ECOLOGY LETTERS

Ecology Letters, (2012) doi: 10.1111/j.1461-0248.2012.01779.x

LETTER

Predicting present and future intra-specific genetic structure through niche hindcasting across 24 millennia

Anahí Espíndola,[*†] **Loïc Pellissier,**[†]
Luigi Maiorano, Wim Hordijk,
Antoine Guisan[‡] **and Nadir Alvarez**[‡]
Department of Ecology and Evolution, Biophore Building, University of Lausanne, 1015 Lausanne, Switzerland

[†] *Anahí Espíndola and Loïc Pellissier equally contributed to this study and should be considered as co-first authors.*
[‡] *Nadir Alvarez supervised Anahí Espíndola and Antoine Guisan supervised Loïc Pellissier.*
[*] *Correspondence: E-mail: anahi.espindola@gmail.com*

Abstract

Paleoclimatic reconstructions coupled with species distribution models and identification of extant spatial genetic structure have the potential to provide insights into the demographic events that shape the distribution of intra-specific genetic variation across time. Using the globeflower *Trollius europaeus* as a case-study, we combined (1) Amplified Fragment Length Polymorphisms, (2) suites of 1000-years stepwise hindcasted species distributions and (3) a model of diffusion through time over the last 24,000 years, to trace the spatial dynamics that most likely fits the species' current genetic structure. We show that the globeflower comprises four gene pools in Europe which, from the dry period preceding the Last Glacial Maximum, dispersed while tracking the conditions fitting its climatic niche. Among these four gene pools, two are predicted to experience drastic range retraction in the near future. Our interdisciplinary approach, applicable to virtually any taxon, is an advance in inferring how climate change impacts species' genetic structures.

Keywords

AFLP, climate change, forecasting, hindcasting, past climate, phylogeography, post-glacial contraction, recolonisation, species distribution models, *Trollius europaeus*.

Ecology Letters (2012)

INTRODUCTION

Current distributions are predominantly the result of the interaction between species' environmental requirements (the niche *sensu* Hutchinson) and geographic variation of key environmental factors (the realised environment; Guisan & Thuiller 2005), among which climate plays a predominant role (Araújo & Pearson 2005). In the context of Quaternary climatic oscillations, the ebb and flow of glacial dynamics caused dramatic species range expansions and retractions, involving local extinction, migration, drift and adaptation (Hewitt 1999). The genetic structure of species is thus intimately related to spatial and temporal variation in their distribution ranges, which in turn shapes the pattern and frequency of inter-population genetic exchanges.

Changes in climate have often fragmented or reconnected populations of the same species, with gene flow among populations being accordingly restricted or enhanced (Hewitt 1999). As a result, the genetic variation within species has been structured spatially in distinct and variably isolated gene pools. Identifying and explaining these genetic structures has been the aim of phylogeography over the last 20 years (Avise 2009) and has gained new attention due to the possible implications for forecasting the distribution of gene pools under climate change (Etterson & Shaw 2001; Davis *et al.* 2005). Phylogeographic patterns in Europe have been summarised into a few paradigms for temperate organisms, such as the ones represented by the European beech *Fagus sylvatica,* the hedgehog *Erinaceus europaeus* and the brown bear *Ursus arctos* (e.g. Hewitt 1999), showing genetically diverse gene pools that survived cold periods in Southern peninsulas.

In contrast, cold-adapted species have only recently received attention, when compared to temperate species, and knowledge of their phylogeographic patterns has not yet been summarised into major paradigms (e.g. Weider & Hobaek 2000; Abbott & Brochmann 2003; Brochmann *et al.* 2003; Ehrich *et al.* 2007). While phylogeographic investigations on temperate organisms provide information on the genetic consequences of post-glacial expansion processes, the study of cold-adapted species that are currently found fragmentarily distributed and under refugial conditions in most of their range (e.g. at high altitudes) unravels the genetic consequences of distributional contractions related to climate warming.

Species distribution models (SDMs) rely on the analysis of the climatic conditions shaping the distribution of species (Guisan & Thuiller 2005). In the last decade, a few studies (Hugall *et al.* 2002; Alsos *et al.* 2009; Freedman *et al.* 2010; Vega *et al.* 2010; Beatty & Provan 2011) have used hindcasted SDMs in combination with phylogeographic surveys to detect past spatial discontinuities in species distributions, and thus to identify possible past barriers to gene flow. However, very few points in time were considered in studies so far – the present, the Last Glacial Maximum (LGM) and sometimes an additional mid-point between these two – providing limited information on past range changes, and therefore neglecting the quantitative identification of detailed spatial dynamics. Circumventing this limitation has been roughly achieved, for instance, by assuming a linear change in temperature and precipitations from 9 to 18 kya (Graham *et al.* 2010). However, the lack of accurate climatic estimates at several time periods between the present and the LGM precluded a thorough interpretation of range-shifts through time. Hindcasting the distribution of species based on a larger number of more continuous time steps (Maiorano *et al.* in press) and combining these with modelled dispersal rules represent an advance to decipher a species' spatial and genetic histories.

Here, we use a global circulation model to reconstruct past climates and thus predict short-time stepwise species potential distributions, going back to the LGM and further into the past across the last 24 millennia. Furthermore, we show how this information can be integrated with large phylogeographic surveys and with a spatial model of expansion/contraction filtered by habitat suitability (hereafter referred to as diffusion model) to understand a species' spatio-temporal history. This approach provides key data on the putative location of the past distribution of gene pools and on the pathways that they followed during range expansions and contractions across the last millennia. We illustrate the approach by investigating the past distributional dynamics of a cold-adapted species, the globeflower, *Trollius europaeus* L. (Ranunculaceae), by using simulations to infer how past range-variation influenced its current intra-specific genetic structure, and finally identifying how it will be affected in the future. This species is a well-suited case-study because (1) it is associated with cold and moist habitats that experienced drastic spatial reshuffling in the last millennia (Hewitt 1996) and (2) a preliminary phylogeographic survey of the species suggested several spatially structured gene pools (Després *et al.* 2002).

Based on the historical and ecological knowledge of the species, we expect the combined use of phylogeography and spatial modelling to provide detailed information on how the current gene pools moved in space and time to reach their present range, and how they could evolve in a warming future. More precisely, we predict that (1) the past range of the current gene pools (retrieved from phylogeographic analyses) should be identified by hindcasting approaches, (2) the most likely pathways followed by the different gene pools should be recognisable using a spatio-temporal niche-based diffusion model, and (3) future genetic structure can be forecasted based on the combination of allele distribution interpolation and range predictions.

