



The role of cryptic diversity and its environmental correlates in global conservation status assessments: Insights from the threatened bird's-eye primrose (*Primula farinosa* L.)

Spyros Theodoridis¹ | David Nogués-Bravo² | Elena Conti¹

¹Department of Systematic and Evolutionary Botany and Botanic Garden, University of Zurich, Zurich, Switzerland

²Department of Biology, Center for Macroecology, Evolution and Climate, University of Copenhagen, Copenhagen Ø, Denmark

Correspondence

Spyros Theodoridis, Center for Macroecology, Evolution and Climate, Department of Biology, University of Copenhagen, Copenhagen Ø, Denmark.
Email: spyros.theodoridis@bio.ku.dk

Funding information

Pro-Doc Research Module in Plant Sciences and Policy of the Swiss National Science Foundation, Grant/Award Number: 132471; Claraz Foundation; Dept of Systematic and Evolutionary Botany of the Univ. of Zurich; Danish National Research Foundation, Grant/Award Number: DNRF96

Editor: Robert Cowie

Abstract

Aim: Most of the fundamental questions in conservation biogeography require the description of species geographic boundaries and the identification of discrete biological units within these boundaries. International conservation efforts and institutions rely mainly on traditional taxonomic approaches for defining these boundaries, resulting in significant cryptic diversity going undetected and often extinct. Here, we combine high-throughput genomic data with publicly available environmental data to identify cryptic diversity in the threatened bird's-eye primrose (*Primula farinosa*). We aim to characterize evolutionary lineages and test whether they co-occur with changes in environmental conditions. These lineages can be used as intraspecific units for conservation to enhance assessments regarding the status of threatened species.

Location: Europe and temperate Asia (latitude, 40–65°N; longitude, 10°E–115°W).

Methods: We genotyped 93 individuals from 71 populations at 1,220 loci (4,089 SNPs) across the Eurasian distribution of *P. farinosa*. We used phylogenomic and population structure approaches to identify intraspecific lineages. We further extracted statistically derived and remotely sensed environmental information, that is land cover, climate and soil characteristics, to define the biotic and abiotic environment inhabited by each lineage and test for niche similarities among lineages. Additionally, we tested for isolation by distance among populations and applied linear and polynomial regressions to identify lineage-environment associations.

Results: Analyses of genomic data revealed six major lineages within *P. farinosa* corresponding to distinct geographic areas. Niche similarity tests indicated that lineages occupy distinct abiotic and biotic space. Isolation by distance indicated that geography alone cannot explain genetic divergence within *P. farinosa*, while lineage-environment associations suggested potential adaptation to different abiotic conditions across lineages. However, relationships with the land cover classes, a proxy for habitat, were weaker.

Main conclusion: Our results highlight the need for incorporating intraspecific diversity in global assessments of species conservation status and the utility of genomic and publicly available environmental data in conservation biogeography.

KEYWORDS

CHELSEA, high-throughput genomic data, IUCN, land cover, MODIS

1 | INTRODUCTION

Most of the fundamental questions in conservation biogeography require knowledge of the geographic distributions and ecological niches of individual species (Riddle, Ladle, Lourie, & Whittaker, 2011; Whittaker et al., 2005). This knowledge is essential to better understand species responses in a rapidly changing world and prevent the ever-increasing loss of their diversity (Waldron et al., 2017). However, conservation biologists and international efforts, such as the Red List of the International Union for Conservation of Nature (IUCN), often face the double challenge of describing species geographic boundaries and identifying discrete biological units within these boundaries (Keith et al., 2015; Riddle et al., 2011). The identification of subspecific units is of major importance in order to protect both evolutionary history and ecological processes below the species level (Faith et al., 2010; Moritz, 2002; Palsbøll, Bérubé, & Allendorf, 2006).

Wide-range species, particularly those with limited vagility or dispersal potential, often show strong phylogeographic structure across their distributions (Avice, 2009). This structure has been mostly shaped by past geological, ecological and evolutionary processes, and the resulting intraspecific units (i.e., historically isolated groups of populations; hereafter referred to as "lineages") frequently display distinct ecological characteristics (Allendorf, Luikart, & Aitken, 2013). Although intraspecific lineages have long been recognized as significant units for conservation (Moritz, 1994; Ryder, 1986), relevant studies have focused more on the genetic (using mostly a few neutral markers) and less on the ecological (and potentially adaptive) distinctiveness of these lineages (Funk, McKay, Hohenlohe, & Allendorf, 2012; De Guia & Saitoh, 2007). However, responding to recent and ongoing destructive impacts of global climate change and human activities on intraspecific diversity (Miraldo et al., 2016; Pauls, Nowak, Balint, & Pfenninger, 2013), and therefore to ecosystems (Des Roches et al., 2018), requires the integration of genetic and environmental data. Conservation studies integrating both biodiversity aspects are still scarce but are expected to increase as more genetic and environmental data are becoming available (Jenkins, Yannic, Schaefer, Conolly, & Lecomte, 2018; Wilting et al., 2015; Yannic et al., 2017).

Characterization of intraspecific diversity is often linked to the identification of morphologically cryptic lineages (Bickford et al., 2007; Riddle et al., 2011). Cryptic diversity is commonly defined as the occurrence of distinct evolutionary lineages that are otherwise morphologically indistinguishable within a nominal species (Bickford

et al., 2007; Struck et al., 2018). Despite the morphological similarities, cryptic lineages may carry not only unique evolutionary trajectories but also the potential of differing responses to ongoing and future global change (Bernardo, 2011; Feckler et al., 2014; Paaby & Rockman, 2014). Therefore, biodiversity assessments that ignore cryptic lineages, including them in a single species or a single conservation unit, may severely undervalue current and future biodiversity patterns (Bálint et al., 2011; Fišer, Robinson, & Malard, 2018; Riddle et al., 2011). Cryptic diversity is typically assessed using genetic data, and the advent of high-throughput sequencing has the potential to revolutionize the discovery of cryptic lineages across the tree of life (Allendorf, Hohenlohe, & Luikart, 2010; Benestan et al., 2016; Funk et al., 2012). However, the number of studies that use genome-wide data to uncover cryptic diversity remains small (see Struck et al., 2018 for a review). Additionally, there is an ever growing need to broaden the definition of cryptic diversity and incorporate environmental data when assessing intraspecific cryptic variation (Espindola et al., 2016; Freudenstein, Broe, Folk, & Sinn, 2017; Sullivan et al., 2019).

Assessment of environmental conditions across populations has traditionally relied on time-consuming and costly field observations and measurements of relevant ecological variables, with time and cost increasing with species range, number of populations and number of variables (Albert et al., 2010). However, the ever-increasing public availability of spatial environmental data is advancing biodiversity research and enables the time- and cost-effective discovery of ecological biodiversity patterns across taxonomic and spatial scales (Franklin, Serra-Diaz, Syphard, & Regan, 2017; La Salle, Williams, & Moritz, 2016). Data on abiotic variables directly involved in species' physiological limitations and distributions, such as climate and soil characteristics, are publicly available from several sources either as spatial interpolations of field measurements, satellite observations of the earth's surface (i.e., remote sensing) or a combination of both (Franklin et al., 2017). Additionally, remote-sensing technologies offer the means to closely approximate local biodiversity patterns and biotic interactions at large to moderate spatial scales in the form of land cover classes (He et al., 2015; Lausch et al., 2016). Coupled with high-throughput genomic data, statistically derived and remotely sensed environmental data can significantly enhance our ability to uncover and document diversity patterns, yet studies that integrate these diverse data are lacking (Bush et al., 2017; Yamasaki et al., 2017).

The bird's-eye primrose (*Primula farinosa* L.) represents an excellent system to investigate potential patterns of cryptic diversity

relevant for conservation. *Primula farinosa* is a cold-adapted herb in the family Primulaceae characterized by having a wide, but disjunct distribution in Europe and north-east Asia (Hamblen & Dixon, 2003; Krasnoborov, 2000; Roskov et al., 2018; Shishkin & Bobrov, 1952). Previous taxonomic treatments based on a wide range of diagnostic traits suggested that *P. farinosa* is largely absent from western and central Siberia, where it is replaced by closely related species, but it occurs from the east of the Altai mountains through Mongolia to eastern Siberia and the Kamchatka peninsula (Fedorov, 2001; Krasnoborov, 2000; Richards, 2003; Shishkin & Bobrov, 1952). The species shows significant variability with many forms and intermediates over its extended Eurasian distribution (Shishkin & Bobrov, 1952; Krasnoborov, 2000; Theodoridis, personal observations), and the lack of distinct morphological differences among these forms has resulted in significant controversy among taxonomists (Hamblen & Dixon, 2003; Krasnoborov, 2000; Richards, 2003; Shishkin & Bobrov, 1952). The species grows in calcareous grasslands and alpine streams or lakes at altitudes between 400 and 3,000 m a. s. l. in southern Europe (Carpathians, Alps, Iberia) and Mongolia, while in northern

Europe (British Isles, Baltic region and southern Scandinavia) the species grows exclusively in wetlands between 0 and 400 m a.s.l. (Figure 1; Hamblen & Dixon, 2003; Richards, 2003; Theodoridis, Randin, Broennimann, Patsiou, & Conti, 2013). Regarding its conservation status, *P. farinosa* has gone extinct or is facing extinction mainly caused by human-driven changes in its wetland microhabitats and is included in the IUCN Red Lists of several European countries (Croatia: Topić & Stančić, 2006; Ukraine: Didukh, 2009; Hungary: Salamon-Albert & Morschhauser, 2003; Poland: Gajewski, Sitek, Stolarczyk, Nowak, & Kapala, 2013; Denmark: Sørensen, Larsen, Orabi, & ørgaard, 2014; United Kingdom: Cheffings et al., 2005; Slovakia: Turis et al., 2014; Romania: Coldea, Stoica, Pușcaș, Ursu, & Oprea, 2009). Additionally, the species is predicted to suffer a dramatic reduction in its climatically suitable habitats in certain regions across its European range (Theodoridis, Patsiou, Randin, & Conti, 2018).

