

Phylogeography and colonization pattern of subendemic round-leaved oxeye daisy from the Dinarides to the Carpathians

Kamil Konowalik (✉ kamil.konowalik@upwr.edu.pl)

Wrocław University of Environmental and Life Sciences

Research Article

Keywords: chloroplast DNA, ecological niche modeling, endemics, European mountains, phylogeography

Posted Date: August 1st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1184231/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The Carpathians are an important biodiversity hotspot and a link between mountain ranges on the European continent. This study investigated the phylogeography of one the Carpathian subendemics, *Leucanthemum rotundifolium*, which is distributed throughout the range and in one isolated population outside it. Range-wide sampling was used to examine phylogeographic patterns by sequencing uniparentally inherited chloroplast markers that exemplify seed dispersal. Reconstruct Ancestral State in Phylogenies (RASP) software, Bayesian binary Markov Chain Monte Carlo (BBM) analysis, and ecological niche modeling based on concatenated results of five algorithms were used to infer migration routes and examine links with other species through phylogeny. The round-leaved oxeye daisy is an example of organisms that reached the Carpathians through a southern “Dacian” migration route, most probably through long-distance dispersal. Dating placed the events in the Pleistocene and supported migrations during cooler periods and stasis/isolation followed by separation in the interglacials. Haplotype diversification indicated that after *L. rotundifolium* reached the area around the Fagaras Mountains, several migration events occurred leading to colonization of the Southern Carpathians followed by migration to the Apuseni Mountains, the Eastern Carpathians, and finally the Western Carpathians. The results are consistent with previous phylogeographic studies in this region and indicate several novel patterns.

Introduction

The Carpathians are the main mountain range in Central Europe at the crossroads between the western–eastern and northern–southern European massifs. Their flora and fauna have been studied for decades, and feature in numerous biogeographic studies on plants and animals^{1,2}. They harbor a remarkably diverse vascular flora, comparable in species richness to that of the lower Amazon basin^{3,4} or Caatinga⁵. Being surrounded mostly by lowlands, they also exhibit a pattern typical of mountain regions and encompass much wider climatic variability than neighboring areas⁶. However, unlike typical high mountain ranges, their specificity lies in the structure, which is dominated by mid-elevation heights and polynyas (montane meadows), whereas high-elevation peaks with an alpine zone create a chain of spatially isolated islands^{7,8}. The biogeography of the Carpathians is also a topic of interest in modern times, and several important works have been published recently^{9,10}. Many biogeographical studies have attempted to describe the connections between the Carpathian and other mountain systems, especially the Alps and the Balkan Ranges⁹. Migrations between these mountain chains shape the constituents and reflect the origins of Carpathian biodiversity. Plants can choose two routes during migration from the Alps. They may reach the Western Carpathians via the Eastern Alps (Illyric-Noric route) or reach the Southern Carpathians via the Dinarides and then migrate northwards (Dacian route)^{11–13}. Connections between neighboring mountain ranges and the Carpathians are important, since colonization from the Carpathians is also occurring⁹. The Carpathians are interesting not only as a migration stop, but also as a biodiversity hotspot that harbors 420 endemic plant taxa at species and subspecies level^{14–16}. Of these taxa, 39 are subendemic, and are characterized by some satellite populations outside the

Carpathians. Recently, a new project was implemented to gain better knowledge about Carpathian endemic plants¹⁷.

Within the Carpathians, there are several taxa of the oxeye daisy genus *Leucanthemum* Mill. (Compositae; Anthemideae), which represent the northern group of the Leucanthemineae heavily influenced by hybridization and polyploidization. More than half of the species of this genus are polyploids, and most of the diploid species show signatures of past hybridization^{18–21}. One of the oxeye daisies growing in the Carpathians is a subendemic taxon, the round-leaved oxeye daisy (*Leucanthemum rotundifolium* (Willd.) DC. (= *L. waldsteinii* (Sch.Bip.) Pouzar), one of the most divergent species in the genus and characterized by unique morphology (leaf shape) and habitat (preference for moist substrate). In various phylogenetic reconstructions, it is often placed as one of the earliest diverging lineages^{18–20}. It owes its subendemic status to its disjunct distribution, with an isolated population found in the Vranica Mountains (Dinarides) ca. 370 km in a straight line from the nearest population in the Southern Carpathians. *L. rotundifolium* is associated with humid soils and is often found close to springs and stream banks. It may be regarded as a montane species because it has never been found below 500 m, which is the approximate lower border of the montane belt in the Carpathians^{22,23}. Although it also occurs at higher elevations in the subalpine zone, it never forms numerous populations as it does in the montane zone. At higher elevations, it is similarly confined to stream banks, local depressions, or suitable microclimates created by shrub stands (typically consisting of *Pinus mugo* as a main component). Due to its disjunct distribution and subendemic status, it is an ideal candidate to study the phylogeography of the Carpathians.

Plants are valuable subjects for phylogeographic study because of their neutral dispersal patterns that reflect connections between areas in the past. This is also true of *L. rotundifolium* whose seeds lack any features that could directly facilitate dispersal²³. After reaching maturity, its small achenes simply fall out of the capitulum, and most lie in proximity to the maternal plant at this stage. They may be fortuitously distributed further by wind, flowing water, or animals^{23,24}. Seed dispersal reflects natural processes, such as those that create suitable habitat corridors that enable species to spread or form barriers that inhibit dispersal. At the species level, the emergence of such corridors or boundaries, frequently corresponding to biogeographical patterns, is often influenced by climate change.

The study of angiosperm seed dispersal is facilitated by the fact that chloroplasts are maternally inherited, and the analysis of chloroplast markers provides a direct method. The chloroplast genome also has other features that make it ideal for such studies, such as lack of hybridization and relatively rapid fixation in the population²⁵. In contrast to nuclear markers, all chloroplast markers are single copy genes, which, therefore, provide an opportunity for analysis with less noise from hybridization or duplicate events. They have also been used to study the phylogeography of other groups of *Leucanthemum*^{26–28}.

This study has two aims: first, investigation of the phylogeography of *L. rotundifolium*, a pan-Carpathian subendemic, which, in addition to its own merits, has the potential to provide general data on biogeographical patterns in the Carpathian Mountains; and second, inferring possible migration routes

and tests for the existence of geographically structured genetic variation. More importantly this work aims to determine how and when the disjunction between the Dinarides and the Carpathians occurred and how migration within the Carpathians took place. The study will put these events into a time frame provided by the dating of phylogenetic trees and analyze changes in potential distribution through ecological niche modeling. This addresses issues raised by Ronikier (2011), including broad sampling of Carpathian populations, inclusion of dating to calibrate the time frame of the observed pattern, and exploration of the relationship between the Carpathians and the Dinarides.

Materials And Methods

Study species

Leucanthemum rotundifolium is relatively common in most of the Carpathian ranges and has one disjunct population in the Vranica Mountains of the Dinarides. In the main part of its range, *L. rotundifolium* forms variously sized populations, from large ones thriving in stream valleys to single specimens at high elevations and in sunny habitats. It occurs between 525 and 2060 m a.s.l., but has mostly been found in the upper montane belt (mean 1250 ($\pm$ 420) m a.s.l.). The fruits of *L. rotundifolium* are small achenes (each weighing approximately 0.4 mg). A single capitulum may contain numerous achenes (approximately 100–300), and a single plant may have one to a few flower heads. These characteristics do not impose any direct mode of dispersal; rather, they may lead to a high number of seeds and dispersal by random events that may include transportation downslope via flowing water and possibly also by wind or animals. In the field, this species is relatively easy to recognize and spot, especially during flowering. It is also morphologically distinct from sympatric congeneric species due to the unique shape of its basal leaves, stem leaves, and capitula.

Sampling strategies

Species identification was carried out in the field. The sampling design followed the historical locations of *L. rotundifolium* recorded in Zelený (1970) and on herbarium sheets. In the field, each *L. rotundifolium* stand that was spatially isolated from other stands by at least 150 m or visible barriers was collected as a separate population. Following this scheme, samples from 68 populations were collected (Supplementary material 1, Fig. 1). Sampling covered the whole distributional area of the species (Fig. 1). In addition, all known sympatric species, *L. vulgare* Lam. ($2n = 2x = 18$), *L. ircutianum* DC. ($2n = 4x = 36$), *L. gaudinii* Dalla Torre ($2n = 2x = 18$), *L. margaritae* (Jáv.) Zelený ($2n = 6x = 54$), and *L. illyricum* (Horvatić) Vogt & Greuter ($2n = 8x = 72$), were collected and later used as outgroups. Field studies on plants, including the collection of plant material, complied with relevant institutional, national, and international guidelines and legislation. The plant material was collected with the permission of the local authorities and according to national laws.

DNA isolation

DNA was extracted from five individuals in each population, except for populations that were closer than 4 km from which only one to three individuals were sampled (depending on sampling density). DNA was extracted from silica-gel-dried leaves that were pulverized in an adjusted plastic rack attached to a reciprocating saw or jigsaw²⁹. I used the “Sherlock AX” DNA extraction kit (A&A Biotechnology, Gdynia, Poland) according to the manufacturer’s protocol.

PCR amplification

In the following analyses, I used two datasets: the first included samples of the representatives of the genus *Leucanthemum*, members of the subtribe Leucantheminae, and selected accessions of *L. rotundifolium*¹⁹; and the second consisted of all *L. rotundifolium* accessions and sympatric species. The first dataset used chloroplast markers and accessions that were used to reconstruct the phylogeny of *Leucanthemum*¹⁹. These included five intergenic spacer regions (*psbA-trnH*, *trnL-trnF*, *trnC-petN*, *petN-psbM*, and *trnQ-rps16*)³⁰. Initial screening revealed very low variability of those markers within the studied species. Therefore, new primer pairs were designed for the most variable regions within the genus *Leucanthemum*³¹. First, I downloaded available whole plastome sequences for the Asteraceae from GenBank (www.ncbi.nlm.nih.gov/genbank) and searched for conserved regions spanning exons pinpointed by Scheunert et al. (2020). Then, I placed primers in places that would yield a PCR product around 1000 bp. Several such primers were constructed and after screening, two of these primer pairs, *ndhCretF* (AAGTTTCTCCGGTCCTTTGC) – *trnV-comp-R* (CTCTTTTCCTGTCCGAAATC) and *trnT(GUU)-comp-F* (AAGTGGACCTGACCCATTG) – *psbD-comp-R* (GACCATTTCCGAACACCTC), were selected based on their level of polymorphism and feasibility of amplification. The first pair was already proposed by Timme et al. (2007) as *ndhCretF-trnVretR* as a marker for the Asteraceae, but the reverse primer was shifted by a few nucleotides to exclude polymorphic sites. Together with *psbA-trnH*, these three primer pairs were amplified for all *L. rotundifolium* individuals and sympatric outgroup species and constitute a second dataset. All PCR reactions were performed in the T100 Thermal Cycler (Bio-Rad, Hercules, USA) with 15 µl volume using Taq 2x Master Mix RED with 1.5 mM MgCl₂ (Ampliqon, Odense M, Denmark) and 10 µmol of each primer. The PCR program consisted of the following steps: 5 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 40 s at 55 °C, and 30 s at 72 °C, followed by 10 min at 72 °C. A negative control was always included, and if any sign of contamination or cross-contamination was visible, all samples were discarded. The PCR products were cleaned with CleanNGS (CleanNA, Waddinxveen, the Netherlands) and sent to MacroGen Europe (Amsterdam, the Netherlands) for sequencing. Sequence chromatograms were visually inspected in Chromas (Technelysium Pty Ltd, South Brisbane, Australia) and then aligned in Bioedit³³. In cases of poor read quality, sequencing was repeated using both forward and reverse primers. The few segments that could not be aligned unambiguously (mostly sites containing poly-T repeats of variable length) were deleted from the final alignment as a polymerase-induced error. Sequence data were submitted to the GenBank database under accession numbers [Numbers after acceptance], and alignments are available as Supplementary Material 2.

