Unique genetic variation at a species’s rear edge is under threat from global climate change

Jim Provan* and Christine A. Maggs

School of Biological Sciences, Queen’s University Belfast, 97 Lisburn Road, Belfast BT9 7BL, UK

Global climate change is having a significant effect on the distributions of a wide variety of species, causing both range shifts and population extinctions. To date, however, no consensus has emerged on how these processes will affect the range-wide genetic diversity of impacted species. It has been suggested that species that recolonized from low-latitude refugia might harbour high levels of genetic variation in rear-edge populations, and that loss of these populations could cause a disproportionately large reduction in overall genetic diversity in such taxa. In the present study, we have examined the distribution of genetic diversity across the range of the seaweed Chondrus crispus, a species that has exhibited a northward shift in its southern limit in Europe over the last 40 years. Analysis of 19 populations from both sides of the North Atlantic using mitochondrial single nucleotide polymorphisms (SNPs), sequence data from two single-copy nuclear regions and allelic variation at eight microsatellite loci revealed unique genetic variation for all marker classes in the rear-edge populations in Iberia, but not in the rear-edge populations in North America. Palaeodistribution modelling and statistical testing of alternative phylogeographic scenarios indicate that the unique genetic diversity in Iberian populations is a result not only of persistence in the region during the last glacial maximum, but also because this refugium did not contribute substantially to the recolonization of Europe after the retreat of the ice. Consequently, loss of these rear-edge populations as a result of ongoing climate change will have a major effect on the overall genetic diversity of the species, particularly in Europe, and this could compromise the adaptive potential of the species as a whole in the face of future global warming.

Keywords: Chondrus crispus; distribution range; glacial refugia; phylogeography; population genetics

1. INTRODUCTION

Global climate change is having a significant effect on the distributions of a wide variety of taxa, with well-documented cases of range shifts [1,2] and even population extinctions [3,4]. Although it is now accepted that knowledge of the levels and patterns of genetic diversity within and between populations is crucial for formulating conservation strategies for potentially threatened species, no clear consensus has emerged on the relative importance of historical and contemporary factors in determining the distribution of genetic variation across species’ ranges. It is generally assumed that population genetic parameters such as effective population size (Ne), genetic diversity and levels of gene flow would reflect the ‘abundant-centre’ or ‘central-marginal’ model of species’ distributions [5,6]. In this model, optimal conditions at the centre of a species’ range lead to greater abundance in central populations than in peripheral populations, which are likely to exist in sub-optimal conditions and thus be smaller and more isolated [7,8] (but see [9]). This should result in high levels of genetic diversity at the centre of the range, with a decrease in diversity coupled with an increase in population differentiation towards the range edge owing to elevated drift and/or selection. It has also been argued, however, that contemporary processes have less effect on the distribution of genetic diversity across species’ ranges than longer-term processes, such as postglacial recolonization events associated with previous fluctuations in Earth’s climate [10–13]. These have generally resulted in temperate terrestrial species showing greatest genetic diversity at lower latitudes owing to the ‘classic’ paradigm of recolonization from low-latitude refugia [11] (but see [13]). Thus, the conservation value of peripheral populations, particularly those at the rear edge, is of ongoing interest and debate since they could contain unique genetic variation. This would be especially important where one or more refugial populations have not contributed significantly to the recolonization of previously glaciated areas [12,14].

The impact of climate change on the distributions of intertidal species is of particular interest for several reasons. As well as increases in air and sea temperatures, coastal ecosystems are subject to additional effects of climate change, such as rises in sea level and a decrease in sea water pH associated with increased atmospheric carbon dioxide levels [15,16]. Many marine species are believed to exist at, or close to, their thermal tolerance limits [17], and thus are considered good early indicators of the overall effects of climate change on species’ distributional ranges [18]. Unlike many terrestrial species, intertidal taxa generally lack the scope for altitudinal migration to counter the effects of warming climate and, on the whole, their ranges appear to be following the same patterns of poleward expansions in distribution as observed in terrestrial species [15,19]. A recent study...
on intertidal algae, however, highlighted the difficulties in making generalizations concerning range shifts, particularly when comparing warm- and cold-adapted species, since warm-water species that exhibited a range shift all expanded northwards, while cold-tolerant species displayed no consistent pattern [20].

