

Congruent evolutionary responses of European steppe biota to late Quaternary climate change: insights from convolutional neural network based demographic modeling

Philipp Kirschner (✉ philipp.kirschner@gmail.com)

University of Innsbruck <https://orcid.org/0000-0002-9263-5112>

Manolo Perez

Universidade Federal de Sao Carlos <https://orcid.org/0000-0002-4642-7793>

Eliška Závěská

Institute of Botany

Isabel Sanmartin

Real Jardin Botánico-CSIC <https://orcid.org/0000-0001-6104-9658>

Laurent Marquer

University of Innsbruck

Birgit Schlick-Steiner

University of Innsbruck

Nadir Alvarez

Geneva Natural History Museum of Geneva

Florian Steinere

University of Innsbruck

Peter Schönswetter

University of Innsbruck

Article

Keywords: Recurrent Glaciations, Vegetation Shifts, Range Alterations, Approximate Bayesian Computation, Genetic Variance

Posted Date: June 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-534143/v1>

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

[Read Full License](#)

Abstract

Quaternary climatic oscillations had a large impact on European biogeography. Alternation of cold and warm stages caused recurrent glaciations, massive vegetation shifts and large-scale range alterations in many species. The Eurasian steppe biome and its grasslands are a noteworthy example; they underwent climate-driven, large-scale contractions during warm stages and expansions during cold stages. Here, we evaluate the impact of these range alterations on the late Quaternary demography of phylogenetically unrelated plant and insect species, typical of the Eurasian steppes. We contrast three explicit demographic hypotheses by applying a novel approach combining Convolutional Neural Networks with Approximate Bayesian Computation. We identified congruent demographic responses of cold stage expansions and warm stage contractions across all species, but also species-specific effects. The demographic history of Eurasian steppe biota reflects major paleoecological turning points of the late Quaternary, and emphasizes the role of the climate as a driving force behind patterns of genetic variance.

Introduction

The recurrent alternation of cold (glacial periods) and warm stages (interglacial periods) during the Quaternary (the last 2.6 million years, myr) was paramount in shaping present day species distribution patterns in Europe. Transitions were marked by large fluctuations of temperature and precipitation occurring within millennia^{1,2} and fueled extensive range expansions and contractions in many biota³. Phylogeography has contributed significantly to our knowledge about the impact of these climatic fluctuations on the European flora and fauna^{4,5}. Pleistocene sea-level changes, glacier advances and retreats, as well as the complex topography of Europe – with high mountain chains such as the Alps and the Pyrenees acting as major dispersal barriers – seemingly led to large-scale extinction within some groups (e.g., the ‘Tertiary’ tree flora⁶), while promoting the formation of novel evolutionary entities in other groups via recurrent isolation^{7–9}.

The Eurasian steppe is a biome that has undergone massive climate-driven contractions and expansions in the Quaternary. Today, it extends over several thousand kilometers, from the northern coast of the Black Sea in Ukraine in the west throughout Central Asia to northwestern China in the east¹⁰. Low annual precipitation is the decisive factor preventing the formation of closed forests, which renders steppe grasslands the ‘zonal’ (i.e. macroclimatically-induced) vegetation in these areas¹¹. During the cold stages of the Pleistocene, such as the Last Glacial Period (LGP), 115 to 12 kya, the Eurasian steppe had a much larger extent, compared with interglacial periods^{12–14}, and repeatedly expanded into large parts of present-day, forest-covered temperate Europe^{15–17} (Figure 1). In present-day temperate Europe, isolated patches of steppe vegetation, the so-called ‘extrazonal steppes’, occur apart from the zonal steppes, resembling ‘steppe islands in a sea of (potential) forest’¹⁸. These extrazonal steppes occur wherever local factors, such as southern exposition and shallow soil cover, act in concert with a continental climate to prevent forest growth¹² (Figure 1).

Extrazonal steppes have traditionally been considered remnants of an extensive, continuous cold stage steppe belt¹⁴ (Figure 1B), which became isolated from each other and from the zonal steppe due to postglacial forest expansion at the start of the Holocene, 11,700 years ago¹². However, Kirschner et al.¹⁹ recently reported that the isolation of the extrazonal steppe biota is in fact much older; vicariant separation occurred as early as the mid-Pleistocene, c. 1.4 mya. Range contractions triggered by forest expansion recurrently forced Eurasian steppe biota into disjunct and – compared with their extent during glacial stages – small-sized interglacial refugia (i.e. the present-day extrazonal steppes), since the very onset of the Pleistocene.

Climate fluctuations have likely driven large demographic transitions across time in both extrazonal and zonal steppes. Yet, the genetic signatures resulting from these processes are not fully understood. An obvious demographic scenario may be that the climate-driven range expansions led to demographic expansion in both extrazonal and zonal steppe populations during the LGP, followed by a demographic contraction in the Holocene (scenario 1, 'Parallel expansion' in Figure 1A). This first scenario captures classical hypotheses of steppe expansion in Europe during the LGP^{12,14}. Under an alternative scenario, demographic expansion took place only in zonal steppe populations but not in the extrazonal ones (scenario 2 'Zonal expansion only' in Figure 1A). In this case, the mountain barriers surrounding many extrazonal steppes, as well as their isolation due to forest spread during warm stages (Figure 1B), would have prevented demographic expansion of extrazonal steppe lineages. A third scenario implicates absence of demographic expansion in both zonal and extrazonal steppe lineages, as a result of slow range shifts in the zonal steppe during the LGP (scenario 3, 'No expansion' in Figure 1A).

Model-based statistical approaches allow a comparison of alternative demographic scenarios in terms of their fit to the data, and the inference of relevant parameters to explain patterns of genetic variation across geographic space, while incorporating the uncertainty in parameter estimation²⁰. One of the most popular approaches is Approximate Bayesian Computation (ABC), a flexible likelihood-free statistical framework based on simulations²¹. The ABC framework allows researchers to incorporate a-priori information about relevant parameters that are used to simulate genetic datasets under alternative demographic scenarios. The simulated data are then compared with the empirically observed data, using genetic summary statistics to discriminate among scenarios²². Recently, machine-learning approaches such as Convolutional Neural Networks (CNN) have emerged as an alternative to ABC methods²³. CNNs can recover information directly from raw genetic datasets by converting them into images, thus overcoming the necessity to select a particular set of statistics to reduce the high dimensionality in the genetic data that affects traditional ABC methods²⁴. Some recent studies have suggested that improved accuracy can be achieved by combining these two methods, that is, by using machine-learning CNN predictions as an input to perform ABC parameter estimation²⁵.

In this study, we applied a statistical modeling approach based on a coupled CNN and ABC framework to five Eurasian steppe species, three insects and two angiosperms. Inferences were based on genomic sequence data obtained via restriction-site associated DNA sequencing (RADseq) by Kirschner et al.¹⁹. Within each of these species, geographic isolation of two genetic groups, an 'extrazonal lineage' and a

‘zonal lineage’, reflecting their main distribution in either zonal or extrazonal steppes, was demonstrated¹⁹. Using pairwise comparisons of zonal and extrazonal lineages within this phylogenetically diverse array of species, we aimed to test which of the three scenarios outlined above (‘Parallel expansion’, ‘Zonal expansion only’, ‘No expansion’) shows a better fit to the late Quaternary population-size dynamics of European steppe biota. We then estimated relevant demographic parameters for the selected scenarios, such as effective population sizes, divergence times, migration rates and timing of expansion/contraction events. Finally, we evaluated congruence in demographic responses across species, using independent palynological and paleoclimatic evidence, as well as hindcasted distribution models reflecting the climatic niche of each species’ extrazonal and zonal lineages during the Last Glacial Maximum (LGM; ca. 21,000 y ago).