Here, we use *P. farinosa* to (a) identify potential cryptic genetic diversity relevant for IUCN conservation status assessments and (b) test for environmental correlates of intraspecific cryptic diversity

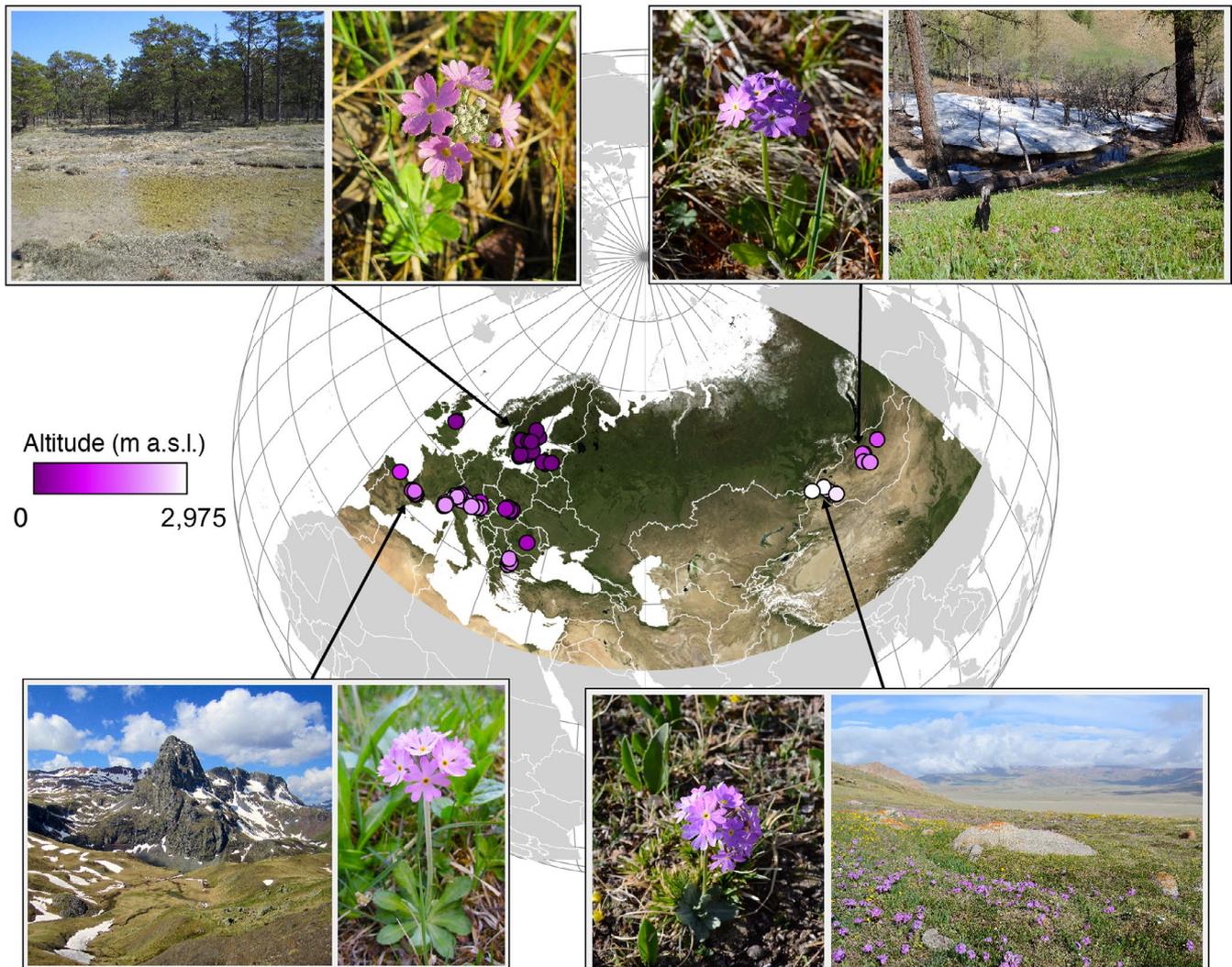


FIGURE 1 Sampling localities of populations of *Primula farinosa* across its Eurasian range. Altitude is indicated by colour (purple to white) at each locality. Pictures represent species morphology and variation of habitat type in four distinct geographic regions

in species with wide continental distributions. To this end, we use previously published and newly generated high-throughput genomic data to identify intraspecific lineages across the species' range. We also extract statistically predicted and remotely sensed environmental variables to approximate the biotic and abiotic habitat of the species. Finally, we test for significant associations between genetic groups and environmental variables. By combining genomic with publicly available environmental data, we aim at contributing novel ways for assessing intraspecific cryptic diversity in conservation biogeography.

2 | METHODS

2.1 | Sampling

To estimate genetic structure and ecological diversity within *P. farinosa*, we sampled 71 populations that encompass most of its known Eurasian distribution (Figure 1). Our sampling design aimed at representing most genetic and ecogeographic variation within *P. farinosa* (Albert et al., 2010; Gotelli & Stanton-Geddes, 2015). For its European distribution, we used 57 populations reported previously (Theodoridis et al., 2018, 2017) and selected one to three individuals per population based on the size of the available genomic data (i.e., maximizing the total number of sequencing reads per individual). For its Asian distribution, we sampled 14 populations for the first time (two individuals per population) across Mongolia. All populations were sampled during the springs and summers of 2011, 2012, and 2013 and were located at least 15 km apart from each other. Additional sampling details are provided in Table S1. Distribution and sampling maps were created using the STRABO library (<https://github.com/spyrostheodoridis/strabo>) in JavaScript.

2.2 | Sequencing and bioinformatics

DNA extraction and sequencing followed the approach described by Theodoridis et al. (2017). Briefly, leaf samples collected from 93 individuals were preserved in silica gel and DNA was extracted using a modified CTAB protocol (Doyle & Doyle, 1990). Genomic library preparation was performed by SNPsaurus (<http://snpsaurus.com/>) using nextRAD libraries with the selective primer sequence GTGTAGAGC. Single-end sequencing was performed in three lanes of an Illumina HiSeq 2000 (University of Oregon High-throughput Sequencing Core), producing raw reads that were 101 bp long.

Raw DNA sequences were filtered and trimmed following the recommendations of Minoche, Dohm, and Himmelbauer (2011) as implemented in GONOSPY v0.1 package (<https://github.com/spyrostheodoridis/gonospy>) in Python. Initial analysis of the raw reads indicated the existence of a few invariable organellar sites with extremely high coverage. Therefore, we further excluded these sites by mapping the filtered reads to the chloroplast genome of *Primula veris* and the mitochondrial genome of *Vaccinium macrocarpon* (Ericaceae) using STAMPY v1.0.28 (Lunter & Goodson, 2011) and applying an expected divergence from the reference (substitution

rate) of 0.05 substitutions per site. After filtering, our final dataset consisted of nuclear reads that were 90 bp long. To identify putative homologous loci across all individuals, we employed the STACKS pipeline v.1.44 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) and applied the following parameter values: we required a minimum coverage (*m*) of five identical reads per stack by allowing a maximum of two gaps (*max_gaps*) between reads; we also removed stacks with coverage of more than two standard deviations above the mean (Catchen et al., 2013); we allowed a maximum distance (*M*) between stacks of two nucleotides and a maximum number of two stacks at a single de novo locus in each individual; a maximum of six mismatches (*n*) and a maximum of four gaps (*max_gaps*) was allowed between putatively homologous loci across all individuals.

2.3 | Phylogenomic inference and genetic structure

To assess intraspecific genetic structure and phylogenetic relationships in *P. farinosa*, we used the Variant Call Format (VCF) file exported by STACKS and filtered it using GONOSPY to exclude SNPs (a) that were represented by $\leq 5\%$ of the total reads for the corresponding locus within each individual, (b) with a frequency ≤ 0.05 across all individuals and (c) absent in more than 30% of the sampled individuals. To infer phylogenetic relationships among the sampled individuals of *P. farinosa*, we used the filtered SNP data (as described above), including all SNPs per locus, and a maximum likelihood approach implemented in RAXML v8.2.11 (Stamatakis, 2014). Alignment sites consisting of only heterozygotes and homozygotes for a single allele (but not homozygotes for the alternative allele) are considered invariant by RAXML; therefore, we further excluded these sites using GONOSPY. The final filtered data matrix consisted of 4,089 SNPs. To account for the lack of invariant sites, we applied an ascertainment bias correction ($--asc-corr = lewis$) to the likelihood calculations as recommended by Leaché, Banbury, Felsenstein, Nieto-Montes de Oca, and Stamatakis (2015). We used the GTR + gamma model of sequence evolution determined by the model selection procedure implemented in IQ-TREE (Nguyen, Schmidt, von Haeseler, & Minh, 2015). We also applied the rapid bootstrap algorithm with 1,000 replicates and performed a full search for the best scoring tree. The final tree with associated branch support values was visualized in ITOOL webserver (Letunic & Bork, 2016).

Genetic structure was inferred using the variational Bayesian framework implemented in the software FASTSTRUCTURE (Raj, Stephens, & Pritchard, 2014). Since FASTSTRUCTURE assumes that the investigated loci are unlinked, we selected one random SNP per locus and repeated this analysis 20 times to account for the stochasticity stemming from the random choice of SNPs (Theodoridis et al., 2017). After filtering, the 20 final data matrices each consisted of 1,220 randomly chosen SNPs. We ran the analyses for numbers of groups, *K*, ranging from 1 to 10 and further applied the chooseK.py program to estimate the most likely number of *K*. For the *K* values that best explained our data, the results of the 20 randomly selected SNP replicates were combined using the "greedy" algorithm

within CLUMPP v1.1.2 (Jakobsson & Rosenberg, 2007). The results of the CLUMPP analysis were visualized using the D3 JavaScript library (Bostock, Ogievetsky, & Heer, 2011).

