Phylogeographic insights into cryptic glacial refugia

Jim Provan¹ and K.D. Bennett²

¹School of Biological Sciences, Queen’s University Belfast, 97 Lisburn Road, Belfast BT9 7BL, UK
²School of Geography, Archaeology and Palaeoecology, Queen’s University Belfast, 42 Fitzwilliam Street, Belfast BT9 6AX, UK

The glacial episodes of the Quaternary (2.6 million years ago–present) were a major factor in shaping the present-day distributions of extant flora and fauna, with expansions and contractions of the ice sheets rendering large areas uninhabitable for most species. Fossil records suggest that many species survived glacial maxima by retreating to refugia, usually at lower latitudes. Recently, phylogeographic studies have given support to the existence of previously unknown, or cryptic, refugia. Here we summarise many of these insights into the glacial histories of species in cryptic refugia gained through phylogeographic approaches. Understanding such refugia might be as important as the Earth heads into another period of climate change, in terms of predicting the effects on species distribution and survival.

Introduction

Over the past ca. two million years (MY) the climate on Earth has undergone periodic fluctuations, resulting most notably in the Ice Ages [1]. The advance and retreat of the ice sheets through multiple glacial cycles, most recently the last glacial maximum (LGM) around 25 000–18 000 years before present, had a major impact on the present-day distribution of species [2,3]. Palaeontological and palynological records, as well as a large body of biogeographical data, suggest that most of the biota, particularly temperate species, persisted through glacial maxima in lower-latitude refugia, where climatic conditions were less extreme [4,5]. More recently, the application of phylogeography has provided new insights into the history of many species throughout glacial cycles. The earliest and most extensive application of phylogeographic methods to reconstruct the periglacial and postglacial history of species has centred around the terrestrial flora and fauna of Europe [6]. Comparative analysis of the levels and patterns of genetic variation across a range of temperate plant and animal taxa has revealed the existence of three major refugia in the Mediterranean peninsulas of Iberia, Italy and the Balkans (Figure 1). It was believed that after the end of the LGM, the rest of Europe was recolonised from these refugia by species moving northward as the climate warmed. This basic expansion–contraction (EC) model provides a simple paradigm for the demography of species through glacial cycles, but is rendered more complex by factors such as different species having different environmental niches and dispersal rates, the scope for altitudinal migration and interactions between communities of species [7,8].

The rates and distances of spread of organisms following the LGM have interested ecologists for more than 100 years. At the end of the 19th century, the botanist Clement Reid recorded his surprise that there were oak trees in northern Britain, despite the fact that they would have had to have travelled ca. 950 km from southern refugia in the time period since the end of the LGM [9]. He believed that this rate exceeded the dispersal capabilities of oaks, and the observation, known as Reid’s Paradox, was initially explained by rare, long-distance dispersal events. Subsequent palaeontological and biogeographical studies have suggested that the inconsistencies could be explained if the species had persisted throughout the LGM in unknown, or

