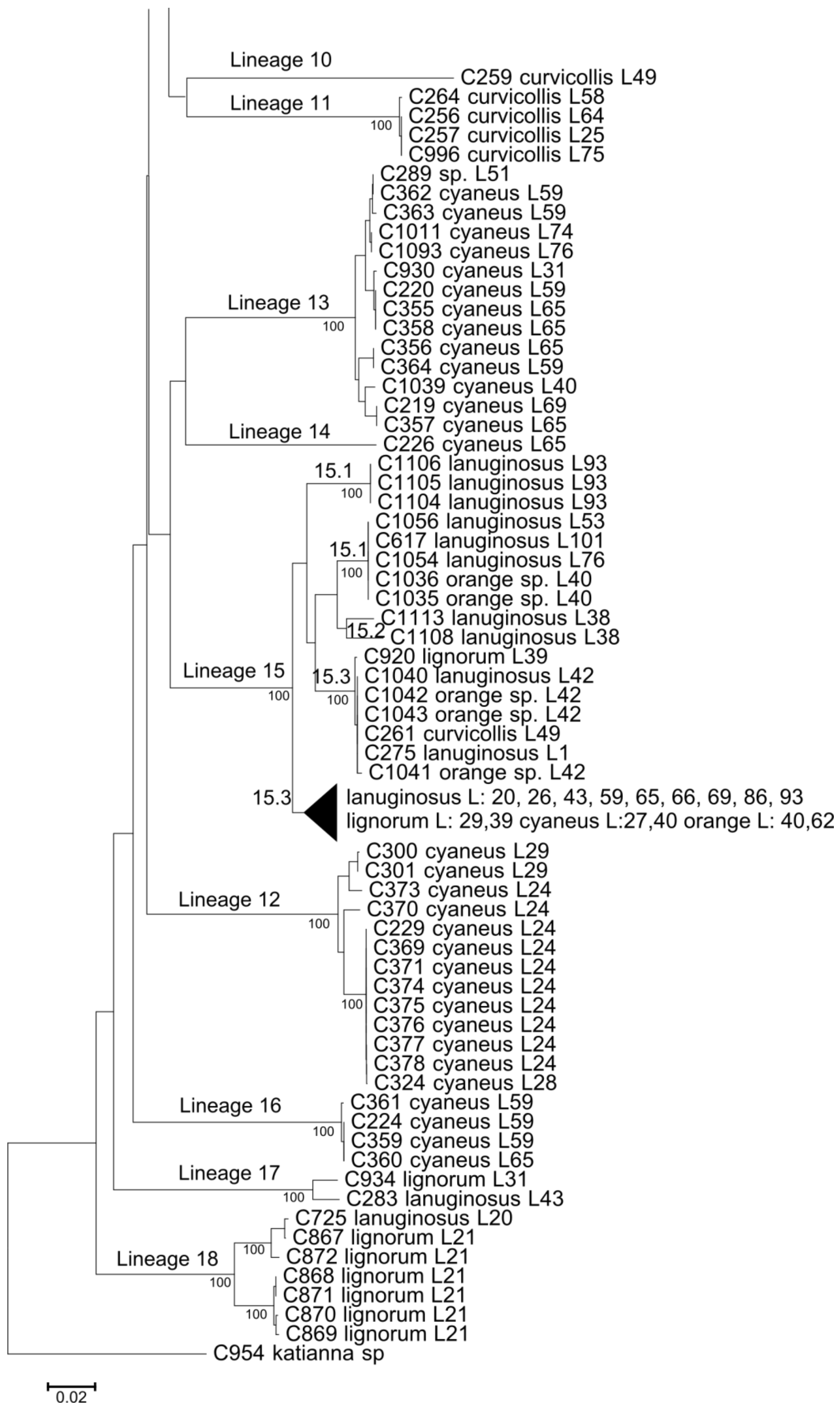


Figure 5.2. Neighbour joining tree of 392 *Lepidocyrtus* mtCOI Sanger sequences derived from 68 sampling sites across Great Britain (see Fig 5.1). A threshold of 96% similarity was used to cluster sequences into molecular lineages (see methods for details). *Katianna* sp. was sampled as an outgroup. Numbers immediately to the right of morphospecies names correspond to sampling localities. Most abundant lineages are collapsed. Further lineage details can be found in Table 5.1

5.4.2 Spatial distribution of lineages

In total, 34 sites were sampled within each of the glaciated and unglaciated areas of the UK (Fig 5.1). Across these sites, three main spatial distribution patterns were identified: lineages either presented spread/widespread ranges (n=3 lineages), disjunct distributions (n=3), or were geographically restricted to a few closely located sites (n=7) (Tab 5.1). The remaining five lineages were found in unique localities and were mainly represented by a single individual (Fig 5.2). *Lepidocyrtus cyaneus* lineages 7.1 and 7.2, *L. cyaneus lineages* 9.1 and 9.2, and *L. lanuginosus* lineages 15.1, 15.2 and 15.3 were considered as single lineages (lineages 7, 9 and 15, respectively) to take into account small sample sizes. The ranges of the widespread lineages 1 and 15 extended across the geographic extent of sampling sites, while lineage 9 has not been sampled from more northerly areas and it is largely distributed from West Yorkshire down to eastern, western and southern regions (Fig 5.3). The following disjunctive distributions were found: North, Wales and South (lineage 11), East Anglia and South West (lineage 17), North, Humberside, South, South East and South West (lineage 13) (Fig 5.4). Within lineages with geographically localised ranges, lineage 3 and 16 are only found in unglaciated sites, while the other lineages (6, 7, 8, 12 and 18) are found in both glaciated and unglaciated sites (Figs 5.5). Widespread lineages were also the most abundant (n > 40 individuals), while localised and disjunctly distributed lineages presented mid to low abundances (Tab 5.1). No pattern of genetic structuring was found within clades of the spread/widespread lineages. However, geographical structuring of genetic variation was found in five out of the seven geographically localised lineages (lineages 3, 7, 8, 12 and 18). These lineages (comprised of 7 to 17 individuals)

presented some level of haplotype variation (ranging from four to eight haplotypes per lineage), which were not found anywhere else but only in a few closely located sites (ranging from two to five sites per lineage) that span both glaciated and unglaciated terrains (Tab 5.1). Glaciated sampling sites were largely located near the limits of the ice sheet extents except for lineage 7 which extended well into the north. Although lineages 6 and 16 were also geographically restricted, they did not present sequence variation.

5.4.3 Species diversity analysis

To assess whether *Lepidocyrtus* genetic diversity in the UK displays a pattern of 'southern richness and northern purity', overall lineage richness and intra-lineage haplotype richness (haplotypes within each of the three widespread lineages) were compared between glaciated (northern) and unglaciated (southern) areas, after controlling for sampling and abundance differences in Ecosim (Gotelli & Entsminger, 2000). This analysis revealed that lineage richness is significantly different between glaciated and unglaciated communities with greater richness found in southern areas (Tab 5.2). After rarefying the unglaciated community data (n=205) down to the abundance level of the glaciated community (n=184), it was found that for 1000 random samples of 184 individuals, there was an average of 16.5 lineages with a variance of 0.4 (Tab 5.2). The confidence interval indicates that 95% of the times that a random sample of 184 individuals is drawn from the unglaciated community, it is expected to find between 15-17 different lineages. The observed diversity of the glaciated community (n=11 lineages) is outside this confidence interval and is below the simulated average, indicating that lineage richness is smaller than expected in the glaciated areas. Looking within the three widespread lineages (1, 9 and 15), observed haplotype richness of the less abundant community (glaciated community for lineages 9 and 15; and unglaciated community for lineage1, Tab 5.2) fell within the simulated 95% confidence interval for all lineages, indicating that haplotype richness does not differ between glaciated and unglaciated communities for these lineages (Tab 5.2).

Table 5.1. Summary data for each lineage, reporting: the number of individuals (n ind), number of mtDNA COI haplotypes (n hap) found, number of sites each lineage was sampled from (n sites), number of localities sampled in glaciated (n glac) and unglaciated (n unglac) areas, the mean and maximum observed intraspecific p-distances, and BOLD information on the locality of occurrence of the identical or near identical sequence found in this database.

Lineage	**Morphospecies**	**n ind**	**n hap**	**n sites**	**n glac**	**n unglac**	**Range**	**Mean/max intralineage p-distances**	**BOLD most frequent sequence**
1	*lanuginosus/lignorum/curvicollis/ cyaneus/ruber* [a]	189	9	41	25	16	widespread	0.02/0.5	Canada
2	*lignorum*	2	1	1	0	1	-	-	France
3	*lignorum/curvicollis/lanuginosus*[b]	9	7	4	0	4	localised	0.86/1.65	not provided
4	*albocincta*	1	1	1	0	1	-	-	<99% match
5	*lanuginosus*	1	1	1	1	0	-	-	<99% match
6	*lanuginosus*	13	2	4	3	1	localised	0.07/0.18	France
7.1	*cyaneus*	16	7	4	3	1		1.76/4.14	Canada
7.2	*cyaneus*	1	1	1	1	0	-	-	Canada
8	*cyaneus*	11	4	5	1	4	localised	2.36/4.14	Tasmania
9.1	*cyaneus*	13	5	7	3	4	spread	0.59/1.98	<99% match
9.2	*cyaneus/violaceus*[c]	36	10	19	5	14		0.59/1.80	Australia, Tasmania
10	*curvicollis*	1	1	1	0	1	-	-	France, Italy
11	*curvicollis*	4	2	4	1	3	disjunct	0.09/0.18	not provided
12	*cyaneus*	13	4	3	2	1	localised	1.09/2.54	no match
13	*cyaneus*/sp. 2[d]	14	8	8	3	5	disjunct	1.12/1.8	Canada, France, Germany
14	*cyaneus*	1	1	1	0	1	-	-	Canada

Table 5.1 (cont)

Lineage	Morphospecies	n ind	n hap	n sites	n glac	n unglac	Range	Mean/max intralineage p-distances	BOLD most frequent sequence
15.1	*lanuginosus*/sp. 1[e]	7	3	4	3	1		1.6/3.6	Canada
15.2	*Lanuginosus*	3	1	1	0	1	spread	-	<99% match
15.3	*lanuginosus/lignorum/cyaneus/ curvicolis*/ sp.1[f]	43	14	17	7	10		1.9/4.68	Canada, USA, France, Poland
16	*cyaneus*	4	2	2	0	2	localised	0.09/0.18	no match
17	*lanuginosus/lignorum*[g]	2	2	2	0	2	disjunct	2.16/ -	no match
18	*lignorum*	7	6	2	1	1	localised	2.36/3.96	Canada

[a] *lanuginosus* (n=83), *lignorum* (n=82), *curvicollis* (n=13), *cyaneus* (n=9), *ruber* (n=1); [b] *lignorum* (n=5), *curvicollis* (n=3), *lanuginosus* (n=1); [c] *cyaneus* (n=35), *violaceus* (n=1); [d] *cyaneus* (n=13), *Lepidocyrtus* sp. 2 (n=1); [e] *lanuginosus* (n=5), *Lepidocyrtus* sp.1 (n=2); [f] *lanuginosus* (n=28), *Lepidocyrtus* sp.1 (n=7), *lignorum* (n=5), *cyaneus* (n=2), *curvicollis* (n=1); [g] *lanuginosus* (n=1), *lignorum* (n=1). Lineages 7.1 and 7.2, 9.1 and 9.2, 15.1, 15.2 and 15.3 are combined into single lineages for description of their ranges.