We further evaluated the degree of genetic divergence among the genetic groups identified from the phylogenomic and genetic structure analyses using Nei's standard genetic distance (Nei, 1972). Following Joly, Bryant, and Lockhart (2015), for each pairwise comparison we first calculated the mean distance of all SNPs within each locus and then calculated the genome-wide distance by estimating the mean across all locus distances using custom Python scripts.

2.4 | Biotic environment

To approximate the type of biotic environments across *P. farinosa* populations, we used the Global Consensus Land Cover database (Tuanmu & Jetz, 2014) that provides information on the prevalence of 12 land cover (including land use) classes at 1-km resolution (<http://www.earthenv.org/landcover.html>). We extracted prevalence values for all classes and populations and further excluded classes with zero prevalence for all populations. The final matrix consisted of the following ten land cover classes: evergreen/deciduous needleleaf trees, deciduous broadleaf trees, mixed/other trees, shrubs, herbaceous vegetation, cultivated and managed vegetation, regularly flooded vegetation, snow/ice, barren, open water. We subsequently reduced the dimensions of the dataset by performing principal components analysis (PCA) implemented in the package MATPLOTLIB (Hunter, 2007) in Python.

2.5 | Abiotic environment

We approximated the abiotic conditions of *P. farinosa* populations using soil, precipitation and temperature variables obtained from two different sources (global reanalysis data from CHELSA and remote sensed data from MODIS; see below). We combined the above variables in two different datasets, one that included the soil, precipitation and the CHELSA temperature data (hereafter abiotic/CHELSA) and a second one that included the soil, precipitation and the MODIS temperature data (hereafter abiotic/MODIS). We then reduced the dimensions of each of the two datasets by performing principal components analysis as described above (see Biotic environment section). The individual abiotic variables are described below.

2.5.1 | Soil attributes

We characterized soil attributes using six soil variables obtained from the web-based global soil information system (SoilGrids; <https://soilgrids.org>) made available by the International Soil Reference and Information Center (ISRIC) at 250 m resolution (Hengl et al., 2017). The SoilGrids system provides global predictions for standard numeric soil properties and is generated using automated statistical mapping and machine learning (Hengl et al., 2017). For each sampled population, we extracted information on the following variables: Predicted most probable class following the World Reference Base

(predClass), soil pH in H₂O at 0 cm (pH1), soil pH in H₂O at 5 cm (pH2), absolute depth to bedrock (absDepth), soil organic carbon content at 0 cm (carbCont1) and soil organic carbon content at 5 cm (carbCont2).

2.5.2 | Temperature and precipitation (CHELSA)

Precipitation and temperature at 2 m above ground for each sampled population were obtained from CHELSA (Climatologies at high resolution for the earth's land surface areas; <http://chelsa-climate.org/>) at 30 arcsec (c. ~1 km on the equator) resolution (Karger et al., 2017). CHELSA includes monthly (mean, maximum and minimum) temperature and precipitation patterns for the time period 1979–2013 derived by downscaling the model output temperature and precipitation estimates of the ERA-Interim climatic reanalysis (i.e., downscaled global reanalysis data; Karger et al., 2017). Using the monthly temperature and precipitation values, we generated the following set of seven climatic variables that describe annual and growing season climate trends for the sampled populations: minimum temperature of growing season (tminGrow), maximum temperature of growing season (tmaxGrow), average temperature range during growing season (tDiffGrow), average temperature of growing season (tmeanGrow), growing degree days (GDD), annual precipitation (precAnnual) and precipitation during the growing season (prcGrow). Growing season for *P. farinosa* was defined as the period from March to September (Hamblen & Dixon, 2003).

2.5.3 | Temperature (MODIS)

In addition to the CHELSA temperature data, we obtained land surface (0 m above ground) temperature derived from thermal infrared measurements of the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua Earth Observing System satellites (Wan & Dozier, 1996). MODIS/Terra (Wan, Hook, & Hulley, 2015, 2015) and MODIS/AQUA (Wan, Hook, & Hulley, 2015, 2015) land surface temperature data are produced daily at 1 km spatial resolution. Because year-round data from both satellites are available since the year 2003, we considered the time period between 2003 and 2017, and for each population, we downloaded and extracted daily temperature data for the growing season (March–September) using the PYMODIS v2.0.9 (<http://www.pymodis.org/>) and CHOROSPY v0.1 (<https://github.com/spyrostheodoridis/chorospy>) packages in Python. We then merged the data from both satellites and calculated mean, maximum and minimum monthly temperatures. Subsequently, we used the monthly values to calculate the same five temperature-related variables as described above (tminGrow, tmaxGrow, tDiffGrow, tmeanGrow and GDD).

2.6 | Niche overlap and niche similarity tests

To assess ecological differentiation within *P. farinosa*, we quantified climatic niche differences among the identified genetic groups using the approach described by Broennimann et al. (2012) in

two-dimensional environmental space (PCA; see also Theodoridis et al., 2013). We first calculated niche overlap between genetic groups using Schoener's D similarity index (Schoener, 1970), a metric that ranges from 0 (completely discordant niches) to 1 (identical niches). We then defined a background extent for each genetic group using a 20 km buffer zone around the sampling localities (Theodoridis et al., 2018) and tested whether the ecological niches of the groups tend to be more similar to each other than would be expected by chance using background (or niche) similarity tests (Broennimann et al., 2011; Warren, Glor, & Turelli, 2008). These tests randomly shift the density of occurrences in one range and calculate similarity (Schoener's D) using the observed niche from the other range. We repeated the process 1,000 times for both ranges under comparison, and the observed D value was compared to the distribution of the randomized D values in a one-sided t test (Broennimann et al., 2011). The tests were conducted for each environmental category (biotic, abiotic/CHELSEA, abiotic/MODIS). For each of the two abiotic categories, the soil variables (and the CHELSEA precipitation variables for the abiotic/MODIS) were reprojected and resampled to match the projection and resolution of the respective temperature layers using the Geospatial Data Abstraction Library v2.3.2 (GDAL, www.gdal.org). All tests were conducted using the "ecospat" package in R (Di Cola et al., 2017).

2.7 | Isolation by distance and ancestry/environment associations

We evaluated the role of geographic distance in explaining the genetic divergence across all sampled Eurasian populations of *P. farinosa*. We calculated pairwise genetic distances as the average nucleotide differences between populations across all SNPs. For each considered SNP that was sampled in both populations under any pairwise comparison, we assigned a distance of 1 when populations shared no alleles and 0 otherwise. Geographic distances among populations were calculated using the haversine formula. To test for the significance of the relationship between genetic and geographic distance, we performed a Mantel test implemented in the scikit-bio v0.5.5 (<http://scikit-bio.org/>) library in Python using the log-transformed geographic and genetic distances and the Pearson correlation coefficient with 10,000 permutations. To account for the large geographic distances between the European and the Asian populations, we performed the tests on two different datasets, one containing all sampled Eurasian populations and one containing only the European populations.

We further tested whether the observed variation in biotic and abiotic environmental variables among populations could be explained by the ancestry profile (i.e., ancestry probabilities) of each individual. Significant ancestry/environment associations would be indicative of distinct environmental niches for each individual or group of individuals with similar ancestry profiles. To this end, we fitted multiple linear and polynomial (quadratic) regression models with the estimated posterior probabilities of each individual belonging to a specific ancestral group as independent

variables, and each of the three main PCA axes for each variable category (biotic, abiotic/CHELSEA, abiotic/MODIS) as response variables. To avoid collinearity among the independent variables, we excluded the last ancestry probability for each individual as linearly dependent (all ancestry probabilities sum up to one). Below, we provide the linear regression formula for the general case of K identified ancestral groups.

$$\hat{E} = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + \dots + \beta_{k-1} P_{k-1}$$

where P_1, P_2, \dots, P_{k-1} are the posterior probabilities of each individual belonging to the respective ancestral group (i.e., group one to group $K - 1$) and \hat{E} is the predicted environmental score along the considered PCA axis. For populations represented by more than one individual (i.e., Balkan and Asian populations), and since all individuals within populations showed identical ancestry profiles (see Results), we randomly selected one individual. Linear and polynomial regression models were fitted using the "lm" and "glm" functions in R version 3.4.0 (R Development Core Team, 2017), respectively. To further test the adequacy of linear models in explaining environmental variation across populations, we tested for residual spatial autocorrelation in each model using Moran's I with kernel distance weights implemented in the python package PYSAL v.2 (<https://pysal.org/>) and 1,000 simulations.

3 | RESULTS

3.1 | Phylogenomic inference and genetic structure

Phylogenomic analyses revealed four major clans/groups (terminology for unrooted trees follows Wilkinson, Mclnerney, Hirt, Foster, & Embley, 2007) separated by long branches with bootstrap support of 100 (Figure 2a). These four clans correspond to four major regions of the Eurasian distribution of *P. farinosa*: western-central-northern Europe, Balkan peninsula, Altai mountains and Khangai-Khentii mountains in central Mongolia (Figure 2). Additionally, RAXML analyses identified three clans within the western-central-northern Europe group (branch bootstrap support of 100), corresponding to major geographic clades reported in previous studies (Theodoridis et al., 2018, 2017), namely the Carpathian, the Iberian and the central-northern European clade. The results of the genetic structure analyses are largely consistent with those obtained from the phylogenomic analyses (Figure 2). Specifically, for ten out of the 20 SNP replicates, the chooseK.py program assigned all individuals to five groups ($K = 5$), while nine fastStructure replicates supported four groups and only one replicate supported three ancestral groups. Additionally, populations within the clans identified in the RAXML unrooted tree showed similar ancestry profiles in the fastStructure analyses (Figure 2b). Moreover, and adding further support to the results of the phylogenomic analyses, Nei's standard genetic distances were lower among the western-central-northern European lineages (0.048–0.098) compared to the distances among all four major Eurasian lineages (0.362–0.517; Table 1).