Results

Sampling

We sampled 48 and 92 populations of two plant species belonging to different angiosperm families (the spurge *Euphorbia seguieriana* from Euphorbiaceae and the grass *Stipa capillata* from Poaceae) and 56, 37 and 60 populations of three arthropod species from different insect orders (the grasshoppers *Omocestus petraeus* and *Stenobothrus nigromaculatus* and the ant *Plagiolepis taurica*). Sampling was carried out over the years 2013–2016 in mainly the western parts of the Eurasian steppes (Figure 2A, Supplementary Data 1). The sampled species are all typical elements of the Eurasian steppe biome, and represent different reproductive, life-history, and dispersal strategies. Collecting permits are given in Kirschner et al.¹⁹.

Clustering analyses

Population grouping into two clusters, corresponding to extrazonal and zonal lineages, was the optimal solution for all investigated species based on Bayesian population assignment (Figure 2A). Though Kirschner et al.¹⁹ found further subgrouping within these clusters, the focus of our study are the contrasting demographic dynamics between extrazonal and zonal steppe lineages, so we constrained all analyses to the clusters at the highest hierarchical level, that is, $K=2$ for all species. Numbers of individuals and loci analysed are given in Supplementary Table 1.

Geographic distribution of, and degree of admixture between, the extrazonal and zonal lineages were found to be species-specific (Figure 2A). The highest level of admixture was found in populations north of the Alps (*E. seguieriana*, *S. capillata*) and in the Pannonian basin (*E. seguieriana*, *P. taurica*). In addition, a few admixed populations were found in the Western Alps (*E. seguieriana*, *P. taurica*, *S. nigromaculatus*, *S. capillata*). Populations from the Apennines show evidence of admixture in two species, *P. taurica* and *S. capillata*; for the latter, an amphi-Adriatic disjunction was found within the zonal lineage, which was not observed in any other species.

Divergence time estimation

Estimates of τ ($\tau = 2\mu t$; μ = mutation rate per site per generation, t = divergence time) were generally consistent among the differently sized alignments (Supplementary Figure 1). Initial divergence between extrazonal and zonal lineages was estimated to have occurred in or after the mid-Pleistocene across all species analysed. These estimations of divergence times were based on τ inferred from the largest subset (500 SNPs). Estimates of absolute divergence times and highest posterior density credibility intervals (HPD) were 0.59 mya (95% HPD: 0.34–0.86 mya) for *E. seguieriana*, 1.39 mya (95% HPD: 1.11–1.7 mya) for *O. petraeus*, 1.07 mya (95% HPD: 0.60–1.60 mya) for *P. taurica*, 0.38 mya (95% HPD: 0.29–0.46 mya) for *S. nigromaculatus*, and 0.8 mya (95% HPD: 0.46–1.12 mya) for *S. capillata*.

Explorative demographic analyses

For most species, stairway plots suggested stable effective population sizes in both extrazonal and zonal lineages during the last 100 ky, followed by a decline of population size between 10 kya and 20 kya which marks the end of the LGP (Supplementary Figure 2). Deviations from this pattern are observed in *E. seguieriana*, for which a stable population size through time was inferred. Population size increases at around 100 kya were found in the zonal and extrazonal lineages of *E. seguieriana*, and in the extrazonal lineage of *P. taurica* and *O. petraeus*; a similar pattern, but somewhat earlier, was observed in *S. nigromaculatus* (Supplementary Figure 2). This result is biologically plausible and concurs with the onset of the LGP. We refrained from interpreting more ancient population size changes, as artificial signals may occur near the method's lower inference limit^{26,27}.

CNN based demographic modeling

Our combined CNN-ABC approach for selecting the best-fit demographic scenario resulted in high overall model accuracy for all study species. The cross-validation procedure, using a test set of simulations not evaluated during the training step, gave a percentage of correct assignment to the simulated scenario larger than 60% in *P. taurica* and than 80% in all other species (Supplementary Table 2). Our training strategy proved also to be effective for avoiding a loss of accuracy in SNP datasets that contain large amounts of missing data (e.g., in *S. nigromaculatus* and *O. petraeus*). The scenario depicting 'Parallel expansion' in zonal and extrazonal lineages was selected as the most explanatory demographic model (with a posterior probability (PP) higher than 0.87), in the angiosperm species *S. capillata* and the two grasshopper species *S. nigromaculatus* and *O. petraeus*. For the other species, *P. taurica* and *E. seguieriana*, the 'Zonal expansion only' scenario was selected as the best model, with a PP value higher than 0.98. The 'No expansion' scenario had the lowest PP value across all species analysed (Supplementary Table 3). Parameter estimation with CNN-ABC estimated large population sizes for *E. seguieriana*, *O. petraeus*, and *P. taurica*, while *S. capillata* had the smallest values. We also inferred lower contemporary population sizes for the extrazonal compared with the zonal lineages across all analysed species. Migration rates between zonal and extrazonal lineages within each time period were species-specific, with higher values during the LGP observed in *E. seguieriana* and *O. petraeus*, higher pre-LGP values estimated in *S. capillata*, and similar values for the two periods in the remaining species, *P. taurica*

and *S. nigromaculatus* (Figure 2). We did not infer common patterns of migration asymmetry between extrazonal and zonal lineages across the analysed species.

Distribution models for extrazonal and zonal lineages

For the zonal lineages of each analysed species, lineage distribution models (we use this term instead of 'lineage range model' to enhance readability) suggested the existence of continuous distribution ranges, extending from the Pontic plains north of the Black Sea to the Pannonian Basin east of the Alps during late Quaternary cold stages (Figure 2B). In *S. capillata* and *S. nigromaculatus*, geographic ranges reached further west along the northern margin of the Alps, into Germany and France. Extensive suitable areas south of the Alps were found only for *O. petraeus* and *P. taurica*. Towards the West, these ranges did not reach further than northeasternmost Italy.

Lineage distribution models for the extrazonal lineages of each species supported continuous ranges south of the Alps for all studied species. Large continuous ranges north of the Alps were modeled for *P. taurica*, and to a lesser extent for *E. seguieriana* and *O. petraeus*. Small potential ranges along the northern margin of the Alps were also inferred for *S. capillata*. Gaps in habitat suitability were found mainly south of the Alps (Figure 2B). In this area, range overlap of extrazonal and zonal lineages was observed only for *P. taurica* and *S. nigromaculatus*. In contrast, north of the Alps, ranges of extrazonal and zonal lineages overlapped in all species except in *O. petraeus*.

Discussion

The classic hypothesis on the Quaternary range dynamics of European steppes assumed an opposite response compared with the range dynamics observed in the well-investigated European temperate forest biota²⁸; that is, a climate-driven interplay of warm stage (including the Holocene) range contractions and cold stage range expansions¹³. Here, we demonstrate that demographic responses in five ecologically similar, but phylogenetically unrelated, steppe species are in line with this hypothesis and seem to be largely climate-driven. However, demographic patterns were not strictly congruent across species, at least for the extrazonal lineages. Whereas large-scale expansions in both extrazonal and zonal lineages during late Quaternary cold stages ('Parallel expansion', Figure 1) were supported in three species (*O. petraeus*, *S. nigromaculatus*, *S. capillata*; Figure 3), a scenario without population expansion in the extrazonal lineage ('Zonal expansion only') was inferred in the other two species (*E. seguieriana*, *P. taurica*; Figure 3).

