ORIGINAL RESEARCH

Ecology and Evolution
WILEY

Pleistocene dynamics of the Eurasian steppe as a driving force of evolution: Phylogenetic history of the genus *Capsella* (Brassicaceae)

Anže Žerdoner Čalasan[1] | Herbert Hurka[1] | Dmitry A. German[2] | Simon Pfanzelt[3,4] | Frank R. Blattner[3] | Anna Seidl[5] | Barbara Neuffer[1]

[1]Department 5: Biology/Chemistry, Botany, University of Osnabrück, Osnabrück, Germany

[2]South-Siberian Botanical Garden, Altai State University, Barnaul, Russia

[3]Experimental Taxonomy, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland-Gatersleben, Germany

[4]Munich Botanical Garden, München, Germany

[5]Institute of Botany, Department of Integrative Biology and Biodiversity Research, University of Natural Resources and Life Sciences, Vienna (BOKU), Austria

Correspondence
Anže Žerdoner Čalasan, Department 5: Biology/Chemistry, Botany, University of Osnabrück, Barbarastraße 11, 49076 Osnabrück, Lower Saxony, Germany.
Email: azerdonercal@uos.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: NE 314/15-1

Abstract

Capsella is a model plant genus of the Brassicaceae closely related to *Arabidopsis*. To disentangle its biogeographical history and intrageneric phylogenetic relationships, 282 individuals of all five currently recognized *Capsella* species were genotyped using a restriction digest-based next-generation sequencing method. Our analysis retrieved two main lineages within *Capsella* that split c. one million years ago, with western *C. grandiflora* and *C. rubella* forming a sister lineage to the eastern lineage consisting of *C. orientalis*. The split was attributed to continuous latitudinal displacements of the Eurasian steppe belt to the south during Early Pleistocene glacial cycles. During the interglacial cycles of the Late Pleistocene, hybridization of the two lineages took place in the southwestern East European Plain, leading to the allotetraploid *C. bursa-pastoris*. Extant genetic variation within *C. orientalis* postdated any extensive glacial events. Ecological niche modeling showed that suitable habitat for *C. orientalis* existed during the Last Glacial Maximum around the north coast of the Black Sea and in southern Kazakhstan. Such a scenario is also supported by population genomic data that uncovered the highest genetic diversity in the south Kazakhstan cluster, suggesting that *C. orientalis* originated in continental Asia and migrated north- and possibly eastwards after the last ice age. Post-glacial hybridization events between *C. bursa-pastoris* and *C. grandiflora/rubella* in the southwestern East European Plain and the Mediterranean gave rise to *C. thracica*. Introgression of *C. grandiflora/rubella* into *C. bursa-pastoris* resulted in a new Mediterranean cluster within the already existing Eurasian *C. bursa-pastoris* cluster. This study shows that the continuous displacement and disruption of the Eurasian steppe belt during the Pleistocene was the driving force in the evolution of *Capsella*.

KEYWORDS
evolution, hybridization, introgression, refugium, shepherd's purse

1 | INTRODUCTION

Capsella Medik. is a small genus within the mustard family (Brassicaceae). Of all wild relatives of the model system _Arabidopsis_ (DC.) Heynh., _Capsella_ is the most closely related and intensively investigated genus. Both genera belong to the tribe Camelineae, which also includes _Camelina_ Crantz and four more genera (https://brassibase.cos.uni-heidelberg.de). The three named genera have become model systems for plant molecular research, as well as the study of genome evolution, and the evolution and development of genotypic and phenotypic traits.

Species delimitation within _Capsella_ is difficult because of the enormous morphological variation especially in leaf and fruit characters, and in hybridization potential studied intensively as early as in the 1900s (Almquist, 1907; Shull, 1909). In most cases, five species are accepted (Chater in Flora Europaea 1964; Koch et al., 2018): _Capsella grandiflora_ (Fauché & Chaub.) Boiss., _C. rubella_ Reut., _C. orientalis_ Klokov, _C. bursa-pastoris_ (L.) Medik., and _C. thracica_ Velen. Less often, some of the species are treated at infraspecific rank, for example, _C. thracica_ (Chater, 1993; Stojanov & Stefanov, 1948) or _C. orientalis_ (Tzvelev, 2000a) as a subspecies of _C. bursa-pastoris_. An ultimately lumping approach recognizes the genus as monotypic, comprising the five abovementioned species (Appel & Al-Shehbaz, 2003; Svensson, 1984). The wealth of available morphological and molecular evidences supports acceptance of the above five entities as distinct species. The first three species are diploid ($2n = 2x = 16$), and the latter two are tetraploid ($2n = 4x = 32$). _Capsella grandiflora_ is an obligate outcrossing species with a sporophytic self-incompatibility system (Paetsch et al., 2006; Riley, 1936), whereas all other species are predominantly selfing.

The evolutionary history of the genus _Capsella_ has been controversial for a long time. However, by including the commonly neglected species _C. orientalis_ and _C. thracica_ in molecular phylogenetic analyses, Hurka et al. (2012) discovered the "missing link" in the understanding of the origin of tetraploid _C. bursa-pastoris_ and provided evidence that in _Capsella_ the transition from outbreeding to selfing at the diploid level occurred twice independently. This was later confirmed by several other studies, for example, Douglas et al. (2015), Bachmann et al. (2019), and Žerdoner Čalasan et al. (2019). The most recent common ancestor (MRCA) of _Capsella_ is presumably of Late Pliocene/Early Pleistocene age and was probably distributed along the Eurasian steppe belt from eastern Europe to western or even Central Asia (Hurka et al., 2012). During the Pleistocene, _Capsella_ split into two lineages, the eastern and the western lineage. The eastern lineage comprises the extant species _Capsella orientalis_ and the western lineage the extant species _Capsella grandiflora_ and _C. rubella_. _Capsella bursa-pastoris_ and _C. thracica_ emerged from hybridization between the western and eastern lineages. Meanwhile, _Capsella_ has become a model system for genomic studies of plant mating systems (e.g., Bachmann et al., 2019, 2021; Brandvain et al., 2013; Mattila et al., 2020; Woźniak et al., 2020) and polyploidization (Douglas et al., 2015; Han et al., 2015; Kryvokhyzha, Salcedo et al., 2019;

Omelchenko et al., 2020). In contrast to the increasing knowledge about _Capsella_ at the genomic level, a detailed phylogeographical aspect of its evolution, especially of the eastern lineage consisting of _C. orientalis_, is currently missing. _Capsella orientalis_ was first described by Klokov in 1926 and for decades eked out a niche existence in botanical research. While the first indications of its diploid nature are dated to the early 2000s (Dorofeyev, 2002; Tzvelev, 2000b), it was only in 2012 when this species was unambiguously placed into the wider evolutionary time frame of _Capsella_ and its distribution range properly delimited (Hurka et al., 2012).

Capsella orientalis ranges from southeastern Europe, the South Urals, northern and eastern Kazakhstan, and southwestern Siberia to the northern part of the Chinese Xinjiang province and Mongolia. Its distribution thus coincides with the Eurasian steppe belt, namely its Euro-Siberian part (Hurka et al., 2012). Prototypes of the Eurasian steppe appeared during the Lower Miocene (23–16 Ma) in Central Asia. Continuous global cooling through the Middle and Upper Miocene accompanied by the uplift of surrounding mountain ranges and the retreat of water bodies have initiated and facilitated the aridification of Central Asia (Hurka et al., 2019 and references therein; Barbolini et al., 2020). This in turn promoted the expansion of drought-adapted steppe forbs and grasses. Spreading toward the west, the Eurasian steppe belt reached the East European Plain in Late Miocene (Akhmetyev et al., 1999; Velichko, 1999; Velichko et al., 2005; 2011b) and expanded in Eastern Europe during the Pliocene (Lang, 1994; Mai, 1995). The Eurasian steppe belt was under strong influence of the Pleistocene glaciations that caused extensive expansions and contractions as well as latitudinal shifts and longitudinal splits of the ranges of steppe plant species during the past 3 million years (Hurka et al., 2019 and references therein). It has been shown that such a biome dynamic is mirrored by molecular signals of typical Eurasian steppe plants (e.g., _Capsella orientalis_) and thus reflects their biogeographical history (Buono et al., 2021; Franzke et al., 2004; Friesen et al., 2016, 2020; Hantemirova et al., 2020; Hurka et al., 2012; Seidl et al., 2019, 2021; Seregin et al., 2015; Volkova et al., 2017).

Using a restriction digest-based NGS method of genotyping-by-sequencing (GBS) on a representative taxon sample of 282 _Capsella_ individuals, we investigated whether and how the Pleistocene glaciations influenced the short but turbulent history of the genus _Capsella_. Furthermore, in accord with Fernández-Mazuecos et al. (2018) we tested to what extent filtering and SNP calling may influence the retrieved topologies and admixture as well as genetic clustering of tested populations. To investigate whether Pleistocene glaciations influenced the genetic structure of the steppe element _C. orientalis_, we calculated _F_-statistics and molecular diversity indices, and conducted an analysis of molecular variance (AMOVA). While Altai has been commonly recognized as a refugial area for steppe flora and fauna, we wanted to assess via ecological niche modeling whether this mountain range had the same refugial role in case of younger taxa, such as _C. orientalis_.

2 | MATERIALS AND METHODS

Seeds of 5–10 randomly selected plants per population collected in the wild (Figure S1) were cultivated in the greenhouse of the University of Osnabrück, Germany (Figure 1). We investigated 18 populations of *Capsella bursa-pastoris*, seven populations of *C. grandiflora*, 57 populations of *C. orientalis*, five populations of *C. rubella*, and four populations of *C. thracica*.

Plants were grown to the flowering stage and ploidy levels were estimated through flow cytometry to minimize species determination errors. Genome size estimations followed the protocols described in Doležel et al. (2007) with some modifications: 1 cm^2 of young fresh leaf tissue was chopped together with 1 cm^2 of tissue from the internal standard and 0.4 ml of cold CyStain® Nuclei Extraction Buffer in a glass petri dish with a razor blade. The cell suspension was vortexed briefly and then stained with 1.6 ml CyStain® Staining Buffer containing 4',6-diamidino-2-phenylindole (DAPI). The samples were incubated for 30 s, mixed gently, and filtered through a 50 µm CellTrics® filter. Flow cytometry was performed with the CyFlow® Ploidy Analyser Typ 24, 2010 (Sysmex Partec GmbH, Germany) with the following settings: gain: 540 V, velocity: 0.4 µl/s, 365 nm UV-LED, 532 nm excitation, and 532 nm emission. Samples were run until 5,000 cells per peak were measured with each measurement repeated thrice (CV < 5%). *Petroselinum crispum* (Mill.) Fuss served as an internal standard (4.46 pg/2C, $SD \pm 0.08$ pg/2C: Yokoya et al., 2000) for diploid as well as for polyploid species. Up to 1 g of fresh leaves were silica-dried and stored at room temperature. Subsequent DNA isolation was carried out using DNeasy Plant Mini Kit (Qiagen), following the manufacturer's protocol.

Genotype-by-sequencing library preparation followed protocols from Poland et al. (2012), and Meyer et al. (2008). We performed single-end sequencing of 100 bp fragments of 282 samples (57 populations) in two runs on a HiSeq 2500 (Illumina, CA, USA) at the Leibniz Institute of Plant Genetics and Crop Plant Research (Gatersleben, Germany). To allow for representative coverage, the sequenced DNA amount of tetraploid *C. bursa-pastoris* and *C. thracica* accessions was doubled. To assess the reproducibility of the sequencing method internally, three samples were loaded twice on the flow cell (depicted with * in Figure S1).

2.1 | SNP filtering/calling

Already demultiplexed raw Illumina DNA sequence data files were analyzed using *ipyrad* v0.9.52 (Eaton & Overcast, 2020). Two separate sub-datasets were analyzed, one containing all *Capsella* species (282 accessions; caps dataset) and one encompassing only *C. orientalis* accessions (235 accessions; ori dataset). Although there are already published reference genomes of *C. rubella* and *C. grandiflora* (Slotte et al., 2013) available, we assembled both of the sub-datasets de novo to prevent SNP (single nucleotide polymorphism) recovery bias. Several analyses were run to test the influence of different parameter settings on the final SNP data matrix. The clust_threshold option was set in parallel to 0.85, 0.90, and 0.95, recognizing reads as homologous if the similarity exceeded 85%, 90% and 95%, respectively. The min_sample_locus was set in a way that only sites with at least 50%, 25%, or 12.5% of data were retained in the final alignment, respectively. The *Capsella* dataset consisted of two tetraploid species; thus, the max_alleles_consens option was set to 4, while this parameter was left at the default value of 2 for the analysis of the *Capsella orientalis* dataset consisting of purely diploid accessions. After initial testing, we set the max_SNPs_locus parameter to 0.1 and 0.05, respectively, in the most stringently filtered datasets (i.e., 0.95 with at least 50% of SNP data per position). The default value of 0.2, indicating that a locus might consist of up to 20% of SNPs, might be set inappropriately high for such a young plant taxon. In total, we analyzed 22 different datasets. All subsequent analyses were carried out using the most stringently filtered datasets with the clustering threshold set to 95%, including SNP positions with at least 50% of available information and allowing for a SNP maximum of 5% per locus.