MATERIAL AND METHODS

Genetic data and analysis of the spatial genetic structure of *T. europaeus*

Samples of *T. europaeus* were collected during the springs and summers of 2006–2008 at 79 locations throughout the European range of the plant (Fig. 1a; see Table S1 in Supporting Information). An Amplified Fragment Length Polymorphism (AFLP; Vos *et al.* 1995) procedure was performed with *Eco*RI and *Mse*I endonucleases. Digested fragments were selected with two primer pairs and genotyped (see Appendix S1 in Supporting Information).

The genetic structure of the dataset was identified with ten Markov chain Monte-Carlo runs of 1 000 000 generations with a 200 000 burn-in period for an *a priori* number of gene pools *K*, ranging from 1 to 20 (200 runs in total) as implemented in STRUCTURE 2.2 (Falush *et al.* 2007). The most probable *K* was identified following Pritchard *et al.* (2000). In order to confirm the genetic structure, a non-model-based approach was applied using the K-means clustering technique as introduced by Hartigan & Wong (1979). Following Burnier *et al.* (2009), we identified the number of clusters (*K*) that optimises the inertia of the dataset. The calculations were repeated 10 000 times starting at different random points, and were run in the R 9.2.1 CRAN environment (R Core Development Group 2009). Finally, we performed Principal Coordinates Analyses (PCoA) using custom R-scripts, and an AMOVA using GENALEX 6.3 (Peakall & Smouse 2006).

Climatic data

The current climatic data (averaged from 1950 to 2000) was obtained from the Climatic Research Unit (Mitchell *et al.* 2004). Simulations of past climate were obtained from a general circulation model based on the Hadley Centre climate model (HAD3; Singarayer & Valdes 2010), and used to produce paleo-temperature and precipitation maps at a 15-km spatial resolution over Europe (see Appendix S1). Simulations of future climate were based on three general circulation models (HAD3, CSRIO2 and CGCM2) and four future emission scenarios (a1, a1FI, b1, b2; 12 projections in total) from the Intergovernmental Panel on Climate Change (IPCC 2001) averaged from 2070 to 2100.

Species distribution modelling

To model the distribution of the species, we combined the available occurrences of *T. europaeus* with a set of six bioclimatic variables

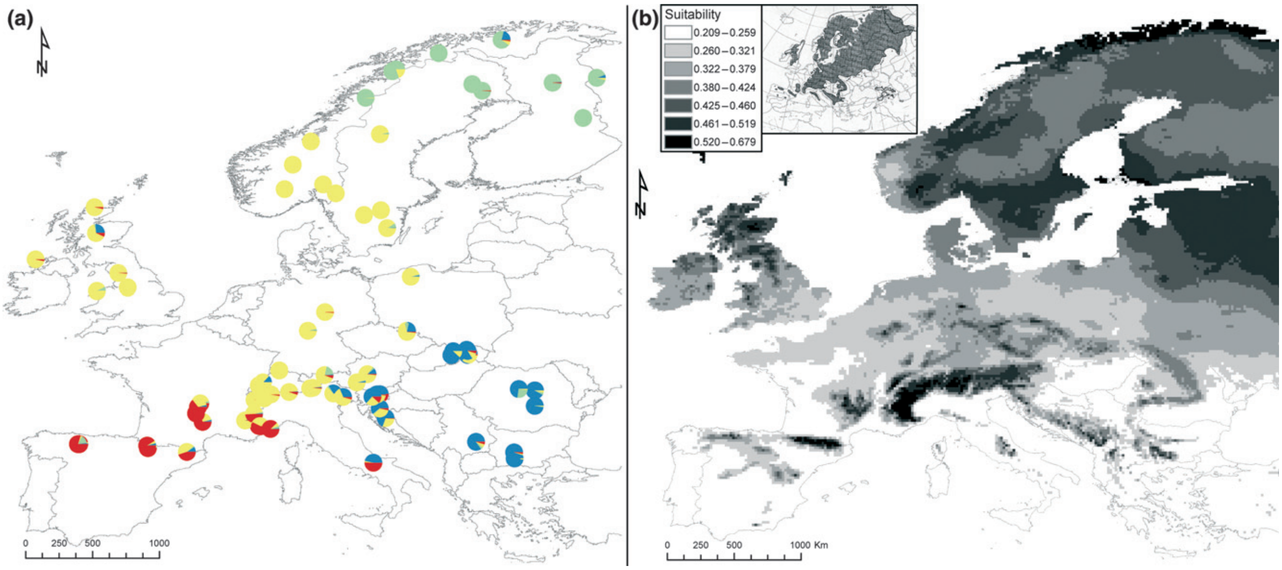

Figure 1 Genetic structure and SDM-based current projection of the potential distribution of *T. europaeus* in Europe. a – Spatial genetic structure of *T. europaeus* according to the STRUCTURE analysis. Proportions in the pies indicate the probability to belong to one of four gene pools (green, Northern Scandinavia; yellow, Western and central Europe; red, South-Western Europe; blue, South-Eastern Europe). b – Predicted current distribution of the European globeflower. Caption: distribution range of the species, modified from Meusel *et al.* (1965).

assumed to be important for alpine plants (Körner 2003): total annual precipitation, summer precipitation, winter precipitation, annual mean temperature, and mean warmest and coldest temperatures. Occurrences of *T. europaeus* with a spatial accuracy of ≤15 km were obtained from published databases and fieldwork (see Appendix S1). This scale was chosen because of its wide application in most biodiversity databases (e.g. GBIF, www. gbif.org) and because it corresponded to the grid-size of all climatic grids. Pseudo-absences were generated by selecting 10 000 random points across Europe and weighted in further analyses to ensure a balance between the prevalence of presences and pseudo-absences. Based on Engler *et al.* (2011), the realised climatic niche was modelled using five modelling techniques (Thuiller *et al.* 2009). We evaluated the predictive performance of each model using a repeated split sampling approach with 50 repetitions, using the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC; Fielding & Bell 1997) as evaluation metric. Following Marmion *et al.* (2009), we calculated for each projection an average of the five modelling techniques weighted by their predictive power (see Appendix S1 and Table S2 in Supplementary Information). The SDM was projected into the past and into the future. We evaluated the accuracy of past range predictions by comparing the hindcasted distribution and the unambiguous fossil record for *T. europaeus* (European Pollen Database; http://pollen.cerege.fr/fpd-epd/).