5.4.4 Long-term persistence analysis

To obtain a temporal framework for the onset of diversification within the geographically localised lineages with signatures of genetic structuring (which provides a rough approximation of the time since the most recent ancestor of these lineages were present in the UK, thus, indicating their temporal persistence), a conservative mtDNA COI rate of 0.0504 substitutions/site/Ma was used under a relaxed clock method (see methods for details). The best model of molecular evolution for lineages 3 and 8 was K2, while for lineages 7, 12 and 18 it was T92. Adopting the closest models in BEAST (HKY for K2 and TN93 for T92), the mean age of the MRCA for these lineages ranged from 112,800 years ago (HPD= 47,000-186,400) to 326,400 years ago (HPD=153,900-506,200) (Tab 5.3). These deep divergence times are more likely to be underestimates than overestimates (i.e. fast, mtDNA COI substitution rate was used), and they substantially predate the Last Glacial Maximum (which occurred between 27,000 and 15,000 years ago), indicating that sequences have persisted in the UK through the extreme climatic changes of this and previous periods. Apart from lineage 3, which is restricted to southern sampling sites, all the other lineages (7, 8, 12 and 18) were sampled from both glaciated and unglaciated sites.

Table 2. Summary of Ecosim data entry and diversity results (mean and 95% confidence interval – CI) for overall lineage richness and for each of the most abundant lineages. For all simulated pairs, the larger community was rarefied down to the abundance level of the smaller community being compared and the observed diversity (n hap) of the smaller community was tested to see if it felt within the estimated 95% confidence interval of the larger community. N ind = number of individuals, N lin/hap = number of lineages or of haplotypes, glac = community in the glaciated area, unglac = community in the unglaciated area.

	All lineages		**lineage 1**		**lineage 9**		**lineage 15**	
	glac	**unglac**	**glac**	**unglac**	**glac**	**unglac**	**glac**	**unglac**
N ind	184	205	117	72	15	34	19	34
N lin/hap	11	17	8	3	6	13	9	13
Mean	-	16.5	6	-	-	7	-	9
95% CI	-	15 to 17	3 to 8	-	-	4 to 9	-	7 to 11

Table 5.3. Estimated times, in thousands of years, of the MRCA of *Lepidocyrtus* mitochondrial lineages sampled across the UK expressed as mean values with 95% highest posterior density (HPD) intervals.

Lineage	Mean value	95% HPD
Lineage 3	112,800	47,700-186,400
Lineage 7	306,000	110,400-494,600
Lineage 8	211,200	66,000-398,800
Lineage 12	164,100	86,500-240,000
Lineage 18	326,400	153,900-506,200

5.5 DISCUSSION

The aim of this chapter was to investigate support for prior evidence of soil dwelling fauna survival through Pleistocene glaciations in the continental island of Great Britain. By further sampling within the Collembola genus *Lepidocyrtus* with the mtDNA COI gene, we assessed lineage richness between glaciated and unglaciated communities, evaluated signatures of disjunct distributions and structuring of genetic variation, and estimated the ages of the onset of diversification within geographically localised lineages as complementary approaches to detect long-term persistence of soil dwelling fauna in Great Britain. Lower than expected lineage richness was found for the glaciated community, suggesting it to follow the 'southern richness and northern purity' pattern. Furthermore, the finding of lineages with disjunct distributions and of lineages with genetic variation geographically structured within a few sites, for which estimated ages largely predate the Last Glaciation Maximum, suggests long term persistence for *Lepidocyrtus* (which is in line with previous findings for the Collembola genus *Entomobrya*), thus indicating that survival through abrupt climatic changes in Great Britain may be more common than previously thought for soil dwelling fauna.

5.5.1 Molecular lineages and geographic distributions

Glacial events greatly altered distributional ranges of European species as ice or permafrost conditions pushed organisms southward or gradually constrained them to various refuges (e.g. Taberlet et al., 1998; Hewitt, 2004). As a consequence, high levels of endemic genetic variation are found in refugial areas (considered as strong evidence

for long-term *in situ* evolution) (e.g. Hewitt, 2000; Provan & Bennett, 2008; Tzedakis et al., 2013), and several taxa show disjunctive modern distributions (e.g Stehlik et al., 2002; Willis & van Andel, 2004; Hedderson & Nowell, 2006; Schmitt et al., 2006). These two basic arguments – endemism and disjunct isolation, have been used to support the glacial survival hypothesis (e.g Brochmann et al., 2003) and they were investigated within the *Lepidocyrtus* genus in order to detect lineages with signatures of long-term persistence. Five mtDNA COI *Lepidocyrtus* lineages presented sequence variation geographically restricted to a few closely located sites (Figs. 5.2 and 5.5) pointing to the existence of endemic genetic variation that has diversified *in situ* within discrete areas in the UK. In total, these lineages represented twenty-nine haplotypes distributed across 18 geographically proximate sampling sites (Figs 5.5). Sequence variation within these sites is mostly concentrated across unglaciated terrains but it also spans into glaciated areas (largely located near the border of the ice sheet except for lineage 7 for which sequence variation extends well into the northern glaciated area). Furthermore, three other lineages presented disjunct geographical distributions mainly between North-South and East-West regions (Fig 5.4). While this disjunctive pattern could be explained by long term dispersal after ice sheet retreat, as noted for other taxa previously studied in different contexts (e.g. Brochmann et al., 2003; Beatty & Provan, 2013; Tzedakis et al., 2013), it seems unlikely for *Lepidocyrtus* (and Collembola in general) due to their limited dispersal capability. This low dispersive ability has been well demonstrated by the strong geographic structuring across very short distances found for *Lepidocyrtus* lineages in previous molecular studies in the Mediterranean Basin and Panama (Cicconardi et al., 2010, 2013), also for other Collembola species studied elsewhere (e.g. Garrick et al., 2007, 2008; McGaughran et al., 2008). A more probable scenario is that these could be native lineages whose origin in Great Britain predates the last glaciation, for which previously broader distributions were reduced by glacial onset, with persistence in a few favourable micro-niches (ice free areas) in otherwise unfavourable habitats. After glaciations, as the ice receded, deglaciated regions would have been rapidly recolonised by a few leading front lineages that effectively pre-empted space, and reduced the chances of the surviving lineages to re-expand their populations.

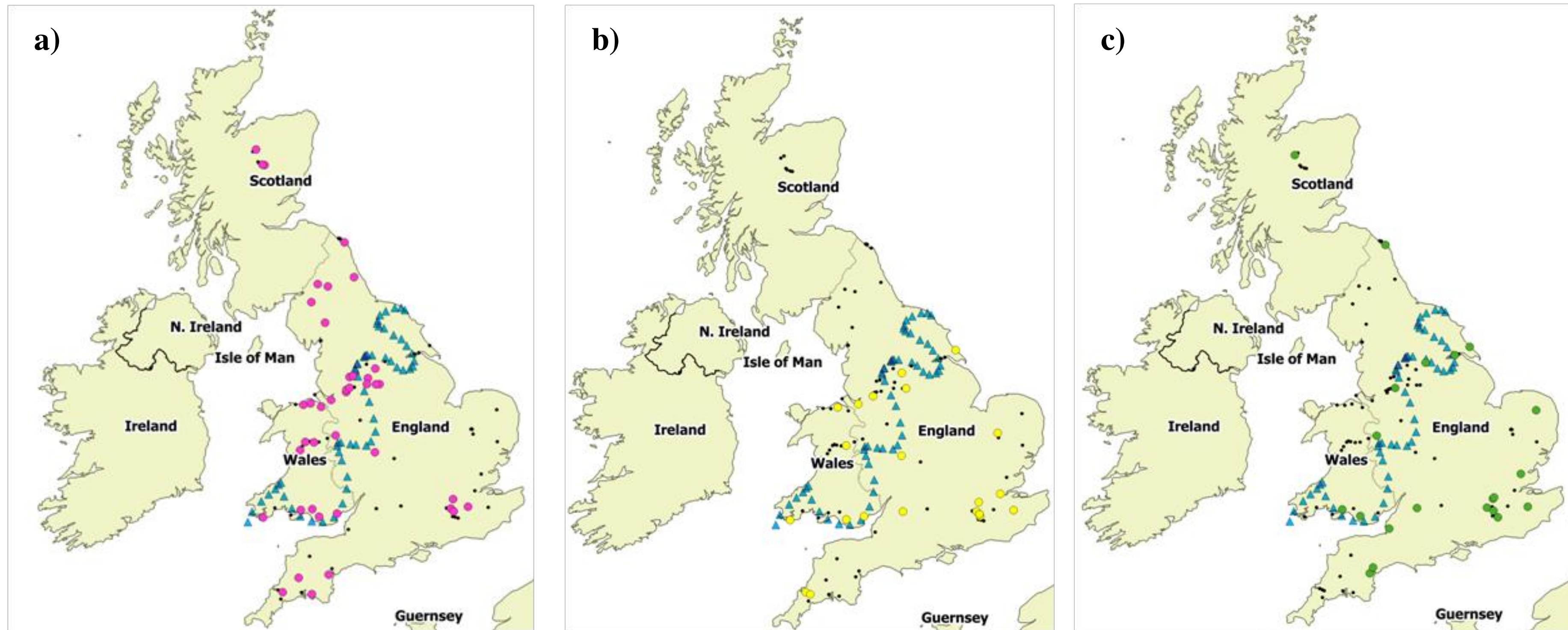
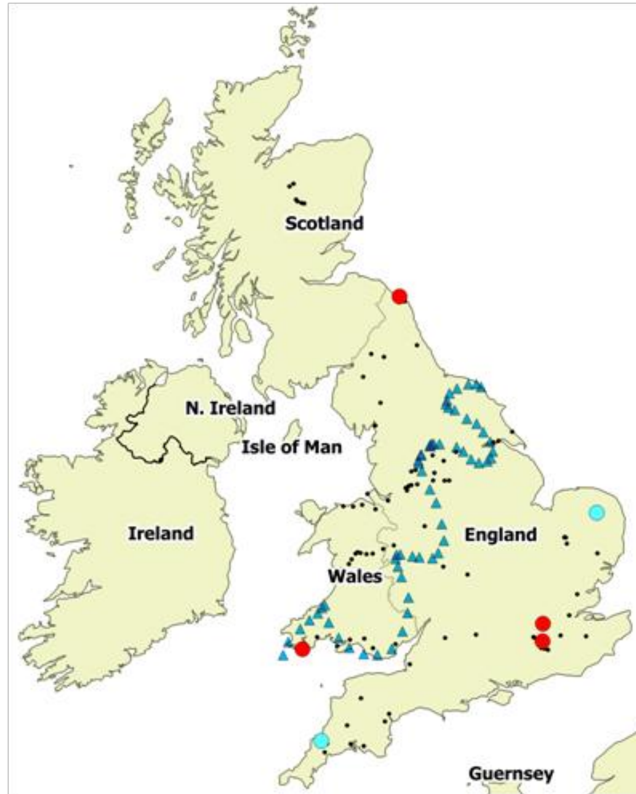