The congruent signal of demographic expansion observed across all zonal steppe lineages and study species (Figure 3) agrees well with the hindcasted lineage distribution models during Quaternary cold stages, and the pattern of climate-driven expansion of Eurasian steppes supported by palynological and paleoclimatic data (Figure 2). Conversely, the pattern of no demographic expansion during cold stages exhibited by the extrazonal lineages of *E. seguieriana* and *P. taurica* (the 'Zonal expansion only' scenario in Figure 1) seems counter-intuitive, given the large-scale availability of climatically suitable habitat during the Last Glacial Period in the hindcasted models (Figure 2B). Smaller population sizes and stronger substructure of source populations previous to expansion, as well as the presence of mountain

barriers preventing effective dispersal^{29–32} are likely key factors that hampered range expansion-induced increase in effective population size in the extrazonal lineages, but less so in the zonal lineages.

In addition, species-specific factors such as dispersal ability are known to affect the demographic response of a population to range expansion³⁰. The two grasshopper species *O. petraeus* and *S. nigromaculatus* and the epizoochorous graminoid species *S. capillata*, all considered effective dispersers followed the 'Parallel expansion' scenario. In contrast, *Euphorbia seguieriana* and *Plagiolepis taurica*, which followed a 'Zonal expansion only' scenario, exhibit a more limited dispersal ability, which might be related to large seed size and myrmecochory in the plant species³³, or to small body size in the ant species³⁴. Thus, a species' capacity for long-range dispersal is likely to have played a role in the observed pattern of disconnected extrazonal steppes, but less so in the more continuous zonal steppe ranges.

Our results support a timeline for the demographic history of genetic lineages in European steppe species during the late Quaternary that was roughly congruent across all study species (Figure 3). For *O. petraeus* and *P. taurica*, we estimated divergence times for the initial split between the extrazonal and zonal lineages within each species that fall within the 95% HPD ranges estimated, based on dated mitochondrial DNA (mtDNA) phylogenies¹⁹; our RADseq-based mean age estimates were on average younger, likely reflecting incomplete lineage sorting during initial divergence and/or inaccuracy in the implemented clock rate prior.

Calibration of genetic divergences to absolute time scales suggests that the timing of events in our demographic scenarios is related to periods that are considered climate turning points based on palynological evidence. We demonstrate this for three specific time horizons. Initial divergence between extrazonal and zonal lineages was estimated to have occurred between 0.37 and 1.39 mya by the Bayesian multispecies coalescent model implemented in BPP³⁵, and between 0.9 and 1.6 mya by CNN modeling (Supplementary Table 4, Figure 3). These estimates roughly fall within a period known as the mid-Pleistocene transition (1.25–0.7 mya), when the 41 ky glacial-interglacial cycles changed to 100 ky cycles³⁶. In this period, an increase in the duration of glacial periods (ca. 80–85 ky), compared with interglacial periods (ca. 15–20 ky), likely favored the expansion of the steppe biota over a large part of Europe. We argue that the extended duration of warm stages during the mid-Pleistocene transition led to equally prolonged range contractions for the European steppe species, which likely facilitated initial allopatric divergence between today's extrazonal and zonal lineages. A similar pattern of intraspecific divergence during the mid-Pleistocene, referred to as the 'Pleistocene species pump', has been highlighted for European butterfly species, based also on genome-wide data⁸.

The CNN modeling inferred a Late Pleistocene demographic expansion between 51 and 82 kya across all study species (Figure 3, Supplementary Table 4). This period was characterized by a significant decrease in global mean temperatures³⁷, Figure 3) and corresponds to the marine isotopic stages 4 and 3, with a gradual cooling during the LGP, in particular during stage 4³⁸. This colder climate likely triggered range expansions in European steppe species, which is also reflected by the increase of pollen attributed to

open land species in the palynological record (Figure 3, Supplementary Figure 3). Pronounced demographic responses on a global scale have been inferred during this period in organisms with different environmental affinities^{39,40}, highlighting the severity and pervasiveness of late Pleistocene climate change. This congruence in demographic events across ecologically divergent species was interpreted as a direct effect of an abrupt global temperature drop induced by the eruption of the Toba supervolcano ca. 74 kya⁴¹. Irrespective of the cause of this global event, we hypothesize that the rapidly cooling climate was key for the massive demographic expansion of the European steppe biota.

Finally, a sharp decline in population size was inferred for all analysed species around the mid to late Holocene (6.7 to 3.2 kya, Figure 3). Interestingly, our data suggest that populations did not collapse immediately after the end of the LGP, ~12 kya, but during or after the Holocene climatic optimum (9 to 5 kya). The warm and humid climate during this period fostered the expansion of deciduous forests^{28,42} and, at the same time, led to a decline of the remaining European steppes and forest steppes (Figure 1). We conclude that the expansion of closed forest vegetation during the Holocene climate optimum was likely the final 'killing blow' for many populations of the European steppe biota, triggering a rapid collapse in population size (Figure 3, Supplementary Figure 3).

Lineage distribution models based on climatic variables indicate adjacency or even overlap of the LGM ranges of extrazonal and zonal lineages in all five study species to the north of the Alps (Figure 2B). This is mirrored by the location of putative contact zones, where hybridization between extrazonal and zonal lineages has caused the frequent occurrence of admixed populations (*E. seguieriana*, *P. taurica* and *S. capillata*, Figure 2). To the south of the Alps, some range overlaps were also modeled, but large suitability gaps clearly prevail (Figure 2B). Admixed populations are virtually absent (Figure 2A); if they indeed existed, they were likely extirpated as the climate became unsuitable in the Holocene.

The dynamic oscillations of steppe vegetation in Europe during the late Quaternary are reflected in the modeled population size changes of the study species. Specifically, population expansions and contractions retrieved by our models were massive and in a similar range across the investigated species (a 62- to 72-fold LGM increase in extrazonal lineages, 55- to 92-fold in zonal lineages; a 49- to 72-fold Holocene decrease in extrazonal lineages, 16- to 34-fold in zonal lineages; Figure 3). CNN modeling also suggests that postglacial population contractions were more pronounced in the extrazonal than in the zonal steppe lineages, in accordance with the proportion of past and present availability of suitable habitat in Europe (Figure 1, Figure 2B). In other words, zonal lineages were able to maintain larger population sizes compared with extrazonal lineages because they had larger continuous ranges throughout the studied time periods (Figure 1).

Simulation-based, likelihood-free modeling approaches, such as ABC or CNN, have become popular in phylogeography because of the ease to explore complex demographic scenarios with multiple interacting parameters without the need to derive the likelihood function of parameter dependencies^{20,22}. However, these approaches may be less efficient for parameter estimation than full likelihood-based methods, such as Maximum Likelihood or Bayesian inference, because they rely on simulations to explore a potentially

large range of parameter values^{43,44}. Machine-learning CNN offers the advantage over ABC methods that information is extracted directly from the entire alignment of SNPs, better capturing patterns of site genetic variation in genome-wide sequence data than the use of a single or multiple summary statistics^{23,45}. In our study, we showed the power of a combined approach, in which CNN is used first to recover information directly from the SNP matrices and to reduce the initial parameter space, followed by an ABC rejection step on CNN predictions²⁵. Our CNN-ABC approach allowed us to disentangle the demographic histories of a diverse array of phylogenetically unrelated European steppe species differing in their natural history, ecology, and reproductive strategy. In essence, we uncovered a congruent signal of climate-driven changes in geographic range and population sizes in zonal steppe lineages, but idiosyncratic genetic histories for extra-zonal lineages that might be linked to species-specific differences in effective dispersal distance.

Material And Methods

Restriction-site associated DNA sequencing

The RADseq data analysed in this manuscript were generated by Kirschner et al.¹⁹ using the original RADseq protocol⁴⁶ with minor modifications⁴⁷. Raw data were downloaded from the NCBI short read archive (Supplementary Data 1) and genomic SNPs were called anew, using the newer version 2.3d of the Stacks package⁴⁸. Several runs of `denovo_map.pl` were done on a subset of raw sequence data to optimize loci yield for each species, following⁴⁹. The following species-specific parameters were used eventually: *E. seguieriana*, -n 3 -M 3 -m 5; *P. taurica* -n 3 -M 3 -m 5; *O. petraeus* -n 8 -M 8 -m 5; *S. nigromaculatus* -n 3 -M 3 -m 5; *S. capillata* -n 8 -M 8 -m 5 (-n number of mismatches allowed between fragments between individuals), -M (number of mismatches allowed per fragment) and -m (minimum depth of coverage required to call a fragment)⁴⁸.