Glossary

- **Allelic richness**: A measure of genetic diversity that allows for comparison of numbers of alleles in samples of unequal size by standardising values to represent a single sample size, usually the smallest used.
- **Ancient DNA (aDNA)**: DNA recovered from fossilised material such as bones and wood. Problems with ancient DNA technology include low yields, degradation of genetic material and contamination by extant DNA.
- **Bottleneck (genetic bottleneck)**: Loss of genetic variation due to a reduction in population size.
- **Coalescent theory**: A model of population genetics that traces the lineages of all alleles in a population back to their most recent common ancestor and allows inferences of population dynamics beyond those available from “classical” population genetics theory (e.g., based on allele frequencies).
- **Contact zone**: An area where separately recolonising lineages meet. If these lineages are genetically differentiated owing to isolation in distinct glacial refugia, the contact zone will be an area of high genetic diversity.
- **Drift (genetic drift)**: Stochastic changes in a gene pool from generation to generation as a result of random sampling of gametes.
- **Expansion–contraction (EC) model**: The demographic scenario in temperate species whereby glacial cycles result in the contraction in size and shift toward lower latitudes during periods of cooling, followed by population growth and recolonisation during postglacial warming.
- **Founder effect**: Establishment of a new population comprising a subset of an original source population, with a concomitant decrease in genetic diversity. Due to random sampling from the source population, the gene pool of the founded population might not be representative of that of the source population.
- **Habitat tracking**: Shifts in the distributional range of species correlated with the availability of suitable habitats as a result of climate change.
- **Haplotype**: A haploid genotype composed of linked alleles on a single chromosome, such as from the chloroplast or mitochondrial genome.
- **Nested clade phylogeographic analysis (NCPA)**: An empirical approach that uses an inference key to identify demographic factors (such as population expansion or restricted gene flow) that best explain the observed relationships between genetic and geographical distance.
- **Phylogeography**: The study of the geographical distribution of genetic lineages.
- **Statistical phylogeography**: An approach to phylogeography that uses models based on coalescent theory to statistically test a set of alternative hypotheses that could explain the current distribution of a species.
cryptic, refugia at higher latitudes [10]. This new information is particularly relevant, given the need to predict the effects of the present period of global warming on species distributions [11,12]. For species whose rates of dispersal are insufficient to accommodate changes in climate, their ultimate fate might be extinction through loss of suitable habitat [13]. If cryptic refugia in northerly latitudes were present during the LGM, then the dispersal capabilities of many organisms might thus far have been overestimated. Here we highlight some of the insights into the existence of cryptic glacial refugia afforded by recent phylogeographic studies. These allow us to refine and extend the earlier models and scenarios by incorporating data from cold-adapted species and from areas beyond terrestrial Europe such as North America and the marine realm. We outline the approaches used to identify refugia, some in surprising locations, and the developments in methodological and analytical approaches that confirm that a simple EC model of species to and from southern glacial refugia might not fully explain the response of species to cycles of climate change.

**Traditional approaches to identifying refugia**

Palaeoecological evidence for refugia has typically taken the form of determining regions from which modern taxa appear to have originated. The most developed form of this approach is the atlas of Huntley and Birks [14], who mapped the distributions of pollen over the whole of Europe for the period <13 000 years ago. Most forest trees typically follow a pattern of spread from areas around the southern (but also western and eastern) fringes of Europe. This led the authors to argue that these regions were glacial refugia for forest trees [14]. A range of additional criteria led to the suggestion that mountainous areas of southern Europe were also critical for the survival of trees in Europe through glacial periods because their high relief offered scope for altitudinal range shifts in response to major climatic changes [4]. Several studies [10,15] based on dated fossils now indicate that a range of mammal and tree taxa might well have survived glacial periods in central Europe, possibly scattered or at low densities, but nevertheless present. The palaeoecological record has thus generated two broad, overlapping, hypotheses of how taxa survive glacial periods: small and localised refugia of southern European mountain areas, and broader, less well-defined, cryptic refugia of scattered populations. However, these hypotheses should probably be seen as end points of a continuum: reality for many taxa might lie in between. The small, localised refugia model might be most appropriate for taxa of what became the Holocene temperate forests of central and northern Europe, whereas cryptic refugia might be more relevant for taxa of other biomes.

The identification of marine refugia using fossil data has been particularly problematic for several reasons.

---

**Figure 1.** Map showing the locations of the three main peninsular refugia during the last glacial maximum (turquoise) and two of the cryptic refugia discussed in the text (yellow): Ib = Iberia; It = Italy; Ba = Balkans; Ca = Carpathians; HD = Hurd Deep. The approximate limits of the ice sheets are shaded white and the black dotted line indicates the southern limit of the permafrost. Redrawn from Ref. [7].
Foremost among these is the fact that many groups of marine organisms, particularly algae and invertebrates, leave little or no fossil record and, where they do, it is often difficult to identify taxa accurately. For intertidal species, a further complicating factor is that the changes in climate were associated with fluctuations in sea levels (down to 120 m below present at LGM), giving rise to different coastlines from those observed today [16]. The subsequent rise in sea levels also obscured large sections of the fossil record. Consequently, phylogeographic studies have been particularly informative in reconstructing the periglacial and postglacial history of marine organisms [17].

