RESEARCH ARTICLE

Long-term continuity of steppe grasslands in eastern Central Europe: Evidence from species distribution patterns and chloroplast haplotypes

Wolfgang Willner1 | Dietmar Moser1 | Kristina Plenk2 | Svetlana Aćić3 | Olga N. Demina4 | Maria Höhn5 | Anna Kuzemko6 | Jan Roleček7,8 | Kiril Vassilev9 | Denys Vynokurov6 | Matthias Kropf2

1Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria
2Institute for Integrative Nature Conservation Research, University of Natural Resources and Life Sciences, Vienna, Austria
3Department of Botany, Faculty of Agriculture, University of Belgrade, Zemun-Belgrade, Serbia
4Karachay-Circassian State University, Karachaevsk, Russia
5Department of Botany, Hungarian University of Agriculture and Life Sciences, Budapest, Hungary
6M.G. Kholodny Institute of Botany, NAS of Ukraine, Kyiv, Ukraine
7Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic
8Department of Paleoecology, Institute of Botany of the Czech Academy of Sciences, Brno, Czech Republic
9Department of Plant and Fungal Diversity and Resources, Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, Sofia, Bulgaria

Correspondence
Wolfgang Willner, Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria.
Email: wolfgang.willner@univie.ac.at

Funding information
Grantová Agentura České Republiky; National Research Foundation of Ukraine; Austrian Science Fund

Handling Editor: Simon Scheiter

Abstract
Aim: The steppe grasslands of eastern Central Europe are exceptionally species rich and valuable from a nature conservation point of view. However, their historical biogeography is still poorly understood. Here we use the regional diversity of habitat specialists and chloroplast DNA data to investigate potential long-term refugia of steppe species in this region.

Location: Pannonian Basin and adjacent regions; SW Russia.

Taxon: Vascular plants.

Methods: After identifying habitat specialists of the three main steppe types (meadow steppes, grass steppes and rocky steppes), we compiled their regional presence–absence in grid cells of 75 km × 75 km. We analysed the dependency of habitat specialist diversity to climate, topographic heterogeneity and geographical distance to potential refugia. For genetic analysis, we sampled three or four habitat specialists of each steppe type and used cpDNA markers to investigate intraspecific diversity and geographical distribution of haplotypes. We also tested for correspondence between the number of habitat specialists and haplotype diversity.

Results: Climate and topography explained between 40% and 63% of the variance in habitat specialist diversity. Adding geographical distance to potential refugia increased the explained variance in the models for all steppe types. Chloroplast haplotypes featured a complex pattern across the study area. Several species showed a strong geographical differentiation, suggesting migration waves from multiple refugia with only limited subsequent genetic intermixture. Maximum haplotype diversity in a region showed a better correlation with the number of habitat specialists per steppe type than mean haplotype diversity.

Main conclusions: We can clearly reject the scenario of a late-Holocene immigration of steppe species from areas outside the Pannonian Basin. Most species must have been present in the region since at least the early Holocene, highlighting the
importance of the lower mountain ranges surrounding the Pannonian Basin as long-term refugia for European steppe species. Dispersal limitation and resulting migration lags seem to have a strong influence on the distribution of steppe species in Central Europe.

**KEYWORDS**
Central Europe, chloroplast DNA, mid-Holocene bottleneck, Pannonian Basin, phylogeography, post-glacial migration, refugia, steppe species

1 | INTRODUCTION

The steppe grasslands of eastern Central Europe are located at the western margin of the Eurasian steppe belt (Erdős et al., 2018; Wesche et al., 2016). They are exceptionally species rich and valuable from a nature conservation point of view (Feurdean et al., 2018; Roleček et al., 2014), which is apparently related to their complex history. Steppe grasslands originated after climate cooling and aridification during the Tertiary and have been the dominant formation in large parts of Eurasia for most of the Quaternary (Binney et al., 2017; Hurka et al., 2019; Prentice et al., 2000). They prevailed during the drier and colder glacial periods, and only the relatively short interglacial periods had a sufficiently moist climate to support forests over large areas (Birks & Birks, 2004).

The last critical period for the long-term persistence of steppe vegetation in Central Europe was the mid-Holocene forest maximum, also known as the mid-Holocene bottleneck (Pokorný et al., 2015). There is growing evidence for a local persistence of steppe-like vegetation in eastern Central Europe throughout the Holocene, even in areas without rocky habitats (Kuneš et al., 2015; Magyari et al., 2010; Moskal del Hoyo et al., 2018; Pokorný et al., 2015). This evidence is mainly based on pollen and macrofossils, including snails. However, the fossil record often provides lower taxonomic resolution than is needed for a reliable reconstruction, and its taxonomic assemblages may be incomplete or biased (Greenwood, 1991; Odgaard, 1999). This opens up opportunities for multi-proxy studies (Birks & Birks, 2006) and use of less common proxies provided, for example, by pedological, phylogeographical or phytosociological studies (Feurdean et al., 2018; Willner et al., 2019).

