DISSERTATION

ZUR ERLANGUNG DES DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)

AM FACHBEREICH 5: BIOLOGIE / CHEMIE

DER UNIVERSITÄT OSNABRÜCK

GENES DOCUMENTING HISTORY:

BIOGEOGRAPHICAL DYNAMICS OF SELECTED BRASSICACEAE TAXA AND CLIMATE-LANDSCAPE HISTORY OF THE EURASIAN STEPPE BELT

VORGELEGT VON

MSC. ANŽE ŽERDONER ČALASAN

AUS

MARBURG AN DER DRAU (SLOWENIEN)

OSNABRÜCK, DEN 19. FEBRUAR 2021

Die vorgelegte Arbeit wurde im Zeitraum vom 1. September 2017 bis zum 19. Februar 2021 in der Abteilung Botanik — Entwicklungsgenetik der Pflanzen der Universität Osnabrück unter der fachlichen Anleitung von apl. Prof. Dr. rer. nat. Barbara Neuffer angefertigt.

Prüfungsausschuss:

apl. Prof. Dr. rer. nat. Barbara NEUFFER **— Erste Gutachterin**
Universität Osnabrück, Fachbereich 5: Biologie/Chemie, Abteilung: Botanik (Evolutionäre Botanik)

Prof. Dr. rer. nat. Birgit GEMEINHOLZER **— Zweite Gutachterin**
Universität Kassel, Institut für Biologie, Fachgebiet: Botanik (Systematik und Morphologie)

Prof. i.R. Dr. rer. nat. Herbert HURKA
Universität Osnabrück, Fachbereich 5: Biologie/Chemie, Abteilung: Botanik (Evolutionäre Botanik)

Dr. rer. nat. Felix ALTHOFF
Universität Osnabrück, Fachbereich 5: Biologie/Chemie, Abteilung: Botanik (Entwicklungsgenetik der Pflanzen)

Dissertation eingereicht am: 19. 2. 2021

Tag der mündlichen Prüfung: 23. 4. 2021

Preface

Erklärung

Diese Dissertation wurde im Sinne von §5 der Promotionsordnung von außerplanmäßige Professorin Barbara Neuffer betreut. Ich erkläre hiermit, dass die Dissertation nicht einer anderen Prüfungsordnung vorgelegt worden ist und dass ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Koautoren der Manuskripte werden an den entsprechenden Stellen in der Dissertation namentlich erwähnt. Weitere Personen waren an der inhaltlichen materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Osnabrück, den 23. April 2021 .....

(Ort. Datum) (Unterschrift)

Declaration of contribution as a co-author

In this cumulative thesis, I present the results of my doctoral research, conducted under the supervision of Professor Barbara Neuffer at the Osnabruck University (Lower Saxony, Germany). The results of my research have been published in international peer-reviewed journals and are presented in the appendix of the thesis. The following articles were included in the cumulative thesis submitted for review and might differ slightly from the published journal articles. All of them have resulted from collaborations with other scientists, and my contributions to each of the four publications are as follows:

Publication I:

Žerdoner Čalasan A, Seregin AP, Hurka H, Hofford NP, Neuffer B. (2019): The Eurasian steppe belt in time and space: Phylogeography and historical biogeography of the false flax (*Camelina*, Camelineae, Brassicaceae). Flora 260: 151477.

Own contribution: For this publication, I carried out field work (10%), performed experiments (80%), analysed the data (100%), led the writing (95%) as well as the correspondence (95%) with a critical input from the other co-authors.

Publication II:

Friesen N, **Žerdoner Čalasan A**, Neuffer B, German DA, Markov M, Hurka H. (2020): Evolutionary history of the Eurasian steppe plant *Schivereckia podolica* (Brassicaceae) and its close relatives. Flora 268: 151602.

Own contribution: For this publication, I analysed the data (85%), assisted with the writing (35%) and contributed to the correspondence (35%).

Publication III:

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

Own contribution: For this publication, I carried out field work (10%), performed experiments (100%), analysed the data (100%), led the writing (95%) as well as the correspondence (95%) with a critical input from the other co-authors.

Publication IV:

Žerdoner Čalasan A, Hurka H, German DA, Pfanzelt S, Blattner F, Seidl A, Neuffer B. (preprint): Evolutionary history and historical biogeography of the model genus *Capsella* (Brassicaceae) with emphasis on the steppe floral element *Capsella orientalis* Klokov.

Own contribution: For this publication, I carried out field work (10%), performed experiments (95%), analysed the data (95%) and led the writing (90%).

Oral presentations:

Genes documenting history: Biogeographic dynamics of Eurasian steppe Brassicaceae and their implications for climate-landscape reconstruction.
Seminar lecture series at the Institute of Systematic Botany and Mycology, Faculty of Biology, Ludwig Maximilian University of Munich, Germany (Invited by Prof. Susanne Renner). Place and time: Munich (Germany) in July 2019.

Dealing with the big guys: Uncovering the floral history of the Eurasian steppe belt.
Talk at the 18th Austrian Botanical Conference & 24th International Symposium on Biodiversity and Evolutionary Biology of the German Botanical Society. Place and time: Klagenfurt (Austria) in September 2018.

Genes documenting history: Uncovering the phylogeographic patterns of *Camelina* Crantz.
Conference talk at the 48th Annual Meeting of the Ecological Society of Germany Austria and Switzerland. Place and time: Vienna (Austria) in September 2018.

The travels of Marco Polo — botanical edition.
Lecture at the Faculty of Natural Sciences and Mathematics, University of Maribor, Slovenia as a part of the course "Biogeography". (Invited by Assoc. Prof. Nina Šajna). Place and time: Maribor (Slovenia) in January 2018.

Additional publications since the official begin of the dissertation

Žerdoner Čalasan A, Kretschmann J, Gottschling M. (2018): Absence of co-phylogeny indicated repeated diatom capture in dinophytes hosting a tertiary endosymbiont. Organismic Diversity & Evolution 18: 29–38.

Kretschmann J, **Žerdoner Čalasan A**, Gottschling M. (2018): Molecular phylogenetics of dinophytes harbouring diatoms and endosymbionts (Kryptoperidiniaceae, Peridiniales), with evolutionary interpretations and a focus on the identity of *Durinskia oculata* from Prague. Molecular Phylogenetics and Evolution 118: 392–402.

Žerdoner Čalasan A, Kretschmann J, Filipowicz NH, Irimia RE, Gottschling M. (2018): Towards global distribution maps of unicellular organisms such as calcareous dinophytes based on DNA sequence information. Marine Biodiversity 49: 749–758.

Kretschmann J, Owsianny PM, **Žerdoner Čalasan A**, Gottschling M. (2018): The hot spot in a cold environment: Puzzling *Parvodinium* (Peridiniopsidaceae, Peridiniales) from the Polish Tatra Mountains. Protist 169: 206–230.

Žerdoner Čalasan A, Kretschmann J, Gottschling M. (2019): They are young and they are many: Dating freshwater lineages in unicellular dinophytes. Environmental Microbiology 21: 4125–4135.

Gottschling M, Chacón Pinilla J, **Žerdoner Čalasan A**, Kretschmann J, Kühne N, Neuhaus S, John U. (2019): Determination of dinophytes detected in environmental samples from Bavarian lakes (Germany). Freshwater Biology 65: 193–208.

Kretschmann J, **Žerdoner Čalasan A**, Meyer B, Gottschling M. (2019): Zero intercalary plates in *Parvodinium* (Peridiniopsidaceae, Peridiniales) and phylogenetics of *P. elpatiewskyi,* comb. nov. Protist 171: 125700.

Žerdoner Čalasan A, Kretschmann J, Gottschling M. (2020): Integrative taxonomy for sexually-deprived protists in the 21st century: A case study of *Trachelomonas* Ehrenb. (Euglenaceae) from Western Ukraine. Taxon 69: 28–42.

Table of contents

Summary

The Eurasian steppe belt is the vastest grassland region worldwide, stretching approximately 8000 km from the Pannonian Basin in the west to the Amur river in the east, altogether covering more than 8 million km^2. Due to its size and location, the vegetation of this biome was under strong influence of past climatic fluctuations that reached their peak in the Pleistocene. Fossil record of different steppe-associated flora and fauna places the onset of Eurasian open grasslands into Central Asia from which the grasslands spread westwards around 20 MYA and reached the East European Plain first in the Late Miocene around 8 MYA. However, while useful as a proxy to infer past vegetation patterns, fossils suffer from low-resolution power and thus fail to elucidate more detailed picture of the onset and the development of the Eurasian steppe belt flora.

The working hypothesis driving the whole project was that molecular signals in typical steppe plant species reflect the climate-landscape history of the steppe and the biogeographic dynamics of steppe taxa and thus allow for a much finer resolution of the history of the steppe belt in comparison to floristic and fossil-based methods. By applying a plethora of different phylogenetic, phylogeographic and biogeographic methods, I first investigated the evolutionary history of four unrelated Brassicaceae taxa that can be nowadays found in the Eurasian steppe belt. Second, I tried to draw parallels with the climate-landscape history of the Eurasian steppe belt as inferred from the fossil record to test the above-mentioned working hypothesis.

The first case study dealt with the economically important Brassicaceae genus *Camelina*, with an emphasis on *C. microcarpa*, which can be found across the whole Eurasian steppe belt. I uncovered that this taxon's contemporary uninterrupted distribution was split along the north coast of the Caspian Sea approximately 1 MYA, dating back to the Apsheron and Baku transgression events. During this time period, a polyploidisation event took place giving rise to a new hexaploid taxon and subsequently preventing gene flow after the regression of the Caspian Sea. The second case study investigated the evolutionary history of *Schivereckia*, which exhibits a highly disjunct distribution along the East European Plain and the Balkans and the whole *Schivereckia* clade that can be nowadays found either at higher latitudes in the subarctic zone or mountain ranges of the northern hemisphere. My analyses placed the radiation of this clade at the beginning of Pleistocene, when low temperatures promoted speciation and radiation of cold-adapted flora and fauna. The study showed that the highly disjunct distribution of *Schivereckia podolica* mirrors the Pleistocene

refugial areas of different ages and points towards a close evolutionary relationship between contemporary steppe and tundra flora.

A third case study focussed on *Sisymbrium*. I uncovered that while *Sisymbrium* did not evolve in the Eurasian steppe belt, it invaded this area as well as the Mediterranean several times independently throughout the Pleistocene in the south-north and east-west trajectory, respectively. This study proved yet again how variable the Eurasian steppe flora is in terms of evolutionary onset and that many of steppe elements originated in its adjacent areas in the south. These then subsequently migrated to the Eurasian steppe belt only after it was already fully developed either towards the west into the Euro-Siberian steppe or to the east into the Mongol-Chinese steppe. The fourth case study investigated the evolutionary history of *Capsella*. A GBS-based approach was proven successful for inferring biological history of young taxa. I showed that contemporary steppe element *Capsella orientalis* invaded the Eurasian steppe belt long after it was already developed in the south-north trajectory. I acquired further insights in the evolutionary history of its cosmopolitan sister tetraploid *C. bursa-pastoris* and placed the origin of the whole genus into the late Pliocene continental Asia.

All studies showed the importance of a taxon sample and offered an alternative nesting dating approach for reliable calibration in cases where no fossil record could be obtained. In conclusion, evolutionary history of all four investigated taxa were shaped by the same environmental changes that played a major role in the biological history of the Eurasian steppe belt. Overall, our working hypothesis that molecular signals in typical steppe plant species reflect the climate-landscape history of the steppe and the biogeographic dynamics of steppe taxa, turned out to be correct.

Zusammenfassung

Der eurasische Steppengürtel ist das größte Graslandbiom der Welt. Er erstreckt sich über ungefähr 8000 km vom Pannonischen Becken im Westen bis zum Amur im Osten und umfasst insgesamt mehr als 8 Millionen km². Aufgrund seiner Größe und Lage stand die Vegetation dieses Bioms stark unter dem Einfluss vergangener Klimaschwankungen, die ihren Höhepunkt im Pleistozän erreichten. Fossilberichte von verschiedenen Steppen-assoziierten Floren und Faunen legen nahe, dass der eurasische Steppengürtel vor ungefähr 20 Ma in Zentralasien entstanden ist, von wo aus er sich nach Westen ausgebreitet und vor ungefähr 8 Ma im späten Miozän zum ersten Mal die osteuropäische Ebene erreicht hat. Obwohl Fossilien als Anhaltspunkte für die Erschließung früherer Vegetationsmuster nützlich sind, leiden sie unter einem geringen Auflösungsvermögen und können daher kein detaillierteres Bild über den Beginn und die Entwicklung der eurasischen Steppengürtelflora liefern.

Die übergreifende Arbeitshypothese, die das gesamte Projekt vorantrieb, war, dass molekulare Signale in typischen Steppenpflanzenarten die Klimalandschaftsgeschichte der Steppe und die damit zusammenhängende biogeografische Dynamik von Steppentaxa widerspiegeln und somit eine viel feinere Auflösung der Geschichte des Steppengürtels im Vergleich zu den floristischen und fossilen Methoden darstellen können. Mit einer Vielzahl verschiedener phylogenetischer, phylogeografischer und biogeografischer Methoden untersuchte ich zunächst die Evolutionsgeschichte von vier nicht nah verwandten Brassicaceae-Taxa, die heutzutage im eurasischen Steppengürtel zu finden sind. Sodann versuchte ich, daraus Parallelen zur Klimalandschaftsgeschichte des eurasischen Steppengürtels zu ziehen, wie sie aus dem Fossilbericht abgeleitet wurden, um die oben erwähnte Arbeitshypothese zu testen.

Die erste Fallstudie befasste sich mit der wirtschaftlich wichtigen Brassicaceae-Gattung *Camelina* mit Schwerpunkt auf *C. microcarpa*, die im gesamten eurasischen Steppengürtel zu finden ist. Wir haben festgestellt, dass die heutzutage ununterbrochene Verbreitung dieses Taxons entlang der Nordküste des Kaspischen Meeres vor ungefähr 1 Ma disjunkt war und auf die Apsheron- und Baku-Transgressionsereignisse zurückgeht. Während dieses Zeitraums fand ein Polyploidisierungsereignis statt, das zu einem neuen hexaploiden Taxon führte und anschließend den Genfluss nach der Regression des Kaspischen Meeres verhinderte. Die zweite Fallstudie untersuchte die Evolutionsgeschichte von *Schivereckia*, die eine disjunkte Verbreitung entlang der osteuropäischen Ebene und des Balkans aufweist. Im evolutionären Kontext haben wir zusätzlich die ganze *Schivereckia*-Gruppe untersucht, die in der heutigen Zeit entweder in höheren Breiten in

der subarktischen Zone oder in Gebirgsketten der nördlichen Hemisphäre vorkommt. Unsere Analysen datierten die Radiation dieser Gruppe auf den Beginn des Pleistozäns, als niedrige Temperaturen die Artenbildung und Radiation der an Kälte angepassten Flora und Fauna förderten. Unsere Studie zeigt, dass die stark disjunkte Verbreitung von *Schivereckia podolica* die Refugialräume unterschiedlichen Alters im Pleistozän widerspiegelt und auf eine enge evolutionäre Beziehung zwischen Steppen- und Tundraflora hinweist.

Eine dritte Fallstudie konzentrierte sich auf *Sisymbrium*. Wir haben festgestellt, dass sich *Sisymbrium* zwar nicht im eurasischen Steppengürtel entwickelt hat, jedoch mehrmals unabhängig während des gesamten Pleistozäns in Süd-Nord- bzw. Ost-West-Richtung in dieses Gebiet sowie in den Mittelmeerraum eingedrungen ist. Diese Studie hat erneut bewiesen, wie variabel die eurasische Steppenflora im Hinblick auf den evolutionären Ursprung ist und dass viele Steppenelemente aus den angrenzenden Gebieten im Süden stammen. Diese Steppenelemente wanderten dann erst in den eurasischen Steppengürtel ein, nachdem dieser bereits vollständig entwickelt war. Diese Einwanderung erfolgte in zwei Richtungen: nach Westen in die euro-sibirische Steppe und/oder nach Osten in die mongolisch-chinesische Steppe. Die vierte Fallstudie untersuchte die Evolutionsgeschichte von *Capsella*. Ein GBS-Ansatz hat sich als erfolgreich erwiesen, um die biologische Geschichte junger Taxa aufzuklären. Wir haben gezeigt, dass das Steppenelement *Capsella orientalis* in der Süd-Nord-Richtung in den eurasischen Steppengürtel eingedrungen ist, auch in diesem Fall lange nachdem dieser bereits vorhanden war. Wir haben weitere Einblicke in die Evolutionsgeschichte ihrer kosmopolitischen, tetraploiden Schwesterart, *C. bursa-pastoris,* gewonnen und konnten den Ursprung der gesamten Gattung in das spätpliozäne Kontinentalasien einordnen.

Alle Studien zeigten die Bedeutung des richtigen Samples von Taxa und boten einen alternativen Verschachtelungsdatierungsansatz für eine zuverlässige Kalibrierung in Fällen, in denen keinerlei fossile Nachweise erhalten waren. Zusammenfassend lässt sich sagen, dass die Evolutionsgeschichte aller vier untersuchten Taxa von denselben Umweltveränderungen geprägt war, die eine wichtige Rolle in der biologischen Geschichte des eurasischen Steppengürtels spielten. Grundsätzlich hat sich unsere Arbeitshypothese, dass molekulare Signale in typischen Steppenpflanzenarten die Klimalandschaftsgeschichte der Steppe und die damit zusammenhängende biogeografische Dynamik von Steppentaxa widerspiegeln, als richtig erwiesen.

Chapter 1:

GENERAL INTRODUCTION

“Nothing in nature could be finer. The whole surface resembled a golden-green ocean, upon which were sprinkled millions of different flowers. Through the tall, slender stems of the grass peeped light-blue, dark-blue and lilac star-thistles; the yellow broom thrust up its pyramidal head; the parasol-shaped white flower of the false flax shimmered on high. A wheat-ear, brought God knows whence, was filling out to ripening. Amongst the roots of this luxuriant vegetation ran partridges with outstretched necks. The air was filled with the motes of a thousand different birds. On high hovered the hawks, their wings outspread, and their eyes fixed intently on the grass. The cries of a flock of wild ducks, ascending from one side, were echoed from God knows what distant lake. ... Oh, steppes, how beautiful you are!”

[N. V. Gogol, in Taras Bulba (1835) Chapter II]

Steppe characteristics

Zonal steppes are a macroclimate-driven vegetation biome dominated by graminoids (predominantly Poaceae but also Cyperaceae and Juncaceae) and to a lesser extent by perennial herbs, geophytes and therophytes, sometimes with a significant admixture of chamaephytes. They are overall too dry to sustain the permanent growth of phanerophytes and are found in semi-arid climates under strong influence of semi-permanent areas of high atmospheric pressure, causing precipitation seasonality and high temperature amplitudes resulting in occasional frost periods (modified from Hurka et al., 2019 and Wesche et al., 2016). This is by no means the only definition that can currently be found in scientific literature, however, it includes this biome’s most representative biotic as well as abiotic features and has thus been put forward in this dissertation (for alternatives, see Discussion). Despite a somewhat wordy definition, the phrasing includes not only north hemisphere grasslands (prairies in North America and steppes in Eurasia) but also South American grasslands (pampas), nevertheless, this dissertation deals with the vegetation dynamics of the zonal Eurasian steppe (Eurasian steppe belt) only. Before going into more detail on the (zonal) Eurasian steppe belt, azonal steppe concept should be briefly introduced. Azonal steppe formations can be established mostly due to edaphic properties outside of the above-mentioned climate properties. These may include mountain steppes of high altitudes, sandy steppes, and saline steppe communities on water-logged or periodically wet grounds. These azonal steppe islands are commonly found in Europe, for example in the Alps, Mainz Sand Dunes, Balkan Karst region and Lesser Poland Uplands.

The Eurasian steppe belt is the largest steppe formation worldwide, stretching approximately 8000 km from the Hungarian Basin in the West to Manchuria (Amur river) in the East (between 27° and 128° eastern longitude) and reaching up to 1000 km in its width, approximately between 48° and 57° northern latitude (Jacobs et al., 1999; Lavrenko and Karalysheva, 1993). Depending on the formal definition, the size of the Eurasian steppes varies between 8 million km^2 and 10.5 million km^2 (Schultz, 2002; Wesche et al., 2016). Due to its sheer size, it should not come as a surprise that this region represents a mosaic of numerous habitats, partially interspersed by the Altai-Sayan mountain range around the 85th east meridian. As for the definition, further steppe classification systems and latitudinal zonation lack standardised jargon and depend mostly on the background education of the researches who introduced the terminology. Three different contemporary approaches try to objectively delimitate the Eurasian steppe belt. The physiognomic approach introduced by geobotanists and plant biogeographers in the 1960s (Ellenberg and Mueller-Dombois, 1967) differentiates between forest steppe, long-grass steppe, mixed steppe, short-grass steppe and desert steppe along the latitudinal gradient. Contrarily, vegetation ecologists of the Braun-Blanquet school are more familiar with the floristic approach, separating the Eurasian steppe belt into vegetation classes, orders, alliances and associations based on the species composition and its coverage (Braun-Blanquet, 1964). The third approach, mostly used by the Russian research community, is a dominant species approach, which is to a certain extent a sort of a mixture of the two previously mentioned approaches. All three approaches have their own advantages and disadvantages and can be used depending on the research question.

In this dissertation, I put forward the subdivision of the Eurasian steppe belt introduced by a Russian geobotanist and plant biogeographer E. M. Lavrenko. Lavrenko (1969, 1970a, b) subdivided the Eurasian steppe belt into two major parts, along the Yenisei River – Russian Altai – Mongolian Altai — Irtysh River border: (I) Black Sea-Kazakhstan steppe subregion (or Euro-Siberian steppe) and (II) the Central Asian steppe Subregion (or Mongol-Chinese steppe). The former can be further divided into (I-a) the Eastern European Block with Balkano-Mesian forest-steppe province, East European forest-steppe province, and Black Sea (Pontic) steppe province and (I-b) Western Siberian-Kazakhstan Block with West Siberian forest-steppe province and Transvolga-Kazakhstan steppe province. The split between these two subregions follows roughly the 50th east meridian, splitting the present-day Caspian Sea into two equal halves. The Central Asian steppe Subregion consists of Dauro-Mongolian Block with Hangay-Daurian mountain forest-steppe province, Mongolian steppe province, and North Gobi desert-steppe province (II-a) and

Manchuro-North-western Chinese Block with Manchurian forest-steppe province, Shanxi-Gansuian forest-steppe and steppe province (II-b). These two subregions are divided by the Yin Mountains and Gobi desert in the south and the Greater Khingan in the east (see Appendix 1).

The western Euro-Siberian steppe and eastern Mongol-Chinese steppe are well defined in terms of climate, edaphic factors and floristic composition (reviewed in Hurka et al., 2019 and Wesche et al., 2016). While both of the subregions experience approximately the same amount of precipitation, the Mongol-Chinese steppe, being under influence of East Asian Monsoon system, receives the highest amount of precipitation in late summer. On the other hand, the Euro-Siberian steppe experiences the peak precipitation period earlier in the summer months, mostly in May and June. Edaphic profile differs extensively between these two regions. The Euro-Siberian steppe exhibits a latitudinal zonation along the precipitation gradient with chernozem soil in the most humid steppe regions in the north, followed by kastanozems in short grass steppes and calcisols in semi-desert and desert steppes. Contrarily, the Mongol-Chinese steppe floor consists mostly of kastanozems, cambisols and calcisols in the driest regions, and chernozems are almost entirely absent (FAO/ UNESCO Soil Map).

Different edaphic factors and shifted climate regime are responsible for extensive changes in the floristic composition. Taxa such as *Acantholimon* (Plumbaginaceae), *Chamaecytisus* (Fabaceae), *Crambe* (Brassicaceae), *Crocus* (Iridaceae), *Ornithogalum* (Liliaceae), *Salvia* (Lamiaceae), *Seseli* (Apiaceae), *Trinia* (Apiaceae), *Tulipa* (Liliaceae), and *Verbascum* (Scrophulariaceae) are exclusively or almost exclusively found in the western part of the Eurasian steppe belt (Appendix 2), while *Cymbaria* (Scrophulariaceae), *Filifolium* (Asteraceae), *Hemerocallis* (Liliaceae), *Lespedeza* (Fabaceae), *Panzeria* (Lamiaceae), *Saposhnikovia* (Apiaceae), *Schizonepeta* (Lamiaceae), and *Stellera* (Thymeleaceae) are more or less restricted to the eastern part of the Eurasian steppe belt (Smelansky and Tishkov, 2012; Wesche et al., 2016 and the literature cited therein; Appendix 3). The clear border in florogenetic composition is also mirrored on a higher ranking class level; while western steppes are dominated by *Festuco-Brometea*, the steppes east of Altai mountains are covered mostly in *Cleistogenetea squarrosae* (Ermakov et al., 2006). Nevertheless, there are a handful of species or species groups that can be found across the entire Eurasian steppe belt and are thus a proper proxy to infer the florogenesis of the Eurasian steppe belt as a whole. These include *Agropyron cristatum* (Poaceae), *Allium strictum* (Alliaceae) and *Scorzonera austriaca* (Asteraceae) species complexes, *Krascheninnikovia ceratoides*

(Chenopodiaceae), *Onobrychis arenaria* (Fabaceae) *Odontarrhena obovata* and *Sisymbrium polymorphum* (Brassicaceae), and *Veronica spicata* (Plantaginaceae) (Appendix 4).

The importance of steppes

It should be of get concern that only about 5% of the total area of temperate grasslands and shrublands are under protection, making them the least protected biome worldwide (Brooks et al., 2004; Smelansky and Tishkov, 2012; Wesche et al., 2016). As of 2020, the following areas are officially recognised as UNESCO World Natural Heritage Sites and are as such legally protected by international treaties: Saryarka saline steppes of Northern Kazakhstan, mountain (highland) steppes of Golden Mountains of Altai (including Ukok Plateau, which is believed to be one of the last remnants of the once extensive mammoth steppe biome), forest steppes of the Lake Baikal Basin and true zonal steppes of Uvs Nuur Basin and Dauria (UNESCO World Natural Heritage; Smelansky and Tishkov, 2012). It is noticeable that a big portion of these protected areas are comprised of small, insignificant azonal steppe patches of poor genetic diversity that are easier to register as protected as they do not interfere with the economic growth of countries in question (Venter et al., 2014). The main treats to the steppe habitats are of anthropogenic nature. For the past couple of millennia, humankind has been actively transforming steppe habitats into arable, genetically poor monocultures or into mine, oil and gas fields (Hurka et al., 2019).

Furthermore, the eastern steppe habitats are notoriously overgrazed and are slowly disappearing under the hooves of massive herds of cattle. Overexploitation of the natural resources and inappropriate land management have led to desertification and soil salinisation, leaving behind so called anthropogenic deserts (Saiko, 2001). While poaching does not pose such as severe threat as in the rainforests, many animal species including the Saiga antelope (*Saiga tatarica*) and the Saker falcon (*Falco cherrug*) have been driven to the verge of extinction and a number of plant species are still being sold on the black market (Moshkin, 2010; Neronov et al., 2012; Smelansky and Tishkov, 2012). To make matters even worse, climate change has had a serious impact on the steppe dynamics. A prime example is Mongolia, where the amount of precipitation declined sharply over the years, leaving the soils parched and highly susceptible to now common steppe fires (Menzel et al., 2016). Consequently, long-term reconstructions suggest that the permafrost layer in Mongolia will be gone in the regions of lower altitudes by the end of this century, further accelerating desertification (Azzaya and Khaulenbek, 2008).

Unfortunately, little effort has been invested to sustain these steppe ecosystems, despite being immensely important. Studies estimated the yearly value of a steppe hectare in 2009 was between 200 and 450 USD (Smelansky and Tishkov, 2012). Together with peat bogs, steppe habitats are one of the most important carbon sinks worldwide (Dass et al., 2018; White et al., 2000; Yang et al., 2019) and as such mitigate climate change. Their sustainable high soil fertility supports the immense crop production and cattle breeding (Smelansky and Tishkov, 2012) without which many of the countries' economies would collapse. Forest steppes and steppes also have a surprisingly high net primary production (Schultz, 2002; Zlotin, 2002) and harbour a notable diversity of flora and fauna (Kier et al., 2005; Kuzemko et al. 2016; Polyakova et al. 2016), despite only a handful of grasslands being considered as biodiversity hotspots (Dengler et al., 2014; Habel et al., 2013). Grasslands can supply additional services, such as erosion control, water supply and flow regulation as well as habitat and food for economically important pollinators (Bengtsson et al., 2019; Lemaire, 2007).

Summary of the onset and development of Eurasian steppe belt

The contemporary Eurasian steppe belt is northwards bordered by temperate deciduous forests, cool mixed forests, cold evergreen needle-leaf forests and cold deciduous forests, and southwards by semi-desert and desert vegetation (sensu Binney et al., 2017). However, this was not always the case. The first reliable physical evidence of steppe prototypes are dated to late Eocene/early Miocene (34-20 MYA) and placed into the Kazakhstan Plains (Akhmetiev and Zaporozhets, 2014; Velichko, 1999). In parallel, claims that central China has been dominated by steppe-like environments as early as in the late Palaeocene (55 MYA) have been put forward and are based on pollen fossil records of Asteraceae, Chenopodiaceae and Polygonaceae (Ma et al., 1998, 2012). This view, however, has been questioned on numerous occasions (Jiang and Ding, 2009; Quan et al., 2012; Wang and Shu, 2013; Utescher and Mosbrugger, 2007).

To fully understand the climatic reasons influencing the onset and development of the Eurasian steppe belt, we must return to the Palaeocene (66-56 MYA). During that time, the northern hemisphere was bathing in photosynthetically productive lush forest vegetation. While the northern and eastern parts were covered in boreal flora of the humid warm-temperate climates, the south and the west were directly influenced by the Tethys Ocean (Akhmetiev and Zaporozhets, 2014). This provided a more humid and subtropical-like climate, in which paratropical rainforests could thrive. The Palaeocene/Eocene transition was accompanied by a maximum warming caused by an

uninterrupted heat transfer between the Tethys and the Atlantic Ocean (Akhmetiev, 2007). This allowed the Tethys flora to extend north- and eastwards, pushing the boreal flora into the contemporary Arctic region (Hurka et al., 2019 and the references therein).

The end of Eocene was demarcated by the global cooling and a tendency towards aridification as a result of the isolation of the West Siberian Sea from the Arctic Basin, partial confinement of the Paratethys, and the onset of Antarctic glaciations (Akhmetiev, 2010; Zachos, 2008). During that time, sclerophyllous shrub vegetation developed in the most continental parts of Central Asia (Harzhauser et al., 2016; Popova et al., 2017; Rögl, 1999; Sun et al., 2014). The steady uplift of the Alpine fold belt (including Alps, Carpathians, Dinarides, Taurus and Alborz mountain ranges) continued limiting the water exchange between the Paratethys and the Mediterranean (Rögl, 1999; Schulz et al., 2005). Oligocene continued with a cooling trend that eradicated the Tethys flora and replaced it with a more robust temperate deciduous mesophyllous coniferous-broadleaved Turgai flora (Akhmetiev and Zaporozhets, 2014). The cooling trend was probably a result of desiccation of the two most important sources of humidity of Central Asia — the Turgai Strait and the West Siberian Sea (Velichko, 1999, Velichko and Nechaev, 1999).

Miocene cooling accompanied by increased seasonality of precipitation and temperature amplitudes, caused a demise of warm-loving Turgai floral elements such as *Glyptostrobus*, *Metasequoia* and *Taxodium* as well as temperate-loving *Acer*, *Alnus*, *Fraxinus*, *Quercus* and *Tilia* dendroflora, and in turn promoted the expansion of open steppe-like habitats in the temperate zone (Popova et al., 2017). In the boreal zone, however, the Turgai flora was replaced by cold-adapted coniferous species and small-leaved dendroflora including *Abies*, *Betula*, *Larix*, *Picea*, *Pinus*, *Populus*, *Salix* taxa (Nikitin, 2007). During this period, the complete isolation of the later brackish Pannonian Lake took place, which continued throughout the Miocene until this body of water finally disappeared in the Pliocene (Popov et al., 2006; Ter Borgh et al., 2013). The eastern portion of the Paratethys also changed into a system of fragmented intracontinental brackish basins that continued to dwindle (Hurka et al., 2019). There are extensive pollen records of *Artemisia*, *Ephedra*, Brassicaceae, Chenopodiaceae and Poaceae as well as faunal fossil records that prove forest steppe/steppe existence in Central Asia and gradual spread into the adjacent areas during the Miocene and Pliocene (Mai et al., 1995, Nikitin, 2007; Popova et al., 2013, 2017). The Eurasian steppe belt we know today is dated to the Miocene/Pliocene border and came into existence 5.33

MYA at the latest (Denk et al., 2018; Mai 1995, Utescher et al., 2017; Velichko and Nechaev 1999; Wu et al., 2007).

Around 2.58 MYA the progressive cooling changed its pace, introducing a new epoch — the Pleistocene (Gibbard and Head, 2009). This era is characterised by continuous more or less extensive glaciation periods, interspersed by warmer interglacial cycles, causing repeated massive vegetation turnovers in Eurasia (Velichko, 1999). Before the first major cool-down event, the northern Eurasia was covered in mixed forests of deciduous broadleaf trees and conifers (Frenzel, 1968). However, when climate began to cool down the overall humidity decreased with it, the forests retreated and gave space to dry adapted steppe-like vegetation. In contrast, the interglacial cycles promoted the growth of forests, which overgrew the existing steppe vegetation. It in part survived in areas where the amount of water was not high enough to allow for forest growth. This dynamics of continuous expansions and contractions of the dendro- and steppe flora, respectively, along the latitude continued until 11700 years ago (Frenzel, 1968).

The Eurasian steppe belt, however, also suffered from repeated longitudinal splits. During the Early Pleistocene (2.6-1.8 MYA) the Akhchagyl transgression of Caspian Sea keeping the Eastern European Block and the Western Siberian-Kazakhstan Block divided for about one million years, followed by a number of additional short-lived transgression events throughout the Pleistocene (Forte and Cowgill, 2013; Popov et al., 2006; Tudryn et al., 2013) might or might not have been caused by the glacial cycles (van Baak et al., 2013). These transgressions represent the only time when the brackish basins of the former Eastern Paratethys Ocean reconnected. Massive glaciers in the north served as dams, causing extensive water spills of Ob, Irtysh and Tobol rivers much further to the south, turning the steppe habitats adjacent to riverbeds into marshes (Arkhipov and Volkova, 1994; Astakhov, 2008). The first concrete evidence on the periglacial steppes during cold phases dates back to around 1 MYA, indicating a new key element in the florogenesis of the Eurasian steppe belt (Hurka et al., 2019). Without a modern analogue, the biome that dominated Eurasia during the glacial periods of the Middle and Late Pleistocene bears an illustrative name: periglacial steppe-tundra. While there have been many different terms describing essentially the same biome, I will follow the definition put forward in Hurka et al., 2019.

As the name suggests, two different contemporary components are included into the definition. While the periglacial steppe was consistently found in the lower latitudes of the temperate zone (~60°N-45°N) and comprised predominantly of steppe-floral elements with a portion of tundra-

floral elements, the periglacial tundra was usually found north of the periglacial steppe (from ~60°N northwards) and was comprised of tundra-floral elements with a smaller portion of steppe-floral elements (Hurka et al., 2019). The line is of course artificial and moved during the glacial cycles. Contemporary analogues of this biome cannot be found (but look Chytrý et al., 2018) due to climate differences. Pleistocene subarctic regions experience far higher temperature amplitudes and less precipitation (Guthrie, 2001; Zimov et al., 2012), thus promoting the coexistence of steppe elements such as *Artemisia* (Asteraceae), Chenopodiaceae, *Astragalus*, *Oxytropis* (both Fabaceae), *Potentilla* (Rosaceae), *Stipa* and *Festuca* (both Poaceae), arctic-alpine taxa such as *Carex*, *Kobresia* (both Cyperaceae), *Draba, Smelowskia* (both Brassicaceae), *Dryas* (Rosaceae) and *Papaver* (Papaveraceae) and less visible moisture-loving *Eriophorum* (Cyperaceae), *Hierochloe* (Poaceae), *Ranunculus* (Ranunculaceae) and *Vaccinium* (Ericaceae) (Kienast et al., 2005; Nimis et al., 1998; Willerslev et al., 2014; Yurtsev, 1982, 2001).

Despite popular opinion, the whole continent of Eurasia was never completely covered in ice during the Pleistocene glaciations. While the Early and Middle Pleistocene glaciations were more extensive (Astakhov, 2011; Ehlers et al., 2013; Velichko et al., 2011), there is an accumulating pile of evidence suggesting that the Late Pleistocene glaciations were restricted to mountain ranges and plateaus only (Gobejishvili et al., 2011; Lehmkuhl et al., 2011; Matoshko, 2011; Stauch and Lehmkuhl, 2011; Svendsen et al., 2014; Urdea et al., 2011). The end of the last significant temperature drop 11.5 kya (Younger Dryas) accompanied by a massive megafaunal extinction (Sandom et al., 2014) demarcated the beginning of the Holocene (Cohen et al., 2013). The temperature dynamics of the Holocene resembles the Pleistocene, albeit with smaller amplitudes (Velichko, 1999). Immense increase in rainfall at the beginning of Holocene caused a demise of periglacial steppe/tundra, which was gradually replaced by boreal and temperate mixed forests and by contemporary tundra vegetation within the Arctic Circle (Binney et al., 2009, 2017; Tarasov et al., 2000, 2007). During the Middle Holocene (8.2-4.2 kya) steppes have found themselves yet again pushed deep into the south and most of the contemporary Eurasian steppe belt was covered in a diverse mosaic of taiga forests in the east, and mixed forests and forest-steppes in the west (An et al., 2008; Magyari et al., 2010; Tarasov et al., 2000; Yang et al., 2011; Zhilich et al., 2017). It is hypothesised that most of the eurythermic steppe species survived the unfavourable climatic conditions in this mosaic, while strictly thermophilic steppe species migrated northwards into the steppes from their respective refugia only recently, when the final increase of aridity pushed the forests towards the north (Hurka et al., 2019, Magyari et al., 2014).

The onset and development of the Eurasian steppe belt was under strong influence of the Cenozoic geological events. The most important ones that have caused a continuous aridification of Asia were: global decrease in the atmospheric temperatures, closure of the Arctic Sea – Turgai Basin strait, continuous retreat of Tethys and Paratethys, uplift of the Tibetan Plateau as well as Pamir, Tian Shan and Altai mountain ranges, and the onset of Asian monsoon systems (reviewed in detail in Hurka et al., 2019 and the references therein).

Steppe refugia

During unfavourable environmental conditions flora of a particular region can either adapt and survive, go extinct or (passively) retreat into an area of any size in which a taxon can endure and survive the unfavourable conditions — a refugium (definition altered by Bennett et al., 1991). There has been a long lasting debate on the definitions and types of refugia, resulting in an introduction of numerous, and in my opinion, sometimes completely valueless terms. For the sake of better understanding the terminology used in the publications, I shall briefly mention the two main refugia types. Whereas the glacial refugia, providing shelter for warm-loving taxa during the glacial periods, are very well investigated, little effort has been invested into elucidating the spatial distribution and extent of the interglacial refugia, which allowed for the cold-adapted taxa to survive the warmer and wetter periods. Reasonably assuming that woody species of the past experienced the lower temperatures and increased aridity in the same way as contemporary arbour vegetation, i.e. as unfavourable, one concludes that they must have survived the unfavourable conditions in glacial refugia. In terms of steppe, however, this assumption should not be made. Contemporary steppe vegetation consists of a mosaic of warm-adapted (or rather dry-adapted) as well as cold-adapted or eurythermic taxa, preventing us to assign them to a simple one-sided refugium. Nevertheless, several steppe refugial centres have been continuously put forward in the recent years. A peculiar plant association has led the scientist to recognise the Ural and the adjacent uplands of the European Plain to represent an interglacial refugium (Gajewski, 1934; Kozo-Poljanski, 1928, 1929; Litvinov, 1902). A recent genetic study supported this idea (Friesen et al., 2020). Furthermore, Altay-Sayan Mountains harbour a number of endemic species, which was secondarily supported by the fossil evidence and modelling of the Last Glacial Maximum environments (Hais et al., 2015; Horsák et al., 2015; Pavelková Řičánková et al., 2014, 2015). Southern Ukraine (especially the kurgans) with adjacent areas along the Black Sea coast are also believed to have played an important role as a refugial centre for steppe vegetation, supported by

exclusively genetic studies (Dembicz et al., 2018; Meindl et al., 2016; Sudnik-Wójcikowska et al., 2011; Wódkiewicz et al., 2016)

Overall high genetic diversity and presence of private alleles are the two genetic indicators of a refugium, followed by (in)direct paleontological evidence and palaeoclimatological reconstructions. The latter two need to be treated with caution, however, due to a possible bias. The research community still lacks knowledge on the taphonomy of the fossil record and thus repeatedly comes across the problem appropriately descried by the aphorism "Absence of evidence is not evidence of absence" (Sober, 2008). Furthermore, palaeoclimatological reconstructions are problematic as they rely mostly on a flawed coexistence approach (Mosbrugger and Utescher, 1997) that has been extensively criticised in the last decade (Grimm and Denk, 2012; Grimm and Potts, 2016). In other words, this approach functions under multiple requirements that are not necessarily fulfilled. The most striking assumptions this model presupposes are that the climatic requirements of a fossil taxon are similar to those of its nearest living relative and that the absence of a fossil in a paleo-assemblage is an evidence of true absence (Utescher et al., 2014). Thus, while this method might uncover dramatic climate changes, minor climate fluctuations of potentially great importance for the floral dynamics remain hidden.

Contemporary problems in steppe research

It is evident from the previous chapter that there is a growing amount of literature that deals with the onset and development of the Eurasian steppe belt on local bases. A closer look, however, reveals that the knowledge on the evolution of the steppe belt as a whole is patchy. While there is substantial amount of information dealing with the consequences the Last Glacial Maximum had on the flora and fauna of Europe, there is proportionally little reference literature on older glacial periods, their extent and influence on the steppe development outside of Europe. What are the reasons behind that? Firstly, due to the sheer size of the Eurasian steppe belt and the biological diversity that comes with it, it is impossible to locally investigate a response of flora and fauna, and apply these results to the whole steppe belt. Secondly, the steppe belt is part of countries with ongoing war conflicts and an unsettled political landscape, making scientific expeditions rather problematic. Thirdly, while published research done in certain areas might be abundant, it is inaccessible to the western scientific communities due to language barriers. Due to the isolation of research societies, each developed its own 'standardised' chronostratigraphy (Hurka et al., 2019),

which further impairs research communication and can lead to some serious misinterpretations. Overall, while the phytosociological aspects of the Eurasian steppe belt flora are well known, the extensive knowledge on its florogenesis (the onset and development of flora) is still lacking.

Present knowledge on the Eurasian steppe belt is based mostly on plant macrofossils, pollen and spore fossil assemblages, sedimentary organic biomarkers, palaeopedological and phytolithic records, fossils of herbivores that grazed on steppe vegetation, stable isotope records in palaeosoils, sediments and biominerals, ice profiling (Berke, 2018; Beverly et al., 2018; Bredenkamp et al., 2002; Higgins, 2018; Jacobs et al., 1999; Mander et al., 2018; Ng et al., 2010; Tarasov et al., 2000), and recently also on molecular (Zimmermann et al., 2019) as well as aDNA-based studies (Epp et al., 2018; Willerslev et al., 2014). However, all these proxies have their own flaws and biases, thus the interpretations based on single one of them should be treated with caution.

Macrofossil records are sporadic and thus provide only a local picture in a narrow period (Kienast et al., 2005). An additional problem is the above-mentioned lack of knowledge on fossil preservation (Binney et al., 2017; Birks and Birks, 2000). Furthermore, pollen records are notoriously biased towards wind-pollinated taxa, which might skew the interpretations on the palaeolandscape due to its high dispersal potential (Binney et al., 2017; Birks and Birks, 2016). In addition, the resolution of pollen analyses differs greatly between different taxonomic groups (Jørgensen et al., 2012). Dental wear of grazers and browsers might be a good indirect indicator of which vegetation these animals fed on. However, a number of factors complicate the interpretation extensively described in Strömberg, 2011. It is also problematic to infer whether the development of the steppe was the trigger due to diversification of herbivores. Maybe it was the other way around or maybe these two systems developed independently from each other. Although the grasses are nowadays dominating the steppe regions worldwide, they might not be the best proxy to infer the steppe development for two reasons. The steppes in the past were not dominated by grasses but could consist of up to 50% of forbs (Willerslev et al., 2014), and the evolutionary history of grasses began in warm tropical or subtropical climates (Edwards and Smith, 2010; Linder and Rudall, 2005).

Contemporary and past soil profiles might be of independent origins and different ages and are thus unreliable as a single proxy (Eckmeier et al., 2007). aDNA and sedDNA-based studies suffer from high contamination levels and computational threshold biases, potentially recording contemporary laboratory contaminants and overlook vegetation signals of less abundant taxa (Alsos, et al., 2016;

Edwards et al., 2018; Parducci et al., 2017). Isotope investigation of different depositions might also be problematic due to poor time constraints and incorrect extrapolations. The onset of the Asian monsoon system serves as an example. Some studies have placed the origin of monsoonal circulation into the Eocene (Licht et al., 2014), while some estimated its age to be much younger and place it at the end of the Miocene, seemingly coinciding with the uplift of the Qinghai-Tibet Plateau (An et al., 2001). We now know that the onset of the Asian monsoon system as well as the Tibetan uplift are of much higher age and that they arose independently, and the young age was an artefact of the methodology and incorrect assumptions (reviewed in Renner, 2016). Moreover, it is known that thickness and ice-volume significantly depend on the elevation (Hassan et al., 2019), however, current glaciological methods do not take elevation dependence into account, making their estimation potentially highly misleading (Osmaston, 2005).

Rough outline and biological history of Brassicaceae

The mustard family (Brassicaceae) belongs to the order Brassicales and is distinctive from other angiosperm families in including representatives of herbs found in temperate regions of the world, with glucosinolates, tetramerous and primarily choripetalous perianth, 2+4 tetradynamous stamens, a gynoecium with a superior, 2 carpellate/loculate ovary, with axile-parietal placentation, and a usually 2-valved, dehiscent fruit with a false septum (Franzke et al., 2011; Simpson, 2019). Nevertheless, despite belonging to one of the middle-sized plant families, they were distributed worldwide and homoplasious features have been observed on several occasions. These include woody lianas in the South American *Polypsecadium* (Thelypodieae) and *Cremolobus* (Cremolobeae), South African *Heliophila scandens* (Heliophileae) and Australian *Lepidium scandens* (Lepideae), deeply dissected petals in Chilean *Schizopetalon* (Schizopetaleae), North American *Ornithocarpa* (Cardamineae) and *Dryopetalon* (Thelypodieae), and Eurasian *Berteroa* (Alysseae) and *Draba* (=*Erophila*) *verna* (Arabideae) (Franzke et al., 2011). Synapomorphies might be present in some taxa, such as wind pollination in *Pringlea antiscorbutica*, monoecy in *Megacarpaea megalocarpa*, deeply bifid cotyledons in *Schizopetalon*, polyandry in *Megacarpaea polyandra* (hence the name), or alterations in stamen morphology, as for example reduction to two stamens in *Lepidium* and unequal stamen lengths in *Streptanthus* (Al-Shehbaz, 2012; Franzke et al., 2011).

Brassicaceae comprises c. 4000 species currently classified into 3+1 (Franzke et al., 2011) or 5+1 (Nikolov et al, 2019) evolutionary lineages and further sorted into 52 monophyletic tribes (Al-

Shehbaz, 2012; Koch et al., 2018). These are, however, only temporary numbers, as a number of the genera are still awaiting tribe assignment, and new species and taxonomic rearrangements are being published in a continuous pace (Al-Shehbaz, 2015; Brock et al., 2019; Chen et al., 2018; German et al., 2016; German and Al-Shehbaz, 2018; Heenan, 2017; Moazzeni et al., 2018; Özüdoğru et al., 2019 to only mention a few new ones from within the last 5 years). After the Poaceae, this taxon is probably one of the most well studied plant families due to its immense economic potential. The most important vegetable crops belong to *Brassica oleracea* and include broccoli (var. *italica*), Brussels sprouts (var. *gemmifera*), cabbage (var. *capitata*), cauliflower (var. *botrytis*), kohlrabi (var. *gongylodes*) and kale (var. *acephala*) (Al-Shehbaz, 2012; Appel and Al-Shehbaz, 2003). The list continues with edible *Raphanus* (radish), *Eruca* (arugula) and *Nasturtium* (watercress) (Al Shehbaz, 2011). Several condiments from this family are used in cuisines worldwide on a daily basis, including *Armoracia* (horseradish), *Eutrema* (wasabi), *Sinapis alba* (white mustard) and *Brassica nigra* (black mustard) (Al-Shehbaz, 1985). Much of the oil industry relies on *Brassica napus* and *B. rapa*, *Camelina sativa*, *Crambe abyssinica*, *Lepidium sativum* and *Physaria* oil-rich seeds (Cartea et al., 2019; Diwakar et al., 2010; Moser, 2010; Salywon et al., 2005; Wang et al., 2000). Moreover, a number of Brassicaceae from the genera *Aubrieta*, *Aurinia*, *Erysimum*, *Iberis*, *Hesperis*, *Lobularia*, *Lunaria*, *Matthiola* have found their way into our gardens, where they thrive in sunny or partially shaded calcareous spots (Al-Shehbaz, 2011). *Arabidopsis thaliana* was the first plant (probably due to its small genome) and the third multicellular organism whose genome was successfully sequenced and published (AGI, 2000).

This achievement revolutionised experimental biology in general and raised interest particularly in the family of Brassicaceae (Al-Shehbaz, 2012). Newly acquired information has been applied on numerous close relatives of *Arabidopsis thaliana* since then, generating a number of new model organisms in plant biology. These include *Capsella grandiflora* and *C. rubella* as well as *Arabidopsis lyrata* to investigate self-incompatibility system (Foxe et al., 2009, 2010; Guo Y-L et al., 2009; Nasrallah et al., 2004), *Arabidopsis halleri* and *Noccaea caerulescens* as potential phytoremediation plants for heavy metal accumulation (Deng et al., 2016; Marie Muehe et al., 2015), *Boechera* and *Phoenicaulis* to investigate apomixis (Lovell et al., 2013; Mandáková et al., 2020), *Capsella bursa-pastoris* to investigate effects of polyploidisation (Cornille et al., 2016; Douglas et al., 2015), *Cardamine* species to infer evolutionary advantages of hybridisation as well as the function of the RCO gene complex (Hay and Tsiantis, 2016; Mandáková et al., 2019),

Thellungiella to search for genes responsible for salt tolerance (Wu et al., 2012) and many others (Franzke et al., 2011).

The sequencing of the *Arabidopsis thaliana* genome revealed that at least 60% of its genome shows signs of some sort of genome duplication (AGI, 2000), which majorly changed the research perspective on the evolution of Brassicaceae genomes. Further studies confirmed the existence of several consecutive major whole duplication events in all angiosperm lineages, with a possible exception of *Amborella trichopoda* (Cui et al., 2006; Soltis et al., 2009). The core Brassicaceae have been subjected to at least three different genome duplication events (Franzke et al., 2011). While the At-γ duplication was linked to the potential diversification of all angiosperms (De Bodt et al., 2005; Van de Peer et al., 2009), the At-β duplication postdates the Caricaceae–Brassicaceae split, at ~124 MYA (Kagale et al., 2016). Traces of the third subsequent At-α whole genome duplication that took place ~47 MYA were recovered in all Brassicaceae representatives, except in the early diverging *Aethionema*, thus specific for core Brassicaceae only (Kagale et al., 2016; Lysak and Koch, 2011; Schranz and Mitchell-Olds, 2006). This 'tale of duplications' (Van de Peer et al., 2009), however, does not end here. There is evidence that numerous other Brassicaceae taxa underwent mesopolyploidisation followed by a subsequent diploidisation event and potential chromosome rearrangements (Mandáková et al., 2010a, b, 2012, 2017a, b, 2018). Nonetheless, several Brassicaceae lineages exhibit evidence of At-α duplication only, including Boechereae, Cardamineae, Descurainieae, Eutremeae, Isatideae, Sisymbrieae and others (Lysak et al., 2005, 2006; Mandáková et al., 2008; Schranz et al., 2007). While it is generally believed that the whole genome duplications facilitated the diversification of the angiosperms (Landis et al., 2018; Soltis et al., 2009; Tank et al., 2015), this causation is not as clear in the mustard family (Hohmann et al., 2015; Huang X-C et al., 2020).

Brassicaceae is one of the most prominent plant families of the steppe habitats (Bone et al., 2015), whose diversification started in cool and dry environments around 34 MYA at the Eocene-Oligocene boundary and continued throughout the Oligocene/Miocene transition, accompanied by cooling and increased continentality (Couvreur et al., 2010; Guo X et al., 2017; Hohmann et al., 2015; Huang C-H et al., 2016, Huang X-C et al., 2020). Coming to a reliable age estimation was, however, extremely difficult. The mustard family lacks fossils that could potentially be used as calibration points (see Discussion) and the ones that are present cannot be reliably placed into the phylogenetic context (Franzke et al., 2012). An alternative source for reliable fossils are the closely

related families that constitute the order Brassicales together with Brassicaceae. These include a flower fossil of *Dressiantha bicarpellata*, dated to early late Cretaceous (Gandolfo et al., 1998), a Palaeocene leaf fossil of an Akaniaceae representative (Iglesias et al., 2007), and a silicified piece of wood of *Capparidoxylon holleisii*, dated to the warm Miocene Burdigalian period (Selmeier, 2005).

An alternative to the primary calibration points using fossils, one can use secondary calibration points by using stem and crown ages estimated from other studies. This approach, however, is as accurate as the study from which the calibration points are used and the reuse of a certain node age might results in a compounded error (Franzke et al., 2012). There are numerous alternatives to test the accuracy of the time estimation analyses. One of them is the use of previously published substitution rates of separate loci (Kay et al., 2006) or the whole genomes (Schwarz et al., 2017; Wicke et al., 2014; Xu et al., 2015). Nevertheless, these methods come with their own disadvantages, as substitution rates of separate loci as well as whole genomes between species might differ due to different biological history of taxa in question (Kay et al., 2006; Wicke et al., 2014). Furthermore, even within the genome, coding and non-coding positions might differ in their substitution rates, albeit this seems to be extremely inconsistent throughout the angiosperms (Sloan et al., 2012; Weng et al., 2016, 2017; Zhu et al., 2016). Another possible alternative to test for the accuracy of time estimation analyses is to compare the time spans with the geographic components. For example, as an island endemic clade(s) could not have occupied the island before it rose out of the ocean, allopatric species did not evolve prior to the allopatric barrier that caused the speculation in the first place, it is highly unlikely that a taxon occupied a region at a certain point in time that exhibited unfavourable conditions for that taxon in question at that exact same time, etc. All these assumptions must nevertheless be treated with extreme caution and the methodology used in my publications, i. e. fossil calibration (if available) coupled with different secondarily calibrated runs accompanied with substitution rates seems to be a proper way to go.

After inferring the approximate time span of the biological history of a taxon in question, another question arises — the historical biogeographic aspect, which aims to explain the geographic distribution of the taxon in question in terms of its evolutionary history, taking into account environmental factors shaping the distribution of individual organisms at local spatial scale (Avise, 2000; Gugger et al., 2010; Morrone et al., 1995). The mustard family can be found on all continents (Appel and Al-Shehbaz, 2003), protruding even deep into the Antarctic Polar Front. The species

Pringlea antiscorbutica of the monotypic genus *Pringlea* (Thelypodieae) can be found in extremely inhospitable environments of Crozet, Heard, McDonald, Kerguelen, and Marion-Prince Edward archipelagos (Hooker, 1947). The highest number of taxa can be found in temperate regions of the Northern Hemisphere, and the family is almost absent from the tropical zone, with an exception of montane and alpine areas (Appel and Al-Shehbaz, 2003). Most important diversification centres of the Brassicaceae is the Irano-Turanian floristic region (sensu Takhtajan, 1986) and Mediterranean region, followed by dry North American West and Saharo-Sindian region (Al-Shehbaz 1984; Hedge 1976; Rollins 1993).

Due to the highest diversity, the Irano-Turanian region is thought to represent the centre of origin of the mustard family. This idea is further supported by the earliest diverging lineage *Aethionema* also having the highest diversity in the western part of the Irano-Turanian region (Al-Shehbaz et al., 2006; Mohammadin et al., 2017). The pattern of high species diversity continues into the sister group of Brassicaceae — Cleomaceae (Edger et al., 2015), which is also believed to have originated in the arid environments of Central Asia (Cardinal-McTeague et al., 2016; Feodorova et al., 2010). High species diversity, however, does not necessarily indicate the centre of origin of a taxon in question, but could point towards a refugium area or generally an area of low extinction rate. Thus, such claims should be investigated rigorously. Tightly connected with the Eurasian steppe belt, Brassicaceae is a proper proxy to infer its past vegetation patterns for the following reasons: (I) its putative origin in the Irano-Turanian floristic region, which the Eurasian steppe belt is also at least partially a part of, (II) its crown age that coincides with the onset of the Eurasian steppe belt, (III) its diversification and consequent radiation, which were highly influenced by the aridification of Central Asia; a pattern also observed in the development of the Eurasian steppe belt, (IV) a plethora of already available genetic information and methods that make further investigations on related Brassicaceae taxa less complicated, and (V) an extensive taxonomically curated databank (https://brassibase.cos.uni-heidelberg.de/), making the taxon comparisons more straightforward and objective.

Specific research questions and aims

The overall dynamics of the Eurasian steppe belt was appropriately summarised by German geographer and botanist B. Frenzel who said that the recent vegetation of northern Eurasia had resulted from a relentless contest between steppe and forest (Frenzel, 1968). Nevertheless, little is known about the florogenesis of the Eurasian steppe belt as well as its refugial areas, which comes

as a surprise as the floristic composition of this area is very well documented (Lavrenko 1969; Meusel et al. 1965, 1978, 1992; Walter, 1974). To extend and deepen our understanding on the florogenesis of the largest grassland region in the world, the genetic footprint of several steppe species has been investigated. The working hypothesis is that molecular signals in typical steppe plant species reflect the climate-landscape history of the steppe and the biogeographic dynamics of steppe taxa, and thus allow for a finer resolution of the florogenetic history of the steppe belt in comparison to the floristic and fossil-based methods. The project aimed to answer the following questions:

Question 1: The fossil record indicates that the Eurasian steppe developed in Central Asia earlier and then gradually expanded into Western Europe. Thus, steppe plants with an Eurasian distribution should have first evolved in Asia and later extended their distribution area to the west. Does molecular evidence support this hypothesis?

Question 2: Despite being tightly connected to the steppe biome, some taxa might have spread through the Euro-Siberian steppe belt only after its complete formation. Does molecular evidence support more recent east-to-west, west-to-east or centre-to-west/east migration patterns?

Question 3: Steppe formation may have occurred independently at different places and in different times. If this were the case, one would expect colonisation of steppe habitats from different source populations. Does molecular footprint show such colonisation patterns?

Question 4: The distribution areas of plant taxa stretching over the entire Eurasian steppe are often divided by their Asian and European parts with a split in the Pontic steppe or in western Siberia/ Kazakhstan, while other steppe plants do not display such a distribution disruption. What are the biogeographical reasons behind these differences and how common are either of these distribution patterns?

Question 6: Where were the steppe refugia during the cold phases of the Quaternary, what was their role in the extension of the steppe habitats during the warm phases, and how did they differ from the forest refugia?

Question 7: Did the Altai-Sayan region indeed play a significant role not only as a forest but also as a steppe refugium during the Quaternary — a view introduced by Lavrenko, 1930, in absence of a molecular proof and is nowadays treated as a common truth?

Chapter 2:

GENERAL DISCUSSION AND CRITICAL POINTS

The importance of geographic distribution of species has been raised as early as in the mid-19th century (Darwin, 1859; Wallace, 1876). Scientific fields of biogeography and phylogeography are thus more than 150 years old. However, only in the past 30 years, these scientific disciplines have prospered due to the development of molecular methods, improved geographical occurrence data and better climate models (Avise, 2000; Renner, 2016). These now allow us to investigate processes that have shaped the geographic distributions of organisms and ecosystems through time and space in more detail, and help us understand the contemporary distribution of organisms and ecosystems. Biogeography itself, however, cannot explain the contemporary distribution patterns alone (Ricklefs, 2004; Wiens and Donoghue, 2004), and additional in-depth knowledge on ecological, climate and evolutionary processes is thus necessary. Nevertheless, understanding the relative role these processes have played in shaping the composition of contemporary biotas is extremely challenging (Lomolino et al., 2006; Wiens, 2012).

In this dissertation, I have tackled this issue and in the four presenting scientific publications illustrated that different taxa in question have been under different influences of the above-mentioned processes depending on time and space in their evolutionary history. In the light of current rapid environmental changes accompanied by an increased extinction trend, understanding of processes that generate and maintain this exact biodiversity is of utmost importance, if we want to prevent ecosystem destruction (Smith et al., 2005). The results of this dissertation provide exactly that — i) a baseline against which the future biodiversity changes can be compared to and ii) they show the past response of particular taxa to climate change, which might help us to predict how these taxa could react to the contemporary anthropogenic changes (Erkens, 2013; Purvis et al., 2005).

Despite recent methodological developments, there are still many open question about how accurate the methods used in the field of biogeography are, as many include simulations of the past and/or the future. In the following chapter, I explore the methodological issues I have encountered during the course of this project, answer critical questions asked by the reviewers, and provide potential in-depth solutions or at least suggest potential ways on how to minimise the liability of the used methods. Furthermore, I will discuss general terminological issues, the importance of taxonomy in organismic research and conclude with a general summary of the inferred results and an outlook.

When is Elster also Oka glaciation and grassland a steppe?

The Eurasian steppe belt is the vastest grassland region worldwide and has thus attracted the attention of three biggest research communities of the Old World — European, Chinese and Russian. Due to the political as well as scientific reticence, these research communities had not cooperated in the past and in their research failed to acknowledge the Eurasian steppe belt as one entity. Investigation of the geological history of only closely located geographical units and a reluctance to cooperate resulted in a development of several independent and incompatible terminologies specific to either the research realm or a geographic unit in question. Current geological scientific society struggles to unify the terminology system for a better scientific communication exchange. While the International Chronostratigraphic Chart helped develop a standardised system of epochs and their time spans (https://stratigraphy.org), older literature still relies mainly on regional stages whose names and time spans differ among geological events specific to contemporary China, East European Plain, Siberian Plain and Europe, respectively. (Hurka et al., 2019 and the references therein). The differences might be attributed to the geological activity of a certain unit in question. For example, there has been an extensive geological activity recorded in the East European Plain during the late Middle and early Upper Miocene, resulting in an introduction of six (according to Russian scientists) standardised substages. Contrarily, during the same period no significant or extensive geological events have been inferred from western China, which divides the time period into only two substages (Table 2 in Hurka et al., 2019). One should thus always rely on the time scale of a certain event and not necessarily on the name of the period, because the name might not be known in another research community or even worse, it might be perceived as a different one.

A similar problem arises when different names essentially represent the same geological event, such as a glaciation cycle. While the term Cromer Complex is known only from the European realm, its dynamics has also been investigated in the East European and Siberian Plain, just not under any specific umbrella term that would combine this series of glacial and interglacial cycles. Furthermore, while the three most massive glaciations in Europe are named Elster, Saale and Weichsel (or Würm in the Alps) after four different rivers in Germany and Poland, essentially the same glaciations are known under the names Oka, Dnieper and Valdai glaciation from the East European Plain, and Shaitan, Samarovo/Tazov and Sartan glaciation from the Siberian Plain (Table

5 in Hurka et al., 2019). These are only a couple of examples of unstandardised terminology that hinders open scientific dialogue among scientists of different nations.

As put forward in the introduction, the literature is overflown with different more or less congruent definitions on what characterises steppes and grassland biomes. Similarly to the German definition of a flower (terminological dispute between “Blüte” vs. “Blume”), the definition of a grassland and/ or steppe might vary depending on the scientific background of the person putting it forward. In this dissertation I follow a combinatorial definition introduced from two individual research aspects — one having a strong evolutionary background (Hurka et al., 2019) and the other one having a more pronounced ecological background (Wesche et al., 2016). While it is clear that steppes represent only one particular type of grasslands found in temperate zones only (Hurka et al., 2019; Wesche et al., 2016), the formal definition of a grassland is the least problematic.

Numerous definitions include phrases such as “the great empty middle of our continent”, “periodic drought and frost”, “absence of trees”, and “domination by grasses” (Dengler et al., 2014; Dixon et al., 2014; Gibson, 2009 and references therein). Continentality (absence of larger waterbodies) responsible for high temperature amplitudes and precipitation reduction results in environments dominated by hardy and drought-adapted graminoids and lack of woody plants. From this definition it follows that steppe is a biome with the aforementioned features located in a temperate zone. Further subdivisions of temperate grasslands were introduced and nowadays we also know prairies and pampas as grasslands of the temperate zone together with the steppes. Due to their physiognomic and biotic diversity, a clear difference between these temperate grassland biomes is missing. Nevertheless, it is generally accepted that prairies are temperate grasslands of the North America, pampas are temperate grasslands of South America, and steppes represent temperate grasslands of Eurasia (Gibson, 2009; Hurka et al., 2019; Wesche et al., 2016 and references therein).

In addition to these climate-driven temperate grasslands, one should not forget about the azonal and extrazonal temperate grasslands (Dengler et al., 2014; Mucina, 2018). Driven by edaphic and local topographic features, these grassland regions are also an important hub for endemic and threatened species, and a potential genetic source for economically important plants, such as *Camelina sativa* (Ambarlı et al., 2016). Depending on the definition and the inclusion/exclusion of azonal grasslands, current grasslands distribution maps are not always congruent. While Olson et al., 2001, based on which WFF terrestrial ecoregion polygons were developed, includes Alps conifer and mixed forests, Faroe Islands boreal grasslands, and Scandinavian Montane Birch forest and

grasslands, as well as entire Mediterranean and Tibetan steppes (sensu Wesche et al., 2016), I, for example, excluded all of them because their evolution and floristic composition differ greatly from the onset and development of the Eurasian steppe belt. Corresponding differences can be inferred in detail from Appendices 5–8. Surprisingly, WFF did not recognise Bohemian steppe exclaves, Kujawy extrazonal xeric grasslands, Moravian steppe exclaves, Oder Valley, Polish Uplands, Rhine Valley and, most importantly, the Pannonian Basin as steppe habitats included into the Global Grassland shapefile. This is not in agreement with the definition of Eurasian steppe belt that I followed in this project. Thus, I included them into the Appendices 4-8, which reflect my understanding of the Eurasian steppe belt and that of the research group, in which this project has been carried out.

Calibration methods in time divergence estimation analyses

In the recent years, the molecular dating has become an important component of phytogeographic studies, allowing researchers to place evolutionary events onto a timescale. A direct proxy to infer spatial and temporal presence of a study subject in questions are fossils, however, isolated fossils often provide only screenshots of continuous evolutionary processes, and for example extinction followed by recolonisation is particularly difficult to infer solely from the fossil record (Renner, 2005). Even if properly identified, fossil calibration points might suffer from phylogenetic, sampling and absolute age uncertainties (Sauquet, 2012), making such 'node dating' time divergence estimations very problematic (Heath, 2012; Ho and Phillips, 2009; Wang and Mao, 2016; Wilf and Escapa, 2015). Recent advances in fossil age integration now allow for inclusion of all available fossils (fossilised-birth-death approach) and more prior-independent time spans (Heath et al., 2014). While this algorithm has been recently applied to numerous study systems (Grimm et al., 2014; Renner et al., 2016; Šmíd and Tolley, 2019), the family of Brassicaceae is not eligible for this approach due to the lack of extensive reliable fossil record.

Nevertheless, Brassicaceae are not completely absent from the fossil record. There are some records of Brassicaceae-like fossils dated to the upper Miocene from France (Naud and Sue, 1975) and later on from Oligocene as well (Cronquist, 1981). In Europe, fossilised seed-bearing structures of *Barbarea*, *Bunias*, *Rorippa* and *Cardamine* were found and dated to Pliocene (Mai, 1995). Rich fossil record of the latter genus was also inferred from Pleistocene and Holocene layers in Australia and New Zealand (Martin, 1986; Moar and Suggate, 1979), together with *Lepidium* (Mildenhall, 1980) and other Brassicaceae species (Colhoun et al., 1982). Seeds of *Capsella* dated to middle

Pleistocene were found in the interglacial deposits in Ilford, United Kingdom (West et al., 1964, Turner, 2000). Seed, seed-bearing structures and aDNA signals assigned to *Alyssum*, *Arabis*, *Cardamine*, *Cochlearia*, *Descurainia*, *Draba*, *Parrya*, *Smelowskia* have been retrieved from late Pleistocene and Holocene deposits. (Binney 2017; Kienast et al., 2005, Willerslev et al., 2014, Yurtsev 2001). A general trend in the fossil recovery of Brassicaceae can be observed — the younger the deposits, the higher the proportion of Brassicaceae taxa. This might be supported by the young age of the family, however, the more probable explanation are the taphonomic features of fossils.

While these Brassicaceae fossils are relatively densely sampled, this information unfortunately cannot be used as calibration points because of their young age. Furthermore, not only does the use of solely young calibration points result in substitution rates and date estimates with higher error and lower precision (Ho et al., 2014), these fossils are due to highly homoplasious nature of seeds and fruits in Brassicaceae unreliable per se and can lead to false results (Franzke et al., 2011). A very infamous example is the use of incorrectly assigned *Thlaspi* fossil (Becker, 1961), pushing the crown age of Brassicaceae into the early Eocene (~54 MYA; Beilstein et al., 2010). This age is approximately 20 million years off from the currently accepted crown age of this family, which is around 30–35 MYA (Edger et al., 2015; Fawcett et al., 2009; Hohmann et al., 2015; Schranz and Mitchell-Olds, 2006; Walden et al., 2020).

Nevertheless, due to their great economic importance, some alternative solutions to primary fossil calibration have been suggested. One includes broadening of the taxon sample, where plant families with a more reliable fossil record are included. This (essentially primary calibration) approach, however, can only be applied on a whole plastome or multi single-copy nuclear marker datasets, because only these can resolve the internal topology of the investigated clade sufficiently. Another strategy, which includes working on smaller datasets, is the use of secondary calibration points from bigger fossil-calibrated time divergence estimations. However, each reuse of the node might result in an increased error and uncertainty in estimated ages, therefore special attention should be given to the quality of the original analysis and its original calibration scheme (Hipsley and Müller, 2014; Sauquet, 2012). The main problems might lie in the use of only single secondary calibration points and/or secondary calibration points, where the constitution of the clade from which the age was taken from does not fully reflect the structure of the clade the age is applied to. These mistakes have

somewhat justifiably made the secondary calibration approach rather unpopular (Graur and Martin, 2004; Pirie and Doyle, 2012; Reisz and Müller, 2004; Sauquet, 2012).

Despite all these potential problems, I have used the secondary calibration approach in all of the studies published in my doctoral dissertation. To allow for representativeness of my results, I have always carried out multiple runs, using several independently selected group of multiple secondary calibration points to assess the potential changes in the time spans. The recovered time spans were then compared with previously independently published multi-locus based studies, such as Couvreur et al., 2010; Guo X et al., 2017; Hohmann et al., 2015; Huang C-H et al., 2016, Huang X-C et al., 2020; Hurka et al., 2012. In most of the investigated cases, the results were highly congruent, indicating that the use of secondary calibration points was reliable. In cases where this was not so, I uncovered flawed assumptions leading to skewed time spans. For example, in the case of St Onge et al., 2011, the split between a self-compatible (*C. rubella*) and a self-incompatible species (*C. grandiflora*) was assumed to have occurred simultaneously with the disruption of the mating system. Assuming that solely the disruption in the mating system was responsible for the speciation process, the latter could not have occurred simultaneously, but rather later on as a result of the already mentioned mating system change. Another example is from Bachmann et al., 2019, where the authors carried out a time estimation analysis based on a single non-neutral S-locus responsible for the disruption of the mating system. In this particular case, the crown age of *Capsella* clade was pushed further back into the Pleistocene, which was never supported by any other time divergence estimation. This age might, however, correspond to the age of S-locus divergence.

Furthermore, as shown in Publications II and III, investigating the stability of the time spans through different taxon samples (a nested dating approach) might help illustrate the robustness of the results, even if they were obtained via secondary calibration. In these publications, I showed that the crown age of the *Schivereckia* group and Sisymbrieae tribe remain comparable across different taxon samples, tree priors and secondary calibration methods. As a side note, while problematic, published ITS-substitution rate calibration might also retrieve congruent results with other calibration methods. Nevertheless, it is of utmost importance to either combine this secondary calibration option with other more robust calibration method or compare the retrieved time spans with other independent studies.

Ancestral area reconstruction approach

Historical biogeography aims to explain species distribution patterns over evolutionary time scales (Sanmartín, 2012). In the past two decades, numerous methods to reconstruct ancestral geographical distributions using a combination of phylogenetic and distributional information have been established, and new software packages are being developed rapidly (Yu et al., 2015). For a successful application, a model of choice must (or at least should) delimitate between four important evolutionary processes that shape the biological history of a study system: extinction, range expansions and two kinds of allopatric speciation events – vicariance and dispersal (Futuyma, 1998). Despite its popularity in the last decade, the parsimony-based DIVA (Dispersal-Vicariance Analysis: Ronquist, 1997) is for example unable to distinguish between range expansion and across-barrier dispersals, and can accurately reconstruct ancestral areas solely in cases where speciation occurred through vicariance (Kodandaramaiah, 2010). Furthermore, simulations have shown that this algorithm is extremely sensitive to the presence or absence of outgroup taxa (Kodandaramaiah, 2010), making these reconstruction dubious (Clark et al., 2008; Clayton et al., 2009).

Contrarily, a likelihood based DEC (dispersal extinction cladogenesis) model does not favour vicariance over speciation by dispersal and takes into account anagenetic as well as cladogenetic changes in geographic ranges (Ree et al., 2005; Ree and Smith, 2008). This model allows for additional inclusion of the following two parameters that might vary among time slices: connectivity between the ancestral areas and information on rates of dispersal between areas (Kodandaramaiah, 2010). These are of extreme importance, especially in my case as I was investigating the influence of Pleistocene glaciations on the evolutionary history of steppe flora. During these epochs, a complete isolation of contemporary adjacent areas took place, as for example the complete isolation of Pontic Steppe and Kazakh Steppe during several consecutive periods of Caspian Sea transgressions (see Publication I), the onset of highly disjunct distribution in the East European Plain during the Pleistocene interglacial cycles (see Publication II), or a disjunct distribution of Sisymbrieae representatives from Central European and Far Asian temperate forests (see Publication III). Surprisingly, while the adjacency matrix strongly determines the outcome of the DEC-based reconstructions, time slices and dispersal probability categories have only minor effects. These findings corroborate the algorithms behaviour previously assessed by Chacón and Renner, 2014.

While DIVA and DEC perform more reliably when the number of ancestral areas is lower, BayArea was developed to cope with a large number of areas, thus allowing reconstructions of more complex biological histories (Landis et al., 2013). Furthermore, while BayArea can also accept trees with polytomies, it accepts only one tree per analysis, thus excluding phylogenetic uncertainty into biogeographic inference (Yu et al., 2015). As described above, all these algorithms have numerous disadvantages. Thus, as suggested by the developers of these programmes and experts in the field (Clark et al., 2008; Lamm and Redelings, 2009; Landis et al., 2013; Ronquist, 1997; Yu et al., 2015), one should always test several of them with different input parameters to assess the robustness of the results and always carry out additional independent analyses to support the data, e.g. time divergence estimations and ecological niche modelling. Personally, I think the most problematic feature of these algorithms is their assumption that the area is an inherited trait, which might not necessarily be the case (Brooks and Van Veller, 2003; Kodandaramaiah, 2010; Posadas et al., 2006; Renner, pers. comm.). Nevertheless, these countless more or less perfected algorithms produce visually extremely appealing figures and are thus favoured in publications (Ronquist et al., 2011; Yu et al., 2015). In the end, it depends on the complexity of the biogeographic history of a taxon in question. The more complex it is, the lower the chance that the reconstruction will retrieve accurate results.