2.2 | Phylogenetic inference

We performed a SVDQuartets analysis (Chifman & Kubatko, 2014) in PAUP* v4.0a (Swofford, 2002) first using all 282 *Capsella* accessions (unreduced SVDQ dataset). However, due to extremely high admixture recovered from the population genetic structure analyses, another analysis was carried out including only accessions showing

FIGURE 1 Habitus of *Capsella orientalis* Klokov. Whole plant (left) and inflorescence (right) of *Capsella orientalis* (Astrakhan region, Bogdinsko-Baskunchaksky nature reserve, Bolshoy Bogdo, cereal-forb steppe at the southern foot of the mountain; 4 May 2011; by M. S. Knyazev)

more than 95% association to a specific cluster (i.e., exhibiting <5% admixture), reducing the dataset from 282 accessions to 57 (reduced SVDQ dataset). To verify the monophyly of species, no accession-to-species attribution was carried out prior to the analyses. We evaluated all possible quartets in the reduced dataset and 10,000 quartets in the unreduced one. The handling of the ambiguities was set to "distribute." QFM was specified as the quartet assembly algorithm. Bootstrapping was performed with the number of replicates set to 1,000. Final trees were exported with saved internal nodes and visualized with FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

2.3 | Time divergence estimation

Time divergence analysis was carried out in SNAPP (Bryant et al., 2012) coupled with a molecular clock model (Stange et al., 2018), which is implemented as an add-on package in BEAST2 (Bouckaert et al., 2019). The SNAPP input file was generated using a ruby script developed by M. Matschiner (available at: https://raw.githubusercontent.com/mmatschiner/snapp_prep/master/snapp_prep.rb). An unlinked *Capsella* SNP dataset generated from ipyrad was used as an input file; however, due to the high computational intensity, further reduced to 25 accessions, including all major taxon groups as retrieved in the SVDQ analysis. The 25 accessions were selected so that the genetically most divergent accessions within selected species and/or genotypes were represented, covering the whole currently known distribution area. We provided a newly reconstructed starting tree from an SVDQ analysis but without a specified out-group to avoid artificial topologies. We made use of four different secondary calibration points that have been independently inferred from the following unrelated studies using different calibration methods, different taxon samples, and different loci: Douglas et al., 2015, Han et al., 2015, Hurka et al., 2012 and Žerdoner Čalasan et al., 2019. The calibration node ages were specified as log-normally distributed in real space. We performed 20,000,000 Markov-chain Monte Carlo (MCMC) iterations per SNAPP analysis, and the stationarity of MCMC chains was assessed using Tracer v1.7.0 (Rambaut et al., 2018). Trees generated in three independent analyses were combined in LogCombiner v.2.5.2 with an additional burn-in of 10%, and the final tree was generated using TreeAnnotator v2.5.2 using the "maximum clade credibility tree with median heights" option. All phylogenetic analyses were carried out at the CIPRES Science Gateway computing facility (Miller et al., 2010).

2.4 | Inference of genetic structure

Population genetic structure was assessed using the clustering algorithm implemented in STRUCTURE v.2.3.4 (Pritchard et al., 2000) by carrying out a run with length of 10,000 MCMC iterations with a burning period of 2,000. For the *Capsella* dataset, we tested a range of K-values from $K = 2$ to $K = 10$ and repeated the analysis 20 times.

For the *Capsella orientalis* dataset, we tested a range of K-values from $K = 2$ to $K = 5$ with the same number of iterations as specified above. All settings were left at default with the exception of the correlated allele frequencies and admixture options. Due to previous experience, we expected notable introgression. Consistency across replicate cluster analyses was assessed with CLUMPP v.1.1.2. (Jakobsson & Rosenberg, 2007) by using Greedy algorithm and 1,000 random input orders. K probability was assessed using the Evanno method (Evanno et al., 2005) as implemented in Structure Harvester (Earl & von Holdt, 2012). Population specific ancestry coefficients were calculated as mean values of the individuals and shown as pie charts plotted on the map. In addition, population genetic structure was also inferred using a non-Bayesian clustering algorithm, namely sparse non-negative matrix factorization as implemented in the R package LEA (an R package for Landscape and Ecological Association Studies: Frichot & François, 2015; R Core Team, 2013). Twenty hypothetical ancestral populations were tested with 50 repetitions each for both datasets, and the K probability was assessed using the cross-entropy approach incorporated in the LEA package. Population specific ancestry coefficients were calculated and depicted as specified above. To confirm the result, additional DAPC analyses for the *Capsella* as well as *C. orientalis* datasets were carried out under default parameters using the adegenet R package (Jombart, 2008).

2.5 | Ecological niche modeling and assessment of the geographical range

To infer the climatic suitability of habitats of *Capsella orientalis* through space and time, ecological niche modeling was employed and implemented in MaxEnt v3.3.4 (Phillips et al., 2006). A set of 19 bioclimatic variables were retrieved from the WorldClim database (http://worldclim.com/paleo-climate1: Hijmans et al., 2005) at a spatial resolution of 2.5 arc minutes (approximately 4.5 km^2 at the equator) and for three different time slices: present, Mid Holocene (8.33–4.2 ka), and Last Glacial Maximum (ca. 21 ka). The geographical distribution of *Capsella orientalis* was assessed based on geographical range information from the literature (Ebel, 2002; German, 2015; German et al., 2012; German & Ebel, 2009; Iljinska et al., 2007; Jalas & Suominen, 1994; Kulikov, 2005; Sheremetova et al., 2011; Verkhozina et al., 2020), Global Biodiversity Information Facility (GBIF: https://www.gbif.org), and herbarium and own field data. The final dataset consisted of 121 entries (Figure 2 and Figure S2). Variable exclusion via cross-correlation was carried out to assure that the used bioclimatic variables were uncorrelated (Pearson's correlation coefficient <0.5). Selected variables were bio4 (temperature seasonality), bio5 (maximum temperature of the warmest month), and bio12 (annual precipitation). To maximize predictive ability and avoid model overfitting, the number of climate variables was kept to a minimum. Models were constructed using Maxent v3.4.1 (Phillips et al., 2006) and evaluated with the R package ENMeval (Muscarella et al., 2014).
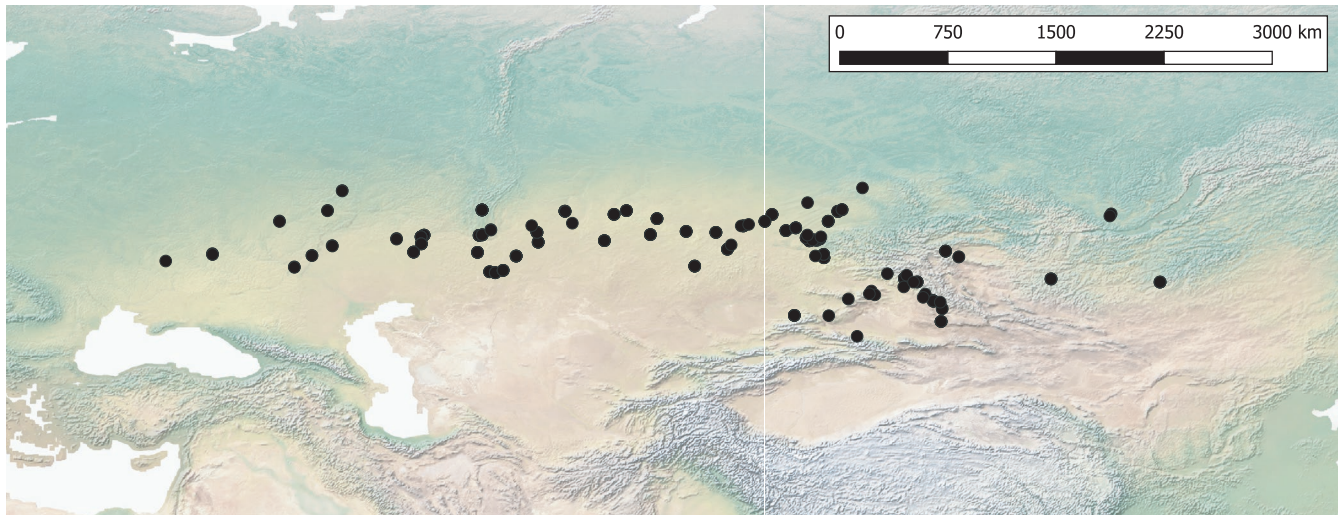
FIGURE 2 Distribution map of *Capsella orientalis* based on geographical range information from the literature (Ebel, 2002; German & Ebel, 2009; German et al., 2012; German, 2015; Han et al., 2015; Hurka et al., 2012; Iljinska et al., 2007; Jalas & Suominen, 1994; Kryvokhyzha, Salcedo et al., 2019; Kulikov, 2005; Sheremetova et al., 2011; Verkhozina et al., 2020), our own field data, and a priori critically assessed GBIF collection points

2.6 | Population genetics analysis

Arlequin v3.5 (Excoffier & Lischer, 2010) was used to determine the genetic diversity of *Capsella orientalis* populations, calculate *F*-statistics and distribution of private alleles, and assess the genetic differentiation on the cluster- as well as population-level via AMOVA (Analysis of molecular variance). Molecular diversity indices and private allele sharing were assessed with the complete ori SNP dataset retrieved directly from the ipyrad analysis, while F-statistics and AMOVA were carried out on a *.loci-dataset also retrieved directly from the ipyrad workflow. The significance of the test results was based on 1,023 permutations. We categorized the populations into five genetic clusters corresponding to the ancestral populations inferred from the LEA analysis. For AMOVA, we used only five populations per genetic cluster/geography, as this was the lowest possible number of investigated populations per genetic cluster/geographical region (only five populations belonged to the Central Asian genetic cluster/region). Five geographically most distant populations were selected for genetic clusters that consisted of more than five populations.

3 | RESULTS

3.1 | SNP filtering/calling

The final analyses were carried out using the *Capsella* (caps) dataset comprising 282 accessions representing different *Capsella* species and *Capsella orientalis* (ori) dataset comprising 235 accessions of *C. orientalis*. Furthermore, the reads had to be at least 95% similar to be recognized as homologous and only positions where at least 50% of samples had a recorded base were retained in the final alignment. Maximum percentage of allowed SNPs per locus retained in the final

alignment was set to 5%. Combination of these parameters consistently outperformed other parameter assemblages (Figure S3), in which no population structure could have been inferred throughout different *K*-values (data not shown). The final *Capsella* dataset consisted of 12,635 SNPs and 13.17% missing sites, while the final *Capsella orientalis* dataset included 4,644 SNPs with 24.23% of missing data. The alignment used for time divergence estimation consisted of 25 entries, 16,398 SNPs (out of which 6,320 were retrieved in the unlinked SNP dataset) and 20.07% missing sites. Further statistics can be inferred from the File S2.

3.2 | Phylogenetic inference

The SVDQ analysis of the nonreduced dataset (Figure S4) retrieved only poorly supported species-specific clades of *Capsella rubella* (57 BS) and *C. grandiflora* (37 BS) but well-supported species-specific clades of *C. orientalis* (100 BS) and *C. bursa-pastoris* (94 BS). Within the latter, the Mediterranean *C. bursa-pastoris* was monophyletic (93 BS), while the Eurasian *C. bursa-pastoris* and *C. thracica* were retrieved as poorly supported clade and grade, respectively (Figure S4). *C. grandiflora* and *C. rubella* were defined as the out-group to the rest of *Capsella* species (100 BS), following the topologies of Douglas et al. (2015), Guo et al. (2017), Hurka et al. (2012), Kryvokhyzha, Salcedo et al. (2019), and Walden et al. (2020). The analysis of the reduced dataset (i.e., dataset with accessions exhibiting <5% admixture) including only 57 accessions retrieved the same topology but with a better bootstrap support within *C. orientalis* (Figure S5). The following clades were inferred: *Capsella rubella* (100 BS), *C. grandiflora* (77 BS), Mediterranean *C. bursa-pastoris* (100 BS), *C. thracica* (96 BS), Eurasian *C. bursa-pastoris* (80 BS), South Kazakhstan *C. orientalis* (83 BS), Tabagartai–Dzhungarian Alatau *C. orientalis* (100 BS), Central Asian *C. orientalis* (100 BS), North Kazakhstan *C. orientalis*

(73 BS), and Altai *C. orientalis* (98 BS). The only well-supported sister group relationship was inferred between Altai and North Kazakhstan *C. orientalis* (100 BS). For easier depiction of the phylogenetic relationships within *Capsella*, a scheme based on the topology retrieved from the unreduced SVDQ analysis was reconstructed (Figure 3).