In a study on European beech forests, Willner et al. (2009) showed that the regional diversity of characteristic understorey species is closely related to the geographical distance to glacial refugia of beech. Similarly, areas of endemism in the European Alps show a clear relationship with hypothetical glacial refugia and are congruent with molecular phylogeographical data (Schönswetter et al., 2005; Tribsch & Schönswetter, 2003). Indeed, broad-scale analyses of species distribution data as well as species distribution models suggest that many species ranges are restricted by migration lags (Dullinger et al., 2012; Svenning et al., 2008). Thus, the analysis of regional diversity patterns of habitat specialists seems to be a promising tool to infer past range dynamics of the habitats, including steppes. Another source of evidence may be provided by the geographical distribution of intraspecific genetic diversity (Avise, 2000; Kirschner et al., 2020). Such analyses allow identification of regions with high unique genetic diversity as possible long-standing refugial areas. Conversely, recently established populations will be characterised by low intraspecific genetic diversity (Hampe & Petit, 2005), reflecting just a fraction of the diversity in the refugia.

An obstacle in applying these approaches to steppe species is that the mid-Holocene bottleneck might have erased signals of earlier migrations. Another difficulty, compared to the study of forest and alpine species, is the lack of agreement on the position of potential refugia for European steppe species. One region that has frequently been suggested as refugial area for the more thermophilous steppe species is the Black Sea coast, in particular the Dobruja region (Magyari et al., 2010). Another potential refugium mentioned by Magyari et al. (2010) is the Vojvodina in the southern Pannonian Basin. Among the more northern regions, Transylvania and western Podolia have been proposed as potential refugia (Roleček et al., 2014, 2019) as well as the lower mountain ranges in Hungary (Fekete et al., 2010). Finally, a recent phylogeographic study (Plenk et al., 2020) suggested the north-western Pannonian region, including the eastern margin of the Alps, as another refugium for steppe species (Figure 1a). All these areas are characterised by a high topographic heterogeneity and mostly calcareous substrates, providing potentially suitable habitats for steppe species throughout the whole glacial-interglacial cycle.

Here, we define primary (i.e. long-term) refugia as those where steppe species have been continuously present since pre-LGM times, and secondary refugia as those which were colonised during the Late Glacial or early Holocene and became refugia only in the middle Holocene. Distinguishing between these two types of refugia is challenging without dated fossil records. In this study, we investigate potential refugia of meadow steppes, grass steppes and rocky steppes (see definitions of these steppe types in the Methods section) in eastern Central Europe using the regional diversity of their character species and the chloroplast haplotype diversity of selected species from these ecological groups. Specifically, we address the following questions:

1. Are the recent environmental factors (climate, topographic heterogeneity) sufficient to explain the regional diversity of meadow steppe, grass steppe and rocky steppe specialists in eastern Central Europe (indicating that there is no migration
lag), or does geographical distance to potential refugia play a significant role?

2. Do cpDNA data of habitat specialists provide evidence for the same refugial areas in eastern Central Europe?

3. Are these areas primary or rather secondary refugia?

4. Are there differences in the diversity patterns and positions of refugia among the three steppe types?

5. Is there a relation between cpDNA haplotype diversity and number of habitat specialists within a given region?

For each steppe type, we test three main scenarios, each one associated with a specific set of expectations (Figure 1b):

Scenario a: No survival of steppe during the mid-Holocene bottleneck in the Pannonian region; subsequent immigration of habitat specialists from eastern and south-eastern Europe. In this case, geographical distance to potential refugia within the Pannonian region should not be a significant explanatory variable. Haplotype diversity should be highest in the east or south-east of the study area, gradually decreasing towards the north-west, and there should be no unique haplotypes within the Pannonian region.

Scenario b: Survival of steppe throughout the mid-Holocene bottleneck in the Pannonian region, but no survival of the LGM; immigration of habitat specialists during the Late Glacial or early Holocene and survival in secondary refugia. Assuming that some habitat specialists still show a migration lag, geographical distance to potential refugia in the Pannonian region should be a significant variable explaining the regional diversity of habitat specialists. Haplotype diversity should have regional maxima in the secondary refugia. However, there should be no unique haplotypes associated with these refugia (or very few single-step mutations) as time was too short to accumulate mutations.

Scenario c: The steppe habitat specialists survived both the LGM and the mid-Holocene bottleneck in primary refugia within the Pannonian region. As in scenario b, geographical distance to potential refugia in the Pannonian region should be a significant variable explaining the regional diversity of habitat specialists. Haplotype diversity should be highest in primary refugia, and there should be unique (and genetically more divergent) haplotypes associated with each primary refugium.

2 | MATERIALS AND METHODS

2.1 | Study area and potential refugia

Our study area comprises the Pannonian Basin in eastern Central Europe and adjacent regions. Sampling for molecular analyses also covered the area of zonal steppes in south-western Russia. Following the proposals of previous studies (Fekete et al., 2010; Magyari et al., 2010; Plenk et al., 2020; Roleček et al., 2014, 2019), we defined the following regions as potential refugia for steppe species (Figure 1a): (1) eastern margin of the Alps and calcareous hills in NE Austria and
S Moravia (Czech Republic); (2) calcareous hills between the Vienna Basin and Lesser Hungarian Plain (Kisalföld); (3) Hungarian Central Ranges and SE Slovakia; (4) W Podolia in Ukraine; (5) Transylvanian Basin in Romania; (6) hills in Vojvodina (Serbia); and (7) Dobruja in Romania and Bulgaria. Focusing on the Pannonian region, more southerly potential steppe refugia (e.g. on the Balkan Peninsula) have not been addressed in our study.