Regardless, during the course of my dissertation I have carried out ancestral area reconstructions using different reticulate models of higher complexity (likelihood based DEC and Bayesian-based BayArea over parsimony-based DIVA) and compared the results afterwards (see Publications I, II and III). In case of incongruent results, the whole ancestral area reconstruction was discarded and treated as unreliable. These analyses were always accompanied by other independent methods, and in the end served only as an additional independent source of information that supported our interpretation.

Ecological niche modelling approach

Ecological niche modelling (ENM) is an approach that aims to predict the relative habitat suitability of a species by linking the occurrence data of the species in question with the environmental variables (Warren and Seifert, 2011). These algorithms can also be used to predict the changes in the potentially suitable habitats through time, under the assumption of niche stability (Ficetola et al., 2010; Jeschke and Strayer, 2008; Nogués-Bravo, 2009). One has to keep in mind, however, that these reconstructions only infer the suitability of the habitat, whereas no information is given about

the species actual realisation of the ecological niches in that habitat. This depends on a combination of different abiotic conditions, biotic interactions, and accessibility by dispersal (Soberón and Peterson, 2005). The latter is of especially great importance for some species with high invasive potential that are confined to a small region solely due to the lack of dispersal — a phenomenon known as "Wallace's Dream" (Saupe et al., 2012). It has been proven that ENM models do not retrieve satisfactory results in such cases and should thus be treated with caution (Owens et al., 2013). Furthermore, ENM algorithms are extremely sensitive to the numerous input information briefly described in the paragraph below (Hijmans 2012; Radosavljevic and Anderson, 2014; Warren and Seifert, 2011).

Every ecological niche modelling highly depends on the curated data matrix describing the current species distribution. Contemporary distribution should ideally be composed of as many georeferenced and equally distributed records as possible. Bias in sampling density and a lack of coverage influence the modelling results greatly (Phillips et al., 2006, 2017). This can of course pose a problem, if one is working on relatively widely distributed species that morphologically highly resembles its cosmopolitan sister species, as it is the case in Publication IV. I thus relied primarily on well-curated university databank on *Capsella orientalis* records and occurrence points from individually recognised herbarium accessions by taxonomist D. A. German. The occurrence points from public databases such as GBIF were taken into account only as a secondary source and all the analyses were ran in parallel once with GBIF accessions and once without. I also made sure that sample density remained the same across the whole core distribution area and that (to the best of our knowledge) extreme west, east, south and north occurrences were included into the dataset.

The analysis was carried out using Maxent, a machine-learning algorithm that uses the principle of maximum entropy on presence-only data to approximate the species' niche (Phillips et al., 2006, 2017). It has been shown that this algorithm is prone to overfitting, thus appropriate regularisation that forces Maxent to focus on only the most important model features, following the Occam's razor principle, is necessary (Phillips et al., 2006). This also reflects in the number of tested climatic variables. To prevent overfitting (see James et al., 2014), I carried out multiple multicollinearity tests to estimate the variables dependancy. In the end, the final analysis presented in Publication IV was carried out using five different negligibly correlated climatic variables. These included: annual mean temperature (bio1), mean diurnal temperature range (bio2), mean temperature of the wettest quarter (bio8), precipitation of the warmest quarter (bio18), and precipitation of the coldest quarter

(bio19). Intuitively, one would expect that at least some of these variables should always correlate, however, the correlation surprisingly depends on the investigated taxon, as different variables influence the distribution of a taxon in question to different degrees. The five above-mentioned climate variables influence the distribution of *Capsella orientalis* the most (according to Han et al., 2015), and reducing the variables to four and three, respectively, did not significantly alter the results. This is also an indication that even a complex model with five variables does not cause overfitting.

Furthermore, relying on presence-only datasets, multiple studies have shown that numerous models implemented in Maxent should be extensively tested before use (Araújo and Guisan, 2006; Elith et al., 2011; Radosavljevic and Anderson, 2014; Shcheglovitova and Anderson, 2013). Thus, I carried out a comprehensive model evaluation using ENMeval (Muscarella et al., 2014). Due to the dubious nature of these reconstructions, ecological niche modelling should always be tested rigorously before taken into consideration and included only as an accompanying evidence supporting/ rejecting a hypothesis in question (Escobar et al., 2018), as illustrated in Publication IV.

Genetic clustering approach

Population structure studies are inevitable for elucidating demographic histories of taxa in question and can help infer the origin of polyploid populations as well as patterns of admixture and gene flow between different genetic entities (Stift et al., 2019). Thus since the early 2000s, many different genetic clustering methods have been developed, including STRUCTURE (Pritchard et al., 2000), ADMIXTURE (Alexander et al., 2009), INSTRUCT (Gao et al., 2009), FASTSTRUCTURE (Raj et al., 2014) and LEA (Frichot and François, 2015), to mention only a few. Over time, robustness of these methods has been questioned and many disadvantages have been uncovered.

The oldest and probably the most popular algorithm STRUCTURE assumes that all the investigated loci are in the Hardy-Weinberg equilibrium (see Structure manual). This should be tested a-priori, however, even when tested, most of the research community decides to ignore it and still analyse the dataset with this algorithm despite the absence of the Hardy-Weinberg equilibrium. It has never been investigated, however, how far away from the Hardy-Weinberg equilibrium a dataset must be for STRUCTURE to produce anomalous clustering. Developed in the early 2000s, the programmers of STRUCTURE focussed on robustness of the algorithm rather than on speed. STRUCTURE is thus outperformed by FASTSTRUCTURE and ADMIXTURE when it comes to time efficiency.

However, both of these clustering algorithms have been developed for diploid datasets only and despite being extremely time efficient, it was proven that the efficiency of correct clustering assignment in mixed-ploidy datasets plateaued at 80% even in cases with low admixture rates (Stift et al., 2019). On the down side, STRUCTURE assumes that markers are not in linkage disequilibrium within subpopulations and therefore all the closely lying genetic markers should be excluded prior to the analysis (see Structure manual). Another bias causing the artificial deltaK inference might be the uneven sampling across the investigated dataset (Kalinowski, 2010).

An alternative algorithm that does not assume the Hardy-Weinberg equilibrium nor linkage disequilibrium is INSTRUCT (Gao et al., 2009). However, in mixed-ploidy datasets where the polyploids cannot be genotyped with the same precision as diploids (Dufresne et al., 2014), INSTRUCT performed the worst and STRUCTURE outperformed all the above-mentioned clustering methods with the exception of LEA, which was not tested in that simulation study (Stift et al., 2019). It should be noted that LEA was developed primarily for landscape and ecological association studies and that its clustering algorithm for estimation of population genetic structure is based on principal component analysis (PCA) and admixture analysis (Patterson et al., 2006; Pritchard et al., 2000), both unsuitable for our *Capsella* datasets (see below). While unfortunately no comparative studies with other clustering algorithms have been performed yet, I tested LEA and STRUCTURE for our *Capsella* datasets in Publication IV.

My *Capsella* SNP datasets defy most of the assumptions under which all the above-mentioned clustering algorithms operate. The genus *Capsella* consists of five species, diploid self-compatible *Capsella orientalis* and *C. rubella*, tetraploid self-compatible *C. bursa-pastoris*, diploid self-incompatible *C. grandiflora*, and a hybrid self-compatible tetraploid *C. thracica*. Thus, the one-ploidy level policy and the Hardy-Weinberg equilibrium assumption were already violated before the analysis even started. In addition, as the focus of the dataset was on steppe element *Capsella orientalis* and other species were included “only” for the completeness of the taxon sample, the sampling was highly biased towards *C. orientalis*. Furthermore, due to its young age, the species exhibit notable introgressions from each other and relatively low F_{ST} values, masking clear species boundaries. Because of all these features, I wanted to test different genetic clustering algorithms and I focussed on LEA and STRUCTURE.

The clustering analyses did not necessarily contradict each other, but rather follow different paradigms. LEA mirrored a bifurcating representation of the taxon set, a very similar one to the

SVDQ-based tree. All the retrieved genetic clusters were either species-specific or revealed intraspecific genetic pools (for example in the cases of *Capsella orientalis* and *C. bursa-pastoris*). Notably, the same genetic pools were retrieved also from other analyses using different dataset (see details in the Publication IV) or other studies (Cornille et al., 2016; Wesse et al., 2021). This is an indication that the retrieved genetic clusters are not of artificial nature. However, LEA tends to produce extensive admixture zones at lower K-values (K=2), which contradict well-supported phylogenetic tree representations. Assuming that extensive admixture zones indeed exist, they should be seen in bifurcating display as a poorly supported subgrouping (Pfanzelt, pers. comm.). Additionally, while clustering proportions retrieved by LEA might point towards a hybridisation zone, further investigations might reveal more of an isolation-by-distance pattern (Pfanzelt, pers. comm.). This view is also supported by LEA tutorial (François, 2016). The optimal number of K in LEA analyses was always either higher or never reached a plateau regardless of the chain length and repetition number, in comparison to the STRUCTURE deltaK analyses, which should also be taken into consideration.

Despite many violations in our dataset, STRUCTURE retrieved biologically relevant optimal K values in all investigated datasets and the retrieved subsequent clustering reflected the biological history of the genus *Capsella* significantly better than LEA. The three most prominent examples are mentioned here. While the hybrid species *Capsella thracica* was retrieved as a separate cluster from K=4 onwards in LEA, STRUCTURE algorithm showed an admixture pattern comprising approximately 80% of sympatrically occurring maternal genetic cluster and 20% of paternal genetic cluster that cannot be found in the region anymore (hence the low genetic proportion). This pattern was consistent throughout different K-values. While it may seem adverse that STRUCTURE failed to delimit two morphologically distinct species, NGS-based-datasets favour the idea that diploid self-compatible *Capsella rubella* is genetically depauperated diploid self-incompatible *C. grandiflora* (Guo Y-L et al., 2009). In higher K-values these two species can be delimited, as the self-incompatible *Capsella grandiflora* exhibits additional cluster affiliation specific only for this particular species and the gene pool, into which *C. grandiflora* partially introgressed. The gene pool in question is the Mediterranean *Capsella bursa-pastoris*, which can be found only in warmer climates around the world (including Mediterranean where *C. grandiflora* is growing). We postulate that it is indeed this particular introgression that allowed Mediterranean *Capsella bursa-pastoris* to not only occupy new habitats in Mediterranean climate but also prevent back-hybridisation with the Eurasian genetic pool. (Hurka and Neuffer, pers. comm.).

(In)complete taxon sample and multi-locus phylogenies

There has been a long-standing debate about the impact of taxon sampling on phylogenetic inference (Nabhan and Sarkar, 2012). While many studies have reported substantial changes in the topology as well as statistical support among different taxon sets (Heath et al., 2008; Hedtke et al., 2006; Rydin and Källersjö, 2002; Stefanović et al., 2004; Wallberg et al., 2004; Xiang et al., 2012; Zhao L et al., 2013), some simulations suggest that longer sequences, rather than extensive sampling, improve the accuracy of phylogenetic studies (Rosenberg and Kumar, 2001). The latter study is, however, problematic, as their data were simulated under extreme theoretical conditions, separate from living study systems (Pollock et al., 2002). One might intuitively believe that it is generally better (or at least not worse) to increase the amount of data to resolve a question in statistical inference (Pollock et al., 2002). However, an increased taxon sample has proven to decrease the accuracy of the phylogenetic inference in some single-locus analyses (Kim, 1996, 1998). In these cases, a reduced taxon sample probably retrieved only a simplified evolutionary history of the investigated taxon. Thus, additional taxa uncover further biological aspects and retrieve topologies, which better resemble the true biological history of the taxon in question, which may or may not be appropriately depicted with a bifurcating tree.

Overall, I personally prefer multiple-gene-based phylogenies with an as complete taxon sample as possible. Nevertheless, in the end it comes down to the study subject. The statistical support of the backbone phylogeny of dinoflagellates does not differ extremely regardless, if it is based on rDNA operon only (Chacón and Gottschling, 2020) or on whole transcriptome data (Janouškovec et al., 2017). Another prime example that more genetic data (as long as the taxon sample remains comparable) does not improve the backbone statistical support is the family of Amaranthaceae s.l. (Kadereit et al, 2003 and 2017 vs. Morales-Briones et al., 2020). Contrarily, phylogenies retrieved from whole genome, plastome and transcriptome datasets (Gitzendanner et al., 2018; Leebens-Mack et al., 2019; Zhao T et al., in rev.) have immensely improved our understanding of the angiosperm evolution in comparison to single locus phylogenetic inferences of the early 2000s (Soltis et al., 1997, 1999, 2000).

All four publications resulting from the PhD project illustrate the importance of the taxon sample for a robust phylogenetic inference. Publication I not only encompasses all the currently available and validly described species of *Camelina* for the first time but also covers their whole distribution range. This strategy allowed us to minimise the influence of the incomplete taxon sample and

include the whole genetic diversity of this taxon. These two aspects are extremely important, as seen from a parallel study of Brock et al., 2018 in which only some *Camelina* taxa from Turkey, Georgia and Armenia were included into their taxon sample. Due to a different taxon sample as well as sequence data, the topological differences in these two studies should be left separate. Contrarily to Brock et al., 2018, we were able to confirm the close relationship between 'wild' *Camelina sativa* and its 'adapted cultivar' *C. alyssum*, and we uncovered new geographically restricted tetraploid populations of plants that were morphologically determined as *C. microcarpa*. Of course, our study is limited by the low number of investigated loci. Another prominent example is seen in the Publication II, where a more comprehensive taxon sample of *Draba* species (227 species or 387 accessions in Friesen et al., 2020 vs. 169 species or 372 accessions in Jordon-Thaden et al., 2010) led us to a conclusion that the *Schivereckia* subclade is a periglacial relic — a discovery, which could not have been made based solely from the taxon sample of Jordon-Thaden et al., 2010.

Furthermore, in Publication III a comprehensive taxon sample led us to a close relationship between a newly discovered species *Sisymbrium malatyanum* with *S. gaubae* and *S. leucocladum*. This newly uncovered relationship is supported also by the morphological similarities among these species as well as biogeography, and questions previously retrieved sister group relationship with species from different biogeographic regions and contradicting morphologies. Publication IV dealt with the evolutionary history of *Capsella orientalis*. Nevertheless, for the representativeness of the taxon sample, an additional whole genus phylogeny was carried out. While numerous attempts have been made to elucidate the evolutionary history of *Capsella*, only one previous study based on a single gene included all the species (Hurka et al., 2012). Our SNP-based phylogeny uncovered not only the elusive genetic diversity within *C. orientalis*, but also irrefutably placed *C. thracica* into a sister group relationship with the Eurasian *C. bursa-pastoris*, thus rendering *C. bursa-pastoris* paraphyletic. This pattern furthermore supports previously suggested yet disputed ideas that i) *C. bursa-pastoris* possibly evolved independently at least two times (see Discussion of Publication IV; Ceplitis et al., 2005; Cornille et al., 2016; Douglas et al., 2015; Han et al., 2015; Hurka et al., 2012; Kryvokhyzha et al., 2019; Slotte et al., 2006; St Onge et al., 2011) and ii) that one of the parental lineages of hybrid *C. thracica* is the Eurasian *C. bursa-pastoris* lineage.

Gene trees, species trees and associated topological incongruences

Single gene phylogenies are adequate to uncover the evolutionary history of a gene in question, however, organismic biologists tend to investigate the evolutionary history of species, not genes.

While species' trees can be seen as a consensus of trees of numerous gene trees (Maddison, 1997), single genes might have a different history than the species they are found in, because they might underwent duplication, might got loss or might were transferred horizontally, possibly through a vector or a hybridisation event (Maddison, 1997). Therefore, phylogeny of these genes will retrieve a different topology, incongruent to the one of the species' tree. One has to keep in mind that basic evolutionary unit is not an individual but a population (Nakleh et al., 2009). Thus, evolutionary forces are influencing the whole gene pool of the population, not only an individual. When investigating young, rapidly evolving lineages with big effective population sizes, the incomplete lineage sorting will thus furthermore obscure the recovery of a robust species' tree (Pamilo and Nei, 1988). Further incongruences or poorly supported bifurcating phylogenetic relationships might be a result of ancient hybridisation events or introgression (Morales-Briones et al., 2020; Publications III and IV). While incongruences might be observed in genes within the same organelle, the incongruences might also occur when comparing genes (loci) between different organelles due to their different inheritance pattern and evolutionary speed under which they develop (numerous practical examples, only a few mentioned herein: Gutiérrez-Larruscain et al., 2018; Kramina et al., 2016; Pérez-Escobar et al., 2016; Sun et al., 2015; Yu et al., 2013). A generally biparentally fast-evolving nuclear locus might thus retrieve a different topology in comparison to generally maternally inherited slowly evolving plastome or chondriome loci (Smith DR, 2015). Such incongruences are traditionally treated as evidence of "chloroplast capture" (Rieseberg and Soltis, 1991) assuming that the plastome functions as a single locus. This assumption, however, has been questioned recently (Gonçalves et al. 2019; Walker et al. 2019) and further in-depth studies on the inheritance dynamic of plastome genes are needed to unequivocally answer this question.

Incongruences between nuclear and plastid gene trees were also inferred in all four case studies. For example, *Camelina* and *Capsella* datasets could only be sufficiently resolved with rDNA operon loci or SNP datasets, as numerous low-copy nuclear loci (Stockenhuber et al., 2015) and chloroplast loci failed to retrieve species-specific phylogenies in the first place. While this does not necessarily point towards a topological incongruence per se, it does point, however, towards a different evolutionary history of the investigated loci.

Draba s.l. is the largest Brassicaceae genus, consisting of around 400 species (sensu BrassiBase). The most comprehensive phylogenetic reconstructions to this day only investigated nuclear ITS and plastid trnL intron and trnL–F intergenic spacer. The analysis of approximately 42% of all the

species retrieved fairly poorly resolved chloroplast tree that resembled the phylogeny retrieved from the ITS only up to some degree (Jordon-Thaden et al., 2010). However, this incongruence is presumably a result of a limited resolution power of the plastid encoded trnL–F region (see previous paragraph) and does not indicate the potential hybrid origin of individual *Draba* s.l. taxa. As mentioned above, an increase of species taxon sample to 56% in Friesen et al., 2020 did not increase the backbone resolution but improved the internal topology. To increase the resolution, an extensive whole plastome phylogeny should be carried out, accompanied with a multi-gene phylogeny of single copy genes (Stockenhuber et al., 2015). These coding genes are generally fast evolving (due to the introns they contain) and rapidly evolve back to single-copy state after duplication (Blattner, 2016). This feature reduces potential issues arising from complex gene lineage evolution (Nikolov et al., 2019). Moreover, fast-evolving nature and the insensitivity to concerted evolution make these loci an appropriate proxy to infer recent allopolyploidisation events (Sang, 2002). It is highly unlikely that a single low-copy gene will reflect the species tree (Blattner, 2016). Furthermore, while Brassicaceae have a plethora of publicly available transcriptomes based on which primers for low-copy markers easily can be developed (Stockenhuber et al., 2015), many other plant families' genomes are not that well investigated. Thus, time-consuming methodological issues concerning primer design as well as PCR programme development and potential cloning might hinder the use of these low-copy markers. Using a couple of low-copy markers and combining them with easily amplifiable ITS, ETS and standard chloroplast markers (Dong et al., 2015; Shaw et al., 2005, 2014) might be a proper way to go.

For the subproject resulting in Publication III, I have used such a combinatorial approach to elucidate the evolutionary history of the tribus Sisymbrieae. In the course of the study, I analysed eight different markers to establish a robust phylogeny of the tribus: fast evolving biparentally inherited internal- and external transcribed spacer regions (ITS and ETS), fast evolving biparentally inherited low-copy markers Bra13, Bra246, Bra813 and Bra1402 and chloroplast encoded trnQ-rps16 spacer, psbA-trnH spacer, and a fast evolving ycf1b locus. This selection of loci allowed for a multi-layered study that resulted in a well-resolved phylogeny and offered insights into hybridisation events within the tribus that could not have been inferred, if only one set of loci had been investigated. A peculiar topological incongruence regarding the position of *Sisymbrium volgense* could have been inferred. This species morphologically resembles *Sisymbrium polymorphum* s.l. and clusters as its sister species in the chloroplast tree. Contrarily, the loci encoded in the genome places this species into a statistically well-supported clade consisting of

predominantly Mediterranean *S. damascenum*, *Sisymbrium macroloma* and *S. orientale*. The latter species can also be found in the western Irano-Turanian region, adjacent to the natural habitat of *S. polymorphum*. This genetic incongruence coupled with the overlaying distribution and morphological similarity to *S. polymorphum* s.l. might indicate a hybrid origin of *S. volgense*. Nevertheless, extensive whole-genome sequencing study or FISH/GISH analysis would be necessary to irrefutably confirm its hybrid nature.

Another completely unrelated reason for topological incongruences could be the use of improper phylogenetic algorithms when datasets that contain multiple genes are analysed with inference methods designed explicitly for single loci matrices (Kubatko and Degnan, 2007). Multiple-gene phylogenies can be inferred in two ways: i) loci sequences can be concatenated in one alignment, which serves as the basis for the species tree, or ii) every single locus generates a gene tree and the final species tree is inferred as the consensus of all generated gene trees (Gadagkar et al., 2005). As long as all the analysed gene trees exhibit the same topology, the consensus approach is superfluous. However, due to different evolutionary processes, such as incomplete lineage sorting, hybridisation, polyploidisation, introgression and horizontal gene transfer, there is a high chance that at least a certain portion of investigated loci will have different evolutionary histories and their trees will thus exhibit different topologies. Finally, the most common gene tree topology should surely resemble the species tree topology. Nevertheless, this might not always be the accurate assumption (Degnan and Rosenberg, 2006) and in such cases, concatenation might outperform simplified coalescent methods (Gatesy and Springer, 2014). Luckily, these phenomena have been recognised and dealt with, and for the phylogenetic inference of the SNP datasets in case study IV, I have used the multi-species-coalescent model implemented in SVDQuartets (Chifman and Kubatko, 2014; Wascher and Kubatko, 2021). This inference method treats single SNP positions as phylogenetic markers, thus allowing multiple phylogenetic histories and avoiding aforementioned setbacks. Another programme implementing the multi-species-coalescent model is SNAPP (Bryant et al., 2012; Stange et al., 2018), which I used to perform the time divergence estimation analysis of *Capsella* dataset under a strict-clock model.

Preliminary testing of different SNP datasets using concatenation method in case study IV resulted in numerous trees that show different well-supported topologies within *Capsella orientalis* clade depending on the taxon sample and degree of filtering. The coalescent approach, however, retrieved the internal topology in *C. orientalis* as a polytomy. While unsatisfying, the polytomy does make

sense in the light of the evolutionary history of highly admixed *C. orientalis*, whose crown age is dated to less than 0.01 MYA. The question remains, whether the extensive SNP datasets failed to retrieve the 'true' topology, or did the geographically well-defined subgroups indeed evolve at the same time (hard polytomy). Hard polytomies are hard to prove (Dillenberger and Kadereit, 2012) but can be present in nature. One example from a representative of Andean Loasaceae is introduced below (Gottschling, pers. comm.). Two genetically compatible and morphologically different species A and species B exhibit partially overlapping distribution patterns. While the F1 hybrids exhibit an intermediate morphology, the F2 hybrids exhibit not only the intermediate morphology but also occupy the extreme points in the hypothetical morphospace. These extreme morphologies allow them to potentially occupy new non-overlapping ecological niches, which facilitates simultaneous allopatric speciation. Investigating the evolutionary history inferred from such an F2 hybrid taxon set would result in a true hard polytomy pattern.

Similar approach used in Publication III yielded topologies that better reflect the biological history of Sisymbrieae. Coalescent-based species tree algorithm ASTRAL (Mirarab et al., 2014) retrieved slightly different backbone topology of Sisymbrieae in comparison to the concatenation-based maximum likelihood implemented in RAxML (Stamatakis, 2014) and Bayesian Inference implemented in MrBayes (Ronquist et al., 2012) software algorithms, with the internal topology remaining the same. The new backbone pointed towards a rapid diversification in Sisymbrieae, a phenomenon known for the whole Brassicaceae family (Huang X-C et al., 2020; Walden et al., 2020).

Is taxonomic activity (still) important?

Taxonomy is a scientific discipline of naming and classifying organisms (Stuessy, 2009) and is thus tightly connected to systematics and nomenclature (de Queiroz, 2006; Stevens, 2003). While the 'formal' beginning of this scientific discipline dates back to the time of Linnaeus more than 250 years ago, the urge of classification of living things started as early as in ancient Greece (Godfray, 2002). Taxonomy has been an inevitable part not only of the taxonomist's everyday life but also of other scientists and people outside of the academia. Nevertheless, contemporary taxonomy is facing a lack of prestige and funding that both hinder continuation of biodiversity assessment greatly (Godfray, 2002). This should not be treated as a trivial problem, because taxonomy plays an integral role in conservation biology, which is getting more funding from private sources but struggles when it comes to governmental funding (Anyango-van Zwieten et al., 2019; Bakker et al., 2010).

Furthermore, unambiguous scientific names are the prerequisite for proper identification and, therefore, any subsequent application of biological species (Žerdoner Čalasan et al., 2020). This might be seen as a minor setback of insignificant importance, yet the problem becomes more evident when the taxonomic confusion concerns economically or medically important, model or harmful organisms, such as *Alexandrium tamarense* species complex (John et al., 2014), *Arabidopsis* (Koch and German, 2013), *Hirudo medicinalis* species complex (Kutschera, 2012) and *Hordeum* (Blattner, 2018). Overall, taxonomy is one of the most important basic scientific disciplines and as organismic biologists, we should strive towards recovering its lost glory.

While not being the focal point of this dissertation, for Publication III I have carried out a formal synonymisation of the monotypic genus *Ochthodium* DC. with the genus *Sisymbrium* L. and subsequent transfer of the only species *Octhodium aegyptiacum* (L.) DC. into the latter genus under the name *Sisymbrium aegyptiacum* (L.) German, Zerdoner & Al-Shehbaz, comb. nov. This treatment renders the genus *Sisymbrium* as monotypic (again), and because human mind tends to operate according to the Occam's razor principle it allows for an easier understanding of the genus concept. Further taxonomic activity to resolve the *Sisymbrium polymorphum* sensu lato was out of the scope of Publication III and will be a part of a separate study outside of this dissertation. Another opportunity for taxonomic activity arose in Publication II, however, a description of a new *Draba* species (in the Publication II labelled as '*Draba* spec.') was not done for several reasons. First is the fact that the through morphometrics we did not recognise any synapomorphic characters that would point towards a new *Draba* species. Thus newly collected fresh material from Caucasus should be investigated to assess, whether this genetically distinct '*Draba* spec.' can be morphologically told apart from other morphologically similar *Draba* species. In some cases, where the morphological character changes between species are minute, the process of drying of herbarium sheets might destroy them, preventing morphological characterisation of the accession in question. If indeed no morphological differences are found, the researcher is left with what might be one of the more "ungrateful" tasks that have been taunting the scientific communities for centuries — to make a decision on what is a (new) species? While it is possible to describe a new species using molecular diagnosis only without clear morphological synapomorphies (Filipowicz et al., 2012), this activity has not been received well in the community of organismic biologists (Renner, pers. comm.).

Leaving the sometimes already philosophical discussion on what a (new) species actually is when it comes to deciding whether a new species should be introduced into the scientific community based on morphological vs. genetic characteristics, I am mostly on the side of genetic characteristics for the two following reasons. Firstly, while some plant species are specialists and are found in ecologically rather similar environments (thus also exhibiting uniform morphology), a lot of plants species have a plethora of adaptive strategies that reflect in their ability to invade new habitats and thus exhibit immense morphological plasticity (González and Gianoli, 2004; Meerts, 1995; Munier-Jolain et al., 2014; Neuffer et al., 2018). In the latter case, it seems to be rather unwise to base the description of a new species based on morphology and since a formal species description is based on a type specimen, it is hard to incorporate potentially wide morphological variability into it. The second reason is of a more informal nature and it probably has to do with the fact that I started my scientific career in the field of evolutionary phycology, where we were actively working on phylogenetics and taxonomy of unicellular organisms. These often lack clear morphological characteristics based on which we would be able to tell the species apart, thus we rely mostly on the genetic information.

Species descriptions based on genetic information allow circumventing the problematics concerning the ecological plasticity. However, assigning single nucleotide polymorphisms (SNPs) to either intraspecific or interspecific variability is not a trivial task. One has to take into consideration what is the nature of the investigated genetic locus. While slowly evolving chloroplast loci without SNPs are often shared among several, otherwise morphologically well-defined species, SNPs in extremely fast evolving loci (such as internal transcribed spacer: ITS) are often treated as information for intraspecific variation but not for the interspecific one (Žerdoner Čalasan et al., 2019). Another problem with the use of ITS as a species delimitation marker is its multi-copy nature and biparental inheritance (Kiss, 2012). Assuming that during sequencing only paternal ITS copy is reported, a barcoding algorithm would not recognise this particular species in the plethora of mixed ITS signals of the environmental sample, if by chance the maternal ITS copy was sequenced. Furthermore, due to its fast evolving nature, different SNPs in the ITS copies might evolve due to some recent allopatric event(s) that prevented the gene flow from one population into another. Of course, one might argue that this is a potential onset of a new speciation process. Nevertheless, it is hard to objectively establish the thin line when diverging populations transit into two closely related species, even with extensive crossing experiments that would determine the degree of genetic isolation between the two biological objects in question.

An instinctive solution would be to add more and more loci, including fast and slow evolving loci from nucleus, chloroplast and mitochondrion. Yet again, a question arises: How many loci is enough? An exaggerated and completely unfeasible (yet theoretically possible) solution would be to sequence and annotate the whole genome and investigate the proportion and position of SNPs in it. However, even if that would be the case, recent studies have shown that there are genomic regions that are extremely variable even among individuals of well-defined species, such as plant model species of *Arabidopsis* (Burns et al., in rev.; Jiao and Schneeberger, 2020). Clearly, there is no clear answer to this problem and each individual case should be treated uniquely.

Biogeographic patterns inferred from the four case studies

Brassicaceae are generally perceived as one of the younger plant families that started to diversify relatively late – approximately 30 MYA (Walden et al., 2020). Fossil record from Kazakhstan Plains suggests the first evidence of an open grassland landscape from around the same time (Akhmetiev and Zaporozhets, 2014; Velichko, 1999). A hypothesised centre of origin of Brassicaceae is the Irano-Turanian floristic region, which is ecologically and geologically extremely diverse and harbours the highest number of Brassicaceae taxa (Al-Shehbaz et al., 2006; Hedge, 1976; Koch and Kiefer, 2006). However, this floristic region is vast and numerous sources that cite this information did not provide further details regarding their understanding of the Irano-Turanian floristic region (see Manafzadeh et al., 2016 for background information) or infer a more precise locality. Thus, the exact place of origin of Brassicaceae remains, and it will probably remain, unknown. More recent publications place the origin of numerous Brassicaceae taxa to the Anatolian region (Ansell et al., 2011; Karl and Koch, 2013; Mohammadin et al., 2017; Özüdoğru et al., 2015) but nevertheless, one has to keep in mind the problematics concerning the ancestral area reconstruction (see chapter ‘Ancestral area reconstruction approach’) and allow for a possibility that the highest genetic diversity does not irrefutably point towards a centre of origin. While it is reasonable to speculate that an overall increase in aridity and decrease of air temperature promoted the expansion of open grasslands as well as diversification of Brassicaceae (Huang X-C et al., 2020), it remains unclear, however, whether these two events happened simultaneously or are by some means connected. The following paragraphs will introduce the reader to the key findings of the evolutionary history of the four investigated taxon groups, according to their crown group ages.

***Sisymbrium* clade (crown age: ~14 MYA: middle Miocene)**

The third case study dealt with the genus *Sisymbrium*, an almost exclusively Old-World, yellow-flowered herbaceous taxon, whose origin was estimated to be around the Middle Miocene Climate Transition (~14 MYA). Around this time, a major Antarctic ice sheet expansion and global cooling took place, resulting in an extinction wave also known as Middle Miocene disruption (Frigola et al., 2018). During the same period, mountain ranges of Alborz, Zagros, Kopet Dagh and Pamir started to form, creating considerable habitat heterogeneity (Manafzadeh et al., 2016). The cooling (and consequent decrease in precipitation) as well as formation of new habitats possibly promoted the diversification of this dry-adapted taxon that took place in the western Irano-Turanian floristic region. Increased aridification that continued throughout the Pliocene and Pleistocene allowed the grasslands to become the dominant biome of the Central and Middle Asia (sensu Cowan, 2007). During this period, forests and their undergrowth were forced to retreat to more secluded areas (see the introductory section). This split is reflected also in the split of the only forest-associated *Sisymbrium* species — *Sisymbrium yunnanense* and *S. luteum* from eastern China, Korea and Japan, and European *S. strictissimum*.

In parallel, Messinian Salinity Crisis caused the Mediterranean grasslands to expand across the realm, which is also observed by two independent invasion events of *Sisymbrium* into the Mediterranean from the western Irano-Turanian floristic region. While not enough accessions of *Sisymbrium volgense* were investigated to make sound conclusions (hence data not included into any publication), preliminary genetic data suggest that populations east and west of the southern East European Plain were separated during the Middle Pleistocene. This coincides with the Oka glaciations, one of the most extensive glaciations of the East European Plain that reached as south as the city of Voronezh, albeit its southern border has not been irrefutably determined yet (Velichko et al., 2011). After the ice retreated, the populations regain contact and contemporary populations in this region exhibit multiple copies of the ITS, corresponding to the eastern and western populations. The original region, from which this species spread to occupied the southern East European Plain remains unclear, however, the phylogeny and ancestral area reconstruction suggest a Irano-Turanian floristic origin. While *Sisymbrium* did not originate in the Eurasian steppe belt, its evolutionary history is tightly connected to the evolutionary dynamics of this biome. Due to their adjacent position, the Irano-Turanian floristic region served as a hub for *Sisymbrium*, which invaded this

biome through Middle Asia as well as adjacent Mediterranean through Anatolia on multiple occasions throughout the past 14 million years.

***Camelina* clade (crown age: ~2.5 MYA: early Pleistocene)**

The first case study dealt with *Camelina*, an economically important plant genus, whose biological history has been marked by whole genome duplications and interspecific hybridisation events. The genus *Camelina* originated ~3.5 MYA and started to diversify ~1 MY later in the western and central regional sub centre of the Irano-Turanian floristic region (sensu Takhtajan, 1986). This main diversification event coincides with the transition into the Pleistocene, which was demarcated by a significant cooling that initiated the first out of many glacial periods. Changes in the distribution range within this region during the continuous exchange of glacial and interglacial cycles possibly contributed to the divergence of gene pools and eventually led to allopatric speciation (similar pattern observed in *Acantholimon* by Moharrek et al., 2019). This would explain the presence of sympatrically occurring three diploid *Camelina* species from this region.

The MRCA of subsequent western and eastern *Camelina microcarpa-rumelica* lineage spread into the Kazakh steppe and uplands ~1.5 MYA. At that time, the Eurasian steppe belt was already fully developed, allowing *Camelina* to occupy it without any extensive biogeographic restrictions. The split between the western *C. microcarpa-rumelica* lineage and eastern *C. microcarpa* lineage occurred approximately 1 MYA around the northern coast of the Caspian Sea. During the same time period, the Caspian Sea underwent Apsheron and Baku transgressions (Hurka et al., 2019 and references therein) that were probably responsible for the genetic split between the *C. microcarpa-rumelica* lineage and eastern *C. microcarpa* lineage. After the regression of the Caspian Sea, the populations regain contact. Nevertheless, the merging of the genetic pool was blocked, due to an allopolyploidisation event of the western *C. microcarpa*, which occurred during the isolation period (Mandáková et al., 2019).

The biogeographic history of *C. alyssum-sativa* cannot be unequivocally attributed to certain geological events. This might be due to the limited amount of used genetic markers, or due to the influence humans had on their history. It has been suggested that *C. sativa* is a morphotype of *C. alyssum,* which developed in flax fields as a result of directional selection, causing morphological changes in the plant's habitus and fruit morphology to resemble flax (Barrett, 1983). Furthermore, it being a synanthropic species, the evolutionary history of *Camelina* has undoubtedly been

influenced by human migration, as the oldest finding of human interaction with *Camelina* date back into the Neolithic (Hovsepyan and Willcox, 2008).

***Schivereckia* clade (crown age: ~2.5 MYA: early Pleistocene)**

The second case study dealt with *Schivereckia*, a mysterious floristic element (Kozo-Poljanski, 1928), which can be found solely on low hillsides on chalk or calcareous substrates across the East European Plain. Despite sharing the same crown age, *Schivereckia* clade shows a significantly different evolutionary history in comparison to *Camelina*. Not being a floristic element of contemporary steppes, the genus *Draba* (a clade *Schivereckia* is a part of) was an important floristic element of periglacial steppes. These biomes are without a modern counterpart (but see Chytrý et al., 2018) and supposedly supported the existence of megafauna throughout the Middle and Late Pleistocene (Zimov et al., 2012). While the genus *Draba* originated probably in the Irano-Turanian floristic region (Jordon-Thaden et al., 2010), approximately 5 MYA, the *Schivereckia* subgroup is of subarctic origin and began to diverge in the mid-Pliocene. The earliest diversification events in this group are dated to the early Pleistocene, at ~2.5 MYA.

The end of the early Pleistocene is demarcated by the extensive growth of the ice sheets, accompanied by the spread of permafrost and onset of periglacial steppes. These became the dominant continuous zonal vegetation biome throughout the northern hemisphere during the glacial periods of the Middle and Late Pleistocene when the continentality increased to its maximum and the climate became significantly drier and colder. Nevertheless, during interglacial cycles, the temperatures rose, the periglacial hyperzone became patchier and the cold-adapted periglacial vegetation secluded into the refugial areas at high latitudes or mountain ranges. During the middle Pleistocene, the ice shields prevented the refugia-confined species to reconnect, thus preventing gene flow and facilitating speciation via allopatry. This is illustrated by the North American *Draba* species, whose contemporary distribution mirrors the distribution of periglacial steppes during the Middle Pleistocene. Other *Draba* species within *Schivereckia* subgroup experienced a similar fate during late Pleistocene. While some *Draba* species, such as *Draba murrayi* and *Draba cinerea,* have reoccupied the subartic regions after the last interglacial period, *Draba baikalensis* and *Schivereckia doerfleri* seem to have never left their respective refugia. *Schivereckia podolica* survived unfavourable (= interglacial) conditions of the middle and late Pleistocene primarily in the Ural Mountains and the uplands of Ukraine and central (European) Russia. Climate-landscape reconstruction suggests that the species probably persisted across the whole area in the cold

continental climate during Dnieper and Valdai (= Saale and Weichsel) glaciation (Simakova, 2006). Its current distribution thus mirrors perfectly the interglacial cold-steppe refugia of different ages.

***Capsella* clade (crown age: ~0.75 MYA: middle Pleistocene)**

The fourth case study dealt with a small but in terms of evolutionary history rather problematic taxon of *Capsella*. Due to its young age and small number of species, the importance of a representative taxon sample and an NGS approach are the two key elements for retrieval of robust conclusions. *Capsella* originated around the Pliocene and split into and eastern and a western lineage during the Pleistocene. Subsequent splits within the two lineages occurred during the middle to late Pleistocene as well as in postglacial times and gave rise to *Capsella grandiflora* and *C. rubella*, and *C. bursa-pastoris*, *C. thracica* and *C. orientalis*, respectively.

During this time the Eastern European Plain, Western Siberian Plain and adjacent areas were under extensive influence of Don and Oka glaciations that in general promoted the expansion of open habitats. Nevertheless, one should keep in mind that these glaciations promoted expansion of cold periglacial steppes and during the interglacial cycles, warm loving steppes had to make space for nemoral forests and were thus pushed deep into the south of Asian inland. Putative origin of *Capsella* lies in the centre of Asia at that time dominated by semi-desert and steppe-like vegetation (Hurka et al., 2019 and the references therein). These shifts southwards caused a fragmentation of the Eurasian steppe belt, resulting in the development of the eastern and western *Capsella* lineage. Subsequent climate change promoted secondary range extensions of the western and eastern lineage that hybridised and gave rise to a tetraploid *Capsella bursa-pastoris.* During Holocene another hybridisation took place, highly resembling the one from the Pleistocene and giving rise to another tetraploid *Capsella* species — *C. thracica*.

Subpopulation structure of *Capsella orientalis* is of a very young age postdating any extensive glacial periods. We were, however, able to infer high genetic diversity in the populations from the south Middle Asia. A significant human influence is reflected in population structure of this species and despite nowadays being treated as a typical steppe floral element, it is evident that *Capsella orientalis* migrated into the Eurasian steppe belt long after it was already fully developed.

Chapter 3:

General conclusion and outlook

Brassicaceae are one of the most well studied organismic groups in the world due to their diverse ecosystem services they provide. Consequently, there is a plethora of genetic information already available for numerous economically important representatives and model species across the whole family. This facilitates further genetic studies, this project included. This dissertation is build upon four cases studies that deal with four different Brassicaceae taxa of different ages and evolutionary histories. The evolutionary history of all four investigated taxa has been proven to be under strong influence of Pleistocene climate changes, thus supporting our working hypothesis.

However, despite the fact that evolutionary histories of many Brassicaceae have been shaped by the same environmental changes that played an important role in the biological history of the Eurasian steppe belt, this plant family might not be the most optimal proxy to infer past vegetational changes of this biome, especially in its earlier evolutionary stages. The main reason for this is their relatively young age and place of origin, which is probably adjacent to the Eurasian steppe belt. Nevertheless, while not originating in this vastest grassland region worldwide, taxa from this family repeatedly invaded the Eurasian steppe belt (predominantly in the south-north trajectory) and successfully established themselves as one of the most prominent plant families of the region.

This dissertation has been funded by a DFG grant limited to three years. There were also two additional sub-projects that developed along the regularly scheduled activities described in detail in the DFG funding proposal but could not have been finished by the time the present thesis was being written. The two projects are mentioned briefly in Publications III and IV and will build upon these two already published studies. Publication III deals with the phylogeny of *Sisymbrium* and serves as a necessary prerequisite for the project in which the phylogeography of *Sisymbrium polymorphum* will be investigated. The second project will deal with the hybrid origin and contemporary genetic profile of *Capsella thracica*. Preliminary results shown in Publication IV will serve as a phylogenetic backbone for further studies on hybridisation and polyploidisation, and their consequences.

There is a long way ahead of us before we will completely understand the florogenesis of this biome due to its complex evolutionary history, intertwined with the evolutionary history of current or former adjacent biomes, including different types of forests, deserts and tundra. The four case studies elucidate the evolutionary history of four distantly related taxa of different ages, thus providing numerous missing pieces of an evolutionary steppe puzzle and help towards building a more comprehensive knowledge on the onset and florogenesis of the Eurasian steppe belt.

References

Akhmetiev MA and Zaporozhets NI. 2014. Paleogene events in central Eurasia: their role in the flora and vegetation cover evolution, migration of phytochore boundaries, and climate changes. Stratigr Geol Correl 22: 312–335.

Akhmetiev MA. 2007. Paleocene and Eocene floras of Russia and adjacent regions: climatic conditions of their development. Paleontol J 41: 1032–1039.

Akhmetiev MA. 2010. Paleocene and Eocene floristic and climatic change in Russia and northern Kazakhstan. Bull Geosci 85: 17–34.

Alexander DH, Novembre J and Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19: 1655–1664.

Al-Shehbaz IA, Beilstein MA and Kellogg EA. Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Plant Syst Evol 259: 89–120.

Al-Shehbaz IA. 1984. The tribes of Cruciferae (Brassicaceae) in the southeastern United States. J Arnold Arbor 65: 343–373.

Al-Shehbaz IA. 1985. The genera of Brassiceae (Cruciferae; Brassicaceae) in the southeastern United States. J Arnold Arbor 66: 279–351.

Al-Shehbaz IA. 2011. Brassicaceae (Mustard Family). In: eLS. John Wiley & Sons, Ltd: Chichester.

Al-Shehbaz IA. 2012. A generic and tribal synopsis of the Brassicaceae (Cruciferae). Taxon 61: 931–954.

Al-Shehbaz IA. 2015. *Aphragmus pygmaeus* and *Cardamine pseudotrifoliolata* (Brassicaceae), new species from himalayan China. Novon 24: 1–5.

Alsos IG, Sjörgen P, Edwards ME, Landvik JY, Gielly L, Forwick M, Coissac E et al. 2016. Sedimentary ancient DNA from Lake Skartjørna, Svalbard: assessing the resilience of arctic flora to Holocene climate change. Holocene 26: 627–642.

Ambarlı D, Zeydanlı US, Balkız Ö, Aslan S, Karaçetin E, Sözen M, Ilgaz Ç et al. 2016. An overview of biodiversity and conservation status of steppes of the Anatolian Biogeographical Region. Biodivers Conserv 25: 2491–2519.

An C-B, Chen F and Barton L. 2008. Holocene environmental changes in Mongolia: a review. Glob Planet Change 63: 283–289.

An Z, Kutzbach JE, Prell WL and Porter SC. 2001. Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan plateau since Late Miocene times. Nature 411: 62–65.

Ansell, SW, Stenoien HK, Grundmann M, Russell SJ, Koch MA, Schneider H and Vogel JC. 2011. The importance of Anatolian mountains as the cradle of global diversity in *Arabis alpina*, a key arctic-alpine species. Ann Bot 108: 241–252.

Anyango-van Zwieten N, Lamers M and van der Duim R. 2019. Funding for nature conservation: a study of public finance networks at World Wide Fund for nature (WWF). Biodivers Conserv 28: 3749–3766.

Appel O and Al-Shehbaz IA. 2003. Cruciferae. In: Kubitzki K (Ed.), Families and Genera of Vascular Plants, vol. 5, pp. 75–174. Heidelberg, Berlin: Springer-Verlag.

Araújo MB and Guisan A. 2006. Five (or so) challenges for species distribution modelling. J Biogeogr 33: 1677–1688.

Arkhipov SA and Volkova VS. 1994. Geological history of Pleistocene landscapes and climate in west Siberia. Novosibirsk, 105 pp [In Russian].

Astakhov V. 2008. Evidence of Late Pleistocene ice-dammed lakes in West Siberia. Boreas 35: 607–621.

Astakhov V. 2011. Ice Margins of Northern Russia Revisited. In: Ehlers J, Gibbard PL and Hughes PD (Eds.), Quaternary Glaciations - Extent and Chronology: a Closer Look. Elsevier, Amsterdam, pp. 323–336.

Avise JC. 2000. Phylogeography: the history and formation of species. Cambridge, Mass: Harvard University Press, 464 pp.

Azzaya D. and Khaulenbek A. 2008. Environmental and climate change in Mongolia. URL: www.ees.hokudai.ac.jp/ coe21/dc2008/EM/azzaya.pdf.

Bachmann JA, Tedder A, Laenen B, Fracassetti M, Désamoré A, Lafon-Placette C, Steige KA et al. 2019. Genetic basis and timing of a major mating system shift in *Capsella*. New Phytol 224: 505–517.

Bakker VJ, Baum JK, Brodie JF, Salomon AK, Dickson BG, Gibbs HK, Jensen OP et al. 2010. The changing landscape of conservation science funding in the United States. Conserv Lett 3: 435–444.

Barrett SH. 1983. Crop mimicry in weeds. Econ Bot 37: 255–282.

Becker HF. 1961. Oligocene plants from the upper Ruby River Basin, southwestern Montana. Mem Geol Soc Am 82: 1–127.

Beilstein MA, Nagalingum NS, Clements MD, Manchester SR and Mathews S. 2010. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proc Natl Acad Sci USA 107: 18724–18728

Bengtsson J, Bullock JM, Egoh B, Everson C, Everson T, O'Connor T, O'Farrell PJ et al. 2019. Grasslands—more important for ecosystem services than you might think. Ecosphere 10: e02582.

Bennett KD, Tzedakis PC and Willis KJ. 1991. Quaternary refugia of north Europe. J Biogeogr 18: 103–115.

Berke MA. 2018. Reconstructing Terrestrial Paleoenvironments Using Sedimentary Organic Biomarkers. In: Croft DA, Simpson SW and Su DF. (Eds.), Methods in Paleoecology: Reconstructing Cenozoic Terrestrial Environments & Ecological Communities. Vertebrate Paleobiology & Paleoanthropology. Springer, pp. 121–149.

Beverly EJ, Lukens WE and Stinchcomb GE. 2018. Paleopedology as a Tool for Reconstructing Paleoenvironments and Paleoecology. In: Croft DA, Simpson SW and Su DF. (Eds.), Methods in Paleoecology: Reconstructing Cenozoic Terrestrial Environments & Ecological Communities. Vertebrate Paleobiology & Paleoanthropology. Springer, pp. 151–183.

Binney HA, Edwards M, Macias-Fauria M, Lozhkin A, Anderson P, Kaplan JO, Andreev A et al. 2017. Vegetation of Eurasia from the last glacial maximum to present: key biogeographic patterns. Quat Sci Rev 157: 80–97.

Binney HA, Willis KJ, Edwards ME, Bhagwat SA, Anderson PM, Andreev AA, Blaauw M et al. 2009. The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. Quat Sci Rev 28: 2445–2464.

Birks HH and Birks HJB. 2000. Future uses of pollen analysis must include plant macrofossils. J Biogeogr 27: 31–35.

Birks HJB and HH Birks 2016. How have studies of ancient DNA from sediments contributed to the reconstruction of Quaternary floras? New Phytol 209: 499–506.

Blattner FR. 2016. TOPO6: a nuclear single-copy gene for plant phylogenetic inference. Plant Syst Evol 302: 239–244.

Blattner FR. 2018. Taxonomy of the Genus *Hordeum* and Barley (*Hordeum vulgare*). In: Stein N and Muehlbauer K (Eds.), The Barley Genome. Compendium of Plant Genomes. Springer, Cham. pp. 11–23.

Bone M, Johnson D, Panayoti K, Kintgen M and Vickerman LG. 2015. Steppes: The Plants and Ecology of the World's Semi-arid Regions. Timber Press, Portland, Oregon, 360 pp.

Braun-Blanquet, J. 1964. Pflanzensoziologie, Grundzüge der Vegetationskunde, Springer-Verlag, Berlin, 631 pp.

Bredenkamp GJ, Spada F and Kazmierzak E. 2002. On the origin of northern and southern hemisphere grasslands. Plant Ecol: 163: 209–229.

Brock JR, Dönmez AA, Beilstein MA and Olsen KM. 2018. Phylogenetics of *Camelina* Crantz. (Brassicaceae) and insights on the origin of gold-of-pleasure (*Camelina sativa*). Mol Phylogenet Evol 127: 834–842.

Brock JR, Mandáková T, Lysak MA and Al-Shehbaz IA. 2019. *Camelina neglecta* (Brassicaceae, Camelineae), a new diploid species from Europe. PhytoKeys 115: 51–57.

Brooks DR and van Veller M. 2003. Critique of parsimony analysis of endemicity as a method of historical biogeography. J Biogeogr 30: 819–825.

Brooks TM, Bakarr MI, Boucher T, Da Fonseca GAB, Hilton-Taylor C, Hoekstra JM, Moritz T et al. 2004. Coverage provided by the global protected-area system: is it enough? BioScience 54: 1081–1091.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Syst Biol 29: 1917–1932.

Bumenthal SA, Cerling TE, Smiley TM, Badgley CE and Plummer TW. 2019. Isotopic records of climate seasonality in equid teeth. Geochim Cosmochim Acta 260: 329–348.

Burns R, Mandáková T, Jagoda J, Soto-Jimenez LM, Liu C, Lysak MA, Novikova PYu et al. in rev. Gradual evolution of allopolyploidy in *Arabidopsis suecica*. bioRxiv preprint

Cardinal-McTeague WM, Sytsma KJ and Hall JC. 2016. Biogeography and diversification of Brassicales: a 103 million year tale. Mol Phylogenet Evol 99: 204–224.

Cartea W, De Haro-Bailón A, Padilla G, Obregón-Cano S, del Rio-Celestino M and Ordás A. 2019. Seed oil quality of *Brassica napus* and *Brassica rapa* germplasm from northwestern Spain. Foods 8: 292.

Ceplitis A, Su Y and Lascoux M. 2005. Bayesian inference of evolutionary history from chloroplast microsatellites in the cosmopolitan weed *Capsella bursa-pastoris* (Brassicaceae). Mol Ecol 14: 4221–4233.

Chacón J and Gottschling M. 2020. Dawn of the dinophytes: a first attempt to date origin and diversification of harmful algae. Harmful Algae 97: 101871

Chacón J and Renner SS. 2014. Assessing model sensitivity in ancestral area reconstruction using LAGRANGE: a case study using the Colchicaceae family. J Biogeogr 41: 1414–1427.

Chen H, Al-Shehbaz IA, Yue J and Sun H. 2018. New insights into the taxonomy of tribe Euclidieae (Brassicaceae), evidence from nrITS sequence data. PhytoKeys 100: 125–139.

Chifman J and Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30: 3317–3324.

Chytrý M, Horsák M, Danihelka J, Ermakov N, German DA, Hájek M, Hájková P et al. 2018. A modern analogue of the Pleistocene steppe-tundra ecosystem in southern Siberia. Boreas 48: 36–56.

Clark JR, Ree RH, Alfaro ME, King MG, Wagner WL and Roalson EH. 2008. A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. Syst Biol 57: 639–707.

Clayton JW, Soltis PS and Soltis DE. 2009. Recent long-distance dispersal overshadows ancient biogeographical patterns in a pantropical angiosperm family (Simaroubaceae, Sapindales). Syst Biol 58: 395–410.

Cohen KM, Finney SM, Gibbard PL and Fan J-X. 2013. The ICS international chronostratigraphic chart. Episodes 36: 199–204.

Colhoun EA, van der Geer G and Mook WG. 1982. Stratigraphy, pollen analysis and palaeoclimatic interpretation of Pulbeena Swamp, Northwestern Tasmania. Quat Res 18: 108–126.

Cornille A, Salcedo A, Kryvokhyzha D, Glémin S, Holm K, Wright SI and Lascoux M. 2016. Genomic signature of successful colonization of Eurasia by the allopolyploid shepherd's purse (*Capsella bursa-pastoris*). Mol Ecol 25: 616–629.

Couvreur TLP, Franzke A, Al-Shehbaz IA, Bakker FT, Koch MA and Mummenhoff K. 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). Mol Biol Evol 27: 55–71.

Cowan PJ. 2007. Geographic usage of the terms Middle Asia and Central Asia. J Arid Environ 69: 359–363.

Cronquist A. 1981. An Integrated System of Classification of Flowering Plants. Columbia University Press, New York, 1262 pp.

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS et al. 2006. Widespread genome duplications throughout the history of flowering plants. Genome Res 16: 738–749.

Darwin C. 1859. On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life. London: J. Murray, 440 pp.

Dass P, Houlton BZ, Wang Y and Warlind D. 2018. Grasslands may be more reliable carbon sinks than forests in California. Environ Res Lett 13: 074027

De Bodt S, Maere S and Van de Peer Y. 2005. Genome duplication and the origin of angiosperms. Trends Ecol Evol 20: 591–597.

de Queiroz K. 2006. The PhyloCode and the distinction between taxonomy and nomenclature. Syst Biol 55: 160–162.

Degnan JH and Rosenberg NA. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet 2: e68.

Dembicz I, Szczeparska L, Moysiyenko II and Wódkiewicz M. 2018. High genetic diversity in fragmented *Iris pumila* L. populations in Ukrainian steppe enclaves. Basic Appl Ecol 28: 37–47.

Deng T-H-B, Tang Y-T, van der Ent A, Sterckman T, Echevarria G, Morel J-L and Qui R-L. 2016. Nickel translocation via the phloem in the hyperaccumulator *Noccaea caerulescens* (Brassicaceae). Plant Soil 404: 35–45.

Dengler J, Janisova M, Török P and Wellstein C. 2014. Biodiversity of Palaearctic grasslands: a synthesis. Agric Ecosyst Environ 182: 1–14.

Denk T, Zohner CM, Grimm GW and Renner SS. 2018. Plant fossils reveal major biomes occupied by the late Miocene Old-World Pikermian fauna. Nat Ecol Evol 2: 1864–1870.

Dillenberger MS and Kadereit JW. 2012. Simultaneous speciation in the European high mountain flowering plant genus *Facchinia* (*Minuartia* s.l., Caryophyllaceae) revealed by genotyping-by-sequencing. Mol Phylogenet Evol 112: 23–35.

Diwakar BT, Dutta PK, Lokesh BR and Naidu KA. 2010. Physicochemical properties of garden cress (*Lepidium sativum* L.) seed oil. J Am Oil Chem Soc 87: 539–548.

Dixon AP, Faber-Langendoen D, Josse C, Morrison J and Loucks CJ. 2014. Distribution mapping of world grassland types. J Biogeogr 41: 2003–2019.

Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T et al. 2015. *ycf*1, the most promising plastid DNA barcode of land plants. Sci Rep 5: 8348.

Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, Josephs EB, Arunkumar R et al. 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. Proc Natl Acad Sci USA 112: 2806–2811.

Dufresne F, Stift M, Vergilino R and Mable BK. 2014. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Mol Ecol 23: 40–69

Eckmeier E, Gerlach R, Gehrt E and Schmidt MWI. 2007. Pedogenesis of chernozems in Central Europe – a review. Geoderma 139: 288–299.

Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. Proc Natl Acad Sci USA 112: 8362–8366

Edwards ME, Alsos IG, Yoccoz N, Coissac E, Goslar T, Gielly L, Haile J et al. 2018. Metabarcoding of modern soil DNA gives a highly local vegetation signal in Svalbard tundra. Holocene 28: 2006–2016.

Ehlers J, Astakhov V, Gibbard PL, Mangerud J and Svendsen JI. 2013. Middle Pleistocene Glaciations in Eurasia. In: Elias SA (Ed.), Encyclopedia of Quaternary Science. Elsevier, Amsterdam, pp. 172–179.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE and Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. Divers Distrib 17: 43–57.

Ellenberg H and Mueller-Dombois D. 1967. A key to Raunkiaer plant life forms with revised subdivisions. Berichte des Geobotanischen Instituts der Eidg Techn Hochschule Stiftung Rübel 37: 56–73.

Epp LS, Kruse S, Kath NJ, Stoof-Leichsenring KR, Tiedemann R, Pestryakova LA and Herzschuh U. 2018. Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling. Sci Rep 8: 17436.

Erkens R. 2013. On why we should teach biogeography and the need for a biogeography compendium. Front Biogeogr 5: 84–87.

Ermakov N, Chytrý M and Valachovič M. 2006. Vegetation of the rock outcrops and screes in the forest-steppe and steppe belts of the Altai and Western Sayan Mts., southern Siberia. Phytocoenologia 36: 509–545.

Escobar LE, Qiao H, Cabello J and Townsend Peterson A. 2018. Ecological niche modeling re-examined: A case study with the Darwin's fox. Ecol Evol 8: 4757–4770.

Fawcett JA, Maere S, and van de Peer Y. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc Natl Acad Sci USA 106: 5737–5742.

Feodorova TA, Voznensenskaya EV, Edwards GE and Roalson EH. 2010. Biogeographic patterns of diversification and the origins of C4 in *Cleome* (Cleomaceae). Systematic Botany 35: 811–826.

Ficetola GF, Maiorano L, Falcucci A, Dendoncker N, Boitani L, Padoa-Schioppa E, Miaud C et al. 2010. Knowing the past to predict the future: land-use change and the distribution of invasive Bullfrogs. Glob Chang Biol 16: 528–537.

Filipowicz N, Nee MH and Renner SS. Description and molecular diagnosis of a new species of *Brunfelsia* (Solanaceae) from the Bolivian and Argentinean Andes. PhytoKeys 10: 83–94.

Forte AM and Cowgill E. 2013. Late Cenozoic base-level variations of the Caspian Sea: a review of its history and proposed driving mechanisms. Palaeogeogr Palaeoclimatol Palaeoecol 386: 392–407.

Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H and Wright SI. 2009. Recent speciation associated with the evolution of selfing in *Capsella*. Proc Natl Acad Sci USA 106: 5241–5245.

Foxe JP, Stift M, Tedder A, Haudry A, Wright SI and Mable BK. 2010. Reconstructing origins of loss of self-incompatibility and selfing in North American *Arabidopsis lyrata*: a population genetic context. Evolution 64: 3495–3510.

François O. 2016. Running Structure-like population genetic analyses with R. R Tutorials in population genetics, University of Grenoble-Alpes, 1–9.

Franzke A, Koch MA, Mummenhoff K. 2012. Turnip time travels: age estimates in Brassicaceae. Trends Plant Sci 21: 554–561.

Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA and Mummenhoff K. 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. Trends Plant Sci 16: 108–116.

Frenzel B. 1968. The Pleistocene vegetation of northern Eurasia. Science 161: 637–649.

Frichot E and François O. 2015. LEA: An R package for landscape and ecological association studies. Methods in Ecol Evol 6: 925–929.

Friesen N, Žerdoner Čalasan A, Neuffer B, German DA, Markov M and Hurka H. 2020. Evolutionary history of the Eurasian steppe plant *Schivereckia podolica* (Brassicaceae) and its close relatives. Flora 268: 151602.

Frigola A, Prange M and Schulz M. 2018. Boundary conditions for the Middle Miocene Climate Transition (MMCT v1.0). Geosci Model Dev 11: 1607–1626.

Futuyma DJ. 1998. Evolutionary biology. Sinauer Associates, Sunderland, MA. 600 pp.

Gadagkar SR, Rosenberg MS and Kumar S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. Mol Dev Evol 304: 64–74.

Gajewski MW. 1934. Avenetum desertorum, a monographical study. Bull Int Acad Polon Sci, Cl Sci Math, Ser B, Sci Nat 1: 1–27.

Gandolfo MA, Nixon K and Crepet W. 1998. A new fossil flower from the Turonian of New Jersey: *Dressiantha bicarpellata* gen. et sp. nov. (Capparales). Am J Bot 85: 964–974.

Gao H, Williamson S and Bustamante CD. 2007. A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176: 1635–1651

Gatesy J and Springer MS. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol Phylogenet Evol 80: 231–266.

German DA and Al-Shehbaz IA. 2018. A reconsideration of *Pseudofortuynia* and *Tchihatchewia* as synonyms of *Sisymbrium* and *Hesperis*, respectively (Brassicaceae). Phytotaxa 334: 95–98.

Gibbard P and Head MJ. 2009. The definition of the Quaternary system/era and the Pleistocene series/epoch. Quaternaire 20: 125–133.

Gibson DJ. 2009. Grasses and Grassland Ecology. Oxford University Press Inc., New York, United States of America, 314 pp.

Gitzendanner MA, Soltis PS, Wong GK-S, Ruhfel BR and Soltis DE. 2018. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. Am J Bot 105: 291–301.

Gobejishvili R, Lomidze N and Tielidze L. 2011. Late Pleistocene (Würmian) Glaciations of the Caucasus. In: Ehlers J, Gibbard PL and Hughes PD (Eds.), Quaternary Glaciations - Extent and Chronology: a Closer Look. Elsevier, Amsterdam, pp. 141–147.

Godfray HCJ. 2002. Challenges for taxonomy. Nature 417: 17–19.

Gogol NV. 1835. Taras Bulba. New York, Modern Library, 311 pp.

Gonçalves DJP, Simpson BB, Ortiz EM, Shimizu GH and Jansen RK. 2019. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. Mol Phylogenet Evol 138: 219–232.

González AV and Gianoli E. 2004. Morphological plasticity in response to shading in three *Convolvulus* species of different ecological breadth. Acta Oecol 26: 185–190.

Graur D and Martin W. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet 20: 80–86.

Grimm GW and Denk T. 2012. Reliability and resolution of the coexistence approach — a revalidation using modern-day data. Rev Palaeobot Palynol 172: 33–47.

Grimm GW and Potts AJ. 2016. Fallacies and fantasies: the theoretical underpinnings of the coexistence approach for palaeoclimate reconstruction. Clim Past 12: 611–622.

Grimm GW, Kapli P, Bomfleur B McLoughlin and Renner SS. 2014. Using more than the oldest fossils: dating Osmundaceae with three Bayesian clock approaches. Syst Biol 64: 396–405.

Gugger PF, González-Rodríguez A, Rodríguez-Correa H, Sugita S and Cavender-Bares J. 2010. Southward Pleistocene migration of Douglas-fir into Mexico: phylogeography, ecological niche modeling, and conservation of 'rear edge' populations. New Phytol 189: 1185–1199.

Guo X, Liu J, Hao G, Zhang G, Mao K, Wang X, Zhang D et al. 2017. Plastome phylogeny and early diversification of Brassicaceae. BMC Genomics 18: 176.

Guo Y-L, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, Weigel D and Schierup MH. 2009. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. Proc Natl Acad Sci USA 106: 5246–5251.

Guthrie RD. 2001. Origin and causes of the mammoth steppe: a story of cloud cover, woolly mammal tooth pits, buckles, and inside-out Beringia. Quat Sci Rev 20: 549–574.

Gutiérrez-Larruscain D, Santos-Vicente M, Anderberg AA, Rico E and Martínez-Ortega MM. 2018. Phylogeny of the *Inula* group (Asteraceae: Inuleae): Evidence from nuclear and plastid genomes and a recircumscription of *Pentanema*. Taxon 67: 149–164.

Habel JC, Dengler J, Janišová M, Török P, Wellstein C and Wiezik M. 2013. European grassland ecosystems: threatened hotspots of biodiversity. Biodivers Conserv 22: 2131–2138.

Hais M, Komprdová K, Ermakov N and Chytrý M. 2015. Modelling the Last Glacial Maximum environments for a refugium of Pleistocene biota in the Russian Altai Mountains, Siberia. Palaeogeogr Palaeoclimatol Palaeoecol 438: 135–145

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

Harzhauser M, Daxner-Höck G, López-Guerrero P, Maridet O, Oliver A, Piller WE, Richoz S et al. 2016. Stepwise onset of the Icehouse world and its impact on Oligo-Miocene Central Asian mammals. Sci Rep 6: 36169.

Hassan SN, Reba MN, Hussain D and Ali A. 2019. Elevation dependent thickness and ice-volume estimation using satellite derived DEM for mountainous glaciers of Karakorum range. IOP Conf Ser: Earth Environ Sci 169: 012115.

Hay A and Tsiantis M. 2016. *Cardamine hirsuta*: a comparative view. Curr Opin Genet Dev 39: 1–7.

Heath TA, Hedtke SM and Hillis DM. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J Syst Evol 46: 239–257.

Heath TA, Huelsenbeck JP and Stadler T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. Proc Natl Acad Sci USA 111: 2957–2966.

Heath TA. 2012. A hierarchical Bayesian model for calibrating estimates of species divergence times. Syst Biol 61: 793–809.

Hedge IC 1976. A Systematic and Geographical Survey of the Old World Cruciferae. In: Vaughan JG, Macleod AJ and Jones BMG (Eds.), The Biology and Chemistry of the Cruciferae. London: Academic Press, 45 pp.

Hedtke SM, Townsend TM and Hillis DM. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst Biol 55: 522–529.

Heenan PB. 2017. A taxonomic revision of *Cardamine* L. (Brassicaceae) in New Zealand. Phytotaxa 330: 1–154.

German DA, Tekin M, Španiel S, Marhold and Al-Shehbaz IA. 206. A brief taxonomic revision of *Physoptychis* (Alysseae, Brassicaceae). Phytotaxa 258: 75–82.

Higgins P. 2018. Isotope Ecology from Biominerals. In: Croft DA, Simpson SW and Su DF. (Eds.), Methods in Paleoecology: Reconstructing Cenozoic Terrestrial Environments & Ecological Communities. Vertebrate Paleobiology & Paleoanthropology. Springer, pp. 99–120.

Hijmans RJ. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93: 679–688.

Hipsley CA and Müller J. 2014. Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. Front Genet 5: 138.

Ho SYW and Duchêne S. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. Mol Ecol 23: 5947–5965.

Ho SYW and Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst Biol 58: 367–380.

Hohmann N, Wolf EM, Lysak MA and Koch MA. 2015. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. Plant Cell 27: 2770–2784.

Hooker JD. 1947. Flora Antarctica, Part II Flora of Fuegia, the Falklands, Kerguellen's land, etc Reeve Brothers, London, 235 pp.

Horsák M, Chytrý M, Hájková P, Hájek P, Danihelka J, Horsáková V, Ermakov N et al. 2015. European glacial relict snails and plants: environmental context of their modern refugial occurrence in southern Siberia. Boreas 44: 638–657.

Hovsepyan R and Willcox G. 2008. The earliest finds of cultivated plants in Armenia: evidence from charred remains and crop processing residues in pisé from the Neolithic settlements of Aratashen and Aknashen. Veg Hist Archaeobot 17: 63–71.

Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q et al. 2016. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. Mol Biol Evol 33: 394–412.

Huang X-C, German DA and Koch MA. 2020. Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. Ann Bot 125: 29–47.

Hurka H, Friesen N, Bernhardt K-G, Neuffer B, Smirnov S, Shmakov A and Blattner F. 2019. The Eurasian steppe belt: status quo, origin and evolutionary history. Turczaninowia 22: 5–71.

Hurka H, Friesen N, German DA, Franzke A and Neuffer B. 2012. 'Missing link' species *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus *Capsella* (Brassicaceae). Mol Ecol 21: 1223–1238.

Iglesias A, Wilf P, Johnson KR, Zamuner AB, Rubén Cúneo N, Matheos SD and Singer BS. 2007. A Paleocene lowland macroflora from Patagonia reveals significantly greater richness than North American analogs. Geology 35: 947–950.

Jacobs BF, Kingston JD and Jacobs LL. 1999. The origin of grass-dominated ecosystems. Ann Missouri Bot 86: 590–643

James G, Witten D, Hastie T and Tibshirani R. 2014. An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Incorporated, 426 pp.

Janouškovec J, Gavelis GS, Burki F, Dinh D, Bachvaroff TR, Gornik SG, Bright KJ et al. 2017. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. Proc Natl Acad Sci USA 114: E171–E180.

Jeschke JM and Strayer DL. 2008. Usefulness of bioclimatic models for studying climate change and invasive species. Ann N Y Acad Sci 1134: 1–24.

Jiao W-B and Schneeberger K. 2020. Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. Nat Commun 11: 989.

Jiang H and Ding Z. 2009. Spatial and temporal characteristics of Neogene palynoflora in China and its implication for the spread of steppe vegetation. J Arid Environ 73: 765–772.

John U, Litaker RW, Montresor M, Murray S, Brosnahan ML and Anderson DM. 2014. Formal revision of the *Alexandrium tamarense* species complex (Dinophyceae) taxonomy: the introduction of five species with emphasis on molecular-based (rDNA) classification. Protist 165: 779–804.

Jordon-Thaden I, Hase I, Al-Shehbaz I and Koch MA. 2010. Molecular phylogeny and systematics of the genus *Draba* (Brassicaceae) and identification of its most closely related genera. Mol Phylogenet Evol 55: 524–540

Yu Y, Harris AJ, Blair C and He X. 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol Phylogenet Evol 87: 46–49.

Jørgensen T, Haile J, Möller P, Andreev A, Boessenkool S, Rasmussen M, Kienast F et al. 2012. A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. Mol Ecol 21: 1989–2003.

Kadereit G, Borsch T, Weising K and Freitag H. 2003. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C4 photosynthesis. Int J Plant Sci 164: 959–986.

Kadereit G, Newton RJ and Vandelook F. 2017. Evolutionary ecology of fast seed germination – a case study in Amaranthaceae/Chenopodiaceae. Perspect Plant Ecol Evol Syst 29: 1–11.

Kagale S, Robinson SJ, Nixon J, Xiao R, Huebert T, Condie J, Kessler D et al. 2016. Polyploid evolution of the Brassicaceae during the Cenozoic Era. Plant Cell 26: 2777–2791.

Kalinowski ST. 2010. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity 106: 625–632.

Karl R and Koch MA. 2013. A world-wide perspective on crucifer speciation and evolution: Phylogenetics, biogeography and trait evolution in tribe Arabideae. Ann Bot 112: 983–1001.

Kay KM, Whittall JB and Hodges SA. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. BMC Evol Biol 6: 36.

Kienast F, Schirrmeister L, Siegert C and Tarasov P. 2005. Palaeobotanical evidence for warm summers in the East Siberian Arctic during the last cold stage. Quat Sci Rev 63: 283–300.

Kier G, Mutke J, Dinerstein E, Ricketts TH, Kuper W, Kreft H and Barthlott W. 2005. Global patterns of plant diversity and floristic knowledge. J Biogeogr 32: 1107–1116.

Kim J. 1996. General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. Syst Biol 45: 363–374.

Kim J. 1998. Large-scale phylogenies and measuring the performance of phylogenetic estimators. Syst Biol 47: 43–60.

Kiss L. 2012. Limits of nuclear ribosomal DNA internal transcribed spacer (ITS) sequences as species barcodes for Fungi. Proc Natl Acad Sci USA 109: E1811.

Koch MA and German DA. 2013. Taxonomy and systematics are key to biological information: *Arabidopsis*, *Eutrema* (*Thellungiella*), *Noccaea* and *Schrenkiella* (Brassicaceae) as examples. Front Plant Sci 4: 267.

Koch MA, German DA, Kiefer M and Franzke A. 2018. Database taxonomics as key to modern plant biology. Trends Plant Sci 23: 4–6.

Koch MA and Kiefer C. 2006. Molecules and migration: biogeographical studies in cruciferous plants. Plant Syst Evol 259: 121–142.

Kodandaramaiah U. 2010. Use of dispersal–vicariance analysis in biogeography – a critique. J Biogeogr 37: 3–11.

Kozo-Poljanski BM. 1928. Glaziale Pflanzenrelikte auf dem Orel-Kurskischen Plateau im Süden der Mittelrussischen Hochebene I. Vegetationsbilder. G. Fischer, Jena. Reihe 19, Heft 1/2. Tafel 1–12.

Kozo-Poljanski BM. 1929. Glaziale Pflanzenrelikte auf dem Orel-Kurskischen Plateau im Süden der Mittelrussischen Hochebene II. Vegetationsbilder. G. Fischer, Jena. Reihe 19, Heft 7/8. Tafel 37–48.

Kramina TE, Degtjareva GV, Samigullin TH, Valiejo-Roman CM, Kirkbride JH Jr, Volis S, Deng T et al. 2016. Phylogeny of *Lotus* (Leguminosae: Loteae): partial incongruence between nrITS, nrETS and plastid markers and biogeographic implications. Taxon 65: 997–1018.

Kryvokhyzha D, Salcedo A, Eriksson MC, Duan T, Tawari N, Chen J, Guerrina M et al. 2019. Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid *Capsella bursa-pastoris* (Brassicaceae). PLoS Genet 15: e1007949.

Kubatko L and Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst Biol 56: 17–24.

Kutschera U. 2012. The *Hirudo medicinalis* species complex. Naturwissenschaften 99: 433–434.

Kuzemko AA, Steinbauer MJ, Becker T, Didukh YP, Dolnik C, Jeschke M, Naqinezdah A et al. 2016. Patterns and drivers of phytodiversity of steppe grasslands of Central Podolia (Ukraine). Biodivers Conserv 25: 2233–2250.

Lamm KS and Redelings BD. 2009. Reconstructing ancestral ranges in historical biogeography: properties and prospects. J Syst Evol 47: 369–382.

Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC and Soltis PS. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. Am J Bot 105: 348–363.

Landis MJ, Matzke NJ and Moore BR. 2013. Bayesian analysis of biogeography when the number of areas is large. Syst Biol 62: 789–804.

Lavrenko EM. 1970a. Division of the Black Sea – Kazakhstan subregion of the Eurasian steppe region. Bot Zhurn (Moscow & Leningrad) 55: 609–625 [In Russian].

Lavrenko EM. 1970b. Division of the Central Asian subregion of the steppe region of Eurasia. Bot Zhurn (Moscow & Leningrad) 55: 1734–1741 [In Russian].

Lavrenko EM and Karamysheva ZV. 1993. Steppes of the Former Soviet Union and Mongolia. In: Coupland RT (Ed.), Natural Grasslands. Eastern Hemisphere and Resume. Elsevier, Amsterdam, pp. 3–59.

Lavrenko EM. 1930. Forest relicts (Tertiary) between the Carpathians and Altai. Zhurn Russk Bot Obshch 15: 351–363 [In Russian].

Lavrenko EM. 1969. Über die Lage des euroasiatischen Steppengebiets in dem System der pflanzengeographischen Gliederung des aussertropischen Eurasiens. Vegetatio 19: 11–20.

Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574: 679–685.

Lehmkuhl F, Klinge M and Stauch G. 2011. The Extent and Timing of Late Pleistocene Glaciations in the Altai and Neighbouring Mountain Systems. In: Ehlers J, Gibbard PL and Hughes PD (Eds.), Quaternary Glaciations - Extent and Chronology: a Closer Look. Elsevier, Amsterdam, pp. 967–979.

Lemaire G. 2007. Research priorities for grassland science: the need of long term integrated experiments networks. R Bras Zootec 36: 93–100.

Licht A, van Cappelle M, Abels HA, Ladant J-B, Trabucho-Alexandre J, France-Lanord C, Donnadieu D et al. 2014. Asian monsoons in a late Eocene greenhouse world. Nature 513: 501–506.

Linder HP and Rudall PJ. 2005. The evolutionary history of Poales. Annu Rev Ecol Evol Syst 36: 107–124.

Edwards EJ and Smith SA. 2010. Phylogenetic analyses reveal the shady history of C4 grasses. Proc Natl Acad Sci USA 107: 2532–2537.

Litvinov DJ. 1902. About the relic character of stony slopes in European Russia. Trav Mus Bot Acad Imp Sci St-Pétersbourg 1: 76–109 [In Russian].

Lomolino MV, Sax DF, Riddle BR and Brown JH. 2006. The island rule and a research agenda for studying ecogeographical patterns. J Biogeogr 33: 1503–1510.

Lovell JT, Aliyu OM, Mau M, Schranz ME, Koch MA, Kiefer C, Song B-H et al. 2013. On the origin and evolution of apomixis in *Boechera*. Plant Reprod 26: 309–315.

Lysak MA and Koch MA. 2010. Phylogeny, Genome, and Karyotype Evolution of Crucifers (Brassicaceae). Genetics and Genomics of the Brassicaceae in: Plant Genetics and Genomics: Crops and Models book series. Springer-Verlag New York. pp. 1–31.

Lysak MA, Berr A, Pečinka A, Schmidt R, McBreen K and Schubert I. 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. Proc Natl Acad Sci USA 103: 5224–5229.

Lysak MA, Koch MA, Pečinka A and Schubert I. 2005. Chromosome triplication found across the tribe Brassiceae. Genome Res 15: 516–525.

Ma X, Jiang H, Cheng J and Xu H. 2012. Spatiotemporal evolution of Paleogene palynoflora in China and its implication for development of the extensional basins in East China. Rev Palaeobot Palynol 184: 24–35.

Ma YZ, Li JJ and Fang XM. 1998. Pollen assemblage in 30.6–5.0 Ma redbeds of Linxia region and climate evolution. Chinese Sci Bull 43: 301–304.

Maddison WP. 1997. Gene trees in species trees. Syst Biol 46: 523–536.

Magyari EK, Chapman JC, Passmore DG, Allen JRM., Huntley JP and Huntley B. 2010. Holocene persistence of wooded steppe in the Great Hungarian Plain. J Biogeogr 37: 915–935.

Magyari EK, Kuneš P, Jakab G, Sümegi P, Pelánková B, Schäbitz F, Braun M et al. 2014. Late pleniglacial vegetation in eastern-central Europe: are there modern analogues in Siberia? Quat Sci Rev 95: 60–79.

Mai DH. 1995. Tertiäre Vegetationsgeschichte Europas. Fischer, Stuttgart. 691 pp.

Manafzadeh S, Staedler YM and Conti E. 2016. Visions of the past and dreams of the future in the Orient: the Irano-Turanian region from classical botany to evolutionary studies. Biol Rev 92: 1365–1388.

Mandáková T and Lysak MA. 2008 Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). Plant Cell 20: 2559–2570.

Mandáková T, Ashby K, Price BJ, Windham MD, Carman JG and Lysak MA. 2020. Genome structure and apomixis in *Phoenicaulis* (Brassicaceae; Boechereae). J Syst Evol 59: 83–92.

Mandáková T, Guo X, Özüdoğru B, Mummenhoff K and Lysak MA. 2018. Hybridization-facilitated genome merger and repeated chromosome fusion after eight million years. Plant J 96: 748–760.

Mandáková T, Heenan PB and Lysak MA. 2010. Island species radiation and karyotypic stasis in *Pachycladon* allopolyploids. BMC Evol Biol 10: 367.

Mandáková T, Joly S, Krzywinski M, Mummenhoff K and Lysak MA. 2010. Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. Plant Cell 22: 2277–2290.

Mandáková T, Li Z, Barker MS and Lysak MA. 2017. Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. Plant J 91: 3–21.

Mandáková T, Mummenhoff K, Al-Shehbaz I, Mucina L, Mühlhausen A and Lysak MA. 2012. Whole genome triplication and species radiation in the southern African tribe Heliophileae (Brassicaceae). Taxon 61: 989–1000.

Mandáková T, Pouch M Harmanová K, Zhan SH, Mayrose I and Lysak MA. 2017. Multispeed genome diploidization and diversification after an ancient allopolyploidization. Mol Ecol 26: 6445–6562.

Mandáková T, Puch M, Brock JR, Al-Shehbaz IA and Lysak MA. 2019. Origin and evolution of diploid and allopolyploid *Camelina* genomes were accompanied by chromosome shattering. Plant Cell 31: 2596–2612.

Mandáková T, Zozomová-Lihová J, Kudoh H, Zhao Y, Lysak MA and Marhold K. 2019. The story of promiscuous crucifers: origin and genome evolution of an invasive species, *Cardamine occulta* (Brassicaceae), and its relatives. Ann Bot 124: 209–220.

Mander L and Punyasena SW. 2018. Fossil Pollen and Spores in Paleoecology. In: Croft DA, Simpson SW and Su DF eds. Methods in Paleoecology: Reconstructing Cenozoic Terrestrial Environments & Ecological Communities. Vertebrate Paleobiology & Paleoanthropology. Springer, pp. 215–234.

Marie Muehe E, Weigold P, Adaktylou IJ, Planer-Friedrich B, Kraemer U, Kappler A and Behrens S. 2015. Rhizosphere microbial community composition affects cadmium and zinc uptake by the metal-hyperaccumulating plant *Arabidopsis halleri*. Appl Environ Microbiol 81: 2173–2181

Martin ARH. 1986. Late Glacial and Early Holocene Vegetation of the Alpine Zone, Kosciusko National Park. In: Barlow BA (Ed.), Flora and Fauna of Alpine Australasia, Ages and Origins. CSIRO, Melbourne. pp. 161–172.

Matoshko A. 2011. Limits of the Pleistocene Glaciations in the Ukraine: a Closer Look. In: Ehlers J, Gibbard PL and Hughes PD (Eds.), Quaternary Glaciations - Extent and Chronology: a Closer Look. Elsevier, Amsterdam, pp. 405–418.

Meerts P. 1995. Phenotypic plasticity in the annual weed *Polygonum aviculare*. Plant Biol 108: 414–424.

Meindl C, Brune V, Listl D, Poschlod P and Reisch C. 2016. Survival and postglacial immigration of the steppe plant *Scorzonera purpurea* to Central Europe. Plant Syst Evol 302: 971–984.

Menzel L, Kopp B and Munkhjargal M. 2016. Does mountain permafrost in Mongolia control water availability? Geophysical research Abstracts at the EGU General Assembly 2016, held 17–22 April, 2016 in Vienna, Austria.

Meusel H and Jäger E. 1992. Vergleichende Chorologie der zentraleuropäischen Flora. Vol. 3. (Text and Maps). Fischer, Jena. 688 pp.

Meusel H, Jäger E and Weinert E. 1965. Vergleichende Chorologie der zentraleuropäischen Flora. Vol. 1. (Text and Maps). Fischer, Jena. 583 pp.

Meusel H, Jäger E, Rauschert S and Weinert E. 1978. Vergleichende Chorologie der zentraleuropäischen Flora. Vol. 2. (Text and Maps). Fischer, Jena. 418 pp.

Mildenhall DC. 1980. New Zealand Late Cretaceous and Cenozoic plant biogeography: a contribution. Paleogeogr Palaeoclim Palaeoecol 31: 197–233.

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P et al. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* L.). Nature 452: 991–996.

Mirarab S, Raez R, Bayzid MdS, Zimmermann T, Swenson MS and Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30: i541–i548.

Mittermeier RA, Turner WR, Larsen, FW, Brooks TM and Gascon C. 2011. Global Biodiversity Conservation: the Critical Role of Hotspots. In: Zachos FE and Habel JC (Eds.), Biodiversity Hotspots – Distribution and Protection of Conservation Priority Areas, Springer, Berlin, pp. 3–22.

Moar NT and Suggate RP. 1979. Contributions to the Quaternary history of the New Zealand flora, 8. Interglacial and glacial vegetation in the Westpoint District, South Island. N Z J Bot 17: 361–387.

Moazzeni, H, Al-Shehbaz IA, German DA, Assadi M, Müller J, Joharchi MR and Memariani F. 2018. A taxonomic revision of the genus *Aethionema* s.l. (Brassicaceae) in Iran. Phytotaxa 356: 241–266.

Mohammadin S, Peterse K, van de Kerke SJ, Chatrou LW, Dönmez AA, Mummenhoff K, Pires JC et al. 2017. Anatolian origins and diversification of *Aethionema*, the sister lineage of the core Brassicaceae. Ann Bot 104: 1042–1054.

Moharrek F, Sanmartín I, Kazempour-Osaloo S and Feliner GN. 2019. Morphological innovations and vast extensions of mountain habitats triggered rapid diversification within the species-rich Irano-Turanian genus *Acantholimon* (Plumbaginaceae). Front Genet 9: 698.

Morales-Briones DF, Kadereit G, Tefarikis DT, Moore MJ, Smith SA, Brockington SF, Timoneda A et al. 2020. Disentangling sources of gene tree discordance in phylotranscriptomic datasets: a case study from Amaranthaceae s.l. Syst Biol 70: 219–235.

Morrone JJ and Crisci JV 1995. Historical biogeography: introduction to methods. Annu Rev Ecol Syst 26: 373–401.

Mosbrugger V and Utescher T. 1997. The coexistence approach — a method for quantitative reconstructions of Tertiary terrestrial palaeoclimate data using plant fossils. Palaeogeogr Palaeoclimatol Palaeoecol 134: 61–86.

Moser BR. 2010. *Camelina* (*Camelina sativa* L.) oil as a biofuels feedstock: golden opportunity or false hope? Lipid Technol 22: 270–273.

Moshkin AV. 2010. Is there any scientific basis for decreasing the conservation status of the Saker Falcon? Raptors Conserv 19: 37–74.

Mucina L. 2018. Biome: evolution of a crucial ecological and biogeographical concept. New Phytol 222: 97–114.

Muller J. 1981. Fossil pollen records of extant angiosperms. Bot Rev Lancaster 47: 1–142.

Munier-Jolain NM, Collard A, Busset H, Guyot SHM and Colbach N. 2014. Investigating and modelling the morphological plasticity of weeds. Field Crops Res 155: 90–98.

Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M and Anderson RP. 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods Ecol Evol 5: 1198–1205.

Nabhan AR and Sarkar IN. 2012. The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. Brief Bioinform 13: 122–134.

Nakleh L, Ruths D and Innan H. 2009. Gene Trees, Species Trees, and Species Networks. In: Guerra R and Goldstein DG (Eds.), Meta-analysis and Combining Information in Genetics and Genomics. Chapman and Hall/CRC, New York, pp. 275–295.

Nasrallah ME, Liu P, Sherman-Broyles S, Boggs NA and Nasrallah JB. 2004. Natural variation in expression of self-incompatibility in *Arabidopsis thaliana*: Implications for the evolution of selfing. Proc Natl Acad Sci USA 101: 16070–16074.

Naud G and Sue J-P. 1975. Contribution à l'étude paléofloristique des Coirons (Ardèche): premières analyses polliniques dans les alluvions sous-basaltiques et interbasaltiques de Mirabel (Miocène supérieur). Bull Soc Géol Fr 17: 820–827.

Neronov VM, Lushchekina AA, Karimova TYu and Arylova NYu. 2012. Population Dynamics of a Key Steppe Species in a Changing World: the Critically Endangered Saiga Antelope. In: Werger MJA and van Staalduinen MA (Eds.), Eurasian Steppes. Ecological Problems and Livelihoods in a Changing World. Springer, Switzerland, pp. 335–356.

Neuffer B, Wesse C, Voss I and Scheibe R. 2018. The role of ecotypic variation in driving worldwide colonization by a cosmopolitan plant. AoB Plants 10: ply005.

Ng FSL, Barr ID and Clark CD. 2010. Using the surface profiles of modern ice masses to inform palaeo-glacier reconstructions. Quat Sci Rev 29: 3240–3255.

Nikitin VP. 2007. Paleogene and Neogene strata in northeastern Asia: paleocarpological background. Russ Geol Geophys 48: 675–682.

Nikolov LA, Shushkov P, Nevado B, Gan X, Al-Shehbaz IA, Filatov D, Donovan Bailey C et al. 2019. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. New Phytol 222: 1638–1651.

Nimis PL, Malyshev LI, Bolognini G and Friesen N. 1998. A multivariate phytogeographic analysis of plant diversity in the Putorana Plateau (N. Siberia). Opera Botanica 163: 1–72.

Nogués-Bravo D. 2009. Predicting the past distribution of species climatic niche. Glob Ecol Biogeogr 18: 521–531.

Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, D'amico J et al. 2001. Terrestrial ecoregions of the world: a new map of fife on Earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51: 933–938.

Osmaston H. 2005. Estimates of glacier equilibrium line altitudes by the area x altitude, the area x altitude balance ratio and the area x altitude balance index methods and their validation. Quat Int 138–139: 22–31.

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N and Clark RM. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327: 92–94.

Owens HL, Campbell LP, Lynnette Dornak L, Saupe EE, Barve N, Soberón J, Ingenloff K et al. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. Ecol Modell 263: 10–18.

Özüdoğru B, Akaydın G, Erik S, Al-Shehbaz IA and Mummenhoff K. 2015. Phylogeny, diversification and biogeographic implications of the eastern Mediterranean endemic genus *Ricotia* (Brassicaceae). Taxon 64: 727–740.

Özüdoğru B, Özgişi K, Tarıkahya-Hacıoğlu B, Ocak A, Mummenhoff K and Al-Shehbaz IA. 2019. Phylogeny of the genus *Noccaea* (Brassicaceae) and a critical review of its generic circumscription. Ann Missouri Bot 104: 339–354.

Pamilo P and Nei M. 1988. Relationships between gene trees and species trees. Mol Biol Evol 5: 568–583.

Parducci L, Bennett KD, Ficetola GF, Alsos IG, Suyama Y, Wood JR, Pedersen MW et al. 2017. Ancient plant DNA in lake sediments. New Phytol 214: 924–942.

Patterson N, Price AL and Reich D. 2006. Population structure and eigenanalysis. PLoS Genetics 2: 20.

Pavelková Řičánková V, Robovský J and Riegert J. 2014. Ecological structure of recent and last glacial mammalian faunas in northern Eurasia: the case of Altai-Sayan refugium. PloS One 9: e85056.

Pavelková Řičánková V, Robovský J, Riegert J and Zrzavý J. 2015. Regional patterns of postglacial changes in the Palearctic mammalian diversity indicate retreat to Siberian steppes rather than extinction. Nat Sci Rep 5: 12682.

Pérez-Escobar OA, Balbuena JA and Gottschling M. 2016. Rumbling orchids: how to assess divergent evolution between chloroplast endosymbionts and the nuclear host. Syst Biol 65: 51–65.

Phillips SJ, Anderson RP and Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecol Modell 190: 231–259.

Phillips SJ, Anderson RP, Dudík M, Schapire RE and Blair ME. 2017. Opening the black box: an open-source release of Maxent. Ecography 40: 887–893.

Pirie MD and Doyle JA. 2012. Dating clades with fossils and molecules: the case of Annonaceae. Bot J Linn Soc 169: 84–116.

Pollock DD, Zwickl DJ, McGuire JA and Hillis DM. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst Biol 51: 664–671.

Polyakova MA, Dembicz I, Becker T, Becker U, Demina ON, Ermakov N, Filibeck G et al. 2016. Scale- and taxon-dependent patterns of plant diversity in steppes of Khakassia. Biodivers Conserv 25: 2251–2273.

Popov SV, Shcherba IG, Ilyina LB, Nevesskaya LA, Paramonova NP, Khondkarian SO and Magyar I. 2006. Late Miocene to Pliocene palaeogeography of the Paratethys and its relation to the Mediterranean. Palaeogeogr Palaeoclimatol Palaeoecol 238: 91–106.

Popova S, Utescher T, Gromyko D, Mosbrugger V, Herzog E and Francois L. 2013. Vegetation change in Siberia and the northeast of Russia during the Cenozoic cooling – a study based on diversity of plant functional types. Palaois 28: 418–432.

Popova S, Utescher T, Gromyko DV, Bruch AA, Henrot A-J and Mosbrugger V. 2017. Cenozoic vegetation gradients in the mid- and higher latitudes of Central Eurasia and climatic implications. Palaeogeogr Palaeoclimatol Palaeoecol 467: 69–82.

Posadas P, Crisci JV and Katinas L. 2006. Historical biogeography: a review of its basic concepts and critical issues. J Arid Environ 66: 389–403.

Pritchard JK, Stephens M and Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Purvis A, Gittleman JL and Brooks TM. 2005. Phylogeny and Conservation. Cambridge University Press, Cambridge, UK, 431 pp.

Quan C, Liu Z, Utescher T, Jin J, Shu J, Li Y-X and Liu Y. 2012. Revisiting the Paleogene climate pattern of East Asia: a synthetic review. Earth-Sci Rev 139: 213–230.

Radosavljevic A and Anderson RP. 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. J Biogeogr 41: 629–643.

Raj A, Stephens M and Pritchard JK. 2014. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics 197: 573–589.

Ree RH and Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst Biol 57: 4–14.

Ree RH, Moore BR, Webb CO and Donoghue MJ. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59: 2299–2311.

Reisz RR and Müller J. 2004. Molecular time scales and the fossil record: a paleontological perspective. Trends Genet 20: 237–241.

Renner SS, Grimm GW, Kapli P and Denk T. 2016. Species relationships and divergence times in beeches: new insights from the inclusion of 53 young and old fossils in a birth–death clock model. Philos Trans R Soc B 371: 20150135.

Renner SS. 2005. Relaxed molecular clocks for dating historical plant dispersal events. Trends Plant Sci 10: 550–558.

Renner SS. 2016. Available data point to a 4-km-high Tibetan Plateau by 40 Ma, but 100 molecular-clock papers have linked supposed recent uplift to young node ages. J Biogeogr 43: 1479–1487.

Ricklefs RE. 2004. A comprehensive framework for global patterns in biodiversity. Ecol Lett 7: 1–15.

Rieseberg LH and Soltis DE. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol Trends Plants 5: 65–84.

Rögl F. 1999. Mediterranean and Paratethys. Facts and hypotheses of an Oligocene to Miocene paleography. Geologica Carpathica 50: 339–349.

Rollins RC. 1993. The Cruciferae of Continental North America. Stanford: Stanford University Press, 996 pp.

Ronquist F and Sanmartín I. 2011. Phylogenetic methods in biogeography. Annu Rev Ecol Evol Syst 42: 441–464.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA and Huelsenbeck J. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61: 539–542.

Ronquist F. 1997. Dispersal–vicariance analysis: a new approach to the quantification of historical biogeography. Syst Biol 46: 195–203.

Rosenberg MS and Kumar S. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proc Natl Acad Sci USA 98: 10751–10756.

Rydin C and Källersjö M. 2002. Taxon sampling and seed plant phylogeny. Cladistics 18: 485–513.

Saiko T. 2001. Environmental Crisis: Geographical Case Studies in Post-socialist Eurasia. Addison-Wesely Longman Pearson, London, 320 pp.

Salywon AM, Dierig DA, Rebman JP and de Rodríguez DJ. 2005. Evaluation of new *Lesquerella* and *Physaria* (Brassicaceae) oilseed germplasm. Am J Bot 92: 53–62.

Sandom C, Faurby S, Sandel B and Svenning J-C. 2014. Global late Quaternary megafauna extinctions linked to humans, not climate change. Proc R Soc B 281: 20133254.

Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit Rev Biochem Mol 37: 121–147.

Sanmartín I. 2012. Historical biogeography: evolution in time and space. Evo Edu Outreach 5: 555–568

Saupe EE, Barve V, Myers CE, Soberón J, Barve N, Hensz CM, Peterson AT et al. 2012. Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. Ecol Modell 237–238: 11–22.

Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, Bayly MJ et al. 2012. Testing the impact of calibration on molecular divergence times using a fossil–rich group: the case of *Nothofagus* (Fagales). Syst Biol 61: 289–313.

Schranz ME and Mitchell-Olds T. 2006. Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. Plant Cell 18: 1152–1165.

Schranz ME, Windsor AJ, Song B-H, Lawton-Rauh A and Mitchell-Olds T. 2007. Comparative genetic mapping in *Boechera stricta*, a close relative of *Arabidopsis*. Plant Physiol 144: 286–298.

Schultz J. 2002. The Ecozones of the World: The Ecological Division of the Geosphere. Springer, Berlin, 451 pp.

Schulz HM, Bechtel A and Sachsenhofer RF. 2005. The birth of the Paratethys during the Early Oligocene: From Tethys to an ancient Black Sea analogue? Glob Planet Change 49: 163–176.

Schulz OE. 1936. Cruciferae. In: Engler A, Prantl K (Eds.) Die natürlichen Pflanzenfamilien, 2nd edn. Vol. 17b. Leipzig: W. Engelmann, pp. 227–658.

Schwarz EN, Ruhlman TA, Weng M-L, Khiyami MA, Sabir JSM, Hajarah NH, Alharbi NS et al. 2017. Plastome-wide nucleotide substitution rates reveal accelerated rates in Papilionoideae and correlations with genome features across legume subfamilies.J Mol Evol 84: 187–203.

Selmeier A. 2005. *Capparidoxylon holleisii* nov. spec., a silicified *Capparis* (Capparaceae) wood with insect coprolites from the Neogene of southern Germany. Zitteliana 45: 199–209.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC et al. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92: 142–166.

Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M and Morris AB. 2014. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. Am J Bot 101: 1987–2004.

Shcheglovitova M and Anderson RP. 2013. Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. Ecol Modell 269: 9–17.

Simakova AN. 2006. The vegetation of the Russian Plain during the second part of the Late Pleistocene (33–18 ka). Quat Int 149: 110–114.

Simpson M. 2019. Plant Systematics, Elsevier Academic Press, Burlington, Massachusetts, 774 pp.

Sloan DB, Alverson AJ, Wu M, Palmer JD and Taylor DR. 2012. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. Genome Biol Evol 4: 294–306.

Slotte T, Ceplitis A, Neuffer B, Hurka H and Lascoux M. 2006. Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. Am J Bot 93: 1714–1724.

Smelansky IE and Tishkov AA. 2012. The Steppe Biome in Russia: Ecosystem Services, Conservation Status, and Actual Challenges. In: Werger MJA and van Staalduinen MA (Eds.), Eurasian Steppes. Ecological Problems and Livelihoods in a Changing World. Springer, Switzerland, pp. 45–101.

Šmíd J and Tolley KA. 2019. Calibrating the tree of vipers under the fossilized birth-death model. Sci Rep 9: 5510.

Smith DR. 2015. Mutation rates in plastid genomes: they are lower than you think. Genome Biol Evol 7: 1227–1234.

Smith TB, Saatchi S, Graham C, Slabbekoorn H and Spicer G. 2005. Putting Process on the Map: Why Ecotones are Important for Preserving Biodiversity. In: Purvis A, Gittleman JL and Brooks TM. (Eds.), Phylogeny and Conservation. Cambridge University Press, Cambridge, UK, pp. 166–197.

Sober E. 2008. Absence of evidence and evidence of absence: evidential transitivity in connection with fossils, fishing, fine-tuning, and firing squads. Philos Stud 143: 63–90.

Soberón J and Peterson AT. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodivers Inform 2: 1–10.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D et al. 2009. Polyploidy and angiosperm diversification. Am J Bot 96: 336–348.

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V et al. 2000. Angiosperm phylogeny inferred from 18s rDNA, *rbc*L, and *atp*B sequences. Bot J Linn Soc 133: 381–461.

Soltis DE, Soltis PS, Nickrent DL, Johnson LA, Hahn WJ, Hoot SB, Sweere JA et al. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Ann Missouri Bot 84: 1–49.

Soltis PS, Soltis DE and Chase MW. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402: 402–404.

St Onge KR, Källman T, Slotte T, Lascoux M and Palmé AE. 2011. Contrasting demographic history and population structure in *Capsella rubella* and *Capsella grandiflora*, two closely related species with different mating systems. Mol Ecol 20: 3306–3320.

Stamatakis A. 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.

Stange M, Sánchez-Villagra MR, Salzburger W and 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. Syst Biol 67: 681–699.

Stauch G and Lehmkuhl F. 2011. Extent and Timing of Quaternary Glaciations in the Verkhoyansk Mountains. In: Ehlers J, Gibbard PL and Hughes PD (Eds.), Quaternary Glaciations - Extent and Chronology: a Closer Look. Elsevier, Amsterdam, pp. 877–881.

Stefanović S, Rice DW and Palmer JD. 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots? BMC Evol Biol 4: 35.

Stevens PF. 2003. History of Taxonomy. In: eLS. John Wiley & Sons, Ltd: Chichester.

Stift M, Kolář F and Meirmans PG. 2019. STRUCTURE is more robust than other clustering methods in simulated mixed-ploidy populations. Heredity 123: 429–441.

Stockenhuber R, Zoller S, Shimizu-Inatsugi R, Gugerli F, Shimizu KK, Widmer A and Fischer MC. 2015. Efficient detection of novel nuclear markers for Brassicaceae by transcriptome sequencing. PLoS One 10: e0128181.

Strömberg CAE. 2011. Evolution of grasses and grassland ecosystems. Annu Rev Earth Planet Sci 39: 517–544.

Stuessy TF. 2009. Plant Taxonomy: The Systematic Evaluation of Comparative Data. Columbia University Press, 568 pp.

Sudnik-Wójcikowska, B, Moysiyenko II, Zachwatowicz M and Jabłońska E. 2011. The value and need for protection of kurgan flora in the anthropogenic landscape of steppe zone in Ukraine. Plant Biosyst 145: 638–653.

Sun J, Ni X, Bi S, Wu W, Ye J, Meng J and Windley BF. 2014. Synchronous turnover of flora, fauna and climate at the Eocene–Oligocene Boundary in Asia. Nat Sci Rep 4: 7463.

Sun M, Soltis DE, Soltis PS, Zhu X, Burleigh JG and Chen Z. 2015. Deep phylogenetic incongruence in the angiosperm clade Rosidae. Mol Phylogenet Evol 83: 156–166.

Svendsen JI, Krüger LC, Magerud J, Astakhov VI, Paus A, Nazarov D and Murray A. 2014. Glacial and vegetation history of the Polar Ural Mountains in northern Russia during the Last Ice Age, Marine Isotope Stages 5–2. Quat Sci Rev 92: 409–428.

Takhtajan A. 1986. Floristic regions of the world. University of California Press, Berkeley, pp. 522.

Tang HB, Bowers JE, Wang X, Ming R, Alam M and Paterson AH. 2008. Perspective – Synteny and collinearity in plant genomes. Science 320: 486–488.

Tank DC, Eastman JM, Pennell MW, Soltis PS, Soltis DE, Hinchliff CE, Brown JW et al. 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. New Phytol 207: 454–467.

Tarasov PE, Volkova VS, Webb T, Guiot J, Andreev AA, Bezusko LG, Bezusko TV et al. 2000. Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern Eurasia. Journal Biogeogr 27: 609–620.

Tarasov PE, Williams JW, Andreev A, Nakagawa T, Bezrukova E, Herzschuh U, Igarashi Y et al. 2007. Satellite- and pollen-based quantitative woody cover reconstructions for northern Asia: verification and application to late-Quaternary pollen data. Earth Planet Sci Lett 264: 284–298.

Ter Borg M, Vasiliev I, Stoica M, Knežević S, Matenco L, Krijgsman W, Rundić L et al. 2013. The isolation of the Pannioan basin (Central Paratethys): new constraints from magnetostratigraphy and biostratigraphy. Glob Planet Change 103: 99–118.

Ter Borgh M, Vasiliev I, Stoica M, Knežević S, Matenco L, Krijgsman W, Rundić L et al. 2013. The isolation of the Pannonian basin (Central Paratethys): New constraints from magnetostratigraphy and biostratigraphy. Glob Planet Change 103: 99–118.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature 408: 796–815.

Tudryn A, Chalié F, Lavrushin YuA, Antipov MP, Spiridonova EA, Lavrushin V, Tucholka P. et al. 2013. Late Quaternary Caspian Sea environment: Late Khzarian and Early Khvalynian transgressions from the lower reaches of the Volga River. Quat Int 292: 193–204.

Turner C. 2000. The Eemian interglacial in north European plain and adjacent areas Neth J Geosci 79: 217–232.

Urdea P, Onaca A, Ardelean F and Ardelean M. 2011. New Evidence on the Quaternary Glaciation in the Romanian Carpathians. In: Ehlers J, Gibbard PL and Hughes PD (Eds.), Quaternary Glaciations - Extent and Chronology: a Closer Look. Elsevier, Amsterdam, pp. 305–322.

Utescher T and Mosbrugger V. 2007. Eocene vegetation patterns reconstructed from plant diversity — a global perspective. Palaeogeogr Palaeoclimatol Palaeoecol 247: 243–271.

Utescher T, Bruch AA, Erdei B, François I, Ivanov D, Jacques FMB, Kern AK et al. 2014. The coexistence approach – theoretical background and practical considerations of using plant fossils for climate quantification. Palaeogeogr Palaeoclimatol Palaeoecol 410: 58–73.

Utescher T, Erdei B, Hably L and Mosbrugger V. 2017. Late Miocene vegetation of the Pannonian Basin. Palaeogeogr Palaeoclimatol Palaeoecol 467: 131–148.

Wu F, Fang X, Ma Y, Herrmann M, Mosbrugger V, An Z and Miao Y. 2007. Plio–Quaternary

stepwise drying of Asia: evidence from a 3-Ma pollen record from the Chinese Loess Plateau. Earth Planet Sci Lett 257: 160–169.

van Baak CGC, Vasiliev I, Stoica M, Kuiper KF, Forte AM, Aliyeva E and Krijgsman W. 2013. A magnetostratigraphic time frame for Plio-Pleistocene transgressions in the South Caspian Basin, Azerbaijan. Glob Planet Change 103: 119–134.

Van de Peer Y, Fawcett JA, Proost S, Sterck L and Vandepoele K. 2009. The flowering world: a tale of duplications. Trends Plant Sci 14: 680–688.

Velichko A. A. 1999. General Features of the Landscape and Climate of Northern Eurasia in the Cenozoic. In: Velichko AA and Nechaev VP (Eds.), Climate and Environment Changes During the Last 65 Million Years; Cenozoic: from Paleocene to Holocene. GEOS, Moscow, pp. 219–233 [In Russian].

Velichko AA and Nechaev VP. 1999. Climate and Environment Changes during the last 65 Million Years; Cenozoic: from Paleocene to Holocene. GEOS, Moscow, 260 pp [In Russian].

Velichko AA, Faustova MA, Pisareva VV, Gribchenko YuN, Sudakova NG and Lavrentiev NV. 2011. Glaciations of the East European Plain: Distribution and Chronology. In: Ehlers J, Gibbard PL and Hughes PD (Eds.), Quaternary Glaciations - Extent and Chronology: a Closer Look. Elsevier, Amsterdam, pp. 337–359.

Venter O, Fuller RA, Segan DB, Carwardine J, Brooks T, Butchart SH, Di Marco M et al. 2014. Targeting global protected area expansion for imperilled biodiversity. PLoS 12: e1001891.

Walden N, German DA, Wolf EM, Kiefer M, Rigault P, Huang X-C, Kiefer C et al. 2020. Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. Nat Commun 11: 3795.

Walker JF, Walker-Hale N, Vargas OM, Larson DA and Stull GW. 2019. Characterizing gene tree conflict in plastome-inferred phylogenies. PeerJ 7: e7747.

Wallace A. 1876. Wallace's geographical distribution of animals. Nature 14: 165–168.

Wallberg A, Thollesson M, Farris J and Jondelius U. 2004. The phylogenetic position of the comb jellies (Ctenophora) and the importance of taxonomic sampling. Cladistics 20: 558–578.

Walter H. 1974. Die Vegetation Osteuropas, Nord- und Zentralasiens. G. Fischer, Stuttgart, 452 pp.

Wang Q and Mao KS. 2016. Puzzling rocks and complicated clocks: how to optimize molecular dating approaches in historical phytogeography. New Phytol 209: 1353–1358.

Wang WM and Shu JW. 2013. Cenozoic xeromorphic vegetation in China and its spatial and temporal development in connection with global changes. Palaeoworld 22: 86–92.

Wang YP, Tang JS, Chu CQ and Tian J. 2000. A preliminary study on the introduction and cultivation of *Crambe abyssinica* in China, an oil plant for industrial uses. Ind Crops Prod 12: 47–52.

Warren DL and Seifert SN. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol Appl 21: 335–342.

Wascher M and Kubatko L. 2021. Consistency of SVDQuartets and maximum likelihood for coalescent-based species tree estimation. Syst Biol 70: 33–48.

Weng M-L, Ruhlman TA and Jansen RK 2017. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. New Phytol 214: 842–851.

Weng M-L, Ruhlman TA and Jansen RK. 2016. Plastid–nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. Genome Biol Evol 8: 1824–1838.

Wesche K and Treiber J. Abiotic and Biotic Determinants of Steppe Productivity and Performance – a View from Central Asia. In: Werger MJA and van Staalduinen MA (Eds.), Eurasian Steppes. Ecological Problems and Livelihoods in a Changing World. Springer, Switzerland, pp. 3–43.

Wesche K, Ambarlı D, Kamp J, Török P, Trieber J and Dengler J. 2016. The Palaearctic steppe biome: a new synthesis. Biodivers Conserv 25: 2197–2231.

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

West RG, Lambert CA and Sparks B. 1964. Interglacial deposits at Ilford, Essex Phil Tran R Soc B 247: 185–212.

White RP, Murray S and Rohweder M. 2000. Pilot Analysis of Global Ecosystems. Grassland Ecosystems. World Resource Institute, Washington, DC, 70 pp.

Wicke S, Schäferhoff B, dePamphilis CW and Müller KF. 2014. Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. Mol Biol Evol 31: 529–545.

Wiens JJ and Donoghue MJ. 2004. Historical biogeography, ecology and species richness. Trends Ecol Evol 19: 639–644.

Wiens JJ. 2012. Why biogeography matters: historical biogeography vs. phylogeography and community phylogenetics for inferring ecological and evolutionary processes. Front Biogeogr 4: 128–135.

Wilf P and Escapa IH. 2015. Green Web or megabiased clock? Plant fossils from Gondwanan Patagonia speak on evolutionary radiations. New Phytol 207: 283–290.

Willerslev E, Davison J, Moora M, Zobel M, Coissac E, Edwards ME, Lorenzen ED et al. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. Nature 506: 47–51.

Wódkiewicz M, Dembicz I and Moysiyenko II. 2016. The value of small habitat islands for the conservation of genetic variability in a steppe grass species. Acta Oecol 76: 22–30.

World Wildlife Fund. 2012. Conservation Science Data and Tools. Terrestrial Ecoregions of the World https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world: last accessed on December 20th 2020.

Wu H-J, Zhang Z, Wang J-Y, Oh D-H, Dassanayake M, Liu B, Huang Q et al. 2012. Insights into salt tolerance from the genome of *Thellungiella salsuginea*. Proc Natl Acad Sci USA 109: 12219–12224.

Xiang QY, Thomas DT and Xiang QP. 2012. Resolving and dating the phylogeny of Cornales – effects of taxon sampling, data partitions, and fossil calibrations. Mol Phylogenet Evol 59: 123–138.

Xu J-H, Liu Q, Hu W, Wang T, Xue Q and Messing J. 2015. Dynamics of chloroplast genomes in green plants. Genomics 106: 221–231.

Yang X, Scuderi L, Paillou P, Liu Z, Li H and Ren X. 2011. Quaternary environmental changes in the drylands of China — a critical review. Quat Sci Rev 30: 3219–3233.

Yang Y, Tilman D, Furey G and Lehman C. 2019. Soil carbon sequestration accelerated by restoration of grassland biodiversity. Nat Commun 10: 718.

Yu W-B, Huang P-H, Li D-Z and Wang H. 2013. Incongruence between nuclear and chloroplast DNA phylogenies in *Pedicularis* section *Cyathophora* (Orobanchaceae). PLoS One 8: e74828.

Yurtsev BA. 1982. Relics of the Xerophyte Vegetation of Beringia in Northeastern Asia. In: Hopkins DM, Matthews Jr JV, Schweger CE, and Young SB (Eds.), Paleoecology of Beringia. Academic Press, New York, pp. 157–177.

Yurtsev BA. 2001. The Pleistocene Tundra-Steppe and the productivity paradox: the landscape approach. Quat Sci Rev 20: 165–174.

Zachos JC, Dickens GR and Zeebe RE. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature 451: 279–283.

Zhao L, Annie ASH, Amrita S, Yi SKF, Rudolf M. 2013. Does better taxon sampling help? A new phylogenetic hypothesis for Sepsidae (Diptera: Cyclorrhapha) based on 50 new taxa and the same old mitochondrial and nuclear markers. Mol Phylogenet Evol 69: 153–164.

Zhao T, Xue J, Kao S-M, Li Z, Zwaenepoel A, Schranz ME and Van de Peer Y. in rev. Novel phylogeny of angiosperms inferred from whole-genome microsynteny analysis. preprint on bioRxiv

Zhilich S, Rudaya N, Krivonogov S, Nazarova L and Pozdnyakov D. 2017. Environmental dynamics of the Baraba forest-steppe (Siberia) over the last 8000 years and their impact on the types of economic life of the population. Quat Sci Rev 163: 152–161.

Zhu A, Guo W, Gupta S, Fan W and Mower JP. 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol 209: 1747–1756.

Zimmermann HH, Harms L, Epp LS, Mewes N, Bernhardt N, Kruse S, Stoof-Leichsenring KR et al. 2019. Chloroplast and mitochondrial genetic variation of larches at the Siberian tundra-taiga ecotone revealed by de novo assembly. PloS one 14: e0216966.

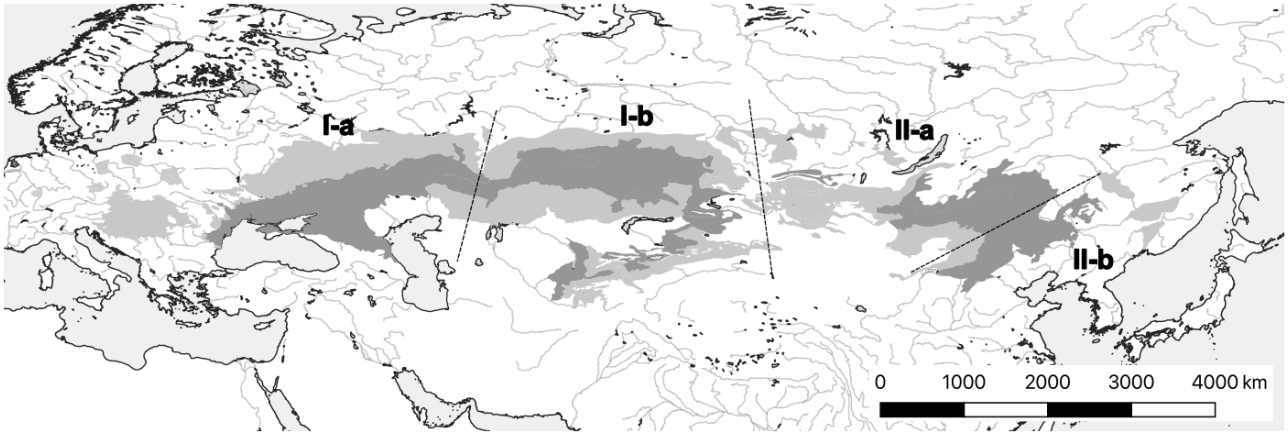
Zimov SA, Zimov NS, Tikhonov AN and Chapin FS. 2012. Mammoth steppe: a high-productivity phenomenon. Quat Sci Rev 57: 26–45.

Zlotin R. 2002. Biodiversity and Productivity of Ecosystems. In: Shahgedanova M. (Ed.), The Physical Geography of Northern Eurasia. Oxford, UK: Oxford University Press, pp. 169–190.

Žerdoner Čalasan A, Kretschmann J and Gottschling M. 2020. Contemporary integrative taxonomy for sexually deprived protists: a case study of *Trachelomonas* (Euglenaceae) from western Ukraine. Taxon 69: 28–42.

Žerdoner Čalasan A, Kretschmann J, Filipowicz NH, Irimia R-E, Kirsch M and Gottschling M. 2019. Towards global distribution maps of unicellular organisms such as calcareous dinophytes based on DNA sequence information. Mar Biodivers 49: 749–758.

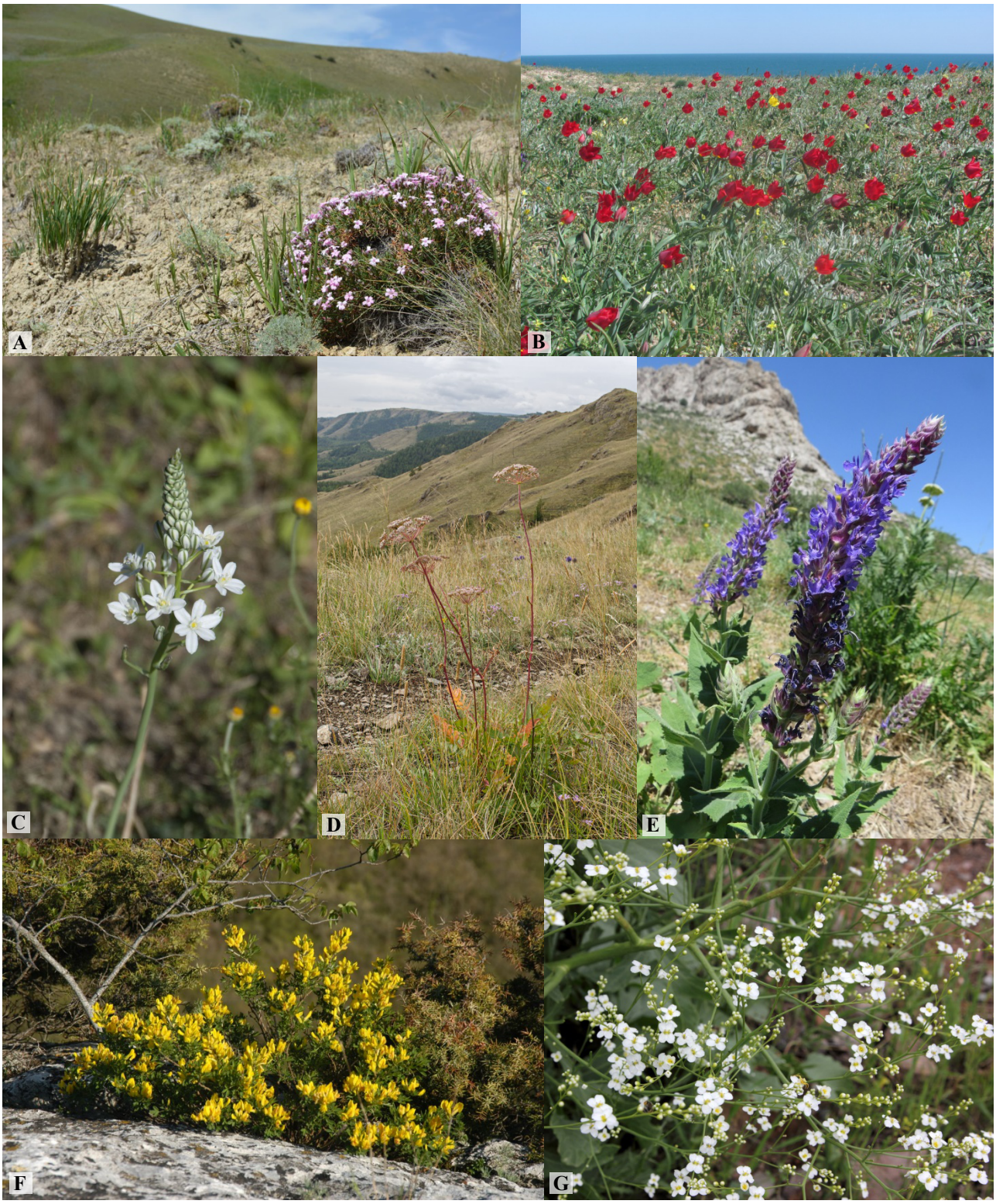
Appendix



Appendix 1: Geographical delimitation of the Eurasian steppe belt
Eurasian steppe belt and its exclaves according to our interpretation. The dark grey shaded areas represent the zonal core part of the Eurasian steppe belt and the light grey shaded areas represent additional ecoregions (for definition see Olson et al. 2001) that share abiotic as well as biotic features of the Eurasian steppe belt.

The dark grey shaded zonal core part of the Eurasian steppe belt consists of the following ecoregions: Alai-Western Tian Shan steppe, Altai steppe and semi-desert, Daurian forest steppe, Emin Valley steppe, Kazakh steppe, Kazakh Upland, Mongolian-Manchurian grassland, Pontic steppe, Sayan Intermontane steppe and Tian Shan foothill arid steppe.

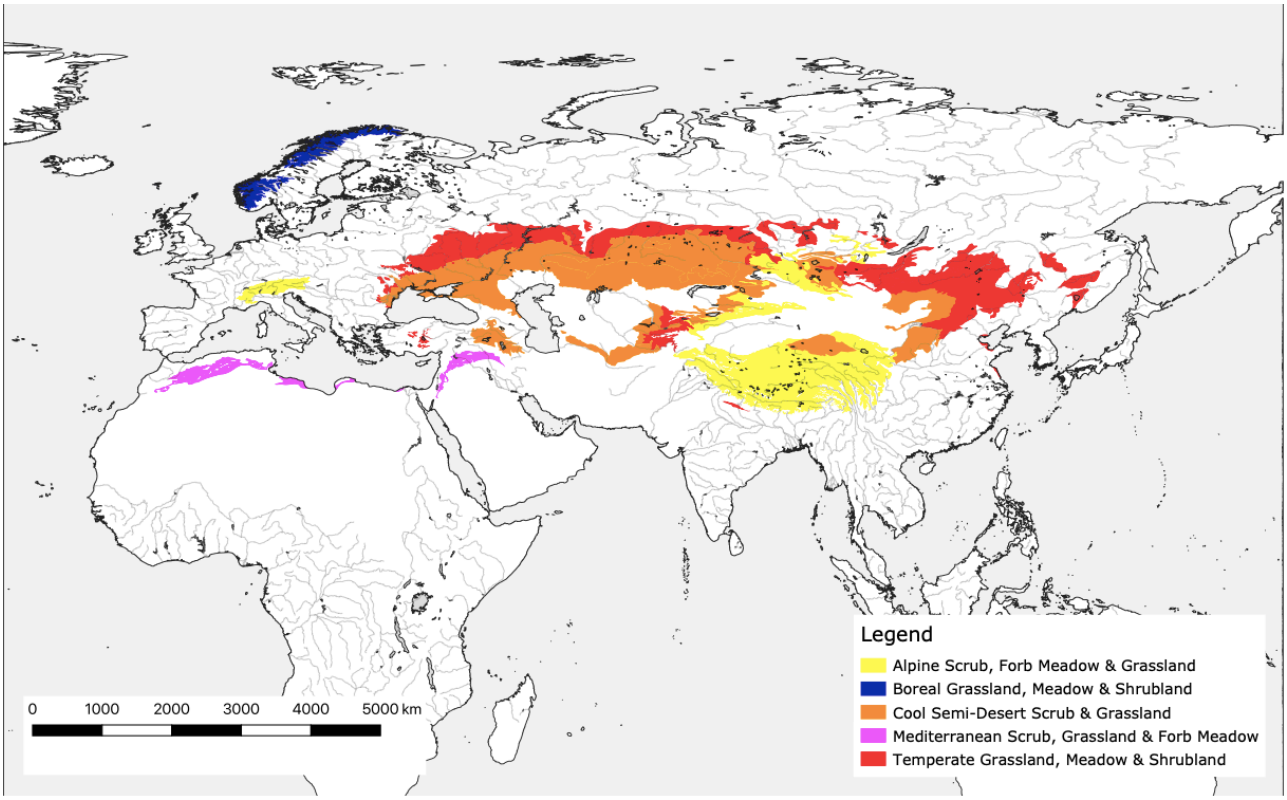
The light grey shaded steppe-accompanying area consists of the following ecoregions: Altai montane forest and forest steppe, Amur meadow steppe, Bohemian steppe exclaves, East European forest steppe, Eastern Gobi desert steppe, Gissaro-Alai open woodlands, Great Lakes Basin desert steppe, Kazakh forest steppe, Kazakh semi-desert, Kujawy extrazonal xeric grasslands, Moravian steppe exclaves, Oder Valley, Pannonian mixed forests (Pannonic steppes and forest steppes), Polish Uplands, Rhine Valley, Selenge-Orkhon forest steppe, South Siberian forest steppe, Suiphun-Khanka meadows and forest meadows, and Tian Shan montane steppe and meadows.



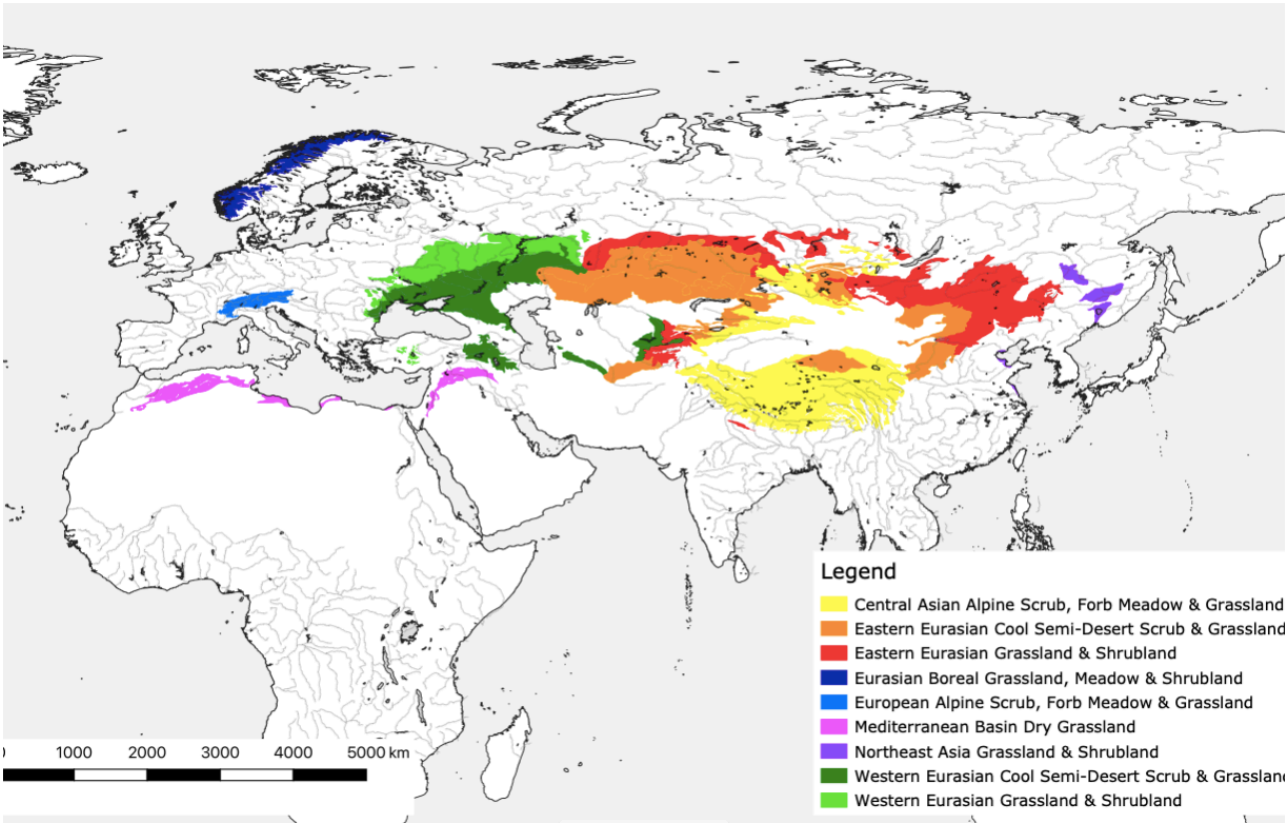
Appendix 2: Photographs of floral elements specific for Euro-Siberian steppe
A, *Acantholimon alatavicum* (Kazakhstan, Almaty region, Raiymbek district, Zhabyrtau mountains, sandy area of the southern slope, photo taken on 06/07/2017 by Pavel Gorbunov); B, *Tulipa suaveolens* (Crimea, Kerch Peninsula, Mount Opuk, photo taken on 04/15/2018 by Pavel Evseenkov); C, *Ornithogalum ponticum* (Mountain Crimea, North Demerdzhi mountain, photo taken on 21/06/2009 by Alexander Lebedev); D, *Seseli krylovii* (Bashkortostan, Uchalinsky district, Ilchigulovsky village council, near Muldashevo village on a rocky slope, photo taken on 07/09/2018 by Maria Zhukova); E, *Salvia deserta* (Kazakhstan, South Kazakhstan region, Syrdarya-Turkestan state regional natural park, Boraldai branch, photo taken on 26/06/2018 by Alexander Ebel); F, *Chamaecytisus ruthenicus* (Crimea, Bakhchisarai district, the rocky slope on which the Donators temple is located, photo taken on 05/05/2007 by Alexander Lebedev); G, *Crambe orientalis* (Uzbekistan, Tashkent region, Bostanlyk district, photo taken on 12/05/2012 by Alim Gaziev).

Appendix 3: Photographs of floral elements specific for Mongol-Chinese steppe
A, *Filifolium sibiricum* (Primorsky Territory, Kavalerovsky District, sea coast between Dubovaya and Nerpa bays, photo taken on 07/08/2018 by Vera Volcotrub); B, *Hemerocallis minor* (Irkutsk region, Irkutsk district, near Bokovo, meadow on the left bank of the river Angara, photo taken on 24/06/2014 by Lyudmila Palamarchuk); C, *Panzerina lanata* (Altai, Chuiskaya steppe, foothills of the South Chuisky ridge, photo taken on 30/06/2010 by Denis Kochetkov); D, *Schizonepeta multifida* (Irkutsk region, Olkhonsky district, western coast of the Lake Baikal, the coast of the Small Sea, a rocky slope between the Chara base and the village Sarma, photo taken on 07/05/2014 by Galina Chulanova); E, *Lespedeza bicolor* (Primorye, Nakhodka, by the forest road, photo taken on 22/07/2016 by Natalia Surovtseva); F, *Stellera chamaejasme* (Mongolia, Uverhangai aimag, river valley Orkhon on a stony dry slope, photo taken on 03/06/2017 by Marina Skotnikova); G, *Cymbaria daurica* (Irkutsk Region, Ust-Ordynsky District, Mount Bulen, photo taken on 15/06/2014 by Lyudmila Palamarchuk).

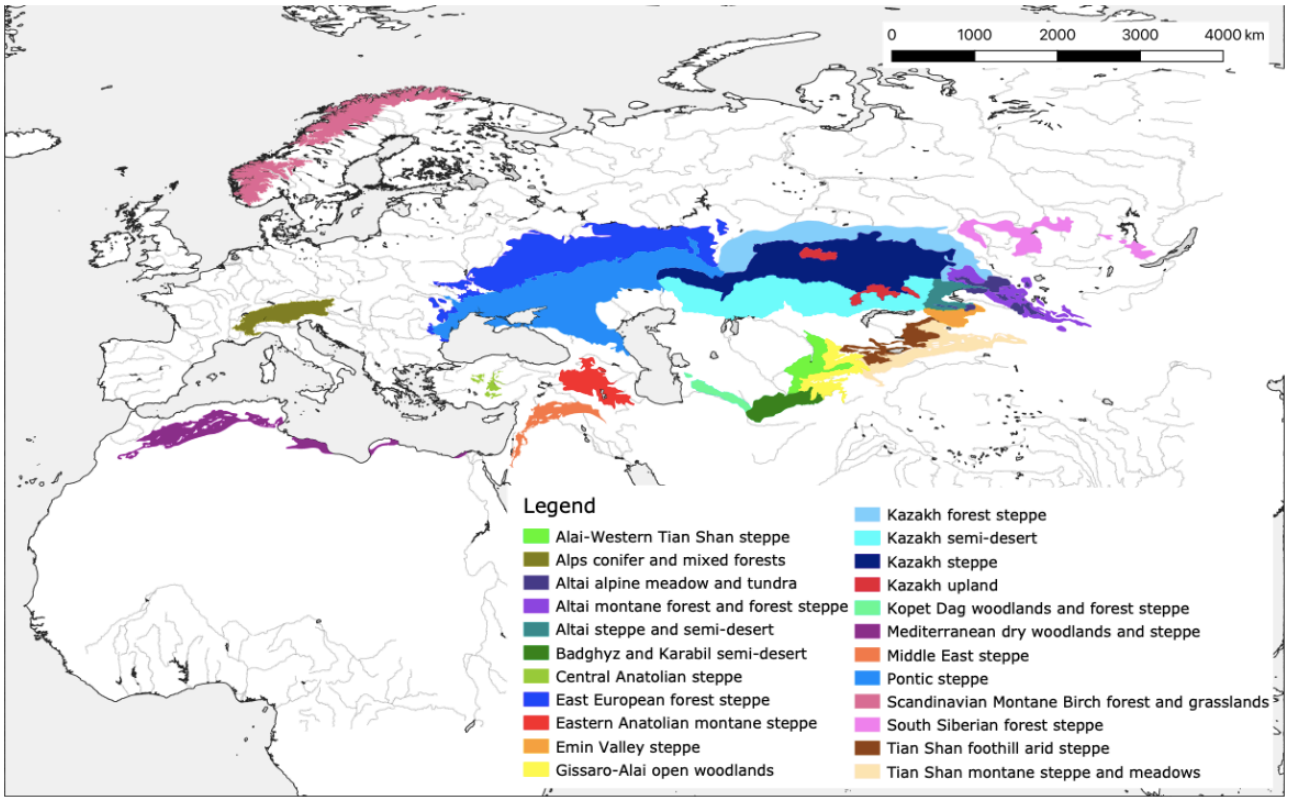
Appendix 4: Photographs of floral elements found across the entire Eurasian steppe belt
A, *Agropyron cristatum* (Irkutsk region, the coast of the Lake Baikal in the Maloye More region, near village Kurkut, photo taken on 22/06/2017 by Lyudmila Palamarchuk); B, *Odontarrhena obovata* (East Kazakhstan, Ust-Kamenogorsk city, Leskhoz, hillside, photo taken on 22/05/2018 by Tamara Rib); C, *Veronica spicata* (Pskov region, Sebezhsky district, near Idritsa settlement, a herb-grass meadow, photo taken on 02/08/2019 by Maria Novikova); D, *Onobrychis arenaria* (Voronezh Region, Liskinsky District, Divnogorye Museum-Reserve, chalk slope, photo taken on 01/06/2019 by Marina Skotnikova); E, *Krascheninnikovia ceratoides* (Altai Territory, Barnaul, southern steppe slope of the river Ob, photo taken on 13/07/2019 by Alexander Fateryga); F, *Scorzonera austriaca* (Irkutsk region, Irkutsk district, near village Zherdovka, photo taken on 14/08/2015 by Lyudmila Palamarchuk); G, *Allium strictum* (Ukraine, Ternopil region, Kremenets district, national park Kremenets mountains, rocky outcrop, photo taken on 26/06/2010 by Andrey Kovalchuk).



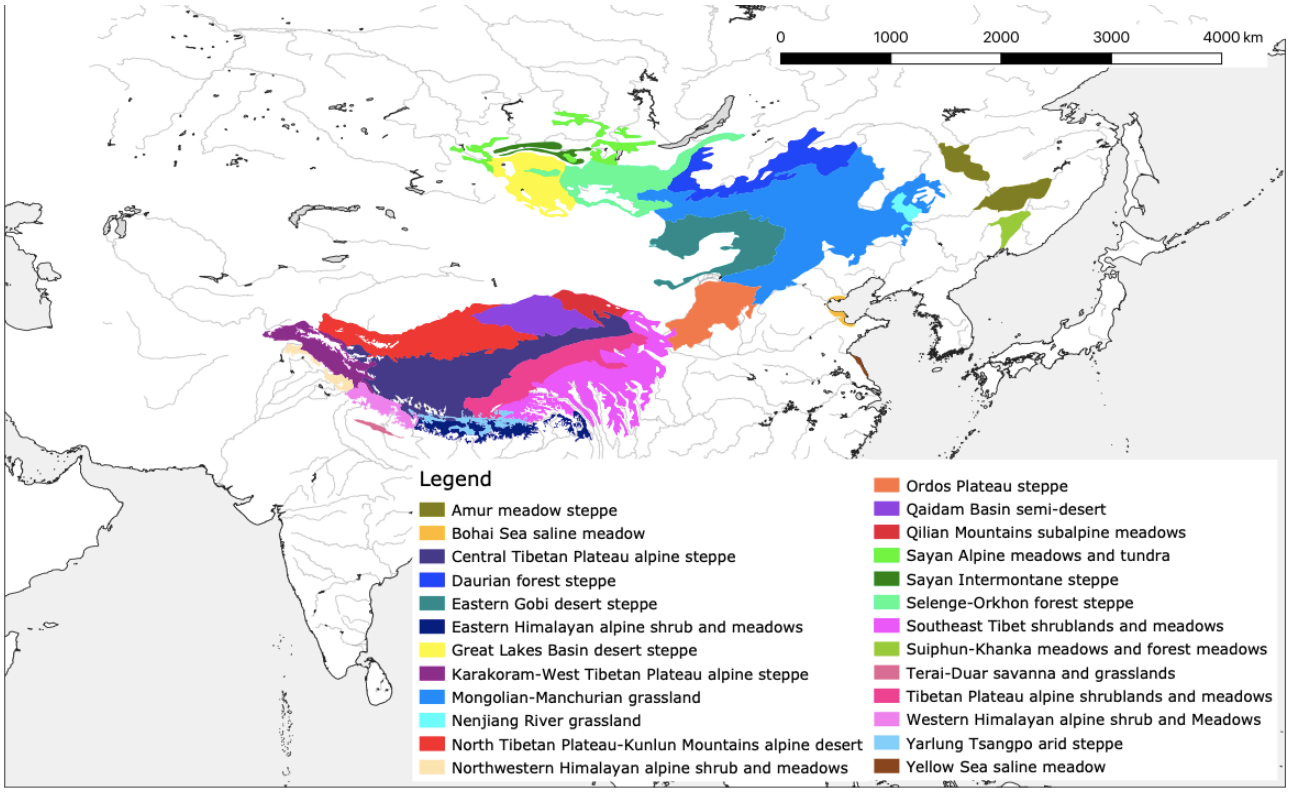
Appendix 5: Division of Eurasian grasslands according to the WWF
Division of Eurasian grasslands according to the World Wildlife Fund (WWF). The Faroe Islands boreal grasslands were excluded from the graphic, due to their small area they cover in comparison to the other subunits.



Appendix 6: Subdivision of Eurasian grasslands according to the WWF
Subdivision of Eurasian grasslands according to the World Wildlife Fund (WWF). The Faroe Islands boreal grasslands were excluded from the graphic, due to their small area they cover in comparison to the other subunits.



Appendix 7: Further subdivision of the western portion of Eurasian grasslands
Further subdivision of the western portion of Eurasian grasslands according to Olson et al. 2001.



Appendix 8: Further subdivision of the eastern portion of Eurasian grasslands
Further subdivision of the eastern portion of Eurasian grasslands according to Olson et al. 2001.

Acknowledgements

First I would like to thank my advisor Prof. Dr. Herbert Hurka and my official supervisor apl. Prof. Dr Barbara Neuffer for giving me a free hand in all the decision making from the beginning of the project until the official end of it. Being in complete control of the project helped me to improve my organisational skills, funding management, student supervision, lecture/seminar preparation, scientific writing ability, as well as damage control, when something did not go according to the plan. In the course of this project I had an opportunity to visit two research institutes and I would like to thank the research group leaders — Dr. Frank Blattner and Prof. Dr. Andreas Houben from the IPK, and Prof. Dr. Martin A. Lysák from the CEITEC — to allow me to carry out a part of my research in their labs, as well as their students and the staff that helped me with laborious lab work. A special thank you goes to Simon, Anna, Christina, and Tobias, four partners-in-crime who helped me to overcome bioinformatics issues or just happened to be the perfect chitchat partners to confide in. Writing about chitchat… Rudi, Claudia, Petra, and Ulrike, thank you for helping me with the greenhouses, plant care, administration and lab tips. Another special thank you goes to my student gang, with whom I spent many many hours in the lab. It was an honour to be your supervisor, thank you for your help in the lab, for all the private conversations we had and last but not least for showing me how darn patient I can be: Christoph, Julia, Madita, Nate, Nils, Samantha and Yasemin. Emotional support, hikes, strolls and coffee breaks initiated by my fiancé David, my Slovene friends Nina, Blaž, Karin and Ana, and my family are also greatly appreciated.

Curriculum Vitae

Date & place of birth:	4. April 1992 in Maribor, Slovenia
Civil & family status:	in a civil union, without children
Institution:	Department 5: Biology / Chemistry — AG Botany & Developmental Biology of Plants University of Osnabrueck, Barbarastrasse 11 49076 Osnabrueck, Lower Saxony, Germany
Email & Phone:	azerdonercal@uos.de / +49 541 969 2878
Research interest(s):	**evolutionary botany & phycology / phylogenomics & herbariomics**

EDUCATION AND TRAINING

04/2021 — ongoing	**Postdoctoral fellow** at the Ludwig Maximilian University of Munich DFG SPP 1991: New approaches to discovering and naming biodiversity: TaxonOMICS TaxonOMICS of Australian Chenopodiaceae based on llddRAD, hyRAD and HybSeq approaches Mentor: Professor Gudrun Kadereit
09/2017 — 03/2021	**Doctoral position in Plant Biogeography** at the University of Osnabruck & Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben (IPK), including a 6-month-long internship at the Mendel Centre for Plant Genomics and Proteomics
	Dissertation title: Genes documenting history: Biogeographical dynamics of selected Brassicaceae taxa and climate-landscape history of the Eurasian steppe belt (Supervisors: Professor Barbara Neuffer & Professor Herbert Hurka)
04/2017 — 08/2017	**PhD in Phycology** at the Ludwig Maximilian University of Munich (interrupted due to the lack of funding)
	Dissertation title: A tale as old as time: Divergence dating of fossil-rich protist lineage of Dinophyta (Alveolata, SAR, Eukayota)
10/2014 — 02/2017	**MSc in Biology (Evolutionary Botany)** at the Ludwig Maximilian University of Munich; graduated with distinction
	Thesis title: Morphological, phylogenetic and taxonomic analysis of selected *Trachelomonas* (Euglenida) Ehrenb. strains (Supervisor: Professor Marc Gottschling)
10/2011 — 09/2014	**BSc in Biology, BSc in Educational Biology and Educational Chemistry** at the University of Maribor (Slovenia); graduated magna cum laude
	Thesis title: Morphological, physiological and biochemical responses of *Apium repens* (Jacq.) Lag. to different selenite concentrations (Supervisor: Professor Jana Ambrožič Dolinšek)

PUBLICATION RECORD

[18] **Žerdoner Čalasan A**, Hurka H, German DA, Pfanzelt S, Blattner F, Seidl A, Neuffer B. (preprint): Evolutionary history and historical biogeography of the model genus *Capsella* (Brassicaceae) with emphasis on the steppe floral element *Capsella orientalis* Klokov.

[17] Seidl A, Tremetsberger K, Pfanzelt S, Blattner F, Neuffer B, Friesen F, Hurka H, Shmakov A, Oyuntsetseg B, **Žerdoner Čalasan A**, Bernhardt K-G. (2021): The phylogeographic history of *Krascheninnikovia* reflects the development of dry steppes and semi-deserts in Eurasia. Scientific Reports 11: 6645.

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

[15] Friesen N, **Žerdoner Čalasan A**, Neuffer B, German DA, Markov M, Hurka H. (2020): Evolutionary history of the Eurasian steppe plant *Schivereckia podolica* (Brassicaceae) and its close relatives. Flora 268: 151602.

[14] **Žerdoner Čalasan A**, Kretschmann J, Gottschling M. (2020): Integrative taxonomy for sexually-deprived protists in the 21st century: A case study of *Trachelomonas* Ehrenb. (Euglenaceae) from Western Ukraine. Taxon 69: 28–42.

[13] Kretschmann J, **Žerdoner Čalasan A**, Meyer B, Gottschling M. (2019): Zero intercalary plates in *Parvodinium* (Peridiniopsidaceae, Peridiniales) and phylogenetics of *P. elpatiewskyi,* comb. nov. Protist 171: 125700.

[12] **Žerdoner Čalasan A**, Seregin AP, Hurka H, Hofford NP, Neuffer B. (2019): The Eurasian steppe belt in time and space: Phylogeny and historical biogeography of the false flax (*Camelina*, Camelineae, Brassicaceae). Flora 260: 151477.

[11] Gottschling M, Chacón Pinilla J, **Žerdoner Čalasan A**, Kretschmann J, Kühne N, Neuhaus S, John U. (2019): Determination of dinophytes detected in environmental samples from Bavarian lakes (Germany). Freshwater Biology 65: 193—208.

[10] **Žerdoner Čalasan A**, Kretschmann J, Gottschling M. (2019): They are young and they are many: Dating freshwater lineages in unicellular dinophytes. Environmental Microbiology 21: 4125—4135.

[9] Kretschmann J, Owsianny PM, **Žerdoner Čalasan A**, Gottschling M. (2018): The hot spot in a cold environment: Puzzling *Parvodinium* (Peridiniopsidaceae, Peridiniales) from the Polish Tatra Mountains. Protist 169: 206—230.

[8] **Žerdoner Čalasan A**, Kretschmann J, Filipowicz NH, Irimia RE, Gottschling M. (2018): Towards global distribution maps of unicellular organisms such as calcareous dinophytes based on DNA sequence information. Marine Biodiversity 49: 749—758.

[7] Kretschmann J, **Žerdoner Čalasan A**, Gottschling M. (2018): Molecular phylogenetics of dinophytes harbouring diatoms and endosymbionts (Kryptoperidiniaceae, Peridiniales), with evolutionary interpretations and a focus on the identity of *Durinskia oculata* from Prague. Molecular Phylogenetics and Evolution 118: 392—402.

[6] **Žerdoner Čalasan A**, Kretschmann J, Gottschling M. (2018): Absence of co-phylogeny indicated repeated diatom capture in dinophytes hosting a tertiary endosymbiont. Organismic Diversity & Evolution 18: 29—38.

[5] Izquierdo López A, Kretschmann J, **Žerdoner Čalasan A**, Gottschling M. (2017): The many faces of *Peridinium cinctum* (Peridiniaceae, Peridiniales): Morphological and molecular variability in a common dinophyte. European Journal of Phycology 53: 156—165.

[4] Kretschmann J, **Žerdoner Čalasan A**, Kusber WH, Gottschling M. (2017): Still curling after all these years: *Glenodinium apiculatum* Ehrenb. (Peridiniales, Dinophyceae) repeatedly found at its type locality in Berlin (Germany). Systematics and Biodiversity 16: 200—209.

[3] Gottschling M, **Žerdoner Čalasan A**, Kretschmann J, Gu H. (2017): Two new generic names for dinophytes harbouring a diatom as endosymbiont, *Blixaea* and *Unruhdinium* (Kryptoperidiniaceae, Peridiniales). Phytotaxa 306: 296—300.

[2] Gottschling M, Kretschmann J, **Žerdoner Čalasan A.** (2017): Description of Peridiniopsidaceae, fam. nov. (Peridiniales, Dinopyhceae). Phytotaxa 299: 293—296.

[1] Mechora Š, **Žerdoner Čalasan A**, Felicijan M, Urbanek Krajnc A, Ambrožič Dolinšek J. (2016): The impact of selenium treatment on some physiological and antioxidant properties of *Apium repens*. Aquatic Botany 138: 16—23.

CONFERENCE TALKS

02/2019 **They are young, and they are many: Dating freshwater lineages in unicellular dinophytes.** Conference talk at the 20th Annual Meeting of the Society for Biological Systematics.

09/2018 **Dealing with the big guys: Uncovering the floral history of the Eurasian steppe belt.** Talk at the 18th Austrian Botanical Conference & 24th International Symposium on Biodiversity and Evolutionary Biology of the German Botanical Society.

09/2018 **Genes documenting history: Uncovering the phylogeographic patterns of *Camelina* Crantz.** Conference talk at the 48th Annual Meeting of the Ecological Society of Germany Austria and Switzerland

07/2017 **Repeated diatom capture in dinophytes hosting a tertiary endosymbiont (Kryptoperidiniaceae, Peridiniales).** Conference talk at the 11th International conference on modern and fossil dinoflagellates, University of Bordeaux, France.

02/2016 **Could tiny dinos take over big waters of the world? – Spatial distributions of unicellular dinophytes.** Conference talk on the 17th Annual Meeting of the Society for Biological Systematics (Gesellschaft für Biologische Systematik).

09/2014 **The impact of selenium treatment on fitness and antioxidant defense system of *Apium repens* (Jacq.) Lag.** Conference talk at the 6th Slovenian Symposium on Plant Biology (undergraduate session).

INVITED TALKS

03/2021 **Cenozoic evolution of the Eurasian steppe flora (in Slovene).**
Lecture at the Faculty of Natural Sciences and Mathematics, University of Maribor, Slovenia as a part of the course "Biogeography". (Invited by Assoc. Prof. Nina Šajna).

07/2019 **Biogeographical dynamics of selected Brassicaceae taxa and climate-landscape history of the Eurasian steppe.**
Oral presentation at the institute seminar organised by the Institute of Systematic Botany and Mycology, Faculty of Biology, Ludwig Maximilian University of Munich, Germany (Invited by Prof. Susanne Renner).