Using phylogeography to determine refugia
The basic EC model of population demography during glacial maxima [6] gives rise to two key predictions concerning the levels and patterns of genetic diversity that will be observed in potential refugial areas and in areas that have subsequently been recolonised from these refugia. First, because populations of organisms that have persisted throughout glacial maxima in refugia will have

Box 1. An example of using phylogeography to identify cryptic refugia

The red seaweed *Palmaria palmata* (dulse) is found on both sides of the North Atlantic Ocean, its current-day distribution ranging from Portugal to Spitzbergen, northern Norway and southern Iceland in Europe, and from Long Island, New York to Arctic Canada on the east coast of North America. As with many algal species, it has no fossil record, so a phylogeographic approach was used to elucidate the postglacial history of the species [42]. Polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP) analysis of two regions of the chloroplast genome revealed the highest levels of haplotype diversity in the English Channel (Figure I), with six haplotypes detected. Sequencing of a further region of the chloroplast genome also found the highest levels of nucleotide diversity in the same area. The rest of the species range (with the exception of Galway Bay, Ireland) was composed of only two of the haplotypes, with no intrapopulation genetic diversity detected. Such patterns are often indicative of glacial refugia and subsequent recolonisation. Admixture within the channel, for example due to shipping-mediated dispersal, could be ruled out, as four of the six PCR-RFLP haplotypes were not found elsewhere in the entire range of the study. Furthermore, previous studies on the seaweeds *Ascophyllum nodosum* [40] and *Fucus serratus* [41] had also found similarly high levels of genetic diversity in populations from the English Channel. The observed distribution of genetic diversity throughout the range of the species, coupled with the concordance of similar patterns across several species, led the authors to conclude that *P. palmata*, and possibly other marine species, had persisted through the LGM in a series of flooded depressions on the floor of the English Channel [74] and that postglacial recolonisation of the eastern North Atlantic has taken place from this area as temperatures and sea levels rose. A possible second refugium was proposed in the Galway Bay area to the southwest of Ireland, but populations there did not appear to have participated in the recolonisation process. Subsequent studies in marine species from the region have also found similar patterns of genetic diversity [43,44].

Figure I. Distribution of *Palmaria palmata* PCR-RFLP haplotypes (colour-coded) found by Ref. [42]. The dotted oval highlights haplotypes found in the English Channel. Reproduced with permission.
a longer demographic history than populations that have evolved during the postglacial, one of the key signatures of glacial refugia should be that they tend to harbour high levels of genetic diversity [7,18]. This gives rise to a scenario of 'southern richness and northern purity' [8], because recolonising populations are usually composed of subsets of the genetic diversity present in the source refugial population and the founding process can result in a series of sequential founder effects and bottlenecks. Second, the long-term isolation of populations within geographically separate refugia will lead to genetic differentiation due to drift. The identification of spatial structuring of distinct genetic lineages, both between putative refugial areas and along possible recolonisation routes, is thus further evidence for the existence of glacial refugia [6] and can be corroborated by phylogeographic concordance across species in comparative studies.