Bayesian clustering analysis

STRUCTURE v. 2.3.4.50 was used to explore patterns of genetic grouping within our datasets. Input files were exported from the Stacks catalogue using the function `populations.pl`⁴⁸. Only a single SNP per RADseq fragment (`-write-single-snp` flag in `populations.pl`) was selected to avoid linked SNPs, which violate the algorithm's assumption that SNPs are in linkage disequilibrium. In an additional filtering step, loci with excess heterozygosity (> 65%) (`-max-obs-het` flag) were removed. This procedure has been suggested as a way to mitigate calling of paralogous loci from RADseq data⁵¹. The final alignments contained only loci present in at least 40% (*E. seguieriana*), 15% (*O. petraeus*), 50% (*P. taurica*), 25% (*S. nigromaculatus*), or 33% (*S. capillata*) of all populations. STRUCTURE⁵⁰ was run assuming a grouping into K = 1 to 5 clusters for 1,000,000 generations, using a burnin of 100,000 generations and ten replicates per K. The optimal K was assessed based on the rate change in likelihood among runs⁵².

Estimation of divergence times

Relative divergence times between the extra-zonal and zonal genetic lineages were inferred by applying a multi-species coalescent (MSC) model as implemented in the software BPP v. 4.2.935, and using the 'fixed topology' approach (option A00). For each species, RADseq tags were exported from the STACKS catalogue via `populations.pl` using the `-fasta-samples` flag and the `-max-obs-het` flag to remove loci with excess heterozygosity⁴⁸ (> 65%, see also above), and were further converted from fasta files to phylip files using the python script `fasta2genotype.py`⁵³. RADseq tags missing in more than 85% (*P. taurica*), 75% (*E. seguieriana*, *S. capillata*), or 50% (*O. petraeus*, *S. nigromaculatus*) were removed in each species. To reduce computational time, random subsets were generated containing 30 (*E. seguieriana*), 33 (*O. petraeus*), 33 (*P. taurica*), 30 (*S. capillata*), or 23 (*S. nigromaculatus*) individuals proportionally sampled from the extra-zonal and zonal group in each instance (Supplementary Table 1). Similarly, the full alignments of each species were randomly subsetting into smaller alignments containing 300, 400 and 500 RADseq loci for the final analysis. Analysing SNP subsets of different size has been suggested as a way to evaluate consistency of estimates within a given dataset^{35,54,55}. The BPP analyses were run under default settings, assuming data to be diploid and unphased³⁵. MCMC chain length was set to 1,000,000 generations, and 10% of the samples were discarded as burnin. All runs were checked in Tracer v. 1.6.056 to evaluate chain convergence to stationarity and adequate mixing, and to check if the effective sample size for estimated parameters reached at least 200.

Next, the function `msc2time.r` implemented in the R package `bppr`^{57,58} was used to calibrate the relative branch lengths obtained in BPP to absolute divergence times. Specifically, this function calculates absolute divergence times based on MSC derived estimates of τ , by sampling mutation rate and generation time from a gamma distribution to obtain estimates of mutation rate per absolute time. Mutation rates for each species were taken from literature-based estimates for genome-wide mutation rates (plants: 7×10^{-9} substitutions per site per generation⁵⁹; animals: 2.8×10^{-9} substitutions per site per generation⁶⁰, and a deviation of 10% was allowed when calibrating τ .

The generation times used to calibrate the time estimates were defined as average time between two successive generations within a lineage or population⁶¹. In the case of the univoltine grasshopper species *O. petraeus* and *S. nigromaculatus*, this generation time is one year. For the ant species *Plagiolepis taurica*, no species-specific data is available. The most thorough study on generation times in ants targeting the red harvester ant *Pogonomyrmex barbatus* found generation times (as defined above) in wild populations to be 7.8 years on average⁶². A slightly longer generation time of 10 ± 2 years was assumed for *P. taurica*, since their colonies are smaller and produce fewer offspring. Generation time estimates were not available for the two studied plant species, and the maximum lifespans of related and ecologically similar species were the only available references. Consequently, generation times of 10 ± 2 years and 25 ± 5 years were used for *E. seguieriana*⁶³ and *S. capillata*⁶⁴ (Podgaevskaya & Zolotareva pers. comm. 2020), respectively.

Exploration of demographic history

SNP data were exported to vcf files from the Stacks catalogue using the `-vcf` flag in `populations.pl`, allowing for a single SNP per locus by using the `-write-single-snp` flag⁴⁸; separate vcf files were generated for the extrazonal lineage and the zonal lineage (n of individuals and sites given in Supplementary Figure 1). From each of these variant files, individuals with an excessive amount of missing data were discarded, and the software `vcftools`⁶⁵ was subsequently used to remove SNPs that were missing in more than 85% (*E. seguieriana*), 75% (*P. taurica*, *S. capillata*), and 50% (*O. petraeus*, *S. nigromaculatus*) of the individuals. Calculation of the joint site frequency spectrum (SFS) was done using a custom Python script written by Isaac Overcast (available at GitHub <https://github.com/isaacovercast/easySFS>). This method is particularly suitable for RADseq data, as it handles missing data in the SNP matrix by downprojecting to a smaller sample size and averaging over all possible resamplings. Following the author's suggestions, downprojection was chosen to retain the maximum number of individuals while avoiding loss of too many SNPs.

The resulting SFS were used to explore population-size changes for each species and for each genetic lineage, using Stairway plots^{26,27}. The blueprint files informing the algorithm were modified for each species, accordingly. Random breakpoints were defined as suggested²⁷. Average generation times and mutation rates were the same as those used for divergence time estimation (see above). The remaining input parameters were not changed from the default settings.

Demographic model testing using Convolutional Neural Networks

To better understand and compare the demographic dynamics of each study species, we evaluated the three potential scenarios for the evolution of the European steppe biota during the Pleistocene climatic oscillations described in the Introduction ('Parallel expansion', 'Zonal expansion only', 'No expansion'; Figure 1). The number of individuals analysed per species and lineage, and the number of sites is given in Supplementary Table 1. We performed 10,000 coalescent simulations per scenario with the software `ms`⁶⁶, with species-specific priors for generation time and mutation rate as described above, and population sizes based on our Stairway plot results. Because our empirical SNP datasets included different levels of missing data (Supplementary Table 1), we randomly inserted similar proportions of missing characters to the simulated SNP matrices for each species. This procedure allowed us to train the CNN to recover information from the genotype matrices, while also recognizing missing data. We used the network architecture from Oliveira et al.⁶⁷, modified to include suggestions from Sanchez et al.²⁵, namely the use of different kernel sizes and intercalation of convolutional layers with batch normalization. The trained networks were then used to predict the most likely model on the empirical SNPs and on a new set of 10,000 independent simulations per scenario. We also predicted parameter values for the empirical SNPs and 10,000 independent simulations for the preferred scenario. The obtained CNN predictions were then used to perform an ABC step (with an approach similar to Mondal et al.⁶⁸; and also recommended by Sanchez et al.²⁵).