06/2019 **Morphology, diversity and evolution of dicotyledonous plants with an emphasis on the indigenous flora; 65 million years of plant evolution in 90 minutes (in German).**
Lecture series as a part of a BSc modul "Heimische Biodiversität: Flora und Fauna Mitteleuropas" organised by the student council of the Osnabrueck University

01/2018 **The travels of Marco Polo — botanical edition (in Slovene).**
Lecture at the Faculty of Natural Sciences and Mathematics, University of Maribor, Slovenia as a part of the course "Biogeography". (Invited by Assoc. Prof. Nina Šajna).

05/2017 **Taxonomy for the sexually-deprived protists in the 21st century.**
Oral presentation at the institute seminar organised by the Institute of Systematic Botany and Mycology, Faculty of Biology, Ludwig Maximilian University of Munich, Germany (Invited by Prof. Susanne Renner).

03/2017 **Protistology — exploring the invisible.**
Lecture at the Faculty of Natural Sciences and Mathematics, University of Maribor, Slovenia as a part of the course "Phylogeny and evolution of lower plants" (Invited by Assoc. Prof. Sonja Škornik).

01/2017 **Dinophytes and their diatom endosymbiotic partners.**
Oral presentation at the institute seminar organised by the Institute of Systematic Botany and Mycology, Faculty of Biology, Ludwig Maximilian University of Munich, German (Invited by Prof. Susanne Renner).

04/2016 **Dinoflagellates — an amazing evolutionary experiment.**
Oral presentation at the Faculty of Natural Sciences and Mathematics, University of Maribor, Slovenia, as a part of the course "Phylogeny and evolution of lower plants". (Invited by Assoc. Prof. Sonja Škornik).

TEACHING AND CURATORIAL EXPERIENCE

04/2017 — 08/2017 **Research assistant (Wissenschaftlicher Mitarbeiter)** at *Übung: Artenvielfalt Botanik* course at Ludwig Maximilian University of Munich, Germany (3 SWS).

10/2016 **Student assistant (Wissenschaftliche Hilfskraft)** at *Wahlpflichtmodul 10.2: Botanik und Mykologie* course at Ludwig Maximilian University of Munich, Germany (3 SWS).

08/2016 **Student assistant (Wissenschaftliche Hilfskraft)** at the Bavarian State Collection for Botany (herbarium curatorial work).

04/2016 — 08/2016 **Student assistant (Wissenschaftliche Hilfskraft)** at *Übung: Artenvielfalt Botanik* course at Ludwig Maximilian University of Munich, Germany (3 SWS).

WORKSHOP ATTENDANCE

05/2019 **Physalia courses: Phylogenomics** — carried out at the Free University of Berlin in the frame of Profund Innovation Programme

12/2018 **Genotype-by-sequencing: Data generation and analysis** — carried out at the Leibnitz Institute of Plant Genetics and Crop Plant Research (IPK Gatersleben)

FUNDING AND AWARDS

02/2021	**Feodor Lynen Research Fellowship (Alexander von Humboldt Foundation)** for a postdoctoral research project on phylo- and cytogenomics of Camelineae at the Central European Institute of Technology, Brno, Czech Republic (declined due to COVID-19 pandemic)
07/2019	**Scholarship for a lab visit** to carry out a cytological analysis of a recent allotetraploid species *Capsella thracica* and its progenitor species at the Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben; awarded by German Botanical Society (Deutsche Botanische Gesellschaft)
02/2018	**Scholarship for a RISE Germany DAAD research internship** for an undergraduate student from North America, Great Britain and Ireland for a period of three months; scholarship went to Nathaniel Hofford (Ohio State University, USA)
11/2017	**LMU Forscherpreis für exzellente Studierende** for one of the best Lehre@LMU financed projects, awarded by Ludwig Maximilian University of Munich (LMU München)
11/2017	**Best Master's Thesis Award in Botany** for one of the best Master's Theses, awarded by German Botanical Society (Deutsche Botanische Gesellschaft)
07/2017	**Best Young Scientist Oral Presentation (PhD section)** for the talk at the 11th International conference on modern and fossil dinoflagellates, University of Bordeaux, France
12/2016 — 03/2017	**Lehre@LMU-Förderpreis** for student research projects
12/2015 — 03/2016	**Lehre@LMU-Förderpreis** for student research projects
10/2014	**Graduated magna cum laude** at the study programme Educational Biology & Educational Chemistry for being in top 5% of the entire graduating class
	Graduated magna cum laude at the study programme Biology for being in top 5% of the entire graduating class

PUBLICATION I:

THE EURASIAN STEPPE BELT IN TIME AND SPACE:

PHYLOGENY AND HISTORICAL BIOGEOGRAPHY OF THE FALSE FLAX

(*CAMELINA*, CAMELINEAE, BRASSICACEAE)

ŽERDONER ČALASAN A, SEREGIN AP, HURKA H, HOFFORD NP, NEUFFER B.

FLORA (2019) 260: 151477

DOI: 10.1016/j.flora.2019.151477

The Eurasian steppe belt in time and space: Phylogeny and historical biogeography of the false flax (*Camelina* Crantz, Camelineae, Brassicaceae)

Anže Žerdoner Čalasan[a,*], Alexey P. Seregin[b], Herbert Hurka[a], Nathaniel P. Hofford[c], Barbara Neuffer[a]

[a] Department 5: Biology/Chemistry, Botany, University of Osnabruck, Barbarastraße 11, 49076 Osnabruck, Lower Saxony, Germany

[b] Herbarium, Department of Geobotany, Faculty of Biology, M.V. Lomonosov Moscow State University, 119991 Moscow, Russia

[c] Department of Evolution, Ecology & Organismal Biology, The Ohio State University, Columbus, Ohio 43210, The United States of America

[*] Corresponding author. *E-mail address*: Anze.ZerdonerCalasan@biologie.uni-osnabrueck.de

Abstract

Stretching 8,000 km from the Pannonian basin and the Danube delta in the West to the Manchuria region in the East and reaching up to 1,000 km in width, the Eurasian steppe belt is the vastest steppe region worldwide. However, our knowledge about the temporal and spatial patterns of floral origin and evolution of the Eurasian Steppe is limited and inconclusive. Case studies on typical steppe flora may help us close such gaps. The study subject of this project was *Camelina* – a taxon which occupies open dry habitats in temperate zones of Eurasia. To infer the evolutionary history of this genus, maximum likelihood optimisation in RAxML and Bayesian Inference approach were carried out, based on the nuclear external transcribed spacer region. Furthermore, we performed a secondarily calibrated time estimation analysis using Bayesian optimisation in BEAST to infer potential influence of climatic shifts and paleogeographic events on the distribution patterns of *Camelina* and carried out an ancestral area reconstruction analysis using a Bayesian Binary Method. Our study resulted in a well-supported phylogeny that corresponds with the species morphology and uncovered several genetically distinct inter- and intraspecific lineages which appear to correlate geographically. Time divergence estimation argue for the diversification of *Camelina* to have taken place in the Middle East around the transition from Pliocene to Pleistocene (3-2 mya), and its historical biogeography to have been under a strong influence of several glacial periods and their palaeoclimatic and palaeoenvironmental consequences. Its young age also explains the subtle morphological character differences among species and high interspecific hybridisation potential. We further discuss the rediscovery of wild *Camelina sativa* populations and propose the external transcribed spacer as a ribotype identifying region for young and rapidly evolving core eudicot lineages.

Key words: florogenesis, Eurasian steppe, dated phylogeny, ancestral area reconstruction, *Camelina*, Brassicaceae

1 Introduction

The Eurasian steppe belt stretches an impressive 8000 km from the Danube basin in Romania all the way to northeast China and can be up to 1000 km wide (Lavrenko, 1969). Despite being the largest grassland region in the world, little is known about the florogenesis (= floral origin and evolution) of this belt. It has been argued that the Eurasian steppe developed during the early Miocene around 20 million years ago (mya) in Central Asia (Jiang and Ding, 2009; Strömberg, 2011), but there is accumulating evidence for an earlier origin (Gomes Rodrigues et al., 2012; Hurka et al., 2019; Mai, 1995; Quan et al., 2014). Regardless of the exact age, the onset and development of the Eurasian steppe belt was under strong influence of worldwide Palaeogene cooling trend interrupted by warming periods and development of the Asian monsoon system. Together with the retreat of Paratethys and the uplift of the Tibetan Plateau, early Oligocene disappearance of the Turgai Strait played a pivotal role in the initial aridification of Central Asia (Hurka et al., 2019 and references therein). However, it remains unclear to what extent each of these features contributed to the climate-landscape history of the Eurasian steppes.

Earliest records of steppe climate and vegetation for the East European Plate are from the late Miocene time, less than 9 mya (Bruch et al., 2011; Ivanov et al., 2011; Jacobs et al., 1999; Velichko, 1999). The steppe ecoregion was strongly influenced by past climatic shifts. One of the most prominent transitions was the Last Glacial Maximum followed by Late Glacial climate warming, in which the vegetation went through dramatic changes caused by extensive flora die out (Binney et al., 2017; Frenzel, 1968; Tarasov et al., 2000). Megafaunal extinctions on local and global levels at the end of the late Pleistocene and early Holocene took place (Barnosky, 2004; Bartlett et al., 2016; Fiedel, 2009; Orlova et al., 2004; Owen-Smith, 1987), resulting in a sudden lack of herbivores likely influencing the development of the Eurasians grasslands (Strömberg, 2011). Recent studies argue that the decline of large herbivores induces major alterations in landscape structure and ecosystem function (Bakker et al., 2016; Sandom et al., 2014). Furthermore, during glacial and interglacial cycles, continuous expansions, contractions and latitudinal range shifts of the Eurasian steppe belt have been documented (Frenzel, 1968; reviewed in Hurka et al., 2019) alongside with massive longitudinal range splits of the Eurasian steppe belt. These were caused by transgressive-regressive events of the Caspian Sea and ice damming of the Siberian river systems (Arkhipov et al., 1999; Mamedov, 1997; Svitoch, 2012; Tudryn et al., 2013).

The genetic footprint of typical dry-adapted steppe species mirrors these changes and allows for detailed studies of the Eurasian steppe belt florogenesis in time and space (Franzke et al., 2004; Friesen et al., 2016; Herden et al., 2016; Volkova et al., 2017). Brassicaceae is one of the biggest and highly diverse plant families of the steppes (Bone et al., 2015), whose representatives have diversified mostly in cool and dry environments around the Eocene-Oligocene boundary (Hohmann et al., 2015). The tribus Camelineae DC. has diversified in the late Miocene (Guo et al., 2017) and is one of the most comprehensively studied Brassicaceae tribes, including important model systems such as *Arabidopsis* and *Capsella* and economically important *Camelina* (Koch et al., 2018).

False flax (*Camelina*, Camelineae, Brassicaceae) is currently comprised of nine species—*Camelina alpkoyensis* Yild., *Camelina alyssum* (Mill.) Thell., *Camelina anomala* Boiss. & Hausskn. ex Boiss., *Camelina hispida* Boiss., *Camelina laxa* C.A. Mey., *Camelina microcarpa* Andrz. ex DC., *Camelina neglecta* J.Brock, Mandáková, Lysak & Al-Shehbaz, *Camelina rumelica* Velen. and *Camelina sativa* (L.) Crantz (Al-Shehbaz, 2012; Brock et al., 2019). They are annual or biennial herbs, presently growing mostly in open and disturbed habitats, including pastures, dry grasslands and along roadsides and railways (Francis and Warwick, 2009). While *C. alyssum*, *C. microcarpa* and *C. sativa* have been introduced to the New World, the rest of the species show a more restricted distribution, not expanding further than the Western Asiatic part of the Irano-Turanian floristic region (sensu Takhtajan 1986). The latter region is also speculated to represent the area of origin not only of *Camelina*, but of the whole Brassicaceae family (Hedge, 1976). Interspecific and intergeneric hybridisation is a common phenomenon in *Camelina* (Julié-Galau et al., 2014; Séguin-Swartz et al., 2013) and together with subtle morphological character differences between species and numerous cytotypes, these make the taxonomy of *Camelina* notoriously difficult (Brock et al., 2018). Additionally, several contradictory cytotypes have been reported for this species group (Francis and Warwick, 2009; Koch et al., 2018; Martin et al., 2017), possibly due to a high hybridisation potential which further confounds accurate taxonomic inference.

Fast-evolving rRNA operon regions such as the internal transcribed spacers 1 and 2 (ITS) (Schultz, 2005; Suh et al., 1993) have failed to delimitate the species in question, further preventing the recovery of true evolutionary relationships between the species (Brock et al., 2018). Until now, no study has investigated other nuclear regions that could be useful as a barcode for species level taxonomic identification within the *Camelina* clade. Species delineation and their evolutionary relationships have been tackled successfully in a recent study using a double digest restriction site associated sequencing (ddRADSeq) approach, recovering distinct and highly supported clades that

were congruent with morphologically assigned species names (Brock et al., 2018). However, not all *Camelina* species were taken into account. Had all the known species been included and sampled across their entire geographic range, the topology of the phylogenetic tree might be different.

Identifying plant species with DNA-barcodes is already a well-established methodology that cost effectively promotes biodiversity assessment, conservation, ecological and life history studies of plants (Kress et al., 2005). As such, it is a widely used tool despite possibly being confounded by plant hybridisation, asexual reproduction, genome duplications and incomplete lineage sorting (Fazekas et al., 2009; Stebbins, 1950). Fast and efficient species determination of *Camelina* is of practical nature due to its agricultural implications and hybridisation potential. For example, important oilseed crop species *C. sativa* is limited by lack of certain desirable agronomic, yield and oil quality traits (Iskandarov et al., 2014). Recovery of potential wild genetic resources would thus come in handy to enhance its breeding pool and potentially improve these inhibiting characteristics. However, because of its young age and supposed rapid speciation, finding an informative barcode for *Camelina* is likely a difficult task.

The overall aim of our studies is to place the development of the flora of the Eurasian steppe belt into time and space. To achieve this goal, case studies are needed. Here, we carried out a phylogenetic and historical biogeographic reconstruction of the *Camelina* genus and consequently characterised the spatial and temporal patterns of this group within the florogenesis of the Eurasian steppe belt. Furthermore, we demonstrated the ribotype identification applicability of a well-known nuclear coding region, the external transcribed spacer (ETS), for the *Camelina* group and investigated the congruency between morphological and genetic species delimitation methods.

2 Materials and methods

2.1 Taxon sampling and distribution range surveys

We compiled a taxon sample covering the whole distribution area and all the known species of *Camelina* (also see Discussion). Distribution maps were put together primarily using geographical range information from the literature (Vassilczenko, 1939; Nikiforova, 2004; Meusel et al., 1965) and our own field data, and only secondarily using online databases (Plants of the World Online, http://www.plantsoftheworldonline.org/), whose entries were critically assessed before being included in the survey. Plant material was obtained from specimens deposited in HBG, JE, MW, NSK and OSBU herbaria (Fig. S1). In addition, several accessions collected as seeds during field trips have been sowed and grown in the greenhouse of the University of Osnabruck. In some cases, fresh leaves were collected in the field, dried in silica gel and used directly for DNA isolation (for the taxon list, see Fig. S1).

Currently, the tribe Camelineae encompasses *Arabidopsis* (DC.) Heynh. nom. cons., *Camelina* Crantz, *Capsella* Medik. nom. cons., *Catolobus* (C.A. Mey.) Al-Shehbaz, *Chrysochamela* (Fenzl) Boiss., *Neslia* Desv. nom. cons. and *Pseudoarabidopsis* Al-Shehbaz, O'Kane & R.A. Price (Koch et al., 2018). To reduce potential topological and time estimation artefacts due to a biased taxon sampling, we included at least one representative per higher taxonomic unit (genus) in our analysis. The taxonomic placement of *Noccidium* endemic to Iran has been controversial, as morphological evidence and some molecular studies (Couvreur et al., 2010) have consistently placed it into Coluteocarpaeae (German, 2018). Therefore, it was excluded from our analysis. For the same reason, all but two currently accepted *Camelina* species were used in this study to ensure our analyses were representative and robust. The following *Camelina* species were included in the analyses: *Camelina alyssum* (Mill.) Thell., *Camelina anomala* Boiss. & Hausskn. ex Boiss., *Camelina hispida* Boiss., *Camelina laxa* C.A. Mey., *Camelina microcarpa* Andrz. ex DC., *Camelina rumelica* Velen. and *Camelina sativa* (L.) Crantz.

The identity of *Camelina alpkoyensis* Yıld. (Yıldırımlı, 2011) remained unresolved and was therefore not included in the analysis. The type specimen is currently at the author's disposal only and neither leaf nor seed material is available for molecular analysis (Yıldırımlı, pers. comm. Sept. 19th 2018). The holotype material was supposedly located in the Herbarium of Hacettepe University (Turkey); however, their database does not hold any information of it (Özüdoğru, pers. comm.). Thus, molecular analysis, which would clarify whether this truly is a species or just an ecotype of

Camelina laxa, could not have been carried out. During the course of this study, a new species *Camelina neglecta* J.Brock, Mandáková, Lysak & Al-Shehbaz was described (Brock et al., 2019). Nevertheless, as the origin of the type material is an unspecified seed collection without geo-referenced data deposited at the USDA and since no original voucher was ever recovered, we also excluded this taxon from the present analysis.

Geographical data was recorded at the specimen collection site, recovered directly from the voucher specimens or reconstructed by using TopGlobus (http://www.topglobus.ru/), Wikimapia (http://www.wikimapia.org/) and Loadmap (http://loadmap.net/) based on text information from the voucher specimens. Distribution maps were produced using QGIS (https://qgis.org/en/site/). The Eurasian steppe belt shape included the following geographic units: Altai Steppe and Semi-Desert, Daurian Forest Steppe (NE Mongolia, S Siberia), Eastern Anatolian Montane Steppe, Emin Valley Steppe (China-Kazakhstan border), Kazakh Forest Steppe, Kazakh Steppe, Kazakh Upland, Mongolian-Manchurian Grassland, Pontic Steppe, Sayan Intermontane Steppe, Selenge-Orkhon Forest Steppe (north-central Mongolia) and South Siberian Forest Steppe. Their descriptions were taken from The Nature Conservancy (http://maps.tnc.org/gis_data.html) and WWF For a Living Planet – Temperate Grasslands, Savannas and Shrubland Ecoregions (https://www.worldwildlife.org/biomes/temperate-grasslands-savannas-and-shrublands). The geographic terminology of the East European Plain followed the floristic division of Komarov's U.S.S.R Flora included in the Flora Europaea (as outlined in Tutin et al., 1993).

2.2 Molecular analyses and ancestral area reconstruction

Total genomic DNA was isolated from herbarium specimens using the InnuPREP Plant DNA Kit (Analytic Jena AG, Jena, Germany) according to the instructions of the manufacturer and used directly in PCR amplifications. PCR was carried out using 10 µl of 2X HS Taq Mix Red (Biozym Scientific, Hessisch Oldendorf, Germany), 7 µl of ddH_2O, 1 µl of each 10 µM primer and 1 µl of template DNA. Alternatively, PCR was carried out using 20.8 µl of ddH_2O, 3 µl of 10x Taq Buffer with $MgCl_2$, 2 µl of 10mM dNTP mixture, 1 µl of DMSO, 0.2 µl of 5U Taq Polymerase, 1 µl of corresponding primer, and 1 µl of the template DNA. The analysed ETS region spanned from a conserved sequence segment in the 18S rRNA operon to a semi-conserved 9-bp motif near the 5' end of ETS (Cordesse et al., 1993). Slightly shortened primer 18S-IGS (5'-GAGACAAGCATATGACTACTGGCAGGATC-3'; Baldwin and Markos, 1998) and ETS-9 primer (5'-CATGGGCGTGTGAGTGGTGA-3'; Wright et al., 2003) were used. PCR conditions included

initial denaturation at 95 °C for 5 min, 35 cycles of denaturation at 95 °C for 30 s, primer annealing at 55 °C and extension at 68 °C for 1 min, and additional extension at 68 °C for 10 min. Samples that yielded single bands once run through gel electrophoresis were sent to Microsynth Seqlab (Gottingen, www.microsynth.ch) for purification and sequencing. Sequences from each accession were manually edited in Chromas Lite 2.6.4 (Chromas | Technelysium Pty Ltd) and aligned and manually corrected in AliView v1.24 (Larsson, 2014). In addition to the ETS region, internal transcribed spacer (ITS) and trnL intron, psbA-trnH, ndh-rpL32 and trnQ-rps16 chloroplast regions were also investigated on a reduced taxon sample to infer potential intraspecific variation, using corresponding primers published previously in Blattner (1999), Taberlet et al. (1991), Sang et al. (1997) and Shaw et al. (2007), respectively.

Maximum likelihood tree inference relied on RAxML-HPC v8.2.10 (Stamatakis, 2014) with 1000 ML bootstrap iterations with the default number of distinct rate categories. Tree searches were executed from a random maximum-parsimony tree employing either the GTRCAT or the GTRGAMMA bootstrapping phase model. In addition, Bayesian Inference was carried out using MrBayes v3.2.6 (Ronquist et al., 2012) under GTR+Γ substitution model using the random-addition-sequence method with 10 replicates. Two independent Markov chain Monte Carlo (MCMC) analyses of four chains were run – one cold and three heated. Runs were carried out for 20 million cycles, with parameters and trees sampled every 1,000th cycle including an appropriate burn-in (10%) as inferred from the evaluation of the trace files using Tracer v1.6.0 (Rambaut et al., 2018).

Molecular dating analyses relied on BEAUti & BEAST v1.8.4 (Suchard et al., 2018) and the strict clock model. The use of the strict clock model was justified, after assessing the coefficient of variation in the trace file, which did not exceed 0.3. We used a Coalescent: Constant Size tree prior, the GTR+Γ substitution model, and three independent Monte Carlo Markov chains (MCMC) runs for 20 million generations, with parameters sampled every 10,000th generation. Effective sample sizes (ESS) for all estimated parameters were assessed using Tracer v1.6.0 (Rambaut et al., 2018). TreeAnnotator v1.8.4 (Suchard et al., 2018) was used to discard 10% of the saved trees and annotate the rest of them. Maximum clade credibility tree with median node heights was visualized using FigTree graphical viewer of phylogenetic trees v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). All sequence evolution models used in this study were assessed using the Akaike information criterion (AIC) implemented in the jModelTest2 v2.1.6 (Darriba et al., 2012). All the

analyses were carried out at the CIPRES Science Gateway computing facility (Miller et al., 2010). The aligned matrices are available as *.nex files upon request.

Time estimation analysis was carried out using only one representative per ribotype, reducing the whole dataset to 26 entries. The following two secondary calibration points were used in the analysis: the crown age of tribus Camelineae dated to 8.16 ± 2.0 mya (Hohmann et al., 2015) and the crown group of *Arabidopsis* dated to 5.72 ± 1.40 mya (Hohmann et al., 2015; Koch, 2019; Novikova et al., 2016). To elucidate the historical biogeography of *Camelina*, an ancestral area reconstruction was carried out, using the Bayesian binary MCMC (BBM) method (Ali et al., 2012) implemented in RASP v4.1.0 (Yu et al., 2015). Distribution ranges were divided into four regions: (A) Europe, including Pontic steppe, (B) western and central regional sub centre of the Irano-Turanian region (Léonard, 1988), (C) Kazakh steppe, uplands and East Kazakhstan and (D) Mongolian-Chinese steppe. The BBM was run with a fixed state frequencies model (JC model) including either equal among-site rate variation or gamma distribution (0.001, 200) for two million generations, ten chains each, and two parallel runs. Maximal number of allowed areas was set to two and we sampled every 1000th tree with a final burn-in of 10%. Furthermore, region combinations A and D, as well as B and D were not allowed.

3 Results

3.1 Phylogeny and geographic distribution

In total, 120 sequences were generated and deposited to GenBank under MN120314—MN120430 over the course of the study. The ETS alignment encompassed 122 entries, was 431 bp long and comprised 280 constant characters, 52 variable characters and 99 parsimoniously informative characters. Fig.S2 shows the best-scoring Maximum Likelihood (ML, LBS values) tree (–ln=1835.01) carried out by employing the GTRCAT bootstrapping phase model and the Bayesian tree (BI, BPP values). Both trees show a well resolved backbone and congruent topologies, but not fully resolved internal topology. This is the first well-supported phylogeny of Camelineae that includes all the representatives of the tribus. *Arabidopsis* (100LBS, 1.00BPP), *Capsella* (100LBS, 1.00BPP) and *Camelina* (97LBS, 1.00BPP) taxa were recovered as monophyletic. *Neslia* was placed as the sister group to *Camelina* (99LBS, 1.00BPP) and *Catolobus* tentatively as the sister group to *Capsella* (80LBS, .94BPP). The exact position of *Chrysochamela* remained unclear.

Figure 1: Habitus of *Camelina microcarpa*. A – Leaves (Tatarstan, Bavly; May 31st 2010; by E. Izmailov), B – inflorescence (Tatarstan, Bavly; May 31st 2010; by E. Izmailov), C – close up of an inflorescence (Krasnodar, Novorossiysk, Markotkh Ridge; May 16th 2016; by A. Malykhina), D – dried silicles (Krasnodar, Anapa, Utrish; May 10th 2016; by S. Banketov), E – seeds (Krasnodar, Anapa, Utrish; May 10th 2016; by S. Banketov).

Camelina laxa was ranked at the base of *Camelina* group (97LBS, 1.00BPP). Internal relationships of *C. sativa* ribotypes (R1 and R2), *C. alyssum*, *C. anomala* and *C. hispida* remained unclear, albeit moderately supported and grouped according to the morphology. Eastern *C. microcarpa* ribotypes (E1 and E2) were placed as sister ribotypes (94LBS, .91BPP) to a combined *C. microcarpa* Western ribotypes (W1 and W2) and *C. rumelica* (98LBS, 1.00BPP). Further subgrouping within Eastern *C. microcarpa* ribotype as well as within Western *C. microcarpa* ribotype was recognised, albeit with a

very low statistical support (66LBS, .57BPP for the eastern ribotypes and 61LBS, .00BPP for the western ribotypes). No delimitating names have been assigned to the *Camelina microcarpa* and *Camelina sativa* ribotypes. The investigated internal transcribed spacer (Fig. S3) as well as the chloroplast regions (data not shown) displayed little to no variation in *Camelina* and were therefore not used for further studies.

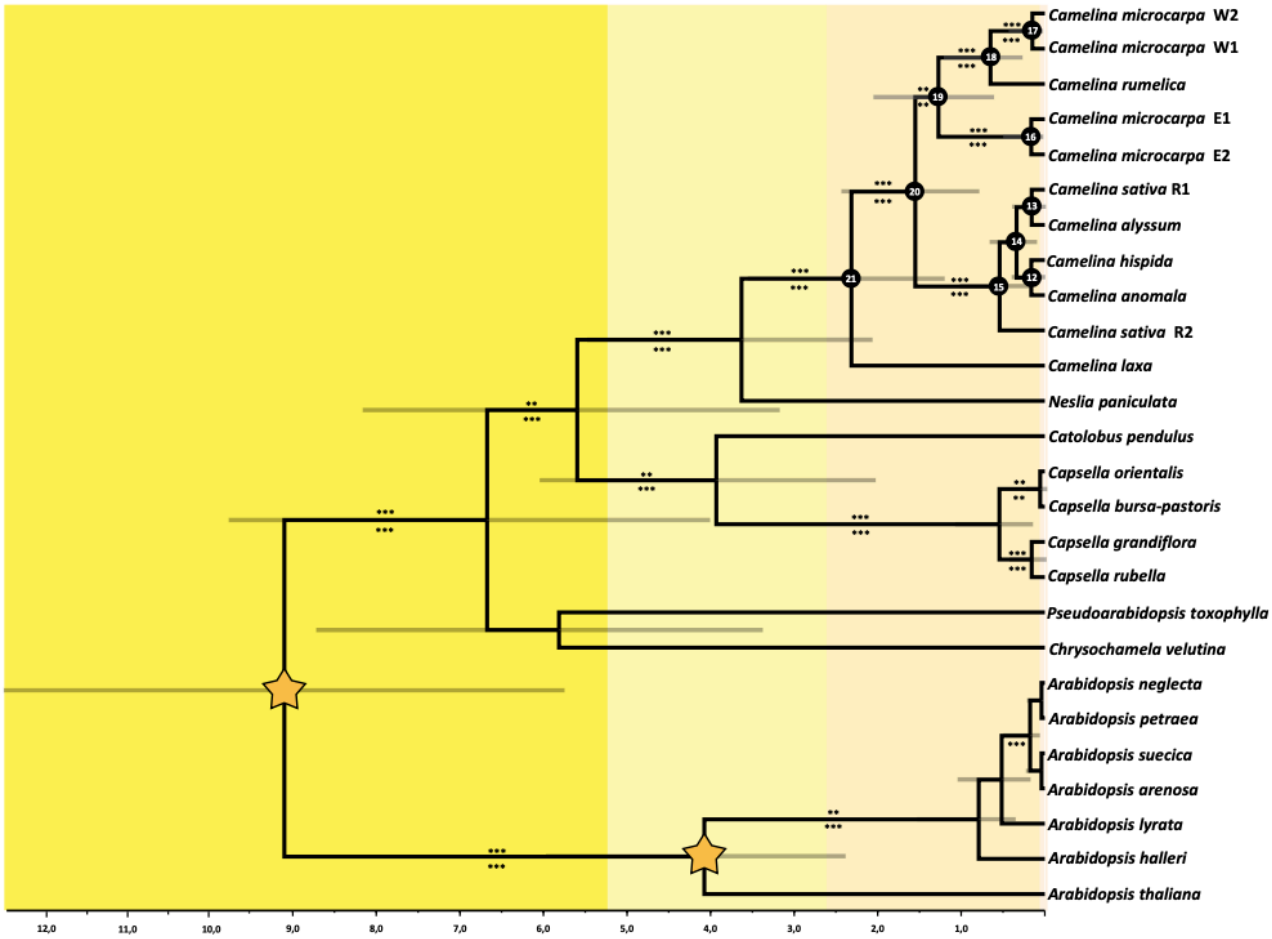


Figure 2: Dated phylogeny of Camelineae. Median rate is given in units of substitutions per million years (including 95% confidence intervals). Absolute ages are in millions of years, and epochs are indicated in the same colours as in (Gradstein et al., 2012). The numbers on the branches are statistical support values (above: ML bootstrap values, values <75 are not shown; below: Bayesian posterior probabilities, values <.90 are not shown). Significance levels: highly supported (***): BI ≥ 0.98, BS ≥ 95%; well supported (**): BI ≥ 0.90, BS ≥ 85% to < 95%; supported (*): BI ≥ 0.90, BS ≥ 75% to < 85%. Orange stars indicate the calibration points.

On the species level, ribotypes and morphological determinations recovered highly congruent results. Morphological determinations were carried out based on the inflorescence structure, fruit size and shape, life strategy, and stem hairiness, following Al-Shehbaz (1987), Brock et al. (2019) and Mirek (1984). *Camelina rumelica* was not only genetically, but also morphologically well separated from the rest of *Camelina* species. Its stem is namely covered in numerous long simple trichomes and its fruits are more densely spaced in the upper than in the lower part of the

infructescence. Life strategy and size of the fruits separated the *C. microcarpa* (Fig. 1) from *C. sativa-alyssum*. While *C. sativa* and *C. alyssum* are the only summer annual species and have (probably through human induced directional selection) developed significantly bigger fruits and seeds, *C. microcarpa* exhibits a winter annual life strategy and produces significantly smaller fruits and seeds. *Camelina sativa* and *C. alyssum* were told apart based on the stem hairs. While *C. sativa* builds mostly stellate trichomes, stems of *C. alyssum* are covered in short simple non-branching trichomes. *Camelina anomala*, *C. hispida* and *C. laxa* were the only three morphologically distinct species. *Camelina anomala* possesses simple stiff hairs only on the lower stem parts and linear-cylindrical siliques with seeds arranged in one row only, whilst all the other *Camelina* species build pyriform fruits with two-rowed seeds. *Camelina hispida* has pubescent middle stems and inflorescences and *C. laxa* exhibits heavily flexuous infructescences. *Camelina hispida* and *C. laxa* are also self-incompatible, while all the other *Camelina* species are self-compatible. The self-compatibility of *C. anomala* is currently unknown.

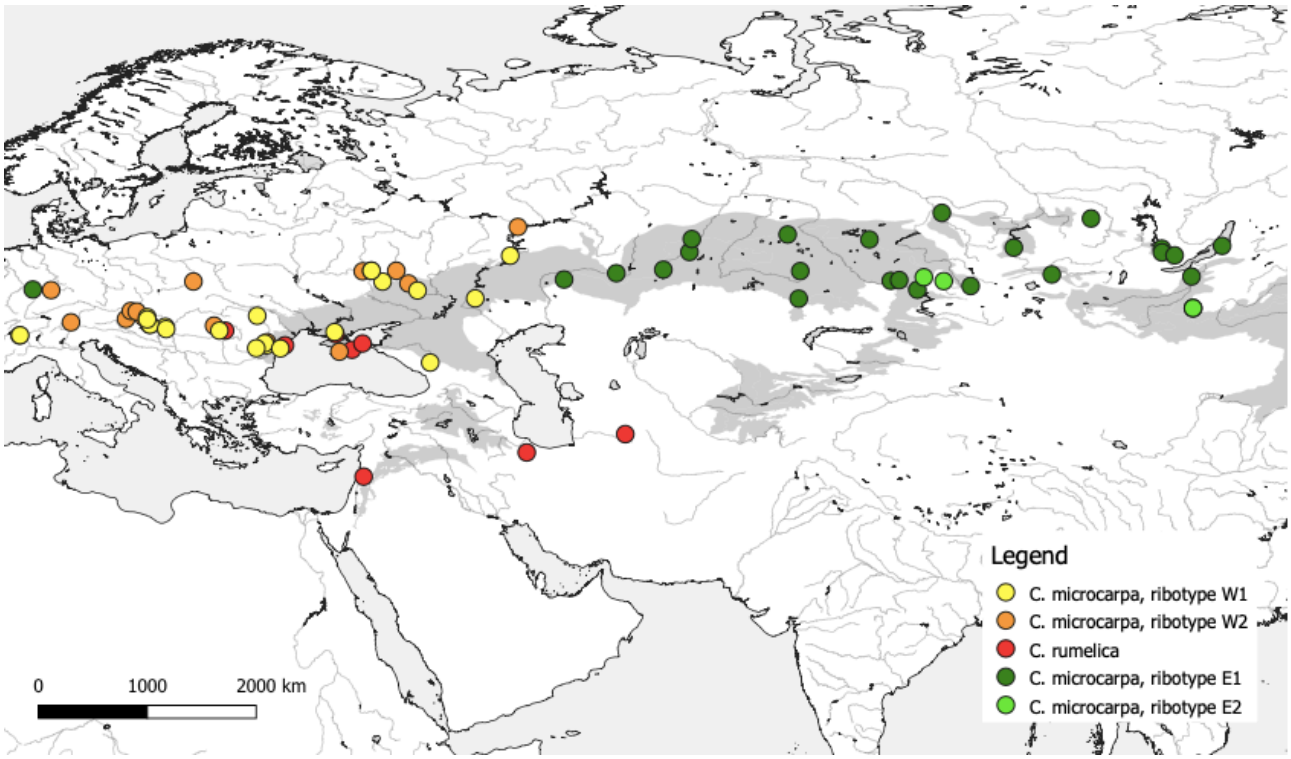


Figure 3: Distribution map of *Camelina rumelica* (ribotype indicated in red) and *Camelina microcarpa* (western ribotype 1 (W1) indicated in yellow, western ribotype 2 (W2) indicated in orange, eastern ribotype 1 (E1) indicated in dark green and eastern ribotype 2 (E2) indicated in light green).

Camelina microcarpa can be found across the whole Eurasian Steppe belt and is the most widely distributed species of *Camelina* in Eurasia (Fig. 3). A similar geographic area was observed for *Camelina sativa*, however its natural distribution range remained unclear. Our surveys found *C.*

sativa across the whole Eurasian steppe belt, with a gap in the Central Kazakh steppe (Fig. 4). Within *C. sativa* two ribotypes have been recognised, where one ribotype was discovered only eastwards from the Pontic steppe region. In the high altitude steppes of the Emin Valley (Kazakhstan-China border) both ribotypes co-occur nowadays. The distribution area of *Camelina rumelica* is more restricted, and encompasses southern parts of the Pontic steppe, lowlands of South-Eastern Europe, Middle East steppes and Western Kazakh steppes (Fig. 3). *Camelina anomala*, *Camelina hispida* and *Camelina laxa* grow only in Central Anatolian steppes and Middle East steppes (data not shown). *Camelina alyssum* was once found across most of Europe and Western Asia but is nowadays treated as extinct in Western and Northern Europe. However, it is still present, albeit scarcely, in the East European Plain.

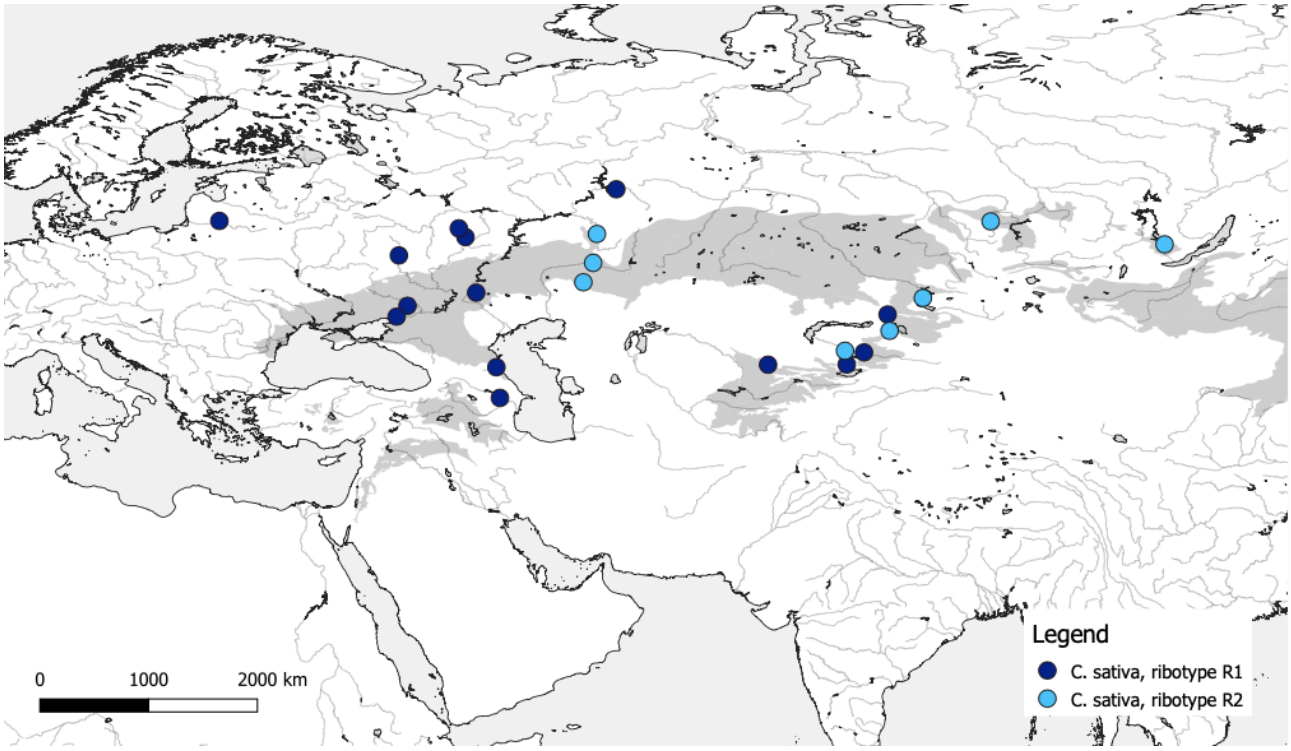


Figure 4: Distribution map of *Camelina sativa* (ribotype R1 indicated in dark blue and ribotype R2 indicated in light blue).

3.2 Time estimation analysis and historical biogeography

The overall topology of the phylogenetic tree and its rooting was in correspondence with other partial and whole plastome studies that have analysed phylogenetic and temporal aspects of Brassicaceae evolution (Guo et al., 2017; Hohmann et al., 2015). Time estimation analysis (Fig. 2) generated a highly congruent topology (compared to Maximum Likelihood and Bayesian Inference analyses), with ESS values exceeding 4000. It placed *Camelina* diversification on the Pliocene/

Pleistocene border. The first major split in *Camelina* occurred at 1.5 ± 1 mya, resulting in a *C. microcarpa-rumelica* lineage and *C. sativa-alyssum* lineage including other species with highly restricted distribution area. The east-west split in the *C. microcarpa-rumelica* lineage was dated to 1.2 ± 1 mya and the split between *C. microcarpa* and *C. rumelica* first at 0.75 ± 1 mya. Irrefutable conclusions on *C. sativa-alyssum* lineage cannot be drawn due to a modest statistical support.

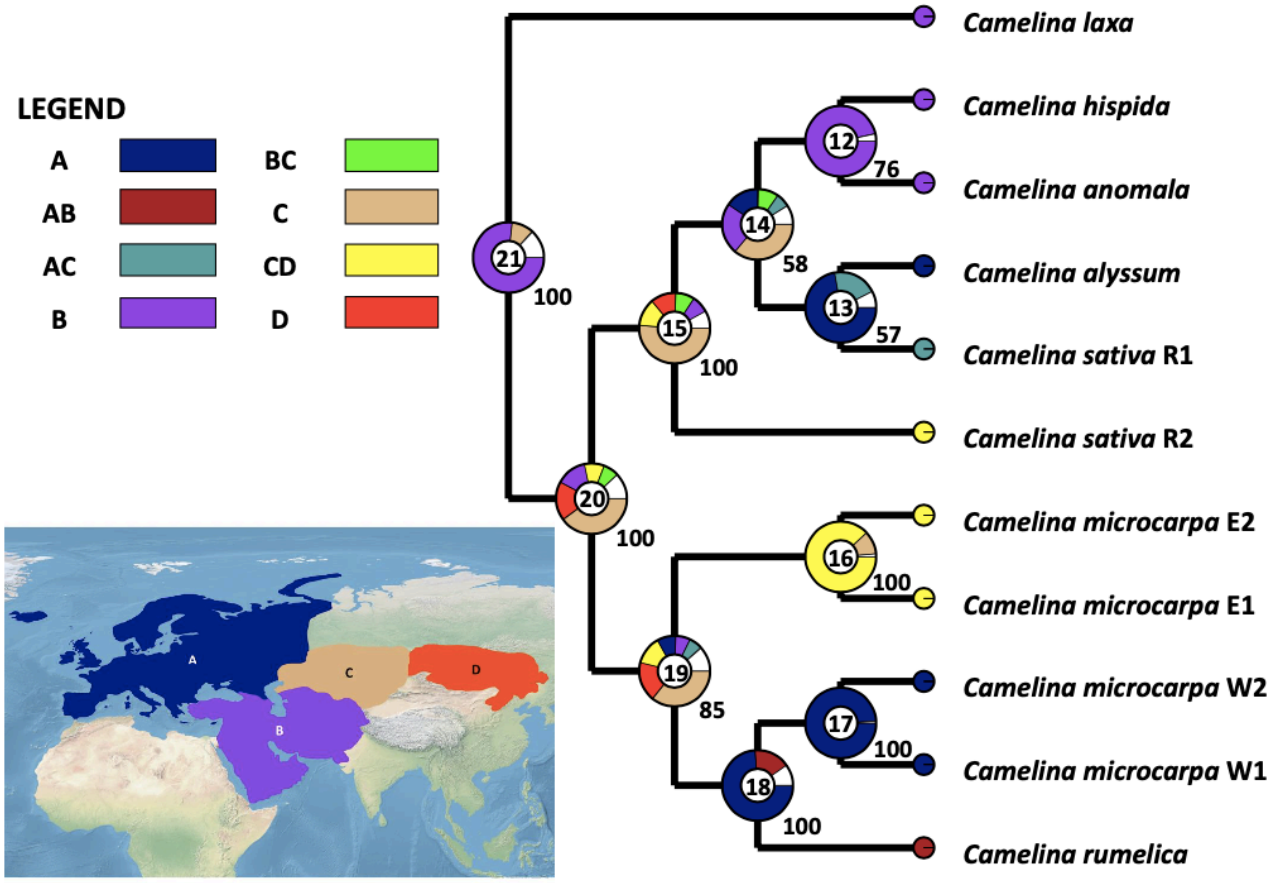


Figure 5: Ancestral area reconstruction of *Camelina* using the Bayesian binary MCMC (BBM) method, based on the time phylogeny derived from the BEAST analysis. The areas are coded as follows: (A) Europe, including Pontic steppe, (B) western and central regional sub centre of the Irano-Turanian region (Léonard, 1988), (C) Kazakh steppe, uplands and East Kazakhstan and (D) Mongolian-Chinese steppe. Unresolved parts are indicated in white. Numbers next to the nodes represent node frequency (%).

Due to a smaller size of reconstructed areas and moderate statistical support on deeper nodes within *Camelina*, a Bayesian Binary Method (BBM) was used for ancestral area reconstruction, which accepts polytomies and tends to suggest single distribution areas for ancestral nodes more often than other methods (Müller et al., 2015). Both BBM models JC and JC+G generated highly congruent results (data not shown). The inferred ancestral area of *Camelina* was shown to be the western and central regional sub centre of the Irano-Turanian region (Fig. 5: node 21), from which

both *Camelina* lineages spread across the Kazakh steppes and uplands (Fig. 5: node 20). Eastern ribotypes of *C. microcarpa* evolved east of the Caspian Sea and potentially spread eastwards (Fig. 5: node 19 → 16). While western ribotypes of *C. microcarpa* and *C. rumelica* lineage never re-occupied the Irano-Turanian region when migrating from east towards west (Fig. 5: node 19 → 18), *C. hispida* and *C. anomala* possibly returned to this floristic region (proportional increase of B in Fig. 5: nodes 15 → 14 → 12). *Camelina rumelica* plausibly developed in the Pontic steppe and later on occupied the Irano-Turanian floristic region (Fig. 5: node 18). Ribotype 2 of *C. sativa* did not migrate with the rest of species and rather stayed in the eastern steppes (east of Caspian Sea; Fig. 5: node 15). Contrary, *C. alyssum* and ribotype 1 of *C. sativa* migrated to the west (Fig. 5: node 14 → 13). For details on dispersal, vicariance and/or extinction events refer to the Fig.S5.

4 Discussion

The use of molecular markers has greatly enhanced our knowledge on past vegetation patterns, including small, sub-regions of the Eurasian steppe belt (Kajtoch et al., 2016; Meindl et al., 2016; Meng et al., 2015; Plenk et al., 2017; Qiu et al., 2011). However, only a handful of studies (Franzke et al., 2004; Friesen et al., 2016; Hurka et al., 2012; Volkova et al., 2017) have focused on the Eurasian steppe belt as a whole. In our study, we put the biogeographical history of *Camelina* into time and space and linked it with past climate-landscape features (Fig. 6). Furthermore, we present the first comprehensive phylogeographic study of *Camelina* encompassing the whole documented distribution ranges of all species. The only potential sampling artefact could be the lack of *C. sativa* accessions from the Central Kazakh steppe. *Camelina sativa* and *C. alyssum* are namely the only two *Camelina* summer annuals (Al-Shehbaz, 1987). As most of our fieldwork in the Kazakh steppe was carried out in early summer, these individuals could have escaped our notice.

4.1 Phylogenetic Analysis of the Genus *Camelina*

The tribus Camelineae was constrained secondarily with a node by Hohmann et al. (2015), where the tribus was represented using only three different genera – *Arabidopsis*, *Camelina* and *Capsella*. To justify this step, an additional time estimation analysis was carried out including only the taxa used in the Hohmann et al. (2015) study (Fig. S4). Topologies including time spans of both analyses were compared and no significant changes in time constraints or topologies were detected. Therefore, using the same node to date the whole Camelineae tribe seemed to be justified. Furthermore, crown group age time spans and time spans of speciation events in *Capsella* and *Arabidopsis* corresponded with previous independently published time span ages: crown age of *Capsella* dated to be up to approximately 1 mya (Douglas et al., 2015), the diversification of *Arabidopsis* to have taken place approximately 6 mya (Novikova et al., 2018), split between *Arabidopsis* and sister clade including *Capsella* dated to approximately 9 mya (Koch et al., 2000) and the Pleistocene origin of *Capsella rubella* and *Capsella grandiflora* (Hurka et al., 2012). Thus, despite using only secondarily calibrated nodes resulting in wider 95% confidence intervals, our results are supported by different independent studies that used different calibration methods.

Time estimation analysis and ancestral area reconstruction placed *Camelina* diversification at approximately 2.5 ± 1 mya on the Pliocene/Pleistocene border into the western and central regional sub-centre of the Irano-Turanian region. This time frame coincides with a global decline in temperatures and humidity (Yang and Ding, 2010), resulting in a cooler and dryer climate that

promoted grassland expansion at the cost of forest decline. As a mostly continuous Eurasian steppe belt already existed on the Miocene/Pliocene border (Mai, 1995; Velichko, 1999), *C. alyssum*, *C. microcarpa*, *C. rumelica* and *C. sativa* consecutively might have spread across an already well-established steppe belt. According to our time estimation analysis, *Camelina* is a young genus, which explains the high levels of hybridisation potential and minute morphological character differences between species. Our study uncovers several ribotypes within *Camelina* group, indicating that despite minute morphological character differences even between species, intraspecific variation (at least in the ETS region) in *Camelina* is present.

4.2. Historical Biogeography of the Genus *Camelina*

4.2.1 *Camelina microcarpa*

The split between *C. microcarpa* ribotypes E1 and E2 and joined *C. microcarpa* ribotypes W1 and W2, and *C. rumelica* is dated to 1.2 ± 1 mya, which coincides with the short-lived Apsheronian transgression of the Caspian Sea (Kroonenberg et al., 1997; Svitoch, 2012; Tudryn et al., 2013). The last common ancestor of these taxa also occupied the Kazakh steppes, which furthermore supports the potential influence of the Caspian Sea transgression. This transgression was a result of a massive freshwater influx caused by the rise of temperatures and a consequent meltdown of glaciers. At its peak, Caspian Sea mean sea level was approximately 100 m higher than nowadays, covering the modern Volga–Kama catchment (Starobogatov, 1994; Van Baak et al., 2013). The western parts of the Pontic–Caspian steppe were not affected by this transgression nor by other consequential events of the glacial periods, and thus provided a refugium for steppe vegetation (Stewart et al., 2010; Varga, 2010).

Within each *C. microcarpa* ribotype subgroups, two additional ribotypes can be detected. However, the statistical support is modest. The two eastern ribotypes (coded in yellow and orange in Fig. 3) and the two western *C. microcarpa* ribotypes (coded in dark and light green in Fig. 3) display only a weak geographical correlation. As the statistical support does not allow for irrefutable conclusions, we can only speculate that the weak geographical correlation of these two ribotypes could be related to human activity. Human influence might also explain that an outlier 12-0076-10-00 has a distinct E1 Kazakh ribotype, but is found in the German state of Rhineland-Palatinate in the Mainz Sand Dunes (dark green point in the left corner of Fig.2). Another explanation could be the dispersal mechanism of *C. microcarpa*. Wet seeds namely produce a thick mucilage coat, which promotes adherence ability and hence long-distance dispersal by either birds or cattle. An alternative

explanation, however, would be that the distribution of western ribotypes is the result of extensive human influence. If this is the case, the Mainz outlier could represent a leftover of a once widely distributed E1 (eastern) ribotype. The Mainz Sand Dunes are namely home to a number of steppe elements (i. e. *Gypsophila fastigiata*, *Onosma arenaria*, *Stipa capillata*), whose distribution, with the exception of Mainz Sand Dunes, do not reach westwards of the Pannonian Basin (Hecker, 1987).

4.2.2 *Camelina rumelica*

Camelina rumelica separated from western ribotypes of *C. microcarpa* approximately 0.75 ± 1 mya. This time span is known for intensified development of permafrost, loess and glacial sheets and consequential radical transformation of the zonal landscape structure. The Baku transgression of the Caspian Sea also belongs to this era (Hurka et al., 2019). The native distribution range of *C. rumelica* encompasses the southern parts of the East European Plain, Central Asia and Middle East, however it has also been introduced and naturalised in Central and western Europe, and in the United States (Plants of the World Online, http://www.plantsoftheworldonline.org/). We argue that this species developed in the Pontic-Caspian steppe in the mid-upper Pleistocene and expanded its distribution area westwards through southern East European Plain and southwards through Asia Minor. Such a scenario likely occurred during only recent glacial periods that favoured dry-adapted steppe species, which would also explain the lack of ribotype differentiation. Despite our extensive field trips and herbarium surveys, no *C. rumelica* plants were discovered east of the 70th meridian. Furthermore, there was also no distribution data on *C. rumelica* that would indicate its presence east of the abovementioned meridian. This together with our ancestral area reconstruction further supports our argument that *C. rumelica* conceivably developed in the western parts of the Eurasian steppe belt and stepwise moved (south)-east and west from the Pontic steppe.

4.2.3 *Camelina sativa* and *Camelina alyssum*

Contrarily to previous assumptions (Brock et al., 2018), populations of *Camelina sativa* have been reported from the grasslands of the East European Plain, the Caucasus, the coastline of Caspian Sea, south-eastern parts of Central Asia (Emin Valley; China-Kazakhstan border) and southern Siberia. It remains unclear, however, if those populations are of anthropogenic origin or indeed reflect the species' natural distribution range. With an exception of Emin Valley, all the other regions are well known for flax production in the past (Vavilov, 1992), thus it is possible that most of the investigated populations are the remnants of the former cultivated flax (*Linum usitatissimum*) fields.

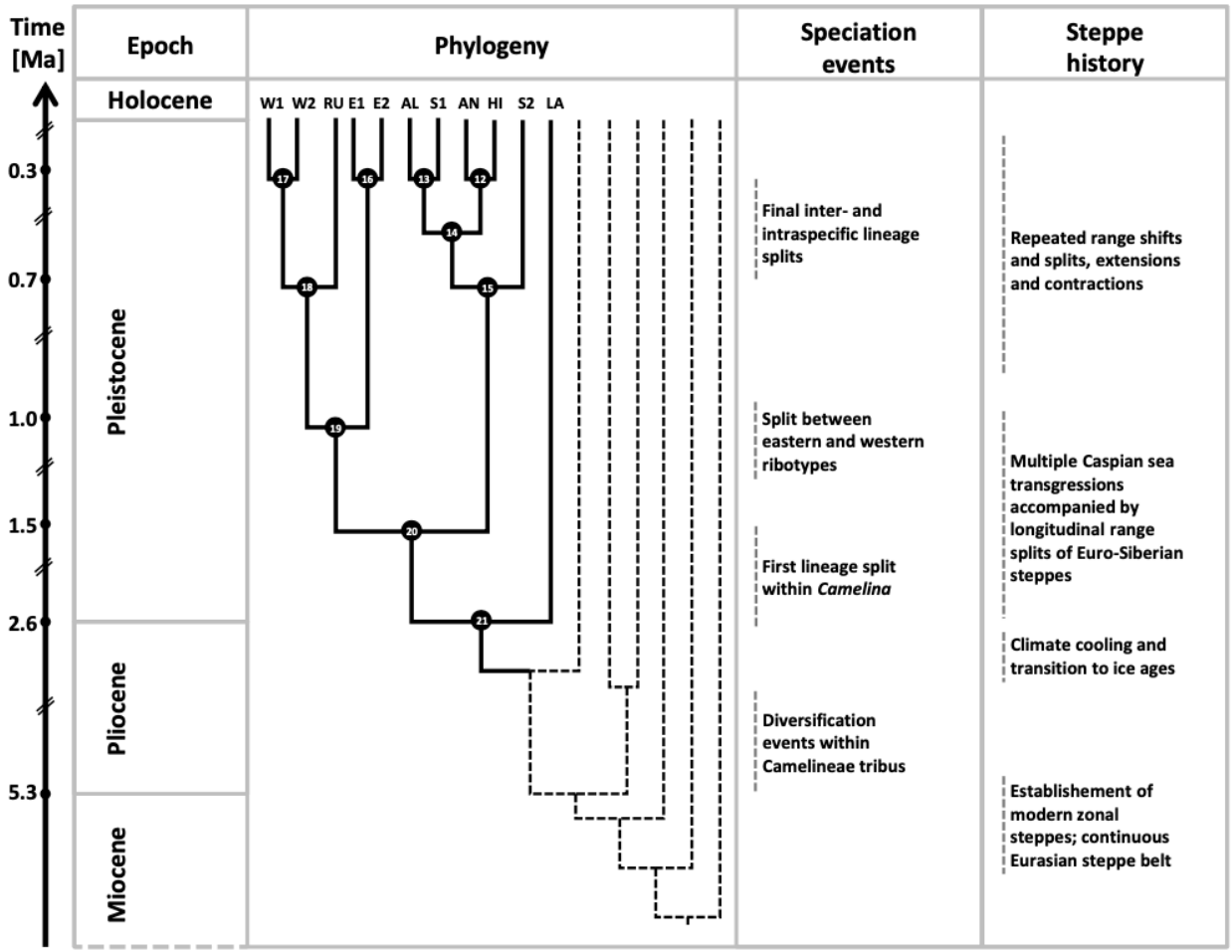


Figure 6: Outline of the evolutionary history of Camelineae. W1 (western *C. microcarpa* ribotype 1), W2 (western *C. microcarpa* ribotype 2), RU (*C. rumelica*), E1 (eastern *C. microcarpa* ribotype 1), E2 (eastern *C. microcarpa* ribotype 2), AL (*C. alyssum*), S1 (*C. sativa* ribotype 1), AN (*C. anomala*), HI (*C. hispida*), S2 (*C. sativa* ribotype 2), LA (*C. laxa*). Taxa indicated with interrupted lines are *Neslia*, *Catolobus*, *Capsella*, *Pseudoarabidopsis*, *Chrysochamela* and *Arabidopsis* (from left to right). Geological epochs according to the International Commission on Stratigraphy (Cohen et al., 2018).

It has been stated that *C. sativa* is the ancestor of *C. alyssum* which has developed in flax fields as a result of directional selection, causing morphological changes in the plant's habitus, fruit morphology and ripening type to resemble flax (Barrett, 1983). Thus, it also shares the same distribution dynamics as flax (Stebbins, 1950; Vavilov, 1992). This is supported by our topology, previous crossing experiments and detailed morphological studies that all point towards a close relationship between *C. sativa* and *C. alyssum* (Sinskaia and Bezluzheva, 1931; Stebbins, 1950; Tedin, 1925; Zinger, 1909). With the abandonment of flax fields, the distribution range of *C. alyssum* also shrank and the species is nowadays considered extinct from most of Central Europe (Francis and Warwick, 2009). Both *C. sativa* ribotypes that were uncovered show a moderate east-west geographical correlation. Similar delineation into two groups has been observed recently, albeit with a different geographical correlation (Luo et al., 2019). The results are, however,

incomparable as their study focused on a different geographic region and used a different set of markers.

The overall geographic signal of *Camelina* has not been overshadowed by the human influence, despite being irrefutably proven to be extensively used already in the Neolithic as an oil source (Hovsepyan and Willcox, 2008). However, it remains unclear which species exactly were used by humans in the Neolithic, if *Camelina* has just been gathered or actively cultivated, and if the latter, when and where the cultivation started (Brock et al., 2018; Hovsepyan and Willcox, 2008).

4.3 Species' Determination and Utility of Molecular Markers for the Genus *Camelina*

Morphological determinations were highly congruent with the ETS sequences; hence the external transcribed region functions as a barcode for *Camelina* at least on a species level. Furthermore, we have uncovered at least two different ribotypes within *C. microcarpa* and at least two within *C. sativa* while amplifying this only 430 bp-long-region. This finding provides a powerful tool for further studies to undoubtedly identify the investigated *Camelina* species on a molecular basis, which is of great importance, as *Camelina* species are hard to delimitate based on morphology only. This locus may even work to delimitate species within other young, morphologically similar and highly plastic species groups, as the flanking regions for primers are highly conserved at least throughout the whole core eudicots (Baldwin and Markos, 1998; Wright et al., 2003).

5 Conclusion

Temporal and spatial aspects of florogenesis within the Eurasian steppe belt are poorly known, despite their great importance in climate-landscape reconstructions. We have uncovered hidden genetic diversity of *Camelina*, which harbours an important species for the agronomic oil industry – *Camelina sativa*. We have shown that phylogeny, differentiation and range evolution of *Camelina* indeed mirrors the climate/landscape dynamics of the steppe (Fig. 6) (Hurka et al., 2019). Furthermore, we put the development of the group into time and space and found a likely correlation between palaeogeographical and palaeoenvironmental events and the evolutionary history of *Camelina*. Current taxonomic treatments of *Camelina* have been confirmed using the ETS locus, however, only on the species level. Genetic dissimilarities of subspecies and their names were incongruent and their nomenclature remains to be completed. We trust that using the ETS will facilitate an easy and reliable species determination, possibly not limited to *Camelina* only. Despite sharing history with humans, *Camelina* has developed on its own prior to its use in Neolithic period, however many aspects of the domestication are still yet to be answered. Our approach proved successful at uncovering another missing point-of-view of florogenesis within the Eurasians steppe belt.

6 Acknowledgements

We would like to thank Simon Pfanzelt to kindly provide us with the Eurasian Steppe belt shape file and all the curators of the HBG, JE and NSK herbaria, who kindly provided the material for sequencing. We also thank Juliana Chacón, who helped us with the ancestral area reconstruction. This work was supported by the Deutsche Forschungsgemeinschaft [grant to BN NE 314/15-1] and the German Federal Foreign Office (via DAAD Rise Germany programme) that supported NPH during his stay in Germany. Furthermore, we would like to acknowledge current digitisation activity in the Moscow University Herbarium [under Government order #AAAA-A16-116021660039-1 headed by APS]. We would also like to thank the 'Plantarium' web-community members, who allowed us to use their photographs: Sergey Banketov, Elvir Izmailov and Anna Malykhina; and Nikolai Friesen, who contacted the photographers on our behalf.

Compliance with ethical standards

Conflict of interest

The authors declare no conflict of interest.

Ethical approval

This article does not contain any studies with animals carried out by any of the authors.

Sampling and field studies

The study was performed in compliance with the Convention on Biological Diversity (CBD).

7 References

Ali, S.S., Yu, Y., Pfosser, M., Wetschnig, W., 2012. Inferences of biogeographical histories within subfamily Hyacinthoideae using S-DIVA and Bayesian binary MCMC analysis implemented in RASP (Reconstruct Ancestral State in Phylogenies). Ann. Bot. 109, 95–107.

Al-Shehbaz, I.A., 2012. A generic and tribal synopsis of the Brassicaceae (Cruciferae). Taxon 61, 931–954.

Al-Shehbaz, I.A., 1987. *Camelina*. J. Arnold Arbor. 68, 234–240.

Arkhipov, S.A., Volkova, V.S., Zolnikov, I.D., Zyklina, V.S., Krukover, A.A., Kulkova, I.A., 1999. West Siberia, in: Climate and environment changes during the last 65 million years; Cenozoic: from Paleocene to Holocene. GEOS, Moscow, pp. 84–109.

Bakker, E.S., Gill, J.L., Johnson, C.N., Vera, F.W.M., Sandom, C.J., Asner, G.P., Svenning, J.-C., 2016. Combining paleo-data and modern exclosure experiments to assess the impact of megafauna extinctions on woody vegetation. Proc. Natl. Acad. Sci. 113, 847–855.

Baldwin, B.G., Markos, S., 1998. Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: congruence of ETS and ITS trees of *Calycadenia* (Compositae). Mol. Phylogenet. Evol. 10, 449–463.

Barnosky, A.D., 2004. Assessing the causes of late Pleistocene extinctions on the continents. Science 306, 70–75.

Barrett, S.H., 1983. Crop mimicry in weeds. Econ. Bot. 37, 255–282.

Bartlett, L.J., Williams, D.R., Prescott, G.W., Balmford, A., Green, R.E., Eriksson, A., Valdes, P.J., Singarayer, J.S., Manica, A., 2016. Robustness despite uncertainty: regional climate data reveal the dominant role of humans in explaining global extinctions of Late Quaternary megafauna. Ecography 39, 152–161.

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J.O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J., Novenko, E., Ryabogina, N., Solovieva, N., Willis, K., Zernitskaya, V., 2017. Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. Quat. Sci. Rev. 157, 80–97.

Blattner, F.R., 1999. Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. BioTechniques 27, 1180–1186.

Bone, M., Johnson, D., Panayoti, K., Kintgen, M., Vickerman, L.G., 2015. Steppes: the plants and ecology of the world's semi-arid regions. Timber Press, Portland, Oregon.

Brock, J.R., Dönmez, A.A., Beilstein, M.A., Olsen, K.M., 2018. Phylogenetics of *Camelina* Crantz. (Brassicaceae) and insights on the origin of gold-of-pleasure (*Camelina sativa*). Mol. Phylogenet. Evol. 127, 834–842.

Brock, J.R., Mandáková, T., Lysak, M.A., Al-Shehbaz, I.A., 2019. *Camelina neglecta* (Brassicaceae, Camelineae), a new diploid species from Europe. PhytoKeys 115, 51–57.

Bruch, A.A., Utescher, T., Mosbrugger, V., 2011. Precipitation patterns in the Miocene of Central Europe and the development of continentality. Palaeogeogr. Palaeoclimatol. Palaeoecol. 304, 202–211.

Brummitt, R.K., Powell, C.E. (Eds.), 1992. Authors of plant names: a list of authors of scientific names of plants, with recommended standard forms of their names, including abbreviations. Royal Botanic Gardens, Kew, London.

Cohen, K.M., Harper, D.A.T., Gibbard, P.L., 2018. ICS International Chronostratigraphic Chart 2019/02. International Commission on Stratigraphy, IUGS. www.stratigraphy.org (accessed last on 2 February 2019).

Cordesse, F., Cooke, R., Tremousaygue, D., Grellet, F., Delseny, M., 1993. Fine structure and evolution of the rDNA intergenic spacer in rice and other cereals. J. Mol. Evol. 36, 369–379.

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). Mol. Biol. Evol. 27, 55–71.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772–772.

Douglas, G.M., Gos, G., Steige, K.A., Salcedo, A., Holm, K., Josephs, E.B., Arunkumar, R., Ågren, J.A., Hazzouri, K.M., Wang, W., Platts, A.E., Williamson, R.J., Neuffer, B., Lascoux, M., Slotte, T., Wright, S.I., 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. Proc. Natl. Acad. Sci. 112, 2806–2811.

Fazekas, A.J., Kesanakurti, P.R., Burgess, K.S., Percy, D.M., Graham, S.W., Barrett, S.C.H., Newmaster, S.G., Hajibabaei, M., Husband, B.C., 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Mol. Ecol. Resour. 9, 130–139.

Fiedel, S., 2009. Sudden deaths: The chronology of terminal Pleistocene megafaunal extinction, in: Haynes, G. (Ed.), American megafaunal extinctions at the end of the Pleistocene. Springer, Netherlands, Dordrecht, pp. 21–37.

Foxe, J.P., Slotte, T., Stahl, E.A., Neuffer, B., Hurka, H., Wright, S.I., 2009. Recent speciation associated with the evolution of selfing in *Capsella*. Proc. Natl. Acad. Sci. 106, 5241–5245.

Francis, A., Warwick, S.I., 2009. The biology of Canadian weeds. 142. *Camelina alyssum* (Mill.) Thell.; *C. microcarpa* Andrz. ex DC.; *C. sativa* (L.) Crantz. Can. J. Plant Sci. 89, 791–810.

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). Mol. Ecol. 13, 2789–2795.

Frenzel, B., 1968. The Pleistocene vegetation of northern Eurasia. Science 161, 637–649.

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. J. Biogeogr. 43, 738–749.

German, D.A., 2018. On the identity of some endemic SW Asian Cruciferae. I. *Thlaspi s. l.* Turczaninowia 21, 180–187.

Gomes Rodrigues, H., Marivaux, L., Vianey-Liaud, M., 2012. Expansion of open landscapes in Northern China during the Oligocene induced by dramatic climate changes: Paleoecological evidence. Palaeogeogr. Palaeoclimatol. Palaeoecol. 358–360, 62–71.

Gradstein, F.M., Ogg, J.G., Schmitz, M., Ogg, G., 2012. The Geologic time scale 2012. Elsevier, Amsterdam.

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.

Hecker, U., 1987. Die Farn- und Blütenpflanzen des Mainzer Sandes. Mainzer Naturw. Archiv 25: 85–133.

Hedge, I.C., 1976. A systematic and geographical survey of the Old World Cruciferae, in: Vaughan, J.G., MacLeod, A.J., Jones, B.M.G. (Eds.), Biology and chemistry of the Cruciferae. Academic Press, pp. 1–45.

Herden, T., Hanelt, P., Friesen, N., 2016. Phylogeny of *Allium* L. subgenus *Anguinum* (G. Don. ex W.D.J. Koch) N. Friesen (Amaryllidaceae). Mol. Phylogenet. Evol. 95, 79–93.

Hohmann, N., Wolf, E.M., Lysak, M.A., Koch, M.A., 2015. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. Plant Cell tpc.15.00482.

Hovsepyan, R., Willcox, G., 2008. The earliest finds of cultivated plants in Armenia: evidence from charred remains and crop processing residues in pisé from the Neolithic settlements of Aratashen and Aknashen. Veg. Hist. Archaeobotany 17, 63–71.

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). Mol. Ecol. 21, 1223–1238.

Hurka, H., Friesen, N., Bernhardt, K.G., Neuffer, B., Smirnov, S., Schmakov, A., Blattner, F. 2019. The Eurasian steppe belt: Status quo, origin and evolutionary history. Turczaninowia 22, 5 - 70.

Iskandarov, U., Kim, H.J., Cahoon, E.B., 2014. *Camelina*: An emerging oilseed platform for advanced biofuels and bio-based materials, in: McCann, M.C., Buckeridge, M.S., Carpita, N.C. (Eds.), Plants and BioEnergy. Springer New York, New York, NY, pp. 131–140.

Ivanov, D., Utescher, T., Mosbrugger, V., Syabryaj, S., Djordjević-Milutinović, D., Molchanoff, S., 2011. Miocene vegetation and climate dynamics in Eastern and Central Paratethys (Southeastern Europe). Palaeogeogr. Palaeoclimatol. Palaeoecol. 304, 262–275.

Jacobs, B.F., Kingston, J.D., Jacobs, L.L., 1999. The origin of grass-dominated ecosystems. Ann. Mo. Bot. Gard. 86, 590–643.

Jiang, H., Ding, Z., 2009. Spatial and temporal characteristics of Neogene palynoflora in China and its implication for the spread of steppe vegetation. J. Arid Environ. 73, 765–772.

Julié-Galau, S., Bellec, Y., Faure, J.-D., Tepfer, M., 2014. Evaluation of the potential for interspecific hybridization between *Camelina sativa* and related wild Brassicaceae in anticipation of field trials of GM camelina. Transgenic Res. 23, 67–74.

Kajtoch, Ł., Cieślak, E., Varga, Z., Paul, W., Mazur, M.A., Sramkó, G., Kubisz, D., 2016. Phylogeographic patterns of steppe species in Eastern Central Europe: a review and the implications for conservation. Biodivers. Conserv. 25, 2309–2339.

Koch, M.A., Haubold, B., Mitchell-Olds, T., 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol. Biol. Evol. 17, 1483–1498.

Koch, M.A., 2019. The plant model system *Arabidopsis* set into an evolutionary, systematic and spatio-temporal context. J. Exp. Bot. 70, 55–67.

Koch, M.A., German, D.A., Kiefer, M., Franzke, A., 2018. Database taxonomics as key to modern plant biology. Trends Plant Sci. 23, 4–6.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. 102, 8369–8374.

Kroonenberg, S.B., Rusakov, G.V., Svitoch, A.A., 1997. The wandering of the Volga delta: a response to rapid Caspian sea-level change. Sediment. Geol. 107, 189–209.

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276–3278.

Lavrenko, E.M., 1969. Über die Lage des eurasiatischen Steppengebiets in dem System der pflanzengeographischen Gliederung des aussertropischen Eurasiens. Vegetatio 19, 11–20.

Léonard , J., 1988. Contribution à l'étude de la flore et de la végtéation des déserts d'Iran: Etude des aires de distribution les phytochories, les chorotype, Fascicules 8–9. Jardin Botanique National de Belgique, Meise. (In French).

Loadmap, 2018. http://loadmap.net/ (accessed last on 2 February 2019).

Luo, Z., Brock, J., Dyer, J.M., Kutchan, T., Schachtman, D., Augustin, M., Ge, Y., Fahlgren, N., Abdel-Haleem, H., 2019. Genetic diversity and population structure of a *Camelina sativa* Spring Panel. Front. Plant Sci. 10.

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

Mamedov, A.V., 1997. The late Pleistocene-Holocene history of the Caspian sea. Quat. Int. 41/42, 161–166.

Martin, S.L., Smith, T.W., James, T., Shalabi, F., Kron, P., Sauder, C.A., 2017. An update to the Canadian range, abundance, and ploidy of *Camelina* spp. (Brassicaceae) east of the Rocky Mountains. Botany 95, 405–417.

Meindl, C., Brune, V., Listl, D., Poschlod, P., Reisch, C., 2016. Survival and postglacial immigration of the steppe plant *Scorzonera purpurea* to Central Europe. Plant Syst. Evol. 302, 971–984.

Meng, H.-H., Gao, X.-Y., Huang, J.-F., Zhang, M.-L., 2015. Plant phylogeography in arid Northwest China: Retrospectives and perspectives. J. Syst. Evol. 53, 33–46.

Meusel, H., Jäger, E., Weinert, E., 1965. Vergleichende Chorologie der Zentraleuropäischen Flora. Vol. 1 Gustav Fischer Verlag, Jena, German Democratic Republic.

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). Presented at the 2010 Gateway Computing Environments Workshop (GCE), IEEE, New Orleans, LA, USA, pp. 1–8.

Mirek, Z., 1984. Monographic studies in genus *Camelina* Cr. 1. *Camelina anomala* Boiss. et Hausskn. Acta Soc. Bot. Pol. 53, 429–432.

Müller, S., Salomo, K., Salazar, J., Naumann, J., Jaramillo, M.A., Neinhuis, C., Feild, T.S., Wanke, S., 2015. Intercontinental long-distance dispersal of Canellaceae from the New to the Old World revealed by a nuclear single copy gene and chloroplast loci. Mol. Phylogenet. Evol. 84, 205–219.

Nikiforova, O.D., 1994. *Camelina* Crantz - Ryzhik., in: Malyschev, L.I., Peschkova, G.A. (Eds.), Flora Sibirica Volume VII: Berberidaceae – Grossulariaceae, Novosibirsk, Nauka, pp. 148-149.

Novikova, P.Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., Guggisberg, A., Paape, T., Schmid, K., Fedorenko, O.M., Holm, S., Säll, T., Schlötterer, C., Marhold, K., Widmer, A., Sese, J., Shimizu, K.K., Weigel, D., Krämer, U., Koch, M.A., Nordborg, M., 2016. Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. Nat. Genet. 48, 1077–1082.

Novikova, P.Y., Hohmann, N., Van de Peer, Y., 2018. Polyploid *Arabidopsis* species originated around recent glaciation maxima. Curr. Opin. Plant Biol. 42, 8–15.

Orlova, L.A., Kuzmin, Y.V., Dementiev, V.N., 2004. A review of the evidence for extinction chronologies for five species of Upper Pleistocene megafauna in Siberia. Radiocarbon 46, 301–314.

Owen-Smith, N., 1987. Pleistocene extinctions: the pivotal role of megaherbivores. Paleobiology 13, 351–362.

Plenk, K., Bardy, K., Höhn, M., Thiv, M., Kropf, M., 2017. No obvious genetic erosion, but evident relict status at the westernmost range edge of the Pontic-Pannonian steppe plant *Linum flavum* L. (Linaceae) in Central Europe. Ecol. Evol. 7, 6527–6539.