Although the genetic signatures of glacial refugia and recolonisation are often obvious from many observed phylogeographic scenarios (Box 1), other factors can act to obscure or confound these. The admixture of genetic lineages due to recolonisation from separate refugia can give rise to regions (i.e. contact zones) of high genetic diversity that might be confused with refugial areas themselves [19]. One key feature that differentiates between these two scenarios is that contact zones will be composed entirely of haplotypes from the source refugia (or their descendants), whereas the refugia themselves tend to harbour high numbers of private haplotypes not found elsewhere and which presumably have not participated in the recolonisation process. This highlights the importance of adequate sampling, because limited sampling within refugial areas might lead to the apparent absence of haplotypes that are found elsewhere. Furthermore, because haplotypes within a refugium will be more closely related genealogically to haplotypes from the same refugium than to those from other refugia, a comparison that incorporates the level of genetic divergence between haplotypes can also differentiate between refugia and admixture zones. One such approach [20] is based on measures of allelic richness and of diversity based only on haplotype frequency ($h_T$) and on both haplotype frequency and the genetic distance between haplotypes ($v_T$). Areas of high allelic richness with $h_T > v_T$ would be indicative of refugia, because the haplotypes found there belong to the same genetic lineage whereas areas of high allelic richness with $h_T \leq v_T$ would indicate admixture of genetically divergent lineages. Under the same model, areas of low to intermediate allelic richness correspond to recolonised areas. This approach, however, might be obscured by the existence of regions of 'refugium-within-refugium,' which are characterised by high levels of between-population differentiation within refugial areas [21]. Such cases further highlight the need for adequate sampling within putative refugial regions as well as recolonised regions. Recolonisation pathways could be less obvious in species with high dispersal capabilities [22] and might be completely confounded by high levels of anthropogenic translocation, for example in commercially important species or those that have spread with humans and/or agriculture over the millennia. Finally, changing regimes of selection over time might give rise to one or more genotypes or haplotypes in a region that are not representative of those found there during the LGM.

### Insights into cryptic refugia from phylogeography

In this section, we highlight examples of where phylogeographic studies have provided new evidence on the existence of hitherto unknown or unconfirmed glacial refugia. We also summarise some of the key advances in methodological approaches that have facilitated these insights.

#### The location of cryptic refugia

One feature that became apparent from the early phylogeographic studies that elucidated the main European refugial areas was the presence of genetic lineages, some widespread across northern Europe, that did not correspond to the lineages present in any of the three peninsular refugia [23]. This suggested that the sources of such lineages lie elsewhere, possibly in western Asia or in more northern cryptic refugia [10]. Subsequent phylogeographic studies have further highlighted the existence of cryptic refugia, often supporting or clarifying earlier theories on the possible existence of these refugia based on biogeography or palaeontology, but sometimes providing de novo evidence for their existence (Table 1).

A large body of phylogeographic evidence now indicates the existence of a major northern refugium for a variety of taxa in the area around the Carpathians. Studies of mammals [24–26], reptiles [27] and amphibians [28,29] have identified divergent lineages in the region, many of which predate the LGM. The existence of a Carpathian refugium characterised by mixed deciduous and coniferous wood-

### Table 1. Examples of cryptic refugia areas revealed by concordance across multiple phylogeographic studies

<table>
<thead>
<tr>
<th>Refugium</th>
<th>Nature of refugial regions</th>
<th>Species</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carpathians</td>
<td>Regions of mixed coniferous and deciduous woodland, often on south-facing mountain slopes</td>
<td>Micrus agrestis (field vole)</td>
<td>[24]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clethrionyns glareolus (bank vole)</td>
<td>[25,26]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vipera berus (adder)</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Triturus spp. (newts)</td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rana arvalis (moor frog)</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dryas integrifolia (plant)</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ovis spp. (mountain sheep)</td>
<td>[36]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lagopus mutus (rock ptarmigan)</td>
<td>[37]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dicrostonyx groenlandicus (collared lemming)</td>
<td>[38]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Packera spp. (plants)</td>
<td>[39]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ascothyllum nodosum (seaweed)</td>
<td>[40]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fucus serratus (seaweed)</td>
<td>[41,43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Palmiria palmata (seaweed)</td>
<td>[42]</td>
</tr>
<tr>
<td>Northern North America</td>
<td>Ice-free regions in the Canadian high Arctic and between the</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Laurentide and Cordillerica ice sheets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Channel</td>
<td>The Hurd Deep, or surrounding trenches, which might have</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>persisted as ‘marine lakes’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
land, often on south-facing slopes, is supported by the fossil record and palynological studies [30,31]. Its role in the postglacial recolonisation of northern and western Europe, however, has only become apparent using comparative phylogeographic approaches.