2.2 Identification of habitat specialists

The term ‘steppe’ denotes a variety of grassland types, which are quite distinct in their habitat, physiognomy and floristic composition (Mucina et al., 2016; Walter & Breckle, 1999; Wescbe et al., 2016). In general, authors distinguish four main eco-floristic types of Eurasian steppes, ranging from relatively humid to extremely dry: (1) meadow steppes, corresponding to semi-dry grasslands dominated by broad-leaved grasses and mesophilous herbs, often occurring in a mosaic with forest patches (‘forest steppe’); (2) grass steppes, representing the typical steppe grasslands, which are less rich in mesic species and mostly dominated by drought-tolerant narrow-leaved bunch grasses (especially from the genera Festuca and Stipa); (3) desert steppes, that is relatively open grasslands with a high portion of dwarf shrubs, representing the transition to deserts; and (4) rocky steppes, that is azonal steppe grasslands on shallow rocky soils, physiognomically often similar to desert steppes. As each steppe type might have responded differently to the climatic oscillations of the Pleistocene, we study them separately. Typical desert steppes are not present in the modern landscapes of Central Europe, and we do not further address them in our study. Pictures of typical stands of the three other types are provided in Appendix S1.

Semi-dry grasslands are distributed throughout the temperate zone (Dierßen, 1996; Ellenberg & Leuschner, 2010; Willems, 1982), but in the western half of Europe they are usually not designated as meadow steppes and they are considered as secondary, man-made vegetation which developed after clearing of the original forests under centuries or even millennia of low-intensity land use (Poschlod et al., 2009). Rocky steppes, on the other hand, are generally accepted as mostly natural, albeit often small-scale habitats, occupying sites that are edaphically too dry for forest growth (Ellenberg & Leuschner, 2010). Patches of grass steppes are scattered all over eastern Central Europe (Borhidi, 2012; Chytrý, 2012; Willner et al., 2017) and also occur in the deep intra-montane valleys of the Alps (Braun-Blanquet, 1961; Kirschner et al., 2020).

We prepared a list of character species of European meadow steppes, grass steppes and rocky steppes, using the diagnostic species of the phytosociological class Festuco-Brometalia (dry grassland and steppe vegetation in the submediterranean, nemoral and hemiboreal zones of Europe), compiled by Mucina et al. (2016), as a starting point. In this checklist, the class Festuco-Brometalia is divided into 11 phytosociological orders, of which five occur in our study area. Meadow steppes are represented by the order Brachypodietalia pinnati, grass steppes by the order Festucetalia valesiaca and rocky steppes by the orders Stipo pulcherrimae-Festucetalia pallentis and Haloclysetalia sendtneri. The Scorzonetalia villosa is a heterogeneous order comprising both meadow steppes (alliance Scorzonierion villosae) and rocky steppes (alliances Chrysopogono grylli-Koelerion splendens and Saturejion subspicatae). We adopted the character species of these orders and alliances from the most recent phytosociological revisions (Aćić et al., 2015; Terzi, 2015; Willner et al., 2017, 2019) and assigned them to the three main steppe types (see Appendix S2). Steppe species with a broad ecological amplitude (e.g. character species of the class Festuco-Brometalia) were not included in the analysis.

2.3 Species distribution data

To analyse the regional diversity of the character species of each steppe type, we divided the study area into grid cells of 75 km × 75 km (Figure 1a) and compiled the presence-absence of each species in each grid cell. The selected grid size should be sufficient to capture the regional species pool of steppe species, so we expect that absences of species mainly reflect either unsuitable macroclimate or incomplete range filling. Some grid cells had to be excluded because no species distribution data were available for these areas (Bosnia and Herzegovina, Moldova, parts of Romania). Along the Adriatic and Black Sea coast, adjacent grid cells were merged in order to reduce differences in the actual area covered by the grid cells. The same was applied along the border between Moldova and Ukraine. Presence-absence data were compiled from various sources: species distribution atlases, national floras, floristic monographs, phytosociological databases, herbarium specimens and personal observations of the authors (see Appendix S3 for a list of all data sources; the full species presence-absence matrix is provided in Appendix S4). We compiled the most complete species lists possible for each grid cell in order to avoid false absences. Finally, we calculated the number of character species of each main steppe type in each grid cell.

2.4 Climate data, topographic heterogeneity and distance to potential refugia

Climatic predictors were based on the CHELSA climatologies 30° dataset (Karger et al., 2017). From the 19 bioclimatic variables, we selected bio1, 4, 5, 6, 10, 11, 12, 15, 16, 17, 18 and 19 as candidate predictors. We did not use bio2 and 3 due to their unclear calculation algorithm and meaning. We also excluded bio8 and 9, which describe the temperature in the wettest and driest period, because in our large study area the wettest period can be either in the warm or the cold season, making these variables difficult to interpret. Moreover, we excluded bio7, 13 and 14 due to their nearly identical meaning as bio4, 16 and 17 respectively. We overlaid the study grid with the climatic surfaces and calculated the average value for each climatic variable per cell.
Topographic heterogeneity was calculated as the ratio between 3D surface area and 2D area of the study grid cells (Irl et al., 2015) based on the EU-DEM V1.1 produced by the Copernicus Land Monitoring Service with funding by the European Union (https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-eu-dem). The minimum distances to potential refugia were calculated as the nearest geodetic distance of a study grid cell border to the border of a refugial area, as shown in Figure 1a.