Distributional dynamics simulation and best-fit scenario identification

We transformed the probability maps obtained from the SDM projections into binary presence/absence maps using the ROC plot method that maximises both sensitivity and specificity (Liu *et al.* 2005). We also considered as unsuitable those regions known to have been covered by ice during each time period (Ehlers & Gibbard 2004; Gyllencreutz *et al.* 2007).

We used simulations to identify the past dynamics of the four currently identified gene pools (see 'Results'). We first randomly chose four suitable pixels to delimitate the original regions occupied by the four gene pools at the cold period previous to the LGM (24 000 years ago, 24 kya). The remaining suitable pixels were then assigned to one of these clusters using a simple proximity rule (i.e. suitable pixels were assigned to a given cluster as a function of their linear distance to the closest starting pixel). This way, the position of the four starting pixels defined the initial distribution of each genetic cluster. Then, for each following time-step (every 1000 years) and up to the present, any suitable pixel in a time frame t could be colonised by the genetic group from the closest suitable pixel in the time frame $t - 1$, a procedure that we refer to as diffusion.

Finally, we compared the fit of the current genetic structure predicted by the simulated scenarios with the empirical population genetic assignments. For this, the majority-rule criterion was applied, assigning each genetically analysed population to a cluster by considering the highest assignment probability obtained when applying genetic clustering approaches (e.g. with STRUCTURE). A population was assumed to be properly recovered by the model if it was assigned to the same genetic cluster both at the end of the simulation process and with the direct molecular approach. The scenario harbouring the highest proportion of recovered populations was considered as the one having most properly recovered the overall current genetic structure. We ran the simulations 10 000 times, providing a sufficient number of possible scenarios to properly examine the most suitable hypotheses. For improved computation speed, we implemented this function, written in C language, in the *MigClim* R package.

Results from the assignments were also evaluated using the assignment test implemented in AFLPOP 1.1 (Duchesne & Bernatchez 2002). Considering the best-fit simulated scenario, four groups were defined *a priori* to further evaluate the samples assignment likelihoods. We used default parameters, with a likelihood threshold sensitivity set to 0.1. In order to compare this output with a standard genetic clustering (i.e. STRUCTURE), we performed the same analysis defining groups based on that clustering method.

Insights into future genetic diversity

We used the population-based observed values of presences and absences of alleles for the whole studied area and interpolated their distribution using an Inverse Distance Weighted (IDW) raster approach (cell size: 15 km) in ArcGIS 9.3 (Environmental Systems Research Institute, Redlands, CA USA). Levels of polymorphism were estimated by combining interpolation of both presences and absences: (1) a locus was considered as present, if the allele was at least present once in a population (identification of monomorphic absences); (2) a locus was considered as absent, if the allele was absent at least once in a population (identification of monomorphic presences). Afterwards, all cells falling within the suitable areas predicted by the different future scenarios were selected and, for each locus, we considered as polymorphic all pixels presenting values of presence and absence ranging between 0.4 and 0.6. Future levels of genetic diversity were calculated by averaging the proportion of polymorphic alleles per cell. Eventual genetic losses were identified by comparing these predicted values with those obtained for current conditions, based on the same interpolated dataset applied to the current projection. We finally identified the regions predicted to be occupied by each gene pool, by assigning each pixel to one gene pool, considering the same proximity rule as for the assignments in the past (see above). We then compared surface occupancy (in number of pixels) of each gene pool, under current and future climatic conditions, for both the whole Europe and each geographic region.

RESULTS

Spatial genetic structure of the globeflower

All clustering methods used to analyse the 374 AFLP fragments amplified in 349 samples consistently identified four gene pools (Fig. 1a; see Fig. S1 and Appendix S1). These gene pools were geographically structured, presented several suture zones (Fig. 1a) and appeared to be well-segregated in the PCoA (see Fig. S2 in Supporting Information). The AMOVA indicated that 4% of the genetic variance was significantly explained by the four gene pools (5.5% explained by variance between populations; 90.5% explained by variance at the intra-population level; see Fig. S2). Our results demonstrated (1) the presence of a cluster specific to South-Eastern Europe, (2) the admixed genetic identity of several locations in the Southern Alps and Eastern Pyrenees and (3) the existence of an exclusive and independent Northern Scandinavian cluster, different from the one found in Southern Scandinavia, the Carpathians and Northern Poland.

Past range dynamics and spatial diffusion of the gene pools

The ensembled SDM properly recovered the current range of the globeflower (Fig. 1b). Projected distributions suggest that, at the earliest time analysed (i.e. 24 kya), *T. europaeus* was more widespread than today (see Fig. S3 in Supporting Information). At that time, the distribution also appeared to be fairly fragmented, with four centres of high climatic suitability observed in the current Balkans, North-Eastern and North-Western Europe and the Baltic zone. Past projections showed that hindcasted regions included the known fossil records (see Fig. S4 in Supporting Information), confirming the accuracy of the model. The 10 000 simulated scenarios provided varying outcomes and fits to the observed data (Fig. 2a; see Fig. S5 in Supporting Information). The two best scenarios properly assigned 83.5% of the current spatial genetic structure. Because they were largely similar, only one of the two is discussed (Fig. 2b; also see Fig. S4). The best 5% of the simulations recovered trends similar to these two best scenarios, assigning the sources of colonization of each of the four clusters to the Pyrenees-Massif Central, to North-Western Europe, to the Balkans and to North-Eastern Europe (Fig. 2b and see Fig. S4). The best range dynamics scenarios (Fig. 2b) indicated that the distribution area of the plant largely increased with the onset of the period generally associated to the LGM (21–18 kya), and got strongly contracted at around 12 kya, to finally reach its current -restricted-range (see Figs S4 and S6 in Supporting Information).

The assignment analysis performed using AFLPOP, with the simulation outputs as a grouping factor, recovered similar assignment fits than our simulations, with 17.6% of samples being assigned to other genetic clusters (vs. 6.7% using the STRUCTURE clusters).