3.3 | Time divergence estimation

All four time divergence analyses led to congruent results, despite the single secondary calibration points originating from different unrelated studies with different taxon sampling and genetic datasets.
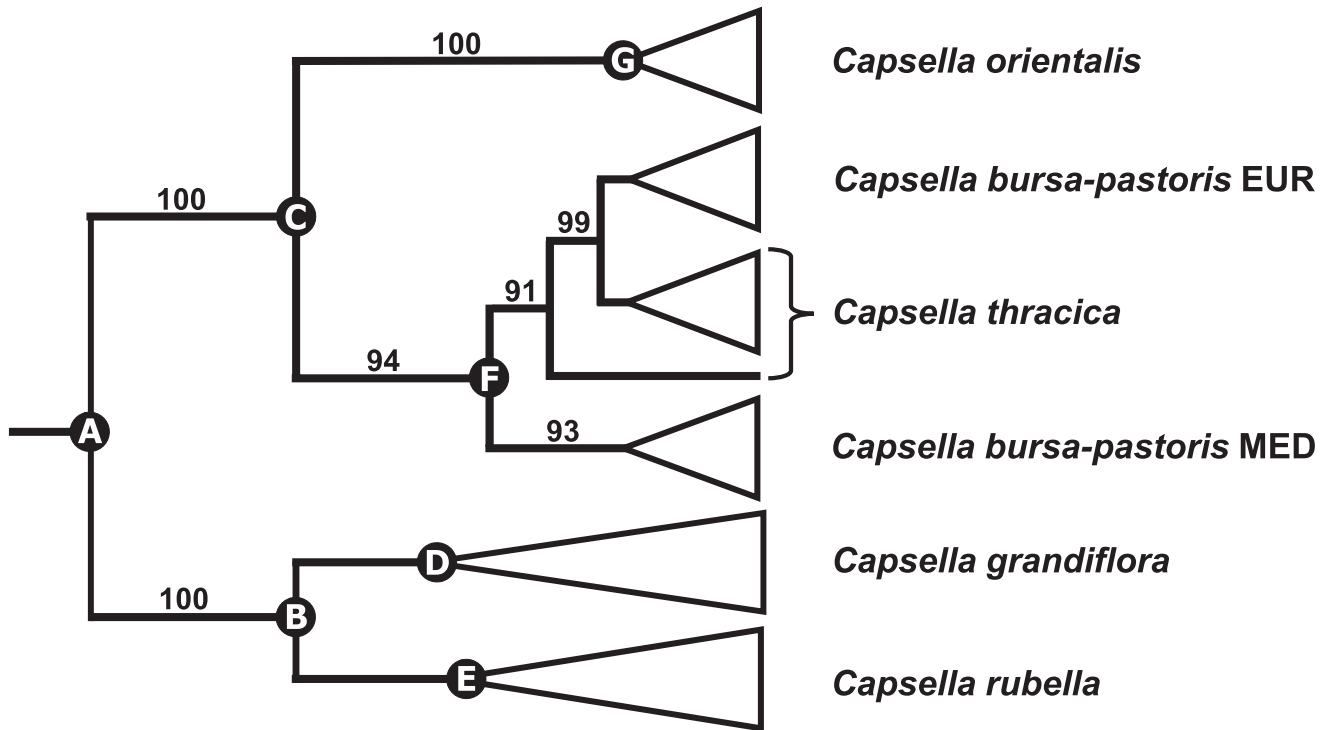
	Node	Calibration I	Calibration II	Calibration III	Calibration IV
Capsella crown age	A	**0.16–0.54–1.10 MYA**	0.07–0.69–3.22 MYA	**0.16–0.93–1.12 MYA**	0.61–0.80–0.98 MYA
Cgr/Cru crown age	B	0.05–0.15–0.31 MYA	0.03–0.21–1.03 MYA	0.08–0.14–0.23 MYA	0.19–0.24–0.31 MYA
Cbp/Cor crown age	C	0.05–0.13–0.28 MYA	**0.01–0.87–2.44 MYA**	0.07–0.12–0.21 MYA	**0.09–0.22–0.44 MYA**
Cgr crown age	D	18.0–55.0–114.6 KYA	10.2–75.0–362.9 KYA	25.1–49.5–86.6 KYA	62.6–87.6–111.8 KYA
Cru crown age	E	4.9–22.0–51.8 KYA	2.9–28.5–135.8 KYA	7.5–19.7–37.0 KYA	16.6–34.9–56.2 KYA
Cbp crown age	F	1.4–5.5–12.4 KYA	0.7–7.6–36.7 KYA	1.6–4.9–8.9 KYA	4.8–8.6–14.1 KYA
Cor crown age	G	0.5–1.9–4.9 KYA	0.3–2.1–12.0 KYA	0.5–1.6–3.4 KYA	1.3–2.8–5.4 KYA

FIGURE 3 Above: Schematic representation of phylogenetic relationships in the genus *Capsella*. Phylogenetic reconstruction is based on a coalescent SVDQ algorithm analysis of all 282 accessions (shown in full in File S3). The values above branches represent bootstrap support values (values below 90 not shown). The node letters correspond to the letters in the table below. Below: Time spans retrieved from SNAPP analyses based on different calibration points. Time spans are represented as median node ages with minimum and maximum age span corresponding to the 95% HPD. The age spans in bold represent the individual calibration points for the time divergence analyses and are retrieved from the following studies: calibration I (Žerdoner Čalasan et al., 2019), calibration II (Hurka et al., 2012), calibration III (Douglas et al., 2015), and calibration IV (Han et al., 2015). Abbreviations: Cgr *Capsella grandiflora*; Cru *Capsella rubella*; Cbp *Capsella-bursa-pastoris*; Cor *Capsella orientalis*

All parameters in the analyses converged to the stationary distribution (ESS > 200) and the overall topologies corresponded to the other phylogenetic studies (Douglas et al., 2015; Guo et al., 2017; Hurka et al., 2012; Kryvokhyzha, Salcedo et al., 2019; Walden et al., 2020). The crown age of *Capsella* was estimated to approximately 0.75 Ma, the split of *C. grandiflora* and *C. rubella* to approximately 0.19 Ma, and the split of *C. bursa-pastoris*, *C. thracica*, and *C. orientalis* to 0.13 Ma (Figure 3). All three splits were placed into the Middle Pleistocene, mostly into MIS15, MIS7, and MIS6 (Figure 3). The crown age of *C. grandiflora* was dated at around 67 kya and the crown age of *C. rubella* at 26 kya, both placed into Late Pleistocene. The crown age of a clade consisting of *C. bursa-pastoris* and *C. thracica* was estimated to the Middle Holocene at around 6.7 kya and the diversification of *C. orientalis* occurred in the Late Holocene approximately 2.1 kya (Figure 3).

3.4 | Inference of genetic structure according to STRUCTURE

The highest ΔK value in the *Capsella* dataset was inferred for $K = 6$, while in the *Capsella orientalis* dataset the highest ΔK value was retrieved at $K = 3$ (Figure S6). The *Capsella* dataset subsequently splits in the following way: $K = 2$ separated *C. orientalis* from the rest, with *C. bursa-pastoris* accessions exhibiting moderate introgression (Figure 4a). At $K = 3$ and $K = 4$, subsequent fragmentation within *C. orientalis* was observed, mirrored also in the LEA analysis in *Capsella* (see below and Figure S7) as well as in *C. orientalis* datasets (Figures S7 and S8). At $K = 5$, the Mediterranean diploid branch got separated, and at $K = 6$, the introgression of *C. grandiflora* was observed in the Mediterranean genetic cluster of *C. bursa-pastoris* (Figure 4d,e). Subsequent clustering delimited Mediterranean *C. grandiflora* from genetically uniform *C. rubella* (Figure 4f). Further clustering could not be inferred from Figure 4 because it resulted in new minor genetic clusters embedded in *C. orientalis*. The *Capsella orientalis* dataset clustered in the following manner. At $K = 2$, northern and southern *C. orientalis* genetic clusters were inferred (Figure 5a). The southern cluster subsequently split up into a Tabagartai–Dzhungarian Alatau and Central Asian cluster, and South Kazakhstan cluster (at $K = 3$). The division ran across the southern foothills of the Altai Mountains, southwards to the Tian Shan Mountains along the Tabagartai–Dzhungarian Alatau mountain chain. At $K = 4$, a new genetic cluster mixed with the Central Asian genetic cluster was inferred along the Tabagartai–Dzhungarian Alatau mountain chain and at $K = 5$ an Altai genetic cluster emerged (Figure 5c,d). STRUCTURE genetic clustering did not strictly follow the *Capsella* species boundaries.

3.5 | Inference of genetic structure according to LEA

The lowest cross-entropy value in the *Capsella* dataset was inferred for $K = 10$ (albeit without a clear signal), while no optimal K value

was inferred for *Capsella orientalis* dataset (Figure S6). The *Capsella* dataset subsequently split in the following way. The deepest split at $K = 2$ was not species specific. *Capsella bursa-pastoris* and its sister species *C. thracica* clustered together, whereas *C. orientalis* clustered separately (Figure S7A). Sister species *C. grandiflora* and *C. rubella* were assigned to both clusters to approximately 50%. At $K = 3$, *C. grandiflora* and *C. rubella* were retrieved as a separate cluster (Figure S7B). At $K = 4$, *C. thracica* was retrieved as an individual species and at $K = 5$ the first intraspecific differentiation took place (Figure S7C,D). Further intraspecific split followed at $K = 6$, $K = 7$, and $K = 8$, inferring subpopulation structure within *C. orientalis* and splitting the *C. bursa-pastoris* clade into a Mediterranean and Eurasian clade (Figure S7E,F,G). At $K = 9$, *C. grandiflora* was recognized as a separate cluster from *C. rubella* (Figure S7H), and at $K = 10$, the population genetic structure of *C. orientalis* reflected the same structuring as inferred at $K = 5$ from the *Capsella orientalis* dataset (Figure S8) and was as follows. At $K = 2$, a northern and southern *C. orientalis* cluster was inferred, which subsequently fragmented into a Central Asian cluster (at $K = 3$, Figure S8A), a Tabagartai–Dzhungarian Alatau cluster (at $K = 4$, Figure S8B), and an Altai cluster (at $K = 5$, Figure S8C). LEA genetic clustering mostly followed the *Capsella* species boundaries.

3.6 | Inference of genetic structure according to DAPC

DAPC analysis of the *Capsella* dataset followed species-specific genetic clustering. However, it failed to delimit *C. rubella* from *C. grandiflora* along all K-values spanning from $K = 3$ to $K = 9$ (Figure S9). Close relationship between *C. bursa-pastoris* and *C. thracica* was inferred from all K-values. No extensive subclustering within *Capsella orientalis* was inferred (Figure S9). DAPC analysis of *Capsella orientalis* dataset retrieved the same substructure as inferred from LEA (Figures S8 and S10) and comparable to the one inferred from STRUCTURE (Figure 5). Recognized as separate clusters already at deeper nodes (inferred at $K = 3$ and $K = 4$ in the *Capsella orientalis* dataset, and at $K = 7$ and $K = 9$ in the *Capsella* dataset in LEA, respectively), the Tabagartai–Dzhungarian Alatau and Central Asian clusters were spatially well defined from the rest of *Capsella orientalis* populations. These were followed by the spatially still well separated Altai cluster, indicating a closer genetic relationship (Figure S10). The North Kazakhstan and South Kazakhstan *C. orientalis* clusters were still recognized as two, albeit partially overlapping, independent clusters (Figure S10).