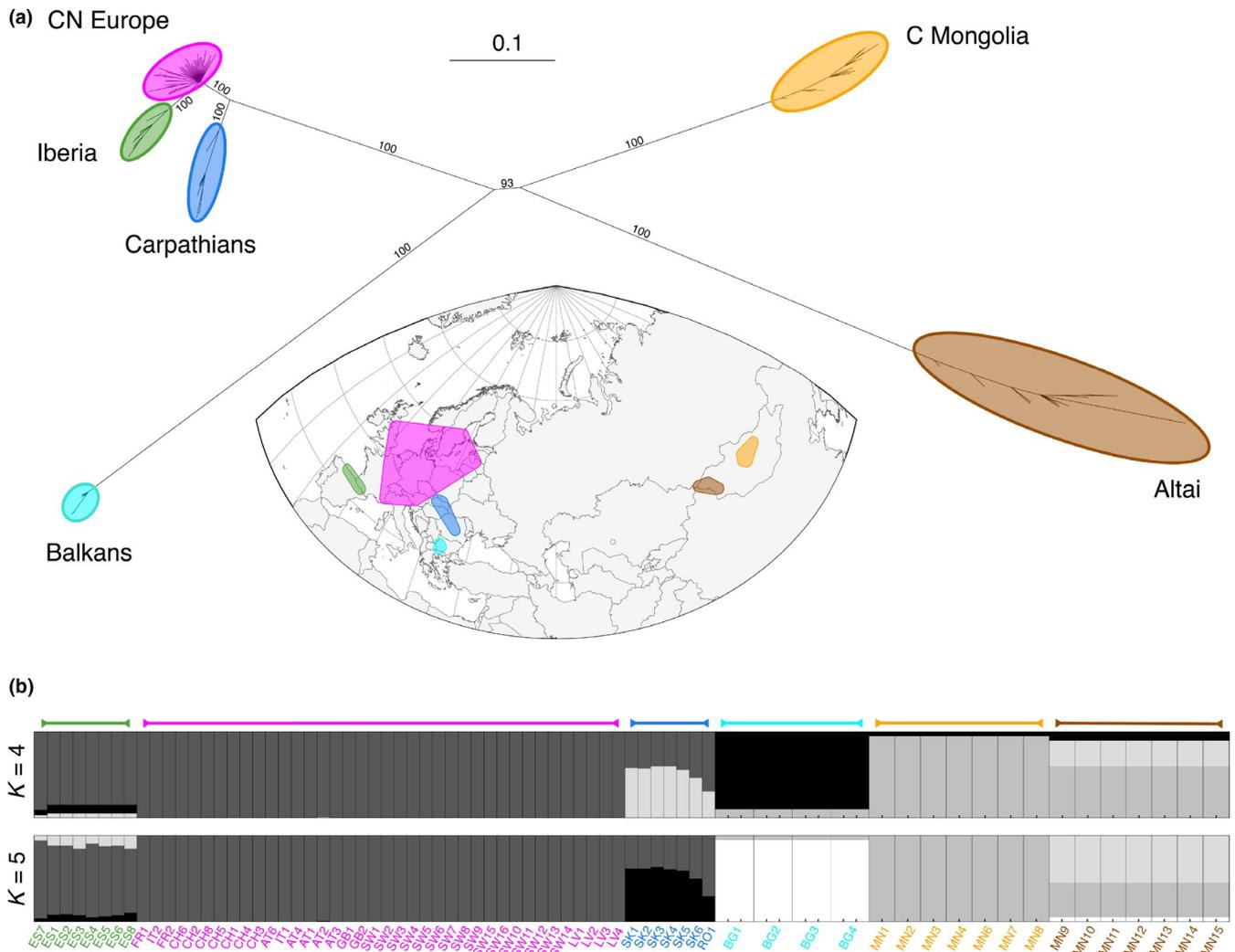


FIGURE 2 Phylogenomic and population structure analyses of *Primula farinosa*. (a) Maximum likelihood (RAxML) unrooted tree inferred from a matrix of 4,089 variant sites (SNPs). Colours indicate distinct lineages (i.e., groups of populations separated by long branches) and correspond to distinct geographic regions (inset map). Numbers on major branches indicate branch support (1,000 bootstrap replicates). Branches with <70% bootstrap support were collapsed. (b) The results of the fastStructure analyses based on 20 repetitions of randomly chosen SNPs from 1,220 loci. The Balkan and Asian populations are represented by three and two individuals, respectively (separated by black inner ticks)

TABLE 1 Nei's standard genetic distances between all identified genetic groups/lineages from the phylogenomic and structure analyses

	Iberia	CN Europe	Carpathians	Balkans	C Mongolia	Altai
Iberia	-	0.048	0.098	0.483	0.386	0.458
CN Europe	-	-	0.076	0.463	0.37	0.44
Carpathians	-	-	-	0.466	0.367	0.442
Balkans		0.457		-	0.443	0.517
C Mongolia		0.362		-	-	0.381
Altai		0.434		-	-	-
	Iberia + CN Europe + Carpathians					

Note: Numbers above the diagonal indicate the distances between all six lineages while numbers below the diagonal (grey box) indicate the distances between the three major lineages (i.e., Balkans, C Mongolia and Altai) and the main European lineage after merging the three European groups, that is Iberia, CN Europe and Carpathians, into one group

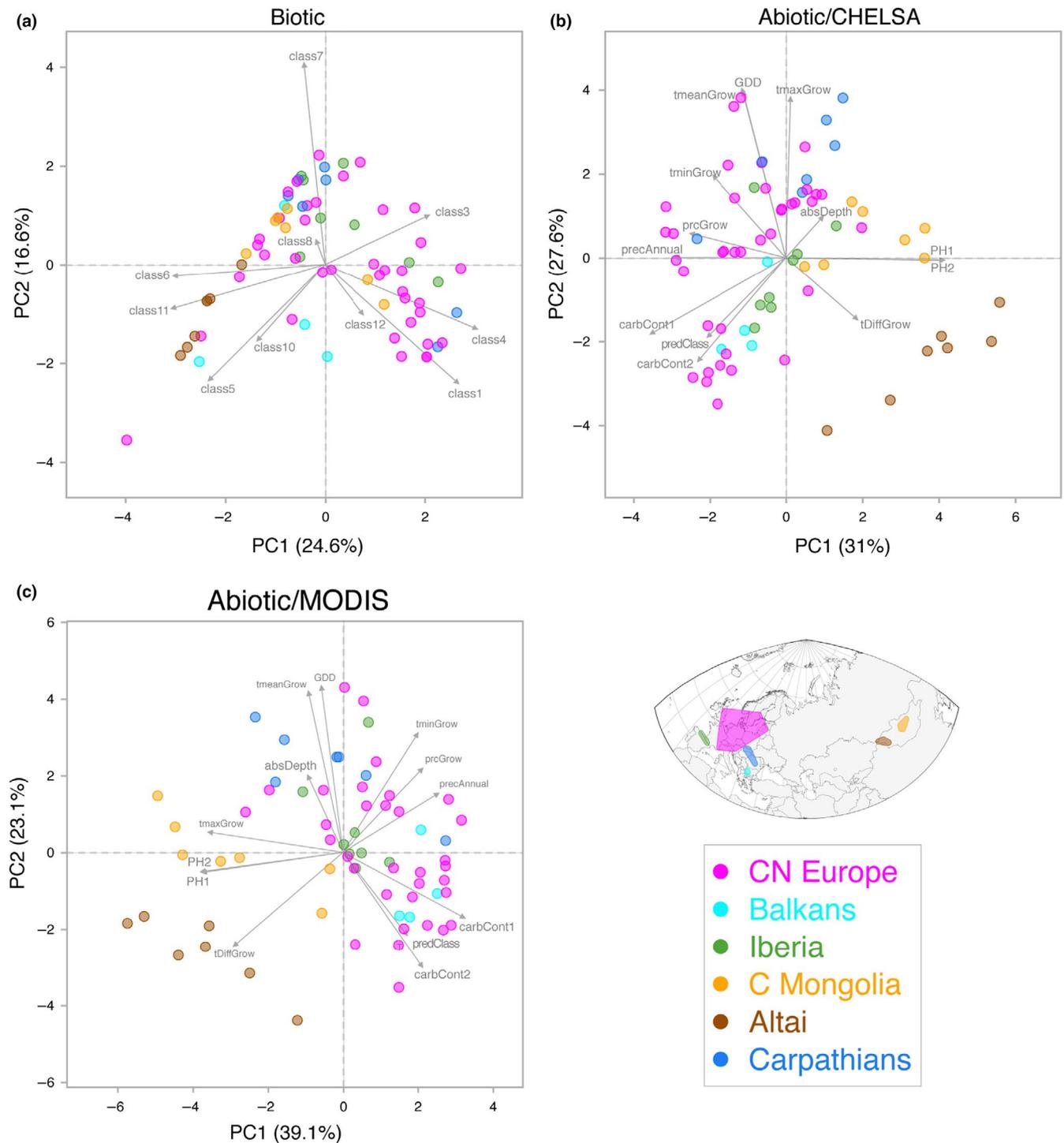


FIGURE 3 The distribution of the sampled populations of *Primula farinosa* in biotic and abiotic multivariate environmental space. (a) Land cover classes: Evergreen/deciduous needleleaf trees (class1), deciduous broadleaf trees (class3), mixed/other trees (class4), shrubs (class5), herbaceous vegetation (class6), cultivated and managed vegetation (class7), and barren (class11). (b) and (c) Climate and soil variables using the CHELSA and MODIS variables, respectively: growing degree days (GDD), minimum temperature of growing season (tminGrow), maximum temperature of growing season (tmaxGrow), average temperature range during growing season (tDiffGrow), average temperature of growing season (tmeanGrow), annual precipitation (precAnnual), precipitation during the growing season (prcGrow), soil pH in H₂O at 0 cm (pH1), and soil pH in H₂O at 5 cm (pH2), absolute depth to bedrock (absDepth), predicted most probable class (predClass), soil organic carbon content at 0 cm (carbCont1), soil organic carbon content at 5 cm (carbCont2)

3.2 | Multivariate biotic and abiotic space and niche similarity

The results of the PCA of the biotic (land use) dataset indicated that the Asian populations, particularly those in the Altai mountains, tend to grow in land mainly covered by shrubs and herbaceous vegetation or land that is mostly barren or snow-covered (Figure 3a). In contrast, the European populations, especially those that belong to the central-northern European group, appear to occupy a wide range of land cover classes, from tree dominated land to land that is mainly cultivated or managed (Figure 3a). In the abiotic PCA space, populations that belong to the same clan/group (i.e., share similar ancestry profiles) appear to grow under similar abiotic conditions (Figure 3b,c). Specifically, in both abiotic datasets, that is CHELSA (global reanalysis temperature data) and MODIS (remotely sensed temperature data), the Altai populations experience much lower temperatures and wider temperature fluctuations during the growing season, followed by the populations in central Mongolia. The Asian populations also appear to receive less precipitation both annually and during the growing season. Moreover, our results indicate that the Asian populations grow in more alkaline (basic) soils compared to the European populations.