POWO, 2018. Plants of the World Online portal. Facilitated by the Royal Botanic Gardens, Kew. http://www.plantsoftheworldonline.org/ (accessed last 2 February 2019)

QGIS Development Team, 2018. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org (accessed last on 2 February 2019).

Qiu, Y.-X., Fu, C.-X., Comes, H.P., 2011. Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. Mol. Phylogenet. Evol. 59, 225–244.

Quan, C., Liu, Z., Utescher, T., Jin, J., Shu, J., Li, Y., Liu, Y.-S., 2014. Revisiting the Paleogene climate pattern of East Asia: A synthetic review. Earth-Sci. Rev. 139, 213–230.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Sandom, C., Faurby, S., Sandel, B., Svenning, J.-C., 2014. Global late Quaternary megafauna extinctions linked to humans, not climate change. Proc. R. Soc. B Biol. Sci. 281, 20133254.

Sang, T., Crawford, D.J., Stuessy, T.F., 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am. J. Bot. 84, 1120–1136.

Schultz, J., 2005. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. RNA 11, 361–364.

Séguin-Swartz, G., Nettleton, J.A., Sauder, C., Warwick, S.I., Gugel, R.K., 2013. Hybridization between *Camelina sativa* (L.) Crantz (false flax) and North American *Camelina* species. Plant Breed. 132, 390–396.

Shaw, J., Lickey, E.B., Schilling, E.E., Small, R.L., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am. J. Bot. 94, 275–288.

Sinskaia, E.N., Bezluzheva, A.A., 1931. The forms of *Camelina sativa* in connection with climate, flax and man. Bull. Appl. Bot. Genet. Plant Breed. 25, 89–200.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Starobogatov, Y.I., 1994. The Freshwater Zebra Mussel; Systematics, Ecology and Applications, in: Systematics and palaeontology. Moscow, pp. 18–46.

Stebbins, G.L.J., 1950. Variation and evolution of plants. Columbia University Press, New York, USA.

Stewart, J.R., Lister, A.M., Barnes, I., Dalén, L., 2010. Refugia revisited: individualistic responses of species in space and time. Proc. R. Soc. B Biol. Sci. 277, 661–671.

Strömberg, C.A.E., 2011. Evolution of grasses and grassland ecosystems. Annu. Rev. Earth Planet. Sci. 39, 517–544.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A., 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4.

Suh, Y., Thien, L.B., Reeve, H.E., Zimmer, E.A., 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of Ribosomal DNA in Winteraceae. Am. J. Bot. 80, 1042–1055.

Svitoch, A.A., 2012. The Caspian Sea shelf during the Pleistocene regressive epochs. Oceanology 52, 526–539.

Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.

Takhtajan, A., 1986. Floristic regions of the world. University of California Press, Berkeley.

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. J. Biogeogr. 27, 609–620.

Tedin, O., 1925. Vererbung, Variation und Systematik in der Gattung *Camelina*. Hereditas 6, 275–386.

The Nature Conservancy, 2018. http://maps.tnc.org/gis_data.html (accessed last on 2 February 2019).

TopGlobus, 2018. http://www.topglobus.ru/ (accessed last on 2 February 2019).

Tudryn, A., Chalié, F., Lavrushin, Y.A., Antipov, M.P., Spiridonova, E.A., Lavrushin, V., Tucholka, P., Leroy, S.A.G., 2013. Late Quaternary Caspian Sea environment: Late Khazarian and Early Khvalynian transgressions from the lower reaches of the Volga River. Quat. Int. 292, 193–204.

Tutin, T.G., Burges, N.A., Chater, A.O., Edmondson, J.R., Heywood, V.H., Moore, D.M., Valentine, D.H., Walters, S.M., Webb, D.A., 1993. Flora Europaea Volume I, 2nd Edition: Psilotaceae to Platanaceae, Cambridge University Press. New York.

Van Baak, C.G.C., Vasiliev, I., Stoica, M., Kuiper, K.F., Forte, A.M., Aliyeva, E., Krijgsman, W., 2013. A magnetostratigraphic time frame for Plio-Pleistocene transgressions in the South Caspian Basin, Azerbaijan. Glob. Planet. Change 103, 119–134.

Varga, Z., 2010. Extra-Mediterranean refugia, post-glacial vegetation history and area dynamics in Eastern Central Europe, in: Relict species. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 57–87.

Vassilczenko, I.T., 1939. Rod 687. Ryzhik – *Camelina* Crantz., in: Komarov, V.L. (Ed.), Flora of the USSR, Volume VIII: Capparidaceae, Cruciferae & Resedaceae, Izdatel'stvo Akademii Nauk SSSR. Moskva – Leningrad. Union of Soviet Socialist Republics. pp. 596–602.

Vavilov, N.I., 1992. Origin and geography of cultivated plants. Cambridge University Press, United Kingdom.

Velichko, A.A., 1999. General features of the landscape and climate of northern Eurasia in the Cenozoic, in: Climate and environment changes during the last 65 million years; Cenozoic: from Paleocene to Holocene. GEOS, Moscow, pp. 219–233.

Volkova, P.A., Herden, T., Friesen, N., 2017. Genetic variation in *Goniolimon speciosum* (Plumbaginaceae) reveals a complex history of steppe vegetation. Bot. J. Linn. Soc. 184, 113–121.

Wikimapia, 2018. http://www.wikimapia.org/ (accessed last on 2 February 2019).

World Wildlife Fund for a living planet, 2018. https://web.archive.org/web/20110401124312/http://wwf.panda.org:80/about_our_earth/ecoregions/about/habitat_types/selecting_terrestrial_ecoregions/habitat08.cfm (accessed last on 2 February 2019).

Wright, S.D., Yong, C.G., Wichman, S.R., Dawson, J.W., Gardner, R.C., 2003. Stepping stones to Hawaii: a trans-equatorial dispersal pathway for *Metrosideros* (Myrtaceae) inferred from nrDNA (ITS+ETS): Stepping stones to Hawaii. J. Biogeogr. 28, 769–774.

Yang, S., Ding, Z., 2010. Drastic climatic shift at ~2.8Ma as recorded in eolian deposits of China and its implications for redefining the Pliocene-Pleistocene boundary. Quat. Int. 219, 37–44.

Yıldırımlı, Şinasi, 2011. Three new species from Turkey. Ot Sist. Bot. Derg. 18, 2–5.

Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol. Phylogenet. Evol. 87, 46–49.

Zinger, H.B., 1909. On the species of *Camelina* and *Spergularia* occurring as weeds in sowings of flax and their origin. Tr. Bot. Muz. Akad. Nauk USSR 6, 1–303.

PUBLICATION II:

EVOLUTIONARY HISTORY OF THE EURASIAN STEPPE PLANT

SCHIVERECKIA PODOLICA (BRASSICACEAE) AND ITS CLOSE RELATIVES

FRIESEN N, **ŽERDONER ČALASAN A**, NEUFFER B, GERMAN DA, MARKOV M, HURKA H.

FLORA (2020) 268: 151602.

DOI: 10.1016/j.flora.2020.151602

Evolutionary history of the Eurasian steppe plant *Schivereckia podolica* (Brassicaceae) and its close relatives

Nikolai Friesen[a,b,*], Anže Žerdoner Čalasan[c], Barbara Neuffer[c], Dmitry A. German[d], Michael Markov[e], Herbert Hurka[c]

[a] Botanical Garden of the University of Osnabrueck, 49076, Osnabrueck, Germany

[b] I. M. Sechenov First Moscow State Medical University of Ministry of Health of the Russian Federation, Izmailovsky Boulevard, 8, Moscow, 105043, Russian Federation

[c] Department of Botany, University of Osnabrueck, 49076, Osnabrueck, Germany

[d] South-Siberian Botanical Garden, Altai State University, 656049, Barnaul, Russia

[e] Department of Botany, Moscow Pedagogical State University, 1/1 M. Pirogovskaya Str., Moscow, 119991, Russia

* Corresponding author. *E-mail address*: nfriesen@uni-osnabrueck.de

Abstract

The genus *Schivereckia* (Brassicaceae) – presently included in the genus *Draba* – comprises two species, *S. podolica* and *S. doerfleri*. *Schivereckia podolica* is an eastern European species with disjunct distribution in Russia, Ukraine, and Podolian-Galician outposts. *Schivereckia doerfleri* is reported from alpine fields of the Balkan Peninsula and from mountain chains of northern Anatolia, Turkey. *Schivereckia podolica* shares its ecological and biogeographic characteristics with a number of other steppe plants. The 'relic hypothesis' interprets the present locations of *Schivereckia* and its associates as relics of a formerly continuous distribution belt of a cold periglacial steppe vegetation.

We employed nuclear ITS and chloroplast trnL-trnF region, and DNA fingerprinting (RAPD) to elucidate the evolutionary history and historical biogeography of *Schivereckia*. RAPD data prove isolation by distance and mirror the present disjunct geographical distribution pattern of *Schivereckia podolica*. Phylogenetic analyses revealed that *Schivereckia* might be polyphyletic and nested within a highly supported clade of central Asian, Caucasian and North American *Draba* species. Time divergence estimation and ancestral area reconstruction support the interpretation of species of the *Schivereckia* subclade as periglacial relic. Our results argue for a continuous persistence of certain steppe species since the Pleistocene, and support the view that the modern steppes of Europe have derived partly from Pleistocene glacial steppes.

Keywords: Periglacial relics, Dated phylogeny, Ancient areal reconstruction, ITS, trnL-trnF, DNA fingerprints

1 Introduction

For more than 100 years, *Schivereckia podolica* has played an important role in the florogenetic literature of Eastern Europe (Taliev, 1897; Litvinov, 1902; Kozo-Poljanski, 1928; Gajeweski, 1934; Tolmachev, 1939; Kazakova, 1984; Radygina, 2003; Iljinska et al., 2007). This is mainly due to its distribution area, its preferred habitat requirements, and due to the fact that it belongs to conspicuous eastern European plant communities. It is a tap-rooted herbaceous perennial, which does not tolerate shade nor competition with other herbaceous plants. It grows on rocks, dry slopes and bare or only weakly turfed outcrops particularly on calcareous substrates such as limestone, chalk, gypsum and chernozem-carbonate soils. *Schivereckia podolica* shares its ecological and biogeographic characteristics with a number of other calcicolous petrophytic steppe plants, for instance *Allium strictum* Schrader, *Androsace villosa* L., *Bupleurum multinerve* DC., *Clausia aprica* (Stephan ex Willd.) Korn.-Tr., *Daphne cneorum* L., *Dendranthema zawadskii* (Herbich) Tzvel. (*Chrysanthemum sibiricum* Turcz. s.l.), *Gypsophila litvinovii* K.-Pol., *Helictotrichon desertorum* (Less.) Nevski (= *Avena desertorum* Less.), and *Scutellaria supina* L. (= *S. alpina* subsp. *supina* (L.) I.B.K. Richardson). These “mysterious elements” (Kozo-Poljanski, 1928) grow on low hillsides on chalk or calcareous substrates with a strongly fragmented distribution area in Eastern Europe and are members of the association Avenetum desertori (Gajeweski, 1934). “These are plants which either in their distribution pattern or in their taxonomic position or physiognomic and ecological characters, or finally in the entirety of all these features give the impression of mountain and even alpine or arctic elements” (Kozo-Poljanski, 1928, translated from German).

Several hypotheses have been put forward to explain the assembly of this striking association of plants in Eastern Europe. The ‘synanthropic hypothesis’ has been put forward by V.I. Taliev already in the year 1897. He explains the disjunct distribution pattern of *Schivereckia podolica* and associated plants by a rather recent colonization of the chalk hills and equivalent habitats starting from refugial areas (which are not further characterized or located). River valleys and steep slopes could have served as natural channels, through which *Schivereckia* spread. Humans who may have preferred this type of landscape could have created additional suitable habitats for the colonizing plants and may have unintentionally fostered their distribution/ dispersal. This hypothesis has presently only very few if any supporters.

The ‘relic hypothesis’ was formulated by Litvinov in 1902. He stressed the role of the glaciers during the last Ice Ages. He interpreted the present locations of *Schivereckia* and its associates as

relics of a formerly continuous distribution belt of a cool periglacial steppe vegetation. This view has since then gained much support, especially by geobotanists in the 1930ies (e.g. Kozo-Poljanski, 1928, 1929, 1931; Gajeweski, 1934; Kleopov, 1941, 1990). *Schivereckia* is nowadays regarded mostly as an ancient periglacial relic species (Tolmachev, 1939; Kazakova, 1984). Radygina (2003) however put forward another hypothesis. She argues that *Schivereckia podolica* is indeed ancient but not a relic species, at least not in its Uralian part areal, and is of Mediterranean origin, a view, which was also held by Alexejenko (1950). Thus, *Schivereckia* remains "one of the most mysterious species in florogenetic aspects" (Radygina, 2003).

The genera *Schivereckia* Andrz. ex DC. and *Draba* L. are closely related, and *Schivereckia* is nowadays included in the genus *Draba* (Al-Shehbaz et al., 2006; Jordon-Thaden et al., 2010). The genus *Draba* belongs to the tribe Arabideae, and within this tribe, *Draba* (including *Schivereckia*) represents a monophyletic evolutionary lineage (Jordon-Thaden et al., 2010; Karl and Koch, 2013). *Draba* is the most species- rich genus of the Brassicaceae comprising ca. 400 species (Koch et al., 2018). It is distributed in the Northern Hemisphere and South America, predominantly in arctic/ subarctic, mountain and alpine regions (Jordon-Thaden et al., 2013).

Schulz (1936) recognized two *Schivereckia* species, *Schivereckia podolica* Andrz. & Bess. ex DC. (Fig. 1), and *S. doerfleri* (Wettst.) Bornm. Later, a number of species were described by Alexejenko (1946, 1950) from the territory of the former Soviet Union, in many ways intermediates between *S. podolica* and *S. doerfleri*: (i) *Schivereckia podolica* occurring in W Ukraine, Moldavia, N.E. Romania, on rocks and dry slopes, particularly on calcareous substrata; (ii) *S. mutabilis* (Alexeenko) Alexeenko occurring in S.C. Russia and N.E. Ukraine, on chalk outcrops; (iii) *S. berteroides* Fisch. ex Alexeenko occurring in S. Ural and Trans-Volga, on rocks; (iv) *S. monticola* Alexeenko occurring in C. Ural and Volga-Kama, on rocks; (v) *S. kusnezovii* Alexeenko occurring in N. Ural, on limestone and dunites. Kazakova (1984) revised the taxonomy of *Schivereckia*, and suggested, based on much more material than had ever been analyzed before, that all described variants are best to be summarized under the same name — *Schivereckia podolica*, which thus became a very polymorphic species in terms of morphology. This view has been adopted in the 2nd edition of the Flora Europaea, vol. I (Tutin et al., 1993), and also e.g. by Jordon-Thaden et al. (2010), and is also followed by us in the present paper.

There was a nomenclatural confusion related to the name *Draba podolica*. Berkutenko (1995) found out that the original material of *Alyssum hyperboreum* L. does not belong to the Amphi-

Beringian *Draba hyperborea* (L.) Desv. but instead refers to *D. podolica* s. l. This means that under the direct application of the ICN (Turland et al., 2018) the name *D. hyperborea* has priority over *D. podolica* and should thus be applied to the latter species. Regarding potentially quite serious negative consequences of this finding, attempts of keeping the nomenclatural stability and saving the use of the name *Draba* (or *Schivereckia*) *podolica* by either rejection of the name *A. hyperboreum* (Mosyakin, 2015) or its conservation with a conserved type (German, 2017) have been made. The first proposal was deferred by the General Committee of IAPT (Wilson, 2017), and no decision has been made on the second proposal at the first attempt (Applequist, 2017) but recently it was finally approved (Applequist, 2019). Following this decision, we keep the existing use of the names, i. e. in applying the epithet "podolica" for the species known under this name (usually as *Schivereckia podolica*) in a vast majority of sources, including biogeographical literature.

Schivereckia podolica (Fig. 1) is a rare Eastern-European species with a disjunct distribution (see Fig. 2). This species can be found (i) from North to South Ural (possibly also in Novaya Zemlya, Tolmachev, 1939); (ii) Volga region (Uljanovsk and Samara oblasts); (iii) Don region (Lipetsk – Voronezh – Kursk oblasts); (iv) Donetsk region in the Ukraine; and (v) Podolian-Galician outposts (Podolian Plateau in the Ukraine; at the river Dnjestr in Moldavia; at the river Prut in NE Romania). *Schivereckia doerfleri* is reported from alpine fields of the Balkan Peninsula (Albania, Serbia and Bulgaria) and from Kuzey Anadolu Daglari mountain chains in the northern parts of Anatolia, Turkey, provinces Kastamonu and Amasya.

The aims of our study are (a) to reveal the closest relatives of *Schivereckia* in order to shed light on its evolutionary history, and (b) to understand the historical biogeography of the enigmatic *Schivereckia podolica*. To achieve these goals, we employed molecular markers nuclear ITS and chloroplast trnL-F, and DNA fingerprinting (RAPD). We analysed *Schivereckia podolica* accessions from all known parts of its disjunct areal. ITS sequences of selected *Draba* and *Schivereckia* accessions were included into an ITS dataset of all currently available sequences of species from the tribe Arabideae, and time estimation analyses were carried out. Ancestral area reconstruction was performed for *Schivereckia* and its closely related species (later on as *Schivereckia* subgroup). Based on all these data, we elucidate the evolutionary history and historical biogeography of *Schivereckia* and its closest relatives against the background of the evolutionary history of the Eurasian steppe.

2 Materials and methods

2.1 Plant material

Draba and *Schivereckia* accessions and their provenances analysed for ITS and cpDNA sequences are listed in Table 1, and corresponding GenBank accession numbers are given. Provenances of *Schivereckia* accessions used for DNA-fingerprinting (RAPD analyses) are shown in Table 2. The ITS dataset of all currently available sequences of species from the tribe Arabideae were extracted from the database on Brassicaceae "BrassiBase" (Koch et al., 2012; Kiefer et al., 2014; Koch et al., 2018).

2.2 DNA sequencing

Total genomic DNA was isolated from herbarium specimens using the NucleoSpin Plant Kit (Macherey-Nagel, Düren, Germany) according to the instructions of the manufacturer and used directly in PCR amplifications. PCR was carried out using 20.8 µL of ddH_2O, 3 µL of 10x Taq Buffer with $MgCl_2$, 2 µL of 10 mM dNTP mixture, 1 µL of DMSO, 0.2 µL of 5U Taq Polymerase, 1 µL of the corresponding primers, and 1 µL of the template DNA. The analysed ITS region was amplified using primers published previously in Blattner (1999). PCR conditions included initial denaturation at 95 °C for 5 min, 35 cycles of denaturation at 95 °C for 30 s, primer annealing at 55 °C and extension at 68 °C for 1 min, and additional extension at 68 °C for 10 min. Samples, that yielded single bands once run through gel electrophoresis, were sent to Microsynth Seqlab (Göttingen, www.microsynth.ch) for purification and sequencing. Sequences from each accession were manually edited in Chromas Lite 2.6.4 (Chromas | Technelysium Pty Ltd) and aligned and manually corrected in MEGA7 (Kumar et al., 2016). In addition to the ITS region, trnL-trnF chloroplast region was also investigated on a reduced taxon sample, using corresponding primers published previously in Taberlet et al. (1991).

2.3 Phylogenetic analyses

The successfully amplified ITS sequences of selected *Draba* and *Schivereckia* accessions were included into a taxonomically revised ITS dataset of all currently available sequences of species, belonging to the tribus Arabideae (Kiefer et al., 2014; Koch et al., 2012, 2018). Maximum likelihood tree inference relied on RAxML-HPC v8.2.10 (Stamatakis, 2014) with 1000 bootstrap iterations with the default number of distinct rate categories. Tree searches were executed from a random maximum-parsimony tree employing either the GTRCAT or the GTRGAMMA

approximation of rate heterogeneity. In addition, Bayesian Inference was carried out using MrBayes v3.2.6 (Ronquist et al., 2012) under the GTR+Γ substitution model using the random-addition-sequence method with 10 replicates. Two independent Markov chain Monte Carlo (MCMC) analyses of four chains were run — one cold and three heated. Runs were carried out for 20 million cycles, with parameters and trees sampled every 1000th cycle including an appropriate burn-in (10 %) as inferred from the evaluation of the trace files using Tracer v1.6.0 (Rambaut et al., 2018). Three different datasets were investigated using the above mentioned inference methods: Arabideae dataset, *Draba* s.l. dataset and *Schivereckia* dataset.

2.4 Time divergence estimation

Three independent time estimation analyses were carried out. The first analysis was carried out on all the Arabideae according to the BrassiBase data bank (Koch et al., 2012; Kiefer et al., 2014; Koch et al., 2018). The second analysis comprised a reduced taxon sample, including only *Draba* s.l., *Drabella* and four *Arabis* species that have been placed as its sister group in the big phylogenetic reconstruction including all Arabideae accessions. The third time estimation analysis was carried on *Schivereckia* and its sister group only. As it is somewhat hard to produce a reliable and stable time estimation analysis especially in the light of fossil absence, all three different taxon-sample-based time estimation analyses — the whole Arabideae clade, *Draba* s.l. as well as *Schivereckia* clade — were carried out using three different tree priors: Yule, Birth-Death and Coalescence. We thus ended up with nine different time estimation analyses. Calibration for all the analyses was based on a published mean ITS substitution rate for herbaceous annual/ perennial angiosperms of 4.13×10^{-9} sub/site/yr (Kay et al., 2006). Molecular dating analyses relied on BEAUti & BEAST v1.8.4 (Suchard et al., 2018) and an uncorrelated relaxed clock model. The use of this model was justified, after assessing the coefficient of variation in the trace file, which exceeded 0.3. We used the GTR+Γ substitution model, and five independent Monte Carlo Markov chains (MCMC) runs for 50 million generations, with parameters sampled every 10,000th generation. Effective sample sizes (ESS) for all estimated parameters were assessed using Tracer v1.6.0 (Rambaut et al., 2018). TreeAnnotator v1.8.4 (Suchard et al., 2018) was used to discard 10 % of the saved trees and annotate the rest of them. Maximum clade credibility tree with mean node heights was visualized using FigTree graphical viewer of phylogenetic trees v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). All sequence evolution models used in this study were assessed using the Akaike information criterion (AIC) implemented in the jModelTest2 v2.1.6 (Darriba et al.,

2012). All the analyses were carried out at the CIPRES Science Gateway computing facility (Miller et al., 2010). The aligned matrices are available as *.nex files upon request.

2.5 Ancestral area reconstruction

To elucidate the historical biogeography of the *Schivereckia* subgroup, an ancestral area reconstruction was carried out, using the Dispersal-Extinction-Cladogenesis (DEC; Ree and Smith, 2008) implemented in RASP v4.1.0 (Yu et al., 2015). The reconstruction was based on the third time estimation analysis dataset, including only representatives of *Schivereckia* subgroup and its sister *Draba* subgroup showing a predominantly Subarctic and Eurasian mountain chain distribution range. Distribution ranges were divided into nine regions: (A) non-arctic North America, (B) Arctic/ Subarctic regions circumpolar, (C) Caucasus Mountains, (D) Ural Mountains, (E) Volga-Don, (F) Ukraine, (G) Balkan and Anatolia, and (H) Altay-Sayan Mountains. Maximal number of allowed areas was set to two and no additional time dependent restrictions were taken into account. However, based on the geographical history of the Eurasian steppe belt and adjacent regions (Hurka et al., 2019), only the following region combinations were allowed: AB, BD, BH, CD, CE, CF, CG, DE, DF, DG, EF, EG and FG.

2.6 DNA fingerprinting

27 accessions of *Schivereckia podolica* and one accession of *Schivereckia doerfleri* have been analysed (Table 2) for RAPDs covering the entire geographical distribution range of *Schivereckia podolica* (Fig. 2). To generate comparable results in RAPD PCR with DNA from herbarium material, the entire isolated DNA was applied on 0.8 % agarose gel and only high molecular weight DNA was isolated from the gel and used for analysis. The concentration of the extracted DNA was checked on an agarose gel. Amplification was carried out using 11 arbitrary 10 bp long primers (A-19, AB-4, AB-18, AC-2, C-7, C-9, C-13, D-1, D-3, G-13 and G-19) obtained from Operon Technologies, Alameda, CA. One third of the reaction mixtures were separated on 1.5 % agarose gels in 0.5x TBE buffer, followed by staining with ethidium bromide. Clearly visible RAPD bands (from 200 bp to 2000 bp) were scored manually for presence (1) or absence (0) from enlarged photographs of the gels. Only bands reproducible in two independent amplification reactions were included in the data analyses. From the resulting 1/0 data matrix, pairwise genetic distances were calculated using the Jaccard coefficient. The genetic distance matrix was subjected to a principal component analysis (PCA). From the distances, new independent axial coordinates are calculated

which represent most of the variability of the original data. These calculations were done with the SPSS 25 Program (SPSS 22, 2017).

The complete RAPD-data set for *Schivereckia* (i.e. samples without missing data, n = 27) were evaluated with the program GenAlEx 6 (Peakall and Smouse, 2006, 2012) by carrying out a Mantel test. This test compares the matrix of pairwise genetic distances among populations and the matrix of geographic distances (km) among populations (Mantel, 1967), trying to establish a correlation between them. Significance tests were based on 99 permutations.

3 Results

3.1 Evolutionary lineages in *Draba*

In total, 86 DNA sequences (ITS and trnL-trnF) were generated and deposited in GenBank under the accession numbers FM957484-FM957524; LR590254-LR590266 for ITS sequences, and LR590427-LR590462 for trnL-trnF sequences (Table 1). The analysed trnL-trnF spacer sequences confirmed the low resolution in *Draba* species (Jordon-Thaden et al., 2010). Only the *Schivereckia* clade was well supported (1.00 BPP). Two nucleotide substitutions were characteristic for the clade, namely A to C transversion in the position 85 and T to G transversion in the position 238 of the alignment. As outlined above (chapter 2.3), our ITS sequences were combined with the most comprehensive ITS dataset for the tribe Arabideae currently available on Brassibase (https://brassibase.cos.uni-heidelberg.de/). Detailed results of the phylogenetic analyses of this dataset are presented in the Supplement 1.

Figure 1: ***Schivereckia podolica*****.** (A) Habitus (Ukraine, Donezk, by A. Bronskov), (B) close up of the inflorescence with distinctive fruits (Russia, Northern Ural, by S. Glotov).

To reduce potential topological artefacts due to the sampling bias, we also carried out Maximum-Likelihood based and Bayesian-Inference based phylogenies for: the whole Arabideae clade (Suppl. 1) and *Draba* s.l. (Suppl. 2). Overall, the ITS-based accessions recognised as *Schivereckia* were polyphyletic and built a well-supported clade (0.97 BPP) with *Draba baicalensis* and *D. cinerea*. Notably, the ribotype of *Schivereckia podolica* from Donetsk region share a mutation with both ribotype accessions recognised as *S. doerfleri* (Fig. 4 and Suppl. 1, 2, 3 and 4). The entire

phylogenetic statistics for the Arabideae, *Draba* s.l. and *Schivereckia* clade are presented in the Supplement 5.

Hitherto, *Schivereckia* DNA sequences were available only from Botanical Garden accessions of unknown or inadequately documented provenances. We analysed *Schivereckia podolica* accessions from documented provenances covering its whole distribution area (Fig. 2; Table 1 and 2). ITS data did not clearly separate *Schivereckia* species from the other species of the *Schivereckia* subgroup as outlined above. However, the trnL-F sequences confirmed *Schivereckia* as a group by itself despite the low resolution of that marker in the whole genus *Draba* (see also Jordon-Thaden et al., 2010) (Fig. S4) as a whole. The ITS and RAPD data reveal intraspecific variation in *Schivereckia podolica* (Figs. 4 and 7 and Supplement 3).

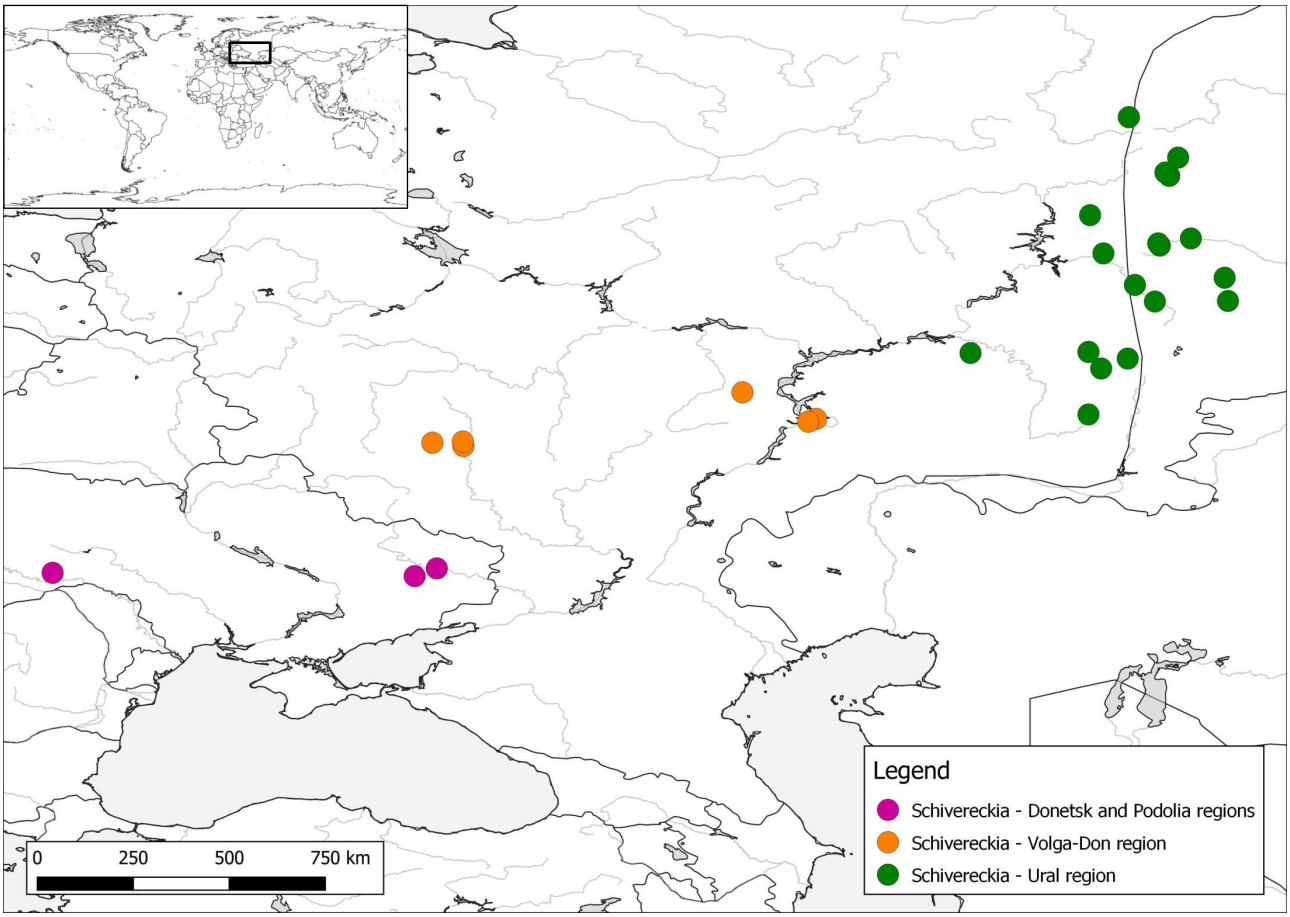


Figure 2: Geographic location of the collected *S. podolica* accessions. Colour coding corresponds to the Figures 5 and 6.

3.2 Time divergence estimation

Time divergence estimation analyses of Arabideae clade, *Draba* s.l. (Fig. 3) and *Schivereckia* (Fig. 4) generated highly congruent topologies (compared to Maximum Likelihood and Bayesian

Inference analyses; Supplements 1, 2 and 3) and time spans (Table 4), with ESS values exceeding 1000. The internal topologies including 388, 278 and 40 taxa respectively (Supplement 5) were moderately to well-supported. The simplified version of different *Draba* s.l. chronograms is a synopsis tree (Fig. 3), in which statistically well-supported branches are named either by taxon names or by the predominant distribution areas of the species belonging to the branch. According to the analysed taxon sets, *Draba* s.l. diversified around 7.4 MYA, in the late Miocene. The first early diverging lineage was *Tomostima* group, which furthermore diversified roughly 5.0 MYA in the early Pliocene. *Draba* s.str. diversified roughly 5.5 MYA. The *Schivereckia* subgroup and its sister group split from the rest roughly 4.4 MYA, and during the mid-Pliocene, *Schivereckia* subgroup split from its sister group (3.9 MYA) and went through diversification roughly 2.6 MYA at the beginning of Pleistocene. The entire time estimation statistics for the Arabideae, *Draba* s.l. and *Schivereckia* clade are presented in the Table 4.

3.3 Ancestral area reconstruction

The distribution range of *Schivereckia* subgroup is depicted in Fig. 5 and is congruent with the current literature on steppe evolution in the Pleistocene (extensively reviewed in Hurka et al., 2019). Ancestral area reconstruction for the *Schivereckia* group is presented in Fig. 6 accompanied by Supplement 6. According to this analysis, the most ancient areal would be a circumpolar arctic/ subarctic region of Pleistocene age (see Fig. 4), including the Ural mountains and present-day nonarctic North America (Fig. 6, Node 55 with 99 % frequency). Contemporary distribution of the North American species has its roots in this ancient areal (Nodes 52–54 in Fig. 6). Distribution areas of the Eurasian species (Caucasus, Altai-Sayan, Eastern Europe) can be also interpreted as derived from an ancient arctic/subarctic vegetation belt (Nodes 49–51 in Fig. 6).

3.4 DNA fingerprinting

The RAPD dataset included 27 individuals of *Schivereckia podolica* and 1 individual of *S. doerfleri* (Table 2). We detected 138 RAPD markers, eight of which were common in all taxa. The *Schivereckia doerfleri* individual generated 32 RAPD bands, 14 of which were private markers occurring only in this individual of the species. Table 3 gives an overview on occurring markers in the three groups Donetsk/Podolia, Ural, and Volga-Don. The Ural-group generated more characters than the other two groups, which might be the effect of the bigger taxon sample. The number of common private alleles is nearly equal for all groups despite the differing taxon sample. There was geographic structure in the distribution of RAPD markers, and genetic distances (Nei's standard

genetic distance) between populations increased significantly with geographic distances (Suppl. 7). For a more detailed view, we analysed the dataset excluding *Schivereckia doerfleri* with a PCA analysis (SPSS 25) (Fig. 7). The first three factors explain 36,171 % of the whole variability of the data set. Factor 1 divided the Ural-cluster from the other two groups. Within the Ural-group provenances from Northern-, Middle-, and Southern Ural clustered together. Factor 2 divided Donetsk/Podolia from the other two groups. The sub-groups Volga-Don 1 and 2 are kept apart by Factors 2 and 3, whereas Volga-Don 2 and the Ural-group clearly come together (Fig. 7). Nevertheless, the two Volga-Don groups are more closely related to each other than Volga-Don 2 is to the Ural-group (Fig. 7). Donetsk/Podolia cluster possess 31 RAPD common markers, 9 of which are private for this group. Volga-Don 1 and 2 have 41 RAPD common markers, 11 of which are private for this group. Volga-Don 2 and Ural share only 24 RAPD common markers and only 3 of which are private markers (Table 3).

4 Discussion

4.1 Phylogenetic analyses of *Draba* s.str. including *Schivereckia*

Rooting and overall topology of our phylogenetic analyses are in accordance with the results of Jordon-Thaden et al. (2010) and Karl and Koch (2013). Jordon-Thaden et al. (2010) recognised three main groups in *Draba* s.l. Our "Asia Minor" group corresponds to their *Draba* group I. Our group including the branches "Asia", "Subarctic + Pacific mountain chain", and "South America" corresponds to *Draba* group II, and the two sister clades "*Schivereckia*" and "Subarctic + Eurasian mountain chains" to *Draba* group III (see Fig. 3). Divergence time estimations are also in accordance with already previously published data (Couvreur et al., 2010; Jordon-Thaden et al., 2013; Karl and Koch, 2013).

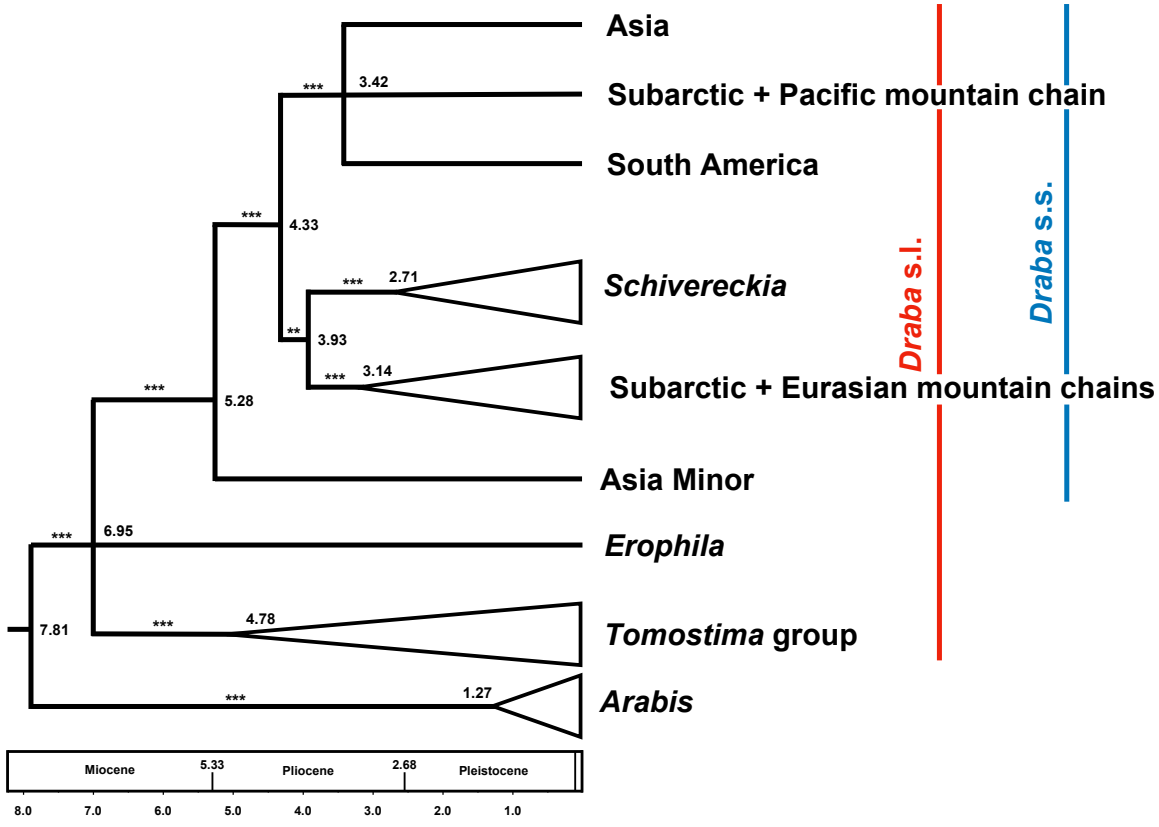


Figure 3: Time estimation outline of *Draba* s.l., *Drabella* and *Arabis*. Numbers on the nodes represent the absolute ages that are in millions of years. The numbers on the branches are statistical support values (Bayesian posterior probabilities; values <.90 are not shown). Significance levels: highly supported (***): BI ≥ 0.98; well supported (**): BI ≥ 0.90. The letters on the nodes correspond to the second column of the Table 4.

Draba group III is the most heterogeneous in the whole genus *Draba* and consists of two subgroups. The majority of group III species belong to the first subgroup and grow in North, Northeast and Central Asia, in circumpolar arctic and subarctic regions as for instance Beringia but include also species from North and South America (Jordon-Thaden et al., 2010). This subgroup is not well resolved and it is represented by the "Subarctic + Eurasian mountain chains" branch (Figs. 3 and S1, S2). The second subgroup within the *Draba* group III is represented in our analyses by the "*Schivereckia*" branch (*Schivereckia* subgroup) in Fig. 3. It is sister clade to the first subgroup (Figs. 3 and S1, S2), and is the most interesting with regard to the focus of the present study. Apart from *Schivereckia*, which is an eastern European genus with outposts in the southern Black Sea mountain ranges, the *Schivereckia* subgroup includes species from arctic/subarctic circumpolar regions (*Draba murrayi*, *D. cinerea*), from the Altai-Sayan and Baikal region (*Draba baicalensis*), from the Caucasus (*Draba* spec.), and from eastern (*Draba ramosissima*) and western North American mountain chains (*Draba pterosperma*, *D. hitchcockii*; Table 1). Otherwise, the *Schivereckia* subgroup is not well-resolved (Figs. 4 and S1, S2, S3, S4).

Jordon-Thaden's et al. (2010) and our *Schivereckia* subgroup differ conspicuously, although we included all species analysed by Jordon-Thaden et al. (2010). Apart from the two *Schivereckia* species, only three species out of nine each (including *Schivereckia*) are common to both subgroups: *Draba cinerea*, *D. murrayi* und *D. rammosissima*. A reason for the mismatch might be our increased sample size compared to Jordon-Thaden et al. (2010). Sample size of Jordon-Thaden included 158 core *Draba* species compared to the 218 *Draba* s. str. species in our study. The differences between the species assemblies of the two subgroups have considerable consequences for biogeographical conclusions. Whereas Jordon-Thaden's *Schivereckia* subgroup includes Himalayan, Tibetan (*Draba glomerata*, *D. surculosa*, *D. winterbottomii*) and coastal Russian Far East and Japanese species (*Draba borealis*), these areas are clearly excluded from our *Schivereckia* subgroup. The Jordon-Thaden subgroup is in historical-biogeographical sense rather heterogeneous, whereas our *Schivereckia* subgroup includes only taxa from periglacial steppe areas and their assumed refuge areas.

The *Draba incana – stylaris* complex needs special attention. The phylogenetic analyses revealed unresolved relationships between the two taxa described as *D. incana* L. and *D. stylaris* J.Gay ex Koch. Three clearly distinct clusters are present (Figs. 4 and S1, S2). One cluster comprises *D. incana* accessions from amphiatlantic arctic to subarctic regions, another cluster those from European Alps and Scandinavia together with *Draba stylaris* accessions of the Alps, and the third

cluster contains *D. stylaris* accessions from the Caucasus (named *Draba* spec. by the authors for reasons given below), clearly distinct from the alpine *D. stylaris* accessions. The Caucasus cluster belongs to the *Schivereckia* subgroup, whereas the two first clusters are members of the "Subarctic + Eurasian mountain chains" subgroup (Figs. 3, 4 and Supplements 1–3). Apparently, the Caucasian *Draba stylaris* represents an undescribed species, named here *Draba* spec. to avoid further confusion. Buttler (1967) already excluded this Caucasian material from the alpine *Draba stylaris* based on morphological characters. He interpreted the *Draba incana* from the Alps as a "Nordic ice age relic" (translated from German). Our molecular data are ambiguous. Alpine *D. incana* accessions cluster with alpine *D. stylaris* and Scandinavian *D. incana* accessions but not with arctic/ subarctic *D. incana* accessions. This issue needs further clarification and taxonomic revisions but is outside our present focus.

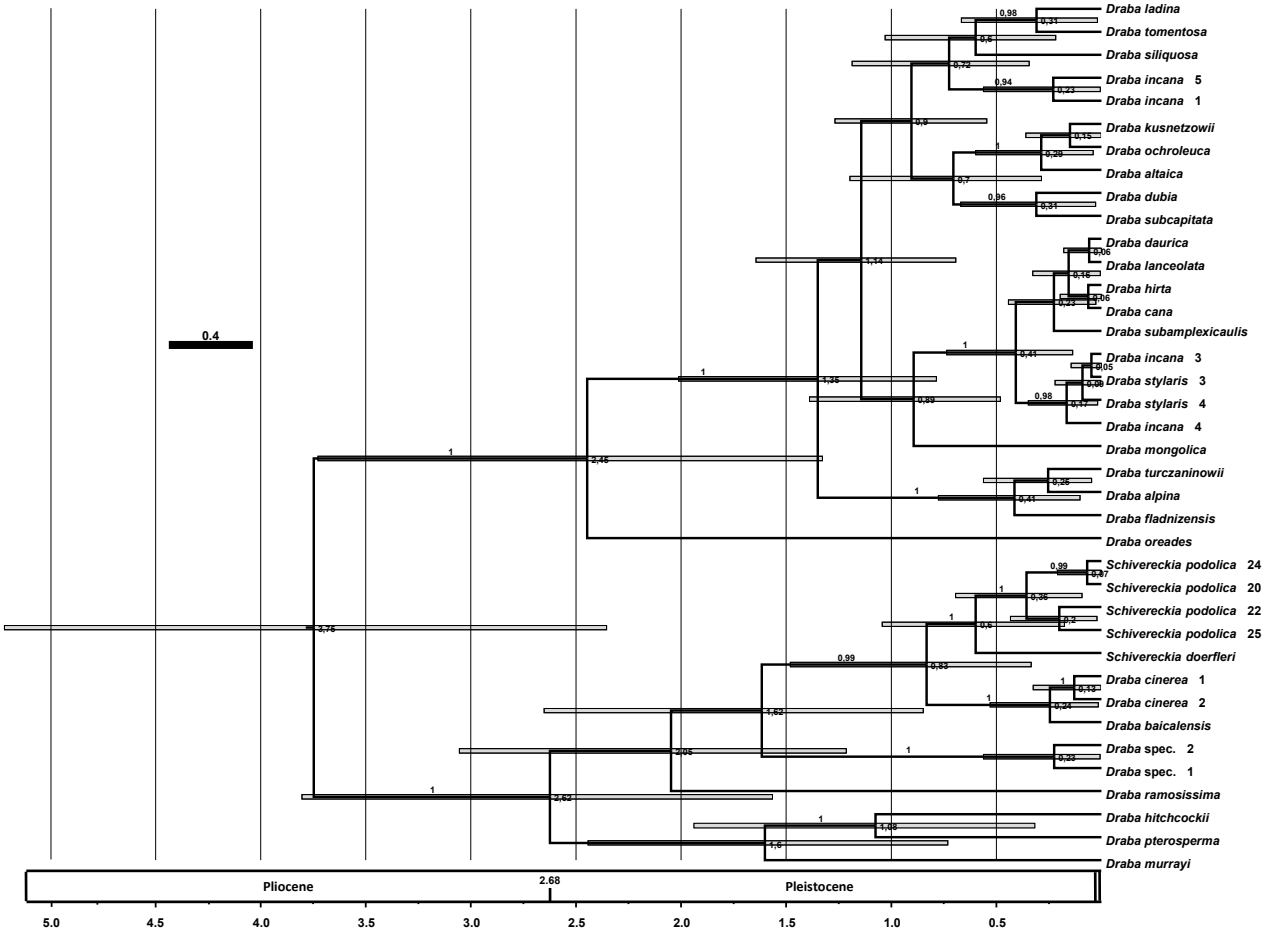


Figure 4: Dated phylogeny of the *Schivereckia* subgroup. Median rate is given in units of substitutions per million years (including 95% confidence intervals). The numbers on the branches are statistical support values (Bayesian posterior probabilities; values <.90 are not shown) and the numbers on the nodes represent the absolute ages that are in millions of years.

4.2. Historical biogeography of the *Schivereckia* subgroup

The *Schivereckia* subgroup challenges the question on how to interpret the peculiar species association and the striking biogeographical pattern. Divergence time estimation and ancestral area reconstruction help to answer this question. The two sister clades "Subarctic + Eurasian mountain chains" and "*Schivereckia*" (Fig. 3) began to diverge in the mid-Pliocene and diversification within both subgroups happened during the Pleistocene (Fig. 4). Driving forces behind the evolutionary split between these two clades remain in the dark. The historical biogeography of the *Schivereckia* subgroup, however, can be inferred from the Pleistocene climate/landscape history and is supported by the ancestral area reconstruction.

During the end of the early Pleistocene (1.8 Ma to 850 ka) dominance of glacial components, spread of permafrost and growing continental ice sheets are recorded throughout the Northern Hemisphere, and prototypes of periglacial steppes during the cold phases appeared. Major continental glaciations were restricted to the last 1.0 Ma to 800 ka or less. Marine Isotope Stage 22 (MIS 22) (ca. 870–880 ka) included the first of the "major" worldwide events with substantial ice volumes that typify the late Pleistocene glaciations. This period is termed 'Middle Pleistocene transition' (Head et al., 2008). Major continental glaciations and waves of deep cooling and warming and periglacial steppes as a zonal landscape structure across the circumpolar regions were characteristic of the middle Pleistocene (850 ka to ca. 130 ka). Severity of continental climate reached a maximum during the late Pleistocene (130 ka to ca. 10 ka), and open-arid landscapes such as periglacial tundra and steppe dominated the environment (for review see Hurka et al., 2019).

In the late Pleistocene, glacial systems in Eurasia were reduced from west to east. Western Siberia and central Siberian lowlands as well the Ural Mountains (with the probable exception of the far most northern parts: Ehlers et al., 2013; Astakhov, 2013) remained ice-free. In northern and north-eastern Siberia, glaciation was largely restricted to the mountains only. Beringia including Alaska remained ice-free (Ehlers and Gibbard, 2011). In North America, major ice sheets repeatedly extended over large regions during the Middle Pleistocene (Illinonian, MIS 16, 12, 8 and 6) and the Late Pleistocene (Wisconsian, MIS 4-2). Most parts of Canada, with the exception of the Yukon Territory, were ice-covered. A continental ice-sheet spread from Ontario and Quebec, Canada, into the Mid-West USA in Kansas, Illinois, Indiana, Ohio, and New York (Ehlers and Gibbard, 2011). Regarding the glacial maximum, there is a striking difference between Eurasia and the North American continent. In Eurasia, glaciation reached its maximum during the penultimate ice age and

not during the last one. In North America, however, ice cover during the last and penultimate glaciation was nearly equal. The reasons for this are still unknown.

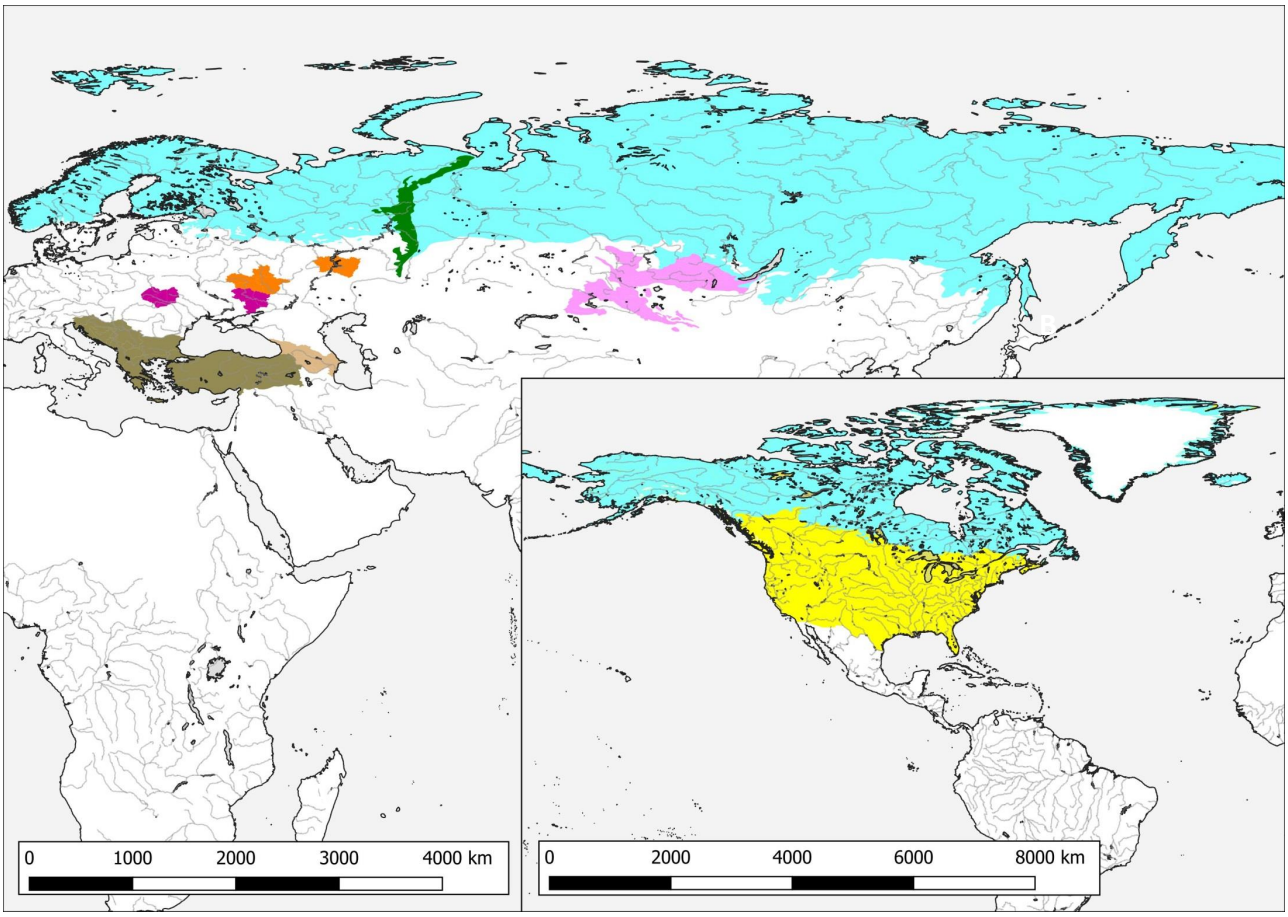


Figure 5: Areal delimitation for the ancestral area reconstruction. The colour coding corresponds to the Figures 2 and 6.

Divergence time estimation (Fig. 4) reveals a first diversification period of the *Schivereckia* subgroup at the end of the early and during the middle Pleistocene when duration of cold stages and climate severity increased and continental ice shields began to build, which had a profound effect on the biota and physical landscape. Final diversification is observed in the Late Pleistocene. The Late Pleistocene has a key position for the understanding of the present-day geographical distribution. The North American species within the *Schivereckia* subgroup seem to split off from the Eurasian species already during the early Pleistocene, but the basal branches are statistically not supported (Fig. 4). The question whether the contemporary North American species represent indeed a basal lineage, remains to be answered. Their geographic origin might well have been elsewhere migrating or being pushed later to North America.

A cold periglacial steppe hyperzone was established during the cold periods. Subarctic and steppe plants were represented almost everywhere (Simakova, 2006) including Brassicaceae as extrapolated from the proof of Brassicacean macrofossils (Kienast et al., 2005) and ancient DNA of *Draba* and other Brassicaceae in LGM records of tundra-steppe from Beringia (Willerslev et al., 2014). The periglacial hyperzone dispersed during the warm periods and was replaced by a warm-adapted interglacial vegetation and re-established during the following cold period. This happened repeatedly during the middle and late Pleistocene climatic macrocycles. The cold-adapted species persisted through interglacials in high-latitude regions known as 'polar refugia' (or northern refugia), and at lower latitudes in southern refugia, which are often the mountains. From these interglacial refugia, they spread again during the glacial periods.

From the middle Pleistocene on (ca. 850 ka), however, ice shields covered many regions of the periglacial hyperzone, which prevented expansion of species from their southern mountain interglacial refugia. The plants were trapped in their refugia and may have eventually evolved into different lineages in geographical isolation. This would explain the evolution history of *Draba hitchcockii* and *D. pterosperma* (western North American mountain chains) and *Draba ramosissima* (eastern North American mountain chains), also being in agreement with the time divergence estimation (Fig. 4). *Draba murrayi*, *D. cinerea* and *D. baicalensis* of artic/subarctic distribution survived in non-glaciated polar interglacial refugia and spread in glacial periods, and, in view of the time divergence estimation, probably experienced repeated contraction-expansion cycles. *Draba* spec. in the Caucasus is a relic species of a former periglacial steppe probably older than the last one as inferred from the time estimation. This appears also to be the case for *Schivereckia doerfleri*.

The programs written for ancestral area reconstructions were originally developed for islands, as they all assume that distribution area is an inherited trait, which is overall untrue. Nevertheless, results of our ancestral area reconstruction still support our time estimation as well as RAPD analysis and help to explain the biological history of *Schivereckia podolica*. According to the ancient area reconstruction, all members of the *Schivereckia* subgroup share a circumpolar arctic/ subarctic ancestral area (Fig. 6, node 55 with 99 % node frequency). It would appear that some species of the *Schivereckia* subgroup migrated (or were pushed) from their circumpolar ancestral distribution area to their present distribution areas. *Draba hitchcockii* and *D. pterosperma* to western North American mountain chains, *Draba ramosissima* to the eastern North American mountain chains, and *Draba* spec. to the Caucasus. Other species seem to have remained in their climate 'envelope' and arctic/ subarctic ancestral area but may have faced range shifts and

contractions (*Draba murrayi*, *D. cinerea*, *D. baicalensis*). The present-day areal of *Schivereckia doerfleri* in the Balkan Peninsula and in northern Anatolia can be traced back to the subarctic/Ural region. This applies also to the disjunct distribution pattern of *Schivereckia podolica* (Fig. 2).

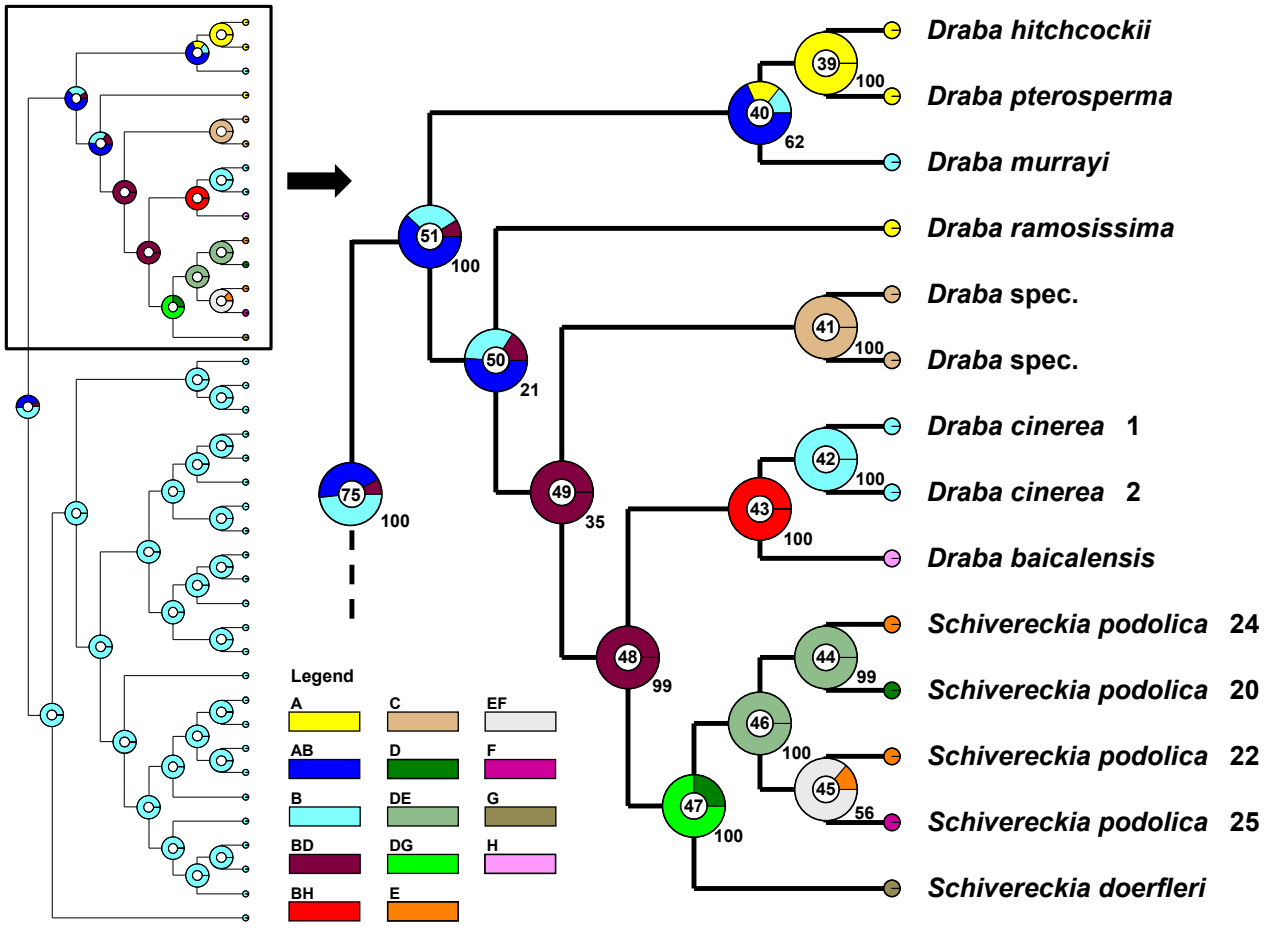


Figure 6: Ancestral area reconstruction of *Schivereckia* subgroup using the Dispersal-Extinction-Cladogenesis (DEC) method, based on the time phylogeny derived from the BEAST analysis. The colour coding corresponds to the Figures 2 and 5. The areas are coded as follows: (A) non-arctic North America, (B) subarctic regions worldwide, (C) Caucasus Mountains, (D) Ural Mountains, (E) Volga-Don region (Samara and Voronesh regions), (F) Southern Middle Russian Plateau (Donetsk and Podolia regions), (G) Balkan and Anatolia and (H) Altay-Sayan Mountains.

4.3. Historical biogeography of *Schivereckia podolica*

Refugial regions of the Eurasian mid-latitudes have been already outlined early in the last century (Lavrenko 1930; Lavrenko et al. 1952), and were supposed to be (from West to East) the Carpathians, the uplands of the Ukraine and central (European) Russia, the Ural Mountains, and the Altai-Sayan mountains. This has been substantiated by many palaeobotanical data, biome reconstructions, palaeoenvironmental modelling, and by genetic studies, yet with a focus on woody forest vegetation. There are considerable gaps in knowledge about grassland, especially steppe refugia, though. One of the main refugial features is that their locations are disjunct or limited to

only one specific region. Distribution of *Schivereckia* and its close relatives (the *Schivereckia* subgroup) match well with these patterns (disjunct distribution or 'endemic' to only one region), specifically by occurring in well-known interglacial refuge areas. *Schivereckia podolica*, e.g., occurs in the Ural Mountains and in the uplands of the Ukraine and central (European) Russia. It was already Gajeweski (1934), who pointed out that the disjunct areal parts of *S. podolica* in the East European Plain are located south of the continental ice shield of the penultimate ice ages (MIS 6, Dnepier glaciation). Even earlier glaciations such as the Don glaciation (MIS 16/15) might have had a massive impact on the distribution of *Schivereckia*, as it stretched as south as to Saratov and Volgograd (Velichko et al., 2004). This is also reflected by our time estimation analysis.

Litvinov (1902) hypothesized that species with current disjunct areas in the uplands of the southern Russia Plateau and Podolia are relics from the subarctic glacial period. This has been corroborated by Kozo-Poljanski (1928, 1929, 1931) and by Gajeweski (1934). *Schivereckia podolica* is one of these species. The distribution area of *Schivereckia podolica* was probably continuous during glacial periods and was disrupted in the interglacial warm period(s). This idea is also supported by several palaeovegetational reconstructions. The location of the disjunct part areals was apparently not much influenced by the LGM (Valdai glaciation). The predominance of open landscapes (periglacial steppe and periglacial forest steppe formations) during the Middle and Late Valdai on the Russian Plain indicates a moderately cold climate during the Briansk interstadial and a cold continental climate during the Late Valdai glaciation (Simakova, 2006). We employed RAPD markers for further analyses. As far as we are aware, it has not been proven so far, that RAPD markers compared to other fingerprint markers like SCoT, AFLP, ISSR, SSR or SNPs produce deviant or even contradictory results. RAPD data reflect the biographical history of *Schivereckia podolica*. There is a strong correlation between geographic and genetic isolation indicating isolation by distance and thus strong phylogeographic structure (Supl. 7). RAPD data confirm three main distribution ranges, the Ural Mountains, the Donetsk-Podolia region, and the Volga-Don region, which is divided into the Volga subregion (Volga-Don 2) and the Don subregion (Volga-Don 1) (Fig. 7). The Ural region was not glaciated during the LGM (see above) and might even represent an ancient areal (see Fig. 6). The Ural refugium and the Volga refugium are closer, which seems reasonable in view of their geographical adjacency, and argues for a younger disjunction age. This is corroborated by the time estimation (late Pleistocene, Fig. 4). The split between Ural/Volga and Donetsk-Podolia/ Don is of older age (middle Pleistocene, Fig. 4). In essence, the present-day disjunct part areals of *Schivereckia podolica* are interglacial cold-steppe refugia of different ages.

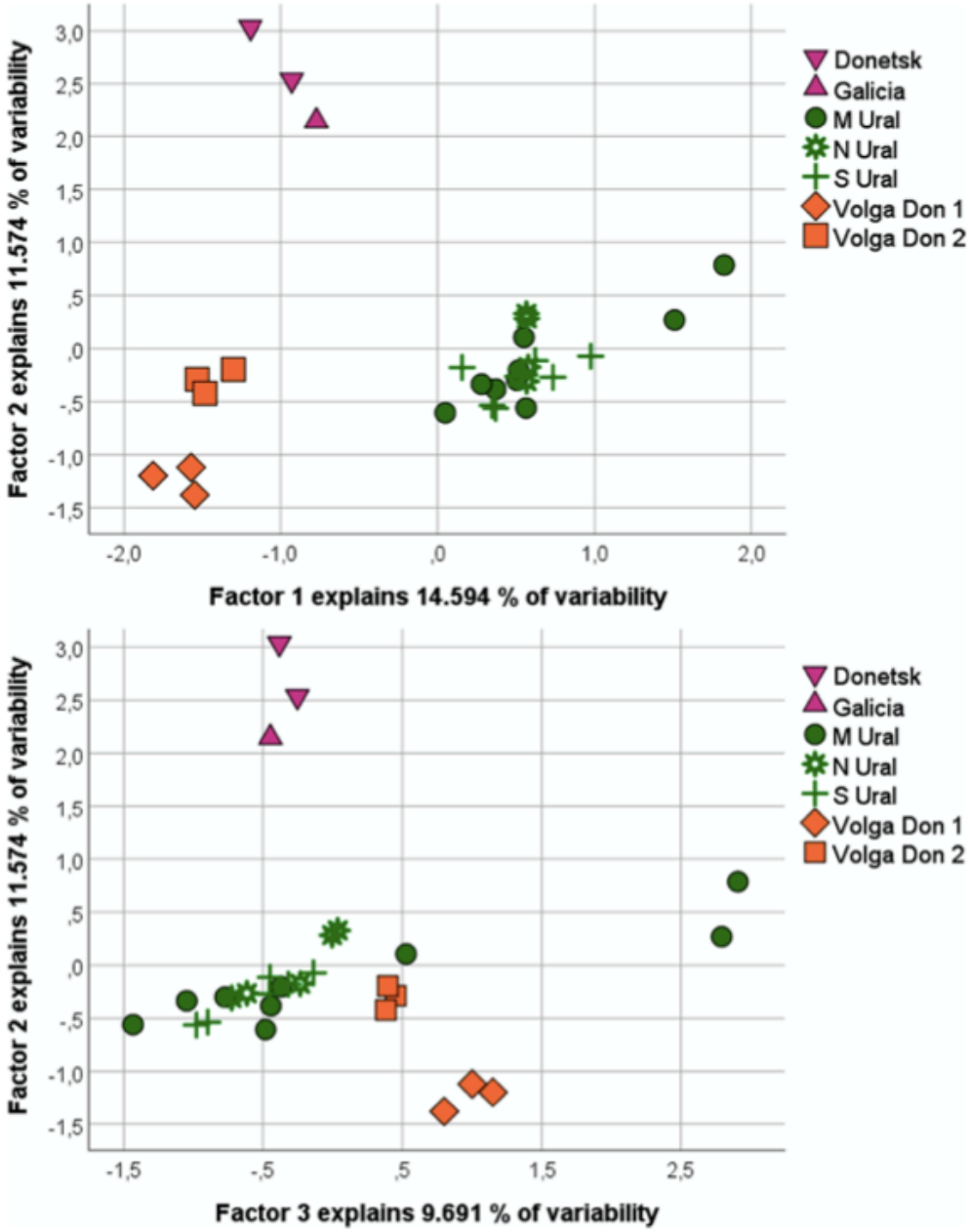


Figure 7: Principal Coordinates Analysis (PCO) of fingerprints (RAPD) data

4.4. Habitats of *Schivereckia podolica*

Can we identify ecological reasons for the disjunct distribution of *Schivereckia podolica*? *Schivereckia podolica* grows on rocks and dry slopes, particularly, but not exclusively on calcareous outcrops. In the Ural Mountains, it is also reported from clay shales and from crystalline ultrabasic igneous rocks (Kulikov et al., 2013). (see also the National Atlas of Ukraine: http:// wdc.org.ua/ atlas/en/4100200.html; where the current distribution of *S. podolica* is not restricted to limestone). Didukh and Vasheniak (2018) described a new plant association Schivereckio podolicae-Seselietum libanotidis from the limestone outcrops in Western and Central Podolia. Temperature and

cryoregime indicators suggest that these communities prefer shaded northern, northwestern and northeastern slopes. In the same study, *Schivereckia podolica* was also found in an association of Aurinio saxatilis-Allietum podolici, occupying limestone outcrops, humps, rocks and shelves. The association is further ecologically characterized as occupying protrusions of limestone deposits, where underdeveloped, shallow crushed stone lithosols are formed, allowing the growth of dry-resilient species, adapted to carbonate-rich soils.

In the Flora of Romania (Sârbu et al., 2013), *Schivereckia podolica* is assigned to the phytosociological alliance of the Potentillion caulescentis Br.-Bl. 1926. This alliance comprises rock plants/petrophytes exposed to high solar radiation. As an adaptation to water stress, such plants often display cushion-like growth and potential leaf succulence as observed in *Schivereckia podolica*, too. The Ellenberg ecological indicator values for *S. podolica* are in line with important climatic factors and substrate properties of its habitats: Heliophyte tolerating half shade (L7), heat indicator (T7), lime-indicating plant (R8), extremely resistant to draught (F2), and indicating extremely low nitrogen content (N2). It would appear that edaphic factors might have played a role in the distribution of *S. podolica*, but they are certainly not solely responsible for this peculiar disjunct distribution pattern. Present knowledge on the ecology of *Schivereckia podolica* argues for poor competitive capacity as a key factor. It grows in rather unfavourable conditions in extreme habitats such as rocks, outcrops or steep slopes, i.e. localities where competition is reduced. Most likely, the present habitat of *S. podolica* is not its original periglacial steppe habitat, but appears to be a secondary biotope into which the species was forced by competitional displacement and where it can nowadays avoid competition. Furthermore, *S. podolica* can be found only on higher altitudes (see topological map of Ukraine), which speaks in favour of an ice relic hypothesis, where it potentially survived unfavourable interglacial conditions, but later never returned into the lower altitudes, due to the competition.

Many other members of the Potentillion caulescentis alliance, which *S. podolica* belongs to, are also interpreted as ice age relics. A similar distribution is shown for example by *Artemisia rupestris* and *A. laciniata* (Hochheimer and Hoffmann, 2016). It is likely that relic taxa are nowadays associated with different species and under different environmental and edaphic conditions than in the true Pleistocene periglacial steppe, comparable to the azonal steppe islands in central Europe.

5 Conclusion

Phylogeny, differentiation and range evolution of the *Schivereckia* clade mirror the climate-landscape dynamics of the steppe from its origin in the early Pleistocene up to the Holocene. The modern structure of species assemblies (e.g. *Schivereckia* group) and genetic diversity across species' ranges was shaped over multiple glacial-interglacial cycles rather than solely by the influence of the last glaciation period. Time divergence estimation and ancestral area reconstruction support the interpretation of *Schivereckia* as a glacial relict and reject its putative Mediterranean origin.

6 Acknowledgements

We thank the curators and staff members of the following herbaria: SVER, NSK, ALTB, OSBU, MW; WU; ZT for their valuable help. We here acknowledge ‘Plantarium’ web-community members S. Glotov (Severouralsk) and A. Bronskov (Donetsk) for permission to publish photographs of living plants of *Schivereckia podolica*. This work has been supported by the Deutsche Forschungsgemeinschaft (DFG; FR 1431/8-1 to NF; NE 314/15-1 to BN and HH) and Russian Science Foundation (RSF 19-14-00071 to DG).

Compliance with ethical standards

Conflict of interest

The authors declare no conflict of interest.

Ethical approval

This article does not contain any studies with animals carried out by any of the authors.

Sampling and field studies

The study was performed in compliance with the Convention on Biological Diversity (CBD).

7 References

Alexejenko, M.I., 1946. New species of the genus *Schivereckia* Andrz. Bot. Mater. Herb. Bot. Inst. Acad. Sci. USSR, Nov. Syst. (Leningrad) 9, 215–231 (In Russian).

Alexejenko, M.I., 1950. On some endemics of the chalk flora of the European part of the USSR. Trudy Nauchn-Issled. Inst. Biol. Kharkov Univ. 13, 95–103 (In Russian).

Al-Shehbaz, I.A., Beilstein, M.A., Kellogg, E.A., 2006. Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Pl. Syst. Evol. 259, 89–120.

Applequist, W.L., 2017. Report of the nomenclature committee for vascular plants: 69. Taxon 66, 500–513.

Applequist, W.L., 2019. Report of the nomenclature committee for vascular plants: 70. Taxon 68, 847–854.

Astakhov, V.I., 2013. Pleistocene glaciations of northern Russia – a modern view. Boreas 42, 1–24.

Berkutenko, A.N., 1995. Detective story about one Linnean species of Cruciferae. Linzer Biologische Beiträge 27, 1115–1122.

Blattner, F.R., 1999. Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. BioTechniques 27, 1180–1186.

Buttler, K.P., 1967. Zytotaxonomische Untersuchungen an Mittel- und Südeuropäischen *Draba*-Arten. Mitteilungen der Botanischen Staatssammlung München 6, 275–362.

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). Mol. Biol. Evol. 27, 55–71.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9 772–772.

Didukh, Y.P., Vasheniak, Y.A., 2018. Vegetation of limestone outcrops in Western and Central Podillia (Ukraine). Tuexenia 38, 419–444.

Ehlers, J., Gibbard, P., 2011. Quaternary glaciation. In: Singh, V.P., Singh, P., Haritashya, U.K. (Eds.), Encyclopedia of Snow, Ice and Glaciers. Springer Science + Business Media BV(V), pp. 873–882.

Ehlers, J., Astakhov, V., Gibbard, P.L., Mangerud, J., Svendsen, J.I., 2013. Late pleistocene in Eurasia. In: second edition. In: Elias, S.A., Mock, C.M. (Eds.), Encyclopedia of Quaternary Science 2. Elsevier, Amsterdam, pp. 224–235.

Gajeweski, M.W., 1934. Avenetum desertorum, a monographical study. Bull. Int. Acad. Pol. Sci. Lett. Classe Sci. Math. Nat. Sér. B: Sci. Nat. I 1–27.

German, D.A., 2017. (2525) Proposal to conserve the name *Alyssum hyperboreum* (*Draba hyperborea*) with a conserved type (Cruciferae). Taxon 66, 755–756.

Head, M.J., Pillans, B., Farquhar, S., 2008. The Early–Middle Pleistocene Transition: characterization and proposed guide for the defining boundary. Episodes 31, 255–259.

Hochheimer, Ja., Hoffmann, M.H., 2016. Footprints of the last Glacial: past competition may have shaped strongly disjunct distribution ranges in *Artemisia rupestris* and *A. laciniata*. Flora 224, 203–210.

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, 5–71.