3.7 | Ecological niche modeling

The best model (LQH 0.5; AUC = 0.8326) inferred from the ENMeval appropriately encompassed the current known distribution area of *Capsella orientalis* (Figure 6a). The maximum entropy algorithm indicated that the climatically suitable habitats were gradually
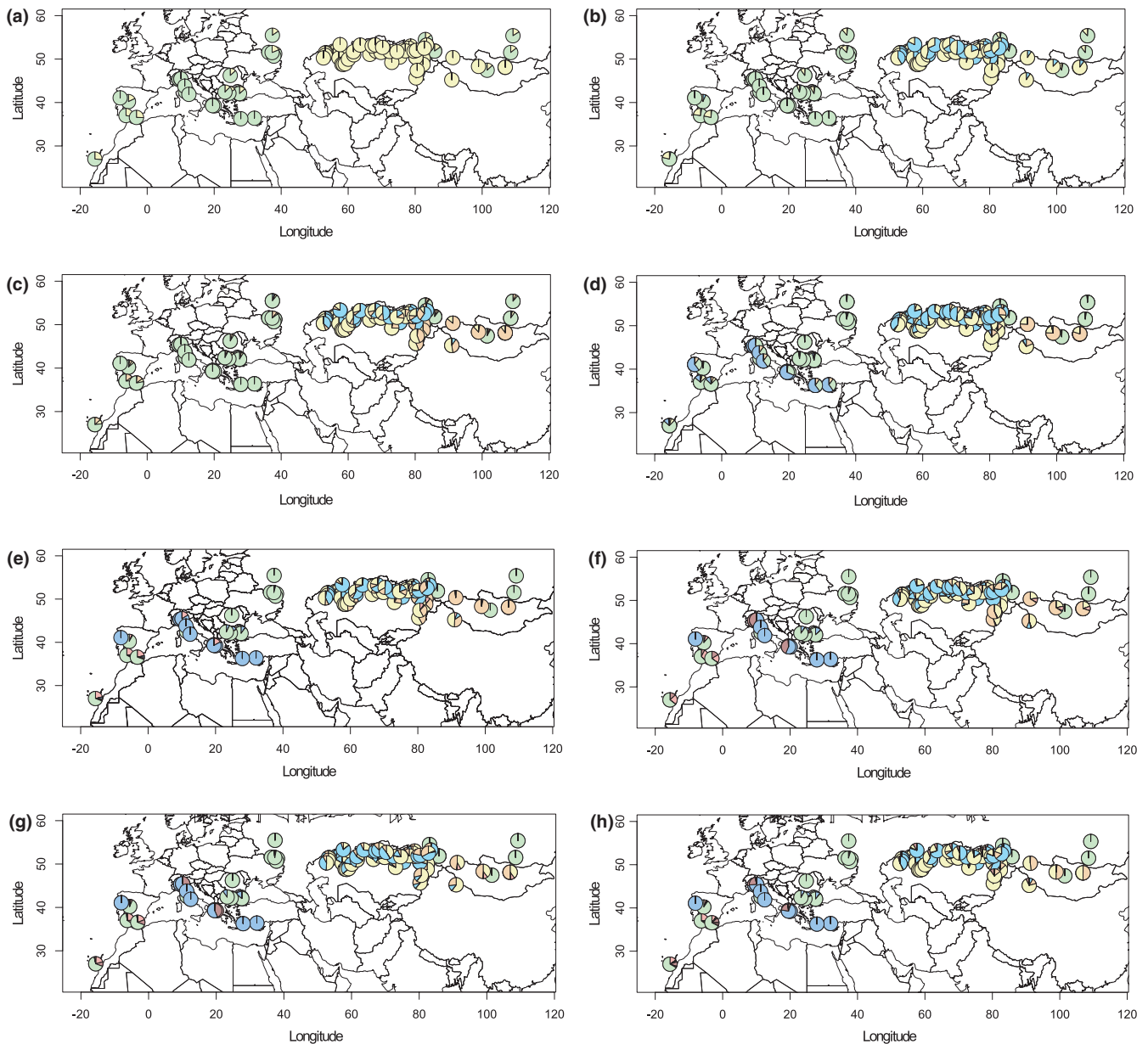
FIGURE 4 Genetic structure of sampled *Capsella* populations. Genetic cluster membership was inferred using STRUCTURE, showed as pie charts for $K = 2$ (a), $K = 3$ (b), $K = 4$ (c), $K = 5$ (d), $K = 6$ (e), $K = 7$ (f), $K = 8$ (g), and $K = 9$ (h)

reduced toward the east and west throughout the Holocene into the south Central Asian Mountain Chains and northern Pontic steppe (Figure 6b). During the LGM, the only climatically suitable habitats according to our niche modeling were the Black Sea coast in the west and South and South-East Kazakhstan in the east, with partially climatically suitable habitats also in Eastern China (Figure 6c).

3.8 | Population genetics analysis of *Capsella orientalis*

No extensive correlation of geography either with the molecular diversity indexes or with the number of private alleles were inferred (Figure S11). Individual populations with the highest genetic

diversity indices (Θ_S and Θ_π) were from Altai, South Kazakhstan, and North Kazakhstan. Individual populations with the lowest genetic diversity according to the Θ_S and Θ_π indices were in Altai and Mongolia but also in North Kazakhstan and Tabagartai/Dzhungarian Alatau (Figure S12–1). The highest number of substitution sites per gene copy were found in Altai populations, North Kazakhstan, and South Kazakhstan, and the lowest number of substitution sites per gene copy were recorded in Mongolian populations, Altai, and North Kazakhstan (Figure S12–1). The highest number of private substitution sites per gene copy, however, was found in South Kazakhstan populations and in two Altai populations. Lowest number of private substitution sites per gene copy were recovered from North Kazakhstan, Altai, and Central Asian populations (Figure S12–1). When populations were pooled together according to geography,
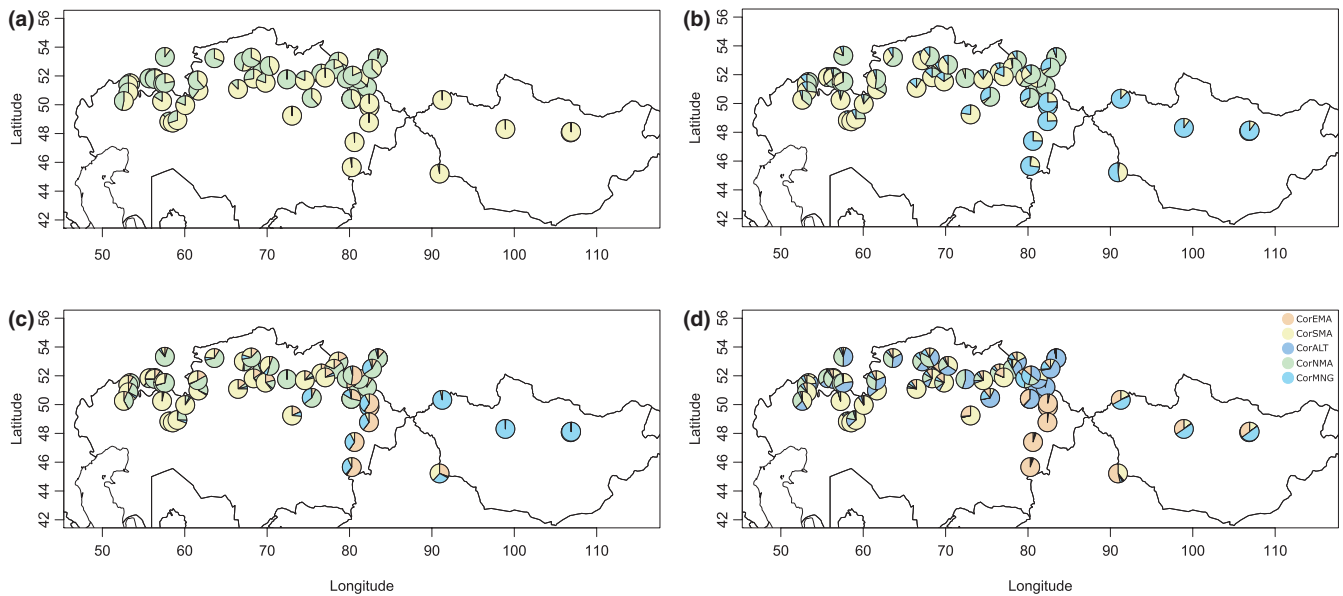
FIGURE 5 Geographic distribution of sampled *Capsella orientalis* populations. Genetic cluster membership was inferred using STRUCTURE, showed as pie charts for $K = 2$ (a), $K = 3$ (b), $K = 4$ (c), and $K = 5$ (d). Color coding as in the Files S4, S7, and S9, and is as follows: North Kazakhstan *C. orientalis* cluster in tea green (CorNMA), South Kazakhstan *C. orientalis* cluster in lemon yellow (CorSMA), Central Asian *C. orientalis* cluster in celeste (CorMNG), Tabagartai–Dzhungarian Alatau cluster *C. orientalis* cluster in deep champagne (CorEMA), and Altai *C. orientalis* cluster in baby blue (CorALT)

the highest genetic diversity in terms of Θ_S, Θ_π, and proportion of substitution sites per gene copy and private substitution sites per gene copy was retrieved from the South Kazakhstan cluster, and the lowest genetic diversity was inferred from the Mongolia genetic cluster (Figure S12–2).

F-statistics uncovered that the most genetically similar populations are in North Kazakhstan with values between 0.08 and 0.09. Contrarily, the highest F_{ST} values ($0.60 > F_{ST} > 0.40$) were inferred between some Central Asian, Altai, and South Kazakhstan populations (Figure S12–3). When pooled together according to geography, F_{ST} analysis indicated that the most similar genetic clusters were the North Kazakhstan and South Kazakhstan genetic clusters ($F_{ST} = 0.039$). Genetically, the most divergent genetic cluster was the Tabagartai/Dzhungarian cluster that differed the most from the Altai ($F_{ST} = 0.115$) and Central Asian genetic cluster ($F_{ST} = 0.110$; Figure S12–4).

AMOVA uncovered that the genetic divergence among the five genetic clusters was low at only about 3.70% (Figure S12–5). 10.60% of genetic diversity could be explained among populations, and the highest percentage of genetic variation could be explained along individuals within populations at 81.83%. The average observed heterozygosity H_O was at 0.071 near zero, and the average expected heterozygosity H_E was 0.489 (Figure S12–5). The degree of inbreeding within populations (F_{IS}) was high at 0.955. Inbreeding coefficient of an individual relative to the total population (F_{IT}) was also high at 0.961. The variance among subpopulations within groups (F_{SC}) was close to zero at 0.110 as it was the variance among groups relative to the total variance (F_{CT}) at 0.037 (Figure S12–5). All AMOVA results were highly significant with $p < .005$.

4 | DISCUSSION

4.1 | Dated phylogeny of evolutionary lineages within *Capsella*

Our SNP-based data clearly confirm the five species concept for *Capsella* (Chater, 1964; Hurka et al., 2012) and corroborate the identification of two ancestral evolutionary lineages within the genus as already outlined in Hurka et al. (2012): *Capsella grandiflora* and *C. rubella* on the Mediterranean side, and *C. orientalis* on the Eurasian side. The two allopolyploids *C. bursa-pastoris* and *C. thracica* are geographically and genetically placed in between, albeit the geographic signal of the former one is obscured due to its immense colonization potential (Figures 3 and 7). Our SNP data also support the two lineages within *C. bursa-pastoris* as previously revealed by multilocus isozyme analyses—a Mediterranean and a temperate Eurasian lineage (Wesse et al., 2021)—and are in agreement with the SNP-based analyses of Cornille et al. (2016). These authors detected a European/Russian and a "Middle Eastern" (which we refer to as "Mediterranean") *C. bursa-pastoris* genetic cluster in the geographic area covered by our study. The origin of *C. thracica* is within the temperate *C. bursa-pastoris* lineage, matching the geographical distribution areas of the two taxa.

The main conclusion from hitherto published *Capsella*-dating analyses and our current analysis is that the tribe Camelineae originated in the second half of the Miocene and the stem age of *Capsella* is of Pliocene age (Figures 3 and 7, Figure S13). During the Pleistocene, *Capsella* split into two basal lineages, an eastern and a western lineage, and subsequent splits within the two lineages occurred during the Middle to Late Pleistocene and in postglacial times (Figures 3 and 7).
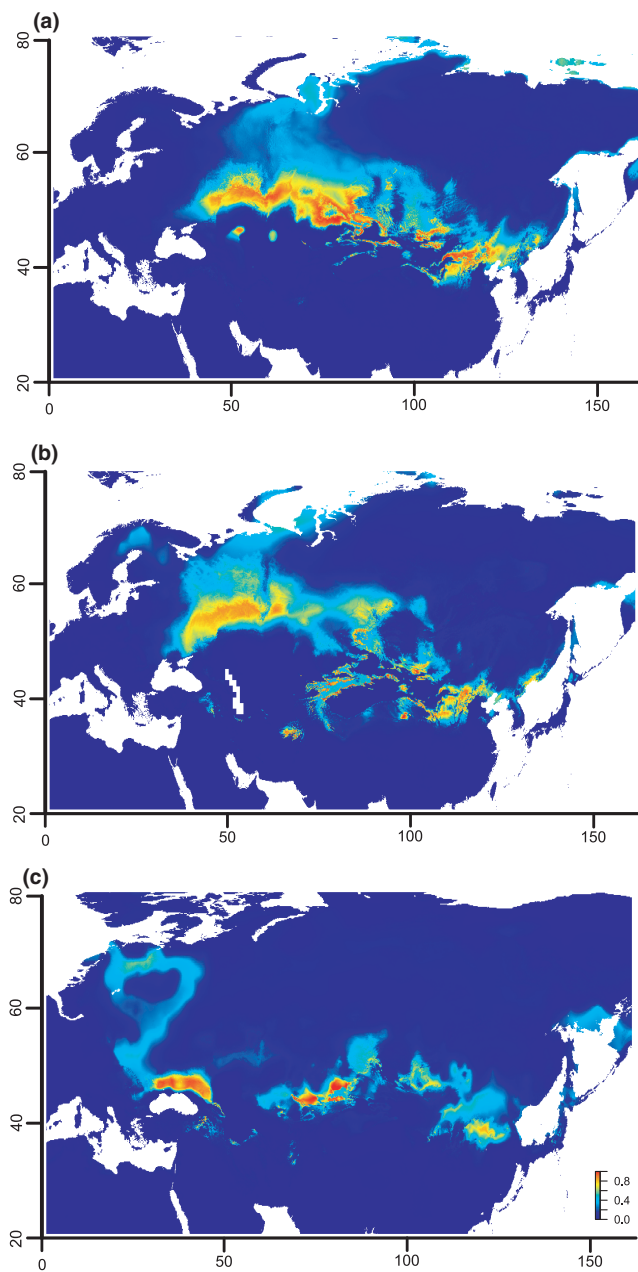
FIGURE 6 Predicted habitat suitability for *Capsella orientalis* during present time (a), Mid Holocene (b) and Last Glacial Maximum (c), respectively, as inferred from ecological niche modeling. Warmer colors represent areas of higher probability for the species' presence, while colder colors represent areas with lower probability for its presence. Note that the analysis retrieves only the habitat suitability, but not necessarily the realized ecological niche of *C. orientalis*