Niche overlap (Schoener's *D*) was generally slightly higher in biotic space across all lineage pairwise comparisons, compared to abiotic space (Table 2). Moreover, the majority of the background

similarity tests did not support significant similarities between the available habitats of the six lineages in both biotic and abiotic environments, suggesting ecological divergence in *P. farinosa* (Table 2).

3.3 | Isolation by distance and ancestry/environment associations

The Pearson correlation coefficient between the log-transformed geographic and genetic distances were relatively high among all Eurasian populations ($r = 0.75$, $p < 0.001$; Figure 4), a result probably reflecting the large geographic distance separating the European and Asian populations (Figure 1). This correlation was much weaker when considering only the European populations ($r = 0.46$, $p < 0.001$), suggesting that geographic distance is a weak predictor of the genetic divergence observed among populations.

In general, the results of the multiple regressions indicated significant associations between individual ancestry proportions and environment (i.e., PCA axes), with varying degrees of fit across datasets, number of groups (*K*) and PCA axes (Table 3). The first PCA axis in both abiotic datasets, that is CHELSA and MODIS (mostly representing soil pH, temperature range during the growing season, minimum and maximum temperatures during the growing season, and annual precipitation; Figure 3c), showed the best fit for $K = 5$ (adj. $r^2 = 0.61$), with the linear model providing a slightly better fit (AIC = 240.71 and AIC = 239.43 for MODIS

TABLE 2 Results of niche similarity (Schoener's *D*) analyses and background (niche) similarity tests in three categories of multivariate environmental space for the six lineages of *Primula farinosa* identified in this study

Lineages	Environmental space								
	Biotic			Abiotic/CHELSA			Abiotic/MODIS		
	<i>D</i>	<i>p</i> (1-2)	<i>p</i> (2-1)	<i>D</i>	<i>p</i> (1-2)	<i>p</i> (2-1)	<i>D</i>	<i>p</i> (1-2)	<i>p</i> (2-1)
Altai–Balkans	0.001	0.071	0.281	0	1	1	0	1	1
Altai–Carpathians	0	1	1	0	1	1	0	1	1
Altai–C Mongolia	0.0001	0.546	0.322	0	1	1	0	1	1
Altai–CN Europe	0.0025	0.233	0.557	0	1	1	1.11e–16	0.082	0.357
Altai–Iberia	0	1	1	0	1	1	0	1	1
Balkans–Carpathians	0.0007	0.208	0.906	0	1	1	0	1	1
Balkans–C Mongolia	0.0298	0.455	0.605	0	1	1	0	1	1
Balkans–CN Europe	0.107	0.402	0.351	0.0086	0.118	0.638	0.06	0.106	0.568
Balkans–Iberia	0.0004	0.561	0.855	0.3746	0.01	0.057	0.18	0.231	0.171
Carpathians–C Mongolia	0.175	0.056	0.147	0	1	1	0	1	1
Carpathians–CN Europe	0.1486	0.035	0.08	0.0363	0.236	0.307	0.3415	0.011	0.027
Carpathians–Iberia	0.218	0.06	0.18	0	1	1	0.0007	0.737	0.662
C Mongolia–CN Europe	0.07	0.072	0.231	0.0179	0.232	0.346	0.0586	0.165	0.092
C Mongolia–Iberia	0.2269	0.116	0.176	0	1	1	0.2298	0.082	0.137
CN Europe–Iberia	0.066	0.265	0.036	0.0193	0.682	0.36	0.0637	0.619	0.069

Note: Backgrounds for the tests were defined by applying a 20 km buffer zone around the sampling localities of each lineage. The background similarity tests were performed twice for each lineage pair. The second number in the parentheses after the *p* value indicates the lineage whose background was used for the respective similarity tests. Numbers in bold indicate significant *p* values (<0.05).

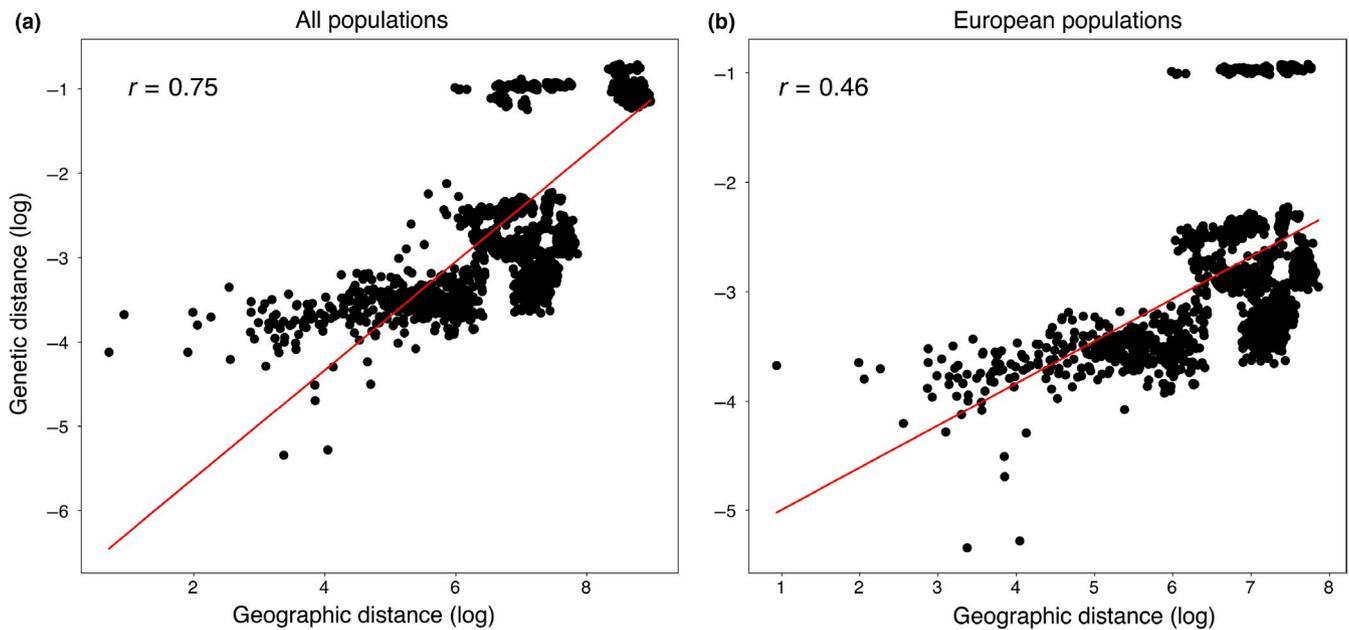


FIGURE 4 Correlations (Pearson's r) between geographic distance (log-transformed km) and genetic distance (log-transformed average nucleotide differences) between all sampled Eurasian populations (a) and only the European populations (b). Significance was assessed using Mantel test ($p < 0.001$ in both cases)

and CHELSA respectively) compared to the quadratic polynomial model (AIC = 241.77 and AIC = 243.46; Table 3). Coefficients of determination were also significant for the second PCA axis of both abiotic datasets (mostly representing soil pH and annual and growing season precipitation) and $K = 5$ (adj. $r^2 = 0.323$ and adj. $r^2 = 0.2945$ for abiotic/MODIS and abiotic/CHELSA respectively). Significant associations were also observed between ancestry proportions and the first PCA axis of the biotic dataset (i.e., land cover classes), with much lower degrees of fit, however, compared to the abiotic datasets (adj. $r^2 = 0.1727$ for $K = 4$ and adj. $r^2 = 0.2747$ for $K = 5$), and significant residual spatial autocorrelation (Moran's $I = 0.32$ for $K = 4$ and Moran's $I = 0.22$ for $K = 5$). Overall, the abiotic/MODIS dataset showed the best linear fit with no significant spatial autocorrelation in residuals across the top explanatory models (PC1 and PC2 for $K = 5$, PC1 for $K = 4$; Table 3).

4 | DISCUSSION

Identification of nature's units and the geographic description of biodiversity variation have been central objectives in conservation biogeography ever since the birth of the field (Whittaker et al., 2005; see also Riddle et al., 2011). However, challenges related to data availability and integration have hindered progress towards these objectives, particularly in taxa with little or uninformative morphological variation. Here, we conducted field work and took advantage of high-throughput sequencing technologies and publicly available environmental data to address the above two objectives in the cold-adapted and threatened plant *P. farinosa*. We identified six major genetic lineages across its Eurasian distribution corresponding to distinct geographic regions and found significant associations between these lineages and

their abiotic environments, indicative of high intraspecific diversity and potential local adaptation to different climatic and soil conditions. Our results highlight the usefulness of genomic and publicly available data in biodiversity research and the need to incorporate intraspecific diversity in global assessments of species conservation status.