2.5 | Analysis of species numbers

We analysed the dependency of the number of character species in each grid cell to climate, topographic heterogeneity and distance to potential refugia by means of generalised linear models with a Poisson-distributed response and the canonical log link function in R (version 3.6.2; R Core Team, 2020). To facilitate the comparisons of regression estimates and variable importance, all variables were scaled and centred by subtracting the mean and dividing by the standard deviation (Harrison et al., 2018). To account for possible nonlinear responses, we always tested the significance of second-order polynomials for each predictor.

We started with a full Poisson family (log link) generalised linear regression model with character species number as the response and bioclimatic variables and the 2D/3D-AreaRatio as predictors. All predictor variables entered as second-order polynomials. At first, we applied the step function (with the search mode set to ‘both’) to select the model by AIC. Next, all remaining variables were checked for multicollinearity by calculating the variance inflation factors (VIFs using vif function of the ‘car’ library; Fox & Weisberg, 2019). We then recursively removed variables with a VIF > 4 until none of VIFs exceeded 4. If models showed signs of overdispersion, that is residual deviance was larger than the degree of freedom, overdispersion was tested by the dispersiontest function from the ‘AER’ library (Kleiber & Zeileis, 2008) and the dispersion.glim function of the ‘bimeco’ library (Korner-Nievergelt et al., 2015). In case overdispersion was detected, we used quasi-Poisson models to correct p-values for overdispersion. Finally, we checked the significance of the variables using the Wald score test for Poisson family models and F-tests for quasi-Poisson models.

After fitting the regression models, we checked for spatial autocorrelation by plotting spline (cross-)correlograms using the function spline.correlog from the R package ‘ncf’ (Bjornstad, 2020). To test whether the inclusion of the geographical distance to potential refugia would improve model performance, we analysed the change in pseudo-\(R^2\) (ps\(R^2\))—as the ratio between residual deviance and null deviance—for models that include the minimum distance to refugia relative to models considering only climate and topographic heterogeneity. Since we did not know which areas acted as actual refugia for steppic species, we tested the effect of all possible combinations of refugia (e.g. for the combination of areas 1, 3 and 5, the predictor variable was the minimum geodetic distance of a study raster cell to any of these three refugia).

2.6 | Molecular data

For the analysis of intraspecific genetic diversity patterns, we studied 3 or 4 character species of each steppe type: Filipendula vulgaris (29 populations), Ranunculus bulbosus (16) and Trifolium montanum (26) for meadow steppes; Astragalus onobrychis (14), Linum austriacum (16), Salvia nemorosa (27) and Stipa capillata (26) for grass steppes; Fumana procumbens (11), Linum tenuifolium (16), Poa badensis (9) and Scorzonera austriaca (9) for rocky steppes. These 11 study species represent eight plant families, with each family represented by only one species per steppe type in order to avoid biases due to phylogenetic constraints. For A. onobrychis, we only included octoploid populations as the di- and tetraploid ones belong to a different lineage, which also has slightly different habitat preferences (Plenk et al., 2020; Záveská et al., 2019). Our primary sampling scheme was based on the same grid cells (75 km × 75 km) as the distribution data (see above), but restricted to 25 grid cells covering different potential refugial and non-refugial regions within the study area. Additional grid cells were sampled within the limits of available resources, as well as some populations in the European part of Russia. Within each grid cell, eight individual plants coming from a single population were sampled representatively across this population (geographical coordinates are given in Appendix S8).

All species were screened for different cpDNA markers, that is, rpl16 intron (Shaw et al., 2005), atpi-atpH, rpl32-trnL (Shaw et al., 2007) and ycf1 (Dong et al., 2012), and finally one to three markers per species were analysed for all individuals and populations. Single markers were aligned using the software BioEdit (7.2.5; Hall, 1999) and occurring indels (\(\text{indel} = \text{insertion and/or deletion}\)) were coded as single-step mutations. Inversions (if present) were handled the same way and single sequence gaps were treated as fifth character state. For all cpDNA analyses, we combined cpDNA sequences derived from different markers into one alignment per species. We used ncs (1.21; Clement et al., 2000) to create statistical parsimony haplotype networks under the default connection limit of 95% (90% for Salvia nemorosa, allowing to demonstrate connections of haplotype groups also for this species). Resulting cpDNA haplotypes were counted per population. Haplotype diversity \(h\) was calculated as genetic diversity index using ARequin (3.5.2.2; Excoffier & Lischer, 2010). To visualise the geographical distribution of haplotypes on maps, we partitioned the parsimony network of each species into a maximum of 10 haplotype groups and calculated their respective proportions for each population.