Phylogenetic analyses and dating

I constructed phylogenetic trees using MrBayes 3.2.7a³⁴ with all markers and indels as a separate partitions. Indels were coded using the simple gap coding³⁵ implemented in SeqState 1.4.1³⁶. For all partitions, the best fitting model was selected using jModelTest^{37,38}, and the Jukes–Cantor model³⁹ was applied for gaps. Two separate runs were performed, each consisting of four chains, 15×10^6 generations, and sampling every 1000th tree. The first 25% of the trees were discarded as a burn-in period. The convergence of chains and ESS values were checked in the Tracer software⁴⁰.

The first data set that employed the alignment of five intergenic spacer regions of the chloroplast genome (*psbA-trnH*, *trnL-trnF*, *trnC-petN*, *petN-psbM*, and *trnQ-rps16*) and contained all diploid *Leucanthemum* species, outgroups from the subtribe Leucantheinae, and selected *L. rotundifolium* accessions in order to produce a dated phylogeny comparable to that from a previous study¹⁹. The second data set contained all *L. rotundifolium* individuals and sympatric species as outgroups to reconstruct phylogeographic patterns for the focal species. The dating was performed using MEGA software employing the RelTime method^{41–44}. RelTime has been shown to perform well in cases where the taxa studied come from data sets containing a mix of samples at specific and subspecific levels⁴⁵. I used four dating points: two geological events that separated an island endemic from a mainland relative, i.e., *Plagius flosculosus* from *P. maghrebinus* and *Maurantheum paludosum* from *M. ebusitanum*¹⁹, the crown age of *Leucanthemum*²⁰, and the split between *Artemisia* and other Anthemideae estimated from fossilized pollen⁴⁶. The details are provided in Table 1. To date the tree obtained from the second dataset, I used nodes from the first analysis that included taxa present in both analyses (Table 1).

Haplotype network and clustering

To produce the haplotype network, I used a TCS algorithm⁴⁷ implemented in the software PopART⁴⁸. To assess the number of genetic clusters, I used Bayesian Analysis of Population Structure (BAPS)⁴⁹, a hierarchical genetic clustering algorithm, available in the RhierBAPS package for R⁵⁰. Three separate runs testing 2–34 populations and two levels were conducted.

Bayesian binary Markov Chain Monte Carlo analysis

To estimate ancestral areas, I used the software Reconstruct Ancestral State in Phylogenies (RASP)⁵¹ and Bayesian binary Markov Chain Monte Carlo analysis (BBM)⁵² with the consensus trees. . The advantage of this method for this data set was that it could deal with polytomies, while this feature of the phylogenetic trees prevented the use of other methods. The BBM method was independently run twice using 1×10^6 generations, sampling every 1000th generation, and discarding the first 25% as a burn-in. The maximum number of areas per node was set to four, the state frequencies were fixed following the Jukes–Cantor model, and the variation between sites was set to equal. The results of both runs were combined within the program.

Species distribution modeling

Modeling followed the gold standards in ecological niche modeling⁵³ and used an ensemble of several algorithms rather than a single model. To select the best-performing algorithms, I used the method proposed earlier for the same species⁵⁴, and compared the same set of 13 algorithms as the original article and chose those with the highest AUC and MAE scores. AUC and MAE were calculated using the package Metrics in R⁵⁵. The five selected algorithms were ANN^{56,57}, CTA^{57,58}, MAXENT⁵⁹, BRT^{60,61}, and BART⁶². Each algorithm was repeated 100 times; their settings were as described in Konowalik and Nosol (2021).

Points for modeling

As a source of points for modeling, I used a combined data set, which joined data from Jasiewicz (1965), Kornaś (1955), and Zelený (1970) and georeferenced herbarium specimens and field trips. Irrespective of the source, I used only locations that could be unambiguously assigned to a certain place with an accuracy higher than the rasters used (0.5 km²). Occurrences within a single raster cell were reduced to one point. To further decrease spatial bias, I applied environmental filtering, which performs better than geographical filtering^{65–67}. I used the algorithm provided by Varela et al. (2014) at <https://github.com/SaraVarela/envSample>. As filters, I used PCA maps computed from previously selected environmental variables and filtered presences and background using the same settings. The background was generated as in Konowalik and Nosol (2021) and consisted of all data points above 400 m within the Carpathians. All data points falling within one unit of principal component axis 1 and principal component axis 2 were trimmed to only one randomly selected data point. After all filtering procedures, 429 presence data points were left for the final analyses (Fig. 1).

Variable selection

The analysis was restricted to the part of Central Europe that includes the Carpathians and neighboring regions, which define the most probable accessible area of the studied species⁶⁸. I downloaded 19 bioclimatic variables for this region from CHELSA 1.2 with resolution of 30 arc seconds^{69,70}. Pearson's correlation coefficient was used to assess associations between variables, which were removed stepwise until all variables with $p < 0.7$ were removed. Variables showing linear dependencies and zero or near-zero variance were removed with the Caret package in R⁷¹. I wanted to further reduce the number of environmental variables to those that are influential for the studied species. A previous study indicated that the use of modeling algorithms to select variables may need further development⁷²; therefore, I used a customized procedure. To reach consensus between different algorithms, an elimination strategy was used in which a variable that was less important than a random variable in at least three out of five algorithms was deleted. For this purpose, the spatialEco package in R⁷³ was used to create random variables with the same range of values as those of the true variables. I ran models with true and random variables jointly and after each run, discarded a variable that performed worse than a random variable. This procedure was repeated for each algorithm until all true variables showed a higher contribution than

any of the random variables. Finally, I compared the variables discarded by different algorithms and removed from the final data set those variables discarded by the majority of the algorithms (Table 2).

Predicting past distribution

To predict past distributions, I used variables available on PaleoClim.org⁷⁴. These variables are computed using CHELSA⁷⁰ and contain the same set of 19 bioclimatic variables available for the periods specified in Table 2. Models for potential past distributions were generated using the same methods and algorithms as for the present time.

Model ensemble

Each algorithm was run individually using the same set of presence and background points. In order to find consensus between modeling algorithms that minimized discrepancies and reduced extreme predictions, I produced an ensemble model. It was calculated using principal component analysis (PCA), which is recommended for producing consensus forecasts^{75,76}. As input, all the models produced for a given period were submitted to a standard PCA calculation. The first PCA component (PC1) shows the central tendency over all individual models and represents a consensus between them. The resulting files were rescaled to the 0–100 range, where 0 showed low suitability and 100 high suitability for the studied plant species. All procedures were carried out with the ‘raster’ package and using basic functions in R⁷⁷. All GIS operations were done in R ver. 3.6.2⁷⁸ and QGIS ver. 3.8⁷⁹.

Results

Phylogenetic analyses and dating

In the phylogenetic tree constructed using the five intergenic spacers and their indels *L. rotundifolium* is part of the lineage composed of diploid *L. virgatum* and polyploids (Posterior probability PP = 0.99, Fig. 2). The stem age of this group is 0.65 Ma (Confidence interval CI 0.99–0.42). Within the *L. rotundifolium* clade, a major divergence dating from 0.38 Ma (CI 0.70–0.21) separates the Dinarides population from the rest of the species (PP = 0.99). This analysis was performed mainly to place the *L. rotundifolium* accessions within the framework previously proposed for the genus. However, due to low resolution, a full sampling was not carried out; more detailed relationships between the focal species are observable in the second data set. The second phylogenetic tree was constructed using three markers and displays a structure similar to that of the first tree (Fig. 3). A notable difference is associated with the MRCA of *L. rotundifolium*, at whose node the Dinarides population, polyploids, and the rest of the *L. rotundifolium* accessions are placed in a polytomy (PP = 1). This node dates from 0.66 Ma (CI 0.87–0.48). This may be the effect of not considering other diploid taxa included in the first analysis and could provide more information at this node. The base of the *L. rotundifolium* clade is formed by the plants from the Southern Carpathians; this node dates from 0.23 Ma (CI 0.40–0.13). The clade containing B3 and B2 haplotypes is significantly supported (PP > 0.96), and its origin is estimated to be around 0.09 Ma (CI

0.4032–0.186). The clade containing the B6 haplotype (PP = 0.96) is estimated to originate from 0.0661 Ma (CI 0.3117–0.014). A clade containing the cluster of D haplotypes (PP = 0.95) is estimated to have emerged 0.084 Ma (CI 0.399–0.017) and the youngest clade within this group containing the D5 haplotype (PP = 1) dates from 0.0036 Ma (CI 0.096–0.0001). The clade that includes the C1 haplotype (PP = 1) originated around 0.0475 Ma (CI 0.23–0.01). The clade containing the E1 haplotype (PP = 0.99) originated 0.0665 Ma (CI 0.3213–0.0138). The clade that stems from this group and contains the cluster of F haplotypes is estimated to originate around 0.0216 Ma (CI 0.1554–0.003). The clade containing haplotype G1 (PP = 0.84) is estimated to originate 0.0211 Ma (CI 0.1934–0.0029). There may be two supported clades within this group: one specific to the Gorgany Mountains originating 0.0062 Ma (CI 0.1483–0.0003) and the other specific to the Bieszczady Mountains originating 0.0061 Ma (CI 0.1474–0.0003) – Table 3.