Figure 5.3. Geographical distributions of widespread *Lepidocyrtus* a) Lineage 1 – pink dots, b) lineage 9 – yellow dots, c) lineage 15 – green dots. Small black dots indicate remaining sampling sites. Blue triangles indicate the maximum extent of the British-Irish ice sheet during the last Pleistocene glacial period.

a)



b)

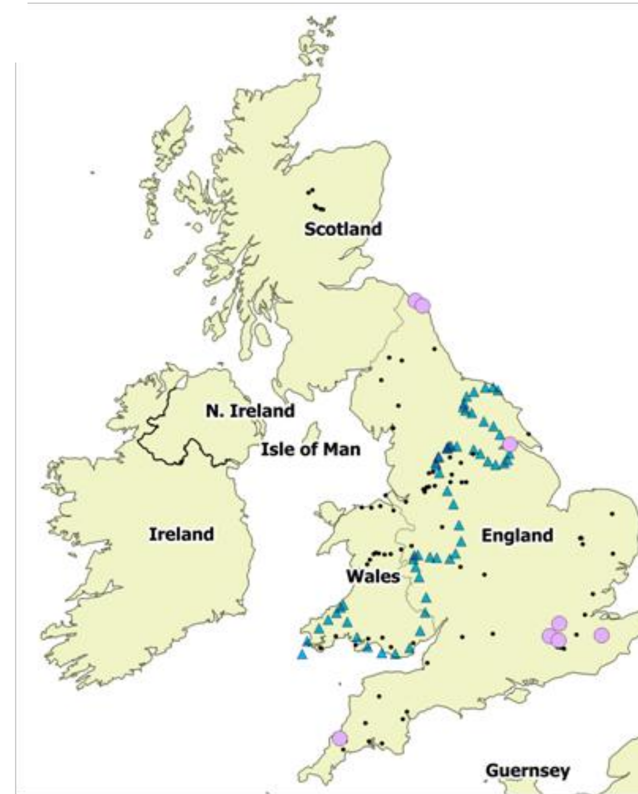
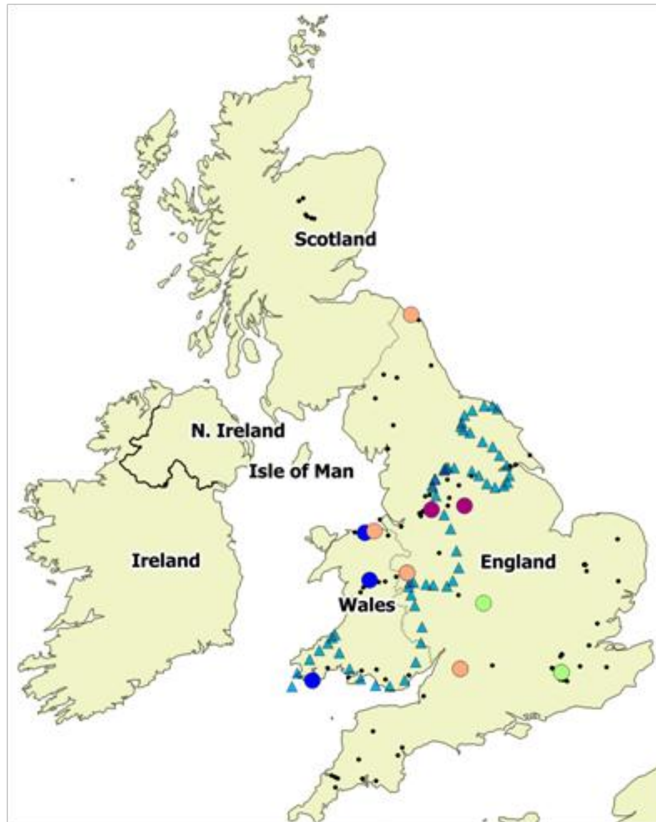


Figure 5.4 Geographical ranges of *Lepidocyrtus* lineages with disjunctive distributions a) Lineage 11 – red, lineage 17 – blue dots; b) lineage 13 – purple dots. Small black dots indicate remaining sampling sites. Blue triangles indicate the maximum extent of the British-Irish ice sheet during the last Pleistocene glacial period.

a)



b)

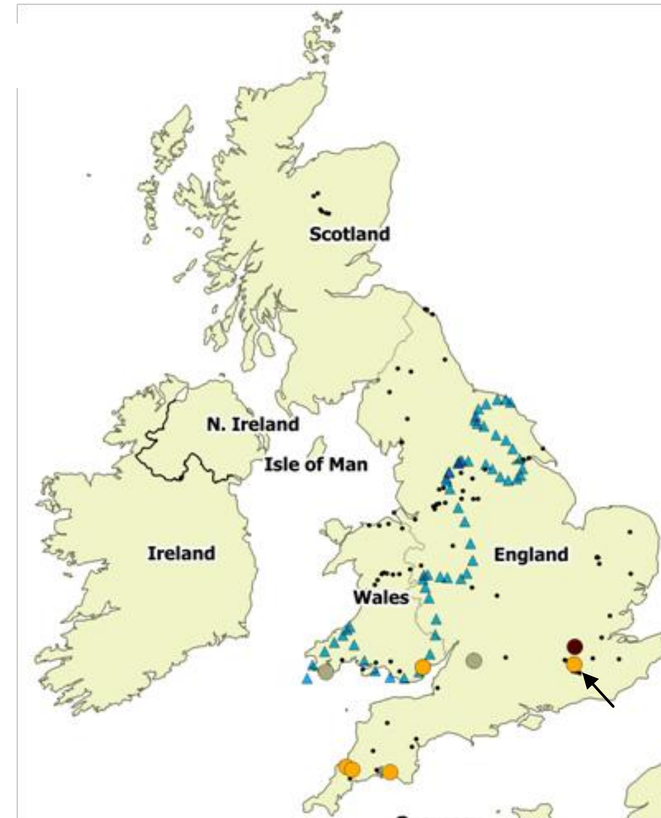


Figure 5.5 Geographical ranges of *Lepidocyrtus* lineages with geographically localized distributions. a) Lineage 3 – light green, lineage 6 – dark blue, lineage 7 – light pink, lineage 18 – dark purple dots; b) lineage 8 – orange, lineage 12 – grey, lineage 16 – brown dots. Black arrow indicates another site for lineage 16, which is shared with lineage 8. Small black dots indicate remaining sampling sites. Blue triangles indicate the maximum extent of the British-Irish ice sheet during the last Pleistocene glacial period.

5.5.2 Species richness analysis

A general prediction emerging from many studies that have explored the influence of glacial and postglacial events on the geographic structure of genetic variation in Europe is that colonisation, after the most recent glacial period, would have left a signature of genetic poverty in northern regions (Hewitt, 1999). While the longer history of southern European refugial areas would have led to greater genetic diversity in those regions (Hewitt, 1996, 1999). The prediction of 'southern richness and northern purity' (which can be seen at the level of species numbers, subspecific division and allelic variation) (Hewitt, 1999) was tested with Great Britain, which was mostly covered by the ice sheet during the LGM but remained free from ice in its southerly regions. *Lepidocyrtus* lineage richness was found to be different between glaciated and unglaciated communities, with a smaller than expected number of lineages in the glaciated areas, indicating recolonisation of the north by a subset of species from the south. This result strengthens the evidence of long-term persistence for soil fauna that previously did not find northern poverty signatures for *Entomobrya* communities, potentially due to the small number of lineages found within this genus. Placing these findings into the European context, both studies extend the list of species that survived in high latitude refugia confirming a more mosaic model of persistence during Pleistocene glaciations, which involves not only southern peninsulas but also northern refugia areas. The unglaciated southern parts of the UK have been previously suggested to act as northern cryptic refugia, e.g. fossil remains (Vincent, 1990), fish (Bernatchez, 2001; Hänfling et al., 2002), and groundwater amphipod (McInerney et al., 2014). Results presented here, and in the previous chapter, go beyond that as they indicate long-term persistence for Collembola lineages in sampling sites distributed not only in unglaciated areas but also far into the glaciated terrains, pointing to the existence of small ice-free areas that allowed survival within northern terrains.