Distribution models for extrazonal and zonal lineages

The potential range occupancy of extrazonal and zonal lineages under climatic conditions of the LGM were estimated using the lineage range estimation method⁶⁹. Species distribution models under LGM climatic conditions based on two general circulation models (MIROC70; CCSM471) were available for all study species¹⁹; lineage ranges of extrazonal and zonal lineages within each species were based on these models. Affiliation of each population to the extrazonal or the zonal lineage was derived from the STRUCTURE results; admixed populations were affiliated using a simple majority rule. Lineage range estimation followed the method by Rosauer et al.⁶⁹, using the R script provided by the authors (github.com/DanRosauer/phylospatial) with default parameters. A relaxed 10th percentile training presence (p10) threshold was applied. This approach was chosen because more stringent thresholds, such as the maximum training sensitivity plus specificity (MTSS) threshold, have been shown to severely under-represent species ranges if a niche is projected from a contracted present-day niche model, which is the case for the Eurasian steppe biota⁷². The suitability values of lineage distribution models were assessed along two transects north and south of the Alps, using the Temporal/Spectral Profile Tool v. 2.0.3 in QGIS v. 3.10. This gradient analysis was done to visualize continuities and gaps of habitat suitability within areas north and south of the Alps, which acted as a major distribution barrier for many species.

Bibliography

1. Shackleton, N. J., Sánchez-Goñi, M. F., Pailler, D. & Lancelot, Y. Marine Isotope Substage 5e and the Eemian Interglacial. *Glob. Planet. Change* 36, 151–155 (2003).
2. Shackleton, N. J., Chapman, M., Sánchez-Goñi, M. F., Pailler, D. & Lancelot, Y. The Classic Marine Isotope Substage 5e. *Quat. Res.* 58, 14–16 (2002).
3. Hofreiter, M. & Stewart, J. Ecological Change, Range Fluctuations and Population Dynamics during the Pleistocene. *Curr. Biol.* 19, R584–R594 (2009).
4. Hewitt, G. M. Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.* 68, 87–112 (1999).
5. Petit, R. J. et al. Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* 300, 1563–1565 (2003).
6. Magri, D., Di Rita, F., Aranbarri, J., Fletcher, W. & González-Sampériz, P. Quaternary disappearance of tree taxa from Southern Europe: Timing and trends. *Quat. Sci. Rev.* 163, 23–55 (2017).
7. Calatayud, J. et al. Pleistocene climate change and the formation of regional species pools. *Proc. R. Soc. B Biol. Sci.* 286, 20190291 (2019).
8. Ebdon, S. et al. The Pleistocene species pump past its prime: evidence from European butterfly sister species. *bioRxiv* 2020.09.04.282962 (2020) doi:10.1101/2020.09.04.282962.
9. Závieská, E. et al. Multiple auto- and allopolyploidisations marked the Pleistocene history of the widespread Eurasian steppe plant *Astragalus onobrychis* (Fabaceae). *Mol. Phylogenet. Evol.* 106572 (2019) doi:10.1016/J.YMPEV.2019.106572.
10. Wesche, K. et al. The Palaearctic steppe biome: a new synthesis. *Biodivers. Conserv.* 25, 2197–2231 (2016).

11. Walter, H. & Breckle, S. *Ökologie der Erde*, Band 1. (Spektrum Akademischer Verlag, 1991).
12. Braun-Blanquet, J. *Die inneralpine Trockenvegetation: von der Provence bis zur Steiermark* / Von J. Braun-Blanquet. (Gustav Fischer, 1961).
13. Hurka, H. et al. The Eurasian steppe belt: Status quo, origin and evolutionary history. *Turczaninowia* 22, 5–71 (2019).
14. Jännicke, W. *Die Sandflora von Mainz, ein Relict aus der Steppenzeit*. (Gebrueder Knauer, 1892).
15. Allen, J. R. M. et al. Rapid environmental changes in southern Europe during the last glacial period. *Nature* 400, 740–743 (1999).
16. Reille, M. & de Beaulieu, J. L. Pollen analysis of a long upper Pleistocene continental sequence in a Velay maar (Massif Central, France). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 80, 35–48 (1990).
17. Sadori, L. et al. Pollen-based paleoenvironmental and paleoclimatic change at Lake Ohrid (south-eastern Europe) during the past 500 ka. *Biogeosciences* 13, 1423–1437 (2016).
18. Ellenberg, H. & Leuschner, C. *Vegetation Mitteleuropas mit den Alpen: in ökologischer, dynamischer und historischer Sicht*. (Stuttgart: Verlag Eugen Ulmer, 2010).
19. Kirschner, P. et al. Long-term isolation of European steppe outposts boosts the biome's conservation value. *Nat. Commun.* 11, 1968 (2020).
20. Fonseca, E. M., Colli, G. R., Werneck, F. P. & Carstens, B. C. Phylogeographic model selection using convolutional neural networks. *Mol. Ecol. Resour.* (2021) doi:<https://doi.org/10.1111/1755-0998.13427>.
21. Beaumont, M. A., Zhang, W. & Balding, D. J. Approximate Bayesian Computation in Population Genetics. *Genetics* 162, 2025–2035 (2002).
22. Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. & François, O. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* 25, 410–418 (2010).
23. Flagel, L., Brandvain, Y. & Schrider, D. R. The Unreasonable Effectiveness of Convolutional Neural Networks in Population Genetic Inference. *Mol. Biol. Evol.* 36, 220–238 (2019).
24. Robert, C. P., Cornuet, J.-M., Marin, J.-M. & Pillai, N. S. Lack of confidence in approximate Bayesian computation model choice. *Proc. Natl. Acad. Sci.* 108, 15112–15117 (2011).
25. Sanchez, T., Cury, J., Charpiat, G. & Jay, F. Deep learning for population size history inference: Design, comparison and combination with approximate Bayesian computation. *Mol. Ecol. Resour.* 00, 1–16 (2020).
26. Liu, X. & Fu, Y.-X. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* 21, 280 (2020).
27. Liu, X. & Fu, Y.-X. Exploring population size changes using SNP frequency spectra. *Nat. Genet.* 47, 555–559 (2015).
28. Magri, D. et al. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytol.* 171, 199–221 (2006).

29. Arenas, M., Ray, N., Currat, M. & Excoffier, L. Consequences of Range Contractions and Range Shifts on Molecular Diversity. *Mol. Biol. Evol.* 29, 207–218 (2012).
30. Excoffier, L., Foll, M. & Petit, R. J. Genetic Consequences of Range Expansions. *Annu. Rev. Ecol. Evol. Syst.* 40, 481–501 (2008).
31. Mona, S., Ray, N., Arenas, M. & Excoffier, L. Genetic consequences of habitat fragmentation during a range expansion. *Heredity* 112, 291–299 (2014).
32. Szűcs, M., Melbourne, B. A., Tuff, T. & Hufbauer, R. A. The roles of demography and genetics in the early stages of colonization. *Proc. R. Soc. B Biol. Sci.* 281, 20141073 (2014).
33. Narbona, E., Arista, M. & Ortiz, P. L. Explosive seed dispersal in two perennial Mediterranean Euphorbia species (Euphorbiaceae). *Am. J. Bot.* 92, 510–516 (2005).
34. Stevens, V. M. et al. A comparative analysis of dispersal syndromes in terrestrial and semi-terrestrial animals. *Ecol. Lett.* 17, 1039–1052 (2014).
35. Flouri, T., Jiao, X., Rannala, B. & Yang, Z. Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Mol. Biol. Evol.* 35, 2585–2593 (2018).
36. Willeit, M., Ganopolski, A., Calov, R. & Brovkin, V. Mid-Pleistocene transition in glacial cycles explained by declining CO₂ and regolith removal. *Sci. Adv.* 5, eaav7337–eaav7337 (2019).
37. Hansen, J., Sato, M., Russell, G. & Kharecha, P. Climate sensitivity, sea level and atmospheric carbon dioxide. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* 371, 20120294 (2013).
38. Martinson, D. G. et al. Age dating and the orbital theory of the ice ages: Development of a high-resolution 0 to 300,000-year chronostratigraphy. *Quat. Res.* 27, 1–29 (1987).
39. O’Connell, K. A. et al. Impacts of the Toba eruption and montane forest expansion on diversification in Sumatran parachuting frogs (*Rhacophorus*). *Mol. Ecol.* 29, 2994–3009 (2020).
40. Theodoridis, S. et al. How Do Cold-Adapted Plants Respond to Climatic Cycles? Interglacial Expansion Explains Current Distribution and Genomic Diversity in *Primula farinosa* L. *Syst. Biol.* 66, 715–736 (2017).
41. Williams, M. The ~73 ka Toba super-eruption and its impact: History of a debate. *Quat. Int.* 258, 19–29 (2012).
42. Marquer, L. et al. Quantifying the effects of land use and climate on Holocene vegetation in Europe. *Quat. Sci. Rev.* 171, 20–37 (2017).
43. Jackson, N. D., Morales, A. E., Carstens, B. C. & O’Meara, B. C. PHRAPL: Phylogeographic Inference Using Approximate Likelihoods. *Syst. Biol.* 66, 1045–1053 (2017).
44. Oaks, J. R. Full Bayesian Comparative Phylogeography from Genomic Data. *Syst. Biol.* 68, 371–395 (2019).
45. Perez, M. F. et al. Coalescent-based species delimitation meets deep learning: Insights from a highly fragmented cactus system. *bioRxiv* 2020.12.23.424219 (2020) doi:10.1101/2020.12.23.424219.
46. Baird, N. a. et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3, 1–7 (2008).