To test for the correspondence of habitat specialist species number and haplotype diversity, we calculated the Pearson correlation between the maximum of haplotype diversity of all sampled species of a particular steppe type in a grid cell and the number of habitat specialists occurring in the same cell. To test whether and at what spatial scale haplotype diversity is autocorrelated, we calculated spline (cross-)correlograms using the function spline.correlog from the R package ‘ncf’ (Bjornstad, 2020). We found no signs for autocorrelation, neither in near distances (neighbouring cells) nor in any other distance up to eight cells (600 km). The 95% confidence intervals for Morans’ \(I\) included zero for every distance tested.
3 | RESULTS

3.1 | Diversity patterns of habitat specialists

Among the species diagnostic for the class Festuco-Brometea and occurring in the study area, we identified 57 character species of meadow steppes, 37 character species of grass steppes and 64 character species of rocky steppes (Appendix S2).

Highest numbers of meadow steppe specialists were found in the north-western margin of the Pannonian Basin (with up to 51 species), the Hungarian Central Range (49), Transylvania (47), the north-western Dinaric Mountains (47) and the western Balkan...
Range (46 species). Local hotspots were also detected in Central Germany and western Podolia, with 43 species each (Figure 2; Figures S5.5–S5.7). Climate and topography explained ca. 40% of the variance of specialist diversity ($psR^2 = 0.402$). Variables with the highest explanatory power were temperature seasonality (bio4; negative relation with species number) and maximum temperature of the warmest month (bio5; positive relation). Annual precipitation (bio12; u-shaped response), precipitation of the warmest quarter (bio18; unimodal response) and topographic heterogeneity (positive relation) also contributed to the final model (Appendix S6).

Highest numbers of grass steppe specialists were observed in the steppe zone of Ukraine (with a maximum of 36 species), followed by the Dobruja region and northern Bulgaria (both up to 28 species). Local hotspots also occurred in the Pannonian Basin and Transylvania, both with a maximum of 23 species (Figure 2). Climate and topography explained c. 60% of the variance of specialist diversity ($psR^2 = 0.618$). Variables with the highest explanatory power were maximum temperature of the warmest month (bio5; positive relation with species number) and annual precipitation (bio12; negative relation), indicating an optimum in dry continental regions with hot summers. Minimum temperature of the coldest month (negative relation) and topographic heterogeneity (positive relation) were also included in the final model (Appendix S6).

Habitat specialists of rocky steppes had diversity hotspots in the mountain ranges of the Central Balkan (with up to 40 character species), the Dinaric Mountains (36), the Hungarian Central Range (35) and the eastern margin of the Alps (32 species) (Figure 2). Climate and topography explained almost two thirds of the variance of specialist diversity ($psR^2 = 0.63$). Climatic variables included in the final model were temperature seasonality (bio4; unimodal response), maximum temperature of the warmest month (bio5; positive relation), annual precipitation (bio12; u-shaped response) and precipitation of the warmest quarter (bio18; negative relation with species numbers). Taken together, these factors indicate a preference for areas with warm and dry summers but only moderate continentality. Topographic heterogeneity showed a strong positive correlation with the number of rocky steppe specialists (Appendix S6).

The residuals for all three models show some amount of spatial autocorrelation (Moran’s $I < 0.5$) within neighbouring cells (i.e. within a distance of 75 km) which could not be accounted for by the predictors used. Beyond the direct neighbours, spatial autocorrelation drops to low and insignificant levels, though (Figure S6.12). Adding geographical distance to potential refugia increased the explained variance in the models for all steppe types, with the highest increase for meadow steppes (Table 1). However, even the most improved model for meadow steppes still had a much lower $psR^2$ than the models for grass steppes and rocky steppes. Among the potential refugia included in the 10 best models for each steppe type, the Hungarian Central Range consistently appeared with the highest frequency, followed by the north-western margin of the Pannonian Basin, western Podolia and Transylvania (Table 1; see Appendix S6 for details).

### 3.2 Molecular data

The number of detected haplotypes varied substantially among the study species, from 11 in $P$. badensis to 63 in $S$. nemorosa (mean 28.3 ± 17.7; Appendix S7 and S8). The second Poaceae species, $S$. capillata, was also characterised by a rather low number of 16 haplotypes. We did not observe a general unidirectional haplotype diversity gradient across the study area (Figure 2). $T$. montanum, $A$. anobrychis, $L$. australicum, $F$. procumbens and $L$. tetuifolium showed a strong geographical differentiation of haplotypes, suggesting migration waves from multiple refugia with only limited genetic intermixture (Figure 3). $F$. vulgaris showed a moderate geographical differentiation with a conspicuous diversity maximum in Transylvania, while $R$. bulbosus, $S$. nemorosa and $S$. capillata had a weak geographical pattern. $P$. badensis and $S$. austriaca were represented by too few populations to draw firm conclusions. In most species, the haplotype diversity varied strongly among populations, with high and low diversity populations often situated in close geographical proximity (Appendix S7). In all study species at least one population was found to be monomorphic (i.e. containing only a single haplotype), with greatly varying proportions of monomorphic populations ranging between 6.3% ($R$. bulbosus) and 55.6% ($P$. badensis). When comparing south-western Russia with the Pannonian region, the most diverse populations showed similar haplotype diversity in both regions. However, for meadow steppe and grass steppe species, extremely low diversity values were only found in the Pannonian region.