Iljinska, A., Diduch, Y., Burda, R., Korotchenko, I., 2007. In: Diduch, Y. (Ed.), Ekoflora of Ukraine Vol. 5 Phytosociocentre, Kiev 584 p.

Jordon-Thaden, I., Hase, I., Al-Shehbaz, I., Koch, M.A., 2010. Molecular phylogeny and systematics of the genus *Draba* (Brassicaceae) and identification of its most closely related genera. Mol. Phylogenet. Evol. 55, 524–540.

Jordon-Thaden, I.E., Al-Shehbaz, I., Koch, M.A., 2013. Species richness of the globally distributed, arctic-alpine genus *Draba* L. (Brassicaceae). Alp. Bot. 123, 97–106.

Karl, R., Koch, M.A., 2013. A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae. Ann. Bot. 112, 983–1001.

Kay, K.M., Whittall, J.B., Hodges, S.A., 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. BMC Evol. Biol. 6, 36.

Kazakova, M.V., 1984. Characterization of the genus *Schivereckia* Andrz. ex DC. within the eastern part European part of its distribution area. Biol. Nauki 4, 57–62.

Kiefer, M., Schmickl, R., German, D.A., Mandáková, T., Lysak, M.A., Al-Shehbaz, I.A., Franzke, A., Mummenhoff, K., Stamatakis, A., Koch, M.A., 2014. BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. Plant Cell Physiol. 55, e3.

Kienast, F., Schirrmeister, L., Siegert, C., Tarasov, P., 2005. Palaeobotanical evidence for warm summers in the East Siberian Arctic during the last cold stage. Quat. Res. 63, 283–300.

Kleopov, Ju.D., 1941. Periglazialnye stepi evropejskoj chast SSSR, in references of the famous summary of the broad-leaved forests book (reprint 1990). Naukova Dumka, Kiev 352 p (In Russian).

Kleopov, Ju.D., 1990. Analysis of the Deciduous Forests of the European Part of the USSR. K. Naukova Dumka 352 p (In Russian).

Koch, M.A., Karl, R., German, D.A., Al-Shehbaz, I.A., 2012. Systematics, taxonomy and biogeography of three new Asian genera from the Brassicaceae, tribe Arabideae: an ancient distribution circle around the Asian high mountains. Taxon 61, 955–969.

Koch, M.A., German, D.A., Kiefer, M., Franzke, A., 2018. Database taxonomics as key to modern plant biology. Trends Plant Sci. 23, 4–6.

Kozo-Poljanski, B.M., 1928. Glaziale Pflanzenrelikte auf dem Orel – Kursk Plateau im Süden der Mittelrussischen Hochebene. I. Vegetationsbilder, 19. Reihe, Heft 1/2. G. Fischer, Jena.

Kozo-Poljanski, B.M., 1929. Glaziale Pflanzenrelikte auf dem Orel – Kursk Plateau im Süden der Mittelrussischen Hochebene. II. Vegetationsbilder, 19. Reihe, Heft 7/8. G. Fischer, Jena.

Kozo-Poljanski, B.M., 1931. *Androsace villosa* L. Als Mitglied der mittelrussischen Flora. Botanische Jahrbücher für Systematik 64, 17–47.

Kulikov, P.V., Zolotareva, N.V., Podgaevskaya, E.N., 2013. In: Mukhin, V.A. (Ed.), Endemic Plants of the Urals in the Flora of the Sverdlovsk Region. Goshchitsky, Yekaterinburg Scientific Ed. 612 p. (In Russian).

Kumar, S., Stecher, G., Tamura, K., 2016. Mega 7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Litvinov, D.I., 1902. On the Relict Character of the Stony Slopes in European Russia. Trudy Botanicheskogo Muzeja Akademii Nauk, pp. 1 (In Russian).

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27, 209–220.

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). Presented at the 2010 Gateway Computing Environments Workshop (GCE), IEEE. New Orleans, LA, USA. pp. 1–8.

Mosyakin, S.L., 2015. (2404) Proposal to reject the name *Alyssum hyperboreum* (*Draba hyperborea*, *Schivereckia hyperborea*) (Brassicaceae). Taxon 64, 1326–1328.

Peakall, R., Smouse, P.E., 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6, 288–295.

Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28, 2537–2539.

Radygina, V.I., 2003. Calciphilous Flora of Middle Russia and the Volga Plateau. Dissertation. Pedagogic University, Moscow (In Russian).

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904.

Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–14.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Sârbu, I., Ştefan, N., Oprea, A., 2013. Plante vasculare din România: determinator ilustrat de teren/ Editura V. Bortaş Bucureşti: Editura Victor B Victor 1320.

Schulz, O.E., 1936. Cruciferae. In: Engler, A. (Ed.), Die natürlichen Pflanzenfamilien. Band 17b. Duncker & Humblot, Berlin, pp. 227–658.

Simakova, A.N., 2006. The vegetation of the Russian Plain during the second part of the late Pleistocene (33-18 ka). Quat. Int. 149, 110–114.

SPSS 25, 2017. IBM Corp. IBM SPSS Statistics for Windows, Version 25.0. IBM Corp., Armonk, NY.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A., 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4.

Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.

Taliev, V.I., 1897. To the Question on Relict Vegetation of the Ice Ages. Trudy Kharkovkogo Obshestva Estestvoispytatelei (In Russian).

Tolmachev, A.J., 1939. *Schivereckia*. In: In: Komarov, V.-L. (Ed.), Flora of the USSR, vol. 8. Moscow & Leningrad, pp. 649–650 (In Russian) (English Translation 1985, Koeltz Scientific Books, Koenigstein, Germany).

Turland, N.J., Wiersema, J.H., Barrie, F.R., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Knapp, S., Kusber, W.-H., Li, D.-Z., Marhold, K., May, T.W., McNeill, J., Monro, A.M., Prado, J., Price, M.J., Smith, G.F. (Eds.), 2018. International Code of Nomenclature for Algae, Fungi, and

Plants (Shenzhen Code) Adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Koeltz Botanical Books, Glashütten XXXVIII + 254 pp. [Regnum Vegetabile 159].

Tutin (Ed.), 1993. Flora Europaea 1 2nd edit. Cambridge Univ. Press, Cambridge.

Willerslev, E., Davison, J., Moora, M., et al., 2014. Fifty thousand years of arctic vegetation and megafaunal diet. Nature 506, 47–51.

Wilson, K.L., 2017. Report of the general committee: 18. Taxon 66 (3), 742–744.

Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol. Phylogenet. Evol. 87, 46–49.

Table 1. Origin, source and GenBank accession numbers of *Draba*, *Schivereckia* and outgroup accessions used for phylogenetic analyses. Herbarium acronyms according to Index Herbariorum.

Species (chromosome counts in brackets)	Provenance (general distribution in brackets)	Collector (Herbarium) or Reference to published sequences	ITS GenBank Accession No	trnL-trnF GenBank Accession No
Aubrieta deltoidea (L.) DC (2n = 16)	D, Bot. Garden Univ. Jena (mts. of Mediterraneis)	Koch et al., 1999	AJ232909	
Arabis alpina L. (2n = 16)	D (arctic-alpine, amphiatl., mts. of SW Asia, Mediterraneis, E Africa)	Koch et al., 1999	AJ232920	
Draba aizoides L. (2n = 16)	D, Bavaria, Fränkische Schweiz (mts. of C. and S. Europe; Britain)	Neuffer, Hurka, 17.5.2002 (OSBU 12710)	FM957484	LR590459
Draba altaica (C.A. Meyer) Bunge (2n = 16)	MGL, Mongolian Altai, Tsambagarev (mts. of C. Asia and Himalayas)	Neuffer, Hurka, 20.7.2000 (OSBU 10287)	FM957486	LR590458
Draba baicalensis Tolm. (2n = 48)	RUS, Altai Republic, Plateau Ukok, Ak-Alakha river valley (Altai – Yenesei - Baikal region)	Kamelin et al., 21.07.1998 (ALTB)	FM957487	LR590456
Draba cinerea Adams (2n = 32, 48)	DK, Greenland (circumpolar, arctic to subarctic; Altai – Yenisei – Baikal region)	Jordon-Thaden et al., 2010	DQ467502	
Draba cinerea Adams (2n = 32, 48)	RUS, East Siberia, Irkutsk Oblast, Mamsko-Chuisky distr. (circumpolar, arctic to subarctic; Altai – Yenisei – Baikal region)	Ivanova, 12.07.1977 (NSK)	FM957488	LR590457
Draba dubia Suter (2n = 16)	(mts. of C. and S. Europe)	Widmer et Baltisberger,1999	AF120722	
Draba eriopoda Turcz. (2n = 16)	RUS, Altai Republic, Ulagan district (mts. of S. Siberia, C. Asia and Himayalas)	German, 05.06.2002 (ALTB)	FM957503, FM957504	LR590462
Draba fladnizensis Wulfen (2n = 16, 32)	N, Jotunheimen (circumpolar, arctic and subarctic; mts. of Scandinavia, C. and S. Europe, and M. Asia – Altai- Baikal region)	Neuffer, Hurka, 25.7.1996 (OSBU 7681)	FM957505, FM957506	

Draba fladnizenisis Wulfen (2n = 16,32)	RUS, Altai Republic, Sailyugem Mts. (circumpolar, arctic and subarctic; mts. of Scandinavia, C. and S. Europe, and M. Asia – Altai- Baikal region)	Neuffer, Hurka, Friesen, 15.8.2002 (OSBU 13211)	FM957485	LR590442
D. glabella Pursh (*Draba hirta* L., nom. ambig.) (2n = 64, ca.80)	RUS, Yakutia, Olekma distr., Arga-Salaa river (circumpolar, arctic to subarctic; Altai – Yenisei – Baikal region)	Vodopyanova, 18.08.1978 (NSK)	FM957489	
Draba glabella Pursh (2n = 64, ca. 80)	D, Bot. Garden Univ. Osnabrück, origin FIN (circumpolar, arctic to subarctic; Altai – Yenisei – Baikal region)	Neuffer, 3.5.1988 (OSBU 101)	FM957501, FM957502	
Draba hispida Willd. (2n = ??)	GEO, Adscharia, Makhumseti (Asia Minor, Caucasus)	Friesen, 22.3.2004 (OSBU 15280)	FM957507, FM957508	LR590460
Draba hitchcockii Rollins (2n = 54)	(North America, Idaho, southern Rocky Mts.)	Koch & Al-Shehbaz, 2002	AF146487	AF146935; AF146986
Draba incana L. 2 (2n = 32)	CDN, Ontario (amphiatlantic, arctic to subarctic; W. and C. Europe from Pyrenees to the E. Alps)	Jordon-Thaden et al., 2010	DQ467496	
Draba incana L. 3 (2n = 32)	N, Sör-Tröndelag, Dovrefjell (amphiatlantic, arctic to subarctic; W. and C. Europe from Pyrenees to the E. Alps)	cultivated in BG Osnabrück, grown from seeds collected by Fremstad, 28.2.2002; (OSBU 19818)	LR590254	LR590427
Draba incana L. 4 (2n = 32)	A, Tyrol, Brenner Pass (amphiatlantic, arctic to subarctic; W. and C. Europe from Pyrenees to the E. Alps)	Huter, June 1882 (WHB 5638)	LR590256	LR590429
Draba incana L. 5 (2n = 32)	IS, Heimaey Island (amphiatlantic, arctic to subarctic; W. and C. Europe from Pyrenees to the E. Alps)	K.-G. Bernhardt, 18.8.1982 (Herbar K.-G. Bernhardt 2604)	LR590255	LR590428
Draba incana L. 6 (2n = 32)	N, Sör-Tröndelag, Dovrefjell (amphiatlantic, arctic to subarctic; W. and C. Europe from Pyrenees to the E. Alps)	K.-G- Bernhardt, 19.7.1986 (Herbar K.-G. Bernhardt 2605)	LR590257	LR590430
Draba incana L. 7 (2n = 32)	CH, Kt. Appenzell - Innerrhoden (amphiatlantic, arctic to subarctic; W. and C. Europe from Pyrenees to the E. Alps)	Sulger Büel, 16.5.1948 (ZT 12483)	LR590258	LR590431

Draba incana L. 8 (2n = 32)	CH, Kt. St. Gallen (amphiatlantic, arctic to subarctic; W. and C. Europe from Pyrenees to the E. Alps)	Seitter und Sulger Büel, 20.7.1953 (ZT 12481)	LR590259	LR590432
Draba kusnetzovii (Turcz.) Hayek (2n = ??)	RUS, Altai Republic, Sailjugem Mt. Range (Yenisey – South Siberia, Russian Far East; Mongolia)	Neuffer, Hurka, Friesen, 13.8.2002 (OSBU 13671)	FM957509, FM957510	LR590443
Draba ladina Braun-Blanq. (2n = 32)	CH, C. Alps (mts. of E. Switzerland)	Widmer et Baltisberger,1999	AF120723	
Draba lanceolata Royle (2n = 32)	RUS, Altai Republic, Samakhu steppe (mts. of M. and C. Asia and Himayalas; Russian Far East, Kamchatka)	Smirnov, 2.6.2005 (OSBU 18564)	FM957511, FM957512	LR590441
Draba lanceolata Royle (2n = 32)	RUS, Altai Republic, confluence of Tschuja and Katun (mts. of M. and C. Asia and Himayalas; Russian Far East, Kamchatka)	Neuffer, Hurka, Friesen, 19.8.2002 (OSBU 13460)	FM957499	
Draba mongolica Turcz. (2n = ??)	RUS, East Sayan, Munku-Sardyk range, Nukhu-Daban plateau (mts. of S. Siberia, Mongolia, China)	Malyshev, 06.08.1983 (NSK)	FM957490	LR590444
Draba murrayi G.A. Mulligan (2n = 48)	(North America, Yukon, Alaska)	Jordon-Thaden et al., 2010	DC467603	DQ467178
Draba nemorosa L. (2n = 16)	RUS, Tverskaja Oblast (throughout Eurasia and N. America)	Neuffer, 29.5.1998 (OSBU 8800)	FM957491	LR590461
Draba ochroleuca Bunge (2n = ca. 80)	MGL, Gobi Altai, Igd Bogd (M. Asia – Altai – Yenisei – Baikal region; Russian Far East; Mongolia)	Neuffer, Hurka, Friesen, 23.8.2001 (OSBU 12044)	FM957513, FM957514	LR590445
Draba oreades Schrenk (2n = 40)	MGL, Mongolian Altai, Tavan Bogd (mts. of M. and C. Asia and Himayalas)	Neuffer, Hurka, 25.7.2000 (OSBU 10433)	FM164651	LR590446
Draba pterosperma Payson	(North America, northern California)	Koch & Al-Shehbaz, 2002	AF146493	
Draba ramosissima Desvaux (2n = 16)	(North America, southern Appalachians)	Jordon-Thaden et al., 2010	DQ467609	DQ467184
Draba scabra C.A. Meyer (2n = ??)	RUS, Caucasus, Rep. Karatschaj-Tscherkessia (Caucasus)	Neuffer, Hurka, 22.7.1999 (OSBU 9134)	FM957519, FM957520	

Draba sibirica (Pallas) Thell. (2n = 16)	RUS, Altai Rep., Yushno Tschuiski Mt. range (Europe, Caucasus, M. Asia, S. Siberia, Russian Far East, Mongolia, W. China)	Neuffer, Hurka, Friesen, 9.8.2002 (OSBU 12973)	FM957492	LR590447
Draba siliquosa Bieb. (2n = 16)	RUS,Caucasus, Rep. Karatschaj-Tscherkessia (mts. of C. and S. Europe; Caucasus)	Neuffer, Hurka, 24.7.1999 (OSBU 9164)	FM957521, FM957522	
Draba stenocarpa J.D. Hooker & Thomson (2n = ca. 40)	RUS, Altai Republic, Kosh-Agatch district (mts. of M. Asia, W. China, W. Mongolia, SW Siberia, and Himayalas)	Smirnov, German & al. 2.6.2005 (OSBU 18563)	FM957515, FM957516	LR590448
Draba spec. 1 (2n = ??)	RUS, Caucasus, Rep. Karatschaj-Tscherkessia, Teberda (Caucasus)	Neuffer, Hurka, 19.7.1999 (OSBU 9048)	FM957493	LR590433
Draba spec. 2 (2n = ??)	RUS, Caucasus, Rep. Kabardino-Balcaria, north slope of Elbrus (Caucasus)	T.Popova, Ju. Menzik, 29.7.1981 (MW)	LR590260	LR590434
Draba stylaris J. Gay 3 (2n = 32)	CH, Kt. Graubünden, Silvretta, Fimbertal (Alps)	Hurka, 29.7.1976 (Herbar H. Hurka 502)	LR590261	LR590435
Draba stylaris J. Gay 4 (2n = 32)	CH, Kt, Graubünden, Silvretta, Fimbertal (Alps)	K.-G. Bernhardt, Juni 1980 (Herbar K.-G. Bernhardt 2597)	LR590262	LR590436
Draba stylaris J. Gay 5 (2n = 32)	I, Prov. Bolzano, Dolomites (Alps)	Handel-Mazzetti, 12.7.1902 (WU OP131/3)	LR590263	LR590437
Draba stylaris J. Gay 6 (2n = 32)	A, Tyrol, near Reschen Pass (Alps)	Handel-Mazzetti, 2.8.1911 (WU OP131/2)	LR590264	LR590438
Draba stylaris J. Gay 7 (2n = 32)	A, Tyrol, Silvretta, Fimbertal (Alps)	Handel-Mazzetti, 24.7.1911 (WU OP131/1)	LR590265	LR590439
Draba stylaris J. Gay 8 (2n = 32)	CH, Kt. Graubünden, Oberengadin (Alps)	Huber, Frey, 9.8.1991 (ZT 12492)	LR590266	LR590440
Draba subamplexicaulis C.A. Meyer (2n = 48)	RUS, Altai Republic, Plateau Ukok, Teply Klyuch (mts. of C. Asia)	Smirnov, 08.2003 (ALTB)	FM957523, FM957524	LR590449

Draba subcapitata Simmons (2n = 16)	RUS, N. Siberia, Taimyr, Sandasko river, Khatangsky Bay (amphiatlantic, arctic)	Vodopyanova, Friesen et al., 23.07.1979 (NSK)	FM957517, FM957518	
Draba tomentosa Clairv. (2n = 16)	CH, C. Alps (mts. of C. and S. Europe)	Widmer et Baltisberger,1999	AF120725	
Draba turczaninovii Pohle et N. Busch (2n = 48)	RUS, Altai Republic, S.-Chuisky Mt. Range, Sebystei river (mountains of Kazakhstan, Kyrgystan Mongolia, Russia and western China)	Khrustaleva, 8.07.2001 (ALTB)	FM957494	LR590450
Schivereckia doerfleri 27 (2n = 16)	D, Bot. Garden Univ. Osnabrück (seeds from Serbia)	Neuffer (OSBU 142) see Table 2	FM164650	LR590451
Schivereckia podolica 25 (2n = 16)	UA (DB): Donetsk Oblast, Gorodeshchina, Khutor Zajachji	MW0372780 12.05.1963 see Table 2	FM957498	LR590455
Schivereckia podolica 20 (2n = 16)	RUS (M Ural): Perm Krai, 125 km northeast of Perm; Tschusovskoy Raion, close to village Malye Voronki near Paschija	(MW) see Table 2	FM957495	LR590452
Schivereckia podolica 24 (2n = 16)	RUS (Volga-Don): Uljanovskaja Oblast, 70 km west of Simbirsk; 8 km south of Schilovka village, rocky steppe, chalk	MW0372747 Maslennikov 10.05.1992 see Table 2	FM957497	LR590454
Schivereckia podolica 22 (2n = 16)	RUS (Volga-Don): Lipetskaja Oblast, 110 km west of Lipetsk; near Vorgol, 35 km northwest of Elez, calcareous outcrops in little river valley	MW0372696 Tichomirov V., Jusuposa L. 13.06.1974 see Table 2	FM957496	LR590453

Table 2. Origin of *Schivereckia* accessions used for fingerprint analysis. (1) Taxon Abbreviation: *S. berteroides*: *Schivereckia berteroides* Fisch. ex Alexejenko; *S. hyperborea*: *Schivereckia hyperborea* (L.) Berkut.; *S. monticola*: *Schivereckia monticola* Alexejenko. (2) Geographic abbreviations: D: Germany; UA: Ukraine; RUS: Russia; (3) Herbarium labels do not record coordinates. Georeferencing was done according to the details of the collecting points given on the labels.

Acc No	Taxon **(1)**	Provenance **(2)**	Collector (Herbarium)	Geographic coordinates **(3)**
1	*Schiv. podolica* (as *S. hyperborea*)	RUS (N Ural): Sverdlovskaya Oblast, 425 km north of Jekaterinburg; 16 km northwest of Ivdel, confluence of rivers Taltija and Ivdel; calcareous rocks	Knyasev 11.07.94 (SVER)	60° 44' N 60° 10' E
2	*Schiv. podolica* (as *S. hyperborea*)	RUS (N Ural): Sverdlovskaya Oblast, 490 km north of Jekaterinburg; ca. 50 km north of Ivdel; at river Talitsa near Burmantovo	Knyasev 4.07.94 (SVER)	61° 17' N 60° 26' E
3	*Schiv. podolica*	RUS (N Ural): Sverdlovskaya Oblast, 430 km north of Jekaterinburg; 25 km northwest of Ivdel; upper reaches of river Losva, Tosemskie Kamen	Knyasev 11.06.95 (SVER)	60° 50' N 60° 03' E
4	*Schiv. podolica* (as *S. hyperborea*)	RUS (M Ural): Sverdlovskaya Oblast, 50 km west of Jekaterinburg; at river Chusovaya, near railway station of Kourovka	Storozheva 15.05.77 (SVER)	56° 57' N 59° 44' E
5	*Schiv. podolica* (as *S. berteroides*)	RUS (S Ural): Tshelyabinskaya Oblast, border to Bashkortostan Rep., 190 km east of Ufa; at river Ay, 10 km upriver from village Mezhevoi	Knyasev 9.05.85 (SVER)	55° 14' N 58° 55' E
6	*Schiv. podolica* (as *S. berteroides*)	RUS (S Ural): Bashkortostan Rep., 130 km northeastern of Ufa; ca. 5 km from village Urmantau, at river Yuryusan (Argazi)	Knyasev 12.05.85 (SVER)	55° 26' N 57° 44' E
7	*Schiv. podolica* (as *S. berteroides*)	RUS (S Ural): Bashkortostan Rep., 180 km southeast of Ufa; confluence of rivers Kungur and Belaya, between Usyan and Kaga	Knyasev 29.04.83 (SVER)	53° 33' N 57° 44' E
8	*Schiv. podolica* (as *S. monticola*)	RUS (M Ural): Sverdlovskaya Oblast, 225 km north of Jekaterinburg; at river Tura, Dyrovatyi rocks, near Verchoturskie Leschos	Not available 28.06.77 (SVER)	58° 51' N 60° 49' E
9	*Schiv. podolica* (as *S. borealis*)	RUS (M Ural): Sverdlovskaja Oblast, 90 km east of Jekaterinburg; at river Pyzhma near Rudjanskoe	Knyasev 12.06.93 (SVER)	56° 58' N 61° 56' E
10	*Schiv. podolica* (as *S. borealis*)	RUS (M Ural): Perm Krai, 200 km north east of Perm; at river Yaiva, near village Suchaya, rocks at confluence of rivers Voronicha and Yaiva	Not available 10.06.94 (SVER)	59° 33' N 57° 47' E
11	*Schiv. podolica* (as *S. monticola*)	RUS (S Ural): Bashkortostan Rep., 140 km north west of Perm, 55 km west of Dyurtyuli, Karabash Mt.	Kler, Vvedensky 16.06.1911 (SVER)	55° 24' N 54° 11' E

12	*Schiv. podolica* (as *S. monticola*)	RUS (M Ural): Sverdlovskaya Oblast, 200 km north of Jekaterinburg; Kamen Dyrovatyi near Nishnaya Tura at river Tura	Not available (SVER)	58° 38’ N 59° 53’ E
13	*Schiv. podolica*	RUS (S Ural): Tshelyabinskaya Oblast; 150 km east of Ufa, at river Yuryusan, 5 km east St. Ust-Katav, calcareous rocks at confluence of rivers Minka and Yuryusan	Knyasev 11.05.84 (SVER)	54° 56’ N 58° 07’ E
14	*Schiv. podolica*	RUS (M Ural): Sverdlovskaya Oblast, 210 km north of Jekaterinburg; rocks at river Tura near village Elkino, 7 km north of Nishnaja Tura	Knyasev 28.07.82 (SVER)	58° 42’ N 59° 50’ E
15	*Schiv. podolica* (as *S. monticola*)	RUS (M Ural): Sverdlovskaya Oblast, 120 km northeast of Jekaterinburg, 20 km southeast of Alapayevsk; from Koptelovo at river Neyva 4 km south, valley of river Rezh, rocks	Not availabe 24.07.82 (SVER)	57° 40’ N 61° 50’ E
16	*Schiv. podolica* (as *S. berteroides*)	RUS (M Ural): Sverdlovskaya Oblast, 110 km northwest of Jekaterinburg; near village Volegova at river Chusovaya, calcareous rocks	Grüner 14.05.49 (SVER)	57° 27’ N 59° 08’ E
17	*Schiv. podolica* (as *S. berteroides*)	RUS (S Ural): Bashkortostan Rep., 130 km northeast of Ufa; upper reaches of river Yuryusan, 5 km upriver Urmantau, Baganatasch tretiy. rocks	Knyasev 12.05.85 (SVER)	55° 26’ N 57° 44’ E
18	*Schiv. podolica*	RUS (N Ural): Komi Rep., 400 km northeast of Syktyvkar, Ural; on the right of the river Ylych, south faced rocky slopes	Govorukhin 12.07.25 MW0372691	62° 30’ N 58° 57’ E **(4)**
19	*Schiv. podolica*	RUS (N Ural): Komi Rep., 400 km northeast of Syktyvkar, Ural; valley of river Ylych, rocky outcrops on the shore	Schyschma 29.07.1940 (MW)	62° 30’ N 58° 57’ E **(4)**
20	*Schiv. podolica*	RUS (M Ural): Perm Krai, 125 km northeast of Perm; Tshusovskoy Raion, close to village Malye Voronki near Paschiya	Not available (MW)	58° 24’ N 58° 11’ E
21	*Schiv. podolica*	RUS (Volga-Don): Lipetskaya Oblast, 40 km west of Lipetsk; near Donskoye, Galichiya Gora reserve, calcareous rocks	MW0372732 Tichomirov V. 20.04.1990	52° 36’ N 38° 55’ E
22	*Schiv. podolica*	RUS (Volga-Don): Lipetskaya Oblast, 110 km west of Lipetsk; near Vorgol, 35 km northwest of Elez, calcareous outcrops in little river valley	MW0372696 Tichomirov V.,Jusuposa L. 13.06.1974	52° 42’ N 37° 59’ E
23	*Schiv. podolica*	RUS (Volga-Don): Lipetskaya Oblast, 50 km northwest of Lipetsk; Krasnoye raion, right-hand side slopes of the Don valley, near village Zasosenki, calcareous rocks	MW0372721 Tikhomirov, Notov, Polevova 7.07.1987	52° 44’ N 38° 54’ E
24	*Schiv. podolica*	RUS (Volga-Don): Ulyanovskaya Oblast, 70 km west of Simbirsk; 8 km south of Shilovka village, rocky steppe, chalk	MW0372747Maslennikov10.05.1992	54° 13’ N 47° 19’ E

25	*Schiv. podolica*	UA (DB): Donetskaya Oblast, 90 km north of Donetsk; Gorodeshchina, Khutor Zayachyi, near Kramatorsk	MW0372780 Not available 12.05.1963	48°41'N 37°27'E
26	*Schiv. podolica*	UA (Podolia): Chmelnizka Oblast, 80 km south west of Chmelnizka; 15 km north of Kamyanez – Podolskie, valley of river Smotrich	MW0372779Khrunkievich June 1935	48° 47' N 26° 33' E
27	*Schiv. doerfleri*	D, Botanic Garden Univ. Osnabrueck, grown from seeds from Serbia	Neuffer B. (OSBU 142)	
28	*Schiv. podolica*	RUS (Volga-Don): Samarskaya Oblast, 50 km north west of Samara; near village Morkvash, Zhiguli hills, calcareous slopes	Not available 30.05.1903 (KAZ)	53° 25' N 49° 32' E
29	*Schiv. podolica*	RUS (Volga-Don): Samarskaya Oblast, 60 km north west of Samara; Zhiguli hills near Zhiguli, calcareous slopes	Janischevskii1906 (KAZ)	53° 20' N 49° 18' E
30	*Schiv. podolica*	UA (DB): Donetska Oblast, 100 km north of Donetsk; valley of the Donets river near village Serebryanka, chalk hills	Taliev 2.05.1903 (KAZ)	48° 55' N 38° 07' E

Table 3. Number of evaluated RAPD marker bands within experimental groups. Markers occurring = markers occurring at least once in one group; common in all = markers shared with other experimental groups; private markers = markers occurring only within the experimental group; common private markers = occurring in a frequency higher than 50% of individuals.

	Markers occurring	Common in all	Private markers	Common private markers
All individuals (27)	120	3		
Donetsk / Podolia	40	31	9	4
Ural	84	43	41	4
Volga-Don 1 + 2	54	41	11	3
Volga-Don 2 + Ural	89	24	3	0

Table 4: Time estimation statistics of the Arabideae, *Draba* s.l. and *Schivereckia* subgroup based on different tree priors. The age spans correspond to the crown ages, unless stated otherwise. The three taxon datasets are in bold and underlined. The clades, whose time spans are described in the table are statistically well-supported and named either after taxon names or after the predominant distribution areas of the species belonging to the clade. The letters in the second column correspond to the letters next to the nodes in the Figure 3. Legend: (OG) outgroup, HPD highest posterior density.

Taxon	**Node name (Fig. 3)**		**Tree Prior**	
		Yule	**Coalescence**	**Birth-Death**
Arabideae				
root age		8.34-14.02 MYA 95% HPD	20.85-46.06 MYA 95% HPD	11.68-23.77 MYA 95% HPD
Clausia & *Arabidopsis* (OG)		3.52-10.38 MYA 95% HPD	14.61-42.48 MYA 95% HPD	6.1-20.45 MYA 95% HPD
Aubrieta		1.36-4.18 MYA 95% HPD	1.40-6.86 MYA 95% HPD	1.57-5.44 MYA 95% HPD
Arabis s.l.		6.27-10.13 MYA 95% HPD	11.59-22.06 MYA 95% HPD	9.39-16.47 MYA 95% HPD
Arabis	A	0.29-2.53 MYA 95% HPD	0.21-2.20 MYA 95% HPD	0.21-2.24 MYA 95% HPD
Tomostima group	B	3.30-6.36 MYA 95% HPD	4.87-10.83 MYA 95% HPD	3.99-8.00 MYA 95% HPD
Asia Minor	C	3.89-6.10 MYA 95% HPD	4.61-8.77 MYA 95% HPD	4.55-7.42 MYA 95% HPD
Subarctic + Eurasia	D	2.50-4.30 MYA 95% HPD	2.53-5.10 MYA 95% HPD	2.52-4.55 MYA 95% HPD
***Schivereckia* group**	**E**	**1.75-3.82 MYA 95% HPD**	**1.64-4.19 MYA 95% HPD**	**1.72-3.98 MYA 95% HPD**
Asia, Subarctic + Pacific mountain chain, South America	F	2.67-4.64 MYA 95% HPD	2.85-5.16 MYA 95% HPD	2.50-4.79 MYA 95% HPD
***Draba* s.l.**				
root age		5.41-9.98 MYA 95% HPD	12.85-50.09 MYA 95% HPD	5.69-11.65 MYA 95% HPD
Arabis (OG)	A	0.16-1.92 MYA 95% HPD	0.11-2.11 MYA 95% HPD	0.18-1.92 MYA 95% HPD
Tomostima group	B	3.09-6.20 MYA 95% HPD	5.25-15.01 MYA 95% HPD	3.24-6.82 MYA 95% HPD
Asia Minor	C	3.08-5.49 MYA 95% HPD	4.89-11.1 MYA 95% HPD	3.19-5.92 MYA 95% HPD
Subarctic + Eurasia	D	2.18-3.87 MYA 95% HPD	2.54-5.73 MYA 95% HPD	2.20-4.02 MYA 95% HPD
***Schivereckia* group**	**E**	**1.60-3.47 MYA 95% HPD**	**1.61-4.97 MYA 95% HPD**	**1.65-3.63 MYA 95% HPD**

Asia, Subarctic + Pacific mountain chain, South America	F	2.37-4.24 MYA 95% HPD	2.56-6.06 MYA 95% HPD	2.38-4.35 MYA 95% HPD
Schivereckia				
root age		1.85-3.87 MYA 95% HPD	2.32-6.12 MYA 95% HPD	1.92-4.16 MYA 95% HPD
Subarctic + Eurasia	D	1.22-2.96 MYA 95% HPD	1.26-4.17 MYA 95% HPD	1.30-3.14 MYA 95% HPD
***Schivereckia* group**	**E**	**1.26-2.91 MYA 95% HPD**	**1.42-4.07 MYA 95% HPD**	**1.32-3.13 MYA 95% HPD**

PUBLICATION III:

A STORY FROM THE MIOCENE:

CLOCK-DATED PHYLOGENY OF *SISYMBRIUM* L. (SISYMBRIEAE, BRASSICACEAE)

ŽERDONER ČALASAN A, GERMAN DA, HURKA H, NEUFFER B.

ECOLOGY AND EVOLUTION (2021) 11: 2573–2595.

DOI: 10.1002/ece3.7217

A story from the Miocene:

Clock-dated phylogeny of *Sisymbrium* L. (Sisymbrieae, Brassicaceae)

Anže Žerdoner Čalasan[a,*], Dmitry A. German[b], Herbert Hurka[a], Barbara Neuffer[a]

[a] Department 5: Biology/Chemistry, Botany, University of Osnabruck, Barbarastraße 11, 49076 Osnabruck, Lower Saxony, Germany

[b] South-Siberian Botanical Garden, Altai State University, Lenina 61, 656049 Barnaul, Russia

* Corresponding author. *E-mail address*: azerdonercal@uos.de

Short title/Running Head: Clock-dated phylogeny of Sisymbrieae

Abstract

Morphological variability and imprecise generic boundaries have hindered systematic, taxonomical and nomenclatural studies of *Sisymbrium* L. (Brassicaceae, Sisymbrieae DC.). The members of this almost exclusively Old-World genus grow mostly on highly porous substrates across open steppe, semi-desert or ruderal habitats in the temperate zone of the Northern Hemisphere and African subtropics. The present study placed the biological history of *Sisymbrium* L. into time and space and rendered the tribus Sisymbrieae as monotypic. Five nuclear encoded and three chloroplast-encoded loci of approximately 85% of all currently accepted species were investigated. Several accessions per species covering their whole distribution range allowed for a more representative assessment of intraspecific genetic diversity. In the light of fossil absence, the impact of different secondary calibration methods and taxon sets on time spans was tested, and we showed that such a combinatorial nested dating approach is beneficial. Multigene phylogeny accompanied with a time divergence estimation analysis placed the onset and development of this tribus into the western Irano-Turanian floristic region during the Miocene. Continuous increase in continentality and decrease in temperatures promoted the diversity of the Sisymbrieae, which invaded the open grasslands habitats in Eurasia, Mediterranean and South Africa throughout the Pliocene and Pleistocene. Our results support the assumption of the Irano-Turanian region as a biodiversity reservoir for adjacent regions.

Key words: florogenesis, Eurasian steppe, dated phylogeny, *Sisymbrium*, Brassicaceae, taxonomy

1 Introduction

Sisymbrium L. (Brassicaceae, Sisymbrieae DC.) belongs to one of the most well-studied and economically important plant families in the world, including important vegetables and forage crops, condiments, decorative plants and model organisms. Nevertheless, morphological homogeneity within the genus on the one hand and extreme intraspecific morphological variability and lack of precise generic boundaries on the other hand have hindered systematic studies of *Sisymbrium* until the last decade (Al-Shehbaz, 2006b; Warwick et al., 2002, 2006). Much effort has been put into elucidating the actual number of *Sisymbrium* species, starting with the first worldwide monographic study by Fournier as early as 1865. His generic concept, however, was based very broadly and is nowadays considered outdated (Warwick et al., 2006). Several other taxonomic studies on *Sisymbrium* followed including Payson (1922) and Schulz (1924, 1928). Schulz's circumscription of the genus was based on non-apomorphic single characters only, including linear and terete fruits (siliques) and seeds with incumbent cotyledons. Thus, it was criticised in subsequent studies on *Sisymbrium* by Rollins (1943) and Romanczuk (1981, 1982). It was Rollins (1982) who first regarded *Sisymbrium* as an exclusively Old World taxon. Nevertheless, under current circumscription *Sisymbrium* indeed includes one single non-Eurasian species, namely *Sisymbrium linifolium* Nutt. Its distribution area was believed to be limited to the dry habitats of the western USA, but it was also recently discovered in north-western China (Chen et al., 2019). While being highly controversial in the past, the generic limits of *Sisymbrium* are nowadays well-understood and generally accepted (following Rollins, 1982; Al-Shehbaz, 2012, 2015) with a notable exception of *Sisymbrium aculeolatum* Boiss. and *S. afghanicum* Gilli. (Warwick et al., 2003). While these two taxa have been formally moved from *Sisymbrium* to *Neotorularia* Hedge & Léonard (Léonard, 1986), there was only little justification for such a transfer (Appel & Al-Shehbaz, 2002). Furthermore, the whole plastome phylogeny placed *Sisymbrium aculeolatum* next to Sisymbrieae (Walden et al., 2020). Hitherto it remains unclear, whether the two aforementioned species belong to the genus *Sisymbrium* or not.

Together with the monotypic *Ochthodium* (Nikolov et al., 2019; Walden et al., 2020) *Sisymbrium* belongs to the tribus Sisymbrieae, which shows a close relationship with the other tribes in the Lineage II of the Brassicaceae phylogeny – Brassiceae, Isatideae and Thelypodieae (BrassiBase: Kiefer et al., 2014; Koch et al., 2012, 2018). Currently, there is no consent neither on which of these tribes are sister group to the Sisymbrieae, nor which of the other clades build a sister group relationship within the Lineage II (BrassiBase: https://brassibase.cos.uni-heidelberg.de). A number

of single, multi-locus and whole genome studies recovered contradicting sister group relationships with Sisymbrieae (Al-Shehbaz et al., 2006; Beilstein et al., 2010; Couvreur et al., 2010; Franzke et al., 2011; German et al., 2009; Huang C.-H. et al., 2016; Huang X.-C. et al., 2020; Liu et al., 2020; Nikolov et al., 2019; Walden et al., 2020; Warwick et al., 2010).

Warwick et al. (2002) published the last and the most comprehensive phylogeny of *Sisymbrium* L. based on nuclear genes, which also served as an anchor point for our study. However, over the past two decades, new species of *Sisymbrium* have been discovered and validly described (Al-Shehbaz, 2015; Blanca et al., 2015; Mutlu and Karakuş, 2015), several taxonomic revisions took place (Al-Shehbaz, 2004a, 2004b, 2006a, 2006b; German and Al-Shehbaz, 2018; Warwick and Al-Shehbaz, 2003), and the shortcomings of the use of single potentially paralogue markers have been uncovered (Koch et al., 2007; Pirie et al., 2007). No comprehensive chloroplast-based *Sisymbrium* phylogeny has been published until now. Furthermore, only scarce genetic data of individual chloroplast regions (Arias and Pires, 2012; Hall et al., 2002; Koch et al., 2007; Warwick et al., 2004, 2006) or whole plastomes (Hohmann et al., 2015; Nikolov et al., 2019; Walden et al., 2020) are available, preventing a comprehensive comparative study. Out of 514 different *Sisymbrium* names that can be found in the literature, currently only 50 of them represent accepted species and subspecies, leaving out 466 synonyms and one name with an unresolved status (Kiefer et al., 2014; Koch et al., 2018).

All *Sisymbrium* representatives exhibit incumbent cotyledons; often pinnately-divided never clasping leaves; indumentum, when developed, of simple trichomes (with the only exception of South African *S. burchellii*, having branched trichomes – Marais, 1970); yellow (or rarely white or pinkish) flowers; two-lobed stigmas; cylindrical fruits with three-veined valves, and seeds that are arranged in a single row (Al-Shehbaz, 2015, Figure 1). In terms of chromosome number, *Sisymbrium* is rather monomorph with x = 7, seldom x = 8 (Al-Shehbaz, 2015). Most of the species exhibit only one ploidy level i.e. diploid with occasional tetraploid exceptions, such as *Sisymbrium luteum*, *S. polyceratium* and *S. strictissumum* (all: 2n = 4x = 28) or hexaploid *S. runcinatum* (Al-Shehbaz, 1988). *Sisymbrium irio* is the most diverse in terms of ploidy (Khoshoo, 1959). While it is mostly diploid throughout its native and naturalised ranges of temperate Eurasia, Mediterranean region and both Americas, tetraploids, hexaploids and octoploids have been reported from the Afghanistan and Indian subcontinent (Khoshoo, 1959; Podlech et al., 1969). Most of the weedy species are self-compatible and autogamous. Some species such as *Sisymbrium irio*, *S. officinale* and *S. orientale*, however, develop self-compatible protogynous flowers (Al-Shehbaz, 1977; Khoshoo, 1959).

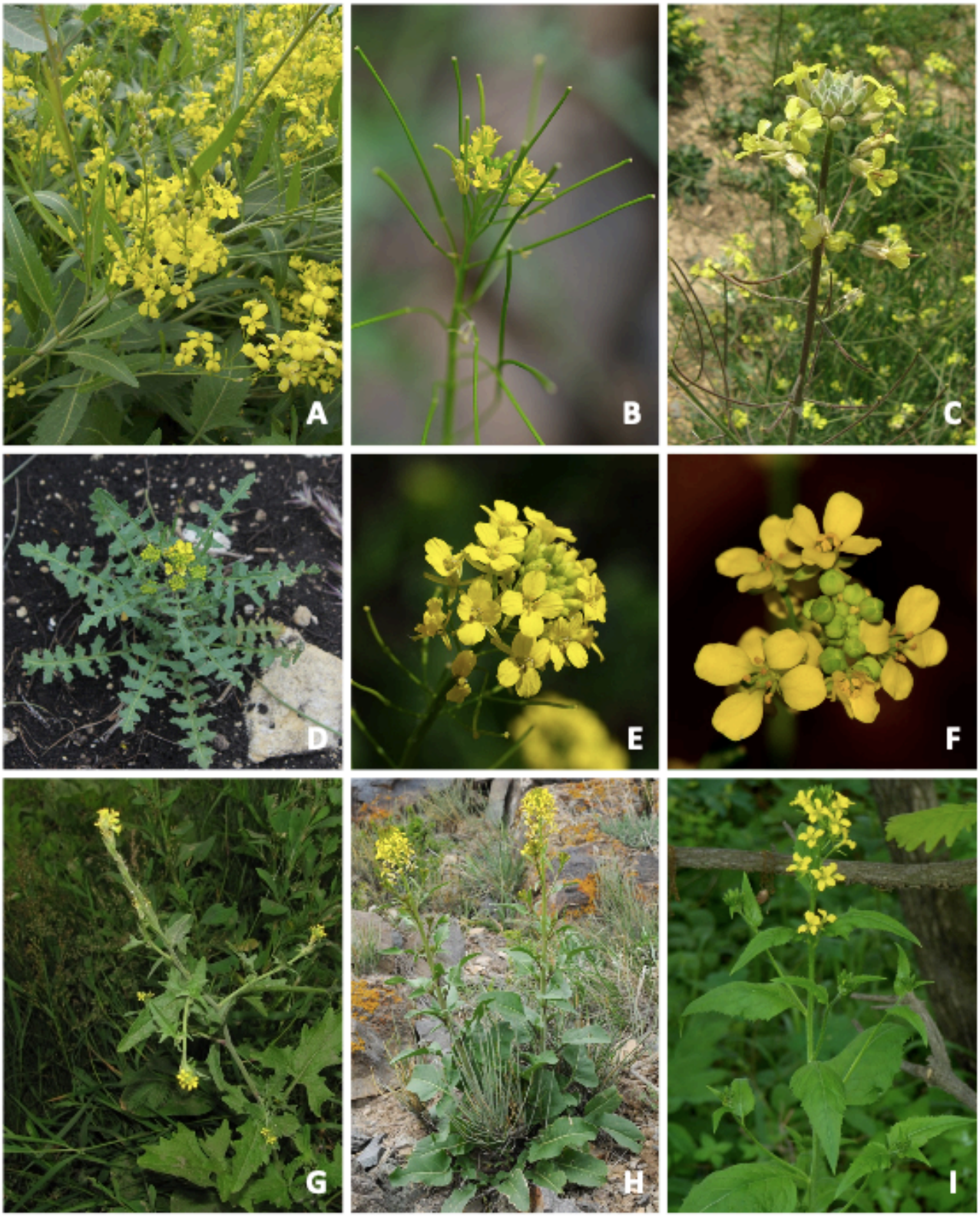


Figure 1: Habitus of different Sisymbrieae representatives. A – *Sisymbrium volgense* (Irkutsk region; June 27th 2012; by A. Manchenko), B – *Sisymbrium heteromallum* (Mongolia, Khovd aimak, Jargalant-Khairkhan; June 25th 2015; by P. Kosachev), C – *Sisymbrium orientale* (Ukraine, Crimea; May 17th 2013; by A. Fateryga), D – *Sisymbrium lipskyi* (Karachay-Cherkesskia, Malokarachaevsky district; June 15th 2013; by I. Tabunova), E – *Sisymbrium loeselii* (Krasnodar Territory, Novorossiysk, Abrau-Dyurso; June 3rd 2009; by V. Gumenyuk), F – *Ochthodium aegyptiacum* (Israel, Lower Galilee, Upper Nazareth; March 25th 2014; author of the photography wanted to remain anonymous), G – *Sisymbrium officinale* (Yaroslavl; June 24th 2012; by E. Zakharov), H – *Sisymbrium brassiciforme* (Kazakhstan,

Kapchagai; April 17th 2019; by V. Epiktetov), I – *Sisymbrium luteum* (Primorsky Territory, Nakhodka; June 14th 2012; by V. Volcotrub).

Sisymbrium is characterised by variable glucosinolate profiles and presence of cardioactive steroid glycosides (cardenolides), which are otherwise extremely scarce in the Brassicaceae (Al-Shehbaz, 1988; Fahey et al., 2001). One of notable exceptions to this rule is the genus *Erysimum* (Züst et al., 2020), which superficially morphologically resembles *Sisymbrium*. A number of medicinal properties have been attributed to several *Sisymbrium* species (one of the most prominent examples is *S. officinale*, hence the name), due to their analgesic, antibacterial, diuretic and myorelaxant activity (Al-Shehbaz, 1988; Blažević et al., 2010; Di Sotto et al., 2010).

The centre of *Sisymbrium* diversity is the Irano-Turanian region (sensu Takhtajan et al., 1986), followed by temperate regions of Europe and the Mediterranean (http://www.plantsoftheworldonline.org/). Only a handful of species can be found in non-Mediterranean North as well as in South Africa and a single already above-mentioned species grows in the dry habitats of the western USA and Canada. Despite showing a variation of lifestyles and distinct morphological features in the vegetative as well as in reproductive organs, most *Sisymbrium* species can be found in or near dry open steppe habitats where they grow on limestone, sand or gravel substrates, or in semi-deserts in the temperate zone of the northern hemisphere and African subtropics. However, some species such as *S. altissimum*, *S. irio*, *S. loeselii*, *S. officinale* and *S. volgense* are known for being successful ruderal-occupying alien species (Jehlík, 1981; Mito and Uesugi, 2004; Oprea and Sîrbu, 2010; Protopopova et al., 2006; Weber et al., 2008) that have spread throughout Europe or Asia from their native habitats in the East European Plain and western Irano-Turanian region.

The southern East European Plain is part of a nowadays continuous Eurasian steppe belt, which is the vastest grassland region in the world with 8000 km in length and up to 1000 km in width. This ecoregion was under strong influence of glacial and interglacial cycles during the past 5 MYR, which continuously caused contractions, expansions and latitudinal range shifts of the Eurasian steppe belt (Hurka et al., 2019). In contrast, the western and central regional sub-centres of the Irano-Turanian region (Léonard, 1988) were not as strongly affected by the Quaternary glaciations as the Eurasian steppe belt. Only intermountanious valley glaciations were recorded from the Pleistocene cycles, leaving ancestral flora at least partially intact (Agakhanjanz and Breckle, 1995). This is corroborated also by high levels of endemic species in Iranian mountain ranges and current distribution patterns of several representative Irano-Turanian plant taxa (Djamali et al., 2012a,

2012b; Noroozi et al., 2008). The western and central regional sub-centres of the Irano-Turanian region may have played an important role by partially “supplying” adjacent regions – northern and eastern Irano-Turanian sub-centres, as well as the Mediterranean Basin and Saharo-Arabian floristic region – with floral elements (Magyari et al., 2008; Manafzadeh et al., 2017, 2014; Takhtajan, 1986; Žerdoner Čalasan et al., 2019).

The overall aim of our studies is to place the biological history of *Sisymbrium* L. into time and space. To achieve that we carried out a multigene phylogeny using bi-parental fast-evolving ITS and ETS regions, as well as low copy nuclear markers Bra13, Bra246, Bra813, Bra1402, and maternal chloroplast markers trnQ-rps16, psbA-trnH and ycfb1 of a representative taxon set, including approximately 85% of all currently recognised species. Due to fossil absence, we further tested the impact of different tree priors, calibration methods (substitution rates vs. secondary calibration points) and different taxon sets on the topology as well as time spans of the Sisymbrieae clade. In addition, aiming for taxonomic stability, we also formally transfer *Ochthodium aegyptiacum* DC. to *Sisymbrium* and investigated the phylogenetic position of *Sisymbirum aculeolatum* and *S. afghanicum*. To achieve the latter and test the robustness of our time estimation, we extended the phylogenetic analysis and time divergence estimation to the whole Lineage II.

2 Materials and methods

2.1 Taxon sampling and distribution range surveys

We compiled a taxon sample covering the whole distribution area of all collected species of *Sisymbrium*. Distribution maps were put together primarily using geographical range information from the literature (Brach et al., 2006; Grubov, 2001; Jalas et al., 1994; Malyschev et al., 2004; Meusel et al., 1965; Post, 1896; Tutin et al., 1993; Vassilczenko, 1939) and our own field data, and only secondarily using online databases (Euro+med PlantBase: http://ww2.bgbm.org/EuroPlusMed/query.asp; GBIF: https://www.gbif.org; Flora of Mongolia: https://floragreif.uni-greifswald.de/category/literature/; Plantarium: https://www.plantarium.ru/; Plants of the World Online: http://www.plantsoftheworldonline.org/; Weeds of Australia : shorturl.at/aduzS), whose entries were critically assessed prior inclusion in the survey. Plant material was obtained from specimens deposited in B, MSB, GDA, GDAC, HBG, HEID, HUJ, K, L, P, MW, PRE, OSBU and WAG herbaria (Appendix 1). Geographical data was recorded at the specimen collection site (in the case of our own collection), recovered directly from the voucher specimens or reconstructed by using TopGlobus (http://www.topglobus.ru/), Wikimapia (http://www.wikimapia.org/) and Loadmap (http://loadmap.net/) based on text information from the voucher specimens. The maps for ancestral range reconstruction were generated with QGIS (https://www.qgis.org/en/site/) and all the figures were adjusted in Inkscape (https://inkscape.org).

2.2 Molecular analyses

Total genomic DNA was isolated from herbarium specimens using the InnuPREP Plant DNA Kit (Analytic Jena AG, Jena, Germany) according to the instructions of the manufacturer and used directly in PCR amplifications. PCR was carried out using 10 µl of 2X HS Taq Mix Red (Biozym Scientific, Hessisch Oldendorf, Lower Saxony, Germany), 7 µl of ddH_2O, 1 µl of each 10 µM primer and 1 µl of template DNA. Alternatively, PCR was carried out using 20.8 µl of ddH_2O, 3 µl of 10x Taq Buffer with $MgCl_2$, 2 µl of 10mM dNTP mixture, 1 µl of DMSO, 0.2 µl of 5U Taq Polymerase, 1 µl of corresponding primers, and 1 µl of the template DNA. We analysed the following nuclear encoded loci: internal transcribed spacer, including the intervening 5.8S region (ITS1-5.8S-ITS2), external transcribed spacer (ETS), Bra264 exon and Bra13, Bra813 and Bra1402 loci all comprised of two introns and three exons (Stockenhuber et al., 2015). Furthermore, three additional chloroplast regions were investigated: trnQ-rps16 intergenic spacer, psbA-trnH intergenic spacer and ycf1b coding locus. Used primers and PCR programmes can be inferred from Appendix

2. Samples that yielded single bands once run through gel electrophoresis were sent to Microsynth Seqlab (Gottingen, www.microsynth.ch) for purification and sequencing. Sequences from each accession were manually edited in Chromas Lite 2.6.4 (Chromas | Technelysium Pty Ltd) and aligned and manually corrected in AliView v1.24 (Larsson, 2014).

2.3 Phylogenetic analyses of Lineage II and Sisymbrieae

The tribus Sisymbrieae is embedded into the Lineage II of the Brassicaceae family (BrassiBase: Kiefer et al., 2014; Koch et al., 2012, 2018). To test the robustness of the retrieved topologies, we performed phylogenetics on two different taxonomic levels. First dataset (from this point onwards referred to as Lineage II) consisted of newly generated ITS-based dataset of Sisymbrieae (see first 55 lines in Appendix 1 and Results section) and from BrassiBase retrieved ITS-based datasets from three other Brassicaceae tribes that together with Sisymbrieae constitute the Lineage II: Brassiceae, Isatideae and Thelypodieae (BrassiBase: Kiefer et al., 2014; Koch et al., 2012, 2018). The complete list of this dataset with corresponding GenBank entries can be inferred from Appendix 3, datasets A and B. The second dataset consisted of Sisymbrieae accessions only (see below).

Due to the problematic nature of *Sisymbrium aculeolatum* and *S. afghanicum*, several datasets were investigated within the Lineage II to test for the phylogenetic placement of the two species. Firstly, we tested the position of these two above-mentioned species by using all four tribes that belong to the Lineage II (sensu BrassiBase). In addition, we tested how presence/absence of *Sisymbrium aculeolatum* and *S. afghanicum* influence the topology when two outgroups to the Lineage II (Eutremeae and Thlaspideae sensu BrassiBase) are included. The ITS sequences from these two tribes that served as outgroup were retrieved directly from the BrassiBase ITS databank and the complete list of this dataset with corresponding GenBank entries can be inferred from Appendix 3, datasets C and D. Thus, we ended up with four different datasets within Lineage II: Lineage II including *Sisymbrium aculeolatum* and *S. afghanicum* (Appendix 3, dataset A), Lineage II excluding *Sisymbrium aculeolatum* and *S. afghanicum* (Appendix 3, dataset B), Lineage II with Eutremeae and Thlaspideae as outgroups including *Sisymbrium aculeolatum* and *S. afghanicum* (Appendix 3, dataset C) and Lineage II with Eutremeae and Thlaspideae as outgroup excluding *Sisymbrium aculeolatum* and *S. afghanicum* (Appendix 3, dataset D).

The phylogenetic analyses of Sisymbrieae dataset (consisting of only *Sisymbrium* accessions, see below) as well as the Lineage II dataset (see above) were carried out using Bayesian Inference (BI) showing Bayesian posterior probabilities (BPP) as well as Maximum Likelihood approach (ML),

showing likelihood bootstrap support (LBS). BI was carried out using MrBayes v3.2.6 (Ronquist et al., 2012) under GTR+Γ substitution model (alongside the HKY+Γ and SYM+Γ model in Sisymbrieae dataset, see below) using the random-addition-sequence method with 10 random perturbations. Two independent Markov chain Monte Carlo (MCMC) analyses of four chains were run – one cold and three heated. Runs were carried out for 20 million cycles, with parameters and trees sampled every 1,000th cycle including an appropriate burn-in (10%) as inferred from the evaluation of the trace files using Tracer v1.7.0 (Rambaut et al., 2018). Maximum likelihood tree inference relied on RAxML-HPC v8.2.10 (Stamatakis, 2014) with 1000 ML bootstrap iterations with the default number of distinct rate categories. Tree searches were executed from a random maximum-parsimony tree employing either the GTRCAT or the GTRGAMMA bootstrapping phase model. In case of multilocus Sisymbrieae dataset, six partitioning schemes were additionally enforced, one combining the three cpDNA loci and one combining rDNA ITS and ETS. Due to the complex nature of the four analysed Bra loci (consisting of different number of exons and introns), each of these were assigned its own partition.

After determining whether *Sisymbrium aculeolatum* and *S. afghanicum* should be incorporated into the dataset based on single-locus ITS phylogenies, we focussed on the tribe Sisymbrieae. We first generated ITS sequences for 209 accessions representing approximately 85% of all currently recognised *Sisymbrium* species (Appendix 1), using the PCR programme and primers specified in Appendix 2. This helped us to assess the genetic diversity of the genus in general. We then reduced the taxon sample to include only one accession per species (if no divergent ITS sequences were retrieved), or alternatively one accession per diverging ribotype. The reduced taxon sample comprised 55 accessions (first 55 entries in Appendix 1). For these accessions, additional loci were amplified (specified above). All accessions of the following investigated species exhibited identical ambiguous positions in the following loci: *Sisymbrium loeselii* in ITS and ETS loci, *S. strictissimum* in Bra13, Bra246, Bra813 and Bra1402 loci, *S. subspinescens* in ITS locus and *S. yunnanense* in Bra13 and Bra246 loci. To test for multiple copies within individuals, cloning of the PCR amplicons of one individual per species (depicted in * in the Appendix 1) was carried out using the Thermo Scientific™ CloneJET PCR Cloning Kit according to the instructions of the manufacturer. An exception was *Sisymbrium strictissimum*, in which two accessions (both exhibiting one diverging ITS copy) were investigated for all four aforementioned Bra loci. 10-15 clones per locus per individual were analysed further. The yielded plasmids were isolated and purified using NucleoSpin® Plasmid EasyPure (Macherey-Nagel, Düren, Germany) following the manufacturer's

instructions. The sequencing was performed by the Microsynth Seqlab (Gottingen, www.microsynth.ch), using universal pJET1.2_F and pJET1.2_R primers.

While the Lineage II phylogenies were based solely on ITS sequence information, the phylogeny of Sisymbrieae was based on two rDNA loci ITS and ETS, four low copy Bra loci and three cpDNA encoded loci (see first 55 lines in Appendix 1 for taxon sample). Single locus phylogenies were carried out and the topologies were assessed using IQ-TREE (Nguyen et al., 2015). We concatenated the six nuclear DNA loci and three cpDNA loci, respectively, as no major topological incongruences were observed within the single loci. However, as concatenation might lead to artificial topologies and bias statistical branch support (Kubatko and Degnan, 2007), we additionally tested both, nuclear DNA based and cpDNA based Sisymbrieae datasets using ASTRAL (Zhang et al., 2017), showing bootstrapping (BS). It was proven that this algorithm is statistically consistent under the multi-species coalescent model (Mirarab et al., 2014). Both IQ-TREE analysis as well as ASTRAL analysis were run under default settings, with increased 1000 bootstrapped gene alignments and 1000 bootstrapping trees, respectively. The topologies retrieved from Bayesian Inference and Maximum Likelihood were then compared with topologies retrieved from ASTRAL.

2.4 Time estimation analyses of Lineage II and Sisymbrieae

To obtain the appropriate crown age of Sisymbrieae, we first dated the whole Lineage II, based on the same ITS-based taxon set as in the phylogenetic studies, one including the four tribes of the Lineage II only (see Appendix 3, dataset A for the whole accession list) and one including the two additional outgroups (Eutremeae and Thlaspideae sensu BrassiBase, see Appendix 3, dataset C for the whole accession list), both now including *Sisymbrium aculeolatum* and *S. afghanicum*. These two different datasets were tested to assess the robustness of the retrieved time spans and to allow for additional secondary calibration points in the dataset including Eutremeae and Thlaspideae as outgroups.

There are no reliable Brassicaceae fossils that could potentially serve as time constraints (Franzke et al., 2011), with a possible exception of East European *Bunias* fossil from the Pliocene (Mai, 1995). Furthermore, in the light of general fossil absence it is hard to establish a stable evolutionary time frame. Thus, we tested three different time constraint approaches, firstly using secondary calibration points retrieved from a whole plastome analysis (Huang et al., 2020) and as an alternative using the internal transcribed spacer substitution rate for herbaceous annual/perennial angiosperms of 4.13 x

10^{-9} sub/site/yr (Kay et al., 2006). The third approach included secondary calibration points as well as the published ITS substitution rate. While Huang et al. (2020) also published the crown age of Sisymbrieae, we did not incorporate this age into our dating analysis of the Lineage II, but used it as an independent checkpoint to estimate the accuracy of the ITS-based substitution rate calibration as a time constraint. Furthermore, we wanted to test how different datasets might influence the time spans. Our multiple layer approach allowed us to test how time spans might be effected by i) the taxon sample outside of Sisymbrieae, by using only Lineage II in our study or alternatively using Lineage II with two outgroups sensu BrassiBase vs. whole family of Brassicaceae used in Huang et al. (2020); ii) the taxon sample within the tribus of interest, by using a comprehensive taxon sample of Sisymbrieae in our study vs. a reduced taxon sample of *Sisymbrium* used in Huang et al. (2020) and iii) calibration strategy, by using single-gene-based nuclear substitution rate vs. primarily calibrated nodes from a whole plastome phylogeny used in Huang et al. (2020).

After obtaining a robust crown age for the tribus Sisymbrieae from our ITS-based time divergence estimations of different datasets of Lineage II, we applied the average median time span of all retrieved time spans onto the multi-locus Sisymbrieae dataset consisting of 55 entries (see first 55 entries in Appendix 1). Due to the topological incongruences between nuclear DNA based trees and cpDNA based trees (in all ML, BI and ASTRAL algorithms: see Results), better resolution obtained from the nuclear based dataset and ability to use ITS substitution rate as calibration method, we carried out time estimation analysis on nuclear DNA based dataset of Sisymbrieae only. Five unlinked datasets were fed to the algorithm, each representing one Bra locus and one representing combined ITS and ETS loci. We left the site and the clock model unlinked and set the trees generated from the independently analysed loci to be linked.

Molecular dating analyses relied on BEAUTi & BEAST v1.10.4 (Suchard et al., 2018) and the uncorrelated relaxed clock model. The use of the uncorrelated relaxed clock model was justified, after assessing the coefficient of variation in all the trace files, which consistently exceed 0.5 in all of the cases – in the dating analyses of the Lineage II as well as of Sisymbrieae. For both Lineage II datasets (the one with putative outgroups Eutremeae and Thlaspideae and the one without) we used the GTR+Γ substitution model, corresponding tree priors (see below), and three independent Monte Carlo Markov chains (MCMC) runs for 100 million generations, with parameters sampled every 20,000th generation. For the datasets of Sisymbrieae we used the GTR+Γ and HKY substitution models partition-wise, corresponding tree priors (see below), and three independent Monte Carlo Markov chains (MCMC) runs for 50 million generations, with parameters sampled every 10,000th

generation. Effective sample sizes (ESS) for all estimated parameters were assessed using Tracer v1.6.0 (Rambaut et al., 2018). We excluded the option +I, as it may lead to over parametrisation (BEAST manual). TreeAnnotator v1.8.4 (Suchard et al., 2018) was used to discard 10% of the saved trees and annotate the rest of them. Maximum clade credibility tree with median node heights was visualised using FigTree graphical viewer of phylogenetic trees v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

All sequence evolution models used in this study were assessed using the Akaike information criterion (AIC) implemented in the jModelTest2 v2.1.6 (Darriba et al., 2012). All the analyses were carried out at the CIPRES Science Gateway computing facility (Miller et al., 2010). The aligned matrices are available as *.nex files upon request. Recent studies have shown that the impact of the tree priors in Bayesian phylogenetics is in general not as strong as previously thought (Ritchie et al., 2017; Sarver et al., 2019). Nevertheless, one should set the priors accordingly, as they might have an impact on the accuracy of the analysis, especially of those based on mixed inter- and intraspecies datasets (Ritchie et al., 2017). Thus, as all of our taxon samples —the Lineage II with and without the putative Eutremeae and Thlaspideae outgroups as well as Sisymbrieae dataset— were not limited to one accession per species only (which would automatically exclude the Coalescence tree prior) and to furthermore allow for potential extinction events (embedded into the Birth-Death tree prior, but not into Yule tree prior), all analyses were done in parallel using the Yule, Coalescence and Birth-Death tree prior.

2.5 Ancestral range reconstruction of Sisymbrieae

The time divergence tree from BEAST analysis performed under coalescent tree prior was used for the ancestral range reconstruction. Distribution maps and contemporary range estimations followed the same methodology described in detail in Section 2.1. The analysis was carried out using RASP4 v4.0 ancestral state reconstruction tool (Yu et al., 2015; 2020). DEC, DIVALIKE, and BAYAREALIKE biogeographic models with and without corresponding jumping parameter '+j' were tested using BioGeoBEARS v1.1.1 algorithm implemented in RASP through R (Matzke, 2013, 2014, 2018). Following the biogeographic division and evolutionary history of Eurasia and Africa (Hurka et al., 2019; Linder et al., 2012; Linder, 2014; Wesche et al., 2016), eight geographic entities were defined: western Irano-Turanian floristic region (A), Euro-Siberian steppe and adjacent semi-deserts (B), Mediterranean and Northern Africa (C), Mongol-Chinese steppe, semi-deserts and deserts (D), European temperate and cool-mixed forests (E), Caucasus (F), Southern

African open grasslands (G), and temperate and subtropical East Asia (H). Due to the geographic distance and evolutionary trajectory, it is highly unlikely that biota of the South African open grasslands migrated from/to areas beside the two most adjacent and climatically similar Mediterranean and/or western Irano-Turanian floristic region. Thus, only combinations of AG and CG were allowed in ancestral range areas including the South African open grasslands with one other region. The maximum number of areas in which a species can coexist was set to three, following the widely distributed *Sisymbrium irio* that can be found in maximally three out of eight areas specified above.

3 Results

Details on individual alignments of Lineage II and Sisymbrieae can be inferred from Appendix 4. Based on the ITS sequence information, three investigated accessions (see first column in Appendix 1) were placed into morphologically similar genus *Erysimum* and were excluded from further analyses. After assessment of the genetic diversity of *Sisymbrium* based on ITS locus, a reduced representative taxon sample set of Sisymbrieae was used for all the subsequent analyses that covered the whole known genetic diversity within this tribus (see Materials & Methods section, for taxon sample refer to the first 55 lines in Appendix 1). The 10-15 sequenced clones per locus per accession consistently retrieved two major individual genetic copies that were included into all subsequent analyses (see Appendix 1, diverging clone copies depicted in *).

3.1 Phylogenetics and divergence time estimation of Lineage II

The intra-tribal relationships within Lineage II without two outgroups (Eutremeae and Thlaspideae) remained unclear and the monophyly of some tribes remained controversial. Isatideae and Sisymbrieae were retrieved monophyletic (93 LBS/1.00 BPP and 78 LBS/0.93 BPP, respectively), whereas the statistical support for Brassiceae and Thelypodieae was poor — 42 LBS/0.58 BPP and 31 LBS/0.59 BPP, respectively (Figure 2). However, in absence of controversial *Sisymbrium aculeolatum* and *S. afghanicum,* the statistical support for the tribes overall increased to reliable 99 LBS/1.00 BPP for Sisymbrieae and 97 LBS/1.00 BPP for Isatideae and still unreliable 51 LBS/0.91 BPP for Brassiceae and 36 LBS/0.93 BPP for Thelypodieae (Figure 2). The above-mentioned datasets comprised accessions presented in detail in Appendix 3, datasets A and B. Further statistical features on individual alignments can be inferred from Appendix 4.

The intra-tribal relationships within Lineage II including two outgroups (Eutremeae and Thlaspideae) remained controversial as well. In dataset without *Sisymbrium aculeolatum* and *S. afghanicum*, all tribes were retrieved as monophyletic, albeit some with a negligible support. Brassiceae (27 LBS/0.53 BPP) were placed as sister group to Thelypodieae (52 LBS/0.80 BPP), Sisymbrieae (89 LBS/1.00 BPP) were placed as sister clade to Isatideae (98 LBS/1.00 BPP), and Eutremeae (88 LBS/0.99 BPP) were placed as sister group to Thlaspideae (98 LBS/1.00 BPP; Figure 2). When *Sisymbrium aculeolatum* and *S. afghanicum* were included, the support of previously moderately supported clades plummeted and previously poorly supported topologies collapsed: Eutremeae (86 LBS/1.00 BPP) remained a sister group to Thlaspideae (97 LBS/1.00 BPP) and the true Lineage II was retrieved on a poorly supported comb comprising of Brassiceae

(29 LBS/1.00 BPP), Thelypodieae (00 LBS/0.00 BPP), Sisymbrieae (00 LBS/0.85 BPP) and Isatideae (00 LBS/0.55 BPP; Figure 2). The above-mentioned datasets comprised accessions presented in detail in Appendix 3, datasets C and D. Further statistical features on individual alignments can be inferred from Appendix 4.

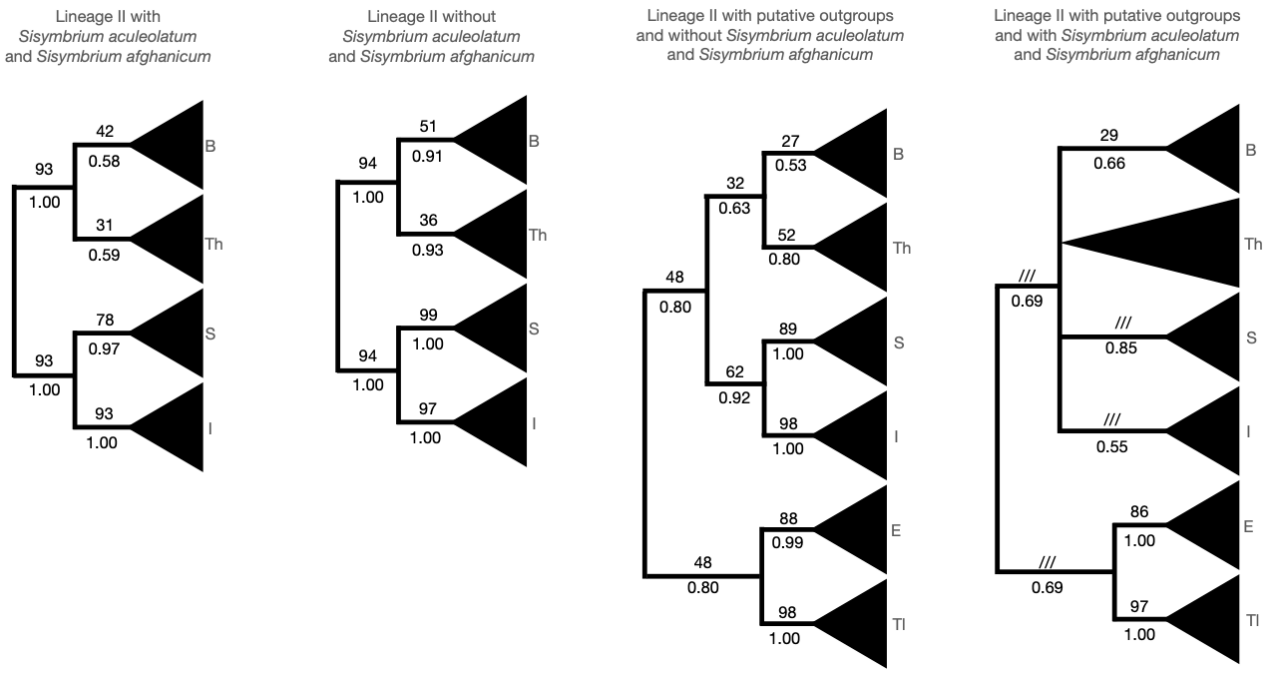


Figure 2: Schematic representation of ITS-based phylogenetic trees based on different taxon sets, illustrating the topological differences. Values above branches correspond to the LBS (likelihood bootstrap support) values from the maximum likelihood analysis and the values below branches correspond to the BPP (Bayesian posterior probabilities) values of the Bayesian Inference analysis. B, Brassiceae; Th, Thelypodieae; S, Sisymbrieae; I, Isatideae; E, Eutremeae; Tl, Thlaspideae.

The Maximum Likelihood analysis and Bayesian Inference analysis of Lineage II consistently placed both *Sisymbrium altissimum* and *S. septulatum* as the sister group to the rest of the Sisymbrieae tribe, albeit with a low statistical support (cannot be inferred from collapsed topologies in Figure 2). Alternatively, depending on the tree prior in the time divergence estimation, an alternative sister group to the rest of Sisymbrieae was recognised, consisting of *Sisymbrium aculeolatum*, *S. afghanicum*, *S. altissimum* and *S. septulatum.* The overall topology corresponded with the previous study of *Sisymbrium* based on an extensive taxon dataset from Warwick et al., 2002. The used taxon sample of Lineage II allowed for three (or five, when the putative outgroups Eutremea and Thlaspideae were included) secondary calibration points from a fossil-calibrated whole plastome analysis (Huang et al., 2020). The crown age of Sisymbrieae was not incorporated into the secondary calibrated Lineage II analysis but was used as an independent checkpoint to estimate the accuracy of the use of ITS-based substitution rates as a time constraint and to test the

influence of completeness of a taxon sample. The ESS values of all the time estimation analyses exceeded 400. Secondarily calibrated analyses and analyses calibrated with substitution rates showed highly congruent results in terms of topology, but in some cases only partially congruent time spans (Tables 1a & 1b). Nevertheless, the crown age of Sisymbrieae was consistently placed in the Miocene. The combinatorial calibration approach in which both secondary calibration points as well as ITS-based substitution rate was used placed the mean crown age of Sisymbrieae at the end of early Miocene and the beginning of middle Miocene in corresponding priors and datasets, respectively (Tables 1a & 1b). These age spans are highly congruent with the Sisymbrieae crown age from Huang et al., 2020 estimated to be 10.2–18.7 MYA 95% HDP. Similar time spans were inferred from calibration using ITS substitution rate only, consistently placing the mean crown age of Sisymbrieae into the early Miocene. Datasets where only secondary calibration points from Huang et al., 2020 were used, estimated the mean age of Sisymbrieae slightly younger, placing it in the late Miocene (Tables 1a & 1b).

3.2 Phylogenetics and divergence time estimation of Sisymbrieae

The separately concatenated nuclear DNA loci and cpDNA loci, respectively, showed only partially congruent results. The most prominent difference was a strong bias in the resolution (Figure 3). While the nuclear DNA loci have contributed substantially to a well-resolved backbone of the Sisymbrieae phylogeny but in some cases failed to unravel internal topology, the three analysed cpDNA loci recovered poorly resolved backbone, but in contrast allowed for better-resolved internal nodes. The ASTRAL analyses (Appendices 5 and 6) retrieved trees with topologies highly comparable to the ones resulting from concatenated datasets (Figure 3).

Similarly to the phylogenetic reconstructions and time estimation analyses of the Lineage II datasets, the time estimation analysis of the Sisymbrieae dataset did not consistently recognise the same Sisymbrieae lineages as sister clade to the rest of the tribe (topologies not shown). Again, depending on the tree prior, a poorly supported group consisting of either *Sisymbrium aculeolatum*, *S. afghanicum*, *S. altissimum* and *S. septulatum* or of only *Sisymbrium aculeolatum* and *S. afghanicum* was placed as the sister clade towards the rest of Sisymbrieae. We reconstructed the outgroup following the topology of some of our time estimation analyses and previous phylogenies with an extensive taxon sample (Warwick et al., 2002), where *Sisymbrium altissimum* and *S. septulatum* constituted a sister clade to the rest of Sisymbrieae (Figures 3, and Appendices 5 and 6). Independently, the phylogenies uncovered numerous geographically well-defined clades that

remained well-supported in nuclear DNA based as well as cpDNA based analyses, regardless of the used dataset (nuclear DNA and cpDNA) as well as algorithms (ML, BI, ASTRAL: Figures 3, and Appendices 5 and 6).