Future distribution

Range predictions for 2070-2100, under three climatic models and four emission scenarios, showed that the distribution of the globeflower will be strongly modified in the future (Fig. 3 and see Fig. S6). The species distribution is predicted to shift to higher elevations and latitudes, leading in some cases (e.g. scenario CGCM-a2) to a final increase in the total occupied area (Table 1), following a wide colonisation of Scandinavia, and despite extinctions in Southern European regions. Such extinctions are expected to strongly reduce the ranges of two of the four gene pools (red and blue; Table 1 and Fig. 3), which are predicted to respectively survive at the Western and Eastern edges of the Alps under some scenarios only (see Fig. S6). Moreover, based on interpolations of allele polymorphism in future

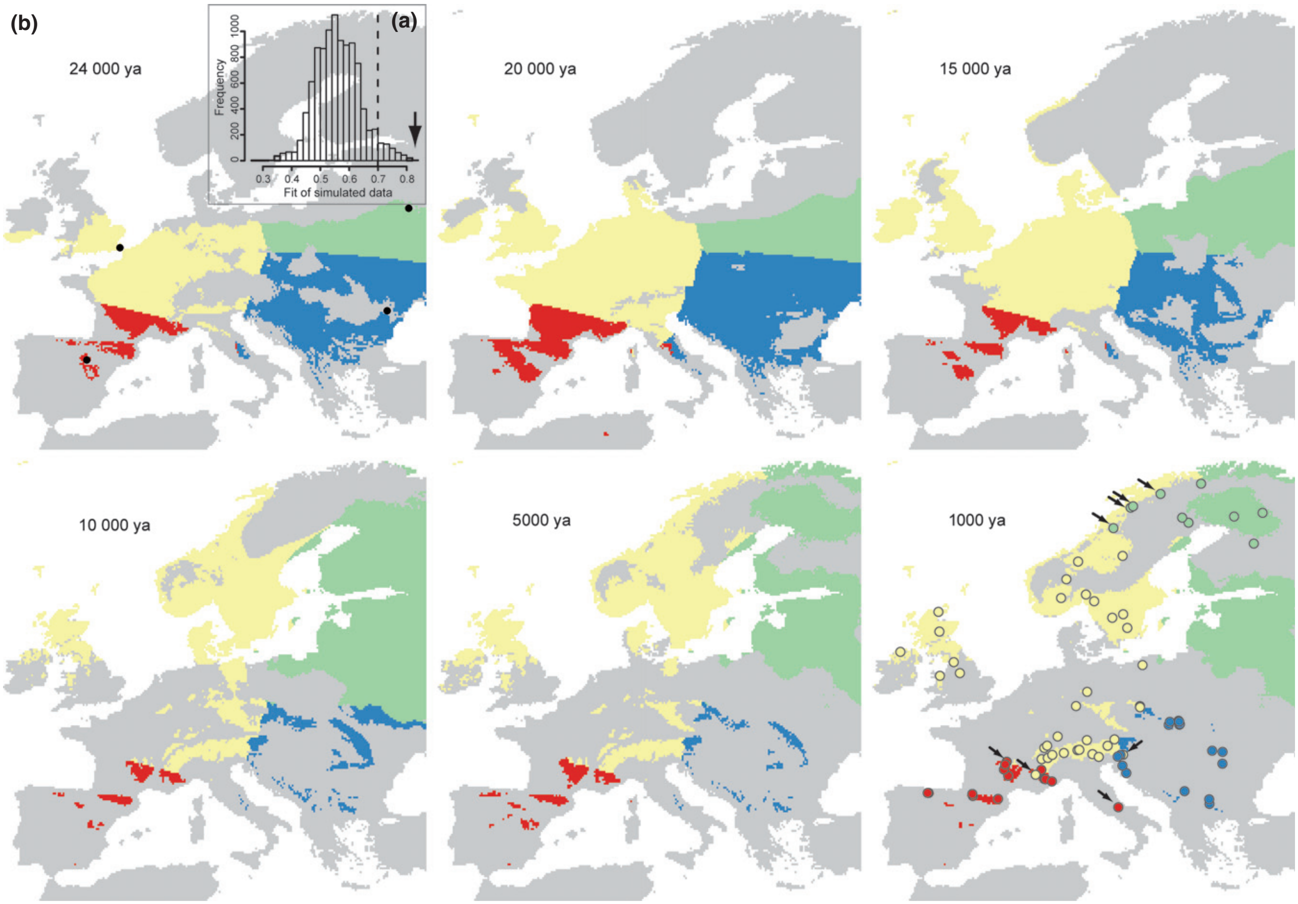

Figure 2 Simulated scenarios. a – Distribution and 95% quantile of fit of 10 000 simulations. The arrow indicates the two fittest simulations that correctly assigned 83.5% of the genetically analyzed populations. b – The hindcasted distribution of the gene pools, as defined by one of the two fittest scenarios for *T. europaeus*. Six time-points are shown. Colours represent the gene pools identified by STRUCTURE. Grey regions represent unsuitable areas; i.e. areas falling below the ROC threshold and/or covered by ice. Dots in the first frame indicate the random starting centers. Circles in the last frame indicate the observed gene pools, as considered for the evaluation of fit; arrows indicate incongruence between results from the diffusion model and the genetic clustering analysis.