Haplotype network and clustering

The clustering applied to the phylogenetic tree constructed using three markers found that the optimal number of partitions is seven. This analysis divides *L. rotundifolium* into groups that correspond to geography and the haplotype network (Fig. 4). Similar to the phylogenetic tree, the haplotype network places a polytomy at the MRCA of *L. rotundifolium*, and the Dinarides population (haplotype: A1) is detached from the species (Fig. 4). The basal node of *L. rotundifolium* (PP = 0.89) consists of haplotype B1, which is placed in a polytomy with several well-supported groups. Haplotype B1 is distributed mainly in the Southern Carpathians. Haplotype B1 is related to several child haplotypes, B2–B6. Most of these haplotypes occur in the Southern Carpathians from the Banat Mountains to the Harghita Mountains. One notable exception is haplotype B6, which is found not only in the Retezat Mountains and the Parang Mountains, but also in the central part of the Eastern Carpathians from the Svydovets to the Rodna Mountains. Two distinct groups emerge from the B1 haplotype: the C1–C2 haplotypes and the D1–D5 haplotypes. Cluster C is found only in the Southwestern Carpathians from the Banat Mountains to the Parang Mountains, and its eastern border is marked by the Olt valley. The D cluster haplotypes occur in the northern part of the Apuseni Mountains and in the Rakhiv Mountains. The basal haplotypes, D1 and D2, which form a clade (PP = 0.95), occur in the Apuseni Mountains; the sister to this group is the clade composed of the D3 and D4 haplotypes in the Rakhiv Mountains and D5, again found only in the Apuseni Mountains. Since all these haplotypes form clades supported by a PP of > 0.95, this peculiar pattern may be explained as two waves of migration: the first colonizing the Rakhiv Mountains from the Apuseni Mountains, and the second colonizing the Apuseni Mountains with intermediate events that led to the separation and divergence of these haplotypes. Another group (PP = 0.99) stemming from the B1 haplotype is E1. This cluster is found in the central part of the Eastern Carpathians. From this haplotype emerged two new clusters, F and G, which spread further north and reached the Western Carpathians.

Despite the large area occupied by the plants belonging to the F and G clusters, they are somewhat less divergent and contain mostly satellite haplotypes that occur without a clear pattern and include only one to a few individuals. However, there are a few exceptions, such as haplotype G5, which groups all individuals in the Gorgany Mountains and does not occur anywhere else. Furthermore, haplotype G6 is

found only in one place in the Bieszczady Mountains. The distribution of both F and G clusters is similar except that cluster G is more widespread. Both clusters occur in a mosaic in the Western Carpathians, especially in the Tatra Mountains. Most of the plants from the Gorce Mountains belong to the F1 haplotype, and all plants from the Żywiec Beskids Mountains belong to the G1 haplotype. During the analyses, a few specimens resembling *L. rotundifolium* were classified into the haplotypes assigned to the outgroup species. It is possible that they are hybrids between *L. rotundifolium* and other species, but this needs further examination.

Bayesian binary Markov Chain Monte Carlo analysis

In the event matrix, RASP suggests the following scenario involving dispersal and then vicariance as the most probable for the MRCA node of *L. rotundifolium*: Dinarides → Dinarides + Southern Carpathians → Dinarides | Southern Carpathians. The Dinarides are also the most probable ancestral area for this node with a probability of 0.53 (while $P = 0.18$ for the Southern Carpathians, $P = 0.13$ for the Eastern Carpathians, and $P = 0.1$ for the Apuseni Mountains). The ancestral area of the node containing, among others, the B1 haplotype, is located in the Southern Carpathians ($P = 0.38$), the Eastern Carpathians ($P = 0.29$), the Apuseni Mountains ($P = 0.12$), or the Southern and Eastern Carpathians ($P = 0.11$). This node is characterized by multiple dispersal events that took place from the Southern Carpathians, which led to the colonization of other parts of the range plus the Apuseni Mountains and the Eastern Carpathians. The D cluster, as described previously, displays a peculiar pattern of origin in the northern part of the Apuseni Mountains ($P = 0.73$), dispersal to the Rakhiv Mountains where two new haplotypes originated (with $P = 0.75$ and $P = 0.99$), and then redispersal to the Apuseni Mountains where a new haplotype emerged ($P = 0.95$). The origin of the C cluster is in the Southern Carpathians ($P = 0.99$). The ancestral area of the node containing haplotype E1 is located in the Eastern Carpathians ($P = 0.99$). Similarly, the ancestral area of nodes containing the G1 and F1 haplotypes is located in the Eastern Carpathians ($P = 0.97$ and $P = 0.96$, respectively). The ancestral areas of nodes containing haplotypes G5 and G6 are also located in the Eastern Carpathians, and as described previously, they are found only locally within a small area.

Species distribution modeling and variable selection

The most important variables selected by majority consensus were mean temperature of the driest quarter (BIO09), mean temperature of the coldest quarter (BIO11), seasonality of precipitation (BIO15), and precipitation of the warmest quarter (BIO18). Generally, temperature-related variables had a highly important association with the area of the suitable niche as exemplified by Figure 5, which shows a close relationship between predicted area and changes in temperature. These variables are also biologically relevant for a mountainous species associated with humid soils. First, temperature limits the vertical distribution of *L. rotundifolium* by excluding competition from lowland species not adapted to low temperatures and confining its range to a suitable area limited by harsher conditions present at higher elevations. Second, the amount of precipitation plays an important role during the warmest period when it is crucial for survival; relatively equal distribution of precipitation throughout the year without severe droughts is also a factor enabling long-term survival of this species.

Model ensemble

In general, the models reconstructed similar patterns, although there were discrepancies in the extent of the potential suitable habitat and the assigned probability of occurrence. In this case, making an ensemble helped to obtain a less equivocal and more balanced results. The derived ensemble maps correlate well with the distribution recorded in the field. The ensemble shows that there are discontinuities within the range. These breaks are chiefly associated with valleys or lowlands. Within the mountains, the pattern is slightly more complicated, but high mountain peaks are not predicted as suitable (especially in the north), and suitability is higher for the eastern and northern slopes. Moderate suitability is shown for the Vranica Mountains in the Dinarides. There are no suitable locations between the Dinarides and the Southern Carpathians, except for a few areas with low suitability scores. Models also predict a high probability of occurrence in areas where species were not found. Such areas include the eastern part of the Apuseni Mountains, the Slovak Ore Mountains, the Sanok-Turka Mountains, the Gurghiu Mountains, and the Buzău Mountains.

Predicting past distribution

The models built for the past show a consistent distribution pattern congruent with haplotype diversification. The earliest available datasets for paleoclimate come from the last interglacial 787 ka, which is later than the main diversification of *L. rotundifolium*. Yet, several haplotypes emerged within this period, for example, some that belong to cluster D. Although today there is no suitable habitat between the Rakhiv Mountains and the Apuseni Mountains, corridors or even larger suitable areas are shown to have been available during past climatic oscillations. The same is true for a connection between the Western and Eastern Carpathians. The main suitable areas are always located within the mountains, but depending on the epoch, additional areas appear or disappear. For example, the Transylvanian Plateau is predicted as a relatively suitable area in the period from 21 ka to 10 ka. Large areas in the east are reconstructed as suitable from 21 to 13.8 ka. During heavier cooling, higher elevations within the Western Carpathians became unsuitable. The suitability of the Eastern Carpathians also varied through time, and most of the time, it displayed a mosaic of areas with low and high suitability. As indicated by the evidence, the extent of the potentially suitable habitat is tightly linked to temperature. During colder times, the species may expand its range, as its niche covers a larger area, whereas during interglacial and warmer periods, it is confined to higher elevations (Fig. 5k). However, the pattern is not straightforward as the maximum extent of the range in the analyzed time period is during the Younger Dryas Stadial (the suitable area is 16× larger compared to the present area), which is not the coldest period. Much colder conditions appeared during the Last Glacial Maximum, and the range in that period is smaller (this may also be seen as retreat from higher elevations). This indicates that climate change does not act linearly on *L. rotundifolium*, and that the best climate is one similar to the present climate of the montane zone of the Carpathians. Other factors, such as those related to precipitation (e.g., BIO15 and BIO18 in this study) may also play a role but are less influential than temperature-related variables.

Discussion

Migration into and colonization of the Carpathians

This paper presents a hypothesis of dispersal and migration of the Carpathian subendemic *L. rotundifolium*. The hypothesis is based on evidence from chloroplast genetic markers, which, therefore, tightly links it to seed dispersal. According to the results obtained, the study was able to reconstruct migration routes; the most plausible hypothesis of migration is illustrated in Figure 6. The ancestral range of *L. rotundifolium* can be confidently located in the Dinarides. Although there is only one confirmed population in this area, its range could be larger in the past (Fig. 5 b-h) and subsequently reduced most probably by climatic changes. From this area, it spread to the Carpathians, likely as a result of long-distance dispersal. The area that was first colonized within the Carpathians is associated with the occurrence of haplotype B1 and includes the Fagaras Mountains and adjacent areas (Fig. 6, event 1). Dispersal to the Carpathians led to vicariance of the *L. rotundifolium* populations; since then, the Dinarides population has remained isolated. Similar findings have been reported for *Heliosperma* (Rchb.) Rchb. in which a major divergence occurred between Dinarides and Carpathian lineages and results similarly suggested colonization of the Carpathians from the Dinarides

⁸⁰. After initial establishment, *L. rotundifolium* spread within the area marked “B1 area” (Fig. 6) in the eastern part of the Southern Carpathians. Haplotype B1 gave rise to several other haplotypes that migrated to other parts of the Carpathians; this includes colonization of the Apuseni Mountains (Fig. 6, event 2), the Retezat Mountains (Fig. 6, event 3), the Parang Mountains (Fig. 6, event 4), and the Eastern Carpathians (Fig. 6, event 5). The Eastern Carpathians were colonized at least three times by plants belonging to different lineages (Fig. 6, events 3, 5 and 6). One of these dispersals is quite interesting as plants belonging to haplotype B6 are found in the Southwestern Carpathians, mainly in the Retezat Mountains, and in the Eastern Carpathians (Fig. 4). A similar pattern of distribution between the Southern and Eastern Carpathians is characteristic of some endemic species and subendemic or mountainous species ^{10,17}. Plants in cluster C colonized the southwestern part of the Carpathians, eventually reaching the Banat Mountains (Semenic Mountain), which is the westernmost location of *L. rotundifolium* in the Southern Carpathians. Very similar patterns (occurrence of a separate haplotype in the westernmost part of the Southern Carpathians and links between the Southern and Eastern Carpathians) have also been observed in *Campanula alpina* Jacq. ⁸¹ and *Arabis alpina* L. ⁸².