*Chondrus crispus* Stackhouse is one species of seaweed that has exhibited a significant range shift in the last few decades, and is thus a potential indicator of the effects of climate change. Lima *et al.* [20] reported a northward shift in the southern limit of Portuguese populations of 180 km since 1971: these populations are morphologically distinct from all other populations of the species (figure S1 in the electronic supplementary material). Portugal is the southern limit of the species’s distribution range in the eastern North Atlantic, where it occurs as far north as approximately 69°N in Norway and on the southern coast of Iceland. In the western Atlantic, it is found between New Jersey and southeastern Labrador (figure 1). Its congeners are Pacific species and, following a trans-Arctic migration, *C. crispus* is now endemic to the northern North Atlantic [21]. The species’s present-day distribution is primarily determined by high temperatures, with spore mortality recorded at temperatures as low as 21°C [22]. Consequently, Iberian populations are probably on the cusp of their upper thermal tolerance limit.

Phylogeographic studies can provide valuable insights into how species have previously responded to episodes of climatic change and the way in which these have shaped the distribution of genetic diversity across entire ranges. North Atlantic marine organisms exhibit complex patterns of species survival in various refugial areas during the last glacial maximum (LGM; ca 18–21 kyr ago), with different refugia contributing to postglacial recolonization to varying degrees [23]. Such complex scenarios have important implications for how species respond to climate change in two ways. Firstly, if species have also persisted in cryptic northern refugia as opposed to surviving solely in low-latitude regions, their dispersal capabilities in the face of previous phases of global warming may be overestimated [13,24,25]. The existence of these northerly refugia, if not recognized as such, would also confound to some degree the expectation of reduced genetic variation at the leading edge [12,13]. Secondly, where southern refugia have not contributed to postglacial recolonization, they may represent reservoirs of unique genetic variation that are under immediate threat from climate change [12,14]. Previous phylogeographic studies on North Atlantic seaweeds with similar distributions to *C. crispus* [26–28] have suggested that both scenarios are likely.

In the present study, we used a phylogeographic approach to test whether populations from the southern limits of the distribution range of *C. crispus* exhibit unique genetic variation that may be lost as a result of climate change. We have sampled across the entire range of the species on both sides of the Atlantic, since the relative importance of leading-edge versus rear-edge peripheral populations is still not fully understood [12,29,30]. Given that extinction of rear-edge populations is more likely than migration [1,29,31,32], they may merit
special conservation effort in the face of global warming, particularly if they are genetically distinct.

2. MATERIAL AND METHODS

(a) Sample collection and DNA extraction

A total of 354 samples were collected from 19 populations (14–23 individuals per population) across the range of C. crispus, including the southern edge of both western and eastern coasts of the North Atlantic (table S1 in the electronic supplementary material; figure 1). Haploid gametophytes were identified using the resorcinol staining test for k-carrageenan [33] and total genomic DNA extracted using the Qiagen Plant DNeasy kit.

(b) Single nucleotide polymorphism analysis of mitochondrial DNA

The trnL intron of the C. crispus mitochondrial genome (EMBL accession no. Z47547) was analysed for SNP polymorphisms. A 498 bp fragment was amplified in an ascertainet set of six individuals from each population using the following primers: trnL5—TGAGTCTAGATTGA TTCCAACCATC and trnL3—GAGTCTATAGCTTAAG GTTAAGAG. PCR was carried out on an MWG Primus thermal cycler using the following parameters: initial denaturation at 94°C for 3 min followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, extension at 72°C for 2 min and a final extension at 72°C for 5 min. Polymerase chain reaction (PCR) was carried out in a total volume of 20 μl containing 200 ng genomic DNA, 20 pmol of each primer, 1× PCR reaction buffer, 200 μM each dNTP, 2.5 mM MgCl2 and 0.5 U GoTaq Flexi DNA polymerase (Promega). 5 μl PCR product were resolved on 1.5 per cent agarose gels and visualized by ethidium bromide staining, and the remaining 15 μl sequenced commercially (Macrogen, Korea).