According to IUCN (<http://www.iucnredlist.org/details/20339/8/1>) and the ITIS Catalogue of Life (Roskov et al., 2018), *P. farinosa* is widespread in Europe and temperate Asia. However, the species is threatened or extinct in many European countries, mainly because of anthropogenic environmental pressure. While these threats are acknowledged by IUCN, the global status assessment is summarized as follows: "This species demonstrates declines in most of its range, being marked as threatened in several national red lists. It suffers from grazing and lack of management in its grassland habitats. However, the species is still widespread and unlikely to severely decline in the near future towards extinction. More information on the current population size, trend and the overall rate of decline is needed to review whether it would qualify for threatened under Criterion A." From this report, it is clear that by considering *P. farinosa* as a widespread species, local extinctions are overlooked because of its relative abundance globally. The countries where the species has already gone extinct or displays the most severe declines are Hungary, Poland, Slovakia and Romania. Both our current and previous studies (Theodoridis et al., 2018, 2017) show that populations in these countries form a distinct Carpathian lineage (Figure 2), occupying to a large extent an ecological niche distinct from the rest of the Eurasian populations (Table 2; Figure 3b,c). Moreover, both the Carpathian and the Iberian lineages are projected to experience a severe loss of their climatically suitable habitat in the next few decades (Theodoridis et al., 2018). These results demonstrate that the projected range

TABLE 3 Statistical relationships between genetic structure (ancestry; see Figure 2b) of the studied populations and the first three PCA axes in their biotic (land cover classes) and abiotic (climate and soil characteristics) environmental space

Environmental Space	Multiple r^2	Adjusted r^2	AIC linear	AIC polynomial	Moran's I
Biotic					
Land cover					
K = 4					
PC1	0.2082**	0.1727	257.82	252.09	0.32***
PC2	0.0379	-0.0051	244.04	237.13	0.16
PC3	0.0914	0.0507	213.89	216.56	0.16
K = 5					
PC1	0.3161***	0.2747	249.41	255.77	0.22**
PC2	0.1179	0.0643	239.88	237.66	0.1
PC3	0.1289	0.076	212.9	220.55	0.12
Abiotic					
Soil, precipitation, temperature (CHELSA)					
K = 4					
PC1	0.556***	0.5362	251.86	240.63	0.22*
PC2	0.1106*	0.0707	293.09	274.05	0.4***
PC3	0.0599	0.0179	269.47	257.03	0.63***
K = 5					
PC1	0.6377***	0.6157	239.43	243.46	0.08
PC2	0.3348***	0.2945	274.46	277.11	0.23**
PC3	0.0858	0.0304	269.49	259.01	0.64***
Soil, precipitation, temperature (MODIS)					
K = 4					
PC1	0.5922***	0.5725	246.88	239.23	0.14
PC2	0.0131*	0.0889	268.57	250.19	0.29***
PC3	0.009	-0.038	224.38	220.55	0.73***
K = 5					
PC1	0.6397***	0.616	240.71	241.77	0.05
PC2	0.365***	0.323	249.86	252.14	0.08
PC3	0.015	-0.049	226.01	222.5	0.73***

Note: r^2 : coefficient of determination obtained using a linear model and significance level (asterisks); Moran's I : residual spatial autocorrelation for the linear models and significance level (asterisks); Linear and polynomial models were fitted using the results of the fastStructure analysis (i.e., ancestry proportions) for $K = 4$ and $K = 5$.

Abbreviation: AIC, Akaike's information criterion.

* $p \leq 0.05$

** $p \leq 0.01$

*** $p \leq 0.001$

reduction of threatened European lineages, such as the Carpathian and Iberian lineages, would translate into the loss of populations with unique genetic and ecological characteristics, and call for immediate revision in the conservation status of *P. farinosa* in the affected European countries. Our results emphasize the need to incorporate genomic and environmental data in IUCN assessments for securing the protection of unique evolutionary lineages.

Identification of cryptic diversity has traditionally relied on the use of a few loci to estimate evolutionary relationships or genetic distances among populations within nominal species (Fišer et al., 2018). Despite the rapid progress in high-throughput sequencing

technologies in the last decade, studies that harness the power of genomic data remain few (Struck et al., 2018). Moreover, studies that integrate genomic with publicly available environmental data, that is statistically derived and/or remotely sensed data, to identify cryptic diversity are scarce (Dufresnes et al., 2018). In this study, we combined genome-wide polymorphisms (4,089 SNPs) with climate, soil and land cover data to identify cryptic genomic diversity and its environmental correlates within the cold-adapted *P. farinosa*. We identified six major lineages (four previously known, two newly discovered; Figure 2) and significant lineage-environment associations (Table 3). Specifically, we uncovered strong relationships between

individual ancestry proportions and the abiotic environment that the individuals occupy. Moreover, the abiotic dataset that includes the remotely sensed surface temperature (abiotic/MODIS) showed a better linear fit across PCA axes and numbers of ancestral groups compared to the abiotic/CHELSA dataset (Table 3). This result further stresses the superiority of satellite-obtained estimates of climatic variables to ground station statistical interpolations in capturing climatic variation in plant species distributions, particularly for regions with complex topography or sparse meteorological station data (Deblauwe et al., 2016). Although we could not experimentally test whether the observed genomic divergence among the Eurasian populations is neutral or adaptive, our IBD results (Figure 4) and the ancestry/environment associations (Table 3) suggest that geographic isolation alone cannot explain the observed differentiation, particularly among the European populations, and that ecological factors may be driving lineage divergence. These results add further evidence for the role of ecological factors in reinforcing cryptic diversity within ecologically diverse plant species (Liu et al., 2013).

Morphological stasis, the tendency of closely related taxa to retain a high degree of morphological similarity over extended periods, has been suggested as a major mechanism to explain high levels of cryptic genetic diversity within species despite the lack of noticeable intraspecific morphological variation (Bickford et al., 2007; Gould & Eldredge, 1977). Taxa experiencing strong selection on traits for adaptation to specific environments might not show morphological disparity, despite their long temporal and spatial isolation (Davis et al., 2014; Estes & Arnold, 2007). Our results show that niche overlap (Schoener's D) among lineages is generally higher in the biotic environmental space compared to the abiotic space, albeit no significant niche similarities were detected (Table 2). Additionally, we found weak relationships between individual ancestry profiles and their occupied biotic space (adj. $r^2 = 0.1727$ for $K = 4$ and adj. $r^2 = 0.2747$ for $K = 5$; Table 3), indicative of the low variation in biotic conditions, that is land cover types, across lineages, despite the long branches separating them (Figure 2a). Although we used a relatively coarse resolution land cover dataset (~1 km), these results may explain the lack of morphological disparity across the Eurasian lineages, adding further evidence for the role of biotic interactions in constraining the evolution of diagnostic traits (i.e., stabilizing selection) over large spatiotemporal scales (Charlesworth, Lande, & Slatkin, 1982).

While the nature and definition of species boundaries continue to be a matter of debate, recent views have focused on the integration of various criteria, including intrinsic reproductive isolation, morphological diagnosability, monophyly and distinct ecological niches (Freudenstein et al., 2017; De Queiroz, 2007). Our results address the last two criteria by revealing strong ecological and evolutionary differentiation among the six *P. farinosa* lineages, although Nei's standard genetic distances between these lineages (Table 1) fall within the range of values previously reported for intraspecific populations of plant species (Alexander, Poll, Dietz, & Edwards, 2009; Li, Wang, Liu, Zhuang, & Huang, 2018). Whether the identified Eurasian lineages could be assigned to different nominal species or not will depend on additional research on intrinsic reproductive

isolation among them, as well as on phylogenomic analyses including closely related and morphologically distinct species within the highly diverse *Primula* sect. *Aleuritia* Duby clade.

5 | CONCLUSIONS

Identifying discrete biological units within threatened species has important implications for conservation. Conservation guidelines rely mainly on traditional taxonomic characters in defining conservation units or species. However, under this approach, significant cryptic diversity may go undetected and often extinct. Integrative approaches that combine genome-wide data with easily accessible environmental information can provide unprecedented resolution below the species level and significantly help conservation agencies in their assessments. Our study demonstrates the utility of these data in the cold-adapted and threatened *P. farinosa* for uncovering cryptic genetic diversity and its environmental correlates and further underlines the need for bringing together principles from the fields of ecology and evolution in conservation planning.

ACKNOWLEDGEMENTS

We thank Per Toräng, Camille Madec, Mihai Puscas, Borja Jiménez Alfaro González, Nikolai Nikolov, Todor Karakiev, Liene Aunina, Michal Hajek, John Richards, Peter Schönschwetter, Gerald Schneeweiss, Michael Hauck and Martin Kretschmer for valuable information on species localities and help during field surveys and the Univ. of Zurich S3IT (Service and Support for Science IT) for technical support and computational resources (www.s3it.uzh.ch). This work was supported by the Pro-Doc Research Module in Plant Sciences and Policy of the Swiss National Science Foundation (grant number 132471 to EC), was affiliated with the Zurich-Basel Plant Science Center ProDoc PhD Program in Plant Sciences and Policy, and was officially endorsed by DIVERSITAS-bio-GENESIS. Further funding was provided by the Claraz Foundation and the Dept of Systematic and Evolutionary Botany of the Univ. of Zurich. D.N.B. thanks the Danish National Research Foundation for support to the Center for Macroecology, Evolution and Climate (DNRF96).