Another interesting pattern is observed in the Apuseni Mountains and the Rakhiv Mountains; it appears that there was connection between these two chains since they are inhabited by haplotypes which are not found in any other part. This migration occurred repeatedly and was influenced by isolation leading to the emergence of several haplotypes specific to cluster D. Corridors or even larger suitable areas were available during the past climatic oscillations (Fig. 5). Mráz et al. (2007) observed a similar pattern that linked the Apuseni Mountains with the Eastern Carpathians in the case of *Hypochaeris uniflora* Vill. Another haplotype emerging from the basal group is E1 (Fig. 6, event 5), which is distributed in the Eastern Carpathians. This connection between the Southern and Central Eastern Carpathians is a pattern

exhibited by other species, e.g., *Campanula alpina*⁸¹; several endemic taxa may be found within this part of the Eastern Carpathians^{8,10}. Two other haplotypes emerged from haplotype E1: clusters F and G. The ancestral area of both these clusters lies in the Eastern Carpathians. Since the admixture of the F and G haplotypes is quite high, the plausible scenario may involve the emergence of the two haplotypes in isolated centers and then subsequent colonization from these two points. The two clusters also reached the Western Carpathians. Plants in cluster G are more widespread and have colonized the southern parts of the Carpathians to a greater degree. Interestingly, parts of the Western Carpathians like the Żywiec Beskids Mountains and the Gorce Mountains could have been colonized by a single dispersal event since their haplotype constitution is quite homogenous.

Consequently, the emergence of new haplotypes may not reflect an ecological pattern or speciation, but migration through the Carpathian arc accompanied by isolation events. Overall linear migration, which may be compared to a stepping-stone model, is supported for movement across the whole mountain range, with the exception of the emergence of several new haplotypes that follow the island model. Linear migration from south to north and east to west is also supported by the lack of occupation of or extreme scarcity of records for other suitable mountain ranges lying along the margins of the Western Carpathians, such as the Malá Fatra Mountains and Silesian Beskids Mountains. This further supports the hypothesis of migration and explains the absence of *L. rotundifolium* in these mountain ranges as time-dependent: in other words, enough time has not elapsed for this species to reach these ranges because it conquered the Western Carpathians only recently.

Implications for the phylogeography of the Carpathians

Leucanthemum rotundifolium exemplifies the colonization of the Carpathians along the arc from the south to the east, and then to the north and the west. This route has also been demonstrated for several other species¹⁰. Similar to our results, Šrámková et al. (2019) found a link between the Dinarides and the Southern Carpathians. In addition, *Heliosperma pusillum* (Waldst. & Kit.) Rchb. populations from the Dinarides and the Southern Carpathians seem to be related⁸⁰. Several species show this distributional pattern and this route should be emphasized as one of the major routes via which the Carpathian Mountains were colonized. However, the agents that enabled this migration or facilitated long-range dispersal are unknown. The timing of the arrival from the Dinarides to the Carpathians is not certain; in this study, it was reconstructed as occurring 0.38 Ma (CI 0.70–0.21) or 0.66 Ma (CI 0.87–0.48). For this node, the first dating is arguably more accurate since it included more species and was not affected by the polytomy. This scenario receives further support from a climate perspective; hypothetical ancestors of *L. rotundifolium* and similar cold-dwelling species expanded their ranges during the Elster glaciation (Marine Isotope Stage 12, 0.478–0.424 Ma). This cold period was followed by the Holstein interglacial (Marine Isotope Stage 11, 0.424–0.374 Ma), which was also one of the longest interglacial periods that could break previously established connections and separate colonizers from source populations.

According to Hurdu et al. (2012) and Ronikier (2011), the distribution of haplotypes within the Carpathians follows a pattern characterized by a high proportion of endemic species. This further

confirms the results of this study and previously inferred floristic barriers and refugial areas⁷ it has limited potential to support or delineate biogeographic regions within the Carpathians. This is especially the case for the border between the Western and Eastern Carpathians, which is often considered in phylogeographic studies^{9,10}. This study found that this border is porous, and during cooler periods, migration between the two mountain chains was possible. The likely ecological characteristics of the species play a significant role: alpine species confined to higher elevations, which represent the most frequently studied group, do not have many chances to migrate across a border composed of hills and low mountains that may act as a bridge only during the coldest phases. In contrast, mid-elevation montane species such as *L. rotundifolium* have more opportunities to establish at lower elevations sooner and require only moderate cooling to start migration. The existence of corridors or areas suitable for montane species at lower elevations is also supported by fossil data⁸⁴. Another established border located between the Eastern and the Southern Carpathians is stronger in a phylogeographical sense (i.e., without populations with mixed haplotypes), but difficult to define geographically. For *L. rotundifolium*, the border is located somewhere between the Călimani Mountains and the Harghita Mountains, but it is difficult to pinpoint any feature related to a valley or river. The exact placement of this border was also difficult for the Orthoptera because a strict boundary between the Eastern and Southern Carpathians was not clear⁸⁵.

Conclusions And Future Directions

This study presents a phylogeographical investigation based on chloroplast haplotypes of a Carpathian subendemic that may be regarded as an exemplar of seed dispersal. The results of this study broaden the understanding of *L. rotundifolium* distribution and shed light on possible migration routes within the Carpathian Mountains.

It is worth mentioning that there are several other subendemic species in the Carpathians¹⁶ that may also provide interesting results and generate further useful hypotheses about migration.

Declarations

Acknowledgements

I am grateful to my family for their help during plant collection and the curators of KRA and KRAM herbaria for access to the specimens. I would like to thank Ivana Rešetnik from the University of Zagreb for photographs of the Ivo Horvat specimens. I would also like to thank Zofia Kowalczywska-Gleńsk for technical help during the final PCR amplifications. The financial support of the National Science Centre (Poland) during the realization of the project “Phylogeography and hybridization within Carpathian oxeye daisies (*Leucanthemum*, Compositae)” no. 2016/23/D/NZ8/00935 is gratefully acknowledged.

Conflict of interest

References

1. Pax, F. *Grundzüge der Pflanzenverbreitung in den Karpathen*. 1–342 (W. Engelmann, 1898). doi:10.5962/bhl.title.20419.
2. ПОРОВ [ПОПОВ], М. Г. [М. Г.]. *ОЧЕРК РАСТИТЕЛЬНОСТИ И ФЛОРЫ КАРПАТ* [ОЧЕРК РАСТИТЕЛЬНОСТИ И ФЛОРЫ КАРПАТ]. vol. 5 (XIII) (IZDATEL'STVO MOSKVSKOGO OBSHCHESTVA NSPYTATELEY PRNRODY [ИЗДАТЕЛЬСТВО МОСКВСКОГО ОБЩЕСТВА НСПЫТАТЕЛЕЙ ПРИРОДЫ], 1949).
3. Barthlott, W., Mutke, J., Rafiqpoor, M. D., Kier, G. & Kreft, H. Global centers of vascular plant diversity. *Nova Acta Leopoldina* **92**, 61–83 (2005).
4. Mutke, J. & Barthlott, W. Patterns of vascular plant diversity at continental to global scale. *Biol. Skr.* **55**, 521–537 (2005).
5. Zappi, D. *et al.* Growing knowledge: an overview of seed plant diversity in Brazil. *Rodriguesia* **66**, 1085–1113 (2015).
6. Rahbek, C. *et al.* Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science* **365**, 1108–1113 (2019).
7. Hurdu, B. *et al.* Patterns of plant endemism in the Romanian Carpathians (South-Eastern Carpathians). *Contrib. Bot.* **47**, 25–38 (2012).
8. Pawłowski, B. Remarques sur l'endémisme dans la flore des Alpes et des Carpates. *Plant Ecol.* **21**, 181–243 (1970).
9. Mráz, P. & Ronikier, M. Biogeography of the Carpathians: evolutionary and spatial facets of biodiversity. *Biol. J. Linn. Soc.* **119**, 528–559 (2016).
10. Ronikier, M. Biogeography of high-mountain plants in the Carpathians: An emerging phylogeographical perspective. *Taxon* 373–389 (2011).
11. Hendrych, R. *Primula vulgaris* in der Slowakei und in den umliegenden Gebieten. *Preslia Praha* **68**, 135–156 (1996).
12. Hendrych, R. & Hendrychová, H. Preliminary report on the Dacian microelement in the flora of Slovakia. *Preslia Praha* **51**, 313–332 (1979).
13. Sramkó, G. „Dunántúli” közép-dunai flóráválasztós fajok a Matricum flórájában. *KITAIBELIA* **9**, 31–56 (2004).
14. Bálint, M. *et al.* The Carpathians as a Major Diversity Hotspot in Europe. in *Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas* (eds. Zachos, F. E. & Habel, J. C.) 189–205 (Springer, 2011). doi:10.1007/978-3-642-20992-5_11.
15. Juříčková, L. *et al.* Early postglacial recolonisation, refugial dynamics and the origin of a major biodiversity hotspot. A case study from the Malá Fatra mountains, Western Carpathians, Slovakia. *The Holocene* **28**, 583–594 (2018).

16. Kliment, J., Turis, P. & Janišová, M. Taxa of vascular plants endemic to the Carpathian Mts. *Preslia - Praha*- **88**, 19–76 (2016).
17. Breman, E. *et al.* Conserving the endemic flora of the Carpathian Region: an international project to increase and share knowledge of the distribution, evolution and taxonomy of Carpathian endemics and to conserve endangered species. *Plant Syst. Evol.* **306**, 59 (2020).
18. Konowalik, K. Reconstructing reticulate relationships in the polyploid complex of *Leucanthemum* Mill. (Compositae, Anthemideae). (Fakultät für Biologie und Vorklinische Medizin, Universität Regensburg, 2014).
19. Konowalik, K., Wagner, F., Tomasello, S., Vogt, R. & Oberprieler, C. Detecting reticulate relationships among diploid *Leucanthemum* Mill. (Compositae, Anthemideae) taxa using multilocus species tree reconstruction methods and AFLP fingerprinting. *Mol. Phylogenet. Evol.* **92**, 308–328 (2015).
20. Wagner, F. *et al.* 'At the crossroads towards polyploidy': genomic divergence and extent of homoploid hybridization are drivers for the formation of the ox-eye daisy polyploid complex (*Leucanthemum*, Compositae-Anthemideae). *New Phytol.* **223**, 2039–2053 (2019).
21. Wagner, F., Härtl, S., Vogt, R. & Oberprieler, C. "Fix Me Another Marguerite!": Species delimitation in a group of intensively hybridizing lineages of ox-eye daisies (*Leucanthemum* Mill., Compositae-Anthemideae). *Mol. Ecol.* **26**, 4260–4283 (2017).
22. Piękoś-Mirkowa, H., Mirek, Z. & Miechowka, A. Endemic vascular plants in the Polish Tatra Mts. - distribution and ecology. *Pol. Bot. Stud.* **12**, (1996).
23. Zelený, V. Taxonomisch-chorologische Studie über die Art *Leucanthemum rotundifolium* (W. K.) DC. *Folia Geobot.* **5**, 369–400 (1970).
24. Piękoś, H. Nowy mieszaniec między *Leucanthemum rotundifolium* (W. et K.) DC. a *L. vulgare* Lam. var. *alpicolum* Greml. - Hybrida nova inter *Leucanthemum rotundifolium* (W. et K.) DC. et *L. vulgare* Lam. var. *alpicolum* Greml. *Fragm. Florist. Geobot.* **16**, 319–326 (1970).
25. Rogalski, M., do Nascimento Vieira, L., Fraga, H. P. & Guerra, M. P. Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front. Plant Sci.* **6**, (2015).
26. Greiner, R., Vogt, R. & Oberprieler, C. Evolution of the polyploid north-west Iberian *Leucanthemum* pluriflorum clan (Compositae, Anthemideae) based on plastid DNA sequence variation and AFLP fingerprinting. *Ann. Bot.* **111**, 1109–1123 (2013).
27. Oberprieler, C., Konowalik, K., Fackelmann, A. & Vogt, R. Polyploid speciation across a suture zone: phylogeography and species delimitation in S French *Leucanthemum* Mill. representatives (Compositae–Anthemideae). *Plant Syst. Evol.* **304**, 1141–1155 (2018).
28. Oberprieler, C., Greiner, R., Konowalik, K. & Vogt, R. The reticulate evolutionary history of the polyploid NW Iberian *Leucanthemum* pluriflorum clan (Compositae, Anthemideae) as inferred from nrDNA ETS sequence diversity and eco-climatological niche-modelling. *Mol. Phylogenet. Evol.* **70**, 478–491 (2014).