One possible mechanism for the existence of northern ice-free habitats in Great Britain comes from recent work suggesting they may result from geothermal activity, i.e. they are located close to geothermal heated ground and water (e.g. heated ground and ponds, steam fields). These areas would have provided thermal gradients with less extreme temperatures (between very high temperatures in the core of geothermal sites, and very low temperatures below the ice-sheets), thus, possibly creating a buffer zone where life

could survive – the "geothermal glacial refugia" hypothesis (Convey & Lewis Smith, 2006). Such geothermal activity usually persists over extended time scales that could encompass glacial cycles and therefore could possibly provide refugia for life (Fraser et al., 2014; Pointing et al., 2014). Using Antarctica as a natural model system and a comprehensive terrestrial database associated with landscape, climate and geothermal data, Fraser et al. (2014) provided strong support for the role of geothermal heated terrain in structuring broad-scale contemporary patterns in Antarctic diversity by providing glacial refugia. They suggest that for many cases where the existence of small refugia within the glaciated regions of northern Europe and northern North America has been inferred from phylogographic data, but their causes and precise locations have not been identified, there is a great possibility that they may have been geothermal (Fraser et al., 2014). A consideration of the geothermal map of Great Britain (http://www.largeimages.bgs.ac.uk/iip/mapsportal.html?id=1004780) suggests that some of our sampling sites, where *Entomobrya* and *Lepidocyrtus* lineages present signatures of *in situ* diversification, coincide with areas around geothermal terrains. For example, the oldest *E. nivalis* lineages (chapter 4) were sampled from localities 76, 77, 78 and 79, which seem to coincide with terrains that receive heat flow from the Northern England Granites. Whereas sampling localities for *L. cyaneus* lineages 7 (loc: 10, 101) and 12 (loc: 24, 28) coincide with areas near thermal springs (e.g. taffs Well) and close to heat flow from the Cheshire Basin. This suggests that long-term persistence of Collembola in glaciated areas of the UK may have been provided by geothermal glacial refugia.

5.5.3 Long-term persistence of *Lepidocyrtus* in the UK

When reconstructing time frames for the evolution of endemic taxa, i.e. endemic variation that has diversified *in situ*, deep divergence times are a key indication of ancient origin and long-term persistence of individual lineages over a fine geographic scale, despite the occurrence of abrupt climatic changes. With divergences dating at least to the Ionian stage of the Pleistocene (~300-100 Kya), it was possible to detect *Lepidocyrtus* lineages differentiation and persistence in isolation over a time span that largely predates the Last Glacial Maximum. The implication is that the present distribution of the five geographically structured lineages has been influenced by relatively old paleogeographic events. These lineages survived through extreme climatic

and ecological changes that characterized Great Britain and remained isolated genetically from neighbouring lineages. This result corroborates previous study that detected signatures of old differentiation and persistence for one *Entomobrya* lineage in the UK. It strengthens evidence of long-term survival due to the greater number of geographically structured lineages found with deeper divergence ages compared to the *Entomobrya* genus finding. Both Collembola genera add to the small list of ancient invertebrate fauna in Great Britain (Nieberding et al., 2005; McInerney et al., 2014). They are also congruent with other studies showing soil dwelling fauna surviving harsh climatic conditions (e.g. Stevens et al., 2006; Garrick et al., 2007; Cicconardi et al., 2010) possibly due to their particular life history traits (e.g. small body size, abundant population in small areas, tolerance to cold) that confer them abilities to resist extreme conditions.

As has been discussed in other papers, “Putative refugia are expected to show higher genetic variability compared with surrounding recolonized regions. This is because populations surviving several glacial cycles should accumulate genetic variation (…). Thus, molecular variation within northern refugia would be expected to contain related and, in some cases, locally endemic alleles, distinct from surrounding regions and other refugial areas (Tzedakis et al., 2013, pg 698)”.

It is important to note that we did not date divergence times among closely related mtDNA lineages as it has been the case in other studies (e.g. extreme mtDNA divergence was found between lineages of the snail *Cepaea nemoralis* in the UK [(Thomaz et al. 1996, Goodacre et al. 2006)]). In fact, we cannot evaluate relationships between our Collembola lineages due to the nature of their mtDNA data, which is very saturated and thus uninformative). What we did was to estimate the time to the most recent common ancestors within lineages that are geographically localised with signatures of genetic structuring, i.e., locally endemic alleles, distinct from surrounding regions. In this case they do tell us about persistence times because they give us a temporal framework for the onset of diversification within these geographically localised lineages. In the case of *Cepaea nemoralis*, no geographic structure of endemic variation has been found within the sampling sites indicating that other drivers than persistence might have caused this deep divergence.

5.5.4 Conclusions, limitations and future work

Our results provided further evidence for lineage diversity differences between glaciated and unglaciated areas and further support for endemism and ancient fauna in the island of Great Britain. The expectations for the four genetic signatures [(i) southern richness and northern purity,' (ii) geographic structuring of genetic diversity, (iii) disjunct distributions and (iv) divergence times predating the LGM] predicted to be found in case of long-term persistence in the UK were met. Together with results for the genus *Entomobrya*, these findings increase the list of species presenting evidence for survival through extreme climatic changes in Great Britain, also providing new evidence for invertebrate persistence in northern European areas during the Pleistocene glaciations.

We argue that explanations of persistence/survival during the ice age are more likely to explain the genetic patterns found in this study for the UK Collembola than post-glaciation colonisation for a number of reasons. First, the finding of two non-random patterns of genetic distributions (i) endemic genetic variation (locally endemic alleles with geographically localised distributions) which have arisen in situ, and (ii) disjunct isolation, are themselves considered to be strong evidence for long-term persistence and they have been used by several authors as the two fundamental arguments to support the glacial survival hypothesis (e.g. Brochmann et al., 2003). Second, the estimated ages for the onset of diversification within geographically localised lineages demonstrate they largely predate the LGM, suggesting their long-term persistence. Similar old ages (deep divergence times within geographically structured lineages) have also been found in other Collembola studies (Australia, Mediterranean, Antartic) and have been indicated as strong evidence for their survival through abrupt climatic and ecological changes (Stevens et al. 2006, Garrick et al. 2007, Cicconardi et al. 2010). Third, the very restricted dispersal ability of Collembola reduces the possibility of long distance migration in explaining disjunct distributions of lineages and this low vagility has already been correlated with survival in small refugial areas (Moritz et al., 2001; Hugall et al., 2002). This trait combined with their small body sizes (only a small area of ice-free habitat is required for survival), high population densities (which confers local demographic stability), and ubiquity (indicate they naturally present a degree of tolerance to inhabit a range of habitats, including those with extreme conditions)

indicate that Collembola naturally hold traits that are expected to promote long term persistence in northern regions greatly affected by ice ages.

Identifying ancestral and derived alleles and exploring their geographic distribution (e.g. Miraldo et al., 2011) could help strengthen the species richness analysis. This is because a high incidence of ancestral haplotypes is expected to be found in refugial areas whereas a greater frequency of derived haplotypes are expected to be found in recently colonised area (Wakeley, 2008). Thus, it could be interesting to assess whether *Lepidocyrtus* ancestral and derived haplotypes are mostly found in unglaciated (source areas) or glaciated areas. Furthermore, this analysis would greatly benefit from a wider sampling strategy in Europe to identify possible origins and compare richness in a broader scale (i.e. the diversity in continental Europe is expected to be greater than the diversity in South Great Britain, which in turn is expected to be greater than the diversity in North Great Britain), which is more likely to provide a more complete history of the evolution of this genus in the UK.

Regarding our species richness analysis, we have used classical rarefaction curves which have long been used to compare species richness among empirical samples that differ in the total number of individuals (e.g. Sanders 1968; Lee et al., 2007) or among sample-based datasets that differ in the total number of sampling units (e.g. Norden et al., 2009; Longino and Colwell 2011). However, Colwell et al. (2012) have recently discussed that the existing variance estimators for individual-based (classical) rarefaction (Heck et al., 1975) and for Coleman rarefaction (Coleman et al., 1982) are not appropriate for this purpose because they are conditional on the reference sample. They have also discussed that one of the limitations of traditional rarefaction method is that, in order to standardize comparisons with the smallest sample in a group of samples being compared, much of the information content of larger samples is “thrown away”. To deal with that, they developed a method that implements unconditional variance estimators (which assumes that the reference sample represents a random draw from a larger, but unmeasured, community or species assemblage) and integrates mathematically distinct approaches that allow the linking of interpolated (rarefaction) curves and extrapolated curves to plot a unified species accumulation curve (Colwell et al., 2012). They argue that this ability to link rarefaction curves with their corresponding extrapolated richness curves, complete with unconditional confidence

intervals (which corrects previous estimators), helps to solve the limitation of throwing away information content on large samples. They were able to demonstrate this with three empirical examples using beetle, tree and ant data (Colwell et al., 2012). This indicates that the rarefaction method used in this thesis (classical) is limited and may complicate interpretation of the results. A re-analysis of the data using this novel approach is thus warranted and it will follow future work.

Whilst this chapter combined sampling sites into two main groups - glaciated and unglaciated, it would be interesting to evaluate whether the different times of retreat (of the different sections of the British-Irish ice sheet) could have an effect on the richness analysis. Different retreat times result in different length of time an area became free of ice and thus available for recolonisation and it could leave a detectable legacy on the contemporary history of lineage richness (Hawkins & Porter, 2003; Montoya et al., 2007). For example, the historical pattern of glacial retreat in response to post-Pleistocene warming has left a signal in the contemporary richness gradient of trees in Europe and North America detectable after at least 14 000 Kya (Montoya et al., 2007). Similar analysis could improve our understanding of the influence of glaciation on the Collembola richness in Great Britain. Furthermore, while the present study provides a general picture of how glaciated and unglaciated areas differ in relation to their *Lepidocyrtus* communities, it does not directly evaluate the correlation between ice sheet retreat and richness pattern, as this was not within our aims. A combination of contemporary and historical variables would need to be acquired and modeled (Montoya et al., 2007; Hortal et al., 2011) to fully explain richness differences across North and South regions. We do not have enough data to perform this analysis but future work taking into account contemporary (e.g. environmental variables) and historical data (including the different times of ice retreat at the different sections of the ice sheet) could give a finer picture of the statistical contribution of historical glaciation on the contemporary pattern of Collembola lineage richness in Great Britain.