### 3.3 Comparison of genetic diversity and number of habitat specialists

The haplotype diversity of individual species indicated no relation to the number of habitat specialists in a given region. The same was true for the mean haplotype diversity of the studied character species of each of the three steppe types. In contrast, when taking the maximum haplotype diversity observed among the character species of a steppe type in a given grid cell, correlations were positive, although not significantly (Figure 4). Correlations were relatively high for grass and meadow steppes (with $p < 0.08$), but much lower for rocky steppes ($p = 0.24$).

### 4 DISCUSSION

#### 4.1 Historical biogeography of the Pannonian steppes

The vegetation history of the Pannonian Basin has been the subject of many discussions and speculations (see Magyari et al., 2010, for a short review). Some of the recurrent questions in this debate are during which period the various steppe species arrived in the
Table 1: Summary of models to explain the number of character species of meadow steppes, grass steppes and rocky steppes in a grid cell, using environmental factors (climate, topographic heterogeneity) and distance to the next potential refugium as predictors. $R^2_{\text{env}}$: proportion of explained variance in model using only environmental factors. $R^2_{\text{dist}}$: proportion of explained variance in model using both environmental factors and distance to a specific set of refugia. Minimum and maximum values for the 10 best models are shown. The numbers in the right half of the table show how often a refugium (columns 1–7) was included in these 10 models.

| Region       | $R^2_{\text{env}}$ | $R^2_{\text{dist}}$ (min) | $R^2_{\text{dist}}$ (max) | Refugia in 10 best models |
|--------------|---------------------|---------------------------|---------------------------|---------------------------|
| Meadow steppes| 0.402               | 0.466                     | 0.488                     | 4 5 10 4 2 0             |
| Grass steppes | 0.618               | 0.641                     | 0.647                     | 7 7 10 6 4 0             |
| Rocky steppes | 0.630               | 0.666                     | 0.671                     | 7 8 9 4 1 0 5            |

region, from where, and whether they have been continuously present since the last glacial period (e.g. Soó, 1929, 1959; Wendelberger, 1954; Zőlyomi, 1953, 1964). Factors allowing the persistence of dry grasslands in eastern Central Europe might have been warmer and drier climate during the early Holocene (Feurdean et al., 2014), fire and grazing by wild herbivores, potentially slowing down the encroachment of forests when the climate became wetter at the onset of the middle Holocene (Feurdean et al., 2018), and the early arrival of Neolithic farmers, which gave the forests little time to gain full dominance (Pokorný et al., 2015; Roleček et al., 2021). The south-facing slopes of the hills and lower mountain ranges surrounding the Pannonian plain might have played a crucial role in past vegetation dynamics, either as refugia for warm-demanding species during the full Glacial or for light-demanding species during the Holocene forest maximum (Fekete et al., 2010). For the Carpathians, which border the Pannonian Basin in the north and east, the presence of taiga and hemiboreal forests during the full Glacial has been suggested by both pollen and macrofossil evidence (Jankovská & Pokorný, 2008; Kuneš et al., 2008). Hemiboreal coniferous forests, currently found in easternmost Europe and southern Siberia, harbour many species of semi-dry grasslands and are often part of a continental type of forest steppe (Chytrý et al., 2008, 2010). The hemiboreal forest steppes of southern Siberia share many species with the Central European meadow steppes (Roleček et al., 2014). Therefore, the hemiboreal forest steppe presumably covering the foothills of the Carpathians during the glacial period might have provided suitable habitats for this group of steppe species.

In our analysis, climate and topographic heterogeneity explained the number of habitat specialists in a region quite well for grass steppes and rocky steppes, but only moderately for meadow steppes. The latter are more widespread in our study area and occur on less extreme sites, so they might be less dependent on specific topography or climate. Adding geographical distance to potential refugia increased model performance for all three steppe types, suggesting that species distributions are still influenced by migration lags, that is limited migration from primary LGM and/or secondary mid-Holocene refugia. The most pronounced increase in model performance was found when adding geographical distance to the Hungarian Central Range, western Podolia and the north-western margin of the Pannonian Basin (refugia 1–4 in Figure 1a). Indeed, all three steppe types have regional character species maxima in the north-western half of the Pannonian Basin (Figure 2), a pattern that seems difficult to explain by recent environmental factors alone. It should be considered, however, that in the absence of migration lags, the regional diversity of habitat specialists would probably bear no signals of refugia. Therefore, independent evidence from molecular data is crucial.

The haplotype diversity of the studied species was well within range of the few previously published values for steppe species (e.g. *Adonis vernalis*—mean 0.39, max. 0.87, Kropf et al., 2020; *A. onobrychis*—mean 0.82, max. 1.0, Plenk et al., 2020; *Carlina acanthifolia* subs. *utzka*—mean 0.24, max. 0.87, Ciesiak & Drobnjak, 2019; *Linum flavum*—mean 0.84, max. 1.0, Plenk et al., 2017; *P. badensis*—mean 0.13, max. 0.52, Plenk et al., 2019). In concordance with the findings of Magyari et al. (2010), we can reject the scenario of a late-Holocene unidirectional migration from refugia near the Black Sea coast (scenario a) for all analysed species. Indeed, all of them either have unique haplotypes in the north-western Pannonian region (*Filipendula vulgaris*, *R. bulbosus*, *Stipa capillata*, *P. badensis*, *Scorzonera austriaca*) or equally high haplotype diversity in the north-west and south-east (*Salvia nemorosa*, *F. procumbens*) or completely different haplotype groups in different parts of the study area (*T. montanum*, *A. onobrychis*, *L. austriacum*, *L. tenuifolium*) (see Figure 3 and Appendix S7). We therefore conclude that these species have been present in the Pannonian Basin since at least the early Holocene. In concordance with Kuneš et al. (2008, 2015), Roleček et al. (2014) and Feurdean et al. (2018), we suggest that forest steppes with a strong meadow steppe component and patches of rocky and grass steppes persisted in the lower mountain ranges surrounding the Pannonian Basin throughout the Holocene. More difficult, however, is the question whether these regions represent primary refugia (scenario c) or were colonised during the Late Glacial or early Holocene, that is representing only secondary refugia (scenario b).