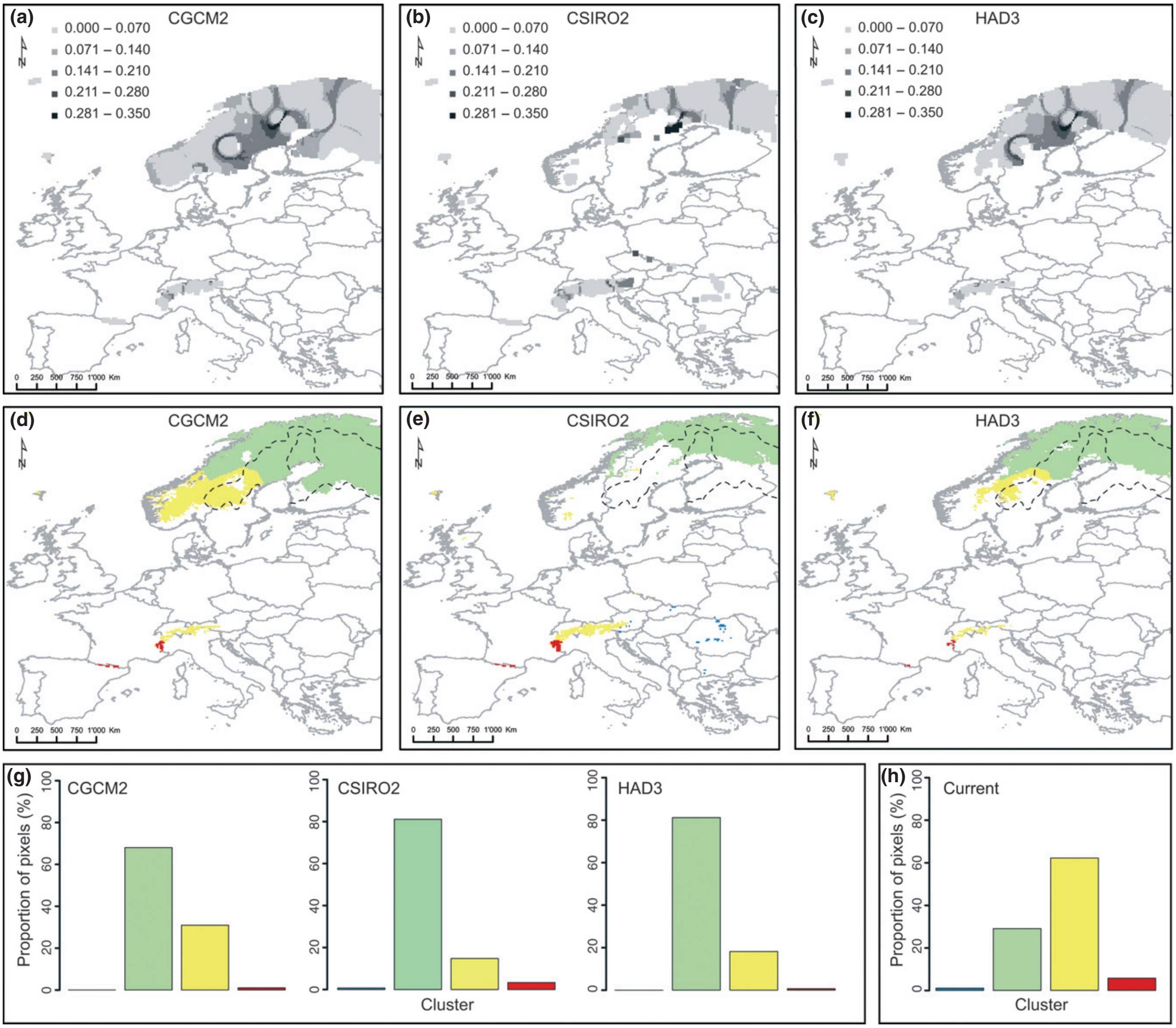

Figure 3 Predicted range, pattern of future genetic diversity (proportion of polymorphic loci; gradient of grey; a–c), future distribution (d–f) and gene pool occupancies (proportion of occupied pixels; g) of *T. europaeus* under three global circulation models for the a1FI future scenario (averaged from 2070–2010; Mitchell *et al.* 2004). a and d – CGCM2 model; b and e – CSIRO2 model; c and f– HAD3 model. h – Current area occupied by each gene pool (colours as in Fig. 1). The dashed line in d–f indicates the current Scandinavian range of *T. europaeus*. Blotches of shaded areas in a–c are a mapping consequence of the coarse-grain IWD interpolation algorithm.

suitable areas, while some scenarios predict almost invariable diversity values compared to current conditions, other forecast losses of up to 28% of the genetic diversity (Table 1). Under the CSIRO2 circulation model, the percentage of genetic diversity loss appeared to be linked to the strong decrease of the occupied surface. In all cases, Scandinavia is predicted to harbour the most diverse areas, causing, under some scenarios, an overall increase in the genetic diversity of *T. europaeus* (Table 1; Fig. 3). Scandinavia is also expected to harbour more than 80% of the future surface occupancy (against 67% nowadays), whatever climatic scenario is considered (see Fig. S6).

DISCUSSION

The interdisciplinary approach proposed in this study allowed modelling the distribution dynamics that shaped the genetic structure of *T. europaeus* over the course of the last 24 millennia. We demonstrated here that, in the case of *T. europaeus*, the main trends in the range dynamics of its gene pools across the last millennia can be reconstructed using spatial simulations closely matching empirical data (83.5%). Our hindcasting-based approach therefore supports the use of niche-based SDMs to predict the fate of species and gene pools in changing climatic conditions. We finally showed that projecting our approach into the future allows predicting the expected variation of the range covered by each gene pool. We also determined which of these gene pools would be the most threatened according to different climate warming scenarios. Our results also identified a variable loss of genetic diversity when associating SDMs to interpolated levels of polymorphism. In the next sections, we discuss these findings in more detail as a way to illustrate the phylo- and biogeographic information one can additionally gather with this approach.

Table 1 Predicted future genetic diversities (proportion of polymorphic loci) and variation of surface occupancy for the species and for each gene pool under three global circulation models and four emission scenarios. Differences to current values are shown.

Model	Average proportion of polymorphic sites	Difference to current genetic diversity (%)	Difference to current occupied surface (%)	Difference to current occupied surface for each gene pool (%)			
				Blue	Green	Yellow	Red
CGCM2-a1FI	0.0639	1.310	1.21	−100.00	58.16	−34.27	−79.75
CGCM2-a2	0.0643	1.846	13.68	−100.00	71.71	−21.41	−70.89
CGCM2-b1	0.0665	5.428	−4.01	−99.16	27.46	−19.56	−57.59
CGCM2-b2	0.0609	−3.487	5.86	−99.58	50.60	−19.71	−60.13
HAD3-a1FI	0.0649	2.847	−33.93	−100.00	23.18	−74.84	−91.14
HAD3-a2	0.0629	−0.290	−12.39	−99.79	39.25	−45.47	−80.38
HAD3-b1	0.0655	3.867	7.69	−95.79	67.00	−31.07	−60.76
HAD3-b2	0.0639	1.301	−8.35	−98.95	48.67	−46.32	−74.05
CSIRO2-a1FI	0.0475	−24.771	−55.36	−91.37	−16.83	−86.14	−70.25
CSIRO2-a2	0.0452	−28.440	−53.96	−96.21	−17.49	−81.94	−72.78
CSIRO2-b1	0.0548	−13.167	−43.53	−93.26	4.97	−81.94	−62.03
CSIRO2-b2	0.0536	−15.062	−43.02	−94.32	−0.18	−76.29	−60.13

As many past centres of distribution as genetic pools

The total hindcasted range at 24 kya was in general congruent with the LGM hindcasted distributions of other cold-adapted species from the region (see Fig. S3; e.g. Svenning *et al.* 2008) and indicated the presence of four centers of high habitat suitability. Based on more samples and loci, a wider and more regular spatial covering and new analyses, the spatial genetic structure identified in our results provides more informative insights into the phylogeographic pattern of the European globeflower, when compared to a previously more restricted study of this species (Després *et al.* 2002). We clearly identified the presence of four gene pools (vs. three supported groups in Després *et al.* 2002), spatially structured (Fig. 1a; see Fig. S2) and restricted to South-Eastern (in blue, Fig. 1a), South-Western (in red), North-Western and Central (in yellow), and North-Eastern Europe (in green).