Genetic models and computer simulations have investigated the importance of different factors in understanding the effect of colonization on geographic structuring of genetic variation (e.g. Ibrahim et al., 1996; Diniz-Filho et al., 2014). Within continuous populations, when the number of individuals moving between populations is small and distances are relatively short, it is well known that chance can lead to spatially clustered

distributions of alleles even when barriers to dispersal or selection are absent (e.g. Crow & Aoki, 1982). Occasional extinction and recolonization are also thought to promote local genetic differentiation (Wright, 1984), and other parameters such as the history of colonists and forms of dispersal have also been demonstrated in simulations to affect population structure (e.g. Agnarsson et al., 2014). Computer simulations could also help to understand the origin of the current distribution patterns of *Lepidocyrtus* haplotypes found here. For example, if Great Britain was completely defaunated due to the ice coverage, and all Collembola were derived from the post-glacial continental recolonisation, we would expect to find lineages largely blended, but the data show non-random patterns of geographic distribution. Simulations taking into account scenarios with different sources of genetic variation (i.e. native or introduced from the mainland) could be performed to see whether similar patterns of genetic structures (i.e. geographically localised lineages and disjunct distributions) would be recovered after a specified number of non-overlapping generations. Three scenarios (1: UK empty, all lineages came from mainland; 2: UK full, all lineages are native; 3: UK mixed, lineages are a mixture of native and introductions) could be simulated using a constant colonization rate, taking into account their low vagility as mode of dispersal, and allowing for random mating, gene flow, extinction and recolonization.

Finally, a formal evaluation of the "geothermal glacial refugia" hypothesis (Convey & Lewis Smith, 2006) would be fruitful. It could help to explain persistence of Collembola in ice-free habitats in glaciated areas of Great Britain during the LGM. Under this hypothesis, a greater contemporary diversity is expected near geothermal sites than in nongeothermal areas. Also, a significant nestedness by distance of this diversity is predicted. To date, only a few studies have suggested possible mechanisms to explain glacial refugia in northern Europe, but none have formally tested the potential hypotheses put forward. For example, Benke et al. (2009) suggested that unfrozen ground within permafrost areas (named taliks) could have enabled the establishment of perennial springs by pouring sub-permafrost groundwater to the surface, and this could have allowed persistence of the *Bythinella* spring snail in northern refugia. McInnerney et al. (2014) suggested that the relative stability of groundwater environment might explain the persistence of amphipod Niphargids in British through changing climates. They also argue that there is little relation between modern-day distributions of *Niphargus* in the British Isles and geothermally heated waters, which makes unlikely

that geothermal heating of groundwater was the only factor enabling their survival during glacial periods. However, the role of geothermal activity in structuring biodiversity patterns has been demonstrated for Antartic species (Fraser et al., 2014), suggesting that similar influences could help to explain the patterns of genetic structuring and persistence found among British soil dwelling fauna.

5.6 REFERENCES

Agnarsson I., Cheng R.C., & Kuntner M. (2014) A Multi-Clade Test Supports the Intermediate Dispersal Model of Biogeography. *PLoS ONE*, **9**, e86780.

Beatty G.E. & Provan J. (2013) Post-glacial dispersal, rather than in situ glacial survival, best explains the disjunct distribution of the Lusitanian plant species *Daboecia cantabrica* (Ericaceae). *Journal of Biogeography*, **40**, 335–344.

Benke M., Brändle M., Albrecht C., & Wilke T. (2009) Pleistocene phylogeography and phylogenetic concordance in cold-adapted spring snails (*Bythinella* spp.). *Molecular Ecology*, **18**, 890–903.

Bennett K.D., Tzedakis P.C., & Willis K.J. (1991) Quaternary Refugia of North European Trees. *Journal of Biogeography*, **18**, 103-115.

Bernatchez L. (2001) The evolutionary history of brown trout (*Salmo trutta* L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. *Evolution; International Journal of Organic Evolution*, **55**, 351–379.

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

Bilton D.T., Mirol P.M., Mascheretti S., Fredga K., Zima J., & Searle J.B. (1998) Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society B: Biological Sciences*, **265**, 1219–1226.

Brochmann C., Gabrielsen T.M., Nordal I., Landvik J.Y., & Elven R. (2003) Glacial survival or tabula rasa? The history of North Atlantic biota revisited. *Taxon*, **52**, 417–450.

Cicconardi F., Fanciulli P.P., & Emerson B.C. (2013) Collembola, the biological species concept and the underestimation of global species richness. *Molecular Ecology*, **22**, 5382–5396.

Cicconardi F., Nardi F., Emerson B.C., Frati F., & Fanciulli P.P. (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology*, **19**, 386–400.

Clark J.S., Lewis M., & Horvath L. (2001) Invasion by extremes: population spread with variation in dispersal and reproduction. *The American Naturalist*, **157**, 537–554.

Coleman B.D., Mares M.A., Willig M.R., & Hsieh Y.-H. (1982) Randomness, Area, and Species Richness. *Ecology*, **63**, 1121–1133.Convey P. & Lewis Smith R. i. (2006) Geothermal bryophyte habitats in the South Sandwich Islands, maritime Antarctic. *Journal of Vegetation Science*, **17**, 529–538.

Colwell R.K., Chao A., Gotelli N.J., Lin S.-Y., Mao C.X., Chazdon R.L., & Longino J.T. (2012) Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. Journal of Plant Ecology, **5**, 3–21.

Cooper S.J. & Gessaman J. A. (2004) Thermoregulation and habitat preference in mountain chickadees and juniper titmice. The Condor, 106, 852–861.

Crow J.F. & Aoki K. (1982) Group selection for a polygenic behavioral trait: A differential proliferation model. *Proceedings of the National Academy of Sciences USA*, **79**, 2628–26631.

Davison J., Ho S.Y.W., Bray S.C., Korsten M., Tammeleht E., Hindrikson M., Østbye K., Østbye E., Lauritzen S.-E., Austin J., Cooper A., & Saarma U. (2011) Late-Quaternary biogeographic scenarios for the brown bear (*Ursus arctos*), a wild mammal model species. *Quaternary Science Reviews*, **30**, 418–430.

Deffontaine V., Libois R., Kotlík P., Sommer R., Nieberding C., Paradis E., Searle J.B., & Michaux J.R. (2005) Beyond the Mediterranean peninsulas: evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Molecular Ecology*, **14**, 1727–1739.

Diniz-Filho J.A.F., Soares T.N., & Telles M.P.D.C. (2014) Pattern-oriented modelling of population genetic structure. *Biological Journal of the Linnean Society*, **113**, 1152-1161.

Drummond A.J., Suchard M.A., Xie D., & Rambaut A. (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Elith J. & Leathwick J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

Emerson B.C., Cicconardi F., Fanciulli P.P., & Shaw P.J.A. (2011) Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2391–2402.

Folmer O., Black M., Hoeh W., Lutz R., & Vrijenhoek R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fonseca V.G., Carvalho G.R., Nichols B., Quince C., Johnson H.F., Neill S.P., Lambshead J.D., Thomas W.K., Power D.M., & Creer S. (2014) Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. *Global Ecology and Biogeography*, **23**, 1293–1302.

Fraser C.I., Terauds A., Smellie J., Convey P., & Chown S.L. (2014) Geothermal activity helps life survive glacial cycles. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 5634–5639.

Garrick R.C., Rowell D.M., Simmons C.S., Hillis D.M., & Sunnucks P. (2008) Fine-Scale Phylogeographic Congruence Despite Demographic Incongruence in Two Low-Mobility Saproxylic Springtails. *Evolution*, **62**, 1103–1118.

Garrick R.C., Sands C.J., Rowell D.M., Hillis D.M., & Sunnucks P. (2007) Catchments catch all: long-term population history of a giant springtail from the southeast Australian highlands — a multigene approach. *Molecular Ecology*, **16**, 1865–1882.

Gaston K.J. (2009) Geographic range limits: achieving synthesis. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1395–1406.

Goodacre S.L., Thomaz D., & Davies E.K. (2006) Mitochondrial gene diversity in *Cepaea* – population structure, history and positive selection. Biological Journal of the Linnean Society, **87**, 167–184.

Gotelli N.J. & Entsminger G.L. (2000) *EcoSim: Null models software for ecology. Version 5.0*. Acquired Intelligence Inc. & Kesey-Bear. Available from http://homepages.together.net/~gentsmin/ecosim.htm.

Hänfling B., Hellemans B., Volckaert F. a. M., & Carvalho G.R. (2002) Late glacial history of the cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. *Molecular Ecology*, **11**, 1717–1729.

Hawkins B.A. & Porter E.E. (2003) Relative influences of current and historical factors on mammal and bird diversity patterns in deglaciated North America. *Global Ecology and Biogeography*, **12**, 475–481.

Heck K.L., van Belle G., & Simberloff D. (1975) Explicit Calculation of the Rarefaction Diversity Measurement and the Determination of Sufficient Sample Size. *Ecology*, **56**, 1459–1461.

Hedderson T.A. & Nowell T.L. (2006) Phylogeography of Homalothecium sericeum (Hedw.) Br. Eur.; toward a reconstruction of glacial survival and postglacial migration. *Journal of Bryology*, **28**, 283–292.

Hewitt G.M. (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.

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

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

Hofreiter M. & Stewart J. (2009) Ecological Change, Range Fluctuations and Population Dynamics during the Pleistocene. *Current Biology*, **19**, R584–R594.

Hopkin S.P. (2007) *A Key to the Collembola (springtails) of Britain and Ireland*. Field Studies Council-AIDGAP, Shrewsbury, UK.

Hopper S.D. (2009) OCBIL theory: towards an integrated understanding of the evolution, ecology and conservation of biodiversity on old, climatically buffered, infertile landscapes. *Plant Soil*, **322**, 49–86.

Hortal J., Diniz-Filho J.A.F., Bini L.M., Rodríguez M.Á., Baselga A., Nogués-Bravo D., Rangel T.F., Hawkins B.A., & Lobo J.M. (2011) Ice age climate, evolutionary constraints and diversity patterns of European dung beetles: Ice age determines European scarab diversity. *Ecology Letters*, **14**, 741–748.

Hugall A., Moritz C., Moussalli A., & Stanisic J. (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 6112–6117.

Ibrahim K.M., Nichols R.A., & Hewitt G.M. (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. **77**, 282–291.

Jetz W. & Fine P.V.A. (2012) Global Gradients in Vertebrate Diversity Predicted by Historical Area-Productivity Dynamics and Contemporary Environment. *PLoS Biol*, **10**, e1001292.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., &

Drummond A. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Lee T.M., Sodhi N.S., & Prawiradilaga D.M. (2007) The Importance of Protected Areas for the Forest and Endemic Avifauna of Sulawesi (Indonesia). Ecological Applications, **17**, 1727–1741.

Li X.-S. & Wang D.-H. (2005) Seasonal adjustments in body mass and thermogenesis in Mongolian gerbils (*Meriones unguiculatus*): the roles of short photoperiod and cold. *Journal of Comparative Physiology B*, **175**, 593–600.