29. Alexander, P. J., Rajanikanth, G., Bacon, C. D. & Bailey, C. D. Recovery of plant DNA using a reciprocating saw and silica-based columns. *Mol. Ecol. Notes* **7**, 5–9 (2007).
30. Sang, T., Crawford, D. & Stuessy, T. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am. J. Bot.* **84**, 1120 (1997).
31. Scheunert, A., Dorfner, M., Lingl, T. & Oberprieler, C. Can we use it? On the utility of de novo and reference-based assembly of Nanopore data for plant plastome sequencing. *PLOS ONE* **15**, e0226234 (2020).
32. Timme, R. E., Kuehl, J. V., Boore, J. L. & Jansen, R. K. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* **94**, 302–312 (2007).
33. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98 (1999).
34. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
35. Simmons, M. P. & Ochoterena, H. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* **49**, 369–381 (2000).
36. Müller, K. SeqState: primer design and sequence statistics for phylogenetic DNA datasets. *Appl. Bioinformatics* **4**, 65–69 (2005).
37. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Meth* **9**, 772 (2012).
38. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
39. Jukes, T. H. & Cantor, C. R. Evolution of Protein Molecules. in *Mammalian Protein Metabolism* 21–132 (Elsevier, 1969). doi:10.1016/B978-1-4832-3211-9.50009-7.
40. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
41. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
42. Tamura, K., Tao, Q. & Kumar, S. Theoretical Foundation of the RelTime Method for Estimating Divergence Times from Variable Evolutionary Rates. *Mol. Biol. Evol.* **35**, 1770–1782 (2018).
43. Tamura, K. *et al.* Estimating divergence times in large molecular phylogenies. *Proc. Natl. Acad. Sci.* **109**, 19333–19338 (2012).
44. Tao, Q., Tamura, K., Mello, B. & Kumar, S. Reliable Confidence Intervals for RelTime Estimates of Evolutionary Divergence Times. *Mol. Biol. Evol.* **37**, 280–290 (2020).
45. Mello, B., Tao, Q., Barba-Montoya, J. & Kumar, S. Molecular dating for phylogenies containing a mix of populations and species by using Bayesian and RelTime approaches. *Mol. Ecol. Resour.* **21**, 122–136 (2021).

46. Wang, wei-M. On the origin and development of *Artemisia* (Asteraceae) in the geological past. *Bot. J. Linn. Soc.* **145**, 331–336 (2004).
47. Clement, M., Snell, Q., Walker, P., Posada, D. & Crandall, K. TCS: Estimating Gene Genealogies. in *Proceedings of the 16th International Parallel and Distributed Processing Symposium* 311 (IEEE Computer Society, 2002).
48. Leigh, J. W. & Bryant, D. popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).
49. Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M. & Corander, J. Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. *Mol. Biol. Evol.* **30**, 1224–1228 (2013).
50. Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res.* **3**, 93 (2018).
51. Yu, Y., Blair, C. & He, X. RASP 4: Ancestral State Reconstruction Tool for Multiple Genes and Characters. *Mol. Biol. Evol.* **37**, 604–606 (2020).
52. Ali, S. S., Yu, Y., Pfosser, M. & Wetschnig, W. Inferences of biogeographical histories within subfamily Hyacinthoideae using S-DIVA and Bayesian binary MCMC analysis implemented in RASP (Reconstruct Ancestral State in Phylogenies). *Ann. Bot.* **109**, 95–107 (2012).
53. Araújo, M. B. *et al.* Standards for distribution models in biodiversity assessments. *Sci. Adv.* **5**, eaat4858 (2019).
54. Konowalik, K. & Nosol, A. Evaluation metrics and validation of presence-only species distribution models based on distributional maps with varying coverage. *Sci. Rep.* **11**, 1482 (2021).
55. Hamner, B., Frasco, M. & LeDell, E. *Metrics: Evaluation Metrics for Machine Learning*. (2018).
56. Ripley, B. & Venables, W. *nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models*. (2020).
57. Thuiller, W., Georges, D., Engler, R. & Breiner, F. *biomod2: Ensemble Platform for Species Distribution Modeling*. (2020).
58. Therneau, T., Atkinson, B., port, B. R. (producer of the initial R. & maintainer 1999-2017). *rpart: Recursive Partitioning and Regression Trees*. (2019).
59. Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. Opening the black box: an open-source release of Maxent. *Ecography* **40**, 887–893 (2017).
60. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **29**, 1189–1232 (2001).
61. Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. *dismo: Species Distribution Modeling*. (2017).
62. Carlson, C. J. embarcadero: Species distribution modelling with Bayesian additive regression trees in R. *Methods Ecol. Evol.* **11**, 850–858 (2020).
63. Jasiewicz, A. Rośliny naczyniowe Bieszczadów Zachodnich [The Vascular Plants of the Western Bieszczady Mts. (East Carpathians)]. *Monogr. Bot.* **20**, 1–340 (1965).

64. Kornaś, J. Charakterystyka geobotaniczna Gorców [Caractéristique géobotanique des Gorces (Karpathes Occidentales Polonaises)]. *Monogr. Bot.* **3**, 3–230 (1955).
65. Oliveira, G. de, Rangel, T. F., Lima-Ribeiro, M. S., Terribile, L. C. & Diniz-Filho, J. A. F. Evaluating, partitioning, and mapping the spatial autocorrelation component in ecological niche modeling: a new approach based on environmentally equidistant records. *Ecography* **37**, 637–647 (2014).
66. Sobral-Souza, T., Lima-Ribeiro, M. S. & Solferini, V. N. Biogeography of Neotropical Rainforests: past connections between Amazon and Atlantic Forest detected by ecological niche modeling. *Evol. Ecol.* **29**, 643–655 (2015).
67. Varela, S., Anderson, R. P., García-Valdés, R. & Fernández-González, F. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* **37**, 1084–1091 (2014).
68. Barve, N. *et al.* The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol. Model.* **222**, 1810–1819 (2011).
69. Karger, D. N. *et al.* Data from: Climatologies at high resolution for the earth's land surface areas. 7266827510 bytes (2018) doi:10.5061/DRYAD.KD1D4.
70. Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 1–20 (2017).
71. Wing, M. K. C. from J. *et al. caret: Classification and Regression Training.* (2019).
72. Smith, A. B. & Santos, M. J. Testing the ability of species distribution models to infer variable importance. *Ecography* **43**, 1801–1813 (2020).
73. Evans, J. S., Murphy, M. A. & Ram, K. *spatialEco: Spatial Analysis and Modelling Utilities.* (2021).
74. Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C. & Haywood, A. M. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Sci. Data* **5**, 180254 (2018).
75. Araújo, M. B., Whittaker, R. J., Ladle, R. J. & Erhard, M. Reducing uncertainty in projections of extinction risk from climate change. *Glob. Ecol. Biogeogr.* **14**, 529–538 (2005).
76. Zhu, G., Fan, J. & Peterson, A. T. Cautions in weighting individual ecological niche models in ensemble forecasting. *Ecol. Model.* **448**, 109502 (2021).
77. Hijmans, R. J. *et al. raster: Geographic Data Analysis and Modeling.* (2021).
78. R Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, 2019).
79. QGIS Development Team. *QGIS Geographic Information System.* (Open Source Geospatial Foundation Project, 2019).
80. Frajman, B. & Oxelman, B. Reticulate phylogenetics and phytogeographical structure of Heliosperma (Sileneae, Caryophyllaceae) inferred from chloroplast and nuclear DNA sequences. *Mol. Phylogenet. Evol.* **43**, 140–155 (2007).
81. Ronikier, M., Cieślak, E. & Korbecka, G. High genetic differentiation in the alpine plant *Campanula alpina* Jacq. (Campanulaceae): evidence for glacial survival in several Carpathian regions and long-

- term isolation between the Carpathians and the Alps. *Mol. Ecol.* **17**, 1763–1775 (2008).
82. Ehrich, D. *et al.* Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Mol. Ecol.* **16**, 2542–2559 (2007).
83. Šrámková, G. *et al.* Phylogeography and taxonomic reassessment of *Arabidopsis halleri* – a montane species from Central Europe. *Plant Syst. Evol.* **305**, 885–898 (2019).
84. Birks & Willis, K. J. Alpines, trees, and refugia in Europe. *Plant Ecol. Divers.* **1**, 147–160 (2008).
85. Jarčuška, B., Kaňuch, P., Naďo, L. & Krištín, A. Quantitative biogeography of Orthoptera does not support classical qualitative regionalization of the Carpathian Mountains. *Biol. J. Linn. Soc.* **128**, 887–900 (2019).
86. Tadono, T. *et al.* Precise Global DEM Generation by ALOS PRISM. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **II–4**, 71–76 (2014).
87. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography* **20**, (2005).

Tables

Table 1. Divergence times and 95% confidence intervals (CI) for nodes used for dating (nodes A–D from the 1st dataset and E–F from the 2nd dataset). All dating points follow a normal distribution. Numbers are in million years.