4.2 | Phylogeography of *Capsella*

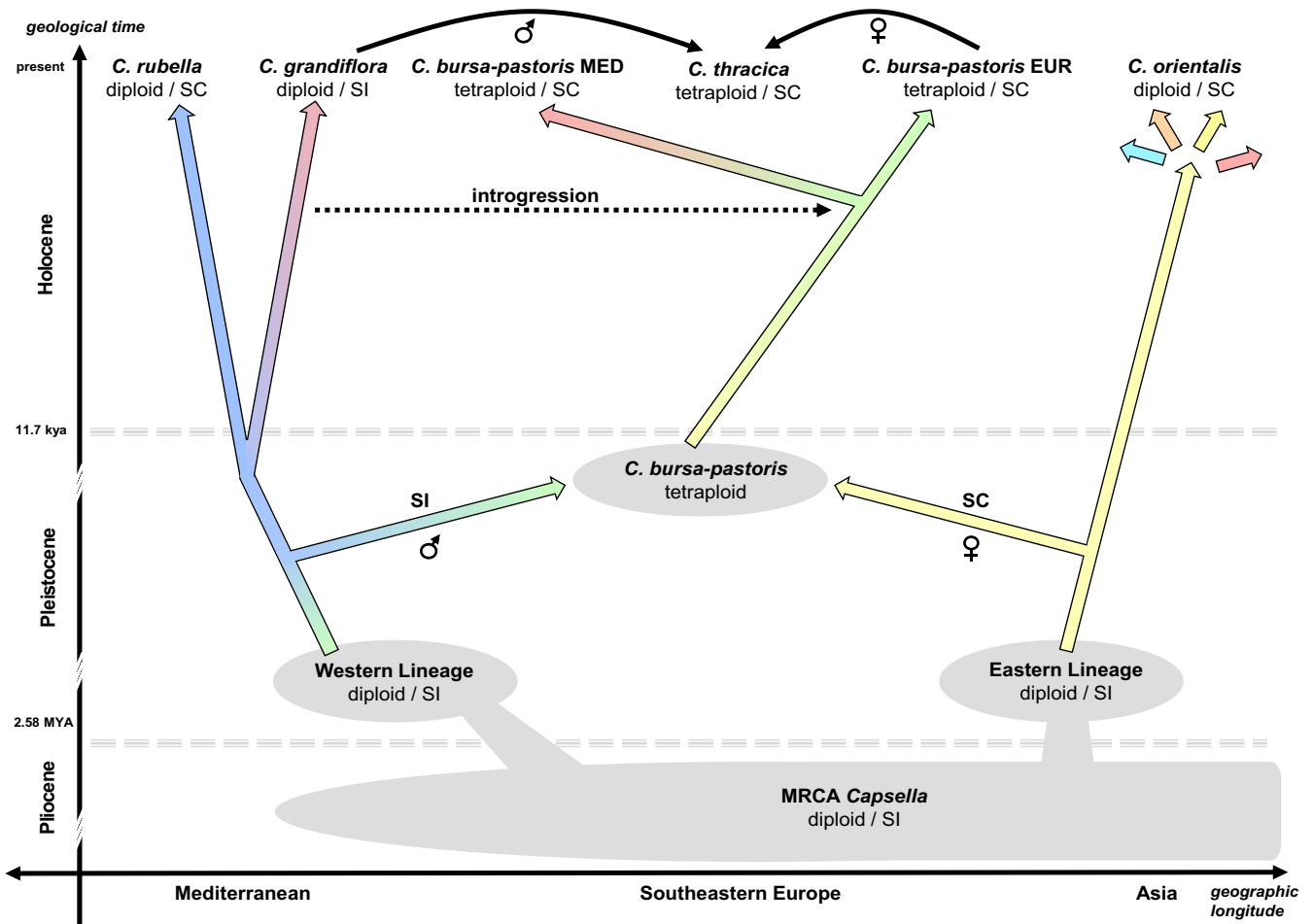### 4.2.1 | The genus *Capsella*

It has already been hypothesized that the distribution area of the most recent common ancestor (MRCA) of *Capsella* stretched across Eastern Europe and/or western Siberia (Hurka et al., 2012). Based on the present results, we can be more precise. The sister genus

Catolobus (Couvreur et al., 2010; Huang et al., 2016) is found in temperate climate zones of Eurasia stretching from Japan, Russian Far East, and China to southeastern Europe, and inhabits woodlands, grasslands, and riverbanks often as a ruderal weed. *Catolobus* thus shares ecogeographical characters and distribution area with *Capsella*. Morphological evidence (indumentum, trichome types, branching patterns) and ploidy level (2n = 2x = 16 for the two basal branching lineages in *Capsella*) also support the close relationship.

We assume the biogeographical roots of the genus *Capsella* in Middle Asia, in the area including the northern Caspian Sea coast region and Kazakhstan (Western Asiatic subregion of the Irano-Turanian floristic region sensu Takhtajan, 1986, regional subcenter IT3 sensu Manafzadeh et al., 2017). This is suggested by the present distribution area of *C. orientalis* (Figure 2). During the Pliocene, the MRCA of *Capsella* extended its range out of Asia and westwards into southeastern Europe and eventually even further west into the Mediterranean region. This timing corresponds with the estimated stem age of *Capsella* (split between *Catolobus* and *Capsella*) of ca. 4 ± 2 Ma (Žerdoner Čalasan et al., 2019). Theoretically, there are two possible east-to-west migration routes, (a) a "southern route" south of the Caspian and Black Sea to the eastern Mediterranean region and from there to southeastern Europe, and (b) a "northern route" along the steppe biome to southeastern Europe and perhaps even further west into the Mediterranean climate zone. Many arguments are in favor of the northern route whereas the southern route appears very unlikely as explained below.

The "southern route" scenario: Mediterranean steppes are thought to be descendants of the Anatolian steppes that expanded throughout the Mediterranean during the Pliocene (Suc et al., 2018). This route, however, is discussed for xeric steppe elements only (Jia & Bartish, 2018; Lauterbach et al., 2019; Manafzadeh et al., 2014; Žerdoner Čalasan et al., 2021), and *Capsella* does not belong to this ecological group. The main argument against an Early Pliocene southern *Capsella* route is the estimated divergence time between the two basally branching lineages, the Mediterranean *C. grandiflora/rubella* and the continental *C. orientalis* lineage of ca. 1 Ma or less, clearly indicating a Pleistocene and not a Pliocene origin. A possible Late Pliocene/Early Pleistocene "southern route" did not exist because of the Akhchagyl and Apsheron transgressions of the Caspian Sea, which were continuously blocking access to the southern route from the north for a period of ca. 2 million years (from ca. 3 to 1 Ma: Popov et al., 2006; Svitoch, 2016). Later, the great deserts of southern Kazakhstan and Turkmenistan prevented migration from northern parts of Kazakhstan to the southwest (southern route). Thus, the scenario of ancestral *Capsella* reaching Europe via a "southern route" is highly improbable.

Instead, there is convincing evidence for the "northern route." The spread of steppe elements from Asia to the Black Sea region since the Pliocene is clearly indicated by pollen records (e.g., Feurdean & Vasiliev, 2019; Popescu et al., 2010; Popova et al., 2017; Tarasov et al., 2000), and the climate/landscape history of the Eurasian steppe belt throughout the Neogene and Quaternary has been intensively studied (Grichuk (1989) and Velichko et al. (2005)

FIGURE 7 Outline of the evolutionary history of the genus *Capsella* placed into space and time. The color coding of the arrows corresponds to the color coding in Figure 4. Note that the geological time axis as well as the geographical latitude axis are not to scale

for East European Plain; Arkhipov et al. (2005) for West Siberian Plain; Akhmetiev et al. (2005) for Kazakhstan Plains; see also Hurka et al. (2019) and references therein). These studies elucidate the importance of Euro-Siberian steppe as a corridor for *Capsella* migration routes, which is mirrored in the present distribution area of *Capsella orientalis* (Figure 2).

In the Middle Pleistocene, the two diploid basal *Capsella* lineages evolved (Figures 3 and 7). Present distribution area and climate envelope identify *C. orientalis* and *C. grandiflora/rubella* as true steppe and Mediterranean elements, respectively. Their Pleistocene biogeography was heavily affected by the numerous cold-warm macrocycles of the Pleistocene. On the East European Plain, extreme changes of landscapes occurred during the last 800 kya (Middle to Late Pleistocene: Don, Oka, Dnieper, Valdai glaciations and corresponding interglacial warm phases). Severity and duration of cold phases increased and development of permafrost, formation of large continental ice sheets, and loess deposition intensified. The Don glaciation was the greatest glaciation on the East European Plain (correlated with the "Cromer Complex" Stage in NW Europe: Velichko et al., 2011a). Periglacial cold steppe spread over the East European Plain north of the Black and Caspian Sea. Interglacial periods were characterized by temperate steppes, forest-steppe, and

forests (Grichuk, 1989, 1992a; Velichko et al., 2005). Contrarily, the Mediterranean region including the Balkan Peninsula was not affected as much. Mediterranean steppes thus continued to exist during the Pleistocene, but around 1 Ma temperature gradients became more important than humidity gradients (turnover from xeric to more thermic determinism sensu Suc et al., 2018).

4.2.2 | *Capsella orientalis*

The distribution map presented in Figure 2 comprises all recorded localities of *C. orientalis* known to us (literature surveys, herbarium studies, and own field collections). The distribution area stretches nearly 5,000 km from Eastern Europe (lower Don) to Central Asia (Mongolia) more or less along the Eurasian steppe belt. Its main distribution is in Kazakhstan.

Present differentiation within *C. orientalis* is of postglacial age and thus very young, which is also expressed by invariable ITS, ETS, and cpDNA sequences (data not shown). The *C. orientalis* lineage itself is of Middle Pleistocene age (Figure 3). Palaeoclimate modeling indicates that *C. orientalis* could survive the LGM in southern Kazakhstan and southeastern Europe (Figure 6). With the beginning

of the Holocene, *C. orientalis* extended its range from the Asian refuge area northwards and eastwards into the Altai/Sayan and Tabagartai/Dzhungarian mountain chains and into Central Asia (Figure 5). This scenario is supported by the high number of private alleles and molecular diversity indexes Θ_S and Θ_π, which on average are higher in southern Kazakhstan than in the recently colonized regions. Lowest genetic diversity was observed in Mongolia. However, discontinuous mosaic variation patterns were also observed (Figure 5; Figures S11 and S12). This overall genetic variation pattern is typical for recent colonization events (for *C. bursa-pastoris* see for instance Hurka, 1984, 1990, Wesse et al., 2021). Interestingly, flowering behavior does not follow this pattern. *Capsella orientalis* flowers in the spring in the lowland steppe and forest-steppe biomes of the Euro-Siberian steppe belt ("planta vernalis," Klokov, 1926) but significantly later in the mountain steppes of the Altai/Sayan Mountain Country and Mongolia (estimated by comparisons of collecting dates, data not shown). Flowering ecology of *C. orientalis* thus varies greatly with the environments and growing seasons. We conclude that the variation pattern at the molecular level reflects population histories, colonization processes like gene flow events, and genetic drift effects whereas variation at the phenotypic level reflects adaption to local environments (including the altitude).

Population genetic data are in accordance with a predominantly selfing breeding system effective in natural populations ($H_O = 0.0071$, $H_E = 0.489$; $F_{IS} = 0.955$, $F_{IT} = 0.961$; see chapter Results) and confirm conclusions drawn from previous research (Bachmann et al., 2019; Hurka et al., 2012; Woźniak et al., 2020). Assuming that self-compatibility (SC) is a derived character state (Durand et al., 2020), the question arises about the timing of the loss of self-incompatibility (SI). Bachmann et al. (2019) argue for a loss of SI in *C. orientalis* between 2.6 Ma and 70 kya but closer to the upper bound, and provide evidence that *C. orientalis* was SC when it contributed to the origin of *C. bursa-pastoris* at about 130 kya.