Longino J.T. & Colwell R.K. (2011) Density compensation, species composition, and richness of ants on a neotropical elevational gradient. *Ecosphere*, **2**, art29.

Marshall D. & Coetzee L. (2000) Historical biogeography and ecology of a Continental Antarctic mite genus, *Maudheimia* (Acari, Oribatida): evidence for a Gondwanan origin and Pliocene-Pleistocene speciation. *Zoological Journal of the Linnean Society*, **129**, 111–128.

McGaughran A., Hogg I.D., & Stevens M.I. (2008) Patterns of population genetic structure for springtails and mites in southern Victoria Land, Antarctica. *Molecular Phylogenetics and Evolution*, **46**, 606–618.

McGaughran A., Stevens M.I., & Holland B.R. (2010) Biogeography of circum-Antarctic springtails. *Molecular Phylogenetics and Evolution*, **57**, 48–58.

McInerney C.E., Maurice L., Robertson A.L., Knight L.R.F.D., Arnscheidt J., Venditti C., Dooley J.S.G., Mathers T., Matthijs S., Eriksson K., Proudlove G.S., & Hänfling B. (2014) The ancient Britons: groundwater fauna survived extreme climate change over tens of millions of years across NW Europe. *Molecular Ecology*, **23**, 1153–1166.

Miraldo A., Hewitt G.M., Paulo O.S., & Emerson B.C. (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, **11**, 170.

Montoya D., Rodríguez M.Á., Zavala M.A., & Hawkins B.A. (2007) Contemporary richness of holarctic trees and the historical pattern of glacial retreat. *Ecography*, **30**, 173–182.

Moritz C., Richardson K.S., Ferrier S., Monteith G.B., Stanisic J., Williams S.E., & Whiffin T. (2001) Biogeographical concordance and efficiency of taxon indicators for establishing conservation priority in a tropical rainforest biota. *Proceedings of the Royal Society of London B: Biological Sciences*, **268**, 1875–1881.

Mortimer E., Vuuren B.J. van, Lee J.E., Marshall D.J., Convey P., & Chown S.L. (2011) Mite dispersal among the Southern Ocean Islands and Antarctica before the last glacial maximum. *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 1247–1255.

Mortimer E., Van Vuuren B.J., Meiklejohn K.I., & Chown S.L. (2012) Phylogeography of a mite, *Halozetes fulvus*, reflects the landscape history of a young volcanic island in the sub-Antarctic. *Biological Journal of the Linnean Society*, **105**, 131–145.

Nieberding C., Libois R., Douady C.J., Morand S., & Michaux J.R. (2005) Phylogeography of a nematode (*Heligmosomoides polygyrus*) in the western Palearctic region: persistence of northern cryptic populations during ice ages? *Molecular Ecology*, **14**, 765–779.

Norden N., Chazdon R.L., Chao A., Jiang Y.-H., & Vílchez-Alvarado B. (2009) Resilience of tropical rain forests: tree community reassembly in secondary forests. Ecology Letters, **12**, 385–394.

Petit R.J., Aguinagalde I., Beaulieu J.-L. de, Bittkau C., Brewer S., Cheddadi R., Ennos R., Fineschi S., Grivet D., Lascoux M., Mohanty A., Müller-Starck G., Demesure-Musch B., Palmé A., Martín J.P., Rendell S., & Vendramin G.G. (2003) Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. *Science*, **300**, 1563–1565.

Pointing S.B., Bollard-Breen B., & Gillman L.N. (2014) Diverse cryptic refuges for life during glaciation. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 5452–5453.

Provan J. & Bennett K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, **23**, 564–571.

QGIS Development Team (2014) *QGIS Geographic Information System*. Open Source Geospatial Foundation Project. Available from http://qgis.osgeo.org.

Rambaut A., Suchard M.A., Xie D., & Drummond A.J. (2014) *Tracer v1.6*. Available from http://beast.bio.ed.ac.uk/Tracer.

Ramirez-Gonzalez R., Yu D.W., Bruce C., Heavens D., Caccamo M., & Emerson B.C. (2013) PyroClean: Denoising Pyrosequences from Protein-Coding Amplicons for the Recovery of Interspecific and Intraspecific Genetic Variation. *PLoS ONE*, **8**, e57615.

Ricklefs R.E. (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, **7**, 1–15.

Sanders H.L. (1968) Marine Benthic Diversity: A Comparative Study. The American Naturalist, **102**, 243–282.

Schmitt T., Hewitt G.M., & Müller P. (2006) Disjunct distributions during glacial and interglacial periods in mountain butterflies: Erebia epiphron as an example. Journal of Evolutionary Biology, **19**, 108–113.

Stehlik I., Blattner F.R., Holderegger R., & Bachmann K. (2002) Nunatak survival of the high Alpine plant *Eritrichium nanum* (L.) Gaudin in the central Alps during the ice ages. *Molecular Ecology*, **11**, 2027–2036.

Stevens M.I., Frati F., McGaughran A., Spinsanti G., & Hogg I.D. (2007) Phylogeographic structure suggests multiple glacial refugia in northern Victoria Land for the endemic Antarctic springtail *Desoria klovstadi* (Collembola, Isotomidae). *Zoologica Scripta*, **36**, 201–212.

Stevens M.I., Greenslade P., Hogg I.D., & Sunnucks P. (2006) Southern Hemisphere Springtails: Could Any Have Survived Glaciation of Antarctica? *Molecular Biology and Evolution*, **23**, 874–882.

Stewart J.R., Lister A.M., Barnes I., & Dalen L. (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 661–671.

Taberlet P., Fumagalli L., Wust-Saucy A.-G., & Cosson J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Tamura K., Stecher G., Peterson D., Filipski A., & Kumar S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Teacher A.G.F., Garner T.W.J., & Nichols R.A. (2009) European phylogeography of the common frog (*Rana temporaria*): routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity*, **102**, 490–496.

Thomaz D., Guiller A., & Clarke B. (1996) Extreme Divergence of Mitochondrial DNA within Species of Pulmonate Land Snails. *Proceedings of the Royal Society of London B: Biological Sciences*, **263**, 363–368.

Tzedakis P.C., Emerson B.C., & Hewitt G.M. (2013) Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, **28**, 696–704.

Ursenbacher S., Carlsson M., Helfer V., Tegelström H., & Fumagalli L. (2006) Phylogeography and Pleistocene refugia of the adder (*Vipera berus*) as inferred from mitochondrial DNA sequence data. *Molecular Ecology*, **15**, 3425–3437.

Valdiosera C.E., García N., Anderung C., Dalén L., Crégut-Bonnoure E., Kahlke R.-D., Stiller M., Brandström M., Thomas M.G., Arsuaga J.L., Götherström A., & Barnes I. (2007) Staying out in the cold: glacial refugia and mitochondrial DNA phylogeography in ancient European brown bears. *Molecular Ecology*, **16**, 5140–5148.

Villesen P. (2007) FaBox: an online toolbox for fasta sequences. *Molecular Ecology Notes*, **7**, 965–968.

Vincent P. (1990) *The Biogeography of the British Isles: An Introduction*. Routledge, London; New York.

Wakeley J. (2008) *Coalescent Theory: An Introduction*. Roberts and Company Publishers, Greenwood Village, Colorado.

Whittaker R.J., Nogués-Bravo D., & Araújo M.B. (2007) Geographical gradients of species richness: a test of the water-energy conjecture of Hawkins et al. (2003) using European data for five taxa. *Global Ecology and Biogeography*, **16**, 76–89.

Willis K.J. & van Andel T.H. (2004) Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews*, **23**, 2369–2387.

Wilson A.B. & Veraguth I.E. (2010) The impact of Pleistocene glaciation across the range of a widespread European coastal species. *Molecular Ecology*, **19**, 4535–4553.

Wright S. (1984) *Evolution and the Genetics of Populations, Volume 3: Experimental Results and Evolutionary Deductions*. University of Chicago Press, Chicago, IL.

5.7 Appendix

Table S5.1. Sampling sites within the island of Great Britain, reporting the geographic coordinates and number of *Lepydocyrtus* individuals collected per locality. Locations coded according to Fig 5.1.

Locality	Latitude	Longitude	N individuals
1	57.226940	-3.744723	1
10	53.270206	-3.332264	14
100	52.672260	-3.407084	17
101	52.722786	-2.839915	19
102	52.362823	-1.945560	8
12	53.191521	-3.080297	1
13	53.269020	-2.800899	12
15	53.348923	-2.387444	2
16	53.383045	-2.369771	2
18	53.403618	-2.281280	1
2	57.250828	-3.646657	2
20	53.400555	-2.160005	6
21	53.374790	-1.511842	12
23	52.215801	-1.503167	1
24	51.653542	-4.957234	11
25	51.647945	-4.937279	8
26	51.678989	-3.993226	4
27	51.534397	-3.574383	3
28	51.539238	-3.128327	17
29	51.511406	-2.158805	13
31	50.512501	-4.822230	5
34	50.470009	-4.720558	9
35	50.654724	-4.290828	3
37	50.376942	-4.034168	6
38	50.633892	-3.565275	5
39	50.716389	-3.465000	3
4	56.992771	-3.484993	6
40	53.723221	-0.479807	5
42	53.800648	-0.057395	5
43	52.623665	1.246741	13
48	52.414166	0.542738	1
49	51.718281	0.523698	2
5	53.567368	-2.232038	1
51	51.321110	-0.475282	1
53	51.316383	-0.464724	1
54	51.299717	-0.374997	18

Table S5.1. (cont)

Locality	Latitude	Longitude	N individuals
58	51.248577	-0.320651	7
59	51.248592	-0.320650	12
6	53.574394	-2.140705	1
60	51.215530	-0.314491	2
62	51.433052	-0.276102	11
64	51.456116	-0.244719	1
65	51.448338	-0.242504	7
66	51.135246	-0.249651	1
67	51.273663	0.042656	2
68	51.490414	0.285921	1
69	51.200829	0.518717	9
73	57.005009	-3.541238	2
74	55.684959	-1.838683	6
75	55.675194	-1.799497	1
76	55.606403	-1.705754	3
77	55.054352	-2.617450	9
78	54.451942	-2.606670	9
79	54.991104	-2.363755	4
8	53.621056	-1.544229	7
80	54.798607	-2.868619	2
81	55.056610	-1.632874	1
85	53.498901	-1.774891	3
86	53.698200	-1.264826	2
89	52.699707	-3.621308	10
9	53.267548	-3.517447	3
91	52.587330	-3.784479	8
92	52.711727	-3.575345	7
93	51.269722	-2.919990	4
95	51.665211	-3.712372	1
96	53.411362	-1.830386	8
97	53.383255	-1.602855	8
98	52.666122	-3.282838	1

Chapter 6 – General conclusions and discussion

Molecular phylogenetic tools are invaluable for the study of evolutionary processes, providing insights about the patterns of colonization and the diversification of species, which ultimately help to gain an understanding of the origin of species and community assemblage. In this thesis, I have applied molecular tools (mtDNA and nuclear DNA Sanger sequencing and high throughput parallel sequencing) to investigate the processes of colonisation and diversification of invertebrates (beetles and springtails) within two island scenarios: (i) the Canary Islands, an oceanic island system and (ii) Great Britain, a continental island setting with a long-term dynamic of geographic and climatic change. The aim of this thesis has been to address three gaps in our understanding about island colonisation, speciation, and biota assembly. Below, I present these gaps explaining how they were filled and how they fit into a broader evolutionary and conservation context. I also discuss future directions and new techniques that could help advance each work.