47. Paun, O. et al. Processes Driving the Adaptive Radiation of a Tropical Tree (*Diospyros*, Ebenaceae) in New Caledonia, a Biodiversity Hotspot. *Syst. Biol.* 65, 212–227 (2016).
48. Catchen, J., Hohenlohe, P., Bassham, S., Amores, A. & Cresko, W. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140 (2013).
49. Paris, J. R., Stevens, J. R. & Catchen, J. M. Lost in parameter space: a road map for stacks. *Methods Ecol. Evol.* 8, 1360–1373 (2017).
50. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945 LP – 959 (2000).
51. O’Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M. & Portnoy, D. S. These aren’t the loci you’re looking for: Principles of effective SNP filtering for molecular ecologists. *Mol. Ecol.* 27, 3193–3206 (2018).
52. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14, 2611–2620 (2005).
53. Maier, P. A., Vandergast, A. G., Ostoja, S. M., Aguilar, A. & Bohonak, A. J. Pleistocene glacial cycles drove lineage diversification and fusion in the Yosemite toad (*Anaxyrus canorus*). *Evolution* 73, 2476–2496 (2019).
54. Ortiz, D., Pekár, S., Bilat, J. & Alvarez, N. Poor performance of DNA barcoding and the impact of RAD loci filtering on the species delimitation of an Iberian ant-eating spider. *Mol. Phylogenet. Evol.* 154, 106997 (2021).
55. Tiley, G. P., Poelstra, J. W., dos Reis, M., Yang, Z. & Yoder, A. D. Molecular Clocks without Rocks: New Solutions for Old Problems. *Trends Genet.* (2020) doi:<https://doi.org/10.1016/j.tig.2020.06.002>.
56. 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).
57. Angelis, K. & Dos Reis, M. The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. *Curr. Zool.* 61, 874–885 (2015).
58. Yoder, A. D. et al. Geogenetic patterns in mouse lemurs (genus *Microcebus*) reveal the ghosts of Madagascar’s forests past. *Proc. Natl. Acad. Sci.* 113, 8049–8056 (2016).
59. Ossowski, S. et al. The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. *Science* 327, 92 LP – 94 (2010).
60. Keightley, P. D., Ness, R. W., Halligan, D. L. & Haddrill, P. R. Estimation of the Spontaneous Mutation Rate per Nucleotide Site in a *Drosophila melanogaster* Full-Sib Family. *Genetics* 196, 313 LP – 320 (2014).
61. Charlesworth, B. *Evolution in Age-Structured Populations*. (Cambridge University Press, 1994). doi:10.1017/CBO9780511525711.
62. Ingram, K. K., Pilko, A., Heer, J. & Gordon, D. M. Colony life history and lifetime reproductive success of red harvester ant colonies. *J. Anim. Ecol.* 82, 540–550 (2013).

63. Lauenroth, W. K. & Adler, P. B. Demography of perennial grassland plants: survival, life expectancy and life span. *J. Ecol.* 96, 1023–1032 (2008).
64. Golubeva, I. V. The age structure and numbers dynamics of feather grass L. (*Stipa pennata*) in the conditions of meadow steppe. *Sci. Proc. Mosc. Reg. Pedagog. Inst. Nat. Geogr. Inst.* 153, 283–303 (1964).
65. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158 (2011).
66. Hudson, R. R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 18, 337–338 (2002).
67. Oliveira, E. A. et al. Historical demography and climate driven distributional changes in a widespread Neotropical freshwater species with high economic importance. *Ecography* 43, 1291–1304 (2020).
68. Mondal, M., Bertranpetit, J. & Lao, O. Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. *Nat. Commun.* 10, 246 (2019).
69. Rosauer, D. F., Catullo, R. A., VanDerWal, J., Moussalli, A. & Moritz, C. Lineage Range Estimation Method Reveals Fine-Scale Endemism Linked to Pleistocene Stability in Australian Rainforest Herpetofauna. *PLOS ONE* 10, e0126274 (2015).
70. Watanabe, S. et al. MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. *Geosci. Model Dev.* 4, 845–872 (2011).
71. Gent, P. R. et al. The Community Climate System Model Version 4. *J. Clim.* 24, 4973–4991 (2011).
72. Richmond, O. M. W., McEntee, J. P., Hijmans, R. J. & Brashares, J. S. Is the Climate Right for Pleistocene Rewilding? Using Species Distribution Models to Extrapolate Climatic Suitability for Mammals across Continents. *PLOS ONE* 5, e12899 (2010).
73. Anhuf, D., Bräuning, A., Burkhard, F. & Max, S. Die Vegetationsentwicklung seit dem Höhepunkt der letzten Eiszeit. in *Nationalatlas Bundesrepublik Deutschland. Band 3. Klima, Pflanzen- und Tierwelt* (ed. Kappas, M.) 88–91 (Spektrum, 2003).
74. Becker, D., Verheul, J., Zickel, M. & Willmes, C. LGM paleoenvironment of Europe - Map. CRC806-Database (2015) doi:10.5880/SFB806.15.
75. de Beaulieu, J.-L. & Reille, M. Long Pleistocene pollen sequences from the Velay Plateau (Massif Central, France). *Veg. Hist. Archaeobotany* 1, 233–242 (1992).
76. Tzedakis, P. C. C., Emerson, B. C. C. & Hewitt, G. M. M. Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends Ecol. Evol.* 28, 696–704 (2013).

Declarations

Acknowledgements

We thank all colleagues listed as collectors in Supplementary Data 1 that provided samples, and also all colleagues that supported us by sharing locality data. We thank C. Lebas for providing his image of *P. taurica* shown in Figure 3, and H. Wiesbauer for his aerial image of an extrazonal steppe shown in Figure

1. The presented study was funded by the Austrian Science Fund (FWF, project P25955 ‘Origin of steppe flora and fauna in inner-Alpine dry valleys’ to P.S.). We thank the centre for Italian studies at the University of Innsbruck that supported our sampling campaign in Italy. We acknowledge the excellent HPC infrastructure LEO at the University of Innsbruck, and also thank the Vienna Scientific Cluster (VSC) - both facilities were central for the success of this study.