Phylogeographic studies have also revealed refugia within the supposed limits of the ice sheets during glacial maxima. At the height of the most recent (Wisconsin) glaciation in North America, the Laurentide and Cordilleran ice sheets covered Canada and the northern USA as far south as around 45°N, and it was believed that regional flora and fauna retreated southward into refugia where they persisted throughout the LGM. In 1937, however, Swedish botanist Eric Hultén published his book Outline of the History of Arctic and Boreal Biota during the Quaternary Period, a survey of comparative plant biogeography (see Ref. [32]). Based on the distribution of ca. 2000 plant species, he proposed that the area between north and west Alaska and Siberia, which were connected at the time by a land bridge, had remained ice free and provided a northern refugium that he called ‘Beringia.’ The first phylogeographic evidence for a Beringian refugium was published in 1999 [33], and a series of studies have since confirmed that a range of plant and animal species persisted throughout the LGM in this area [34–36]. Although the role of Beringia as a refugium for North American species is now widely accepted, several studies have shown that many species also occupied other northern refugia. For example, a cryptic refugium in the Canadian high Arctic has been suggested for the rock ptarmigan Lagopus mutus [37], the plant Dryas integrifolia [33] and the collared lemming Dicrostonyx groenlandicus [38]. The existence of an ice-free corridor between the Laurentide and Cordilleran ice sheets in the area of the Mackenzie Mountains in Alberta [36], another much-debated subject, is also supported by phylogeographic studies [36,39].

Recently, surprising evidence has emerged for a marine glacial refugium in the English Channel. A succession of studies on seaweeds consistently highlighted the region as an area of particularly high genetic diversity, despite the fact that the region between England and France was dry land during the LGM due to the lowering of sea levels [40–43]. It has been postulated that during the LGM, a trench in the English Channel known as the Hurd Deep might have persisted as a marine lake and acted as a refugium [42]. A study on thornback rays (Raja clavata) [44] also found high levels of genetic diversity in this area but could not conclude for certain that these were due to the existence of a refugium. Nevertheless, these findings have highlighted the potential importance of the region for future phylogeographic studies.

**Insights from ancient DNA**
The essence of the phylogeographic approach involves utilising DNA sequence information from extant populations to reconstruct the genealogical history of these populations in a spatial context. The recent development of ancient DNA (aDNA) technology, however, has allowed the direct analysis of sequence variation from specific points in history [45]. aDNA studies in a range of organisms are now providing often surprising insights into the demography of populations during glacial and interglacial periods and have allowed a more detailed analysis of how organisms have responded to changes in climate at the population level.

Although phylogeographic approaches have successfully identified patterns of recolonisation during expansion phases following glacial maxima, they cannot provide information on population dynamics during contraction phases. It has long been unclear whether organisms migrated to refugia in response to the changes in climate, a phenomenon called habitat tracking, or whether populations outside refugial areas simply became extinct. A study on the Arctic fox Alopex lagopus (Box 2, Figure I) using aDNA [46] suggests that populations did not shift their distributions during their postglacial contraction phase to follow available habitat, but rather became extinct (Arctic species largely respond to climate change in a fashion converse to that of temperate species, i.e. their ranges contract rather than expand during interglacial warming periods; see Box 2). This suggests that population extinction, rather than migration, might be the primary driver of range shifts, and this has important implications in current climate change scenarios, particularly for cold-adapted species.

Studies on aDNA in temperate species have also cast doubt on the idea that populations were isolated from each other in refugia during glacial maxima. The spatial distribution of distinct genetic lineages in many extant taxa suggest the action of genetic drift in isolated gene pools within refugia, but aDNA studies in brown bears suggest more complex population dynamics [47,48]. Brown bears are now extinct in central Europe, but analysis of aDNA from fossils from the region indicate high genetic diversity and gene flow in areas north of the traditional peninsular refugia. This is particularly ironic, as the brown bear has long been held as one of the ‘model’ species in reconstructing the recolonisation of Europe from these refugia. As well as providing insights into population persistence and dynamics beyond the southern refugia, aDNA analysis has also cast some doubt on the assumption that extant gene pools are representative of those found in refugia during glacial episodes [49]. The use of aDNA technology to gain insights into the genetic composition of refugial (and non-refugial) populations of temperate species might be limited by the fact that climatic conditions at lower latitudes are far from ideal for the preservation of intact DNA amenable to analysis, but they still represent a potentially important tool in understanding the complexities and subtleties of periglacial and postglacial population dynamics.