The first gap addressed was the possibility that genomic admixture among multiple founding lineages has featured in the recent history of diversification of a very species rich coleopteran genus – *Laparocerus*, in the Canary Islands (chapter 2). To that end, I have used a combination of sequence data from one mitochondrial and one nuclear gene, molecular dating techniques and the spatial and temporal context provided by an oceanic island system to infer the history of colonisation and differentiation within the *L. tessellatus* species complex. This species complex is comprised of nine closely related species and I have studied the relationships among individuals sampled from four different islands. I have also evaluated the fit of mtDNA and nuclear sequence data to a colonisation history where species are the product of a single founding event. Using the geographic context of the islands themselves, and relative temporal information from the gene trees, I was able to identify the geographic origin of the complex, and dismiss explanations of incomplete lineage sorting to reveal a history of colonisation and speciation involving genetic admixture. I found two instances of shared mtDNA variation among species from different islands: each of the single species on La Palma (*Laparocerus* sp1) and El Hierro (*L. bimbache*) were found to be the product of more than one colonisation event from more than one source island. In both cases nuclear

ITS2 data revealed these multiple colonisations to have been followed by genomic admixture. My results are consistent with the other few studies to date that have detected species within an insular setting to be the result of genomic admixture following two multiple distinct colonisation events (Shaw, 2002; Jordal et al., 2006; Nietlisbach et al., 2013; Garrick et al., 2014; Lamichhaney et al., 2015). Together, these studies indicate that genomic admixture among independent founding populations or species may be more common than previously thought for island biota, although its detection is not easy. These findings point to the possibility that admixture could be a driver of speciation itself and it could be a representative feature of the evolutionary process on islands (Emerson & Faria, 2014).

Genomic data from other species, mainly plants but also animal taxa, support the idea that hybridisation is an important factor in the origin of species. As reviewed by Soltis and Soltis (2009, pg 561) "recent developments in genomics are revolutionizing the way we see angiosperm genomes, demonstrating that […] hybridization has been an important force in generating angiosperm species diversity. Hybridization and polyploid formation continue to generate species diversity, with several new allopolyploids having originated just within the past century or so". Although recognized as a common phenomenon in plants (Abbott et al. 2010), the role of homoploid hybrid speciation has only recently gained support as a speciation mechanism for animal taxa (Gompert et al. 2006; Mallet 2007) and it remains controversial (Hochkirch 2013, Harrison and Larson 2014). This is because many think that homoploid hybrid speciation will not become an important mode of speciation in animals (Barton 2013; Servedio et al. 2013; Schumer et al. 2014). However, novel animal genomic data, especially from island systems, are providing strong support for genomic admixture in generating diversity (Shaw, 2002; Jordal et al., 2006; Nietlisbach et al., 2013; Garrick et al., 2014; Lamichhaney et al., 2015). Thus indicating that hybridization among independent founding populations or species may be more common than previously thought for island biota (at least).

The evolutionary consequences of admixture have been demonstrated in a variety of systems, including breeding programs and field studies (e.g. Rieseberg et al., 2003; Grant & Grant, 2008; Nietlisbach et al., 2013), and it is thought to "potentially facilitate evolutionary divergence and speciation by refuelling populations and reorienting their evolution when their environments change" (Grant et al., 2005, pg 63). Its importance in

affecting both the speed and direction of evolution in new environments has been related to the fact that admixture increases levels of genetic variation (Abbott et al., 2003; Garrick et al., 2014), relaxes genetic covariation among traits (Grant & Grant, 1994), and produces novel genotypes (Rieseberg et al., 2003a; Mallet, 2007). Thus there are also conservation implications with a clear potential for genomic admixture to rescue island populations from the deleterious effects of founder events in small populations (since islands are typically colonised by only one or a few individuals). It may also increase long-term viability as it can rapidly enhance genetic and morphological variation (Grant & Grant, 2014) . Detecting genomic admixture may be challenging but it is an important area of research and advanced techniques are proving very useful in this respect. In a recent study, using whole-genome re-sequencing to analyse intra and interspecific genome diversity among all species of the Darwin's finches radiation, Lamichhaney et al. (2015) were able to detect extensive evidence for interspecies gene flow, with species of mixed ancestry originating from genomic admixture. The *L. tessellatus* complex could be further addressed by applying next generation sequencing, such as RAD-seq (restriction-site associated DNA sequencing) which has proved to be an economic and efficient tool for sequencing hundreds of individuals at thousands of loci for nonmodel organisms (e.g. Lexer et al., 2013; Mastretta-Yanes et al., 2014). Applying it to the *L. tessellatus* complex would provide a much larger number of markers, and much greater genome coverage. Such genomic data would allow a more in depth quantification of the extent of admixture among parental different genomes, and how this variation has segregated both spatially and ecologically within founded islands.

The second gap addressed, also in the Canary Islands system, was the lack of information regarding the colonisation history and dynamics of the small arthropod soil dwelling fauna (chapter 3). To address this, I have analysed next generation sequencing data (220 bp of the barcode gene COI obtained from a 454 sequencing platform) from 2500 individuals collected across the island of Tenerife. With these data, I characterised Collembola evolutionary diversity within Tenerife, estimated the genetic relatedness of island and mainland taxa (by comparing them to a database of sequences sampled from outside the archipelago), and inferred the distribution of lineage colonisation times. My analyses recovered broad patterns that indicate the Collembola fauna of Tenerife to be represented by a mosaic of very old lineages and a large number of very recently arrived

lineages, presumably the result of human mediated introductions. This pattern is likely to be quite different from that of large arthropods, although a direct comparison would require a parallel study sampling larger above ground arthropods from the same sampling locations. However, as an indirect example, in the laurel forest of Anaga, northeast of Tenerife, the beetle fauna is found to be essentially native. In a recent study, of the 160 coleopteran species sampled from plots across the laurel forest of Anaga, only 6 species were found to be introduced taxa (Emerson, unpublished). Many of these non-natives were sampled as single individuals, such as a ladybird that probably just by chance blew into the forest, rather than residing within it. In contrast, the Collembola data, which was mostly sampled from forest areas, suggest great penetrance of exotic Collembola into the laurel forest. Regarding diversity in the island, despite the fact that most lineages could not be assigned to taxonomy (because they did not have any close match to sequences on the public database), I reported a remarkable increase in the number of OTUs for the island of Tenerife compared to previous surveys based solely on morphology, with several new taxa being added to the current list. These results confirm the ability of short 454 mini barcodes to rapidly assess the biodiversity of poorly known groups such as soil invertebrates. Results also provide a baseline for future studies investigating specific details of colonisation, diversification and community structure within the Collembola community of the Canary Islands. This is fundamental in a time of rapid global change and pervasive anthropogenic threats to biological communities worldwide.

The main limitation of the work was the short length of DNA sequence reads. Although sequences as long as 500bp were recovered, I was able to use only 220bp, which greatly limited the amount of information that could be used. As sequence lengths get shorter, the ability to detect limited genetic divergence among sequences is reduced, and as such genetic similarity among individuals may be overestimated. Thus, to resolve relationships among sequences of limited divergence, longer sequences are really required. It is clear that this work would greatly benefit from recent technical developments (aimed to obtain longer sequences) that were not available at the time when the data was generated. One example is PCR free techniques such as the parallel de novo mitogenome assembly from a single library of pooled genomic DNA from a bulk sample consisting of many species. This technique has recently been applied to investigate higher-level phylogeny of Curculionoidea beetles (Gillett et al., 2014).

DNA extracts of 173 species belonging to a large number of tribes of Curculionidae were included in a single sequencing run with an Illumina MiSeq. As a result, a total of 92 complete or near complete mitogenomes were obtained without any enrichment of PCR amplification steps, proving it to be an economic, efficient and reliable technique to obtain long sequences reads (sequences in the final aligned data matrix were > 13,000bp long) (Gillett et al., 2014). This is a very promising approach that could be applied to generate mitogenomes for Collembola studied here, as they have a small nuclear genome providing a greater representation of mtDNA sequence data. Complete genomes may even not be needed, for example, by getting sequences of 3000bp, the whole COI gene may be inside this fragment, which would be enough for comparisons against the barcode database. Besides providing long sequences (which enhances taxonomic assignment due to the better matches to the public database), this technique also improves the detection of species within a pool of DNA samples because it removes PCR bias. One complication compared to the work of Gillett et al. (2014) is that intraspecific variation may confound mtDNA genome assembly, but preliminary work attempting to obtain mitogenomes from a bulk sample of many Collembola species suggests this can be overcome (Cicconardi et al., unpublished).

The third and final knowledge gap addressed was the under-explored possibility that the island of Great Britain was not completely defaunated during glaciations and thus subsequently recolonized exclusively from external sources after glaciations (chapters 4 and 5). A more complex pattern involving persistence within small cryptic refugia for small soil dwelling arthropods, was predicted and tested with the Collembola fauna. To assess this possibility, I sequenced the mtDNA COI gene for 722 specimens belonging to the genus *Entomobrya* sampled from across Britain and Wales (chapter 4). I evaluated signatures of persistence through the last Pleistocene glaciation and evidences for species diversity differences between glaciated and unglaciated areas. I found non-random geographic patterns of genetic variation, revealed by the geographically localized range of monophyletic groups, which provided evidence for genetic variation that evolved within Great Britain. Estimated dates for the onset of these *in situ* diversification events revealed lineages to have been present for extended periods of time (22,000 and 45,000 years), thus implying survival through historical climatic and environmental changes. No difference in species diversity between glaciated and unglaciated communities was found, possibly due to the small number of *Entomobrya*

lineages found in Great Britain.