Author Contributions

P.S., I.S., M.P. and P.K. designed the study. E.Z., M.P. and P.K. analysed the data. M.P. conceived the CNN method. F.M.S., P.S., I.S., M.P., and P.K. co-wrote the manuscript. L.M. provided paleoecological expertise and data, and wrote corresponding parts of the manuscript. B.C.S.-S. and N.A. contributed to the development of the manuscript and improved early drafts of the paper. Members of the Steppe Consortium contributed in manuscript writing, sample collection and provided lab expertise.

Figures

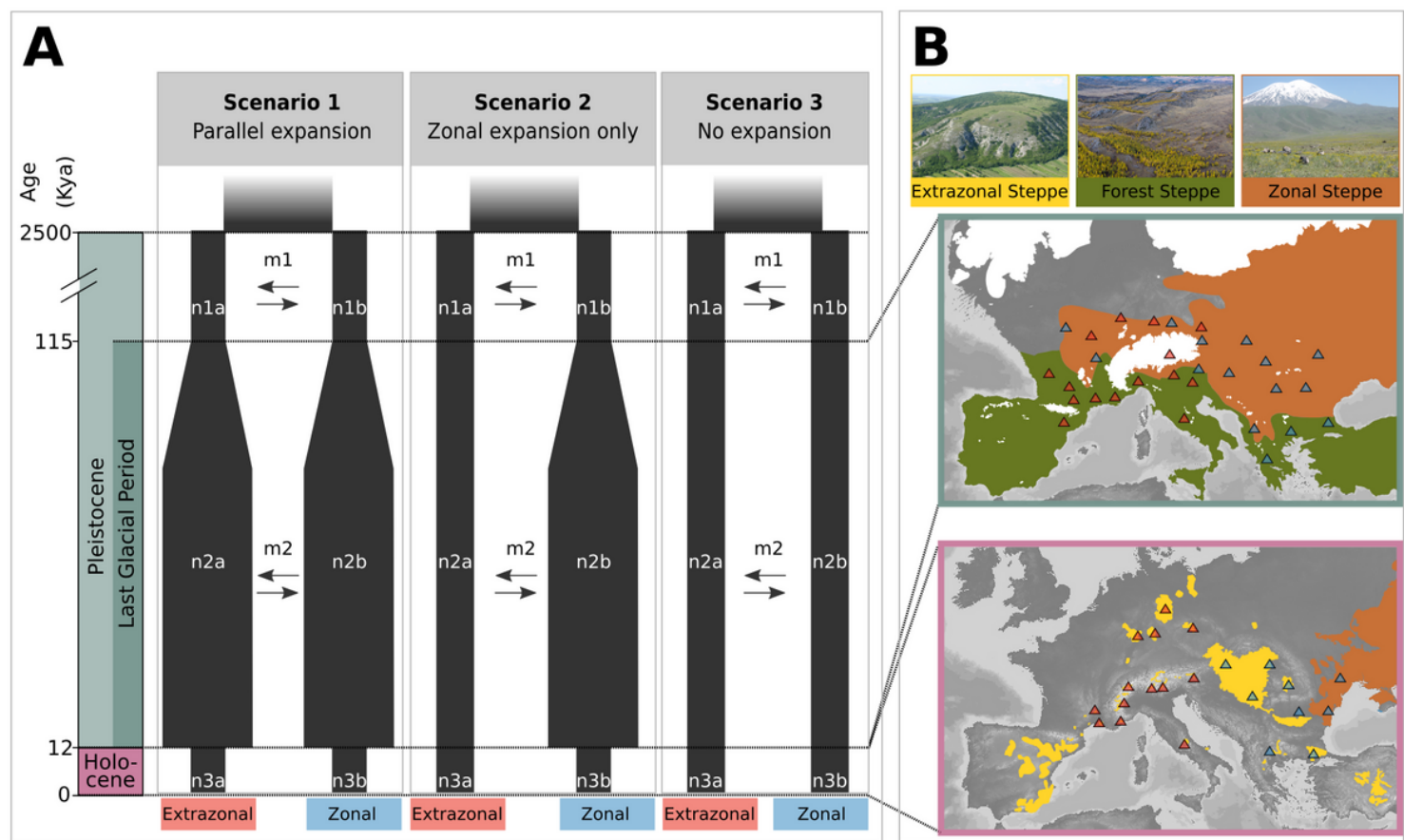


Figure 1

A. Overview of the three demographic scenarios evaluated for the European steppe biota in relation to age and geological epoch (scaling not continuous). In Scenario 1, ‘Parallel expansion’, an increase in effective population size (n) in extrazonal and zonal steppe lineages during the Last Glacial Period (LGP), followed

by a contraction of n in the Holocene, is assumed. Scenario 2, 'Zonal expansion only', implies expansion and contraction for zonal steppe lineages, but a constant population size for extrazonal steppe lineages. Scenario 3, 'No expansion', involves no change in effective population size in neither extrazonal nor zonal lineages. All scenarios include bidirectional migration (m) between the extrazonal and zonal steppe lineages in the pre-LGP and LGP periods. B. Vegetation in Europe under climatic conditions of the Last Glacial Maximum ~21,000 kya (upper map, modified from Anhuf et al.⁷³), and under present-day (that is, warm stage) conditions (modified from Kirschner et al.¹⁹; Wesche et al.¹⁰) brown areas represent zonal steppe (i.e. macroclimatically induced continuous steppe), yellow areas represent extrazonal steppes (patchy steppe occurrences embedded in a matrix of forest), green areas represent forest steppe (open forest with an understory of steppe biota), and white areas show glaciated areas⁷⁴. Blue and red triangles show hypothetical occurrences of populations of steppe species pertaining to extrazonal and zonal lineages, respectively. Image credit: Extrazonal steppe, H. Wiesbauer; forest steppe: M. Zabarovsky / Shutterstock, zonal steppe: P. Schönschwetter