4.2 Scenarios for individual steppe types

The haplotype diversity maximum of *Filipendula vulgaris* and *R. bulbosus* in Transylvania and of *T. montanum* near the Iron Gate (south-western Romania), including unique and divergent haplotypes in each of the three species, most likely indicate primary refugia, suggesting a long-term continuity of meadow steppe flora...
in the southern Carpathians and adjacent regions. This is in line with the fossil evidence of full-glacial hemiboreal forests in the Carpathians (Jankovská & Pokorný, 2008; Kuneš et al., 2008). However, the strong geographical west-east differentiation observed in *T. montanum*, and the presence of unique haplotypes in *F. vulgaris* and *R. bulbosus* provide evidence for additional primary refugia of meadow steppe species in the western half of the Pannonian Basin and in the north-western Dinaric mountains.
We have not analysed the genetic structure of populations from western Podolia, but the analysis of habitat specialists supports the existence of at least a secondary refugium in this region as well.

The habitat specialists of grass steppes show a pattern quite different from meadow steppes, with a character species maximum in the most continental part of the study area in southern Ukraine (Figure 2). This finding is well explicable by the current climatic conditions. The Pannonian Basin is usually considered as part of the forest steppe biome (Erdős et al., 2018; Wesche et al., 2016). Therefore, grass steppes are extrazonal in this region. Some species with pronounced continental distribution might have been more widespread in Central Europe during the early Holocene, the Late Glacial or even the LGM, but gone extinct when the climate became more humid. However, our data support the hypothesis that at least some of these species persisted throughout the Holocene in the Pannonian Basin. Indeed, the strong geographical haplotype differentiation in A. onobrychis (Appendix S7; see also Kirschner et al., 2020; Plenk et al., 2020; Záveská et al., 2019) and L. austriacum (Figure 3d) suggest that these two species have been continuously present in the western Pannonian Basin since pre-LGM times. Thus, a primary refugium of grass steppes may be postulated for this region. The genetic data of Stipa capillata and Salvia nemorosa, which are characterised by more shallow structuring, do not contradict such a scenario. The slightly higher genetic diversity in south-western Russia that was observed in these two species might hint at an old east-west expansion (probably pre-dating the LGM; see also Kirschner et al., 2020). However, it could also result from the mid-Holocene bottleneck, which probably only affected the populations in Central Europe (Wagner et al., 2011).

Rocky steppes often occur on sites unsuitable for forests. Therefore, the mid-Holocene forest maximum probably did not pose a threat to their (local) survival (Feurdean et al., 2018; Poschlod & WallisDeVries, 2002). Most of their habitat specialists have a more or less submediterranean distribution. Accordingly, species diversity hotspots of this group were found in the southern part of our study area. However, almost equally high species numbers were found in the Hungarian Central Range and along the eastern margin of the Alps (Figure 2). The increase in model performance when adding geographical distance to these regions, as well as the strong geographical haplotype differentiation in F. procumbens (Figure 3f) and Linum tenuifolium (Figure 3g), indicate primary refugia of rocky steppes also in the centre and north-western periphery of the Pannonian Basin. This finding

(Figure 3a–c).
might seem at odds with the fact that most members of this species group have a submediterranean distribution, being absent in continental regions representing modern analogues of Central European glacial conditions (such as southern Siberia; Chytrý et al., 2019). However, several character species of rocky steppes reach the upper montane belt of the Carpathian Mountains where they get into contact with subalpine larch forests (Kliment et al., 2005). Therefore, these species seem well adapted to severe winter frosts, and their absence from strongly continental regions might have other (e.g. historical) reasons than their physiological amplitude.

4.3 | Habitat specialists and haplotypes

Unexpectedly, maximum haplotype diversity among the studied species of a given steppe type showed better correlation with the number of habitat specialists in a region than mean haplotype diversity. A possible explanation for this might be the sampling design. Genetic data are derived from eight individuals of a single population within the grid cell and, therefore, will hardly represent the full haplotype diversity present in this region, whereas the number of habitat specialists represents a complete census of the grid cell. Thus, the maximum value of several species likely gives a better approximation of the full genetic diversity present than the average values, let alone the haplotype diversity of individual species.