In order to obtain further support for endemism and ancient fauna in the UK, and improve the strength of evidence for species diversity differences between glaciated and unglaciated areas, I then sequenced the mtDNA COI gene for 428 specimens of the genus *Lepidocyrtus,* a more species rich genus found across Great Britain (chapter 5). I identified lower than expected lineage richness for the glaciated community, suggesting it to follow the 'southern richness and northern purity' pattern. I also identified lineages with disjunct distributions and lineages with genetic variation geographically structured within a few sites, which are two basic genetic signatures of long-term persistence. Furthermore, estimated ages for these lineages with endemic genetic variation largely predated the Last Glaciation Maximum (between 100 and 300 kya), indicating long-term persistence for *Lepidocyrtus* in Great Britain. With these results, I was able to corroborate the findings for the genus *Entomobrya*, and demonstrate a more complex evolutionary history for the British soil dwelling fauna, which implicates persistence within small cryptic refugia. Both Collembola genera add to the small list of ancient invertebrate fauna in Great Britain (Nieberding et al., 2005; McInerney et al., 2014), and are congruent with several studies demonstrating soil dwelling fauna to have survived harsh climatic conditions elsewhere (e.g. Stevens et al., 2006; Garrick et al., 2007; Cicconardi et al., 2010; Mortimer et al., 2011). Results also reinforce the alternative hypothesis of northern cryptic glacial refugia for European biodiversity, which has been demonstrated to have several implications for the way we think about Pleistocene phylogeography, and the way we plan conservation priorities (Provan & Bennett, 2008; Stewart et al., 2010). New information on cryptic refugia helps to rethink species dispersal abilities and postglacial migration histories, which are particularly relevant, given the need to predict the effects of the present period of global warming on species distributions (Provan & Bennett, 2008). It also influences our understanding of spatial organization of genetic diversity with consequences for conservation strategies planning, which are aimed to promote long-term sustainability of temperate and boreal ecosystems (Tzedakis et al., 2013). To develop our understanding of the results from chapters 4 and 5 further, sampling needs to be expanded into neighbouring continental areas. This would achieve a more complete understanding of shared and non-shared genetic variation between Great Britain and potential source areas for postglacial recolonisation. A detailed sampling within continental Europe will help to shed light on

our understanding of the sources of genetic variation, and the phylogeographical consequences of Pleistocene glaciations, for the soil dwelling terrestrial invertebrate fauna of the UK. A formal evaluation of the hypothesis that geothermal activity provided glacial refugia for Collembola in Great Britain is also warranted. To achieve this, a sampling strategy based on the geothermal map of Great Britain can be used as a predictive framework to test whether endemic variation is associated with geothermal sites. Finally, next generation sequencing could also be applied to accelerate assessment of Collembola diversity in Great Britain and continental areas. Sanger sequences obtained here are now an important reference database that can be used to identify clusters obtained from bulk sequencing approaches, as detailed above.

To conclude, I have demonstrated the value of molecular tools in addressing gaps in our knowledge about colonisation and diversification of invertebrates, particularly on islands. I have attempted to place my results in the contexts of evolutionary biology and conservation genetics, and demonstrated the importance of studying the consequences of multiple colonisation events in insular systems; the relevance of applying clever ways to study unknown small soil dwelling fauna; and the potential for Collembola to be used as a good model system to investigate genetic signatures of Pleistocene climatic changes. Developments in laboratory and computational methods will continue to extend the scope of questions and depth of answers that can be obtained, therefore enhancing our ability to obtain insights into the evolutionary pattern and process that promote species diversity and community structure.

References

Abbott R.J., Hegarty M.J., Hiscock S.J., & Brennan A.C. (2010) Homoploid hybrid speciation in action. Taxon, **59**, 1375–1386.

Abbott R.J., James J.K., Milne R.I., & Gillies A.C.M. (2003) Plant Introductions, Hybridization and Gene Flow. *Philosophical Transactions: Biological Sciences*, **358**, 1123–1132.

Barton N.H. (2013) Does hybridization influence speciation? *Journal of Evolutionary Biology*, **26**, 267–269.

Cicconardi F., Nardi F., Emerson B.C., Frati F., & Fanciulli P.P. (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Molecular Ecology*, **19**, 386–400.

Emerson B.C. & Faria C.M.A. (2014) Fission and fusion in island taxa – serendipity, or something to be expected? *Molecular Ecology*, **23**, 5132–5134.

Garrick R.C., Benavides E., Russello M.A., Hyseni C., Edwards D.L., Gibbs J.P., Tapia W., Ciofi C., & Caccone A. (2014) Lineage fusion in Galápagos giant tortoises. *Molecular Ecology*, **23**, 5276–5290.

Garrick R.C., Sands C.J., Rowell D.M., Hillis D.M., & Sunnucks P. (2007) Catchments catch all: long-term population history of a giant springtail from the southeast Australian highlands - a multigene approach. *Molecular Ecology*, **16**, 1865–1882.

Gillett C.P.D.T., Crampton-Platt A., Timmermans M.J.T.N., Jordal B.H., Emerson B.C., & Vogler A.P. (2014) Bulk De Novo Mitogenome Assembly from Pooled Total DNA Elucidates the Phylogeny of Weevils (Coleoptera: Curculionoidea). *Molecular Biology and Evolution*, **31**, 2223–2237.

Grant B.R. & Grant P.R. (2008) Fission and fusion of Darwin's finches populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 2821–2829.

Grant P.R. & Grant B.R. (1994) Phenotypic and Genetic Effects of Hybridization in Darwin's Finches. *Evolution*, **48**, 297–316.

Grant P.R. & Grant B.R. (2014) *40 years of evolution. Darwin's finches on Daphne Major Island*. Princeton University Press, New Jersey.

Grant P.R., Grant B.R., & Petren K. (2005) Hybridization in the Recent Past. *The American Naturalist*, **166**, 56–67.

Harrison R.G. & Larson E.L. (2014) Hybridization, Introgression, and the Nature of Species Boundaries. *Journal of Heredity*, **105**, 795–809.

Hochkirch A. (2013) Hybridization and the origin of species. *Journal of Evolutionary Biology*, **26**, 247–251.

Jordal B.H., Emerson B.C., & Hewitt G.M. (2006) Apparent "sympatric" speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology*, **15**, 2935–2947.

Lamichhaney S., Berglund J., Almén M.S., Maqbool K., Grabherr M., Martinez-Barrio A., Promerová M., Rubin C.-J., Wang C., Zamani N., Grant B.R., Grant P.R., Webster M.T., & Andersson L. (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **advance online publication, doi: 10.1038/nature14181.**

Lexer C., Mangili S., Bossolini E., Forest F., Stölting K.N., Pearman P.B., Zimmermann N.E., & Salamin N. (2013) "Next generation" biogeography: towards understanding the drivers of species diversification and persistence. *Journal of Biogeography*, **40**, 1013–1022.

Mallet J. (2007) Hybrid speciation. *Nature*, **446**, 279–283.

Mastretta-Yanes A., Arrigo N., Alvarez N., Jorgensen T.H., Piñero D., & Emerson B.C. (2014) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, **doi: 10.1111/1755-0998.1229.**

McInerney C.E., Maurice L., Robertson A.L., Knight L.R.F.D., Arnscheidt J., Venditti C., Dooley J.S.G., Mathers T., Matthijs S., Eriksson K., Proudlove G.S., & Hänfling B. (2014) The ancient Britons: groundwater fauna survived extreme climate change over tens of millions of years across NW Europe. *Molecular Ecology*, **23**, 1153–1166.

Mortimer E., Vuuren B.J. van, Lee J.E., Marshall D.J., Convey P., & Chown S.L. (2011) Mite dispersal among the Southern Ocean Islands and Antarctica before the last glacial maximum. *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 1247–1255.

Nieberding C., Libois R., Douady C.J., Morand S., & Michaux J.R. (2005) Phylogeography of a nematode (*Heligmosomoides polygyrus*) in the western Palearctic region: persistence of northern cryptic populations during ice ages? *Molecular Ecology*, **14**, 765–779.

Nietlisbach P., Wandeler P., Parker P.G., Grant P.R., Grant B.R., Keller L.F., & Hoeck P.E.A. (2013) Hybrid ancestry of an island subspecies of Galápagos mockingbird explains discordant gene trees. *Molecular Phylogenetics and Evolution*, **69**, 581–592.

Provan J. & Bennett K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, **23**, 564–571.

Rieseberg L.H., Raymond O., Rosenthal D.M., Lai Z., Livingstone K., Nakazato T., Durphy J.L., Schwarzbach A.E., Donovan L.A., & Lexer C. (2003a) Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. *Science*, **301**, 1211–1216.

Rieseberg L.H., Widmer A., Arntz A.M., & Burke J.M. (2003b) The Genetic Architecture Necessary for Transgressive Segregation Is Common in Both Natural and Domesticated Populations. *Philosophical Transactions: Biological Sciences*, **358**, 1141–1147.

Schumer M., Rosenthal G.G., & Andolfatto P. (2014) How Common Is Homoploid Hybrid Speciation? Evolution, **68**, 1553–1560.

Servedio M.R., Hermisson J., & van Doorn G.S. (2013) Hybridization may rarely promote speciation. Journal of Evolutionary Biology, **26**, 282–285.

Shaw K.L. (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences*, **99**, 16122–16127.

Soltis P.S. & Soltis D.E. (2009) The Role of Hybridization in Plant Speciation. *Annual Review of Plant Biology*, **60**, 561–588.

Stevens M.I., Greenslade P., Hogg I.D., & Sunnucks P. (2006) Southern Hemisphere Springtails: Could Any Have Survived Glaciation of Antarctica? *Molecular Biology and Evolution*, **23**, 874–882.

Stewart J.R., Lister A.M., Barnes I., & Dalen L. (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 661–671.

Tzedakis P.C., Emerson B.C., & Hewitt G.M. (2013) Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, **28**, 696–704.