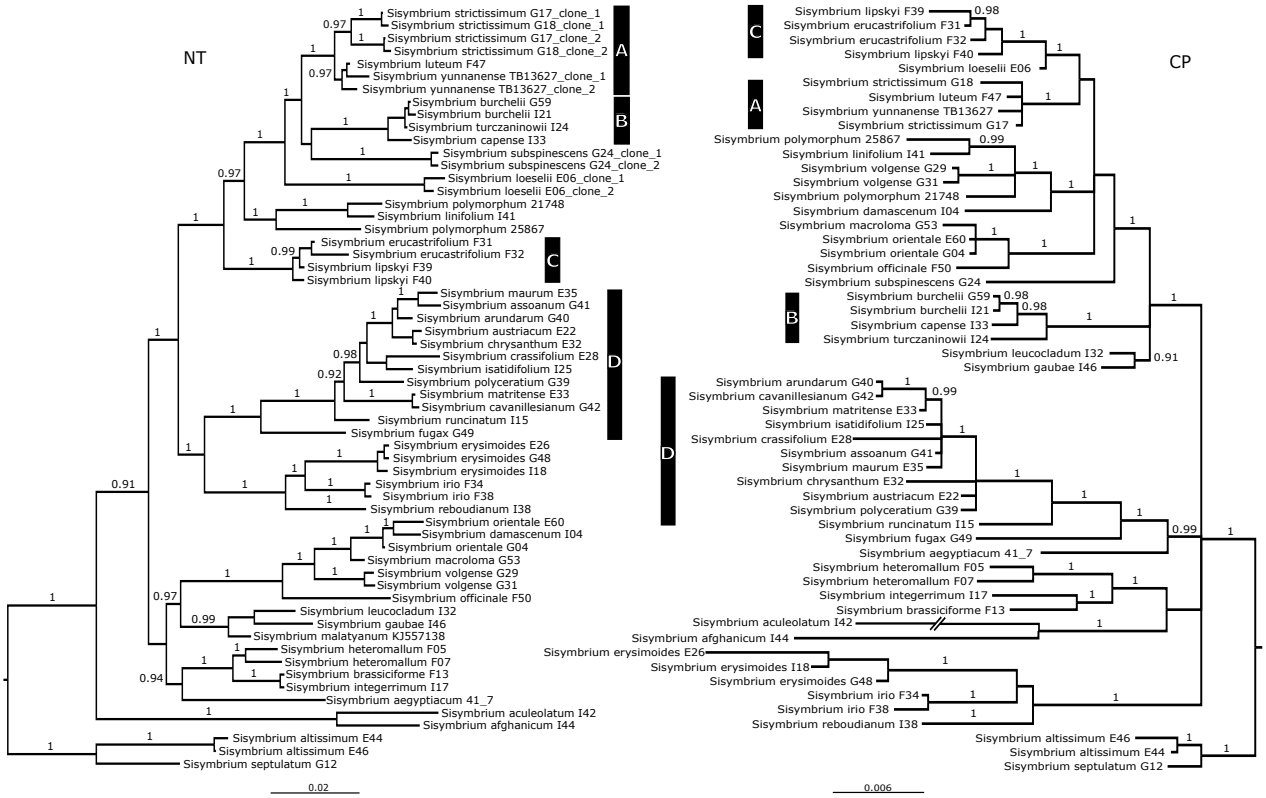


Figure 3: Bayesian Inference phylogenetic tree of Sisymbrieae, the nuclear DNA tree (left) and the cpDNA tree (right). Values below branches correspond to the LBS (likelihood bootstrap support) values from the maximum likelihood analysis (values <0.90 are not shown) and the values above branches correspond to the BPP (Bayesian posterior probabilities) values of the Bayesian Inference analysis (values <0.90 are not shown). A, disjunct forest Eurasian clade; B, South-African clade; C, Caucasian clade; D, pure Mediterranean clade. Two slashes indicate a branch that has been artificially shortened. Codes next to species names refer to the isolation codes from Appendix 1.

These included central Asian *Sisymbrium brassiciforme*, *S. heteromallum* and Iranian endemic *S. integerrimum* (0.98 LBS/1.00 BPP/99 BS NT; 0.92 LBS/1.00 BPP/00 BS CP), Mediterranean and Middle Eastern *S. erysimioides*, *S. irio* and *S. reboudianum* (100 LBS/1.00 BPP/100 BS NT; 100 LBS/1.00 BPP/00 BS CP), pure (primarily West) Mediterranean *S. austriacum*, *S. assoanum*, *S. arundanum*, *S. cavanillesianum*, *S. chrysanthum*, *S. crassifolium*, *S. fugax*, *S. isatidifolium*, *S. matritense*, *S. maurum*, *S. polyceratium* and *S. runcinatum* (100 LBS/1.00 BPP/96 BS NT; 0.95 LBS/1.00 BPP/94 BS CP), South-African *S. capense*, *S. burchelii* and *S. turczaninowii* (100 LBS/ 1.00 BPP/100 BS NT; 100 LBS/1.00 BPP/100 BS CP), Caucasian *S. erucastrifolium* and *S. lipskyi* (100 LBS/1.00 BPP/100 BS NT; 0.91 LBS/1.00 BPP/90 BS CP) and disjunct forest Eurasian *S. luteum*, *S. strictissimum* and *S. yunnanense* (100 LBS/1.00 BPP/100 BS NT; 0.98 LBS/1.00 BPP/97

BS CP). Three major incongruences between the nucleotide and plastid phylogenetic signal were spotted. While the *Sisymbrium polymorphum* s.l. (consisting of *S. polymorphum* and *S. linifolium*) represented a fairly isolated lineage in the nuclear-based phylogenies, it was placed in chloroplast phylogenies as sister group to *S. volgense* (0.89 LBS/1.00 BPP/00 BS). Another contradicting position concerned *Sisymbrium damascenum*, which was nested within the *Sisymbrium macroloma — S. officinale — S. orientale — S. volgense* clade in the nuclear tree, while in the chloroplast tree, it clustered together with *S. polymorphum* s.l. and *S. volgense*. An additional minor topological change was observed in fairly isolated species in nuclear-based phylogenies *S. loeselii* that was placed as sister group to the Caucasian clade in chloroplast-based phylogenies (0.90 LBS/1.00 BPP/ 86 BS).

The position of *Ochthodium aegyptiacum* remained controversial, albeit the whole Lineage II phylogeny doubtlessly placed this taxon in *Sisymbrium*. All species with multiple accessions, except *Sisymbrium polymorphum*, were resolved as monophyletic with maximal support or in the cases of *S. orientale* and *S. yunnanense* paraphyletic with low statistical support. Notably, *Sisymbrium brassiciforme* and *S. integerrimum* exhibited almost identical nucleotide sequences, yet showed extensive differences in the investigated chloroplast regions. Nevertheless, these species showed a sister group relationship in nucleotide as well as chloroplast set (Figure 3, Appendices 5 and 6). The ESS values of time estimation analyses exceeded 600 and the different tree priors did not majorly influence the time spans (see Table 2). Four geographically well-defined clades were retrieved from the analyses. The mean stem age of the disjunct forest Eurasian clade was dated to the Pliocene/ Pleistocene, while its mean crown age was dated to early Pleistocene. The mean stem age of the South-African clade was dated to the Pliocene, while its mean crown age was dated to the Calabrian in the Pleistocene (Figure 4). Furthermore, the mean stem age of the Caucasian clade was dated earlier to the late Miocene, while its mean crown age was dated as well to the Calabrian in the Pleistocene, and the mean stem as well as crown age of the Mediterranean clade was dated to the late Miocene (Figure 4). Retrieved clone copies from *Sisymbrium strictissimum* (100 LBS/1.00 BPP), *S. subspinescens* (100 LBS/1.00 BPP) and *S. loeselii* (100 LBS/1.00 BPP) formed well-supported sister group relationships (Figure 3 and Figure 4). Poorly supported paraphyly between *Sisymbrium yunnanense* clones and its sister species *S. luteum* (48 LBS/0.92 BPP) was observed (Figure 3, Figure 4).

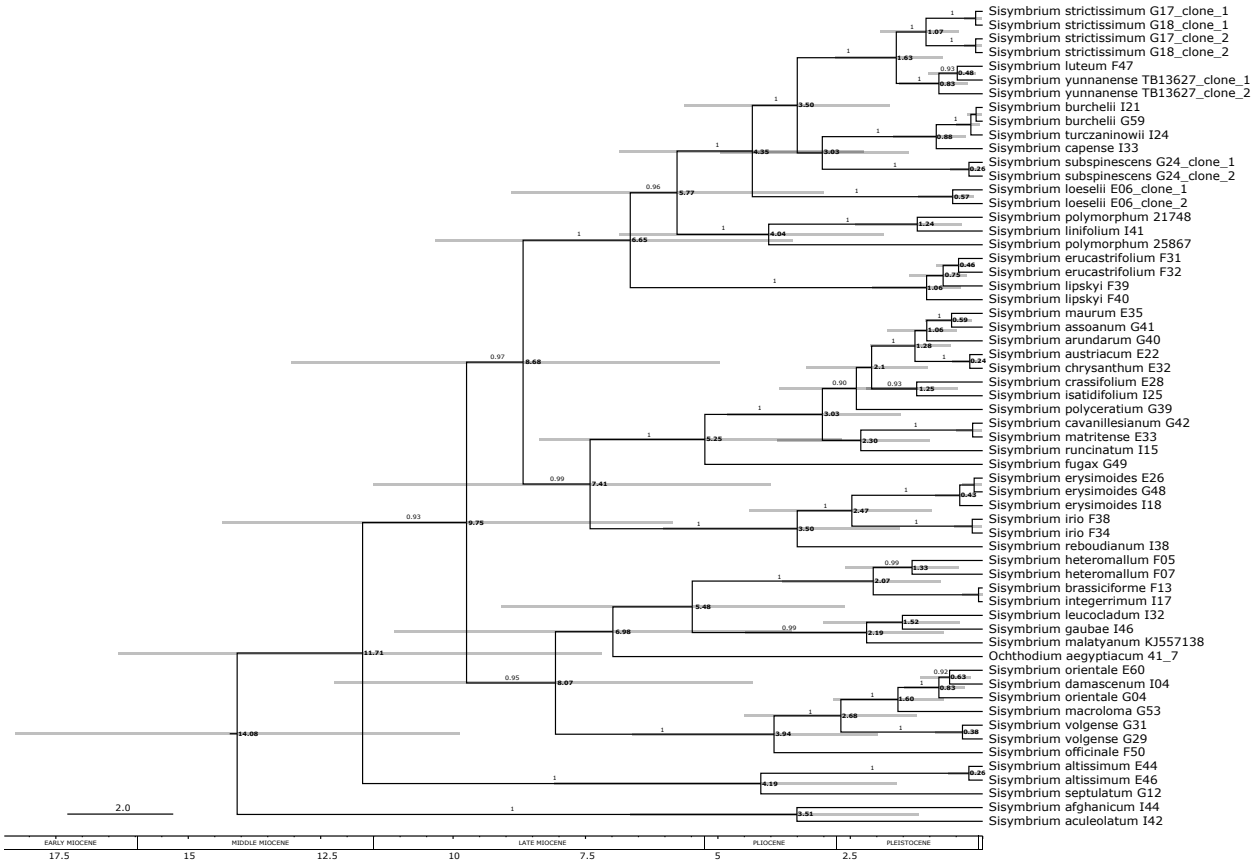


Figure 4: Dated nuclear DNA phylogenetic tree of Sisymbrieae retrieved from BEAST analysis under Coalescent tree prior. The intervals next to the nodes represent the time span of the node age, including the 95% HDP. Absolute ages are in millions of years. The numbers on the branches are statistical support values (Bayesian posterior probabilities, values <0.90 are not shown). The bold numbers next to nodes represent median ages in millions of years. Codes next to species names refer to the isolation codes from Appendix 1.

3.3 Ancestral range reconstruction of Sisymbrieae

The most optimal model according to the AIC algorithm was DEC+j (Appendix 7). Ancestral range reconstruction placed the origin of the tribus Sisymbrieae into the western Irano-Turanian floristic region and Mediterranean (Figure 5A and 5B). In the latter the highest number of speciation events i. e. 23 was inferred, followed by Euro-Siberian steppe and adjacent semi-deserts with 12 and western Irano-Turanian floristic region with 10 speciation events (Figure 5D). These three areas were also the major source for dispersal events, 13 of which took place from the Mediterranean region (Figure 5D). Dispersal between areas index also pointed towards a close relationship between Mediterranean and western Irano-Turanian floristic region, retrieving eight such dispersal

events, albeit with a strong bias, as seven out of eight dispersal events had their onset in the Mediterranean and dispersed into the western Irano-Turanian floristic region (Figure 5B and 5C). Most of the dispersal routes followed from Mediterranean and/or through western Irano-Turanian floristic region (Figure 5C) from which the routes separated into western ones that invaded Europe through the Eastern European Plain (Figure 5B nodes 81→80 and 74→73) and eastern ones that invaded eastern portion of continental Asia (Figure 5B nodes 77→76 and 113→105→100). The node transition 72→66 illustrates an expansion to the east into temperate and subtropical East Asia as well as to the west into European temperate and cool-mixed forests from central laying contemporary Euro-Siberian steppes and adjacent semi-deserts that were at that point also covered in forests (see Discussion). Europe was also invaded through Mediterranean (Figure 5B nodes 85→84). The Southern African grasslands were invaded from climatically similar Mediterranean and adjacent dry habitats of Middle Asia (Figure 5B nodes 71→69). The corresponding statistics from the ancestral range reconstruction can be further inferred from Appendix 7. Please note that *Sisymbrium linifolium* is the only *Sisymbrium* species whose distribution extends into the North American prairies, however, for easier depiction North America is not illustrated on the Figure 5.

4 Discussion

While the onset and development of flora of the Mediterranean is relatively well investigated (Barres et al., 2013; Guzmán and Vargas, 2005; Manafzadeh et al., 2014; Molins et al., 2011; Ortiz et al., 2008; Scheunert and Heubl, 2014; Tremetsberger et al., 2016), the florogenesis of the Eurasian steppe and Middle East is rather poorly understood. This study focuses on *Sisymbrium* L. – yellow-flowering herbaceous Brassicaceae, which grows predominantly in dry steppe, semi-desert and rural habitats of Northern hemisphere (Figure 1).

4.1 Taxon sampling strategy and differences in the topology

Following the contemporary big scale phylogenies, genetic as well as cytological evidence, we excluded taxa *Orychophragmus* (Al-Shehbaz, 1985; Couvreur et al., 2010; German et al., 2009; Gómez-Campo, 1980; Liu et al., 2012; Lysak et al., 2007) and *Sinalliaria* (Kiefer et al., 2014; Koch et al., 2018, 2012) from the analysis, whereas *Sisymbrium leucocladum* (Boiss.) D.A.German & Al-Shehbaz (= *Pseudofortunya leucoclada* (Boiss.) Khosravi) and *Ochthodium aegyptiacum* remain included (Walden et al., 2020; Nikolov et al., 2019). Our Lineage II phylogenetics and the high morphological resemblance of *Sisymbrium aculeolatum* and *S. afghanicum* to other *Sisymbrium* species (with a notable exception of purple flowers vs. yellow flowers in all other *Sisymbrium* species), indicates that the two aforementioned species belong to the genus *Sisymbrium*. The strong impact these two species have on tree topologies might be an indication of an ancient hybrid origin or intense introgression — both known to disrupt the phylogenetic structure of bifurcating trees. This would also explain why *Sisymbrium aculeolatum* was not retrieved as a part of Sisymbrieae in Walden et al. (2020). Further analyses, however, are necessary for any robust conclusions on this issue. The overall topology of all Lineage II phylogenetic trees (ML and BI-based) was in correspondence with other studies that have analysed phylogenetic aspects of Brassicaceae evolution and included *Sisymbrium* species (Huang et al., 2020; Nikolov et al., 2019; Warwick et al., 2002). Due to the topological incongruences, the nuclear- and chloroplast-based matrices in Sisymbrieae dataset were not concatenated and are presented as separate phylogenetic reconstructions in Figure 3 (ML and BI analyses) and Appendices 5 and 6 (ASTRAL analyses).

While the clades A (European *Sisymbrium strictissimum* and East Asian *S. luteum* and *S. yunnanense*), B (South African clade) and C (Caucasian clade) were recognised already by Warwick et al., 2002, an additional geographically well-defined Mediterranean clade D was retrieved in this study (Figures 3 and 4, Appendices 5 and 6). The minor topological change in the placement of

Sisymbrium loeselii and closer relationship to *S. erucastrifolium* in the chloroplast-based phylogenies fits well with the morphological characteristics shared between these two species. Contrarily, major topological difference in the placement of *Sisymbrium volgense* might be attributed to its hybrid nature, where the progenitor not only genetically (this study) but also morphologically highly resembles the putative maternal lineage *Sisymbrium polymorphum* s.l. (Dorofeyev, 1997). Surprisingly, according to the nuclear signals, genetically distantly related *Sisymbrium damascenum* is placed in close proximity to the *S. polymorphum* s.l. and *S. volgense* in the chloroplast-based tree. This relationship cannot be explained neither by morphological similarities nor by geographical proximity and further genome-wide research would be necessary to irrefutable elucidate this peculiar topological incongruence.

The initial study of Mutlu and Karakuş (2015) associated *Sisymbrium malatyanum* endemic to Turkey with central Asian *Sisymbrium brassiciforme* and *S. heteromallum*, albeit with poor statistical support. This placement could not be explained well either by morphology or by biogeography. Contrarily, a more comprehensive taxon sampling of this study placed it into a clade with S*isymbrium gaubae* and *S. leucocladum* with a high statistical support (0.75 LBS/0.99 BPP/90 BS) (Figure 3). Both of these species are endemic to Iran, and *S. gaubae* habitually resembles *S. malatyanum*, further supporting this phylogenetic placement. As for the pair *Sisymbrium integerrimum* and *Sisymbrium brassiciforme,* as early as 1968 Hedge already speculated that the two are very closely related species, where the former might represent a marginal form of the latter. Five shared identical nuclear-based loci support the formal synonymisation of these two names (Khodashenas and Assadi, 2007).

4.2 Differences in the dating and retrieval of multiple copies

There have been some substantial differences in the employed calibration methods, datasets and matrices. Huang et al. (2020) used a generally slowly evolving whole plastome datasets of a representative taxon sample of rosids, where reliable fossils are available, to constrain early nodes. Subsequently, they applied these now secondary constraints to the ITS-based tribe levels under the assumption that plastome and genome-based time estimations produce converging results (Hohmann et al., 2015; Huang et al., 2016). Contrarily, we applied the average ITS-based substitution rate for herbaceous annual/perennial plant species on a reduced taxon sample, covering not all Brassicaceae tribes, but only Lineage II with and without the two putative outgroups (Eutremeae and Thlaspideae). There were some differences in the time spans of crown ages of

Brassiceae, Isatideae and Thelypodieae between Huang et al. (2020) and our analysis (Tables 1a & 1b). These time frame shifts can be easily explained by different time constraint strategies with different backgrounds as well as by differences in the taxon sets. Interestingly, when only secondary calibration points from Huang et al. (2020) are used in our study, the crown age of Sisymbrieae does not correspond with the crown age inferred from Huang et al. (2020). This illustrates how much of an influence different taxon samples has on the retrieved time spans. Nevertheless, the crown age of Sisymbrieae remained relatively stable in the present study, regardless of the taxon sample and used constraints, indicating that at least within Sisymbrieae, secondary calibration based on plastome datasets indeed recovers highly congruent results with the ITS-based substitution rate calibration (Huang et al., 2020).

The whole Brassicaceae family is characterised by an α-WGD, which is only a recent one in the series of two additional older known WGDs that predated the split of Brassicaceae. These were followed by an extensive diploidisation, and shaped the evolutionary history of whole angiosperms and Brassicales, respectively (Barker et al., 2009; Fawcett et al., 2009; Franzke et al., 2011; Huang et al., 2020; Mandáková et al., 2017; Schranz et al., 2012). Recent studies have shown that approximately 40% of all Brassicaceae underwent an additional neopolyploidisation event that has not yet been followed by diploidisation (Hohmann et al., 2015; Kagale et al., 2014). Despite the fact that there is no evidence that *Sisymbrium* underwent a recent polyploidisation event, several nuclear encoded multi-copy loci (marked with * in the Appendix 1) in four species possessed two different copies, however, in all of the cases the copies clustered monophyletically. This indicates that the copies are probably remnants of polyploidisation events and are not of hybrid nature. While there were always only two copies that were retrieved from the majority of the clones, individual SNPs and chimeric copies were also detected (however only in 1–2 clones out of 15 analysed). These can be attributed to sequencing errors, paralogous nature of loci (as none of the investigated nuclear DNA loci were single copy) or cloning artefacts.

4.3 Biogeography of *Sisymbrium* in South Africa

We hypothesise that *Sisymbrium* migrated to southern Africa from the Irano-Turanian floristic region through the East African Riff Mountains sometime in the Pliocene (stem ages dated to approximately 3.5-4.0 MYA: Figure 4 and Figure 5). During that time, this region was under strong influence of Pliocene uplift, which arguably had the most prominent effect on its present day topography (Partridge and Maud, 1987). During the Late Pliocene, an asymmetrical uplift of the

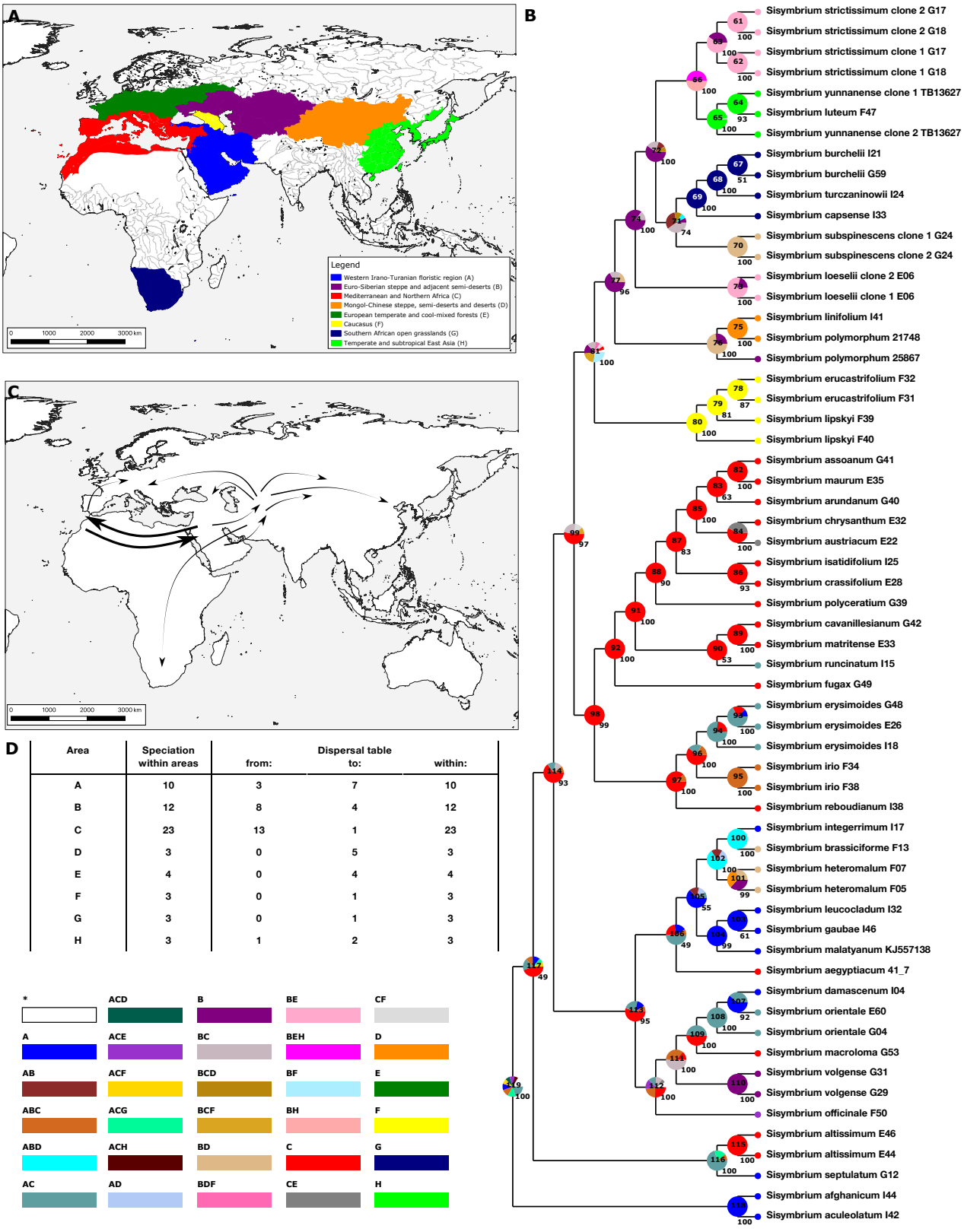


Figure 5: Ancestral range reconstruction, based on a dated nuclear DNA phylogenetic tree of Sisymbrieae retrieved from BEAST analysis under Coalescent tree prior. (A) Geographic representation of biogeographic entities, (B) tree representation of ancestral range reconstruction based on a time divergence tree retrieved from BEAST analysis under Coalescent tree prior, numbers in the middle of pie charts indicate node identity and correspond to the node identity in Appendix 7. Numbers next to pie charts correspond to node frequencies and codes next to species names refer to the isolation codes from Appendix 1, (C) visual representation of the most common range changes,

where the arrowheads indicate the direction of the distribution shift/expansion and thickness of the arrows the frequency, (D) speciation and dispersal per biogeographic entity defined in 5A. Legend in the bottom left corner correspond to the colours in 5A and 5B.

whole continent and the consequent formation of the Post-African II erosion surface took place (Partridge and Maud, 1987). These resulted in a development of a temperature as well as precipitation gradient. Middle and Late Pleistocene climate fluctuations have caused, together with the eustatic sea-level changes, a reduction and shift in the precipitation profile. During glacial cycles, the regions that nowadays experience summer rainfall were under winter rainfall regime (Hoag and Svenning, 2017; van Zinderen Bakker, 1978). Cold air masses that were produced more often due to the development of the Antarctic polar cap were penetrating deeply into the African subcontinent (van Zinderen Bakker, 1978). During glacial periods, open dry grassland and semi-desert environments were favoured (Knight and Grab, 2016; Partridge and Maud, 1987) in which dry-adapted *Sisymbrium* species could thrive. This is also supported by our time estimation analysis, which placed the crown age of South African *Sisymbrium* species into the same period (Figure 4). Plant disjunctions between the Mediterranean, Irano-Turanian floristic region, and southern Africa are a common phenomenon tightly connected to the xeric conditions that connect all three areas (Goldblatt, 1978). While the most common explanation for such a disjunction is the spread from southern Africa northwards (Désamoré et al., 2011; Durvasula et al., 2017; Klak et al., 2017; Moore et al., 2010; Valente et al., 2011), our dataset indicates that *Sisymbrium* occupied southern Africa from the Irano-Turanian floristic region, a more seldom pattern observed in only some other plant species (Carlson et al., 2012; Pyankov et al., 2002).

4.4 Biogeography of *Sisymbrium* in the Mediterranean

The Mediterranean climate regime we know today was established around 3.2 MYA (Suc, 1984). Changes in the precipitation reduction and development of seasonality resulted in a compositional and structural change of the Mediterranean forests that lost subtropical elements and started to resemble contemporary Mediterranean forests with drought-adapted species (Kondraskov et al., 2015; Thompson, 2005). Similarly to the Eurasian steppe belt, the Mediterranean basin was also under strong influence of Pleistocene climatic oscillations causing recurrent range shifts that have shaped current species' distributions (Hewitt, 2011; Weiss and Ferrand, 2007). Prior to the *Sisymbrium* migration into the Mediterranean, this basin was already under strong influence of gradual global cooling and aridification, which have been initiated as early as in the middle Miocene (van Dam, 2006; Zachos et al., 2008).

The Mediterranean migration of *Sisymbrium* from the western Irano-Turanian floristic region coincides with the Messinian Salinity Crisis, characterised by massive desiccation of the Mediterranean Sea, which started in the middle Miocene (early Serravallian, 13.0 MYA) and ended with the Zanclean Reflooding around 5.33 MYA (Fauquette et al., 2006; Krijgsman et al., 2010). Extensive pollen record dated to the same period also indicates that Anatolian steppes dominated by *Artemisia* and *Ephedra* spread into the Mediterranean from the western Irano-Turanian floristic region. The earliest documentations of open dry habitats in the southern Mediterranean are dated much later to roughly 3.5 MYA, when the mangrove forests dried up and gave space to open steppe vegetation. Approximately 2.5 MYA, the mesophilous forests of the northern Mediterranean drew northwards making room for more open environments known as *Artemisia* steppes (Suc et al., 2018). The peak of the Messinian salinity crisis followed by stepwise expansion of drier open habitats coincides with the deeper nodes indicating diversification of Mediterranean *Sisymbrium* species, which started around the transition to the Pleistocene. The onset of modern open grasslands on the Iberian Peninsula is dated to around 1.0 MYA (González-Sampériz et al., 2010). With its complex topography and relatively stable climate during the Late Quaternary, the Iberian Peninsula served as a Pleistocene refugial centre for numerous plant species (Dubreuil et al., 2008; Heredia et al., 2007; Médail and Diadema, 2009; Pérez−Collazos et al., 2009) and could have potentially endorsed diversification of *Sisymbrium* species. High number of *Sisymbrium* species (up to 12, depending on the taxonomic treatments) on the Iberian Peninsula (Ball 1964, Pujadas Salvá 1993) also points towards this region being a shelter during the unfavourable conditions (Figures 4 and 5). High number of speciation and dispersal events within and from this region furthermore reflects an extremely dynamic nature of the Mediterranean (Figure 5B and D). A clade consisting of Mediterranean *Sisymbrium damascenum*, *S. orientale* and *S. macroloma* points towards another independent invasion of the Mediterranean that took place after the end of the Messinian salinity crisis and continued throughout the Pleistocene. Contemporary distribution patterns of these species again indicate that the invasion probably started from the western Irano-Turanian region (Figures 4 and 5).

4.5 Biogeography of *Sisymbrium* in the Irano-Turanian region & Caucasus

The western part of the Irano-Turanian floristic region is one of the global biodiversity hotspots of the Old World (Mittermeier et al., 2000) and a putative origin of several important crop wild relatives, including the family of Brassicaceae (Franzke et al., 2011; Hedge, 1976; Karl and Koch, 2013; Noroozi et al., 2019). Furthermore, due to its position connecting the eastern and western

Eurasian floras, this region has also been proposed to be the main source of plant taxa for adjacent floristic regions, especially the Mediterranean (Herrera, 2010; Jabbour and Renner, 2012; Manafzadeh et al., 2014; Quezel, 1985; Thompson, 2005, this study).

This region has experienced a predominantly dry and stable climate since the early Eocene (Manafzadeh et al., 2016). Mountain ranges of Alborz, Zagros, Kopet Dagh and Pamir had formed from the middle Miocene through the early Pliocene creating considerable habitat heterogeneity (Manafzadeh et al., 2016). Furthermore, climate stability of this region might also sustain genetic diversity, as it deters from climate-dependent extinction (Cowling et al., 2015; Galley et al., 2009; Schwery et al., 2015). Thus, it is not surprising that western Irano-Turanian floristic region harbours approximately 27000 species, and up to 40% of representatives that constitute this flora are endemic (Sales and Hedge, 2013; Takhtajan et al., 1986). This pattern can also be applied to *Sisymbrium*. Fourteen Sisymbrieae species (approximately 30% of the whole tribe) occur in the western part of the Irano-Turanian floristic region, and half of them (sensu BrassiBase) are confined to this floristic region. Furthermore, ten individual speciation events within this area were inferred (third highest after Mediterranean) and the ancestral range reconstruction put the origin of this tribus somewhere into the Mediterranean and Irano-Turanian floristic region (Figure 5), again supporting the hypothesis that this tribe (as well as the whole family) originated in the Irano-Turanian floristic region.

Adjacent to the Irano-Turanian floristic region is another biodiversity hotspot of the Old World — the Caucasus. Currently, there are around 7500 plant taxa described from the Caucasus region, with 35% having a status of an endemic species (Gagnidze et al., 2002; Schatz et al., 2009), including four *Sisymbrium* taxa. Palaeontological, palaeoclimatical and genetic data point out that the Caucasus region was one of the main glacial refugial centres for fauna and dendroflora during the Quaternary together with Iberian, Italian and Balkan Peninsula (Hewitt, 2000; 2011; Stewart et al., 2010). Furthermore, intermontane basins and high mountain plateaus of the greater Caucasus region were also recognised as a cryptic southern refugium for dry- and cold-adapted species during the interglacial periods (Fjellheim et al., 2006; Skrede et al., 2006). The exact extent of Pliocene and Pleistocene glaciations of the Caucasus is, however, unclear (Milanovsky, 2008; Mitchell and Westaway, 1999). Nevertheless, recent isotope dating studies and an overall high level of endemism in this region (Gagnidze et al., 2002; Nakhutsrisvili et al., 2017) suggest that the Caucasian glaciations were not as vast as previously thought. Pliocene to Pleistocene transition had a tremendous effect on the vegetation composition of the Greater Caucasus area. With the decrease in

temperature and humidity, the thermophilic forests flora was gradually replaced by a more robust temperate dendroflora without a modern analogue (Klopotovskaya, 1973; Naidina and Richards, 2016). Gradual intensification of the continental climate and reduced humidity caused an expansion of forest steppes and steppes through the Greater Caucasus area. Palaeovegetational analyses dated this expansion to be not older than 1.8 MYA (Connor and Kvavadze, 2009; Joannin et al., 2010; Messager et al., 2013; Naidina and Richards, 2016; Tagieva et al., 2013). This coincides with *Sisymbrium* clade endemic to Caucasus region, which started to diverge around the same time and its most recent common ancestor that invaded this area through East European Plain in the Pliocene (Figures 4 and 5).

4.6 Biogeography of *Sisymbrium* in Eurasia

A peculiar relationship is observed between European *Sisymbrium strictissimum* and its sister clade, consisting of East Asian *Sisymbrium luteum* and *S. yunnanense*. All three species exhibit similar morphology (Vassilczenko, 1939; Zhou et al., 2001) and distinctive habitat preference. While most of the other *Sisymbrium* species are either widely distributed weeds or show a tendency towards open dry habitats, these three species can be found in more humid and predominantly shaded habitats in forests, thickets, ravines or mountain slopes near water bodies (Brach and Song, 2006; Brandes, 1991). Assuming that the ecology of the last common ancestor of these species did not vary greatly, the expansion of the clade through the late Miocene/early Pliocene forests of central Asia is highly probable. During this era, most of Europe was covered in warm-temperate evergreen broadleaved and mixed forest, Western and Central Asia (between 45°N and 53°N) in temperate deciduous broadleaf forests, while the coastline of Eastern Asia was dominated by similar forest types as present in Europe, which transitioned into temperate deciduous broadleaved savanna biome in the inland (Bezrukova et al., 2003; Haywood et al., 2002; Ivanov et al., 2011; Pound et al., 2012). According to our study, the diversification into two distinct contemporary eastern (*Sisymbrium luteum* and *S. yunnanense*) and western lineage (*Sisymbrium strictissimum*) took place around early Pleistocene (Figures 4 and 5). This coincides with the era when the continuous forest belt started to fall apart due to the increased aridity and lower temperatures (Demske et al., 2002; Frenzel, 1968; Velichko, 2005).

Despite their wide distribution range, the fast evolving ITS locus in *Sisymbrium loeselii* and *S. officinale* remained monomorphic across the species' whole distribution area. While the latter showed no ambiguous signals in either of the investigated loci, all *Sisymbrium loesellii* accessions

had to be cloned to retrieve two unambiguous ITS copies. Because the multi copy ITS region is under strong influence of concerted evolution (Koch et al., 2003), the persisting two ITS copies might be an indicator of a recent hybridisation event through which *S. loesellii* emerged. The young age is also supported by our time estimation analyses, placing the split between the two coexisting ITS copies into the western Euro-Siberian steppe of the middle Pleistocene (Figures 4 and 5). Both species are nowadays distributed worldwide and can be commonly found near human settlements, along roadsides and hedges, or growing on heavily nitrogen contaminated dry and loamy soils (Bobrov et al., 1985; Malyschev et al., 2004). Exhibiting a weed-like ecology, these two species are excellent competitors that can successfully occupy wide ranges of disturbed habitats in a short period of time (Holzner and Numata, 2013). Therefore, quick and extensive repeated migrations through human migratory pathways might be accounted for the lack of geographically correlated genetic footprint even in the fast evolving ITS region. Nevertheless, sound conclusions on this issue can be drawn based only on extensive next generation sequencing datasets and proper sampling efforts.

In contrast to *Sisymbrium loeselii* and *S. officinale*, some *Sisymbrium* species exhibit geographically correlated diverging ITS copies. Genetic splits found in widely distributed species were mostly of younger age indicating a potential influence of middle and late Pleistocene events. This era is characterised by an intensified development of permafrost and continental climate, major trans- as well as regressions of water bodies and major continental glaciations that got gradually reduced towards the end of the Pleistocene (Head et al., 2008a, b; Velichko, 2005). These parameters have shaped the evolutionary history also of other plant taxa, such as *Adonis* (Kropf et al., 2019), *Camelina* (Žerdoner Čalasan et al., 2019), *Linum* (Plenk et al., 2017) and *Schivereckia* (Friesen et al., 2020). This could likewise be the case in a two-way split in *Sisymbrium altissimum*, *S. irio*, *S. volgense* or three-way split in predominantly Mediterranean *S. erysimioides*, all dated to the late or the end of middle Pleistocene. A slightly older split can be inferred from *Sisymbrium heteromallum* dating to the middle Pleistocene in the continental Asia, and an even older split between *S. polymorphum* accessions, dated to late Miocene/early Pliocene into Middle and Central Asia (Figures 4 and 5). The latter one is of great interest for two reasons. Firstly, *Sisymbrium polymorphum* is, as the epithet already suggests, morphologically as well as genetically an extremely variable species. This is also illustrated in our case, where *Sisymbrium linifolium* is retrieved as a sister group to the eastern group of *S. polymorphum* s.l. accessions, rendering the *S. polymorphum* paraphyletic. This paraphyly also found in Chen et al., 2019, coupled with immense

morphological variability of this species group, indicates a complex problem, which is out of the scope of this study. Secondly, *Sisymbrium polymorphum* is a widely distributed species found solely across the entire Eurasian steppe belt. Assuming that its ecology has not changed extensively since the past makes the species a suitable proxy to infer past florogenetic patterns of the Eurasian steppe belt. Molecular signals in typical steppe plant species reflect the climate-landscape history of the steppe and thus allow for a finer resolution of the history of the steppe belt in comparison to floristic and fossil-based methods. Overall, all these splits cannot be rigorously assigned to certain geological events due to a limited taxon sampling. Nevertheless, our study might serve as the starting point for other *Sisymbrium*-specific studies that investigate different aspects of the onset and development of different Eurasian geographical entities.

4.7 Taxonomic update on *Sisymbrium*

In view of the above-mentioned paraphyly of *Sisymbrium* that embeds the single species of *Ochthodium*, the latter genus is synonymised here with the prior one. Before molecular phylogenetic studies, proximity of the two genera has never been assumed due to considerable differences in the fruit and seed characters (many-seeded dehiscent linear-cylindric or conical siliques with smooth papery valves and seeds with incumbent cotyledons in *Sisymbrium* vs. two-seeded indehiscent ellipsoid to subglobose silicles with verrucose corky valves and seeds with accumbent cotyledons in *Ochthodium*) that were traditionally the key features in the systematics of the family. It is therefore understandable that Warwick et al. (2010), who first revealed the phylogenetic proximity of the two genera, plead for the additional study of *Ochthodium* and left it unassigned to a tribe instead of including it in Sisymbrieae. However, the rest of characters such as habit and life form (rosette-forming biennial), indumentum of simple trichomes, leaf morphology (lyrate-runcinate), flowers (yellow, middle-sized, widely open) fit well the molecular signals. Hence, it is just another case of evolutionary recent drastic deviation of fruit character(s) not accompanied by serious genetic changes and alterations of other morphological features, a phenomenon known in various other groups of Brassicaceae, e.g., *Twisselmannia* Al-Shehbaz vs. *Tropidocarpum* Hook. of Descurainieae Al-Shehbaz et al. (Al-Shehbaz, 2003), *Drabopsis* K. Koch vs. *Draba* L. of Arabideae DC. (Al-Shehbaz & Koch, 2003), several segregate genera vs. *Heliophila* L. of Heliophileae DC. (Mummenhoff et al., 2005) or *Tchihatchewia* Boiss. vs. *Hesperis* L. of Hesperideae DC. (German & Al-Shehbaz, 2018), to name a couple of examples. Thus, the fact that in anthesis *Ochthodium* is indistinguishable from *Sisymbrium* is obviously not a coincidence.

Sisymbrium L., Sp. Pl.: 657. 1753.

= *Ochthodium* DC., in Mém. Mus. Hist. Nat. 7: 236. 1821, syn. nov.

Sisymbrium aegyptiacum (L.) German, Zerdoner & Al-Shehbaz, comb. nov. ≡ *Bunias aegyptiaca* L., Syst. Nat., ed. 12, 3: 231. 1768. ≡ *Ochthodium aegyptiacum* (L.) DC., Syst. Nat. [Candolle] 2: 423. 1821.

5 Acknowledgements

The authors thank all the curators of the following herbaria, who kindly provided the material for sequencing: B, MSB, GDA, GDAC, HBG, HEID, HUJ, K, L, P, MW, PRE and WAG. Furthermore, we would also like to thank the 'Plantarium' web-community members, who allowed us to use their photographs for Figure 1: A. Manchenko, P. Kosachev, A. Fateryga, I. Tabunova, V. Gumenyuk, E. Zakharov, V. Epiktetov and V. Volcotrub. This work was supported by the German Research Foundation [grant to BN and HH: NE 314/15-1].

Conflict of interest

The authors declare no conflict of interest.

Authors' contributions

Anže Žerdoner Čalasan: conceptualisation (equal); formal analysis (lead); investigation (lead); visualisation (lead); writing – original draft preparation (lead); writing – review and editing (lead). **Herbert Hurka:** conceptualisation (equal); writing – review and editing (equal); funding acquisition (equal). **Dmitry A. German:** writing – review and editing (equal). **Barbara Neuffer:** conceptualisation (equal); writing – review and editing (equal); funding acquisition (equal)

Data accessibility statement

DNA sequences: Genbank accessions MW270940–MW270994, MW271628–MW271779, MW280305–MW280306, MW281257–MW281310, MW319077–MW319132, MW319133–MW319187, MW345417–MW345468, MW345469–MW345520, MW355774–MW355817 & MW331329–MW331436.

Ethical approval

This article does not contain any studies with animals carried out by any of the authors.

Sampling and field studies

The study was performed in compliance with the Convention on Biological Diversity (CBD).

6 References

Agakhanjanz, O. & Breckle, S.-W. (1995). Origin and Evolution of the Mountain Flora in Middle Asia and Neighbouring Mountain Regions. In: Chapin, F. S. & Körner, C. (Eds.). Arctic and Alpine Biodiversity: Patterns, Causes and Ecosystem Consequences. Springer Berlin Heidelberg, Berlin, Heidelberg, Germany, pp 63–80.

Al-Shehbaz, I. A. (1977). Protogyny in the Cruciferae. *Systematic Botany*, **2**, 327–333. https://www.jstor.org/stable/2418468

Al-Shehbaz, I. A. (1985). The genera of Brassiceae (Cruciferae; Brassicaceae) in the southeastern United States. *Journal of Arnold Arboretum*, **66**, 279–351. https://www.jstor.org/stable/43782168

Al-Shehbaz, I. A. (1988). The genera of Sisymbrieae (Cruciferae; Brassicaceae) in the southeastern United States. *Journal of Arnold Arboretum*, **69**, 213–237. https://www.jstor.org/stable/43817458

Al-Shehbaz, I. A. (2003). A synopsis of *Tropidocarpum* (Brassicaceae). *Novon*, **13**, 392–395.

Al-Shehbaz, I. A. & Koch, M. A. (2003). *Drabopsis* is united with *Draba* (Brassicaceae). *Novon*, **13**, 172–173.

Al-Shehbaz, I. A. (2004a). A Synopsis of the South American *Neuontobotrys* (Brassicaceae). *Novon*, **14**, 253–257. https://www.jstor.org/stable/3393279

Al-Shehbaz, I. A. (2004b). A Synopsis of the South American *Weberbauera* (Brassicaceae). *Novon* **14**, 258–268. https://www.jstor.org/stable/3393280

Al-Shehbaz, I. A. (2006a). *Sisymbrium lactucoides* belongs to *Dicryophragmus* (Brassicaceae). *Harvard Papers in Botany*, **11**, 89–90.

Al-Shehbaz, I. A. (2006b). The genus *Sisymbrium* in South America, with synopses of the genera *Chilocardamum*, *Mostacillastrum*, *Neuontobotrys*, and *Polypsecadium* (Brassicaceae). *Darwiniana*, **44**, 341–358.

Al-Shehbaz, I. A. (2012). A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon*, **61**, 931–954. https://doi.org/10.1002/tax.615002

Al-Shehbaz, I. A. (2015). Brassicaceae. In: Hong, D. Y. (Ed.). Flora of Pan-Himalaya, Vol. 50. Cambridge University Press, Cambridge; Science Press, Beijing, China, 595 pp.

Al-Shehbaz, I. A., Beilstein, M. A. & Kellogg, E. A. (2006). Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Systematics and Evolution*, **259**, 89–120. https://doi.org/10.1007/s00606-006-0415-z

Appel, O. & Al-Shehbaz, I. A. (2002). Cruciferae. In Kubitzki, K. (Ed.). Families and Genera of Vascular Plants, vol. 5. Springer-Verlag, Berlin, Heidelberg, Germany, pp 75–174.

Arias, T. & Pires, J. C. (2012). A fully resolved chloroplast phylogeny of the brassica crops and wild relatives (Brassicaceae: Brassiceae): novel clades and potential taxonomic implications. *Taxon*, **61**, 980–988. https://doi.org/10.1002/tax.615005

Avise, J. C. (2004). What is the field of biogeography, and where is it going? *Taxon*, **53**, 893–898. https://www.jstor.org/stable/4135555

Ball, P. W. (1964). *Sisymbrium* L. In: Tutin, T. G., Heywood, V. H., Burges, N. A., Valentine, D. H., Walters, S. M. & Webb, D. A. (Eds.) Flora Europaea, vol. 1. Cambridge University Press, Cambridge, United Kingdom, pp 264–266.

Barker, M. S., Vogel, H. & Schranz, M. E. (2009). Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biology and Evolution*, **1**, 391–399. https://doi.org/10.1093/gbe/evp040

Barres, L., Sanmartín, I., Anderson, C.L., Susanna, A., Buerki, S., Galbany‐Casals, M. & Vilatersana, R. (2013). Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). *American Journal of Botany*, **100**, 867–882. https://doi.org/10.3732/ajb.1200058

Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R. & Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 18724–18728. https://doi.org/10.1073/pnas.0909766107

Bezrukova, E. V., Kulagina, N. V., Letunova, P. P., Karabanov, E. B., Williams, D. F., Kuzmin, M. I., Krapivina, S. M., Vershinin, K. E. & Shestakova, O. N. (2003). Pliocene-Quaternary Vegetation and Climate History of the Lake Baikal Area, Eastern Siberia. In: Kashiwaya, K. (Ed.), Long Continental Records from Lake Baikal. Springer Japan, Tokyo, Japan, pp 111–122.

Birky, C. W. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the United States of America,* **92**, 11331–11338. https://doi.org/10.1073/pnas.92.25.11331

Blanca, G., Cueto, M. & Fuentes, J. (2015). *Sisymbrium isatidifolium* (Brassicaceae): a new species from southern Spain, and the identity of *S. hispanicum* Jacq. *Phytotaxa*, **220**, 43–53. https://doi.org/10.11646/phytotaxa.220.1.3

Blažević, I., Radonić, A., Mastelić, J., Zekić, M., Skočibušić, M. & Maravić, A. (2010). Hedge mustard (*Sisymbrium officinale*): chemical diversity of volatiles and their antimicrobial activity. *Chemistry & Biodiversity*, **7**, 2023–2034. https://doi.org/10.1002/cbdv.200900234

Bobek, H. (1937). Die Rolle der Eiszeit in Nordwestiran. *Zeitschrift für Gletscherkunde und Glazialgeologie*, **25**, 130–183

Bobek, H. (1953). Zur eiszeitlichen Vergletscherung des Alborzgebirges, Nordiran. *Carinthia*, **2**, 97–104

Brach, A. R. & Song, H. (2006). eFloras: New directions for online floras exemplified by the Flora of China Project. *Taxon*, **55**, 188–192. https://doi.org/10.2307/25065540

Brandes, D. (1991). Untersuchungen zur Ökologie und Soziologie von *Sisymbrium strictissimum* in Mitteleuropa. *Tuexenia*, **11**, 35–48.

Carlson, S. E., Linder, H. P. & Donoghue, M. J. (2012). The historical biogeography of *Scabiosa* (Dipsacaceae): implications for Old World plant disjunctions. *Journal of Biogeography*, **39**, 1086–1100. https://doi.org/10.1111/j.1365-2699.2011.02669.x

Chen, H., Al-Shehbaz, I. A., Yue, J. & Sun, H. (2019). *Sisymbrium linifolium* and *Sisymbriopsis schugnana* (Brassicaceae), two new records from Xinjiang, China. *PhytoKeys*, **119**, 39–52. https://doi.org/10.3897/phytokeys.119.32985

Connor, S. E. & Kvavadze, E. V., 2009. Modelling late Quaternary changes in plant distribution, vegetation and climate using pollen data from Georgia, Caucasus. *Journal of Biogeography,* **36**, 529–545. https://doi.org/10.1111/j.1365-2699.2008.02019.x

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

Cowling, R. M., Potts, A. J., Bradshaw, P. L., Colville, J., Arianoutsou, M., Ferrier, S., Forest, F., Fyllas, N. M., Hopper, S. D., Ojeda, F., Procheş, Ş., Smith, R. J., Rundel, P. W., Vassilakis, E. & Zutta, B. R. (2015). Variation in plant diversity in mediterranean-climate ecosystems: the role of climatic and topographical stability. *Journal of Biogeography*, **42**, 552–564. https://doi.org/10.1111/jbi.12429

Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and high-performance computing. *Nature Methods*, **9**, 772. https://doi.org/10.1038/nmeth.2109

Demske, D., Mohr, B. and Oberhänsli, H. (2002). Late Pliocene vegetation and climate of the Lake Baikal region, southern East Siberia, reconstructed from palynological data. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **184**, 107–129. https://doi.org/10.1016/S0031-0182(02)00251-1

Désamoré, A., Laenen, B., Devos, N., Popp, M., González−Mancebo, J. M., Carine, M. A. and Vanderpoorten, A. (2011). Out of Africa: north-westwards Pleistocene expansions of the heather

Erica arborea. *Journal of Biogeography*, **38**, 164–176. https://doi.org/10.1111/j.1365-2699.2010.02387.x

Di Sotto, A., Vitalone, A., Nicoletti, M., Piccin, A. & Mazzanti, G. (2010). Pharmacological and phytochemical study on a *Sisymbrium officinale* Scop. extract. *Journal of Ethnopharmacology*, **127**, 731–736. https://doi.org/10.1016/j.jep.2009.12.001

Djamali, M., Baumel, A., Brewer, S., Jackson, S. T., Kadereit, J. W., López-Vinyallonga, S., Mehregan, I., Shabanian, E. & Simakova, A. (2012a). Ecological implications of *Cousinia* Cass. (Asteraceae) persistence through the last two glacial-interglacial cycles in the continental Middle East for the Irano-Turanian flora. *Review of Palaeobotany and Palynology*, **172**, 10–20. https://doi.org/10.1016/j.revpalbo.2012.01.005

Djamali, M., Brewer, S., Breckle, S. W. & Jackson, S. T. (2012b). Climatic determinism in phytogeographic regionalization: a test from the Irano-Turanian region, SW and Central Asia. *Flora: Morphology, Distribution, Functional Ecology of Plants*, **207**, 237–249. https://doi.org/10.1016/j.flora.2012.01.009

Dorofeyev, V. I. (1997). Genus *Sisymbrium* (Brassicaceae) of the Caucasian flora. *Botanicheskii Zhurnal SSSR*, **82**, 106–110.

Dubreuil, M., Riba, M. & Mayol, M. (2008). Genetic structure and diversity in *Ramonda myconi* (Gesneriaceae): effects of historical climate change on a preglacial relict species. *American Journal of Botany*, **95**, 577–587. https://doi.org/10.3732/ajb.2007320

Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., Tsuchimatsu, T., Burbano, H. A., Picó, F.X., Alonso-Blanco, C. & Hancock, A. M. (2017). African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 5213–5218. https://doi.org/10.1073/pnas.1616736114

Fahey, J. W., Zalcmann, A. T. & Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*, **56**, 5–51. https://doi.org/10.1016/S0031-9422(00)00316-2

Fauquette, S., Suc, J.-P., Bertini, A., Popescu, S.-M., Warny, S., Bachiri Taoufiq, N., Perez Villa, M.-J., Chikhi, H., Feddi, N., Subally, D., Clauzon, G. & Ferrier, J. (2006). How much did climate force the Messinian salinity crisis? Quantified climatic conditions from pollen records in the Mediterranean region. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **238**, 281–301. https://doi.org/10.1016/j.palaeo.2006.03.029

Fawcett, J. A., Maere, S. & Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 5737–5742. https://doi.org/10.1073/pnas.0900906106

Fjellheim, S., Rognli, O. A., Fosnes, K. & Brochmann, C. (2006). Phylogeographical history of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. *Journal of Biogeography*, **33**, 1470–1478. https://doi.org/10.1111/j.1365-2699.2006.01521.x

Franzke, A., Lysak, M. A., Al-Shehbaz, I. A., Koch, M. A. & Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in Plant Science*, **16**, 108–116. https://doi.org/10.1016/j.tplants.2010.11.005

Frenzel, B. (1968). The Pleistocene Vegetation of Northern Eurasia. *Science*, **161**, 637–649. https://doi.org/10.1126/science.161.3842.637

Friesen, N., Žerdoner Čalasan, A., Neuffer, B., German D. A., Markov, M. & Hurka, H. (2020). Evolutionary history of the Eurasian steppe plant *Schivereckia podolica* (Brassicaceae) and its close relatives. *Flora: Morphology, Distribution, Functional Ecology of Plants,* **268**, 151602. https://doi.org/10.1016/j.flora.2020.151602

Gagnidze, R., Gviniashvili, T., Shetekauri, S. & Margalitadze, N. (2002). Endemic genera of the Caucasian flora. *Feddes Repertorium*,**113**, 616–630. https://doi.org/10.1002/fedr.200290009

Galley, C., Linder, H. P. & Zimmermann, N.E. (2009). *Pentaschistis* (Poaceae) diversity in the Cape mediterranean region: habitat heterogeneity and climate stability. *Global Ecology and Biogeography*, **18**, 586–595. https://doi.org/10.1111/j.1466-8238.2009.00468.x

German, D. A. & Al-Shehbaz, I.A. (2018). A reconsideration of *Pseudofortuynia* and *Tchihatchewia* as synonyms of *Sisymbrium* and *Hesperis*, respectively (Brassicaceae). *Phytotaxa*, **334**, 95–98. https://doi.org/10.11646/phytotaxa.334.1.17

German, D. A., Friesen, N., Neuffer, B., Al-Shehbaz, I. A. & Hurka, H. (2009). Contribution to ITS phylogeny of the Brassicaceae, with special reference to some Asian taxa. *Plant Systematics and Evolution*, **283**, 33–56. https://doi.org/10.1007/s00606-009-0213-5

Goldblatt, P. (1978). An analysis of the flora of Southern Africa: its characteristics, relationships, and origins. *Annals of the Missouri Botanical Garden*, **65**, 369–436. https://doi.org/10.2307/2398858

Gomez-Campo, C. (1980) Morphology and Morpho-taxonomy of the Tribe Brassiceae. In: Tsunoda, S., Hinata, K. & Gomez-Campo, C. (Eds.). *Brassica* Crops and Wild Allies, Biology and Breeding. Japan Scientific Societies Press, Tokyo, Japan, pp 3–31.

González-Sampériz, P., Leroy, S. A. G., Carrión, J. S., Fernández, S., García-Antón, M., Gil-García, M. J., Uzquiano, P., Valero-Garcés, B. & Figueiral, I. (2010). Steppes, savannahs, forests and phytodiversity reservoirs during the Pleistocene in the Iberian Peninsula. *Review of Palaeobotany and Palynology, Iberian Floras through Time: Land of Diversity and Survival*, **162**, 427–457. https://doi.org/10.1016/j.revpalbo.2010.03.009

Grubov, V. I. (2001). Key to the vascular plants of Mongolia. Science Publishers Enfield, New Hampshire, United States of America, 817 pp.

Guzmán, B. & Vargas, P. (2005). Systematics, character evolution, and biogeography of *Cistus* L. (Cistaceae) based on ITS, trnL-trnF, and matK sequences. *Molecular Phylogenetics and Evolution*, **37**, 644–660. https://doi.org/10.1016/j.ympev.2005.04.026

Hall, J. C., Sytsma, K. J. & Iltis, H. H. (2002). Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. *American Journal of Botany*, **89**, 1826–1842. https://doi.org/10.3732/ajb.89.11.1826

Haywood, A. M., Valdes, P. J., Francis, J. E. and Sellwood, B. W. (2002). Global middle Pliocene biome reconstruction: A data/model synthesis. *Geochemistry, Geophysics, Geosystems*, **3**, 1–18. https://doi.org/10.1029/2002GC000358

Head, M. J., Gibbard, P. & Salvador A. (2008a). The Quaternary: its character and definition. *Episodes*, **31**, 234–238.

Head, M. J., Pillans, B. & Farquhar S. (2008b). The Early-Middle Pleistocene transition: characterization and pro- posed guide for the defining boundary. *Episodes*, **31**, 255–259.

Hedge, I. C. (1976). Systematic and Geographical Survey of the Old World Cruciferae. In: Vaughan, J. G., Macleod, J. & Jones, B. M. G. (Eds.). Biology and Chemistry of the Cruciferae. New York Academic Press, London, United Kingdom, pp 1–45.

Heredia, U. L. de Carrión, J. S., Jiménez, P., Collada, C. & Gil, L. (2007). Molecular and palaeoecological evidence for multiple glacial refugia for evergreen oaks on the Iberian Peninsula. *Journal of Biogeography*, **34**, 1505–1517. https://doi.org/10.1111/j.1365-2699.2007.01715.x

Herrera, C. M. (2010). The Mediterranean region: biological diversity in space and time. *The Quarterly Review of Biology*, **85**, 497–497. https://doi.org/10.1086/656852

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913. https://doi.org/10.1038/35016000

Hewitt, G. M. (2011). Mediterranean Peninsulas: The Evolution of Hotspots. In: Zachos, F. E. & Habel, J. C. (Eds.). Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas. Springer Berlin, Heidelberg, Germany, pp 123–147.

Hoag, C. & Svenning, J.-C. (2017). African environmental change from the Pleistocene to the Anthropocene. *Annual Review of Environment and Resources*, **42**, 27–54. https://doi.org/10.1146/annurev-environ-102016-060653

Hohmann, N., Wolf, E. M., Lysak, M. A. & Koch, M. A. (2015). A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell*, **27**, 2770–2784. https://doi.org/10.1105/tpc.15.00482

Holzner, W. & Numata, M. (2013). Biology and Ecology of Weeds. Springer Science & Business Media, Cham, Switzerland, 461 pp.

Huang, C.-H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M. A., Al-Shehbaz, I. A., 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

Huang, X.-C., German, D. A. & Koch, M. A. (2020). Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. *Annals of Botany*, **125**, 29–47. https://doi.org/10.1093/aob/mcz123

Hurka, H., Friesen, N., Bernhardt, K.-G., Neuffer, B., Smirnov, S., Shmakov, A. & Blattner, F. (2019). The Eurasian steppe belt: status quo, origin and evolutionary history. *Turczaninowia*, **22**, 5–71. https://doi.org/10.14258/turczaninowia.22.3.1

Ivanov, D., Utescher, T., Mosbrugger, V., Syabryaj, S., Djordjević-Milutinović, D. & Molchanoff, S. (2011). Miocene vegetation and climate dynamics in Eastern and Central Paratethys (Southeastern Europe). *Palaeogeography, Palaeoclimatology, Palaeoecology, The Neogene of Eurasia: Spatial gradients and temporal trends*, **304**, 262–275. https://doi.org/10.1016/j.palaeo.2010.07.006

Inkscape Project. (2020). Inkscape. Retrieved from https://inkscape.org

Jabbour, F. & Renner, S. S. (2012). A phylogeny of Delphinieae (Ranunculaceae) shows that *Aconitum* is nested within *Delphinium* and that Late Miocene transitions to long life cycles in the Himalayas and Southwest China coincide with bursts in diversification. *Molecular Phylogenetics and Evolution*, **62**, 928–942. https://doi.org/10.1016/j.ympev.2011.12.005

Jalas, J. & Suominen, J. (1994). Atlas Florae Europaeae. Distribution of Vascular Plants in Europe, vol. 10. Cruciferae (*Sisymbrium* to *Aubrieta*). — The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki, Finland, 224 pp.

Jehlík, V. (1981). Chorology and ecology of *Sisymbrium volgense* in Czechoslovakia. *Folia Geobotanica & Phytotaxonomica*, **16**, 407–421. https://www.jstor.org/stable/4180295

Joannin, S., Cornée, J.-J., Münch, P., Fornari, M., Vasiliev, I., Krijgsman, W., Nahapetyan, S., Gabrielyan, I., Ollivier, V., Roiron, P. & Chataigner, C. (2010). Early Pleistocene climate cycles in continental deposits of the Lesser Caucasus of Armenia inferred from palynology, magnetostratigraphy, and 40Ar/39Ar dating. *Earth and Planetary Science Letters*, **291**, 149–158. https://doi.org/10.1016/j.epsl.2010.01.007

Kagale, S., Robinson, S. J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W. E., Edger, P. P., Links, M. G., Sharpe, A. G. & Parkin, I. A. P. (2014). Polyploid evolution of the Brassicaceae during the Cenozoic Era. *Plant Cell*, **26**, 2777–2791. https://doi.org/10.1105/tpc.114.126391

Karl, R. & Koch, M. A. (2013). A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae. *Annals of Botany*, **112**, 983–1001. https://doi.org/10.1093/aob/mct165

Kay, K. M., Whittall, J. B. & Hodges, S. A. (2006). A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evolutionary Biology*, **6**, 36. https://doi.org/10.1186/1471-2148-6-36

Khodashenas, M. & Assadi, M. (2007). A taxonomic revision of the genus *Sisymbrium* (Brassicaceae) in Iran. *Iranian Journal of Botany*, **13**, 49-52.

Khoshoo, T. N. (1959). Biosystematics of *Sisymbrium irio c*omplex. IV: Origin of triploid, hexaploid and octoploid. *Caryologia*, **11**, 319–329. https://doi.org/10.1080/00087114.1959.10797052

Kiefer, M., Schmickl, R., German, D. A., Mandáková, T., Lysak, M. A., Al-Shehbaz, I. A., Franzke, A., Mummenhoff, K., Stamatakis, A. & Koch, M. A. (2014). BrassiBase: Introduction to a novel knowledge database on Brassicaceae evolution. *Plant and Cell Physiology*, **55**, e3. https://doi.org/10.1093/pcp/pct158

Klak, C., Hanáček, P. & Bruyns, P. V. (2017). Out of southern Africa: origin, biogeography and age of the Aizooideae (Aizoaceae). *Molecular Phylogenetics and Evolution*, **109**, 203–216. https://doi.org/10.1016/j.ympev.2016.12.016

Klopotovskaya, N. B. (1973). The Main Regularities of the Formation of Spore-pollen Spectra in the Mountainous Regions of the Caucasus. Publishing House Metsniereba, Tbilisi, Georgia, 184 pp.

Knight, J. & Grab, S. W. (2016). A Continental-scale Perspective on Landscape Evolution in Southern Africa During the Cenozoic. In: Knight, J. (Ed.). Quaternary Environmental Change in Southern Africa Physical and Human Dimensions. Cambridge University Press, Cambridge, United Kingdom, pp 30–46.

Koch, M. A., Dobeš, C. & Mitchell-Olds, T. (2003). Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). *Molecular Biology and Evolution*, **20**, 338–350. https://doi.org/10.1093/molbev/msg04

Koch, M. A., Dobeš, C., Kiefer, C., Schmickl, R., Klimeš, L. & Lysak, M. A. (2007). Supernetwork identifies multiple events of plastid trnF(GAA) pseudogene evolution in the Brassicaceae. *Molecular Biology and Evolution*, **24**, 63–73. https://doi.org/10.1093/molbev/msl130

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

Koch, M. A., Kiefer, M., German, D. A., Al-Shehbaz, I. A., Franzke, A., Mummenhoff, K. & Schmickl, R. (2012). BrassiBase: tools and biological resources to study characters and traits in the Brassicaceae—version 1.1. *Taxon*, **61**, 1001–1009. https://doi.org/10.1002/tax.615007

Kondraskov, P., Schütz, N., Schüßler, C., Sequeira, M. M., de Guerra, A. S., Caujapé-Castells, J., Jaén-Molina, R., Marrero-Rodríguez, Á., Koch, M. A., Linder, P., Kovar-Eder, J. & Thiv, M. (2015). Biogeography of Mediterranean hotspot biodiversity: re-evaluating the "Tertiary relict" hypothesis of Macaronesian laurel forests. *PLOS One*, **10**, e0132091. https://doi.org/10.1371/journal.pone.0132091

Krijgsman, W., Stoica, M., Vasiliev, I. & Popov, V. V. (2010). Rise and fall of the Paratethys Sea during the Messinian Salinity Crisis. *Earth and Planetary Science Letters*, **290**, 183–191. https://doi.org/10.1016/j.epsl.2009.12.020

Kropf, M., Bardy, K, Höhn, M & Plenk, K. (2019). Phylogeographical structure and genetic diversity of *Adonis vernalis* L. (Ranunculaceae) across and beyond the Pannonian region. *Flora: Morphology, Distribution, Functional Ecology of Plants*, **262**, 151497. https://doi.org/10.1016/j.flora.2019.151497

Kubatko, L. S. & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, **56**, 17–24. https://doi.org/10.1080/10635150601146041

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, **30**, 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Léonard, J. (1986). *Neotorularia* Hedge & J. Leonard nom generique nouveau de Cruciferae. *Bulletin du Jardin Botanique National de Belgique*, **56**, 389–395. https://doi.org/10.2307/3668200

Léonard, J. (1988). Contribution à l'étude de la flore et de la végtéation des déserts d'Iran: Etude des aires de distribution les phytochories, les chorotype. Jardin Botanique National de Belgique, Meise, Belgium, 767 pp.

Linder, H. P. (2014). The evolution of African plant diversity. *Frontiers in Ecology and Evolution*, **2**, 38. https://doi.org/10.3389/fevo.2014.00038

Linder, H. P., de Klerk, H. M., Born, J., Burgess, N. D., Fjeldså, J. & Rahbek, C. (2012). The partitioning of Africa: statistically defined biogeographical regions in sub−Saharan Africa. *Journal of Biogeography*, **39**, 1189–1205. https://doi.org/10.1111/j.1365-2699.2012.02728.x

Liu, L-M, Du, X-Y, Guo, C & Li, D-Z. (2020). Resolving robust phylogenetic relationships of core Brassicaceae using genome skimming data. *Journal of Systematics and Evolution*, xx, xxx–xxx. https://doi.org/10.1111/jse.12666

Liu, L., Zhao, B., Tan, D. & Wang, J. (2012). Phylogenetic relationships of Brassicaceae species based on matK sequences. *Pakistan Journal of Botany*, **38**, 731–739. https://doi.org/10.1016/j.bse.2010.06.003

Lysak, M. A., Cheung, K., Kitschke, M. & Bureš, P. (2007). Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiology*, **145**, 402–410. https://doi.org/10.1104/pp.107.104380

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. Fischer, Stuttgart, Germany, 691 pp.

Malyschev, L. I. & Peschkova, G. A. (2004). Flora of Siberia Volume VII: Berberidaceae–Grossulariaceae, Science Publishers, Enfield, United States of America, 318 pp.

Manafzadeh, S., Salvo, G. & Conti, E. (2014). A tale of migrations from east to west: the Irano-Turanian floristic region as a source of Mediterranean xerophytes. *Journal of Biogeography*, **41**, 366–379. https://doi.org/10.1111/jbi.12185

Manafzadeh, S., Staedler, Y. M. & Conti, E. (2017). Visions of the past and dreams of the future in the Orient: the Irano-Turanian region from classical botany to evolutionary studies. *Biological Reviews of the Cambridge Philosophical Society*, **92**, 1365–1388. https://doi.org/10.1111/brv.12287

Mandáková, T., Li, Z., Barker, M. S. & Lysak, M. A. (2017). Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *The Plant Journal*, **91**, 3–21. https://doi.org/10.1111/tpj.13553

Marais, W. (1970). Cruciferae. In: Codd, L. E., De Winter, B., Killick, D. J. B. & Rycroft, H. B. (Eds.). Flora of Southern Africa. Vol. 13. National Botanic Gardens, Kirstenbosch, South Africa, 118 pp.

Matzke, N. J. (2013). BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. University of California, Berkeley, Berkeley, CA

Matzke, N. J. (2013). Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers in Biogeography*, **5**, 4. https://doi.org/10.21425/F5FBG19694

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, **63**, 951–970. https://doi.org/10.1093/sysbio/syu056

Médail, F. & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, **36**, 1333–1345. https://doi.org/10.1111/j.1365-2699.2008.02051.x

Messager, E., Belmecheri, S., Von Grafenstein, U., Nomade, S., Ollivier, V., Voinchet, P., Puaud, S., Courtin-Nomade, A., Guillou, H., Mgeladze, A., Dumoulin, J.-P., Mazuy, A. & Lordkipanidze, D. (2013). Late Quaternary record of the vegetation and catchment-related changes from Lake Paravani (Javakheti, South Caucasus). *Quaternary Science Reviews*, **77**, 125–140. https://doi.org/10.1016/j.quascirev.2013.07.011

Meusel, H., Jäger, E. & Weinert, E. (1965). Vergleichende Chorologie der Zentraleuropäischen Flora. Gustav Fischer Verlag, Jena, Germany. 583 pp.

Milanovsky, E. E. (2008). Origin and development of ideas on Pliocene and Quaternary glaciations in northern and eastern Europe, Iceland, Caucasus and Siberia. *Geological Society, London, Special Publications*, **301**, 87–115. https://doi.org/10.1144/SP301.6

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). Presented at the 2010 Gateway Computing Environments Workshop (GCE), pp 1–8. https://doi.org/10.1109/GCE.2010.5676129

Mirarab, S., Reaz, R., Bayzid, Md. S., Zimmermann, T., Swenson M. S. & Warnow T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, **30**, i541–i548. https://doi.org/10.1093/bioinformatics/btu462

Mitchell, J. & Westaway, R. (1999). Chronology of Neogene and Quaternary uplift and magmatism in the Caucasus: constraints from K–Ar dating of volcanism in Armenia. *Tectonophysics*, **304**, 157–186. https://doi.org/10.1016/S0040-1951(99)00027-X

Mito, T. & Uesugi, T. (2004). Invasive alien species in Japan: the status quo and the new regulation for prevention of their adverse effects. *Global Environmental Research*, **8**, 171–191.

Mittermeier, R. A., Gil, P. G., Hoffman, M. T., Pilgrim, J. D., Brooks, T. M., Mittermeier, C. G., Lamoreux, J. F. & da Fonseca, G. A. B. (2000). Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions. University of Chicago Press, Conservation International, Chicago, United States of America, 392 pp.

Molins, A., Bacchetta, G., Rosato, M., Rosselló, J. A. & Mayol, M. (2011). Molecular phylogeography of *Thymus herba-barona* (Lamiaceae): insight into the evolutionary history of the flora of the western Mediterranean islands. *Taxon*, **60**, 1295–1305. https://doi.org/10.1002/tax.605006

Moore, T. E., Verboom, G. A. & Forest, F. (2010). Phylogenetics and biogeography of the parasitic genus *Thesium* L. (Santalaceae), with an emphasis on the Cape of South Africa. *Botanical Journal of the Linnean Society*, **162**, 435–452. https://doi.org/10.1111/j.1095-8339.2010.01032.x

Mutlu, B. & Karakuş, Ş. (2015). A new species of *Sisymbrium* (Brassicaceae) from Turkey: morphological and molecular evidence. *Turkish Journal of Botany*, **39**, 325–333. https://doi.org/10.3906/bot-1404-28

Naidina, O. D. & Richards, K. (2016). Pollen evidence for Late Pliocene – Early Pleistocene vegetation and climate change in the North Caucasus, North-Western Caspian Region. *Quaternary International, From the Caspian to Mediterranean: Environmental Change and Human Response during the Quaternary: IGCP 610*, **409**, 50–60. https://doi.org/10.1016/j.quaint.2015.12.018

Nakhutsrisvili, G., Abdaladze, O., Batsatsashvili, K., Spehn, E., Körner, C. (2017). Plant Diversity in the Central Great Caucasus: A Quantitative Assessment, Geobotany Studies. Springer International Publishing, Cham, Switzerland, 177 pp.

Nikolov, L. A., Shushkov, P., Nevado, B., Gan, X., Al−Shehbaz, I. A., Filatov, D., Bailey, C. D. & Tsiantis, M. (2019). Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytologist*, **222**, 1638–1651. https://doi.org/10.1111/nph.15732

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A & Minh B. Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274. https://doi.org/10.1093/molbev/msu300.

Noroozi, J., Akhani, H. & Breckle, S.-W. (2008). Biodiversity and phytogeography of the alpine flora of Iran. *Biodiversity and Conservation*, **17**, 493–521. https://doi.org/10.1007/s10531-007-9246-7

Noroozi, J., Talebi, A., Doostmohammadi, M., Manafzadeh, S., Asgarpour, Z. & Schneeweiss, G. M. (2019). Endemic diversity and distribution of the Iranian vascular flora across phytogeographical regions, biodiversity hotspots and areas of endemism. *Scientific Reports*, **9**, 1–12. https://doi.org/10.1038/s41598-019-49417-1

Oprea, A. & Sîrbu, C. (2010). *Sisymbrium volgense* (Brassicaceae in the flora of Romania.*The Annals of Oradea University Biology Fascicle*, **17**, 296–299.

Ortiz, M. Á., Tremetsberger, K., Terrab, A., Stuessy, T. F., García−Castaño, J. L., Urtubey, E., Baeza, C. M., Ruas, C. F., Gibbs, P. E. & Talavera, S. (2008). Phylogeography of the invasive weed *Hypochaeris radicata* (Asteraceae): from Moroccan origin to worldwide introduced populations. *Molecular Ecology*, **17**, 3654–3667. https://doi.org/10.1111/j.1365-294X.2008.03835.x

Partridge, R. R. & Maud, T. C. (1987). Geomorphic evolution of southern Africa since the Mesozoic. *South African Journal of Geology*, **90**, 179–208.