Mitochondrial SNPs identified in more than a single individual in the ascertainet set were screened in all 354 samples using allele-specific PCR (AS-PCR), which uses a competitive PCR approach to amplify either of two alleles at a specific position. AS-PCR primers were designed as described by Provan et al. [34] and are given in table S2 in the electronic supplementary material. The AS-PCR protocol was as follows: initial denaturation at 94°C for 3 min followed by 11 touchdown cycles of denaturation at 94°C for 1 min, annealing at 65°C for 1 min (0.7°C per cycle), extension at 72°C for 1 min followed by 24 cycles of denaturation at 94°C for 1 min, annealing at 58°C for 1 min, extension at 72°C for 1 min and a final extension at 72°C for 5 min. PCR was carried out in a total volume of 10 μl containing 100 ng genomic DNA, 10 pmol of each primer, 1× PCR reaction buffer, 200 μM each dNTP, 2.5 mM MgCl2 and 0.25 U GoTaq Flexi DNA polymerase (Promega). PCR products were resolved on 2 per cent agarose gels and visualized by ethidium bromide staining.

(c) Analysis of single-copy nuclear DNA sequences

Primers were designed to amplify 20 single-copy nuclear DNA (scnDNA) regions based on expressed sequence tag (EST) data from GenBank. Accession numbers and primers are given in table S3 in the electronic supplementary material. After an initial examination of levels of polymorphism and use of the Hudson & Kaplan [35] test in DNAasp (v. 4.90.1) [35,36] to eliminate regions exhibiting recombination (nine of the 20 regions), the two most variable loci (CO653303 and CO653370) were selected for analysis.

PCR was carried out on eight individuals from each population as described for the mitochondrial DNA (mtDNA) except that an annealing temperature of 58°C was used, and the PCR products sequenced commercially (Macrogen, Korea).

(d) Microsatellite analysis

Microsatellite markers were developed by searching C. crispus EST sequences in the GenBank database for all trinucleotide repeat motifs, as described by Provan et al. [37]. Primer sequences are given in table S4 in the electronic supplementary material. All 354 individuals were genotyped. PCR was carried out in a total volume of 10 μl containing 100 ng genomic DNA, 10 pmol of dye-labelled M13 primer (6-FAM or HEX), 1 pmol of tagged forward primer, 1 pmol reverse primer, 1× PCR reaction buffer, 200 μM each dNTP, 2.5 mM MgCl2 and 0.25 U GoTaq Flexi DNA polymerase (Promega). PCR was carried out on an MWG Primus thermal cycler using the following parameters: initial denaturation at 94°C for 3 min followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 58°C for 1 min, extension at 72°C for 2 min and a final extension at 72°C for 5 min. Genotyping was carried out on an ABI3730xl capillary genotyping system. Allele sizes were scored using ROX-500 size standards and were checked by comparison with previously sized control samples. Scoring was carried out using the GENEMAPPER software package (v. 4.1; Applied Biosystems).

(e) Palaeodistribution modelling

Ecological niche modellong was carried out to determine suitable climate envelopes for C. crispus at the LGM (~18 kyr ago) using the maximum entropy approach implemented in the MAXENT software package (v. 3.2.1; [38]). Present-day species occurrence data for C. crispus were downloaded from the Global Biodiversity Information Facility data portal (www.gbif.org). A distribution model based on current-day sea-surface temperature (SST) at 1° resolution [39] was generated using MAXENT with the default parameters for convergence (10−5) and number of iterations (500), and projected onto reconstructed LGM data [39] to identify potential refugial areas. Duplicate records from the same locality were removed to reduce the effects of spatial autocorrelation. A presence threshold was determined using the sensitivity–specificity sum maximization approach [40] and the performance of the model was tested using 25 per cent of the occurrence data points to determine the area under the receiver operating characteristic (ROC) curve.