Node	Description	Prior distribution		Estimated distribution		Source
		Mean age	95% CI	Mean age	95% CI	
A	Root age, divergence between <i>Artemisia</i> and <i>Leucanthemineae</i>	31.0	32.0–30.0	31.035	31.84–30.239	41,80
B	Divergence between <i>Mauranthemum paludosum</i> and <i>M. ebusitanum</i>	5.3	6.3–4.30	4.380	5.278–4.216	14
C	Divergence between <i>Plagius flosculosus</i> and <i>P. maghrebinus</i>	7.25	8.9–5.6	5.852	7.255–5.596	14
D	Crown age of <i>Leucanthemum</i>	1.93	2.71–1.15	1.417	2.111–1.208	15
E	Divergence between “ <i>L. vulgare</i> group” and <i>L. rotundifolium</i>	1.42	2.21–0.63	1.397	1.87–0.953	This study
F	Divergence between “ <i>L. ircutianum</i> group” and <i>L. rotundifolium</i>	0.65	1.0–0.30	0.657	0.866–0.479	This study

Table 2. Sets of variables used for modeling.

Period	Age/Duration	Source
Pleistocene, Marine Isotope Stage 19 interglaciation (MIS19)	ca. 787 ka	Brown et al., (2018)
Pleistocene, Last Interglacial	ca. 130 ka	Otto-Bliesner et al., (2006)
Pleistocene, Last Glacial Maximum	ca. 21 ka	Karger et al., (2021)
Pleistocene, Heinrich Stadial 1	17.0–14.7 ka	Fordham et al., (2017)
Pleistocene, Bølling–Allerød	14.7–12.9 ka	Fordham et al., (2017)
Pleistocene, Younger Dryas Stadial	12.9–11.7 ka	Fordham et al., (2017)
Pleistocene, early Holocene, Greenlandian	11.7–8.326 ka	Fordham et al., (2017)
Pleistocene, mid-Holocene, Northgrippian	8.326–4.2 ka	Fordham et al., (2017)
Pleistocene, late Holocene, Meghalayan	4.2–0.3 ka	Fordham et al., (2017)
Anthropocene	1979–2013	Karger et al., (2018, 2017)

Table 3. Divergence times of haplotypes obtained from the analysis of the phylogenetic tree constructed with three intergenic spacer regions in cpDNA (*psbA-trnH*, *ndhCret-trnV-comp*, and *trnT(GUU)-comp-psbD-comp*). Numbers are in million years.

Description	Estimated distribution		Posterior probability
	Mean age	95% CI	
<i>L. rotundifolium</i> stem age	0.6571	0.866–0.4787	1.00
A1 crown age	0.1345	0.2598–0.0697	1.00
B1 crown age	0.2262	0.4032–0.1269	0.89
<u>B2</u> crown age	0.0891	0.4032–0.0191	0.96
B3 crown age	0.0879	0.4032–0.0186	0.99
B6 crown age	0.0661	0.3117–0.014	0.96
C1 crown age	0.0475	0.2298–0.0098	1.00
D1 crown age	0.084	0.3992–0.0167	0.95
D3 crown age	0.0159	0.144–0.0817	1.00
D4 crown age	0.0176	0.1507–0.002	1.00
D5 crown age	0.0036	0.0961–0.0001	1.00
E1 crown age	0.0665	0.3213–0.0138	0.99
F1 crown age	0.0216	0.1554–0.003	0.99
G1 crown age	0.0211	0.1934–0.0029	0.84
G5 crown age	0.0062	0.1483–0.0003	0.99
G6 crown age	0.0061	0.1474–0.0003	0.99

Figures

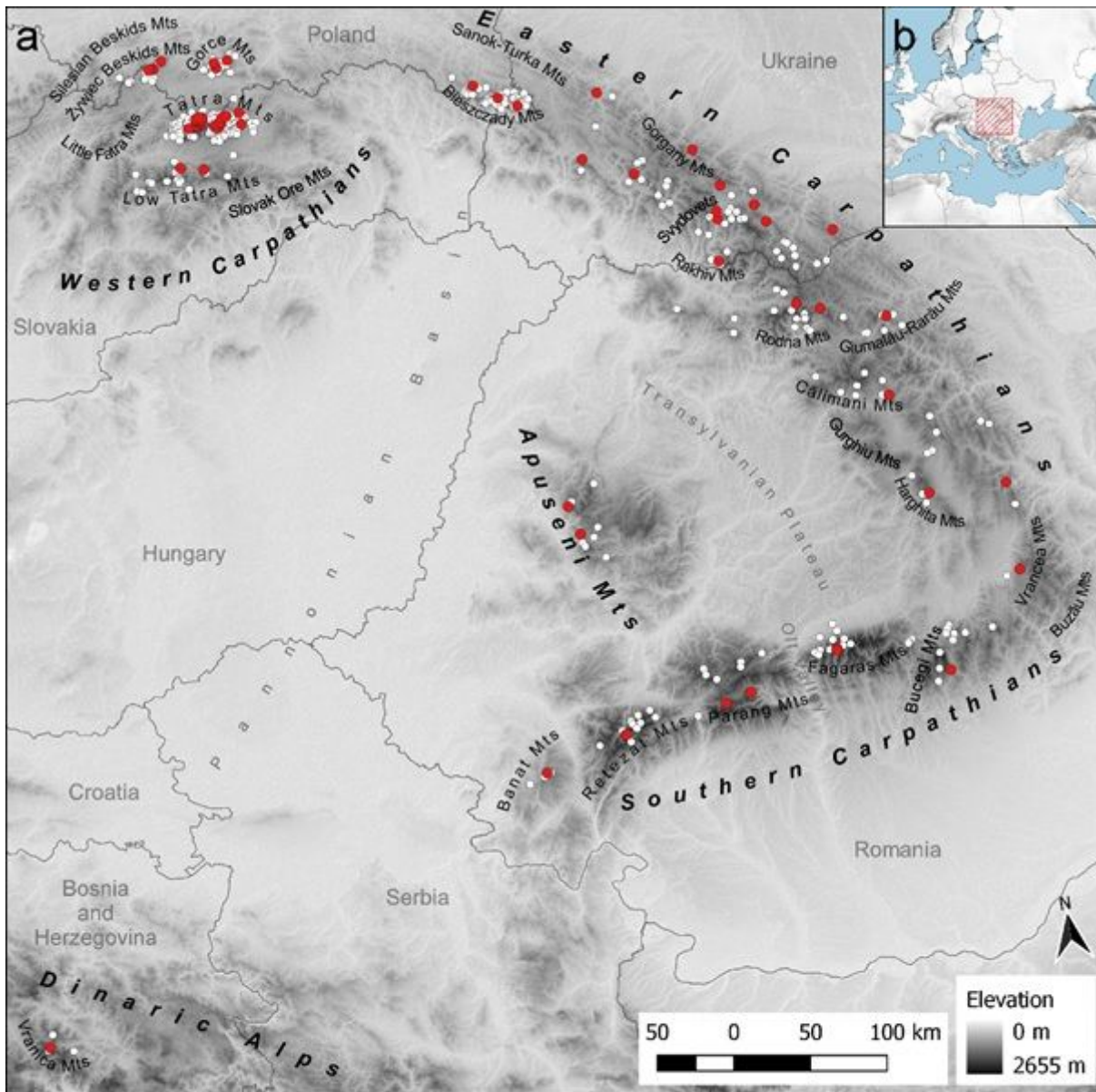


Figure 1

a) The Carpathian region, the area of investigation of this study, with all geographic locations mentioned in the text. Red dots indicate locations sampled for the phylogeographic (cpDNA) study. White dots indicate locations used for ecological niche modeling. Country borders and names are shown for reference. b) Placement of the study area. The underlying image represents the altitude derived from the ALOS DEM dataset⁸⁶. The map is projected in ETRS89 (EPSG: 3035). The maps are drawn using R ver. 3.6.2⁷⁸ and QGIS ver. 3.8⁷⁹.

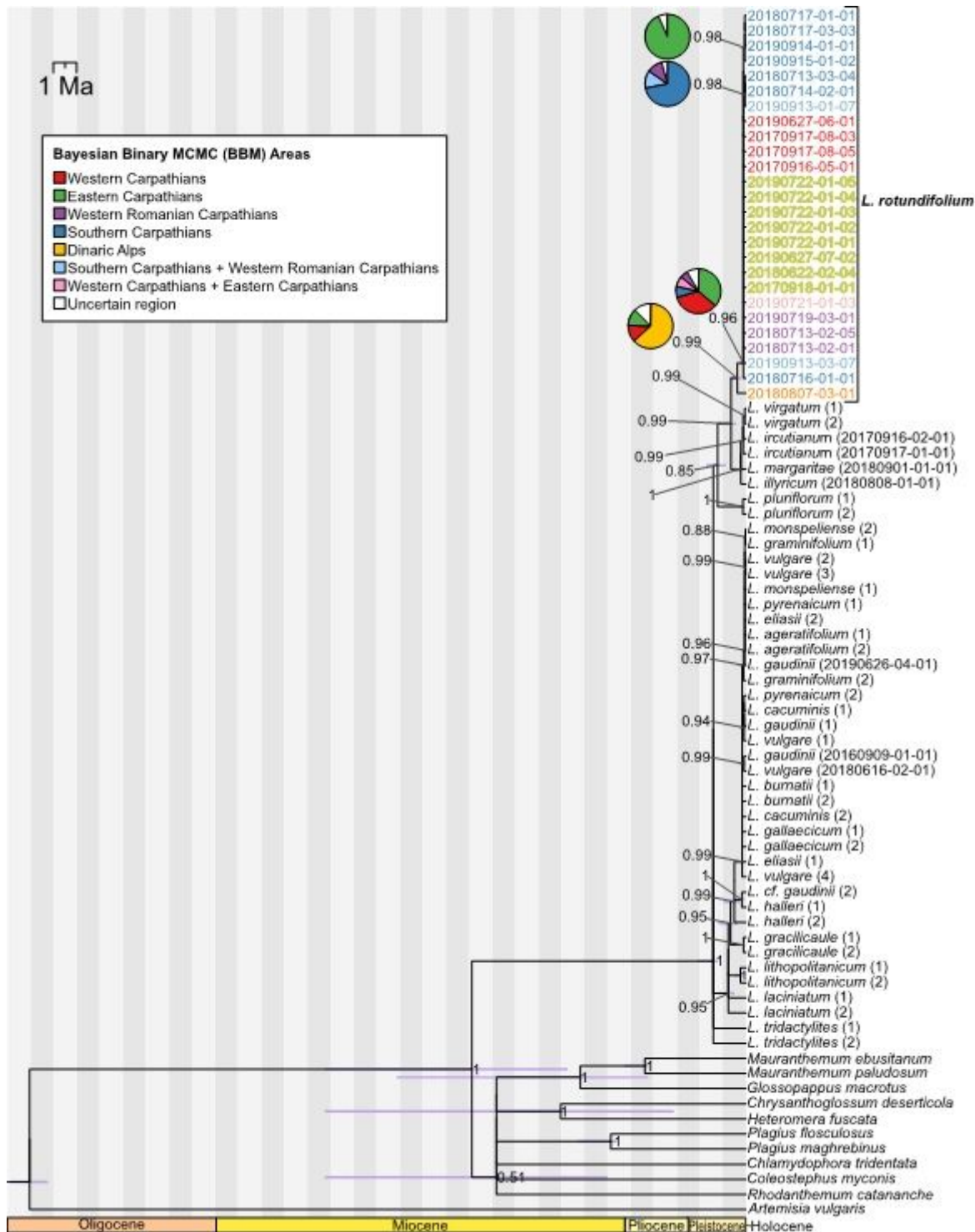


Figure 2

Phylogenetic tree constructed using five intergenic spacer regions in cpDNA (*psbA-trnH*, *trnL-trnF*, *trnC-petN*, *petN-psbM*, and *trnQ-rps16*). The numbers associated with the nodes indicate posterior probabilities from Bayesian analysis. Bars represent uncertainty in the dating of a particular node. Pie charts represent the most plausible ancestral locations inferred by the Bayesian binary Markov Chain Monte Carlo method (BBM), and are shown only for nodes associated with the focal species. The samples of *L. rotundifolium* are colored according to the results of the clustering algorithm (Fig. 4).