**Analytical approaches**
Recent developments in the analysis of phylogeographic data have provided new insights into the demographic histories of refugial species. The application of coalescent theory can better accommodate the stochastic demographic processes associated with postglacial recolonisation, such as population expansion and migration, than ‘classical’ population genetics approaches [50,51]. Furthermore, the newly emerging field of statistical phylogeography enables coalescent-based models to be applied to test...
explicit hypotheses that underlie observed phylogeographic patterns, and holds great promise as an analytical tool [52,53]. Nested clade phylogeographic analysis (NCPA) has also been extensively used to provide insights into the historical processes associated with glacial cycles such as population fragmentation and range expansion, but has been criticised for being post hoc and lacking in statistical rigour [54]. In addition, a recent study that applied NCPA to simulated data sets [55] found that the majority (76%) of the inferences arrived at were false-positives, and it has even been suggested that the use of NCPA be discontinued, at least until the procedure has been more thoroughly evaluated [56].

As with many phylogeographic approaches to reconstructing the past history of organisms, the synergistic information yielded by comparative studies can provide real insights into the larger-scale forces shaping the current-day distributions of species. One particularly interesting approach could involve combining modelling studies with genetic approaches [57,58]. The major advantage of such integrative approaches is that ecological and/or palaeontological models can be used to derive hypotheses concerning species distribution which can subsequently be tested using statistical phylogeography [59–61]. A recent example of this was a study on the hairy land snail (*Trochulus villosus*) [62], which first used ecological niche modelling to identify putative refugial and recolonisation scenarios, and then used genetic data to confirm which of their hypotheses was most likely. A similar approach was used in cod (*Gadus morhua*) [63] to confirm the existence of a marine glacial refugium in the northwest Atlantic [64]. Likewise, the combined analysis of palaeontological and genetic data [65], rather than carrying out genetic analysis and then referring to fossil records for a posteriori explanations, can provide novel information on refugial survival and postglacial recolonisation, such as further evidence for a cryptic refugium in the Carpathians [66].

**Conclusions and prospects**

Phylogeographic studies of both extant and extinct populations have provided new insights into the nature and locations of refugial areas during glacial periods. Most importantly, they have increasingly cast doubt on the idea that temperate areas were recolonised solely from southern refugia after the LGM. The identification of cryptic refugia has important implications in current and future periods of global climate change. Knowledge of changes in the distribution of genetic diversity over time will allow the identification of conservation units [67,68]. In addition, the existence of higher-latitude refugia will require a reevaluation of the migration rates of species based on the apparent recolonisation from low-latitude refugia, thus influencing predictions on the rates [69,70], and even the role [48], of habitat tracking and the response of species to climate change. Many of the recent insights into cryptic glacial refugia have resulted from advances in both methodology (particularly aDNA technology) and analytical approaches to phylogeography. In plant species, the synthesis of palynological data and the ability to amplify DNA from fossil pollen [71,72] (even single pollen grains [73]) will allow novel insights into both the chronological and geographical distribution of genetic lineages in the postglacial. Furthermore, the use of palaeontological and ecological modelling to formulate hypotheses that can be subsequently tested using phylogeographic methods is an attractive prospect. An increasingly sophisticated understanding of the responses of species to climate change, both future and past, is devel-
oping at a time when predicting such responses is becoming increasingly important.

Acknowledgements
We are extremely grateful to Christine Maggs for helpful and interesting discussions on many aspects of glacial refugia. We are also grateful to John Charles, Godfrey Hewitt and four anonymous referees for their comments on earlier versions of the manuscript.

References
5 Williams, D. et al. (1998) *Quaternary Environments*, Arnold
15 Reid’s paradox of rapid plant migration. *Bioscience* 48, 13–24