(f) Data analysis

Gene diversity (H) based on mtDNA haplotypes and microsatellites and numbers of unique haplotypes/alleles were calculated for each population. In addition, populations were also grouped into regions based on putative refugia and recolonized areas representing Iberia (population nos 1 and 2), English Channel (population nos 3–6), Ireland (population nos 7–9), Scandinavia (population nos 10 and 13), Northern Europe (population nos 11 and 12) and North America (population nos 14–19). Levels of within-population and within-region gene diversity were calculated using the ARLEQUIN software package (v. 3.01) [41]. Owing to small sample sizes (n = 8 in each case), diversity statistics based on the two scnDNA regions were not calculated.

To evaluate alternative recolonization scenarios for Europe, we employed a model testing approach as described...
Putative refugial regions were identified based on a combination of the palaeodistribution modelling, the presence of private haplotypes and previously published phylogeographic evidence (summarized in table S5 in the electronic supplementary material). For computational feasibility, populations were pooled into groups representing refugia (Iberia and English Channel—two separate groups as defined previously) and recolonized areas (Ireland, Northern Europe and Scandinavia—three separate groups), and maximum-likelihood migration rates between groups and values of $Q$ were calculated from the two-locus scnDNA data using MIGRATE-N (v. 3.0.3) [43]. For each recolonization model, runs were started using values of $Q$ and $M$ (migration rate) calculated from $F_{ST}$ values and consisted of 10 short chains of 50 000 generations from which trees were sampled at 50 generation intervals (1000 trees sampled) followed by three long chains of 1 million generations from which trees were sampled at 100 generation intervals (10 000 trees sampled). The first 20 000 generations were discarded as a burn-in period and the results of the final chains were combined to generate values of $Q$ and $M$. The analysis was then re-run based on an unconstrained migration model to carry out likelihood ratio tests relative to the various recolonization models. These were then used to calculate the Akaike information criterion (AIC) [44] to obtain relative measures of support for each model. The equivalent analysis was not carried out for North America owing to the difficulty in identifying precisely the location of refugial areas.

3. RESULTS

(a) Mitochondrial single nucleotide polymorphism analysis

Seven SNPs were detected in the ascertainment set and AS-PCR of these SNPs in 356 individuals from 19 populations revealed a total of 10 haplotypes (figure 2; table S6a,b in the electronic supplementary material). The dominant haplotype (shown in red) was found in 18 of the 19 populations studied, the exception being the Portuguese population. Three of the remaining haplotypes were endemic to Iberia, and were found in...
Genetic diversity at a species’s rear edge  J. Provan & C. A. Maggs 5

approximately 75 per cent of the individuals studied from the Portugal (no. 1) and Spain (no. 2) populations. A further two haplotypes were endemic to North America, and a single haplotype was found only in the Torquay (no. 6) population in the English Channel. All populations from the recolonized areas of Europe displayed one or both of the two most common haplotypes with the exception of the Galway (no. 7) population, where two individuals each possessed a haplotype found in the Iberian populations. Gene diversity values based on mtDNA SNP haplotypes (table S6a,b in the electronic supplementary material) by population ranged from zero (several populations) to 0.692 (Spain), and by region ranged from 0.214 (Northern Europe) to 0.714 (Iberia).

(b) Single-copy nuclear DNA analysis
The two scnDNA loci, CO653303 and CO653370, exhibited 10 and 9 haplotypes, respectively (figure 2; table S6a in the electronic supplementary material). The Iberian populations possessed four and two private haplotypes at the two loci, while North American populations contained two and one endemic haplotypes, respectively (figure 2; table S6b in the electronic supplementary material). For both loci, recolonized European populations shared all their haplotypes with English Channel populations, with the exception of a single private haplotype in the Øygarden (no. 11) population at locus CO653370.

(c) Microsatellite analysis
Analysis of eight microsatellite loci revealed between 2 and 12 alleles per locus (average 5.25; table S6a,b in the electronic supplementary material). Gene diversity values by population ranged from 0.105 (Øygarden) to 0.436 (Portugal), and by region ranged from 0.181 (Northern Europe) to 0.343 (Iberia). Six populations exhibited private alleles, ranging from one in the Portugal and New Brunswick (no. 17) populations to three in the Roscoff (no. 4) population. Based on regions, two private alleles were found in Ireland, Northern Europe and North America, three in the English Channel and four in Iberia.