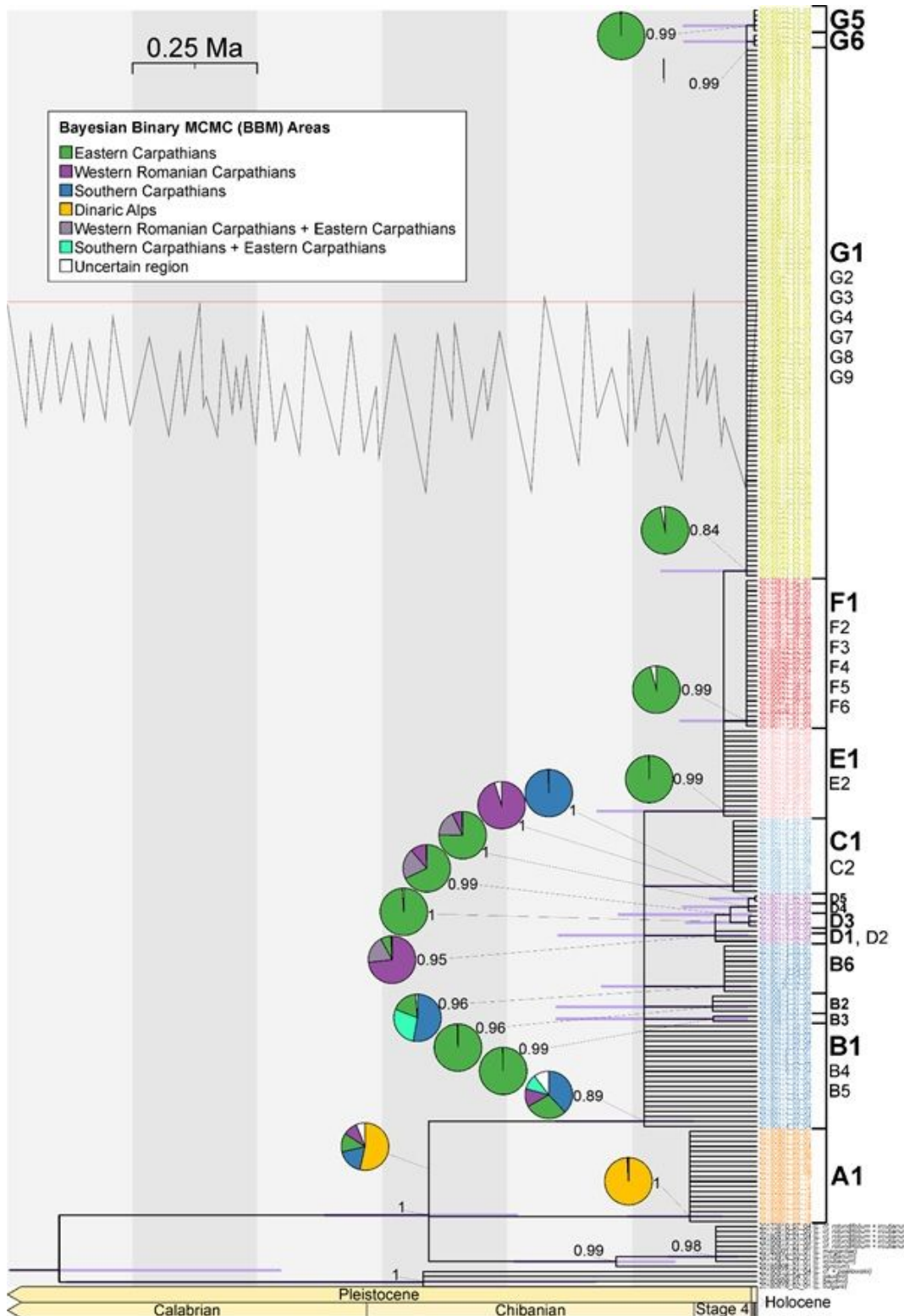


Figure 3

Phylogenetic tree constructed with three intergenic spacer regions in cpDNA (*psbA-trnH*, *ndhCret-trnV-comp*, and *trnT(GUU)-comp-psbD-comp*). The numbers associated with the nodes indicate posterior probabilities from Bayesian analysis. Bars represent uncertainty of dating for a particular node. Pie charts

represent the most plausible ancestral locations inferred by Bayesian binary Markov Chain Monte Carlo method (BBM) and are shown only for nodes associated with the focal species. The *L. rotundifolium* samples are colored according to the results of the phylogenetic analysis of the clustering algorithm (Fig. 4). The tree includes all sympatric species of *L. rotundifolium* (*L. vulgare* 2x, *L. gaudinii* 2x, *L. ircutianum* 4x, *L. margaritae* 6x, and *L. illyricum* 6x). The line in the upper part of the image represents changes in global temperature in the past. The horizontal red line indicates modern thermal conditions, while the peaks indicate higher temperatures and the troughs lower temperatures. The original figure was created by Robert A. Rohde, based on Lisiecki and Raymo (2005), and is available at https://commons.wikimedia.org/wiki/File:Five_Myr_Climate_Change.png.

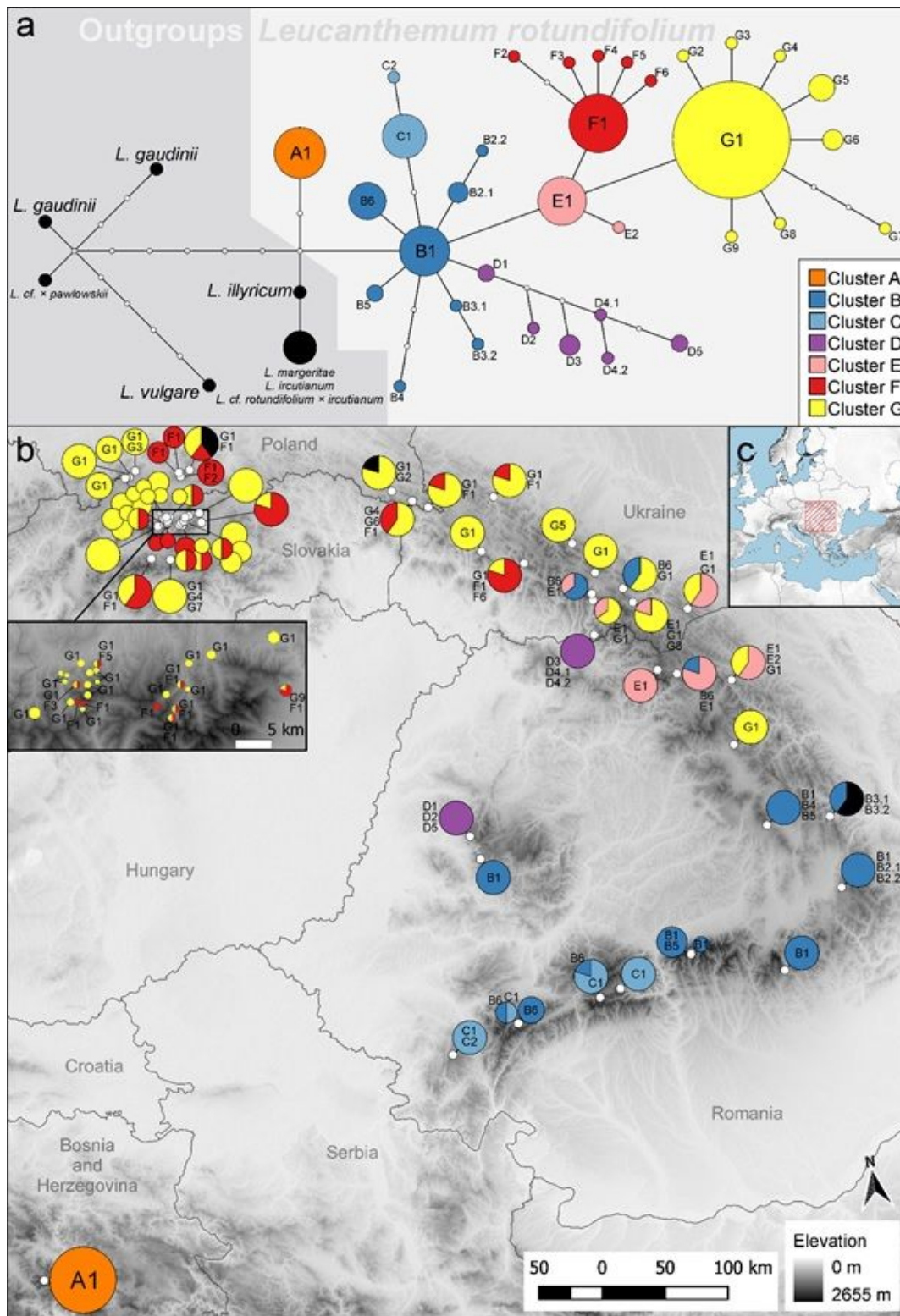


Figure 4

Haplotype network and haplotype distribution. a) Haplotype network constructed with the TCS algorithm using three intergenic spacer regions in cpDNA (*psbA-trnH*, *ndhCret-trnV-comp*, and *trnT(GUU)-comp-psbD-comp*). The network contains all representatives of *L. rotundifolium* included in this study and all sympatric species, but for the sake of clarity, sample names are not shown (for haplotype membership, see Supplementary Material 3). b) Map showing the placement of haplotypes. The size of the pie charts

is proportional to the number of sequenced individuals and the size of the slices is proportional to the percentage of individuals belonging to a certain haplotype. Due to dense sampling in the Tatra Mountains, the area is enlarged within an inset on the left. c) Location of the study region. The underlying image (b, c) represents the altitude derived from the ALOS DEM dataset⁸⁶. The map is projected in ETRS89 (EPSG: 3035). The maps are drawn using R ver. 3.6.2⁷⁸ and QGIS ver. 3.8⁷⁹.