4.2.3 | *Capsella bursa-pastoris*

It is evident from the STRUCTURE analysis (Figure 4) that *Capsella bursa-pastoris* retained SNP sequences from both ancestral lineages. This argues for a hybrid origin of *C. bursa-pastoris* between the two lineages, which has been confirmed by other genomic studies that show two subgenomes in *C. bursa-pastoris*, one originating from the western and one from the eastern lineage (Douglas et al., 2015; Han et al., 2015; Kryvokhyzha, Salcedo et al., 2019). Furthermore, subsequent evidence argues that the maternal parent came from the eastern lineage represented by *Capsella orientalis* (Douglas et al., 2015; Hurka et al., 2012; Omelchenko et al., 2020). To interpret the polyploidization event in a phylogeographic framework, we have to postulate firstly, a split of the continuous MRCA distribution area of *Capsella* dividing it into two separated areas, and secondly, subsequent expansion out of the two part areas leading to a contact zone where the tetraploid ancestor of *C. bursa-pastoris* arose through allopolyploidy.

Arguments for a MRCA range split during the Middle and Late Pleistocene as suggested by our time divergence estimation (Figure 3) are numerous and outlined above (see Discussion on the genus *Capsella*). During the cold stages, the interglacial steppe belt was replaced by periglacial cold steppe and tundra-steppe north of the Black Sea. In the following interglacial, steppe and forest-steppe expanded again in east-to-west direction. This eastward contraction and westward expansion of the steppe belt occurred several times during the climatic macrocycles (Grichuk, 1989; Velichko et al., 2005). In Asia, the steppe could move to the south whereas in southeastern Europe, the steppe could not migrate to the south and the continuous range of the MRCA was disrupted due to geophysical barriers (Black Sea and Caucasus). Its western partial distribution areal was pushed west of the Black Sea into the Balkan Peninsula and into the Mediterranean region. Westward range extensions of Pontic-South Siberian steppe and forest-steppe geoelements into the Balkan Peninsula are numerous (Horvat et al., 1974). The Thracian Plain connecting the southwestern East European Plain and Anatolia with the Mediterranean region is a well-known gateway for plant migration in and out of the Balkan Peninsula (Horvat et al., 1974; Magyari et al., 2008; Suc et al., 2018; Turrill, 1929), and might have also facilitated the westward migration of *Capsella* into the Balkan Peninsula.

According to our divergence time estimations (Figure 3), the hybridization event leading to *C. bursa-pastoris* took place during the last Pleistocene transition from glacial to interglacial, about 120,000 years ago, corresponding approximately to the last interglacial (LIG) and MIS 5e (Eem or Mikulino, respectively). During the climatic optimum of the LIG (about 120,000 years ago), nemoral forest grew around the Black Sea and northern parts of the Balkan Peninsula down to 40°N. It thus appears that the steppe belt was interrupted not only in glacial but also in interglacial times, and one might question a secondary contact zone. However, the nemoral forests mirror the condition only during climatic/vegetation optima of the interglacials. Outside of these optima, longer periods with open vegetation in form of grasslands or forest-steppe existed (Leroy et al., 2011), and this opened up a short window for steppe plants to migrate out of their glacial refugia (western Balkan and eastern Black Sea/Kazakhstan), resulting in a secondary contact.

Predicted distribution of *C. grandiflora* at present, LGM and LIG based on ecological niche modeling support our assumption of range extension from Balkan Peninsula refugium toward the east into the Black Sea region (Han et al., 2015). We hypothesize that the gene pools of the two diploid basal lineages, which had been separated since the Middle Pleistocene, had already differentiated into the *C. grandiflora* and *C. orientalis* lineages before coming into secondary contact in the northern Black Sea coast region. In the secondary contact zone, *C. bursa-pastoris* originated via hybridization and whole-genome duplication (allotetraploidy). This was not a single event and must have taken place several times. Otherwise, the polymorphism within *C. bursa-pastoris* as inferred from isozymes cannot be explained (Hurka et al., 2012). Douglas et al. (2015) also "clearly rule out a single polyploid origin from two haploid gametes".

Recently uncovered spontaneous somatic tetraploidization following hybridization between *C. orientalis* and *C. grandiflora* experimentally supports the origin of *C. bursa-pastoris* (Bachmann et al., 2021) and offers further interpretation on the polyplodization process.

Genome duplications and their consequences as drivers of major evolutionary innovations are in focus of current research at the genomic level (Nieto Feliner et al., 2020; Soltis & Soltis, 2020; Spoelhof et al., 2020; Walden et al., 2020). In the recent years, allopolyploid *C. bursa-pastoris* received special attention because of its young age (Douglas et al., 2015; Kasianov et al., 2017; Kryvokhyzha, Salcedo et al., 2019, Kryvokhyzha, Milesi et al., 2019). The genome duplication in *C. bursa-pastoris* coincides with its enormous colonization ability to migrate into different climate zones and colonize regions far above those inhabited by its parent species (Hurka & Neuffer, 1997; Wesse et al., 2021). It would appear that *C. bursa-pastoris* quickly spread over Eurasia shortly after its origin in the Late Pleistocene. This is supported by fossil seeds of *C. bursa-pastoris* reported from interglacial deposits at Ilford, Essex, England (West et al., 1964). The fossil-bearing sediments correlate with MIS 5e (Eemian of continental Europe). Distribution of *C. bursa-pastoris* on the British Isles at the LIG was predicted by palaeomodelling as well (Han et al., 2015) and thus supports the assumption of quick and early range extension after its origin in the southwestern East European Plain.

4.2.4 | *Capsella bursa-pastoris*, MED, Mediterranean lineage

As outlined above, two intraspecific lineages are recognized in *C. bursa-pastoris*, one predominantly in Mediterranean-like climate regions (*C. bursa-pastoris* MED) and the other in more temperate climate regions (*C. bursa-pastoris* EUR). These two lineages are of Holocene age (Figure 3). We hypothesize that the Mediterranean *C. bursa-pastoris* had contact with the diploid Mediterranean species *C. grandiflora/rubella* and that the MED lineage originated from introgression of *C. grandiflora/rubella* gene material into the *C. bursa-pastoris* genome (Figure 4e). This gene flow may have supported the adaptation to Mediterranean climates during the postglacial colonization. Distribution models of the MED and EUR lineages are clearly in favor of respective climate preferences (Wesse et al., 2021). The observed worldwide distribution patterns of genetic variation in the range of *C. bursa-pastoris* are the outcome of both migration/colonization history and environmental filtering due to climate preadaptation.

Mixed populations of diploid *C. grandiflora/rubella* and tetraploid *C. bursa-pastoris* are common in the Mediterranean region. These species often hybridize, and the triploid hybrid, known under the name *C. gracilis* Gren., can produce flowers. The hybrid is mostly sterile but fertile intermediates sometimes occur (reported by Chater, 1993). Thus, this triploid stage might eventually served as a bridge for introgression of diploid gene material across ploidy barriers into *C. bursa-pastoris*. Slotte et al. (2008) record introgression of *C. rubella* into *C. bursa-pastoris*, and adaptive introgression from

the diploid congeners into *C. bursa-pastoris* has been demonstrated before as well by Han et al. (2015).

4.2.5 | *Capsella thracica*

In the Early Holocene, *C. thracica* originated by hybridization between the Eurasian lineage of *C. bursa-pastoris* and *C. grandiflora* in southwestern Europe. This is clearly confirmed by our time divergence estimation (Figure 3) and by the genetic structure analysis (Figure 4). During the LGM (20,000–18,000 years), periglacial steppes covered the landscape from the northern Black Sea coast to the Altai-Tian Shan-Pamir mountain systems (Grichuk, 1992a, 1992b; Velichko & Isayeva, 1992) and divided the preceding Last Interglacial (LIG) grassland/dry shrubland/nemoral forest belt into a Mediterranean and Asian area, which became closed again in the Holocene. This in turn opened up again migration routes for steppe and Mediterranean elements. A strong Mediterranean influence on the western Pontic steppe vegetation has been observed before (Magyari et al., 2008; Turrill, 1929). The origin of the sub-Mediterranean geoelements in the steppe vegetation has long been discussed, and it is hypothesized that many of them invaded the region during the Holocene from their Pleistocene refuge areas in the Balkan Peninsula (Tzonev et al., 2006 and references therein).

The evolutionary history of *C. grandiflora* fits well into this picture. This species is a typical Mediterranean element whose distribution area was drastically reduced during the glaciation periods. We assume that during the Early Holocene *C. grandiflora* extended its range from its LGM Balkan refuge to the east into the Thracian Plain similarly to the possible pre-LIG expansion (Han et al., 2015). In the Thracian Plain, the hybridization occurred analogously to the one resulting in the origin of *C. bursa-pastoris* as supported by the two ITS copies, one from *C. bursa-pastoris* and one from *C. grandiflora* (Hurka et al., 2012). In agreement with these findings are our SNP data. STRUCTURE analysis retrieved a signal of the Mediterranean diploid *Capsella* species from all the investigated *Capsella thracica* populations (Figure 4). This Mediterranean SNP signal is with all probability from *C. grandiflora* since no *C. rubella* ITS sequence signal was ever inferred in *C. thracica*. Due to continuous backcrossing of *C. thracica* with its maternal EUR *Capsella bursa-pastoris*, the proportion of the Mediterranean signal is minor and might even eventually disappear.

ORCID

Anže Žerdoner Čalasan (iD) https://orcid.org/0000-0003-2081-2076
Dmitry A. German (iD) https://orcid.org/0000-0001-7951-1644
Frank R. Blattner (iD) https://orcid.org/0000-0001-7731-7741
Anna Seidl (iD) https://orcid.org/0000-0002-7177-296X
Barbara Neuffer (iD) https://orcid.org/0000-0003-4971-9991

REFERENCES

Akhmetyev, M. A., Dodonov, A. E., Sotnikova, M. V., Spasskaya, I. I., Kremenetzky, K. V., & Klimanov, V. A. (1999). Kazakhstan and central Asia – Plains and foothills. In A. A. Velichko & V. P. Nechaev (Eds.), _Climate and environment changes during the last 65 million years; Cenozoic: From Paleocene to Holocene_ (pp. 165–191). GEOS (in Russian).

Almquist, E. (1907). Studien über die _Capsella bursa-pastoris_ (L.). _Acta Horti Bergiani Stockholm_, _4_, 1–92.

Appel, O., & Al-Shehbaz, I. A. (2003). Cruciferae. In K. Kubitzki, & C. Bayer (Eds.), _Families and genera of vascular plants_ (Vol. _5_, pp. 75–174). Springer-Verlag.

Arkhipov, S. A., Volkova, V. S., Zolnikov, I. D., Zykina, V. S., Krukover, A. A., & Kulkova, L. A. (2005).West Siberia. In A. A. Velichko, & V. P. Nechaev (Eds.), _Climate and environment changes in Russia._ (pp. 67–88). Geological Society of America, Special Paper 382.

Bachmann, J. A., Tedder, A., Fracassetti, M., Steige, K. A., Lafon-Placette, C., Köhler, C., & Slotte, T. (2021). On the origin of the widespread self-compatible allotetraploid _Capsella bursa-pastoris_ (Brassicaceae). _Heredity_, _127_(1), 124–134. https://doi.org/10.1038/s41437-021-00434-9

Bachmann, J. A., Tedder, A., Laenen, B., Fracassetti, M., Désamoré, A., Lafon-Placette, C., Steige, K. A., Callot, C., Marande, W., Neuffer, B., Bergès, H., Köhler, C., Castric, V., & Slotte, T. (2019). Genetic basis

and timing of a major mating system shift in _Capsella_. _New Phytologist_, _224_, 505–517.

Barbolini, N., Woutersen, A., Dupont-Nivet, G., Silvestro, D., Tardif, D., Coster, P. M. C., Meijer, N., Chang, C., Zhang, H.-X., Licht, A., Rydin, C., Koutsodendris, A., Han, F., Rohrmann, A., Liu, X.-J., Zhang, Y., Donnadieu, Y., Fluteau, F., Ladant, J.-B., … Hoorn, C. (2020). Cenozoic evolution of the steppe-desert biome in Central Asia. _Science Advances_, _6_, eabb8227. https://doi.org/10.1126/sciadv.abb8227

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchene, S., Fourment, M., Gavryushkina, A., & Drummond, A. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. _PLOS Computational Biology_, _15_, e1006650.

Brandvain, Y., Slotte, T., Hazzouri, K. M., Wright, S. I., & Coop, G. (2013). Genomic identification of founding haplotypes reveals the history of the selfing species _Capsella rubella_. _PLoS Genetics_, _9_, e1003754. https://doi.org/10.1371/journal.pgen.1003754

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. _Molecular Biology and Evolution_, _29_, 1917–1932. https://doi.org/10.1093/molbev/mss086

Buono, D., Khan, G., von Hagen, K. B., Kosachev, P. A., Mayland-Quellhorst, E., Mosyakin, S. L., & Albach, D. C. (2021). Comparative phylogeography of _Veronica spicata_ and _V. longifolia_ (Plantaginaceae) across Europe: Integrating hybridization and polyploidy in phylogeography. _Frontiers Plant Science_, _11_, 588354.