(d) Palaeodistribution modelling
For the ecological niche model generated by MAXENT under current SSTs, the value of 0.823 for the area under the ROC curve indicated a prediction that was better than random. A binomial test of omission also indicated that the model predictions were significantly better than random (p < 0.001) for all typical threshold values. Palaeodistribution modelling based on reconstructed SSTs suggested the presence of refugial areas for C. crispus on either side of the North Atlantic during the LGM (figure 1). On the western side of the North Atlantic, the potential distribution was limited to latitudes between approximately 37°N and approximately 44°N, which includes the area around Georges Bank and extends northwards to the Nova Scotian shelf. These latitudes correspond to those inhabited by the present-day Long Island Sound, New Hampshire, Nova Scotia and New Brunswick (nos 14–17) populations, although their current locations would have been dry land at the LGM owing to the decrease in sea levels. The modelled LGM distribution on the eastern side of the North Atlantic was more extensive, ranging from approximately 30°N to approximately 50°N, and included the present-day Iberian populations. It should be borne in mind, however, that the relatively low resolution (1°) of the SST grids used means that they might not completely reflect true intertidal temperatures, either in the current day or at the LGM. Nevertheless, they do provide a broad-scale indication of species occurrence.

(e) Postglacial recolonization model selection
Based on the results of the palaeodistribution modelling, the occurrence of private haplotypes/alleles (table S6a,b in the electronic supplementary material) and evidence from previous phylogeographic studies (see §2), we identified three putative refugia: Iberia and the English Channel in the northeastern North Atlantic, and a further refugium in the northwestern North Atlantic. The precise location of the refugial area(s) on the western side of the Atlantic could not be determined based on the same criteria, but the occurrence of private haplotypes/alleles for all three marker types analysed suggests in situ persistence of C. crispus during the LGM. For Europe, the model testing (figure 3) indicated that a stepping-stone pattern of recolonization of Europe from the English Channel refugium with the Iberian populations remaining isolated was the most likely scenario (AIC = 1067; weight = 0.535). All scenarios involving stepping-stone recolonization were deemed more probable than any of those based on direct recolonization from refugia.

4. DISCUSSION
The findings of the present study suggest that while rear-edge populations of C. crispus in the European North Atlantic are reservoirs of unique genetic variation, this is not the case for the corresponding rear-edge populations in North America. Because the Iberian refugium did not contribute to the postglacial recolonization of the northern North Atlantic by C. crispus to the same extent as the English Channel populations, probably owing to the general lack of suitable substrate in the Bay of Biscay, populations from the area possess a particularly high proportion of endemic haplotypes and alleles across all the marker types analysed (mtDNA SNPs, scnDNA and microsatellites). A similar scenario in the brown seaweed Fucus serratus based on mtDNA sequences led Hoarau et al. [27] to conclude that Iberian populations of the species represent ‘unstable remnants’ of a glacial refugium. Populations of both species in the region frequently suffer stochastic extinction events during hot summers, and although C. crispus thalli exhibit an upper thermal tolerance of 28–29°C (water temperature) [45], these temperatures may be exceeded in intertidal rock pools. Given that predictions of future climate indicate a rise in both air and SSTs [46], there is thus a real danger that the unique genetic diversity maintained in southern-edge populations will be lost. Range-edge populations, particularly those at low latitudes, have been previously highlighted as possibly requiring special conservation efforts [12,14,47]. They often encompass a disproportionate amount of the total genetic diversity across a species’s range [48,49]. As these populations frequently comprise the most evolutionarily divergent
lineages (e.g. the light green haplotype at locus CO653303), they are often active areas of speciation [50,51]. Consequently, loss of the Iberian populations of *C. crispus* might have a major detrimental effect on the evolutionary potential of the species. This might also apply to other marine organisms that have recently been shown to possess unique genetic variation in Iberian populations, such as the ascidian *Botryllus schlosseri* [52], the bryozoan *Celleporella hyalina* [46], the sand goby *Pomatoschistus minutus* [53], and the mysid *Neomysis integer* [54].