The best-fit scenarios identified the Pyrenees-Massif Central, North-Western Europe, the Balkans and North-Eastern Europe as the most likely sources of colonisation of the current range (Fig. 2b; see Fig. S4), which largely corresponds to the distribution of high suitability values for *T. europaeus* populations at 24 kya, during the cold period preceding the LGM in Europe (see Fig. S3). This phase corresponded to a cold maximum, which in Europe, was associated with an arid period (Watts *et al.* 1996) that might have been too dry for the globeflower to widely colonise the region.

It is worth noting that, besides identifying the most likely past sources of colonisation, our procedure allowed for the recognition of a genetic relationship between regions spatially close but separated in the past by a narrow unsuitable zone, such as the region laying between the Pyrenees and the Massif Central (see Fig. S3).

Niche-based diffusion modelling helps identify the most likely phylogeographic dynamics

The diffusion model allowed identifying two best-fit scenarios that correctly assigned 83.5% of the genetically analysed locations and, therefore, accurately tracked the paths followed by the globeflower's gene pools across the last millennia. The two scenarios concurred in showing that the climatic variations characterizing the last 24 ky largely drove the range contractions and expansions experienced by *T. europaeus* (Fig. 2b; see Figs S4 and S6). While the species had a relatively restricted range at 24 kya, its distribution showed an expansion between 20 and 16 kya, probably related to the decrease of aridity in Europe (Watts *et al.* 1996). After the beginning of the LGM, the species remained widespread for several thousand years, with contact zones appearing between the gene pools. The establishment of a long lasting warming period at around 15 kya caused the start of a strong contraction in the species range. From then on, the gene pools moved to higher latitudes, and to higher elevations in the Southern European mountain ranges. The cold Dryas period induced a modest range expansion around 12 kya since the aridity did not increase (Watts *et al.* 1996), which likely favoured the spread of *T. europaeus.* After the end of this cold phase, the general tendency towards a climatic warming became established (Raymo 1997; Labeyrie *et al.* 2003). This warmer period induced the final deglaciation of Northern Europe and of the Southern mountain ranges, allowing a progressive colonisation of regions previously covered by ice, such as Scandinavia and the British Islands. As the climate began to warm up, the North-Western and North-Eastern European clusters, which reached Scandinavia between 10 and 5 kya, continuously expanded their ranges until a contact zone arose in Northern Scandinavia. In contrast, Southern gene pools of the species (i.e. South-Eastern and South-Western) remained restricted to Southern European massifs and suffered a progressive range reduction.

Besides allowing for the identification of the most likely phylogeographic scenarios (Fig. 2), our results explain the origin of contact zones. These were identified in the Alpine range, in the Sudetes, in central Scandinavia and in Northern Scotland (Fig. 1a). While the first three regions correspond to contact zones between neighbouring genetic lineages (yellow, blue and red clusters in the Alps; yellow and blue clusters in the Sudetes; green and yellow clusters in central Scandinavia; Fig. 1a), the fourth represents a more unexpected case of admixture that may be explained by natural or anthropogenic long-distance dispersal. The Alpine contact zone might be both a consequence of the ancient presence of the lineages in the region (particularly for the Southernmost Italian population; Figs 1a and 2b) and the centripetal progressive colonisation of higher elevations, related to the temperature increase of the last thousands years (Fig. 2b). The best phylogeographic scenarios (Fig. 2b) help explaining the contact zone in the Sudetes (Fig. 1a), suggesting the presence of the two gene pools in the region during the LGM. They also allow explaining the formation of the Scandinavian contact zone. Indeed, this region was unavailable (i.e. under ice-sheets) or only locally available (Parducci *et al.* 2012) during the coldest glacial phases, and appears to have been largely colonised during the climate warming that happened between 10 and 5 kya. Our results are thus a new evidence in plants (contrasting with results obtained in Schönswetter *et al.* 2006; Skrede *et al.* 2006; Schmitt 2009; and in agreement with Parducci *et al.* 2012) that Scandinavia was colonised by two distinct lineages that established a contact zone at mid latitudes. It is however important to note that, in contrast to Parducci *et al.* (2012), we did not find any genetic signature (e.g. presence of private alleles) suggesting a survival of *T. europaeus* in Scandinavian glacial refugia.

Incongruence between empirical and modelled spatial genetic structures

The comparison of simulation- and genetic-based assignments showed some incongruences (Fig. 2b), most likely corresponding to (1) incorrect assignment of populations at contact zones and (2) long-distance dispersal events. Overall, however, incongruent assignments accounted for less than one-sixth of all populations, attesting the relevance of the approach used here. The most apparent incongruence occurred at the Scandinavian contact zone, which was shifted to the North in our simulations. This may be due to the fact that in our simulations South-Western Scandinavia became suitable for the species before the Northern Scandinavian edge was free of ice. Furthermore, the projection of our SDM at 24–10 kya indicated that conditions in Northern Scandinavia were not suitable for *T. europaeus*, probably because of the extreme drought associated to the proximity of ice-sheets. However, recent genetic and paleoecological studies demonstrated that some areas at high latitudes were locally ice-free during the last glaciations, providing putative shelter to cold-adapted species (Westergaard *et al.* 2011; Parducci *et al.* 2012) and suggesting that an early local colonisation of the North-Western Scandinavian coasts by the North-Eastern gene pool could have been possible. However, local ice-cover anomalies are not yet taken into account in the Eurasian ice-sheet reconstructions available for the area and thus could not be considered in our simulations. Alternative scenarios may include a colonisation of the region through a long-distance dispersal event (not considered in our model), and/or more recent and quick demographic dynamics, which could have shuffled the precise location of this contact zone. In order to further investigate these points, once more detailed and updated paleogeographical and paleoglaciological maps become available, we recommend their implementation in simulations modelling diffusion of gene pools.

Predicting species range-shifts and genetic losses in response to climate warming

When forecasting the range of the species for the next 60 years, our results indicated that global warming is expected to negatively affect the range of the globeflower under some scenarios only (Table 1), differentially influencing each gene pool (Fig. 3; see below and Fig. S6), and sometimes leading to a strong decrease in the European genetic diversity of the species (Table 1). In contrast to temperate species, cold-adapted species mainly show contracted refugial distributions during interglacial periods, such as the ongoing one (Stewart *et al.* 2010). Under the climatic warming scenarios considered here, the range of the plant is predicted to be more restricted to high elevations and latitudes than today, with the species going almost extinct in most Southern mountain ranges (Fig. 3; see Fig. S6). The Alps are predicted to represent the widest future South-European refugia, presenting the largest suitable area in the region and confirming an idea previously suggested by Alsos *et al.* (2009). Our predictions additionally indicate that central Scandinavia will likely harbour the largest genetic diversity in Europe (Fig. 3). Most future scenarios predict that a large part of the genetic diversity will be lost in the Southern mountain ranges, and only a few scenarios predict fair genetic diversity preservation in the Eastern Alps and the Sudetes (see Fig. S6).