Chater, A. O. (1964). _Capsella_ Medicus. In T. G. Tutin, V. H. Heywood, N. A. Burges, D. H. Valentine, S. M. Walters & D. A. Webb (Eds.), _Flora Europaea_ (vol. _1_, 1st ed). Cambridge University Press.

Chater, A. O. (1993). _Capsella_ Medicus. In T. G. Tutin, N. A. Burges, A. O. Chater, J. R. Edmonson, V. H. Heywood, D. M. Moore, D. H. Valentine, S. M. Walters & D. A. Webb (Eds.), _Flora Europaea_ (vol. _1_, 2nd ed). Cambridge University Press.

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent. _Bioinformatics_, _30_, 3317–3324.

Cornille, A., Salcedo, A., Kryvokhyzha, D., Glémin, S., Holm, K., Wright, S. I., & Lascoux, M. (2016). Genomic signature of successful colonization of Eurasia by the allopolyploid shepherd's purse (_Capsella bursa-pastoris_). _Molecular Ecology_, _25_, 616–629.

Couvreur, T. L. P., Franzke, A., Al-Shehbaz, I. A., Bakker, F. T., Koch, M. A., & Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). _Molecular Biology and Evolution_, _27_, 55–71. https://doi.org/10.1093/molbev/msp202

Doležel, J., Greilhuber, J., & Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. _Nature Protocols_, _2_, 2233–2244. https://doi.org/10.1038/nprot.2007.310

Dorofeyev, V. I. (2002). Cruciferae of European Russia. _Turczaninowia_, _5_(3), 5–114. (In Russian). http://old.ssbg.asu.ru/turcz/p5-114.pdf

Douglas, G. M., Gos, G. G., Steige, K. A., Salcedo, A., Holm, K., Josephs, E. B., & Wright, S. I. (2015). Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid _Capsella bursa-pastoris_. _Proceedings of the National Academy of Sciences USA_, _112_, 2806–2811.

Durand, E., Chantreau, M., Le Veve, A., Stetsenko, R., Dubin, M., Genete, M., Llaurens, V., Poux, C., Roux, C., Billiard, S., Vekemans, X., & Castric, V. (2020). Evolution of self-incompatibility in the Brassicaceae: Lessons from a textbook example of natural selection. _Evolutionary Applications_, _13_, 1279–1297. https://doi.org/10.1111/eva.12933

Earl, D. A., & von Holdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. _Conservation Genetics Resources_, _4_, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Eaton, D., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*, 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Ebel, A. L. (2002). New data on distribution of Brassicaceae species in South Siberia and East Kazakhstan. *Turczaninowia*, *5*, 60–68. (in Russian).

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fernández-Mazuecos, M., Mellers, G., Vigalondo, B., Sáez, L., Vargas, P., & Glover, B. J. (2018). Resolving recent plant radiations: Power and robustness of genotyping-by-sequencing. *Systematic Biology*, *67*, 250–268. https://doi.org/10.1093/sysbio/syx062

Feurdean, A., & Vasiliev, J. (2019). The contribution of fire to the late Miocene spread of grasslands in eastern Europe (Black Sea region). *Scientific Reports*, *9*, 6750.

Franzke, A., Hurka, H., Janssen, D., Neuffer, B., Friesen, N., Markov, M., & Mummenhoff, K. (2004). Molecular signals for Late Tertiary/Early Quaternary range splits of an Eurasian steppe plant: *Clausia aprica* (Brassicaceae). *Molecular Ecology*, *13*, 2789–2795. https://doi.org/10.1111/j.1365-294X.2004.02272.x

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*, 925–929. https://doi.org/10.1111/2041-210X.12382

Friesen, N., German, D. A., Hurka, H., Herden, T., Oyuntsetseg, B., & Neuffer, B. (2016). Dated phylogenies and historical biogeography of *Dontostemon* and *Clausia* (Brassicaceae) mirror the palaeogeographical history of the Eurasian steppe. *Journal of Biogeography*, *43*, 738–749.

Friesen, N., Smirnov, S. V., Shmakov, A. I., Herden, T., Oyuntsetseg, B., & Hurka, H. (2020). *Allium* species of section *Rhizomatosa*, early members of the Central Asian steppe vegetation. *Flora*, *263*, 151536. https://doi.org/10.1016/j.flora.2019.151536

German, D. A. (2015). Cruciferae (Brassicaceae): Alternative treatment for the "Conspectus of the vascular plants of Mongolia". *Turczaninowia*, *18*(2), 39–67. http://turczaninowia.asu.ru/article/view/793/469

German, D. A., Chen, W.-L., Smirnov, S. V., Liu, B., Kutzev, M. G., Wang, J., Shmakov, A. I., & Kamelin, R. V. (2012). Plant genera and species new to China recently found in northwest Xinjiang. *Nordic Journal of Botany*, *30*(1), 61–69. https://doi.org/10.1111/j.1756-1051.2011.01341.x

German, D. A., & Ebel, A. L. (2009). Some interesting findings of the Cruciferae in Asia. *Sistematicheskie Zametki Po Materialam Gerbarii Imeni P. N. Krylova Pri Tomskom Gosudarstvennom Universitete*, *101*, 5–11.

Grichuk, V. P. (1989). *History of flora and vegetation of the Russian Plain during the Pleistocene*. Nauka Press. (In Russian).

Grichuk, V. P. (1992a). Last interglacial (about 120,000 yr B.P.): Vegetation. In B. Frenzel, M. Pési, & A. A. Velichko (Eds.), *Atlas of paleoclimates and paleoenvironments of the northern hemisphere*. G. Fischer.

Grichuk, V. P. (1992b). Maximum cooling of the last glaciation (about 20,000 to 18,000 yr B.P.): Main types of vegetation (ecosystems). In B. Frenzel, M. Pési, & A. A. Velichko (Eds.), *Atlas of paleoclimates and paleoenvironments of the northern hemisphere*. G. Fischer.

Guo, X., Liu, J., Hao, G., Zhang, L., Mao, K., Wang, X., Zhang, D., Ma, T., Hu, Q., Al-Shehbaz, I. A., & Koch, M. A. (2017). Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics*, *18*, 176. https://doi.org/10.1186/s12864-017-3555-3

Han, T.-S., Wu, Q., Hou, X.-H., Li, Z.-W., Zou, Y.-P., Ge, S., & Guo, L. (2015). Frequent introgressions from diploid species contribute to the adaptation of the tetraploid shepherd's purse (*Capsella bursa-pastoris*). *Molecular Plant*, *8*, 427–438. https://doi.org/10.1016/j.molp.2014.11.016

Hantemirova, E. V., Belyaev, A. Y., Korchagina, O. S., & Laletina, I. A. (2020). Genetic differentiation and phylogenetic relationships of *Glycyrrhiza glabra* L., *G. uralensis* Fisch., and *G. korshinskyi* Grig. inferred from chloroplast DNA variation. *Russian Journal of Genetics*, *56*, 810–821. https://doi.org/10.1134/S1022795420070066

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/joc.1276

Horvat, I., Glavać, V., & Ellenberg, H. (1974). *Vegetation of Southeast Europe*. VEB G. Fischer. (In German).

Huang, C.-H., Sun, R., Hu, Y. I., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M. A., Al-Shehbaz, I., Edger, P. P., Pires, J. C., Tan, D.-Y., Zhong, Y., & Ma, H. (2016). Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Molecular Biology and Evolution*, *33*, 394–412. https://doi.org/10.1093/molbev/msv226

Hurka, H. (1984). Influence of population parameters on the genetic structure of *Capsella* populations. In K. Wöhrmann, & V. Loeschcke (Eds.), *Population biology and evolution* (pp. 183–192). Springer.

Hurka, H. (1990). Differentiation and adaptation in the genus *Capsella* (Brassicaceae). In S. Kawano (Ed.), *Biological approaches and evolutionary trends in plants* (pp. 19–32). Academic Press.

Hurka, H., Friesen, N., Bernhardt, K.-G., Neuffer, B., Smirnov, S. V., Shmakov, A. I., & Blattner, F. R. (2019). The Eurasian steppe belt: Status quo, origin and evolutionary history. *Turczaninowia*, *22*(3), 5–71. http://turczaninowia.asu.ru/article/view/6471/5179

Hurka, H., Friesen, N., German, D. A., Franzke, A., & Neuffer, B. (2012). 'Missing link' species *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus *Capsella* (Brassicaceae). *Molecular Ecology*, *21*(5), 1223–1238.

Hurka, H., & Neuffer, B. (1997). Evolutionary processes in the genus *Capsella* (Brassicaceae). *Plant Systematics and Evolution*, *206*, 295–316. https://doi.org/10.1007/BF00987954

Iljinska, A., Didukh, Y. A., Burda, R., & Korotchenko, I. (2007). *Ecoflora of Ukraine*, Vol. *5*. Phytosociocentre.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*, 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jalas, S., & Suominen, J. (1994). (Eds.), *Atlas Florae Europaea. Distribution of Vascular Plants in Europe. 10: Cruciferae (Sisymbrium to Aubrieta) - Verbreitungskarten*. The Academic Bookstore.

Jia, D.-R., & Bartish, I. V. (2018). Climatic changes and orogeneses in the Late Miocene of Eurasia: The main triggers of an expansion at a continental scale? *Frontiers in Plant Science*, *9*, 1400. https://doi.org/10.3389/fpls.2018.01400

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kasianov, A. S., Klepikova, A. V., Kulakovskiy, I. V., Gerasimov, E. S., Fedotova, A. V., Besedina, E. G., & Penin, A. A. (2017). High-quality genome assembly of *Capsella bursa-pastoris* reveals asymmetry of regulatory elements at early stages of polyploid genome evolution. *The Plant Journal*, *91*, 278–291.

Klokov, M. V. (1926). Sur une nouvelle espèce, *Capsella orientalis*, trouvée dans le Gouvernem de Voronéjè. *Byulleten' Obshchestva Estestvoispytatelei Pri Voronezhskom Gosudarstvennom Universitete*, *1*, 119–122. (In Russian).

Koch, M. A., German, D. A., Kiefer, M., & Franzke, A. (2018). Database taxonomics as key to modern plant biology. *Trends Plant Science*, *23*, 4–6. https://doi.org/10.1016/j.tplants.2017.10.005

Kryvokhyzha, D., Milesi, P., Duan, T., Orsucci, M., Wright, S. I., Glémin, S., & Lascoux, M. (2019b). Towards the new normal: Transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (Capsella bursa-pastoris). *PLOS Genetics*, 15, e1008131. https://doi.org/10.1371/journal.pgen.1008131

Kryvokhyzha, D., Salcedo, A., Eriksson, M. C., Duan, T., Tawari, N., Chen, J., Guerrina, M., Kreiner, J. M., Kent, T. V., Lagercrantz, U., Stinchcombe, J. R., Glémin, S., Wright, S. I., & Lascoux, M. (2019). Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid *Capsella bursa-pastoris* (Brassicaceae). *PLoS Genetics*, 15, e1007949. https://doi.org/10.1371/journal.pgen.1007949

Kulikov, P. V. (2005). *Synopsis of flora of the Chelyabinsk Region (vascular plants)*. Geotour. (In Russian).

Lang, G. (1994). *Quartäre Vegetationsgeschichte Europas: Methoden und Ergebnisse*. Gustav Fischer Verlag Jena.

Lauterbach, M., Veranso-Libalah, M. C., Sukhorukov, A. P., & Kadereit, G. (2019). Biogeography of the xerophytic genus *Anabasis* L. (Chenopodiaceae). *Ecology and Evolution*, 9, 3539–3552.

Leroy, S. A. G., Arpe, K., & Mikolajewicz, U. (2011). Vegetation context and climatic limits of the Early Pleistocene hominin dispersal in Europe. *Quaternary Science Reviews*, 30, 1558–1463. https://doi.org/10.1016/j.quascirev.2010.01.017

Magyari, E. K., Chapman, J. C., Gaydarska, B., Marinova, E., Deli, T., Huntley, J. P., Allen, J. R. M., & Huntley, B. (2008). The 'oriental' component of the Balkan flora: Evidence of presence on the Thracian Plain during the Weichselian late-glacial. *Journal of Biogeography*, 35, 865–883. https://doi.org/10.1111/j.1365-2699.2007.01849.x

Mai, D. H. (1995). *Tertiäre Vegetationsgeschichte Europas: Methoden und Ergebnisse*. Gustav Fischer Verlag Jena.

Manafzadeh, S., Salvo, G., & Conti, E. (2014). A tale of migrations east to west: The Irano-Turanian region floristic region as a source of Mediterranean xerophytes. *Journal of Biogeography*, 41, 366–379.

Manafzadeh, S., Staedtler, Y., & Conti, E. (2017). Visions of the past and dreams for the future in the Orient: The Irano-Turanian region from classical botany to evolutionary studies. *Biological Reviews*, 92, 1365–1388.