For sessile species with limited dispersal capabilities in particular, such as seaweeds and bryozoans, the inability to track changes in climate may further compound the risk of extinction and subsequent loss of the unique genetic diversity found at the rear edge. The statistical phylogeographic analysis indicated that Europe was recolonized by *C. crispus* via a stepping-stone model, rather than by direct long-distance dispersal from refugial areas. Fertilization in the genus *Chondrus* occurs in situ and dispersal is via diploid carpospores or haploid tetraspores, and occasionally via vegetative fragments [55]; but spore viability in seaweeds is generally limited to 24–48 h [56], and a previous population genetic study in *C. crispus* indicated limited gene flow and isolation by distance [57]. Furthermore, surveys of the volcanic island of Surtsey [58] at intervals of 5 years had not recorded the species as late as 1997, 32 years after its initial colonisation by *C. crispus*. Phylogeographic analysis indicated that Europe was recolonized from an *in situ* refugium, the precise location of the refugial area or areas is difficult to identify. The area south of the southern limit of the Laurentide ice sheet (south of Long Island Sound) lacks suitable rocky substrate for intertidal species [59,60] and it was generally assumed that the present-day occurrence of such taxa on Northwest Atlantic coastlines was the result of postglacial recolonization from Europe following extirpation of local populations during the LGM [60]. Recent geological reconstructions and modelling studies, however, have indicated that the area around the Grand Banks remained unglaciated at the LGM [61–63]. Nevertheless, the exposure of the continental shelf owing to the drop in sea level may have provided suitable habitat for rocky intertidal species [60], and this idea is supported by phylogeographic evidence for persistence of several intertidal taxa around the northeastern coast of North America throughout the LGM [27,29,60–66].

The occurrence of suitable refugial habitats around the George’s Bank and the southern Nova Scotian Shelf area as indicated by the palaeodistribution modelling is less likely, since most reconstructions suggest that the Laurentide ice sheet extended to the edge of the continental shelf in the Gulf of Maine [62,63,67]. Nevertheless, our findings based upon multiple molecular markers support the idea of a northeastern North American refugium for *C. crispus*, rather than recolonization from Europe, although this American refugium did not include present-day rear-edge populations. The most probable location for the refugium, based on the balance of evidence across the different markers studied, would appear to be in the area of the Canadian Maritime provinces, or further north in Newfoundland. This region has previously been proposed as a refugium for
the hermit crab *Pagurus longicarpus* [64], the hydrozoan *Obelia geniculata* [65] and the periwinkle *Littorina littorea* [66].

This fact that the genetic signatures of glacial refugia were not obvious in North America, unlike Europe, might also reflect the prediction that ‘stable’ rear-edge populations (i.e. those that have persisted in the same areas throughout glacial and postglacial periods) should preserve a higher proportion of genetic variation than ‘trailing’ rear edges, which have migrated since the LGM [12]. The palaeodistribution modelling indicated that the present-day Iberian populations lie within the putative refugial area on the eastern side of the North Atlantic, and thus represent a stable rear edge. Current-day populations of *C. crispus* in the western North Atlantic, however, are situated north of the putative shoreline at the LGM and must have recolonized after the post-LGM rise in sea levels, which would class them as trailing rear edges. Likewise, populations that persisted in the Hurd Deep refugium in the English Channel [26] must have migrated to their current locations since the LGM. Under the two scenarios proposed by Hampe & Petit [12] (stable versus trailing), this would explain the fact that the Portuguese population exhibits by far the highest level of microsatellite diversity in Europe, whereas levels of microsatellite diversity are more consistent across the North American populations. The Iberian populations also exhibited far higher levels of mtDNA diversity than those found elsewhere in Europe. As many of these rear-edge populations are under threat of extinction from climate change, our findings suggest that such populations merit special attention in order to preserve the range-wide evolutionary potential of the species, and in particular those that might represent stable rear edges. These findings also further confirm the importance of historical forces in shaping present-day patterns of genetic variation across species ranges, particularly for species with limited potential for dispersal.

The authors are extremely grateful to everybody who provided samples for this study. We would also like to thank Ruth Kelly for assistance in the laboratory, Marcus Pfenninger for advice on data analysis, Grant Bigg for providing the GIS layers used in the distribution modelling and four anonymous referees for valuable comments on an original version of the manuscript.

**REFERENCES**