Our genetic diversity inferences for the future are similar (no loss, or losses of <30%) to a recent predictive analysis done on a set of arctic-alpine plants (Alsos *et al.* 2012). Indeed, the lack of genetic loss in some future scenarios (e.g. CGCM2) is a result of the final surface gain, which is particularly high in the diverse central Scandinavia (Fig. 3d–h). Genetic diversity values fall only when the range becomes patchy in the whole region, what was observed under all emission scenarios applied to the CSIRO2 circulation model (Table 1). It is thus not straightforward to infer a simple relationship between genetic and surface losses, since it appears that several factors (e.g. level of occupancy of central Scandinavia, range patchiness) bring variation to these values.

With regard to the predicted ranges of current gene pools (Fig. 3d–g and Fig. S6), the South-Eastern one is expected to experience the greatest range contraction, especially at the core of its range (i.e. the Balkans). Our predictions further indicate that the South-Western gene pool is likely to become extremely fragmented in the Pyrenees and nearly disappearing from the Iberian Peninsula, which is in agreement with the predictions made for other European cold-adapted species (Alsos *et al.* 2009). In contrast, the Alpine region is in some cases likely to continue to harbour the three gene pools (red, yellow and blue) that have colonised the Western, Eastern and central edges of the Alps in the last millennia, confirming what has been recently proposed for other cold-adapted species from the region (Treier & Muller-Scharer 2011).

CONCLUSION

The novel simulation approach used in this study allowed unravelling the most likely scenarios of intraspecific gene pool diffusion across time for the last 24 ky. By being virtually applicable to any species, and assuming that enough occurrences are available for calibrating SDMs, this predictive approach opens exciting new research avenues. Besides its obvious application for identifying the most likely phylogeographic scenario for a given species, it may also be used to statistically test user-defined scenarios based on the generation of appropriate null niche-based diffusion models (see Fig. S7 in Supporting Information). This point is appealing when using genetic markers such as AFLPs, for which coalescent-based methods, such as those proposed by Carstens & Richards (2007) are not yet applicable. Moreover, the method may complement coalescent approaches when using sequence-based data, allowing demographic events to be directly dated, mapped and further tested. From this perspective, this study is a breakthrough in interdisciplinary projects and opens new doors in the understanding of the spatial and temporal evolution of species.

ACKNOWLEDGEMENTS

The authors would like to thank R. Arnoux, P. Bowler, O. Broennimann, R. Dafydd, P. Duchesne, R. Gyllencreutz, D. Gyurova, H. Hipperson, M. Howe, L. Juillerat, R. Lavigne, P. Lazarevic, N. Magrou, N. Revel, M. Ronikier, N. Russell, A. Sarr, T. Suchan, Y. Triponez and N. Villard for their invaluable field and technical assistance, as well as J. R. Litman, Y.-H. E. Tsai, the members of B.C. Carstens' laboratory, two anonymous referees, I. G. Alsos, and J. Hille Ris Lambers, whose comments largely improved an earlier version of this manuscript. They also thank the Vital-IT High Performance Computing Center for granting access to the computer cluster, and the Centre du Réseau Suisse de Floristique and the Alpine Conservatory of Gap-Charance for allowing access to their databases. This work was funded by the Swiss National Science Foundation (SNSF grant No. 3100A0-116778 awarded to NA). AE and NA were funded by a SNSF Ambizione fellowship (PZ00P3_126624). AG, LM, WH and LP were supported by the European Commission (ECOCHANGE, FP6 2006 GOCE 036866) and the SNSF (BIOASSEMBLE, Nr. 31003A-125145).

AUTHORSHIP

AE, LP and NA designed the study, AE and LP collected data, AE performed the genetic analyses, LP, LM and WH performed the modelling work, AE and LP performed the meta-analysis, AE wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

REFERENCES

Abbott, R.J. & Brochmann, C. (2003). History and evolution of the arctic flora: in the footsteps of Eric Hulten. *Mol. Ecol.*, 12, 299–313.

Alsos, I.G., Alm, T., Normand, S. & Brochmann, C. (2009). Past and future range shifts and loss of diversity in dwarf willow (*Salix herbacea* L.) inferred from genetics, fossils and modelling. *Global Ecol. Biogeogr.*, 18, 223–239.

Alsos, I.G., Ehrich, D., Thuiller, W., Eidesen, P.B., Tribsch, A., Schonswetter, P. *et al.* (2012). Genetic consequences of climate change for northern plants. *Proc. Biol. Sci. B*, DOI: 10.1098/rspb.2011.2363.

Araújo, M.B. & Pearson, R.G. (2005). Equilibrium of species' distributions with climate. *Ecography*, 28, 693–695.

Avise, J.C. (2009). Phylogeography: retrospect and prospect. *J. Biogeogr.*, 36, 3–15.

Beatty, G.E. & Provan, J. (2011). Comparative phylogeography of two related plant species with overlapping ranges in Europe, and the potential effect of climate change on their intraspecific genetic diversity. *BMC Evol. Biol.*, 11, 29.

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.

Burnier, J., Buerki, S., Arrigo, N., Kupfer, P. & Alvarez, N. (2009). Genetic structure and evolution of Alpine polyploid complexes: *Ranunculus kuepferi* (Ranunculaceae) as a case study. *Mol. Ecol.*, 18, 3730–3744.

Carstens, B.C. & Richards, C.L. (2007). Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, 61, 1439–1454.

Davis, M.B., Shaw, R.G. & Etterson, J.R. (2005). Evolutionary responses to changing climate. *Ecology*, 86, 1704–1714.

Després, L., Loriot, S. & Gaudeul, M. (2002). Geographic pattern of genetic variation in the European globeflower *Trollius europaeus* L. (Ranunculaceae) inferred from amplified fragment length polymorphism markers. *Mol. Ecol.*, 11, 2337–2347.

Duchesne, P. & Bernatchez, L. (2002). AFLPOP: a computer program for simulated and real population allocation based on AFLP data. *Mol. Ecol. Notes*, 3, 380–383.