DATA ACCESSIBILITY

Data available from the Dryad Digital Repository. doi:<https://doi.org/10.5061/dryad.62np6jd>

ORCID

Spyros Theodoridis  <https://orcid.org/0000-0001-5188-7033>

REFERENCES

- Albert, C. H., Yoccoz, N. G., Thomas, C., Edwards, J., Graham, C. H., Zimmermann, N. E., & Thuiller, W. (2010). Sampling in ecology and evolution - Bridging the gap between

- theory and practice. *Ecography*, 33, 1028–1037. <https://doi.org/10.1111/j.1600-0587.2010.06421.x>
- Alexander, J. M., Poll, M., Dietz, H., & Edwards, P. J. (2009). Contrasting patterns of genetic variation and structure in plant invasions of mountains. *Diversity and Distributions*, 15, 502–512. <https://doi.org/10.1111/j.1472-4642.2008.00555.x>
- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11, 697–709. <https://doi.org/10.1038/nrg2844>
- Allendorf, F. W., Luikart, G., & Aitken, S. N. (2013). *Conservation and the genetics of populations* (2nd ed.). Oxford, UK and Hoboken, NJ: Wiley-Blackwell.
- Avise, J. C. (2009). Phylogeography: Retrospect and prospect. *Journal of Biogeography*, 36, 3–15. <https://doi.org/10.1111/j.1365-2699.2008.02032.x>
- Bálint, M., Domisch, S., Engelhardt, C. H. M., Haase, P., Lehrian, S., Sauer, J., ... Nowak, C. (2011). Cryptic biodiversity loss linked to global climate change. *Nature Climate Change*, 1, 313–318. <https://doi.org/10.1038/nclimate1191>
- Benestan, L. M., Ferchaud, A.-L., Hohenlohe, P. A., Garner, B. A., Naylor, G. J. P., Baums, I. B., ... Luikart, G. (2016). Conservation genomics of natural and managed populations: Building a conceptual and practical framework. *Molecular Ecology*, 25, 2967–2977. <https://doi.org/10.1111/mec.13647>
- Bernardo, J. (2011). A critical appraisal of the meaning and diagnosability of cryptic evolutionary diversity, and its implications for conservation in the face of climate change. In T. R. Hodkinson, M. B. Jones, S. Waldren, & J. A. N. Parnell (Eds.), *Climate change, ecology and systematics* (pp. 380–438). Cambridge, UK: Cambridge University Press.
- Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., ... Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22, 148–155. <https://doi.org/10.1016/j.tree.2006.11.004>
- Bostock, M., Ogievetsky, V., & Heer, J. (2011). D3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17, 2301–2309.
- Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., ... Guisan, A. (2011). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, 21, 481–497. <https://doi.org/10.1111/j.1466-8238.2011.00698.x>
- Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., ... Yu, D. W. (2017). Connecting Earth observation to high-throughput biodiversity data. *Nature: Ecology and Evolution*, 1, 0176.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140. <https://doi.org/10.1111/mec.12354>
- Charlesworth, B., Lande, R., & Slatkin, M. (1982). A neo-Darwinian commentary on macroevolution. *Evolution*, 36, 474–498. <https://doi.org/10.1111/j.1558-5646.1982.tb05068.x>
- Cheffings, C. M., Farrell, L., Dines, T. D., Jones, R. A., Leach, S. J., McKean, D. R., ... Taylor, I. (2005). *The vascular plant red data list for Great Britain. Species status*. Peterborough, UK: Joint Nature Conservation Committee (pp. 1–116), 7.
- Coldea, G., Stoica, I. A., Pușcaș, M., Ursu, T., Oprea, A., & The IntraBioDiv Consortium (2009). Alpine-subalpine species richness of the Romanian Carpathians and the current conservation status of rare species. *Biodiversity Conservation*, 18, 1441–1458.
- Davis, C. C., Schaefer, H., Xi, Z., Baum, D. A., Donoghue, M. J., & Harmon, L. J. (2014). Long-term morphological stasis maintained by a plant-pollinator mutualism. *Proceedings of the National Academy of Sciences*, 111, 5914–5919. <https://doi.org/10.1073/pnas.1403157111>
- De Guia, A. P. O., & Saitoh, T. (2007). The gap between the concept and definitions in the Evolutionarily Significant Unit: The need to integrate neutral genetic variation and adaptive variation. *Ecological Research*, 22, 604–612. <https://doi.org/10.1007/s11284-006-0059-z>
- De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, 56, 879–886. <https://doi.org/10.1080/10635150701701083>
- Deblauwe, V., Droissart, V., Bose, R., Sonké, B., Blach-Overgaard, A., Svenning, J.-C., ... Couvreur, T. L. P. (2016). Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Global Ecology and Biogeography*, 25, 443–454. <https://doi.org/10.1111/geb.12426>
- Des Roches, S., Post, D. M., Turley, N. E., Bailey, J. K., Hendry, A. P., Kinnison, M. T., ... Palkovacs, E. P. (2018). The ecological importance of intraspecific variation. *Nature Ecology & Evolution*, 2, 57–64. <https://doi.org/10.1038/s41559-017-0402-5>
- Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., ... Guisan, A. (2017). ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, 40, 774–787. <https://doi.org/10.1111/ecog.02671>
- Didukh, Y. P. (2009). *Red data book of Ukraine. Vegetable Kingdom*. Kyiv, Ukraine: Globalconsulting.
- Doyle, J. J., & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissues. *Focus*, 12, 13–15.
- Dufresnes, C., Mazepa, G., Rodrigues, N., Brelsford, A., Litvinchuk, S. N., Sermier, R., ... Jeffries, D. L. (2018). Genomic evidence for cryptic speciation in tree frogs from the Apennine Peninsula, with description of *Hyla perrini* sp. nov. *Frontiers in Ecology & Evolution*, 6, 44. <https://doi.org/10.3389/fevo.2018.00144>
- Espíndola, A., Ruffley, M., Smith, M. L., Carstens, B. C., Tank, D. C., & Sullivan, J. (2016). Identifying cryptic diversity with predictive phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, 283, 20161529.
- Estes, S., & Arnold, S. J. (2007). Resolving the paradox of stasis: Models with stabilizing selection explain evolutionary divergence on all timescales. *The American Naturalist*, 169, 227–244. <https://doi.org/10.1086/510633>
- Faith, D. P., Magallón, S., Hendry, A. P., Conti, E., Yahara, T., & Donoghue, M. J. (2010). Ecosystem services: An evolutionary perspective on the links between biodiversity and human-well-being. *Current Opinion in Environmental Sustainability*, 2, 66–74. <https://doi.org/10.1016/j.cosust.2010.04.002>
- Feckler, A., Zubrod, J. P., Thielsch, A., Schwenk, K., Schulz, R., & Bundschuh, M. (2014). Cryptic species diversity: An overlooked factor in environmental management? *Journal of Applied Ecology*, 51, 958–967. <https://doi.org/10.1111/1365-2664.12246>
- Fedorov, A. (2001). *Flora of Russia: The European part & bordering regions* (Vol. 5). Bosa Roca: CRC Press.
- Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, 27, 613–635. <https://doi.org/10.1111/mec.14486>
- Franklin, J., Serra-Diaz, J. M., Syphard, A. D., & Regan, H. M. (2017). Big data for forecasting the impacts of global change on plant communities. *Global Ecology and Biogeography*, 26, 6–17. <https://doi.org/10.1111/geb.12501>
- Freudenstein, J. V., Broe, M. B., Folk, R. A., & Sinn, B. T. (2017). Biodiversity and the species concept-lineages are not enough. *Systematic Biology*, 66, 644–656.
- Funk, W. C., McKay, J. C., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, 27, 489–495. <https://doi.org/10.1016/j.tree.2012.05.012>
- Gajewski, Z., Sitek, E., Stolarczyk, P., Nowak, B., & Kapała, K. (2013). A current status of the population of *Primula farinosa* L. (Primulaceae) at the only one known site in Poland. *Polish Journal of Ecology*, 61, 797–804.