Mattila, T. M., Laenen, B., & Slotte, T. (2020). Population genomics of transitions to selfing in Brassicaceae model systems. In J. Y. Dutheil (Ed.), *Statistical population genomics, methods in molecular biology*, Vol. 2090 (pp. 269–287). Humana.

Meyer, M., Stenzel, U., & Hofreiter, M. (2008). Parallel tagged sequencing on the 454 platform. *Nature Protocols*, 3, 267–278. https://doi.org/10.1038/nprot.2007.520

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)* (pp. 1–8). Presented at the 2010 Gateway Computing Environments Workshop.

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENM eval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5, 1198–1205.

Nieto Feliner, G., Casacuberta, J., & Wendel, J. F. (2020). Genomics of evolutionary novelty in hybrids and polyploids. *Frontiers in Genetics*, 11, 792. https://doi.org/10.3389/fgene.2020.00792

Omelchenko, D. O., Makarenko, M. S., Kasianov, A. S., Schelkunov, M. I., Logacheva, M. D., & Penin, A. A. (2020). Assembly and analysis of the complete mitochondrial genome of *Capsella bursa-pastoris*. *Plants*, 9, 469. https://doi.org/10.3390/plants9040469

Paetsch, M., Mayland-Quellhorst, S., & Neuffer, B. (2006). Evolution of the self-incompatibility system in the Brassicaceae: Identification of S-locus receptor kinase (SRK) in self-incompatible *Capsella grandiflora*. *Heredity*, 97, 283–290. https://doi.org/10.1038/sj.hdy.6800854

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, 7, e32253. https://doi.org/10.1371/journal.pone.0032253

Popescu, S.-M., Biltekin, D., Winter, H., Suc, J.-P., Melinte-Dobrinescu, M. C., Klotz, S., Rabineau, M., Combourieu-Nebout, N., Clauzon, G., & Deaconu, F. (2010). Pliocene and Lower Pleistocene vegetation and climate changes at the European scale: Long pollen records and climatostratigraphy. *Quaternary International*, 219, 152–167. https://doi.org/10.1016/j.quaint.2010.03.013

Popov, S. V., Shcherba, I. G., Ilyina, L. B., Nevesskaya, L. A., Paramonova, N. P., Khondkarian, S. O., & Magyar, I. (2006). Late Miocene to Pliocene palaeogeography of the Paratethys and its relation to the Mediterranean. *Palaeogeography, Plaeoclimatology, Palaeoecology*, 238, 91–106. https://doi.org/10.1016/j.palaeo.2006.03.020

Popova, S., Utescher, T., Gromyko, D. V., Bruch, A. A., Henrot, A.-J., & Mosbrugger, V. (2017). Cenozoic vegetation gradients in the mid- and higher latitudes of Central Eurasia and climatic implications. *Palaeogeography Palaeoclimatology Palaeoecolgy*, 467, 69–82. https://doi.org/10.1016/j.palaeo.2016.09.016

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959. https://doi.org/10.1093/genetics/155.2.945

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67, 901–904. https://doi.org/10.1093/sysbio/syy032

Riley, H. P. (1936). The genetics and physiology of self-sterility in the genus *Capsella*. *Genetics*, 21, 24–39. https://doi.org/10.1093/genetics/21.1.24

Seidl, A., Pérez-Collazos, E., Tremetsberger, K., Carine, M., Catalán, P., & Bernhardt, K.-G. (2019). Phylogeny and biogeography of the Pleistocene Holarctic steppe and semi-desert goosefoot plant *Krascheninnikovia ceratoides*. *Flora*, 262, 151504.

Seidl, A., Tremetsberger, K., Pfanzelt, S., Blattner, F. R., Neuffer, B., Friesen, F., Hurka, H., Shmakov, A., Batlai, O., Čalasan, A. Ž., Vesselova, P. V. & Bernhardt, K.-G. (2021). The phylogeographic history of *Krascheninnikovia* reflects the development of dry steppes and semi-deserts in Eurasia. *Scientific Reports*, 11, 6645.

Seregin, A., Anackov, G., & Friesen, N. (2015). Molecular and morphological revision of the *Allium saxatile* group (Amaryllidaceae): Geographical isolation as the driving force of underestimated speciation. *Botanical Journal of the Linnean Society*, 178, 67–101.

Sheremetova, S. A., Ebel, A. L., & Buko, T. E. (2011). Supplement to the flora of Kemerovo region since 2001 till 2010. *Turczaninowia*, 14(1), 65–74. http://old.ssbg.asu.ru/turcz/turcz_14_1_65-74.pdf

Shull, G. H. (1909). *Bursa bursa-pastoris* and *Bursa heegeri*: Biotypes and hybrids. *Carnegie Institution of Washington Publication*, 112, 3–56.

Slotte, T., Hazzouri, K. M., Ågren, J. A., Koenig, D., Maumus, F., Guo, Y.-L., Steige, K., Platts, A. E., Escobar, J. S., Newman, L. K., Wang, W., Mandáková, T., Vello, E., Smith, L. M., Henz, S. R., Steffen, J., Takuno, S., Brandvain, Y., Coop, G., … Wright, S. I. (2013). The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature Genetics*, 45, 831–835. https://doi.org/10.1038/ng.2669

Slotte, T., Huang, H., Lascoux, M., & Ceplitis, A. (2008). Polyploid speciation did not confer instant reproductive isolation in *Capsella*. *Molecular Biology and Evolution*, 25(7), 1478–1481.

Soltis, P. S., & Soltis, D. E. (2020). Plant genomes: Markers of evolutionary history and drivers of evolutionary change. *Plants People Planet*, *3*, 74–82. https://doi.org/10.1002/ppp3.10159

Spoelhof, J. P., Keeffe, R., & McDaniel, S. F. (2020). Does reproductive assurance explain the insidence of polyploidy in plants and animals? *New Phytologist*, *227*, 14–21.

Stange, M., Sanchez-Villagra, M. R., Salzburger, W., & Matschiner, M. (2018). Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian isthmus. *Systematic Biology*, *67*, 681–699. https://doi.org/10.1093/sysbio/syy006

Stojanov, N., & Stefanov, B. (1948). *Flora of Bulgaria III*. Sofia. (In Bulgarian)

Suc, J.-P., Popescu, S.-M., Fauquette, S., Bessedik, M., Jiménez-Moreno, G., Bachirir Taoufiq, N., & Klotz, S. (2018). Reconstruction of Mediterranean flora, vegetation and climate for the last 23 million years based on an extensive pollen data set. *Ecologia Mediterranea, Faculté Des Sciences Et Techniques De St. Jérôme, Institut Méditerranéen D'ecologie Et De Paléoécologie*, *44*, 53–85.

Svensson, S. (1984). Chromosome numbers and morphology in the *Capsella bursa-pastoris* complex (Brassicaceae) in Greece. *Willdenowia*, *13*, 267–276.

Svitoch, A. A. (2016). Regressive periods of the Great Caspian. *Water Resources and the Regime of Water Bodies*, *43*, 134–148. https://doi.org/10.1134/S0097807816020160

Swofford, D. L. (2002). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*. Sinauer Associates.

Takhtajan, A. (1986). *Floristic regions of the world*. University of California Press.

Tarasov, P. E., Volkova, V. S., Webb, T., Guiot, J., Andreev, A. A., Bezusko, L. G., Bezusko, T. V., Bykova, G. V., Dorofeyuk, N. I., Kvavadze, E. V., Osipova, I. M., Panova, N. K., & Sevastyanov, D. V. (2000). Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern Eurasia. *Journal of Biogeography*, *27*, 609–620. https://doi.org/10.1046/j.1365-2699.2000.00429.x

Turrill, W. B. (1929). *The plant life of the Balkan Peninsula: A phytogeographical study*. Clarendon Press.

Tzonev, R., Roussakova, V., & Dimitrov, M. (2006). The Western-Pontic steppe vegetation in Bulgaria. *Hacquetia*, *5*, 5–23.

Tzvelev, N. N. (2000a). Combinationes novae taxorum plantarum vascularium. *Novosti Sistematiki Vysshikh Rasteniy*, *32*, 181–185.

Tzvelev, N. N. (2000b). *Manual of the Vascular Plants of North-West Russia*. St. Petersburg State Chemical-Pharmaceutical Academy Press.

Velichko, A. A. (1999). General features of the landscape and climate of northern Eurasia in the Cenozoic. In A. A. Velichko & V. P. Nechaev (Eds.), *Climate and environment changes during the last 65 million years; Cenozoic: From Paleocene to Holocene* (pp. 219–233). GEOS.

Velichko, A. A., Akhlestina, E. F., Borisova, O. K., Gribchnko, Y. N., Zhidovinov, N. Y., Zelikson, E. M., Iosifova, Y. I., Klimanov, V. A., Morosova, T. D., Nechaev, V. P., Pisareva, V. V., Svetlitskaya, T. V., Spasskaya, I. I., Udartsev, V. P., Faustova, M. A., & Shik, S. M. (2005). East European plains. In A. A. Velichko, & V. P. Nechaev (Eds.), *Climate and environment changes in Russia* (pp. 31–66). Geological Society of America, Special Paper 382.

Velichko, A. A., Faustova, M. A., Pisareva, V. V., Gribchenk, Y. N., Sudakova, N. G., & Lavrientev, N. V. (2011a). Glaciations of the East European Plain: Distribution and chronology. *Development in Quaternary Science*, *15*, 337–359.

Velichko, A. A., & Isayeva, L. L. (1992). Maximum cooling of the last glaciation (about 20,000 to 18,000m yr B.P.): Landscape types. In B. Frenzel, M. Pésci, & A. A. Velichko (Eds.), *Atlas of paleoclimates and paleoenvironments of the northern hemisphere*. G. Fischer.

Velichko, A. A., Timireva, S. N., Kremenetski, K. V., MacDonald, G. M., & Smith, L. C. (2011b). West Siberian Plain as a late glacial desert. *Quaternary International*, *237*, 45–53. https://doi.org/10.1016/j.quaint.2011.01.013

Verkhozina, A. V., Chernysheva, O. A., Ebel, A. L., Erst, A. S., Dorofeev, N. V., Dorofeyev, V. I., Grebenjuk, A. V., Grigorjevskaja, A. Y., Guseinova, Z. A., Ivanova, A. V., Khapugin, A. A., Korolyuk, A. Y., Korznikov, K. A., Kuzmin, I. V., Mallaliev, M. M., Murashko, V. V., Murtazaliev, R. A., Popova, K. B., Safronova, I. N., ... Krivenko, D. A. (2020). Findings to the flora of Russia and adjacent countries: New national and regional vascular plant records, 2. *Botanica Pacifica. A Journal of Plant Science and Conservation*, *9*, 139–154. https://doi.org/10.17581/bp.2020.09115

Volkova, P. A., Herden, T., & Friesen, N. (2017). Genetic variation in *Goniolimon speciosum* (Plumbaginaceae) reveals a complex history of steppe vegetation. *Botanical Journal of the Linnean Society*, *184*, 113–121.

Walden, N., German, D. A., Wolf, E. M., Kiefer, M., Rigault, P., Huang, X.-C., Kiefer, C., Schmickl, R., Franzke, A., Neuffer, B., Mummenhoff, K., & Koch, M. A. (2020). Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. *Nature Communications*, *11*, 3795. https://doi.org/10.1038/s41467-020-17605-7

Wesse, C., Welk, E., Hurka, H., & Neuffer, B. (2021). Geographical pattern of genetic diversity in *Capsella bursa-pastoris* (Brassicaceae)—a global perspective. *Ecology and Evolution*, *11*, 199–213.

West, R. G., Lambert, C. A., & Sparks, B. (1964). Interglacial deposits at Ilford, Essex. *Philosophical Transactions of the Royal Society B*, *247*, 185–212.

Woźniak, N. J., Kappel, C., Marona, C., Altschmied, L., Neuffer, B., & Sicard, A. (2020). A similar genetic architecture underlies the convergent evolution of the selfing syndrome in *Capsella*. *The Plant Cell*, *32*, 935–949.

Yokoya, K., Roberts, A. V., Mottley, J., Lewis, R., & Brandham, P. E. (2000). Nuclear DNA amounts in Roses. *Annals of Botany*, *85*, 557–561. https://doi.org/10.1006/anbo.1999.1102

Žerdoner Čalasan, A., German, D. A., Hurka, H., & Neuffer, B. (2021). A story from the Miocene: Clock-dated phylogeny of *Sisymbrium* L. (Sisymbrieae, Brassicaceae). *Ecology and Evolution*, *11*, 2573–2595.

Žerdoner Čalasan, A., Seregin, A. P., Hurka, H., Hofford, N. P., & Neuffer, B. (2019). The Eurasian steppe belt in time and space: Phylogeny and historical biogeography of the false flax (*Camelina* Crantz, Camelineae, Brassicaceae). *Flora*, *260*, 151477. https://doi.org/10.1016/j.flora.2019.151477

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.