Ehlers, J. & Gibbard, P.L. (2004). Quaternary glaciations: extent and chronology. In: *Developments in Quaternary Science* (eds Ehlers, J. & Gibbard, P.L.). Elsevier Amsterdam, San Diego, CD 1.

Ehrich, D., Gaudeul, M., Assefa, A., Koch, M.A., Mummenhoff, K., Nemomissa, S. *et al.* (2007). Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Mol. Ecol.*, 16, 2542–2559.

Engler, R., Randin, C.F., Thuiller, W., Dullinger, S., Zimmermann, N.E., Araujo, M.B. *et al.* (2011). 21st century climate change threatens mountain flora unequally across Europe. *Global Change Biol.*, 17, 2330–2341.

Etterson, J.R. & Shaw, R.G. (2001). Constraint to adaptive evolution in response to global warming. *Science*, 294, 151–154.

Falush, D., Stephens, M. & Pritchard, J. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes*, 7, 895–908.

Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.*, 24, 38–49.

Freedman, A.H., Thomassen, H.A., Buermann, W. & Smith, T.B. (2010). Genomic signals of diversification along ecological gradients in a tropical lizard. *Mol. Ecol.*, 19, 3778–3788.

Graham, C.H., VanDerWal, J., Phillips, S.J., Moritz, C. & Williams, S.E. (2010). Dynamic refugia and species persistence: tracking spatial shifts in habitat through time. *Ecography*, 33, 1062–1069.

Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.*, 8, 993–1009.

Gyllencreutz, R., Mangerud, J., Svendsen, J.-I. & Lohne, Ø. (2007). DATED – a GIS-based reconstruction and dating database of the Eurasian deglaciation. In: *Applied Quaternary Research in the Central Part of Glaciated Terrain* (eds. Johansson, P & Sarala, P). *Geological Survey of Finland*, 46, 113–120.

Hartigan, J.A. & Wong, M.A. (1979). A K-means clustering algorithm. *Appl. Stat.-J. Roy. St. C.*, 28, 100–108.

Hewitt, G.M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.*, 58, 247–276.

Hewitt, G.M. (1999). Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.*, 68, 87–112.

Hugall, A., Moritz, C., Moussalli, A. & Stanisic, J. (2002). Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snails *Gnarosophila bellendenkerensis*. *Proc. Natl. Acad. Sci. U.S.A.*, 99, 6112–6117.

IPCC (2001). Climate Change 2001: Impacts, Adaptation, and Vulnerability: Contribution of Working Group II to the Third Assessment Report of the IPCC. Cambridge University Press, Cambridge, UK.

Körner, C. (2003). *Alpine Plant Life – Functional Plant Ecology of High Mountain Ecosystems*, 2nd edn. Springer, Heidelberg.

Labeyrie, L., Cole, J., Alverson, K. & Stocker, T. (2003). The history of climate dynamics in the Late Quaternary. In: *Paleoclimate, Global Change and the Future* (eds. Alverson, K.D., Bradley, R.S. & Pedersen, T.F.). Springer, Heidelberg, pp. 33–61.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28, 385–393.

Maiorano, L., Cheddadi, R., Zimmerman, N.E., Pellissier, L., Petitpierre, B., Pottier, J. *et al.* (in press). Building the niche through time: using 13,000 years of data to predict the effects of climate change on tree species in Europe. *Glob. Ecol. Biogeogr.*, DOI: 10.1111/j.1466-8238.2012.00767.x.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modeling. *Div. Dist.*, 15, 59–69.

Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M. & New, M. (2004). A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100). *Tyndall Centre Working Paper*, 55, 1–25.

Parducci, L., Jørgensen, T., Tollefsrud, M.M., Elverland, E., Alm, T., Fontana, S.L. *et al.* (2012). Glacial survival of boreal trees in Northern Scandinavia. *Science*, 335, 1083–1086.

Peakall, R. & Smouse, P.E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*, 6, 288–295.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org.

Raymo, M.E. (1997). The timing of major climate terminations. *Paleoceanography*, 12, 577–585.

Schmitt, T. (2009). Biogeographical and evolutionary importance of the European high mountain systems. *Front. Zool.*, 6, 9.

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

Singarayer, J.S. & Valdes, P.J. (2010). High-latitude climate sensitivity to ice-sheet forcing over the last 120 kyr. *Quater. Sci. Rev.*, 29, 43–55.

Skrede, I., Eidesen, P.B., Portela, R.P. & Brochmann, C. (2006). Refugia, differentiation and postglacial migration in arctic-alpine Eurasia, exemplified by the mountain avens (*Dryas octopetala* L.). *Mol. Ecol.*, 15, 1827–1840.

Stewart, J.R., Lister, A.M., Barnes, I. & Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proc. R. Soc. B*, 277, 661–671.

Svenning, J.-C., Normand, S. & Kageyama, M. (2008). Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *J. Ecol.*, 96, 1117–1127.

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, 32, 369–373.

Treier, U.A. & Muller-Scharer, H. (2011). Differential effects of historical migration, glaciations and human impact on the genetic structure and diversity of the mountain pasture weed *Veratrum album* L. *J. Biogeogr.*, 38, 1776–1791.

Vega, R., Flojgaard, C., Lira-Noriega, A., Nakazawa, Y., Svenning, J.-C. & Searle, J.B. (2010). Northern glacial refugia for the pygmy shrew *Sorex minutus* in Europe revealed by phylogeographic analyses and species distribution modelling. *Ecography*, 33, 260–271.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M. *et al.* (1995). AFLP: a new technique for DNA fingerprinting. *Nucl. Acids Res.*, 23, 4407–4414.

Watts, W.A., Allen, J.R.M. & Huntley, B. (1996). Vegetation history and palaeoclimate of the last glacial period at Lago Grande di Monticchio, Southern Italy. *Quater. Sci. Rev.*, 15, 133–153.

Weider, L.J. & Hobaek, A. (2000). Phylogeography and arctic biodiversity: a review. *Ann. Zool. Fennici*, 37, 217–231.

Westergaard, K.B., Alsos, I.G., Popp, M., Engelskjon, T., Flatberg, K.I. & Brochmann, C. (2011). Glacial survival may matter after all: nunatak signatures in the rare European populations of two west-arctic species. *Mol. Ecol.*, 20, 376–393.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).

Editor, Janneke Hille Ris Lambers
Manuscript received 7 December 2011
First decision made 9 January 2012
Second decision made 5 March 2012
Manuscript accepted 15 March 2012