- Gotelli, N. J., & Stanton-Geddes, J. (2015). Climate change, genetic markers and species distribution modelling. *Journal of Biogeography*, 42, 1577–1585. <https://doi.org/10.1111/jbi.12562>
- Gould, S. J., & Eldredge, N. (1977). Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology*, 3, 115–151. <https://doi.org/10.1017/S0094837300005224>
- Hambler, D. J., & Dixon, J. M. (2003). *Primula farinosa* L. *Journal of Ecology*, 91, 694–705. <https://doi.org/10.1046/j.1365-2745.2003.00797.x>
- He, K. S., Bradley, B. A., Cord, A. F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., ... Pettorelli, N. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation*, 1, 4–18. <https://doi.org/10.1002/rse2.7>
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., ... Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, 12, e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Jenkins, D. A., Yannic, G., Schaefer, J. A., Conolly, J., & Lecomte, N. (2018). Population structure of caribou in an ice-bound archipelago. *Diversity and Distributions*, <https://doi.org/10.1111/ddi.12748>
- Joly, S., Bryant, D., & Lockhart, P. J. (2015). Flexible methods for estimating genetic distances from single nucleotide polymorphisms. *Methods in Ecology and Evolution*, 6, 938–948. <https://doi.org/10.1111/2041-210X.12343>
- Karger, D. N., Conrad, O., Böhrer, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., ... Kessler, M. (2017). CHELSA climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122. <https://doi.org/10.1038/sdata.2017.122>
- Keith, D. A., Rodríguez, J. P., Brooks, T. M., Burgman, M. A., Barrow, E. G., Bland, L., ... Spalding, M. D. (2015). The IUCN Red List of Ecosystems: Motivations, challenges, and applications. *Conservation Letters*, 8, 214–226. <https://doi.org/10.1111/conl.12167>
- Krasnoborov, I. M. (2000). *Flora of Siberia*. Enfield, NH: Science Publishers.
- La Salle, J., Williams, K. J., & Moritz, C. (2016). Biodiversity analysis in the digital era. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20150337.
- Lausch, A., Bannehr, L., Beckmann, M., Boehm, C., Feilhauer, H., Hacker, J. M., ... Cord, A. F. (2016). Linking Earth Observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. *Ecological Indicators*, 70, 317–339. <https://doi.org/10.1016/j.ecolind.2016.06.022>
- Leaché, A. D., Banbury, B. L., Felsenstein, J., Nieto-Montes de Oca, A., & Stamatakis, A. (2015). Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, 64, 1032–1047. <https://doi.org/10.1093/sysbio/syv053>
- Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44, W242–W245. <https://doi.org/10.1093/nar/gkw290>
- Li, B. J., Wang, J. Y., Liu, Z. J., Zhuang, X. Y., & Huang, J. X. (2018). Genetic diversity and ex situ conservation of *Loropetalum subcordatum*, an endangered species endemic to China. *BMC Genetics*, 19, 12. <https://doi.org/10.1186/s12863-018-0599-6>
- Liu, J., Moller, M., Provan, J., Gao, L.-M., Poudel, R. M., & Li, D.-Z. (2013). Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist*, 199, 1093–1108. <https://doi.org/10.1111/nph.12336>
- Lunter, G., & Goodson, M. (2011). Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, 21, 936–939. <https://doi.org/10.1101/gr.111120.110>
- Minoche, A. E., Dohm, J. C., & Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biology*, 12, R112. <https://doi.org/10.1186/gb-2011-12-11-r112>
- Miraldo, A., Li, S., Borregaard, M. K., Florez-Rodriguez, A., Gopalakrishnan, S., Rizvanovic, M., ... Nogues-Bravo, D. (2016). An Anthropocene map of genetic diversity. *Science*, 353, 1532–1535. <https://doi.org/10.1126/science.aaf4381>
- Moritz, C. (1994). Defining 'evolutionarily significant units' for conservation. *Trends in Ecology and Evolution*, 9, 373–375. [https://doi.org/10.1016/0169-5347\(94\)90057-4](https://doi.org/10.1016/0169-5347(94)90057-4)
- Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, 5, 238–254. <https://doi.org/10.1080/10635150252899752>
- Nei, M. (1972). Genetic distance between populations. *American Naturalist*, 106, 283–292. <https://doi.org/10.1086/282771>
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32, 268–274. <https://doi.org/10.1093/molbev/msu300>
- Paaby, A. B., & Rockman, M. V. (2014). Cryptic genetic variation: Evolution's hidden substrate. *Nature Reviews Genetics*, 15, 247–258. <https://doi.org/10.1038/nrg3688>
- Palsbøll, P. J., Bérubé, M., & Allendorf, F. W. (2006). Identification of management units using population genetic data. *Trends in Ecology & Evolution*, 22, 11–16. <https://doi.org/10.1016/j.tree.2006.09.003>
- Pauls, S. U., Nowak, C., Balint, M., & Pfenninger, M. (2013). The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, 22, 925–946. <https://doi.org/10.1111/mec.12152>
- R Development Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197, 573–589. <https://doi.org/10.1534/genet.ics.114.164350>
- Richards, J. (2003). *Primula*. London, UK: BT Batsford Ltd.
- Riddle, B. R., Ladle, R. J., Lourie, S. A., & Whittaker, R. J. (2011). Basic biogeography: Estimating biodiversity and mapping nature. In R. J. Ladle, & R. J. Whittaker (Eds.), *Conservation biogeography* (pp. 45–92). Chichester, UK: John Wiley & Sons Ltd.
- Roskov, Y., Abucay, L., Orrell, T., Nicolson, D., Bailly, N., Kirk, P. M., & Penev, L. (2018). *Species 2000 & ITIS catalogue of Life, 2018 annual checklist*. Retrieved from www.catalogueoflife.org/annual-checklist/2018. Species 2000: Naturalis, Leiden, the Netherlands. ISSN 2405–884X.
- Ryder, O. A. (1986). Species conservation and systematics: The dilemma of subspecies. *Trends in Ecology and Evolution*, 1, 9–10. [https://doi.org/10.1016/0169-5347\(86\)90059-5](https://doi.org/10.1016/0169-5347(86)90059-5)
- Salamon-Albert, É., & Morschhauser, T. (2003). Habitat preferences of a unique specialist plant species (*Primula farinosa* subsp. *alpigena*) in Hungary. *Acta Botanica Hungarica*, 45, 193–215. <https://doi.org/10.1556/ABot.45.2003.1-2.17>
- Schoener, T. (1970). Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology*, 51, 408–418. <https://doi.org/10.2307/1935376>
- Shishkin, B. K., & Bobrov, E. G. (1952). *Flora of the USSR* (Vol. XVIII). Metachlamydeae. Translated (1967) from Russian. Jerusalem: Israel Program for Scientific Translation.
- Sørensen, S. B., Larsen, B., Orabi, J., & ørgaard, M. (2014). *Primula farinosa* in Denmark; genetic diversity and population management. *Nordic Journal of Botany*, 8, 1–8.

- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., ... Dimitrov, D. (2018). Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, *33*, 153–163. <https://doi.org/10.1016/j.tree.2017.11.007>
- Sullivan, J., Smith, M. L., Espíndola, A., Ruffley, M., Rankin, A., Tank, D., & Carstens, B. (2019). Integrating life history traits into predictive phylogeography. *Molecular Ecology*, <https://doi.org/10.1111/mec.15029>
- Theodoridis, S., Patsiou, T. S., Randin, C., & Conti, E. (2018). Forecasting range shifts of a cold-adapted species under climate change: Are genomic and ecological diversity within species crucial for future resilience? *Ecography*, *41*, 1357–1369. <https://doi.org/10.1111/ecog.03346>
- Theodoridis, S., Randin, C., Broennimann, O., Patsiou, T., & Conti, E. (2013). Divergent and narrower climatic niches characterize polyploid species of European primroses in *Primula* sect. *Aleuritia*. *Journal of Biogeography*, *40*, 1278–1289.
- Theodoridis, S., Randin, C., Szövényi, P., Boucher, F. C., Patsiou, T. S., & Conti, E. (2017). How do cold-adapted plants respond to climatic cycles? Interglacial expansion explains current distribution and genomic diversity in *Primula farinosa* L. *Systematic Biology*, *66*, 715–736. <https://doi.org/10.1093/sysbio/syw114>
- Topić, J., & Stančić, Z. (2006). Extinction of fen and bog plants and their habitats in Croatia. *Biodiversity Conservation*, *15*, 3371–3381. <https://doi.org/10.1007/s10531-005-4874-2>
- Tuanmu, M. N., & Jetz, W. (2014). A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, *23*, 1031–1045. <https://doi.org/10.1111/geb.12182>
- Turis, P., Kliment, J., Feráková, V., Dítě, D., Eliášjun, P., Hrivnák, R., ... Bernátová, D. (2014). Red list of vascular plants of the Carpathian part of Slovakia. *Thaiszia - Journal of Botany*, *24*, 35–87.
- Waldron, A., Miller, D. C., Redding, D., Mooers, A., Kuhn, T. S., Nibbelink, N., ... Gittleman, J. L. (2017). Reductions in global biodiversity loss predicted from conservation spending. *Nature*, *551*, 364–367. <https://doi.org/10.1038/nature24295>
- Wan, Z., & Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, *34*, 892–905.
- Wan, Z., Hook, S., & Hulley, G. (2015). MOD11A1 MODIS/terra land surface temperature/emissivity daily L3 global 1 km SIN grid V006 [Data set]. NASA EOSDIS LP DAAC. <https://doi.org/10.5067/MODIS/MOD11A1.006>
- Wan, Z., Hook, S., & Hulley, G. (2015). MYD11A1 MODIS/aqua land surface temperature/emissivity daily L3 global 1 km SIN grid V006 [Data set]. NASA EOSDIS LP DAAC. <https://doi.org/10.5067/MODIS/MYD11A1.006>
- Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, *62*, 868–883. <https://doi.org/10.1111/j.1558-5646.2008.00482.x>
- Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M., & Willis, K. J. (2005). Conservation biogeography: Assessment and prospect. *Diversity and Distributions*, *11*, 3–23. <https://doi.org/10.1111/j.1366-9516.2005.00143.x>
- Wilkinson, M., Mclnerney, J. O., Hirt, R. P., Foster, P. G., & Embley, T. M. (2007). Of clades and clans: Terms for phylogenetic relationships in unrooted trees. *Trends in Ecology and Evolution*, *22*, 114–115. <https://doi.org/10.1016/j.tree.2007.01.002>
- Willing, A., Courtiol, A., Christiansen, P., Niedballa, J., Scharf, A. K., Orlando, L., ... Kitchener, A. C. (2015). Planning tiger recovery: Understanding intraspecific variation for effective conservation. *Science Advances*, *1*, 1–13. <https://doi.org/10.1126/sciadv.1400175>
- Yamasaki, E., Altermatt, F., Cavender-Bares, J., Schuman, M. C., Zuppinger-Dingley, D., Garonna, I., ... Shimizu, K. K. (2017). Genomics meets remote sensing in global change studies: Monitoring and predicting phenology, evolution and biodiversity. *Current Opinion in Environmental Sustainability*, *29*, 177–186. <https://doi.org/10.1016/j.cosust.2018.03.005>
- Yannic, G., Ortego, J., Pellissier, L., Lecomte, N., Bernatchez, L., & Côté, S. D. (2017). Linking genetic and ecological differentiation in an ungulate with a circumpolar distribution. *Ecography*, <https://doi.org/10.1111/ecog.02995>

BIOSKETCHES

Spyros Theodoridis is a biogeographer with a strong focus on patterns and processes of intraspecific biodiversity as well as on conservation biogeography.

David Nogués-Bravo is a macroecologist working on the effects of past environmental change on species distributions and genetic diversity.

Elena Conti is a plant evolutionary biologist particularly interested in the evolution of speciation mechanisms and reproductive barriers.

Author contributions: S.T. and E.C. designed the study; S.T. performed the analyses; S.T., D.N.B., and E.C. interpreted the results; and all authors contributed to writing the manuscript.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Theodoridis S, Nogués-Bravo D, Conti E. The role of cryptic diversity and its environmental correlates in global conservation status assessments: Insights from the threatened bird's-eye primrose (*Primula farinosa* L.). *Divers Distrib*. 2019;25:1457–1471. <https://doi.org/10.1111/ddi.12953>