5 | CONCLUSIONS

Our findings support the crucial role of the lower mountain ranges surrounding the Pannonian Basin as providing both primary and secondary refugia for European steppe species. Topographic heterogeneity is probably one of the most important prerequisites for a region to become a refugial area. The high predictive power of this parameter in explaining the number of habitat specialists in a region could reflect historical contingencies just as current ecological conditions. Dispersal limitation and resulting migration lags probably influence the distribution patterns of steppe species in a similar way as the species ranges in other vegetation types such as deciduous forests. Therefore, geographical distance from refugia is a factor to be taken into account when explaining the regional diversity of steppe specialists. Chloroplast haplotypes provide important evidence for past range dynamics and location of refugia. The number of steppe species studied with phylogeographical methods is still unsatisfactory, and future studies using next-generation sequencing techniques will undoubtedly provide new and more detailed insights. We believe that the results presented here will help create a consistent picture of the long-term development of steppes in western Eurasia.

ACKNOWLEDGEMENTS

The authors thank Sándor Barabás, Katharina Bardy, Yuriy Burlakov, Monika Kriechbaum, Bertalan Lendvay, Nicole Penke, Nikita Tikhomirov, Polína Volkova, Slavica Vuković, Klemens Wernisch, Michaela Wernisch and Silvia Winter for help with collecting plant material, and the authorities (Burgenland no. 5/N.AB-10207-15-2016; Lower Austria no. RU5-BE-370/004-2016; Rostov Nature Reserve no. 02-01/86) for granting us permission. Plant leaves were preferentially collected outside the protected areas. Stefanie Edelmüller and Gerhard Kadlec helped with DNA extraction. W.W., K.P. and M.K. were supported by the Austrian Science Fund (FWF) (P 27955-B29). A.K. and D.V. were supported by the National Research Foundation of Ukraine (project no. 2020.01/0140). J.R. was supported by the Czech Science Foundation (project 20-09895S) and the long-term developmental project of the Czech Academy of Sciences (RVO 67985939).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Species distribution data are included in the Supporting Information. Haplotype sequences were stored at the EMBL Nucleotide Archive (www.ebi.ac.uk/ena) under the following accession numbers. Astragalus onobrychis: atpI-H (LR760132-38, -43-45, -50, -52-54), rpL16 intron (LR760231-49, -61-69, -84-90), ycf1 (LR760299-304, -09-12, -19-23); Filipendula vulgaris: atpI-H (HG993770-73), rpL16 intron (HG993774-83); Fumana procumbens: atpI-H (HG994045-47), rpL32-trnL (HG994053-67); Linum austriacum: rpL16 intron (HG993736-52), rpL32-trnL (HG993753-69); Linum tenuifolium: atpI-H (HG993885-96), rpL16 intron (HG993897-913); Poa badensis: atpI-H (HG994082-90), rpL16 intron (LR735975-76); Ranunculus bulbosus: atpI-H (HG993938-48), rpL32-trnL (HG993814-47); Scorzonera austriaca: rpL16 intron (HG993410-16); rpL32-trnL (HG993248-55); Stipa capillata: atpI-H (HG994048-76), rpL16 intron (HG994048-52); Salvia nemorosa: rpL16 intron (HG99339-401); Trifolium montanum: atpI-H (HG994029-44).

ORCID

Wolfgang Willner https://orcid.org/0000-0003-1591-8386
Jan Roleček https://orcid.org/0000-0002-0023-1111

REFERENCES

Aćić, S., Šilc, U., Petrović, M., Tomović, G., & Dajić Stevanović, Z. (2015). Classification, ecology and biodiversity of Central Balkan dry grasslands. Txexenia, 35, 329–353.
Avise, J. C. (2000). Phylogeography - The history and formation of species. Harvard University Press.
Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A., Bezrukova, E., Bljakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J. O., Novenko, E., Ryabogina, N., Solovieva, N., Willis, K., & Zrnitskaya, V. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. Quaternary Science Reviews, 157, 80-97. https://doi.org/10.1016/j.quascirev.2016.11.022
Birks, H. J. B., & Birks, H. H. (2004). The rise and fall of forests. Science, 305, 484–485.
Birks, H. H., & Birks, H. J. B. (2006). Multi-proxy studies in palaeolimnology. Vegetation History and Archaeobotany, 15, 235–251. https://doi.org/10.1007/s00334-006-0066-6
BIOSKETCH

Wolfgang Willner is a vegetation scientist interested in the current and historical biogeographical factors shaping the species composition of plant communities, with a special focus on Eurasian steppes and temperate forests.

Author contributions: W.W. and M.K. designed the research; W.W., S.A., A.K., J.R., K.V. and D.V. compiled the species distribution data; M.K., K.P., O.D., M.H., A.K., K.V. and W.W. collected plant material; K.P. and M.K. analysed the molecular data; D.M. performed most of the statistical analyses; W.W. led the writing of the manuscript with significant contributions by M.K., D.M., K.P. and J.R.; all authors commented on earlier versions of the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of the article at the publisher’s website.

How to cite this article: Willner, W., Moser, D., Plenk, K., Aćić, S., Demina, O. N., Höhn, M., Kuzemko, A., Roleček, J., Vassilev, K., Vynokurov, D., & Kropf, M. (2021). Long-term continuity of steppe grasslands in eastern Central Europe: Evidence from species distribution patterns and chloroplast haplotypes. Journal of Biogeography, 48, 3104–3117. https://doi.org/10.1111/jbi.14269