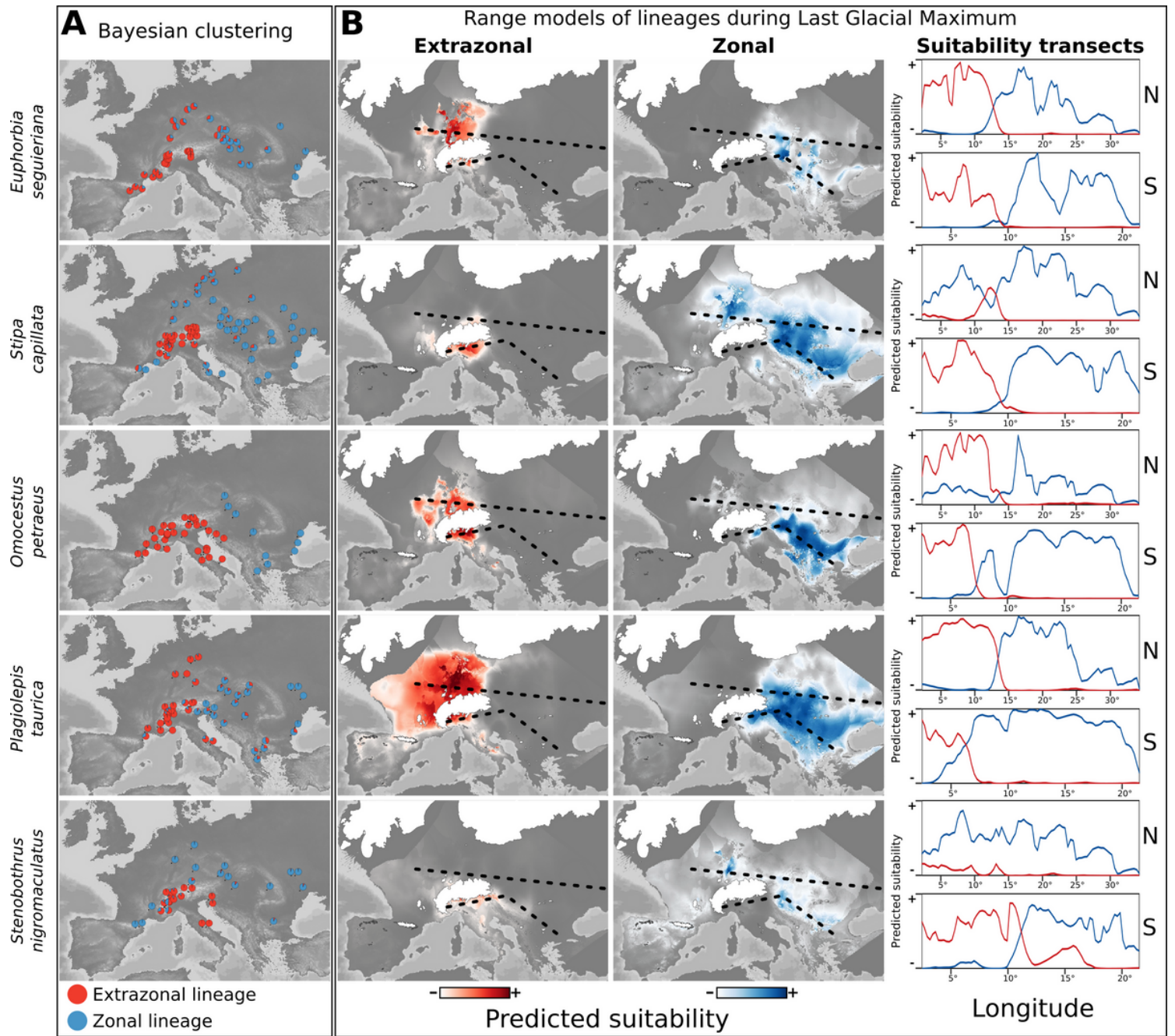


Figure 2

A. Distribution of extrazonal (red) and zonal (blue) lineages of steppe species based on genetic clusters inferred via Bayesian clustering. Each pie chart represents a population. B. Distribution of predicted species-specific habitat suitability (non-standardized) during the Last Glacial Maximum (21 kya) based on niche models using climatic variables. Suitability transects show predicted habitat suitability for the extrazonal lineage (red line) and the zonal lineage (blue line) along the dashed lines depicted in the maps; x-axes represent longitude; N, northern transect; S, southern transect.

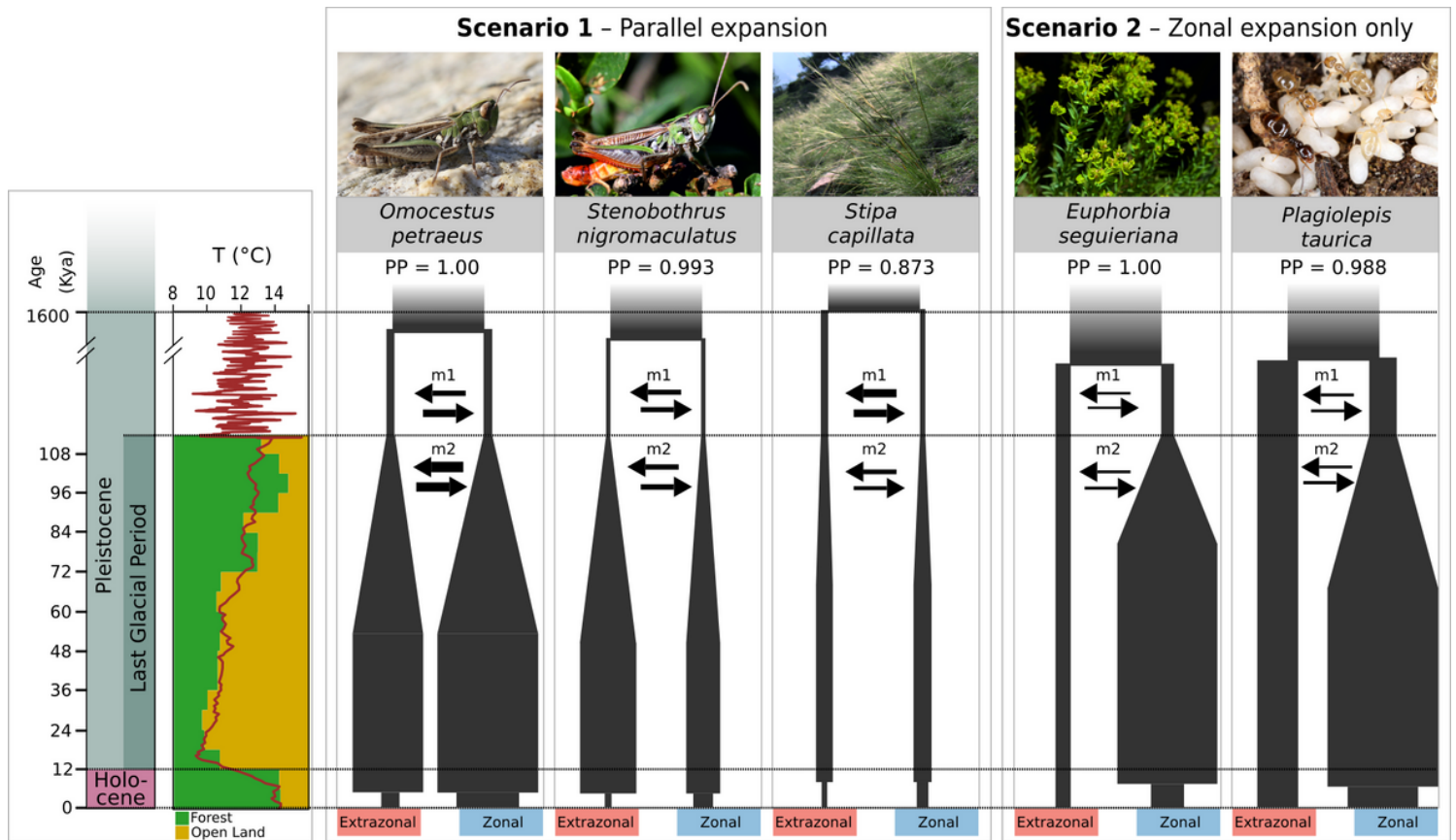


Figure 3

Results showing selected demographic scenarios and associated parameters using Convolutional Neural Network (CNN) modeling in five species endemic to the European steppes. Each species comprises an extrazonal (red) and a zonal (blue) lineage. The left inset figure shows the time scale with geological epochs and temperature fluctuations during the Late Quaternary³⁷: forest (green) and open land (yellow) pollen percentages throughout the Last Glacial Period (LGP) are shown as barplots; each bar represents 6000 year averages of the pollen percentages from five major pollen records covering the entire last glacial-interglacial cycle (data from European Pollen Database and PANGEA database, details in Supplementary Material)^{15–17,75,76}. In each demographic scenario, the timing of the initial split between the zonal and extrazonal lineages, the LGP expansion and the Holocene contraction are proportional to the time scale. The five species are assigned to their best-fitting model. Posterior probabilities (PP) values supporting the best-fitting model for each species are given above the demographic layout. In each layout, the diameters of branches and arrows are proportional to the inferred effective population sizes (n), and the inferred migration rates (m), respectively. For better visualization, branch diameters representing LGP population size were scaled down by 1/10 compared to pre- and post-LGP population sizes. Terminology for estimated parameters in Figure 1; results in Supplementary Table 4. Image credit: *E. seguieriana*, B. Frajman; *P. taurica*, C. Lebas; other species, P. Kirschner.

Supplementary Files

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

- [SupplementaryInformationKirschnerPerezetal.pdf](#)
- [Supplementarydata1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable4.xlsx](#)