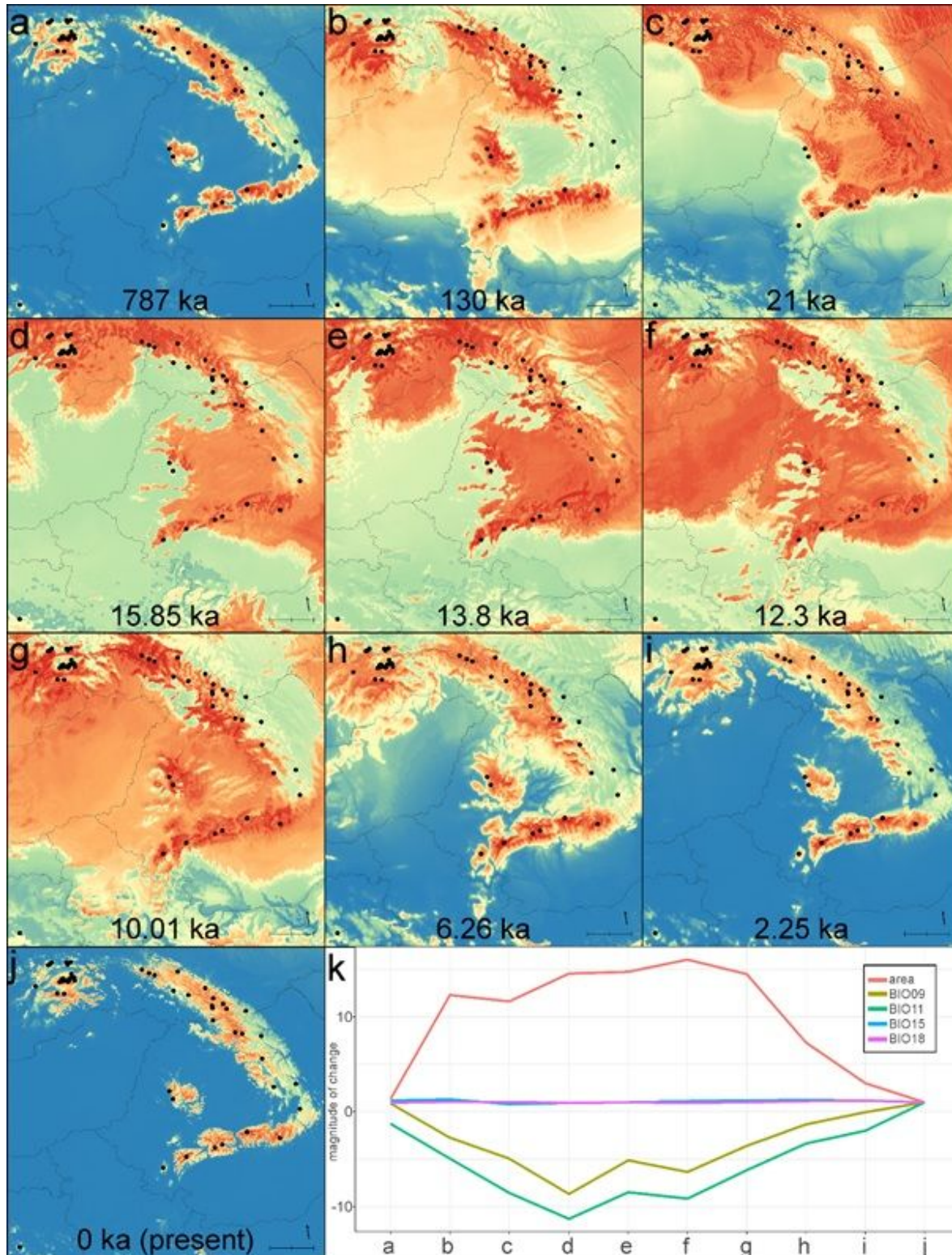


Figure 5

Potential distribution of *L. rotundifolium* throughout the Pleistocene epoch determined by five niche modeling algorithms (MAXENT, BRT, BART, CTA, and ANN) summarized by their first principal component axis. Warmer colors indicate higher suitability, while blue indicates unsuitable areas. a) MIS 19 interglaciation ca. 787 ka; b) last interglacial ca. 130 ka; c) last glacial maximum ca. 21 ka; d) Heinrich Stadial 1 17.0–14.7 ka; e) Bølling-Allerød 14.7–12.9 ka; f) Younger Dryas Stadial 12.9–11.7 ka; g) Greenlandian 11.7–8.326 ka; h) Northgrippian 8.326–4.2 ka; i) Meghalayan 4.2–0.3 ka; j) the Anthropocene, 1979–2013; and k) trends in variables used in the modeling displayed as a mean across the map area, with x-axis labels corresponding to the facets of the figure. BIO09 = mean temperature of the driest quarter; BIO11 = mean temperature of the coldest quarter; BIO15 = seasonality of precipitation; BIO18 = precipitation of warmest quarter; and “area” corresponds to the number of pixels with suitability greater than 30%. The y-axis shows the magnitude of the change, where 1 is the value given for the present conditions. The scale bar in the bottom right corner of each map is equivalent to 100 km. The maps are projected using ETRS 89 (EPSG: 3035). Sampled populations used in this study are shown as black circles to indicate their present distributions. The maps are drawn using R ver. 3.6.2⁷⁸ and QGIS ver. 3.8⁷⁹.

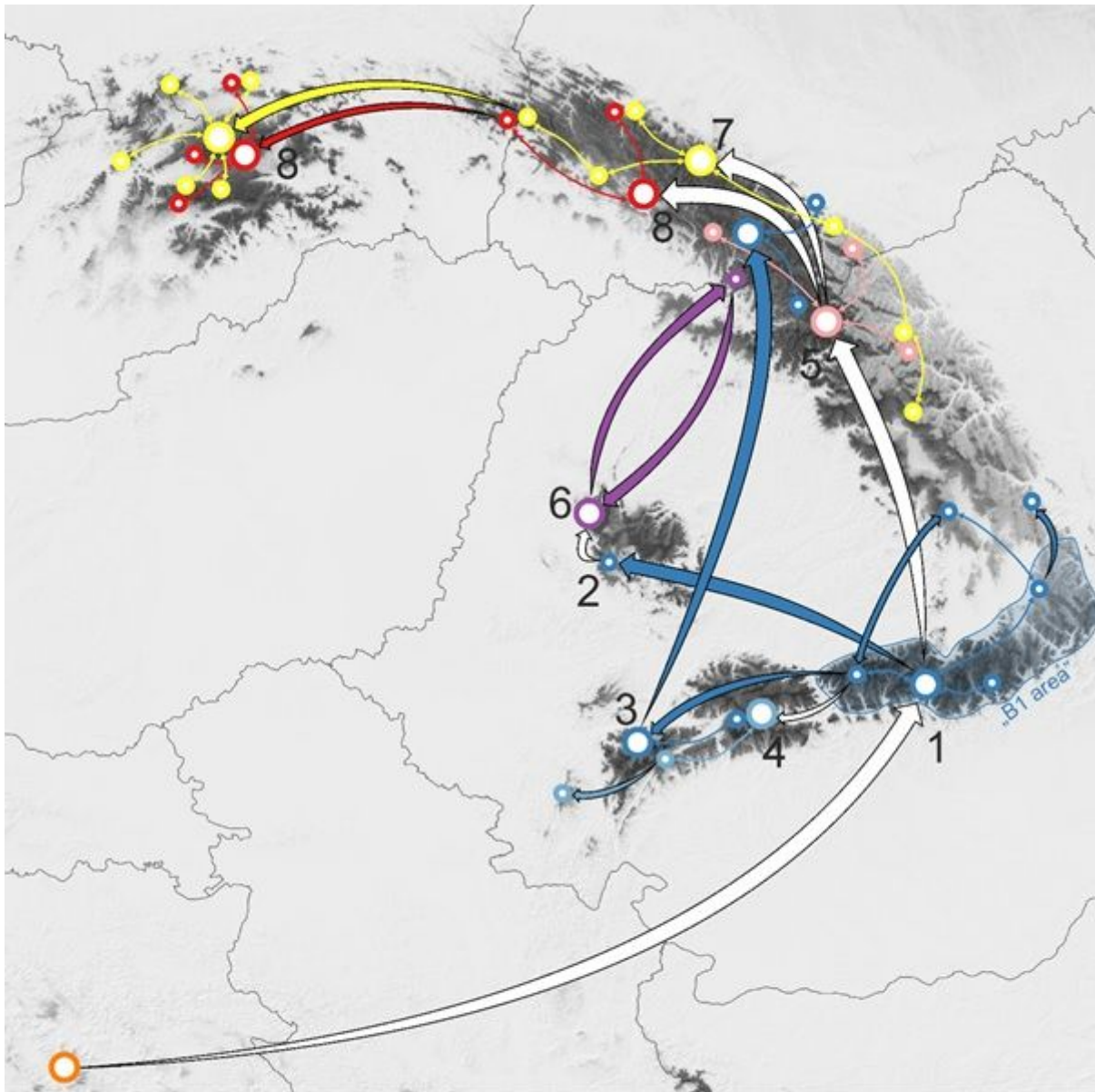


Figure 6

Hypothetical phylogeographic history and migration routes as inferred from haplotype network reconstruction, ecological niche modeling, and RASP analysis (Bayesian binary Markov Chain Monte Carlo). Dispersal and/or vicariance events are numbered and discussed in the text. Thick arrows indicate the emergence of a new haplotype (white arrows) or long-distance dispersal of the haplotypes belonging to the same cluster (arrows in color). Smaller bidirectional arrows indicate connections between populations where the same haplotypes were found. The colors correspond to those assigned to the haplotypes by the clustering algorithm (Fig. 4). The underlying map is generated from ecological niche modeling for the current period, and it is the first axis of the principal component analysis of the five algorithms (as in Fig. 5j) with darker colors indicating higher suitability. For ease of interpretation, political borders between countries are included. The maps are drawn using R ver. 3.6.2⁷⁸ and QGIS ver. 3.8⁷⁹.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial1new.xlsx](#)
- [Supplementarymaterial2.docx](#)
- [Supplementarymaterial3new.xlsx](#)