Payson, E. B. (1922). Species of *Sisymbrium* native to America north of Mexico. *University of Wyoming Publications in Science*, **1**, 1–27.

Pérez−Collazos, E., Sɑnchez−Gómez, P., Jiménez, J. F. & Catalán, P. (2009). The phylogeographical history of the Iberian steppe plant *Ferula loscosii* (Apiaceae): a test of the abundant-centre hypothesis. *Molecular Ecology*, **18**, 848–861. https://doi.org/10.1111/j.1365-294X.2008.04060.x

Pirie, M. D., Vargas, M. P. B., Botermans, M., Bakker, F. T. & Chatrou, L. W. (2007). Ancient paralogy in the cpDNA trnL-F Region in Annonaceae: implications for plant molecular systematics. *American Journal of Botany*, **94**, 1003–1016. https://doi.org/10.3732/ajb.94.6.1003

Plenk, K., Bardy, K., Höhn, M, Thiv, M. & Kropf, M. (2017). No obvious genetic erosion, but evident relict status at the westernmost range edge of the Pontic−Pannonian steppe plant *Linum flavum* L. (Linaceae) in Central Europe. *Ecology and Evolution*, **7**, 6527–6539. https://doi.org/10.1002/ece3.2990

Podlech, D. & Dieterle, A. (1969). Chromosomenstudien an Afghanischen Pflanzen. *Candollea*, **24**, 185–243.

Post, G. E. (1896). Flora of Syria, Palestine and Sinai; from the Taurus to Ras Muhammad, and from the Mediterranean Sea to the Syrian desert. Nabu Press, Charleston. South Carolina, United States of America, 932 pp.

Pound, M. J., Haywood, A. M., Salzmann, U. & Riding, J. B. (2012). Global vegetation dynamics and latitudinal temperature gradients during the Mid to Late Miocene (15.97–5.33Ma). *Earth-Science Reviews*, **112**, 1–22. https://doi.org/10.1016/j.earscirev.2012.02.005

Protopopova, V. V., Shevera, M. V. & Mosyakin, S. L. (2006). Deliberate and unintentional introduction of invasive weeds: a case study of the alien flora of Ukraine. *Euphytica*, **148**, 17–33. https://doi.org/10.1007/s10681-006-5938-4

Pujadas Salvá, A. (1993). *Sisymbrium* L. In: Castroviejo, S., Aedo, C., Gómez Campo, C., Laínz, M., Montserrat, P., Morales R., Muñoz Garmendia, F., Nieto Feliner, G., Rico, E., Talavera, S. & Villar, L. (Eds.). Flora Iberica, Vol. 4. Real Jardín Botánico, C.S.I.C., Madrid, Spain, pp 12–27.

Pyankov, V., Black, C., Stichler, W. & Ziegler, H. (2002). Photosynthesis in *Salsola* species (Chenopodiaceae) from Southern Africa relative to their C4 syndrome origin and their African-Asian arid zone migration pathways. *Plant Biology*, **4**, 62–69. https://doi.org/10.1055/s-2002-20437

QGIS.org (2020). QGIS Geographic Information System. QGIS Association. retrieved from http://www.qgis.org

Quezel, P. (1985). Definition of the Mediterranean Region and the Origin of its Flora. Gomez-Campo, C. (Ed.). Plant Conservation in the Mediterranean Area. Junk, Dordrecht, Netherlands, pp 9–24.

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

Renner, S. S. (2016). Available data point to a 4-km-high Tibetan Plateau by 40 Ma, but 100 molecular-clock papers have linked supposed recent uplift to young node ages. *Journal of Biogeography*, **43**, 1479–1487. https://doi.org/10.1111/jbi.12755

Ritchie, A. M., Lo, N. & Ho, S. Y. W. (2017). The impact of the tree prior on molecular dating of data sets containing a mixture of inter- and intraspecies sampling. *Systematic Biology*, **66**, 413–425. https://doi.org/10.1093/sysbio/syw095

Rollins, R. C. (1943). Generic revisions in the Cruciferae: *Halimolobos*. *Contributions from Dudley Herbarium*, **3**, 241–265.

Rollins, R. C. (1982). *Thelypodiopsis* and *Schoenocrambe* (Cruciferae). *Contributions from the Gray Herbarium of Harvard University*, **212**, 71–102.

Romanczuk, M. C. (1981). Novedades en el género *Sisymbrium* (Cruciferae) y sinopsis de las especies Patagonicas. *Parodiana*, **1**, 1–29.

Romanczuk, M. C. (1982). El género *Sisymbrium* (Cruciferae) en la Argentina. *Darwiniana*, **24**, 75–156. https://www.jstor.org/stable/23216515

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542. https://doi.org/10.1093/sysbio/sys029

Sales, F. & Hedge, I. C. (2013). Generic endemism in South-West Asia: an overview. *Rostaniha*, **14**, 22–35. https://doi.org/10.22092/botany.2013.101306

Sarver, B. A. J., Pennell, M. W., Brown, J. W., Keeble, S., Hardwick, K. M., Sullivan, J. & Harmon, L. J. (2019). The choice of tree prior and molecular clock does not substantially affect phylogenetic inferences of diversification rates. *PeerJ*, **7**, e6334. https://doi.org/10.7717/peerj.6334

Schatz, G. E., Shulkina, T., Nakhutsrishvili, G., Batsatsashvili, K., Tamanyan, K., Ali-zade, V., Kikodze, D., Geltman, D. & Ekim, T. (2009). Development of Plant Red List Assessments for the Caucasus Biodiversity Hotspot. In: Zazanashvili, N. & Mallon, D. (Eds.). Status and Protection of Globally Threatened Species in the Caucasus. CEPF, WWF. Contour Ltd., Tbilisi, Georgia, pp 188–192.

Scheunert, A. & Heubl, G. (2014). Diversification of *Scrophularia* (Scrophulariaceae) in the Western Mediterranean and Macaronesia – phylogenetic relationships, reticulate evolution and biogeographic patterns. *Molecular Phylogenetics and Evolution*, **70**, 296–313. https://doi.org/10.1016/j.ympev.2013.09.023

Schranz, M. E., Mohammadin, S. & Edger, P. P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Current Opinion in Plant Biology*, **15**, 147–153. https://doi.org/10.1016/j.pbi.2012.03.011

Schulz, O. E. (1924). Cruciferae–Sisymbrieae. In Engler, A. (Ed.). Pflanzenreich 86 (IV. 105). Verlag von Wilhelm Engelmann, Leipzig. Germany, 388 pp.

Schulz, O. E. (1928). Cruciferae: *Werdermannia*. *Notizblatt des Königl. botanischen Gartens und Museums zu Berlin*, **10**, 470.

Schwery, O., Onstein, R. E., Bouchenak−Khelladi, Y., Xing, Y., Carter, R. J. & Linder, H. P. (2015). As old as the mountains: the radiations of the Ericaceae. *New Phytologist*, **207**, 355–367. https://doi.org/10.1111/nph.13234

Skrede, I., Eidesen, P. B., Portela, R. P. & Brochmann, C. (2006). Refugia, differentiation and postglacial migration in arctic-alpine Eurasia, exemplified by the mountain avens (*Dryas octopetala* L.). *Molecular Ecology*, **15**, 1827–1840. https://doi.org/10.1111/j.1365-294X.2006.02908.x

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stewart, J. R., Lister, A.M., Barnes, I. & Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 661–671. https://doi.org/10.1098/rspb.2009.1272

Stockenhuber, R., Zoller, S., Shimizu-Inatsugi, R., Gugerli, F., Shimizu, K. K., Widmer, A. & Fischer M. C. (2015). Efficient detection of novel Nuclear markers for Brassicaceae by transcriptome sequencing. *PLOS One*, **10**, e012818. https://doi.org/10.1371/journal.pone.0128181

Suc, J.-P. (1984). Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature*, **307**, 429–432. https://doi.org/10.1038/307429a0

Suc, J.-P., Popescu, S.-M., Fauquette, S., Bessedik, M., Jiménez-Moreno, G., Taoufiq, N. B., Zheng, Z., Medail, F. & Klotz, S. (2018). Reconstruction of Mediterranean flora, vegetation and climate for the last 23 million years based on an extensive pollen dataset. *Ecologia Mediterranea*, **44**, 53–85. https://doi.org/10.3406/ecmed.2018.2044

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J. & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, **4**, vey016. https://doi.org/10.1093/ve/vey016

Svenning, J.-C., Normand, S. & Kageyama, M. (2008). Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology*, **96**, 1117–1127. https://doi.org/10.1111/j.1365-2745.2008.01422.x

Taberlet, P., Fumagalli, L., Wust−Saucy, A.-G. & Cosson, J.-F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464. https://doi.org/10.1046/j.1365-294x.1998.00289.x

Tagieva, E. N., Aleskerov, B. D., Veliev, S. S. & Kuliev, M. Yu. (2013). Transformation of vegetation cover on the territory of Azerbaijan during the Pleistocene. *Geography and Natural Resources*, **34**, 395–400. https://doi.org/10.1134/S187537281304015X

Takhtajan, A. L., Crovello, T. J. & Cronquist, A. (1986). Floristic Regions of the World. University of California Press, Berkeley. United States of America, 522 pp.

Thompson, J. D. (2005). Plant Evolution in the Mediterranean. Oxford University Press, Oxford. United Kingdom, 302 pp.

Tremetsberger, K., Ortiz, M. Á., Terrab, A., Balao, F., Casimiro-Soriguer, R., Talavera, M. & Talavera, S. (2016). Phylogeography above the species level for perennial species in a composite genus. *AoB PLANTS*, **8**, plv142. https://doi.org/10.1093/aobpla/plv142

Tutin, T. G., Heywood, V. H., Burges, N. A., Valentine, D. H., Walters, S. M. & Webb, D. A. (1993). Flora Europaea, vol. 1: Lycopodiaceae to Platanaceae. Cambridge University Press, Cambridge, United Kingdom, pp 1–581.

Valente, L. M., Savolainen, V., Manning, J. C., Goldblatt, P. & Vargas, P. (2011). Explaining disparities in species richness between Mediterranean floristic regions: a case study in *Gladiolus* (Iridaceae). *Global Ecology and Biogeography*, **20**, 881–892. https://doi.org/10.1111/j.1466-8238.2010.00644.x

Vassilczenko I. T. (1939). *Sisymbrium* L. In: Komarov V. L. (ed.). Flora SSSR [Flora of USSR] 8. Moscow & Leningrad: Publishers of Academy of Sciences of USSR, pp 38–55.

van Dam, J. A. (2006). Geographic and temporal patterns in the late Neogene (12–3 Ma) aridification of Europe: The use of small mammals as paleoprecipitation proxies. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **238**, 190–218. https://doi.org/10.1016/j.palaeo.2006.03.025

van Zinderen Bakker, E. M. (1978). Quaternary Vegetation Changes in Southern Africa. In: Werger, M. J. A. (Ed.). Biogeography and Ecology of Southern Africa, Monographiae Biologicae. Springer Netherlands, Dordrecht, Netherlands, pp 131–143.

Velichko, A. A. (2005). Chapter 10: General Features of the Landscape and Climate of Northern Eurasia in the Cenozoic. In: Velichko, A. A. & Wright, H. E. (Eds.). Cenozoic Climatic and Environmental Changes in Russia. Geological Society of America, Boulder, Colorado, The United States of America, pp 189–200.

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

Warwick, S. I. & Al-Shehbaz, I. A. (2003). Nomenclatural notes on *Sisymbrium* (Brassicaceae). *Novon*, **13**, 265–267. https://doi.org/10.2307/3393529

Warwick, S. I. & Al-Shehbaz, I. A. (2003). Nomenclatural notes on *Sisymbrium* (Brassicaceae). *Novon*, **13**, 265–267. https://doi.org/10.2307/3393529

Warwick, S. I., Al-Shehbaz, I. A., Price, R. A. & Sauder, C. (2002). Phylogeny of *Sisymbrium* (Brassicaceae) based on ITS sequences of nuclear ribosomal DNA. *Canadian Journal of Botany*, **80**, 1002–1017. https://doi.org/10.1139/b02-089

Warwick, S. I., Al-Shehbaz, I. A., Sauder, C. A., Harris, J. G. & Koch M. A. (2004). Phylogeny of *Braya* and *Neotorularia* (Brassicaceae) based on nuclear ribosomal internal transcribed spacer and chloroplast trnL intron sequences. *Canadian Journal of Botany*, **82**, 376–392. https://doi.org/10.1139/b04-012

Warwick, S. I., Mummenhoff, K., Sauder, C. A., Koch, M. A. & Al-Shehbaz, I. A. (2010). Closing the gaps: phylogenetic relationships in the Brassicaceae based on DNA sequence data of nuclear ribosomal ITS region. *Plant Systematics and Evolution*, **285**, 209–232. https://doi.org/10.1007/s00606-010-0271-8

Warwick, S. I., Sauder, C. A. & Al-Shehbaz, I. A. (2006). Molecular Phylogeny, Morphology and Cytological Diversity of *Sisymbrium* (Brassicaceae). In: Sharma, A. K. & Sharma, A. (Eds.). Plant Genome: Biodiversity and Evolution. Phanerogams (Angiosperm – Dicotyledons). Oxford & IBH Publishing Co. Pvt. Ltd., New Delhi with Science Publishers, United States of America, pp 219–250.

Weber, E., Sun, S.-G. & Li, B. (2008). Invasive alien plants in China: diversity and ecological insights. *Biological Invasions*, **10**, 1411–1429. https://doi.org/10.1007/s10530-008-9216-3

Weiss, S. & Ferrand, N. (2007). Phylogeography of Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity. Springer Netherlands, Dordrecht, 377 pp.

Wesche, K., Ambarlı, D., Kamp, J., Török, P., Trieber, J. & Dengler, J. (2016). The Palaearctic steppe biome: a new synthesis. *Biodiversity and Conservation*, **25**, 2197–2231. https://doi.org/10.1007/s10531-016-1214-7

Yu, Y., Blair, C. & He, X. J. (2020). RASP 4: Ancestral State Reconstruction Tool for Multiple Genes and Characters. *Molecular Biology and Evolution*, **37**, 604–606. https://doi.org/10.1093/molbev/msz257

Yu, Y., Harris, A. J., Blair, C. & He, X. J. (2015). RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution*, **87**, 46–49. https://doi.org/10.1016/j.ympev.2015.03.008

Zachos, J. C., Dickens, G. R. & Zeebe, R. E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, **451**, 279–283. https://doi.org/10.1038/nature06588

Zhang, C., Sayyari, E. & Mirarab S. (2017). ASTRAL-III: Increased Scalability and Impacts of Contracting Low Support Branches. In: Meidanis, J. & Nakhleh L. (Eds.). Comparative Genomics. RECOMB-CG 2017. Lecture Notes in Computer Science, vol 10562. Springer, Cham. https://doi.org/10.1007/978-3-319-67979-2_4

Zhou, T.-Y., Lu, L.-L., Yang, G. & Al-Shehbaz, I.A. (2001) Brassicaceae (Cruciferae). In: Wu, Z.-Y. & Raven, P. H. (Eds.). Flora of China 8 (Brassicaceae through Saxifragaceae). Science Press, Beijing; Missouri Botanical Garden Press, St. Louis, pp 1–193.

Züst, T., Strickler, S. R., Powell, A. F., Mabry, M. E., An, H., Mirzaei, M., York, T., Holland C. K., Kumar, P., Erb, M., Petschenka, G., Gómez, J.-M., Perfectti, F., Müller, C., Pires, J. C., Mueller, L. A. & Jander, G. (2020). Independent evolution of ancestral and novel defenses in a genus of toxic plants (*Erysimum*, Brassicaceae). *eLife* **9**, e51712. https://doi.org/10.7554/eLife.51712

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

Table 1a: Statistics of the crown ages of tribes depending on the calibration method based on the dataset of Lineage II. Time span is indicated by the lower boundary, median age and the upper boundary. Yule, Yule Tree Prior; Coal, Coalescent Tree Prior; BiDe, Birth-Death Tree Prior.

Calibration method	Tree Prior	Crown age of Brassiceae	Crown age of Isatideae	Crown age of Sisymbrieae	Crown age of Thelypodieae
Secondary calibration (Huang et al., 2020)	**BiDe**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**6.87–10.57–13.03 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD
Secondary calibration (Huang et al., 2020) + ITS substitution rate (Kay et al., 2006)	**BiDe**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**10.93–16.30–18.54 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD
ITS substitution rate (Kay et al., 2006)	**BiDe**	**27.23–32.94–38.99 MYA 95% HPD**	**7.95–11.20–15.07 MYA 95% HPD**	**12.46–16.62–21.08 MYA 95% HPD**	**10.07–18.48–28.25 MYA 95% HPD**
Secondary calibration (Huang et al., 2020)	**Coal**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**5.68–9.52–12.86 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD
Secondary calibration (Huang et al., 2020) + ITS substitution rate (Kay et al., 2006)	**Coal**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**11.50–17.66–25.15 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD
ITS substitution rate (Kay et al., 2006)	**Coal**	**37.69–48.36–61.05 MYA 95% HPD**	**7.53–12.55–17.35 MYA 95% HPD**	**13.91–21.15–28.60 MYA 95% HPD**	**9.48–28.70–48.00 MYA 95% HPD**
Secondary calibration (Huang et al., 2020)	**Yule**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**7.39–10.61–13.1 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD
Secondary calibration (Huang et al., 2020) + ITS substitution rate (Kay et al., 2006)	**Yule**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**11.39–15.96–18.87 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD
ITS substitution rate (Kay et al., 2006)	**Yule**	**26.4–33.01–36.86 MYA 95% HPD**	**7.95–12.66–15.3 MYA 95% HPD**	**12.64–17.92–20.90 MYA 95% HPD**	**9.65–18.52–26.51 MYA 95% HPD**

Table 1b: Statistics of the crown ages of tribes depending on the calibration method based on the dataset of Lineage II with two putative outgroups. Time span is indicated by the lower boundary, median age and the upper boundary. Yule, Yule Tree Prior; Coal, Coalescent Tree Prior; BiDe, Birth-Death Tree Prior.

Calibration method	Tree Prior	Crown age of Brassiceae	Crown age of Isatideae	Crown age of Sisybrieae	Crown age of Thelypodieae	Crown age of Eutremeae	Crown age of Thlaspideae
Secondary calibration (Huang et al., 2020)	**BiDe**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**8.09–11.62–14.00 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD	**Fixed at** 6.9–15.1 MYA 95% HPD	**Fixed at** 9.5–17.1 MYA 95% HPD
Secondary calibration (Huang et al., 2020) + ITS substitution rate (Kay et al., 2006)	**BiDe**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**11.59–15.92–19.39 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD	**Fixed at** 6.9–15.1 MYA 95% HPD	**Fixed at** 9.5–17.1 MYA 95% HPD
ITS substitution rate (Kay et al., 2006)	**BiDe**	**26.09–32.16–35.90 MYA 95% HPD**	**7.77–13.04–16.37 MYA 95% HPD**	**13.14–18.70–22.34 MYA 95% HPD**	**9.33–18.27–29.94 MYA 95% HPD**	**11.44–16.95–20.7 MYA 95% HPD**	**13.35–18.48–21.52 MYA 95% HPD**
Secondary calibration (Huang et al., 2020)	**Coal**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**6.8–10.95–14.49 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD	**Fixed at** 6.9–15.1 MYA 95% HPD	**Fixed at** 9.5–17.1 MYA 95% HPD
Secondary calibration (Huang et al., 2020) + ITS substitution rate (Kay et al., 2006)	**Coal**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**12.23–19.56–24.81 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD	**Fixed at** 6.9–15.1 MYA 95% HPD	**Fixed at** 9.5–17.1 MYA 95% HPD
ITS substitution rate (Kay et al., 2006)	**Coal**	**35.23–44.74–53.28 MYA 95% HPD**	**7.77–13.58–18.77 MYA 95% HPD**	**14.29–22.2–29.90 MYA 95% HPD**	**11.53–38.52–49.07 MYA 95% HPD**	**11.77–19.75–27.07 MYA 95% HPD**	**16.41–23.27–28.9 MYA 95% HPD**
Secondary calibration (Huang et al., 2020)	**Yule**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**8.14–11.85–19.00 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD	**Fixed at** 6.9–15.1 MYA 95% HPD	**Fixed at** 9.5–17.1 MYA 95% HPD
Secondary calibration (Huang et al., 2020) + ITS substitution rate (Kay et al., 2006)	**Yule**	**Fixed at** 14.3–19.8 MYA 95% HPD	**Fixed at** 5.3–11.2 MYA 95% HPD	**11.76–15.70–19.61 MYA 95% HPD**	**Fixed at** 8.8–17.5 MYA 95% HPD	**Fixed at** 6.9–15.1 MYA 95% HPD	**Fixed at** 9.5–17.1 MYA 95% HPD
ITS substitution rate (Kay et al., 2006)	**Yule**	**25.23–30.74–34.95 MYA 95% HPD**	**8.07–12.32–15.99 MYA 95% HPD**	**12.71–17.81–21.68 MYA 95% HPD**	**8.93–16.50–28.26 MYA 95% HPD**	**11.15–16.16–20.25 MYA 95% HPD**	**13.35–17.66–21.04 MYA 95% HPD**

Table 2: Statistics of the crown and stem ages of geographically defined clades depending on the calibration method based on the dataset of Sisymbrieae. Time span is indicated by the lower boundary, median age and the upper boundary. Yule, Yule Tree Prior; Coal, Coalescent Tree Prior; BiDe, Birth-Death Tree Prior.

Geographic entity	Age	Coal	Yule	BiDe
disjunct forest Eurasian clade (A)	**stem**	1.76–3.50–5.64 MYA 95% HPD	2.44–4.40–6.73 MYA 95% HPD	2.23–4.20–6.56 MYA 95% HPD
	crown	0.76–1.63–2.79 MYA 95% HPD	1.10–2.19–3.60 MYA 95% HPD	0.97–2.03–3.43 MYA 95% HPD
South-African clade (B)	**stem**	1.41–3.03–4.96 MYA 95% HPD	2.11–3.85–6.03 MYA 95% HPD	1.92–3.66–5.89 MYA 95% HPD
	crown	0.32–0.88–1.69 MYA 95% HPD	0.47–1.19–2.22 MYA 95% HPD	0.43–1.10–2.11 MYA 95% HPD
Caucasian clade (C)	**stem**	3.6–6.65–10.34 MYA 95% HPD	4.69–8.06–12.18 MYA 95% HPD	4.29–7.72–11.81 MYA 95% HPD
	crown	0.4–1.06–2.08 MYA 95% HPD	0.62–1.43–2.7 MYA 95% HPD	0.54–1.33–2.51 MYA 95% HPD
Mediterranean clade (D)	**stem**	4.97–8.68–13.06 MYA 95% HPD	6.1–10.07–14.76 MYA 95% HPD	5.68–9.67–14.44 MYA 95% HPD
	crown	4.00–7.41–11.5 MYA 95% HPD	5.1–8.86–13.36 MYA 95% HPD	4.72–8.48–13.18 MYA 95% HPD

PUBLICATION IV:

EVOLUTIONARY HISTORY AND HISTORICAL BIOGEOGRAPHY OF THE MODEL GENUS *CAPSELLA* (BRASSICACEAE) WITH EMPHASIS ON THE STEPPE FLORAL ELEMENT *CAPSELLA ORIENTALIS* KLOKOV.

ŽERDONER ČALASAN A, HURKA H, GERMAN DA, PFANZELT S, BLATTNER F, SEIDL A, NEUFFER B.

(PREPRINT)

Evolutionary history and historical biogeography of the model genus *Capsella* (Brassicaceae) with emphasis on the steppe floral element *Capsella orientalis* Klokov.

Anže Žerdoner Čalasan[a,*], Herbert Hurka[a], Dmitry A German[b], Simon Pfanzelt[c,d], Frank Blattner[d], Anna Seidl[e], Barbara Neuffer[a]

[a] Department 5: Biology/Chemistry, Botany, University of Osnabruck, Barbarastraße 11, 49076 Osnabruck, Lower Saxony, Germany

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

[c] Experimental Taxonomy, Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstraße 3, 06466 Seeland, Saxony-Anhalt, Germany

[d] Munich Botanical Garden, Menzinger Straße 67, 80638 München, Bavaria, Germany

[e] Institute of Botany, Department of Integrative Biology and Biodiversity Research, University of Natural Resources and Life Sciences, Vienna (BOKU), Gregor-Mendel-Straße 33, 1180 Vienna, Austria

* Corresponding author. *E-mail address*: azerdonercal@uos.de

Abstract

Capsella is a small model plant genus of the Brassicaceae family, closely related to *Arabidopsis*. Young age coupled with high hybridisation potential and extensive morphological variability have hindered a straight-forward species delimitation in *Capsella*. Unsurprisingly, the evolutionary history of the genus has been thus controversial for a long time. A restriction digest-based NGS method of genotyping-by-sequencing of 282 individuals was carried out, including all the currently recognised *Capsella* species covering large parts of their distribution areas. We performed time estimation analysis and ecological niche modelling to place the evolutionary history of Eurasian steppe floral element *Capsella orientalis* into time and space. Our analysis uncovered that the eastern lineage out of which *Capsella orientalis* evolved was involved in the hybridisation event that gave rise to a tetraploid *C. bursa-pastoris*. Furthermore, extensive genetic variation within *Capsella orientalis* postdated any significant glacial events and surprisingly, this species found its refuge during the Last Glacial Maximum not in Altai as many other species, but around the north coast of the Black Sea in the west and in the South-East Kazakhstan in Tian Shan Mountains and the Tabagartai–Dzhungaria Alatau mountain chain. This idea is also supported by our population genetic study that uncovered the highest genetic diversity in the south Kazakhstan genetic cluster, suggesting that *C. orientalis* originated in the continental Asia and migrated northwards after the last ice age.

Key words: Eurasian steppe belt, genotype-by-sequencing (GBS), phylogeography, young lineage

1 Introduction

Capsella Medik. is a small genus within the mustard family (Brassicaceae). From all wild relatives of the model genus *Arabidopsis* Heynh., *Capsella* is the most closely related and intensively investigated genus. The latter two genera belong to the tribe Camelineae, which also includes the genus *Camelina* Crantz together with four more genera (Koch et al. 2018). All three named genera have become model genera for plant molecular research, for the study of genome evolution, and for the evolution and development of genotypic and phenotypic traits.

Species delimitation within the genus *Capsella* is difficult and controversial because of the enormous morphological variation, especially in leaf and fruit characters, and in hybridisation potential as intensively studied by Almquist and Shull in the early 1900ies (see Hurka and Neuffer, 1997 and references therein). In most cases, up to 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. (Later, Chater in Flora Europaea 1993 reduced *C. thracica* to subspecies rank of *C. bursa-pastoris*.) The first three species are diploid and the latter two are tetraploid. *C. grandiflora* is an obligate outcrossing species due to a sporophytic self-incompatibility system 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 hitherto only poorly known species *C. orientalis* and *C. thracica* in molecular phylogenetic analyses (Hurka et al. 2012), the evolutionary history of *Capsella* started to become clearer, and was confirmed by several other studies, e.g. Douglas et al. (2015), Bachmann et al. (2019), Žerdoner Čalasan et al. (2019). During the Pleistocene, *Capsella* split into two lineages, an eastern and a western lineage. The eastern lineage comprises the extant species *Capsella orientalis*, 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. The most recent common ancestor (MRCA) of both lineages is presumable 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).

Capsella orientalis was first described by Klokov in 1926 but for decades eked out a niche existence in botanical research. It was only in 2012 that the ploidy level of this species (diploid) was resolved and assessment of its distribution area was presented. *C. orientalis* turned out to be a

'missing link' in the understanding of the evolution of the model genus *Capsella,* shedding new light on the origin of the tetraploid *C. bursa-pastoris* and providing evidence that in *Capsella* the transition from outbreeding to selfing at the diploid level occurred two times independently (Hurka et al. 2012). Meanwhile, *Capsella* has become a model for genomic studies of plant mating systems and polyploidisation, e.g., Mattila et al. (2020); Bachmann et al. (2019); Woźniak et al. (2020) for mating shifts, and Douglas et al. (2015); Han et al. (2015); Kryvokhyzha et al. (2019a); Omelchenko et al. (2020) for polyploidisation. In contrast to the increasing knowledge about *Capsella* at the genomic level, a detailed historical biogeographical aspect of its evolution is currently missing.

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; May 4th 2011; by M. Knyazev).

Capsella orientalis is distributed in southeastern Europe, the South Urals, northern and eastern Kazakhstan, southwestern Siberia, northern part of the Chinese province of Xinjang 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 Asia. Continuous global cooling through the Middle and Upper Miocene accompanied with 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 the more dry-adapted steppe forbs and grasses. Spreading towards the west, Eurasian steppe belt reached the East European Plain in late Miocene (Velichko,

1999; Velichko et al. 2005) and expanded to Eastern Europe during the Pliocene epoch (Lang, 1994; Mai, 1995). The Eurasian steppe belt was under strong influence of Pleistocene glaciations that have caused extensive expansions and contractions as well as latitudinal range shifts and longitudinal range splits during the past 3 million years (Hurka et al. 2019 and references therein).

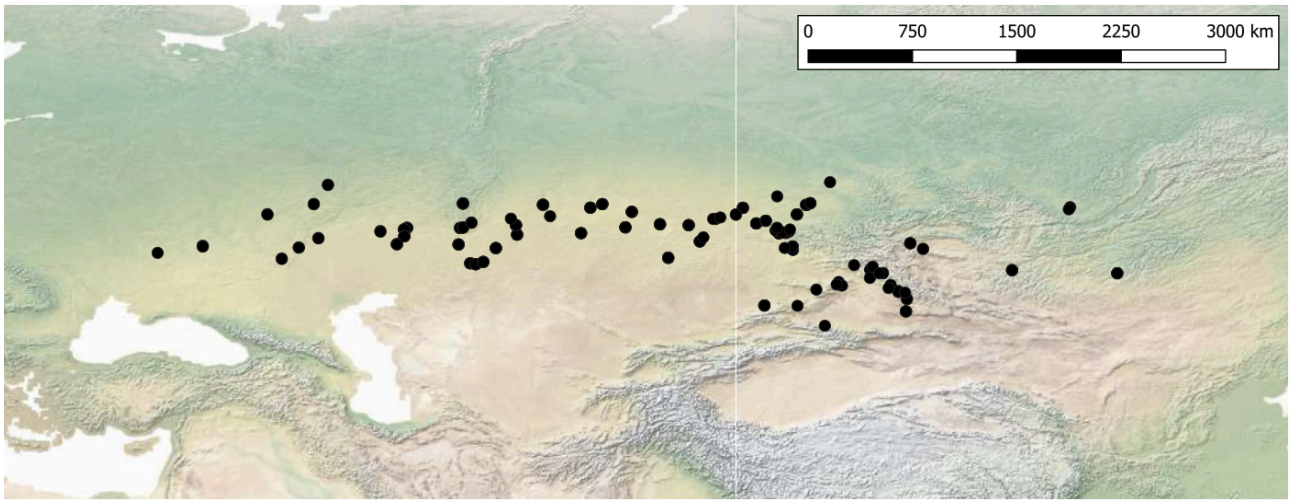


Figure 2: Distribution map of *Capsella orientalis* based on geographical range information from the literature (Bobrov et al., 1985; Ebel, 2002; Geltman et al., 2009; German et al., 2012; Jalas et al., 1994; Sheremetova et al., 2011), our own field data, vouchers from Hurka et al., 2012, Han et al., 2015 and Kryvokhyzha et al., 2019 as well as a priori critically assessed GBIF collection points.

It has been shown that such a biome dynamic is mirrored in molecular signals of typical Eurasian steppe plants and thus reflect 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). The aim of our study was to elucidate the short but turbulent history of the genus *Capsella* and the potential influence of the Pleistocene glaciations on its biological history. We carried out a restriction digest-based NGS method of genotyping-by-sequencing (GBS) of 282 individuals, including all the currently recognised *Capsella* species covering large parts of their distribution areas, and analysed the data using ipyRAD software (Eaton and Overcast, 2020). In accord with contemporary literature (Fernández-Mazuecos et al. 2018), we investigated the robustness of our dataset based on different assembly parameters and different taxon samples. We also assessed the levels of admixture and genetic diversity of *Capsella* and by carrying out secondarily calibrated time divergence analysis place its history into time and space. To infer the population structure of *C. orientalis*, additional F-statistics, AMOVA and molecular diversity indexes assessment were conducted.

2 Materials and Methods

Seeds of 5 to 10 randomly selected plants per population collected in the wild (Supplementary File 1) were cultivated in the greenhouse of the University of Osnabrueck (Figure 1). Plants were grown until the flowering stage and ploidy levels were estimated through flow cytometry to keep the species determination errors to a minimum. *Petroselinum crispum* served as an internal standard (Yokoya et al. 2000) for diploid as well as for polyploid species and extracted nuclei were stained with DAPI. 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 Poland et al. (2015) and Meyer et al. (2008). We performed a single-end sequencing of 100 bp fragments of 282 samples (57 populations) in two runs on a HiSeq 2500 (Illumina, California, 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 internally assess the reproducibility of the sequencing method, three samples were loaded twice on the flow cell (depicted with * in Supplementary File 1).

ipyrad: Already sorted raw Illumina DNA sequence data files were analysed using ipyrad v0.9.52 (Eaton and Overcast, 2020). Two separate sub-datasets were analysed, 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 *Capsella 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, recognising 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 max_alleles_consens option was set to 4, while in the *Capsella orientalis* dataset consisting of purely diploid accessions, this parameter was left at the default value of 2. After initial testing, we furthermore 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

analysed 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.

SVDQ: We performed the SVDQuartets analysis (Chifman et al. 2014) in PAUP* v5.0 (Swofford, 2002). Firstly, a phylogenetic analysis was carried out using all 282 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 more than 95% association to a specific cluster (i.e., exhibiting less than 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 a quartet assembly algorithm. Bootstrapping was performed with the number of replicates set to 1000. Final trees were exported with saved internal nodes and visualised with FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Time estimation: The 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 SNP dataset generated from ipyrad was used as an input file. Due to the high computational intensity, the unlinked SNP dataset was further reduced to 25 accessions, including all major taxon groups retrieved from the SVDQ analysis. The 25 accessions were selected in a way to present genetically most divergent accessions within selected species and/or genotypes. We provided a newly reconstructed starting tree from an SVDQ analysis, but without a specified outgroup 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), after an appropriate burn-in of 10%. 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 v.2.5.2 using '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).

STRUCTURE, LEA & DAPC: 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 10000 MCMC with a burning period of 2000. For *Capsella* dataset we have tested a range of K values from K=2 to K=10 and repeated the analysis 20 times. For *Capsella orientalis* dataset we tested a range of K values from K=2 to K=5 with the same repetition specified above. All settings were left on default with an 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 clump v.1.1.2. (Jakobsson and Rosenberg, 2007) by using Greedy algorithm and 1000 random input orders. K probability was assessed using Evanno method (Evanno et al. 2005) implemented in Structure Harvester (Earl and von Holdt, 2012). Population specific ancestry coefficients were calculated as a means over the individuals and shown in a form of pie charts plotted on the world map. In addition, population genetic structure was also inferred using LEA-clustering analysis implemented in an R package LEA (Frichot and 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 cross-validation approach in snmf function incorporated in the LEA package. Population specific ancestry coefficients were calculated and depicted as specified above. To confirm the result, an additional DAPC analyses for *Capsella* as well as *C. orientalis* dataset were carried out under default parameters using the adegenet R package (Jombart, 2008).

Ecological niche modelling: To infer the climatic suitability of habitats of *Capsella orientalis* through space and time, ecological niche modelling approach was employed, 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. 2015). The bioclimatic variables were used at the spatial resolution of 2.5 arc minutes (approximately 4.5 km^2 at the equator) and we estimated three different time slices: contemporary niche availability, Mid Holocene niche availability (8.33–4.2 ka), and Last Glacial Maximum niche availability (ca. 21 ka). The geographical distribution of *Capsella orientalis* was assessed based on geographical range information from the literature (Ebel, 2002; Geltman et al. 2009; German et al. 2012; Jalas et al.

1994; Sheremetova et al. 2011) and our own field data. To fill in the gaps, accessions from herbarium vouchers from Hurka et al. 2012, Han et al. 2015 and Kryvokhyzha et al. 2019a as well as a priori critically assessed GBIF collection points were added to the final dataset, consisting of 124 entries (Figure 2). Variable exclusion via cross-correlation was carried out to assure that the used climate variables were uncorrelated and the lowest correlation was inferred from the variables bio4, bio5 and bio12. To maximise a predictive ability and to avoid model overfitting, the number of climate variables was kept to a minimum and all the subsequent modelling reconstructions were evaluated with an R package ENMeval (Muscarella et al. 2014).

Arlequin: To determine the genetic diversity of *Capsella orientalis* populations, F-statistics, distribution of private alleles and analysis of molecular variance (AMOVA), assessing genetic differentiation on the cluster as well as population level, we utilised Arlequin v3.5 (Excoffier et al. 2010). Molecular diversity indexes and private allele sharing were assessed with linked SNP dataset retrieved directly from the ipyrad analysis, and the 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 1023 permutations. We categorised 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, as this was the lowest number of investigated populations per genetic cluster. Five geographically most distant populations were selected for genetic clusters, originally consisting of more than five populations. To allow for representativity the populations whose individuals were assigned to different genetic clusters were excluded from this analysis.

3 Results

ipyrad stats: 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*, only. Furthermore, the reads had to be at least 95% similar to be recognised 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, in which no population structure could have been inferred throughout different K-values (data not shown). The final *Capsella* dataset consisted of 12635 SNPs and 13.17% missing sites, while the final *Capsella orientalis* dataset included 4644 SNPs with 24.23% of missing data. The alignment used for time divergence estimation consisted of 25 entries, 16398 SNPs (out of which 6320 were retrieved in the unlinked SNP dataset) and 20.07% missing sites. Further statistics can be inferred from the Supplementary File 2.

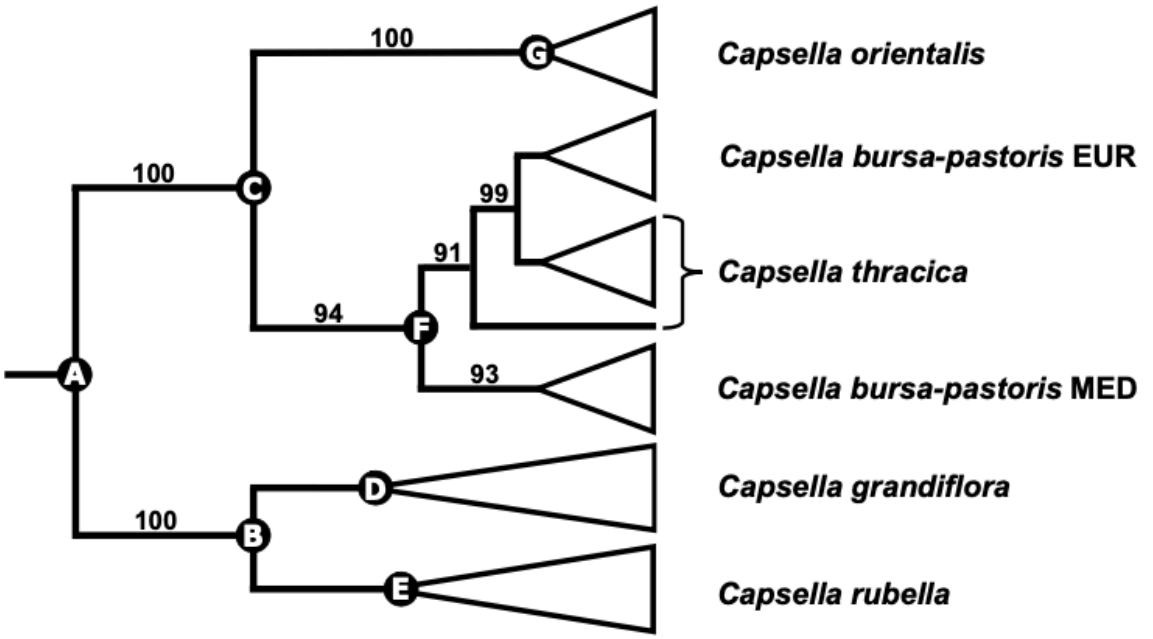
SVDQ: The SVDQ analysis of the non-reduced dataset retrieved only poorly supported species specific clades of *Capsella rubella* (57 BS) and *C. grandiflora* (37 BS) but well supported species clades of *C. orientalis* (100 BS) and *C. bursa-pastoris* clade (94 BS). Within the latter, Mediterranean *C. bursa-pastoris* was rendered monophyletic (93 BS), while the Eurasian *C. bursa-pastoris* and *C. thracica* were retrieved as poorly supported clade and grade, respectively (Supplementary File 3). *Capsella grandiflora* and *C. rubella* were manually retrieved as the out-group to the rest of *Capsella* species (100 BS), following the topologies of Douglas et al. (2015), Guo X et al. (2017), Hurka et al. (2012), Kryvokhyzha et al. (2019a) and Walden et al. (2020). The analysis of the reduced dataset (i.e., dataset with accessions exhibiting less than 5% admixture) including only 57 accessions retrieved the same topology but with a better bootstrap support within *Capsella orientalis*. The following clades were inferred: *Capsella rubella* (100 BS), *C. grandiflora* (71 BS), Mediterranean *C. bursa-pastoris* (100 BS), *C. thracica* (96 BS), Eurasian *C. bursa-pastoris* (100 BS), South Middle Asian *C. orientalis* (83 BS), East Middle Asian *C. orientalis* (100 BS), Mongolian *C. orientalis* (100 BS), North Middle Asian *C. orientalis* (73 BS) and Altai *C. orientalis* (98 BS). The only well-supported sister group relationship was inferred between Altai and North Middle Asian *C. orientalis* (98 BS). No robust contradicting topological differences in sister group relationships were inferred (Supplementary Figure 4). 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).

Time estimation: All four time divergence analyses led to congruent results, despite the single secondary calibration points originated from different unrelated studies with different taxon samples and genetic datasets. All analyses converged to the stationary distribution (ESS > 200) and the overall topologies corresponded to the other phylogenetic studies (Douglas et al. 2015; Guo X et al. 2017; Hurka et al. 2012; Kryvokhyzha et al. 2019a; Walden et al. 2020). The crown age of *Capsella* was estimated to approximately 0.75 Ma, the split of *Capsella 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, specifically into MIS15, MIS7 and MIS6, respectively (Figure 3). The crown age of *Capsella grandiflora* was dated to around 67 KYA and the crown age of *Capsella rubella* at 26 KYA, both placed into Late Pleistocene. The crown age a clade consisting of *Capsella bursa-pastoris* and *C. thracica* was estimated to the Middle Holocene at around 6.7 KYA and the diversification of *Capsella orientalis* occurred in the Late Holocene approximately 2.1 KYA (Figure 3).

Population genetic clustering according to STRUCTURE: The highest delta K value in *Capsella* dataset was inferred for K=6, while in *Capsella orientalis* dataset the highest delta K value was retrieved at K=3 (Supplementary Figure 5). *Capsella* dataset subsequently split in the following way. K=2 separated *Capsella orientalis* from the rest, with *C. bursa-pastoris* accessions exhibiting moderate introgression (Figure 4A). At K=3 and K=4, subsequent fragmentation within *Capsella orientalis* was observed, mirrored also in the LEA analysis in *Capsella* (see below and Supplementary File 6) as well as in *C. orientalis* datasets (Supplementary Files 6 and 7). At K=5, the Mediterranean diploid branch got separated and at K=6 the introgression of *Capsella grandiflora* was observed in the Mediterranean genetic cluster of *C. bursa-pastoris* (Figures 4D and 4E). Subsequent clustering delimited Mediterranean *Capsella grandiflora* from genetically uniform *C. rubella* (Figure 4F). Further clustering could not be inferred from Figure 4 because it resulted in new very minor genetic clusters embedded in *Capsella orientalis*. *Capsella orientalis* dataset clustered in the following manner. At K=2, northern and southern *Capsella orientalis* genetic clusters were inferred (Figure 5A). Souther cluster subsequently split up into an East Middle Asian and Mongolian cluster, and South Middle Asian cluster (at K=3) and the the division ran from the southern foothills of the Altai Mountains, southwards to the Tian Shan Mountains along the

Tabagartai–Dzhungaria Alatau mountain chain. At K=4, a new genetic cluster mixed with the Mongolian genetic cluster was inferred along the East Middle Asian Mountain Chain and at K=5 an Altai genetic cluster emerged (Figures 5C and 5D). Contrarily to LEA, STRUCTURE genetic clustering did not strictly follow the *Capsella* species boundaries.



	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 Supplementary File 3). The values above branches represent bootstrap support values. The node letters correspond to the letters in the table below. **Figure 3 below:** Time spans retrieved from SNAPP analysis 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*.

Population genetic clustering according to LEA: The lowest cross-entropy value in *Capsella* dataset was tentatively inferred for K=10 (albeit without a clear signal), while no optimal K value was inferred for *Capsella orientalis* dataset (Supplementary File 5). *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 (Supplementary File 6A). Sister species *Capsella grandiflora* and *C. rubella* were assigned to both clusters to approximately 50%. At K=3, *Capsella grandiflora* and *C. rubella* were retrieved as a separate cluster (Supplementary File 6B). At K=4, *Capsella thracica* was retrieved as an individual species and at K=5 the first intraspecific differentiation took place (Supplementary Files 6C and 6D). Further intraspecific split followed at K=6, K=7 and K=8, inferring subpopulation structure within *Capsella orientalis* and splitting the *C. bursa-pastoris* clade into a Mediterranean and Eurasian clade (Supplementary Files 6E, 6F and 6G). At K=9, *Capsella grandiflora* was recognised as a separate cluster from *C. rubella* (Supplementary File 6H) and at K=10 the population genetic structure of *Capsella orientalis* reflected the same structuring as inferred at K=5 from the *Capsella orientalis* dataset (Supplementary File 7) and was as follows. At K=2, a northern and southern *Capsella orientalis* cluster was inferred, which subsequently fragmented into a Mongolian cluster (at K=3, Supplementary File 7A), an East Middle Asian cluster (at K=4, Supplementary File 7B) and Altai cluster (at K=5, Supplementary File 7C). Overall, the results of LEA were not contradicting the genetic clustering in STRUCTURE but resolution power was weaker and results were not always in accordance with phylogenetic relationships.

Population genetic clustering according to DAPC: DAPC analysis of *Capsella* dataset followed species specific genetic clustering. However, it failed to delimit *Capsella rubella* from *C. grandiflora* along all K-values spanning from K=3 to K=9 (Supplementary File 8). Close relationship between *Capsella bursa-pastoris* and *C. thracica* was inferred from all K-values. No extensive sub-clustering within *Capsella orientalis* was inferred (Supplementary File 8). DAPC analysis of *Capsella orientalis* dataset retrieved the same substructure as inferred from LEA (Supplementary Files 7, 9) and comparable to the one inferred from STRUCTURE (Figure 5). Recognised as separate clusters already at deeper nodes (inferred at K=3 and K=4 in *Capsella orientalis* dataset, and at K=7 and K=9 in *Capsella* dataset in LEA, respectively), the East Middle Asian and Mongolian clusters were spatially well defined the rest of *Capsella orientalis* populations. These were followed by spatially still well separated Altai cluster, indicating a closer

genetic relationship (Supplementary File 9). The North and South Asian *C. orientalis* clusters albeit partially overlapping, were still recognised as two independent clusters (Supplementary File 9).

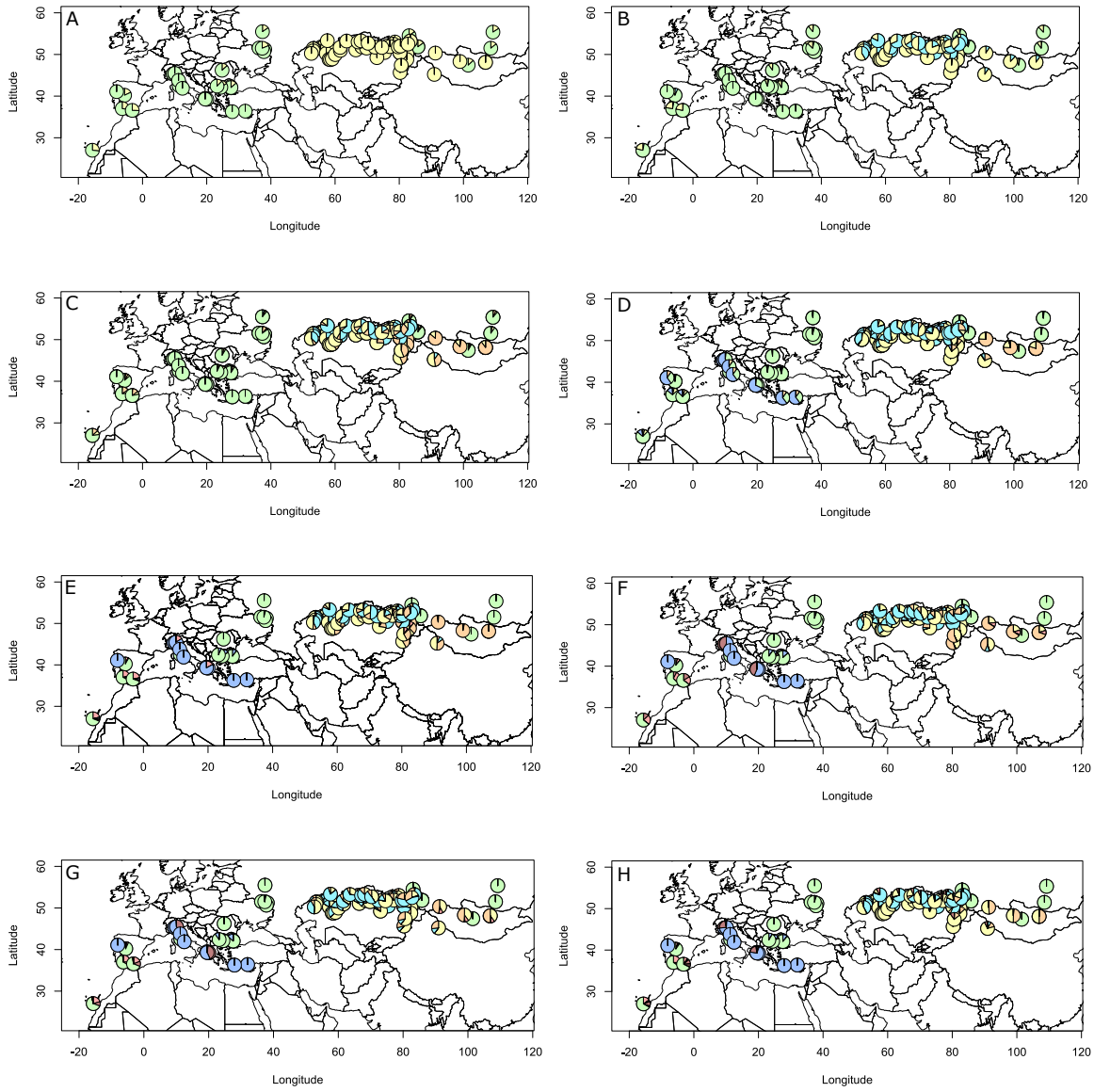


Figure 4: Geographic distribution of sampled *Capsella* populations. Genetic cluster membership was inferred from STRUCTURE-like algorithm, showed as pie charts for K=2 (A), K=3 (B), K=4 (C), K=5 (D), K=6 (E), K=7 (F), K=8 (G) and K=9 (H).

Ecological niche modelling: The best model (LQH 0.5) 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 reduced towards the east and west throughout the Holocene into the south Central Asian Mountain Chains and northern Pontic steppe, respectively (Figure 6B). During the LGM, the only climatically suitable

habitats according to our niche modelling were the Black Sea coast in the west and South and South-East Kazakhstan in the east, with climatically partially suitable habitats also in Eastern China (Figure 6C).

Population genetics inferred from Arlequin: No extensive correlation of geography either with the molecular diversity indexes or with the number of private alleles were inferred (Supplementary File 10). Individual populations with the highest genetic diversity indices (Theta S and Theta Pi) were from Altai, South Kazakhstan and North Kazakhstan. Individual populations with the lowest genetic diversity according to the Theta S and Theta Pi indices were in Altai and Mongolia but also in North Kazakhstan and Tabagartai/Dzhungaria (Supplementary File 11-1). 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 (Supplementary File 11-1). The highest number of private substitution sites per gene copy, however, were 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 Mongolian populations (Supplementary File 11-1). When populations were pooled together according to geography, the highest genetic diversity in terms of Theta S, Theta Pi, as well as proportion of substitution sites per gene copy and private substitution sites per gene copy were retrieved from South Kazakhstan cluster, and the lowest genetic diversity was inferred from the Mongolia genetic cluster (Supplementary File 11-2).

F statistics uncovered that genetically the most 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 Mongolian, Altai and South Kazakhstan populations (Supplementary File 11-3). When pooled together according to the 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 Mongolian genetic cluster ($F_{ST} = 0.110$; Supplementary File 11-4). AMOVA analysis uncovered that the genetic divergence explained among the five genetic clusters was low at only about 3.70% (Supplementary File 11-5). 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

(Supplementary File 11-5). 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 (Supplementary File 11-5). All AMOVA results were highly significant with $p < 0.005$.

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 establishment of two evolutionary lineages within the genus as already outlined in Hurka et al. (2012): *Capsella grandiflora* and *C. rubella* on the Mediterranean side, and *Capsella orientalis*, *C. bursa-pastoris* and *C. thracica* on the Eurasian side (Figures 3 and 7). Our SNP data also supports the two lineages within *C. bursa-pastoris* as previously revealed also by multilocus isozyme genotypes — a Mediterranean and a temperate lineage (Wesse et al. 2021). The findings are in agreement with the SNP-based analyses of Cornille et al. (2016) as well, who detected a European/Russian and a 'Middle Eastern' (which we refer to as 'Mediterranean') *Capsella bursa-pastoris* genetic cluster in the geographic area covered by our study. The origin of *Capsella thracica* is within the temperate *C. bursa-pastoris* lineage, matching the geographical distribution areas of the two taxa.

Comparing our time divergence estimations (Figure 3) with published data reveals the following picture. The stem age of *Capsella* (split between *Capsella* and its monotypic sister genus *Catolobus* (C.A. Mey.) Al-Shehbaz) is estimated to approximately 4 ± 2 Ma (Žerdoner Čalasan et al. 2019). The crown age of *Capsella* (split within *Capsella* into the western and eastern lineage) is estimated at 3.62 Ma (1.15–7.77 Ma) by Han et al. (2015), at 3.2 (0.58–6.98) by Hurka et al. (2012), at 2.6 Ma (3.0–2.4 Ma) by Bachmann et al. (2019), at approximately 1 ± 0.7 Ma by Žerdoner Čalasan et al. (2019), and at 900 kya (161–1159 kya) by Douglas et al. (2015). Timing of the origin of *Capsella bursa-pastoris*, i.e., the split between *C. orientalis* and *C. bursa-pastoris*, is dated to approximately 0.87 Ma (0.006–2.44 Ma) by Hurka et al. (2012), 0.22 (0.09-0.44) Ma by Han et al. (2015), 150 kya by Žerdoner Čalasan et al. (2019), 130 kya (22–177 kya) by Douglas et al. (2015), and 70 kya by Bachmann et al. (2019). Divergence time estimates for the split between *Capsella grandiflora* and *C. rubella* are thought to be extremely young at 20.000 years ago (Foxe et al. 2009), 30.000–50.000 years ago (Guo Y-L et al. 2009), 50.000–100.000 years ago (Brandvain et al. 2013), 0.07 Ma (0.02–0.14 Ma; St Onge et al. 2011), and 0.86 Ma (0.006–2.44 Ma; Hurka et al. 2012). Time estimations based on cpDNA are here excluded due to comparability reasons.

It is obvious that while the evolutionary timescales estimated for *Capsella* are rather divers, they are never seriously conflicting. Time divergence estimates may vary upon taxon coverage, the used molecular markers, assumption about the molecular clock, employed statistical inference methods,

accuracy and number of used calibration points, and assumed substitution rates of DNA sequences. Taken this into account, the main conclusion from hitherto published *Capsella*-dating analyses as well as 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). During the Pleistocene, *Capsella* split into two 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).

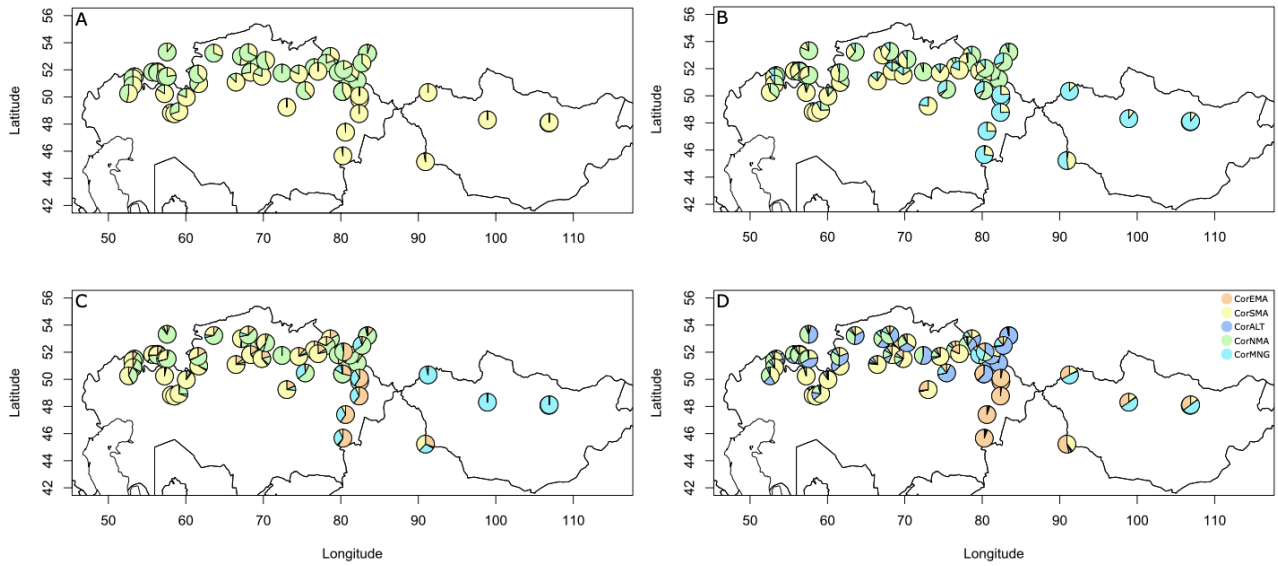


Figure 5: Geographic distribution of sampled *Capsella orientalis* populations. Genetic cluster membership was inferred from STRUCTURE algorithm, showed as pie charts for K=2 (A), K=3 (B), K=4 (C) and K=5 (D). Colour coding reflects the colour coding in the Supplementary Files 4, 7 and 9, and is as follows: North Middle Asian *C. orientalis* subcluster in tea green (CorNMA), South Middle Asian *C. orientalis* subcluster in lemon yellow (CorSMA), Mongolian *C. orientalis* subcluster in celeste (CorMNG), East Middle Asian *C. orientalis* subcluster in deep champagne (CorEMA) and Russian Altai *C. orientalis* subcluster in baby blue (CorALT).

4.2 Origin of the genus *Capsella*

It has already been hypothesised 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, China (with the exception of the humid subtropical southeastern provinces) to southeastern Europe and inhabits woodlands, grasslands, deserts and riverbanks, often as a ruderal weed. *Catolobus* thus shares ecogeographical characters and distribution area with *Capsella*. Morphological evidence (indumentum, trichome types,

branching pattern) and ploidy level (2n=16 as the basal lineages in *Capsella*) also support the close relationship, thus it is reasonable to assume that both genera share a common Asian ancestral area.
We assume that *Capsella* developed in the Western Asiatic sub-region of the Irano-Turanian floristic region sensu Takhtajan (1986), probably in the regional sub-center IT3 (Manafzadeh et al. 2017) as suggested by the present distribution area of *C. orientalis* (Fig. 2). During the Pliocene, the MRCA of *Capsella* extended its range out of Asia 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 by Žerdoner Čalasan et al. (2019). Theoretically, there are two possible east to west migration routes, (i) a 'southern route' south of the Caspian and Black Sea to the eastern Mediterranean region and from there to southeastern Europe, and (ii) a 'northern route' along the steppe biome to southeastern Europe and perhaps even further west into the Mediterranean climate zone.

(i) 'southern route': 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 (Manafzadeh et al. 2014; Jia & Bartish 2018; Lauterbach et al. 2019; Žerdoner Čalasan et al. 2021), and *Capsella* certainly does not belong to this ecological group. The main argument against an early Pliocene southern *Capsella* route, however, is the estimated divergence time between the two basal lineages, the Mediterranean *C. grandiflora/ rubella* and the continental *C. orientalis* lineage of ca. 1 Ma or less, clearly indicating the Pleistocene and not the Pliocene origin. A possible late Pliocene/early Pleistocene 'southern route' furthermore did not exist, because of the Akhchagyl and Apsheron transgressions of the Caspian Sea, which blocked excess 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 'southern route' for ancestral *Capsella* to reach Europe is highly improbable and will not be considered further on.

(ii) 'northern route': Instead, there is convincing evidence for the 'northern route' from the following sources. (a) Present distribution area of *Capsella orientalis* along the Euro-Siberian steppe belt from the Altai/Tabagartai/Dzhungarian Alatau mountain ranges to southeastern Europe (Fig. 2). (b) Spread of steppe elements since the Pliocene from Asia to the Black Sea region clearly indicated by pollen records (e.g. Tarasov et al. 2000; Popova et al. 2017; Popescu et al. 2010; Feurdean and Vasiliev, 2019). (c) Flora and vegetation palaeoreconstructions by Grichuk, 1989. (d)

Climate/landscape history throughout the Neogene and Quaternary (Velichko et al. 2005 for East European Plain; Arkhipov et al. 2005 for West Siberian Plain; Akhmetiev et al. 1999 for Kazakhstan Plains; see also Hurka et al. 2019 and references therein).

In the Middle Pleistocene, the two diploid basal *Capsella* lineages evolved (Figs. 3 and 7). Present distribution area and climate envelope identify *Capsella orientalis* and *C. grandiflora*/*rubella* as true steppe and Mediterranean elements, respectively. Thus, 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 ka (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, b). Periglacial cold steppe spread over the East European Plain north of the Black and Caspian Sea. Interglacial periods were characterised by temperate steppes, forest-steppe and forests (Grichuk, 1989, 1992a, b; 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 were turned over from xeric to more thermic determinism around 1 Ma (Suc et al. 2018).

4.3 *Capsella orientalis*

The distribution map presented in Fig. 2 comprises all hitherto recorded localities of *C. orientalis* known to us (literature surveys, herbarium studies and own field collections). The distribution area stretches nearly 5000 km from eastern Europe (lower Don) to central Asia (Mongolia) more or less along the Eurasian steppe belt. Main distribution is in Kazakhstan.

Present differentiation within *C. orientalis* is of post-glacial age and thus very young, which is also expressed by non-varying of ITS, ETS and cpDNA sequences (data not shown). The *C. orientalis* lineage itself is of Middle Pleistocene age (Fig. 3). It would appear that *C. orientalis* survived the LGM in southern Kazakhstan and southeastern Europe (Fig. 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 Mongolia (Fig. 5). This scenario is supported by the number of private alleles and molecular diversity indexes Theta S and

Pi, which on an average, are higher in southern Kazakhstan than in the recently colonised regions. Lowest diversity was observed in Mongolia. However, discontinuous mosaic variation patterns are also observed (see chapter Results). This overall variation pattern without clear allopatry is typical for recent colonisation events. Interestingly, flowering behaviour 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). *C. orientalis*' flowering ecology thus varies greatly with the environments and growing seasons. We conclude, that the variation pattern at the molecular level mirrors demography, but adaptation at the phenotypic level.

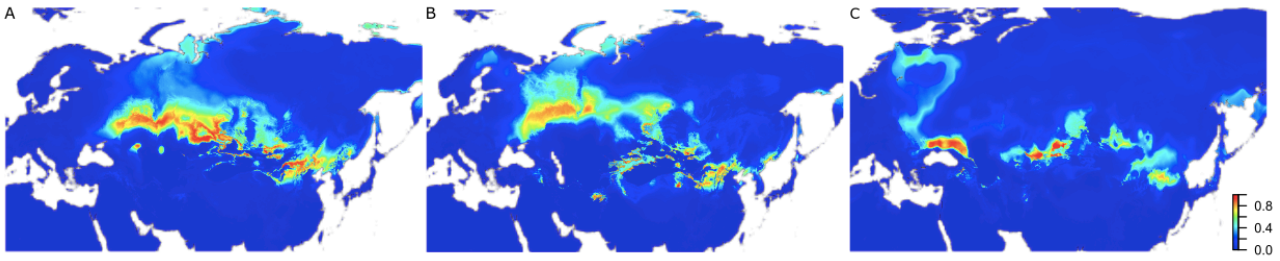


Figure 6: Predicted habitat suitability for *Capsella orientalis* during present time (A), Holocene (B) and Last Glacial Maximum (C), respectively. Warmer colours represent areas of higher probability for the species' presence, while colder colours represent areas with lower probability for its presence. Note that the analysis retrieves only the habitat suitability, but not necessarily the realised ecological niche of *C. orientalis*.

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 the selfing-syndrome research (Hurka et al. 2012; Bachmann et al. 2019; Woźniak et al. 2020). Assuming that self-incompatibility (SC) is a derived character state (Durand et al. 2020), a question arises about the timing of the loss of self-compatibility (SI). Bachmann et al. (2019) argue for a loss of SI in *C. orientalis* between 2.6 Ma and 70 ka but closer to the upper bound, and provide convincing evidence that *C. orientalis* was SC when it contributed to the origin of *C. bursa-pastoris* at about 130 ka.

4.4 Origin of *Capsella bursa-pastoris*

It is evident from the STRUCTURE analysis (Figure 4) that *Capsella bursa-pastoris* retained SNP sequences from both basal lineages. This argues for a hybrid origin of *C. bursa-pastoris* between the two lineages and this has been confirmed by other genomic studies, which show two

subgenomes in *Capsella bursa-pastoris*, one originating from the western and one from the eastern lineage (Douglas et al. 2015; Han et al. 2015; Kryvokhyzha et al. 2019a). Furthermore, subsequent evidence argue that the maternal parent came from the eastern lineage represented by *Capsella orientalis* (Hurka et al. 2012; Douglas et al. 2015; Omelchenko et al. 2020).

To interpret the polyploidization in terms of historical biogeography, we have to postulate first, a split of the continuous MRCA distribution area dividing it into two separated part areas, and second, subsequent extension out of the two part areas leading to a contact zone where the two genomes came back together in a whole-genome duplication. Arguments for a MRCA range split during the Middle and Late Pleistocene as suggested by our time divergence estimation (Fig. 3) are numerous and outlined above. During glaciations, the interglacial steppe belt was replaced north of the Black Sea by periglacial cold steppe and tundra-steppe. In the following interglacial, steppe and forest-steppe expanded again in the east-to-west direction. This eastward contraction and westward expansion of the steppe belt occurred several times during the climatic macrocycles. In Asia, the steppe could move to the south whereas in southeastern Europe due to geophysical barriers (Black Sea and Caucasus) steppe could not escape to the south and the continuous range of the MRCA was disrupted. Its western part areal was pushed west of the Black Sea into the Balkan Peninsula and thus 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 (Magyari et al. 2008; Turrill, 1929; Horvat et al. 1974; Suc et al. 2018), and might have also facilitated the westward migration of *Capsella* into the Balkan Peninsula. According to our divergence time estimations (Fig. 3), the hybridisation event leading to *C. bursa-pastoris* took place during the last Pleistocene transition from glacial to interglacial, about 120,000 yr ago, corresponding approximately to the last interglacial (LIG) MIS 5e (Eem or Mikulino respectively). During the climatic optimum of the LIG (about 120,000 yrs ago), nemoral type forest grew around the Black Sea and northern parts of the Balkan, down to 40° latitude. 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 situation only during climatic/vegetation optima phases of the interglacials. At the transition between glacials and interglacials optima phases, vegetation went through longer periods with open vegetation as grasslands or forest-steppe (Leroy et al. 2011), and

this opened up a short time window for steppe plants to migrate out of their preceding glacial refugia (western Balkan and eastern Black Sea/Kazakhstan) resulting in a secondary contact.

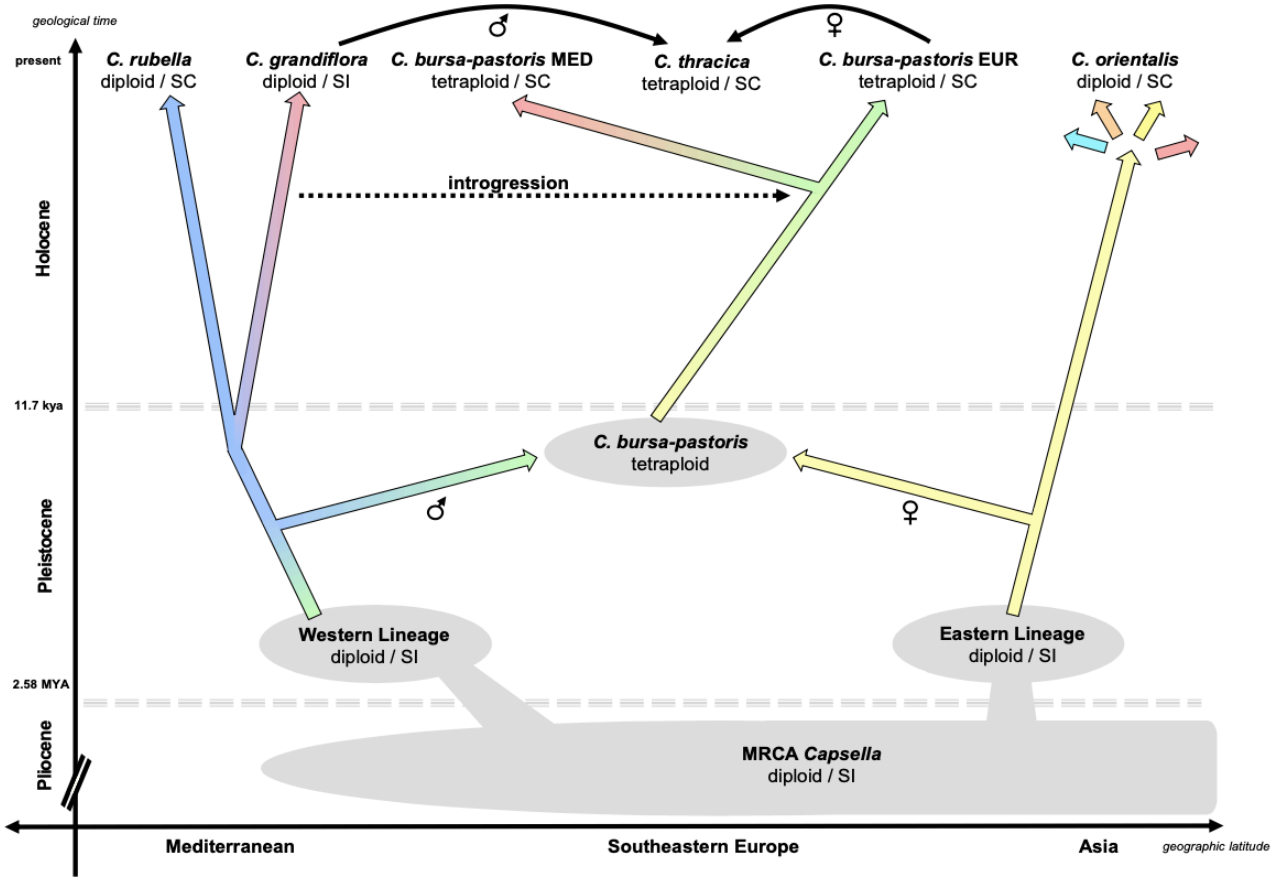


Figure 7: Outline of the evolutionary history of the genus *Capsella* placed into space and time. The colour coding of the arrows correspond to the colour coding in the Figure 4. Note that the geological time axis as well as the geographical latitude axis are not to scale.

Predicted distribution of *Capsella grandiflora* at Present, LGM and LIG based on ecological niche modelling support our assumption of range extension the east into the Black Sea region (Han et al. 2015). We hypothesise 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. Here, in the secondary contact zone, *Capsella bursa-pastoris* originated by hybridisation and whole-genome duplication (allotetraploidy). Under the assumption that two haploid gametes fused followed by polyploidisation, this was not a single event but must had taken place repeatedly several times during the Pleistocene. 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".

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 and Soltis 2020; Spoelhof et al. 2020; Walden et al. 2020), and *Capsella bursa-pastoris* because of its young age received special attention in the recent years (Douglas et al. 2015; Kasianov et al. 2017; Kryvokhyzha et al. 2019a, b). It is another prime example of colonisation success of a polyploid plant species. It gained greatest colonisation ability enabling it to migrate into different climate zones and colonise far above those of its parent species areas and quickly spread over Eurasia. Macrofossils (seeds) of *C. bursa-pastoris* have been reported from the interglacial deposits at Ilford, Essex, England (West et al. 1964). The sediments where the fossils have been found are correlated with MIS 5e (Eemian of continental Europe). Distribution of *C. bursa-pastoris* on the British Isles at the LIG is predicted by palaeomodelling as well (Han et al. 2015).

4.5 Post-glacial hybridisations

Capsella thracica: In the early Holocene, *C. thracica* originated by hybridisation between *C. bursa-pastoris* and *C. grandiflora* in southwestern Europe (Hurka et al. 2012). This is clearly confirmed by our time divergence estimation (Fig. 3) and by the structure analysis (Fig. 4). During the LGM (20.000 – 18.000 yrs) periglacial steppes covered the landscape from the northern Black Sea coast to the Altai-Tien Shan-Pamir mountain systems (Grichuk, 1992 a, b; Velichko and Isayeva, 1992) and divided the preceding LIG grassland/dry shrubland/nemoral forest belt into a Mediterranean and Asian area, which became closed again in the Holocene. We assume that during the early Holocene *C. grandiflora* extended its range from its LGM Balkan refuge to the east into the Thracian Plain where the hybridisation occurred analogue to the one resulting in the origin of *C. bursa-pastoris*. There is a strong Mediterranean influence in the western Pontic steppe vegetation. The origin of the submediterranean geoelements in the steppe vegetation is since long been discussed and it is hypothesised 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 *C. grandiflora* scenario fits into this picture very well.

***Capsella bursa-pastoris* MED**: Wesse et al. (2021) detected two main clusters within *C. bursa-pastoris* based on isozyme variation patterns, one predominantly in Mediterranean-like climate regions (*C. bursa-pastoris* MED), the other in more temperate climate regions (*C. bursa-pastoris* EUR). These two within species lineages are confirmed by our SNP data (Fig. 4) and their origin is of Holocene age (Fig. 3). We hypothesise, that in the Mediterranean *Capsella bursa-pastoris* had

contact with the diploid Mediterranean species *C. grandiflora/rubella* and by introgression of *C. grandiflora/rubella* gene material into the *C. bursa-pastoris* genome, the MED lineage originated (Fig. 3E). This gene flow may have supported adaptation to Mediterranean climates during the postglacial colonisation. To what extent the present distribution of *C. bursa-pastoris* MED mirrors demography and/or adaptation remains an open question but species distribution models of the MED and EUR lineages are clearly in favour of respective climate preferences (Wesse et al. 2021).

Mixed populations of diploid *C. grandiflora/rubella* and tetraploid *C. bursa-pastoris* are common in the Mediterranean region. These species often hybridise, 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 hybridogenous stage may eventually serve as a bridge for introgression of diploid gene material across ploidy barriers into *C. bursa-pastoris*. Introgression of *C. rubella* into *C. bursa-pastoris* is recorded by Slotte et al. (2008), and adaptive introgression from the diploid congeneric *Capsella* species into *C. bursa-pastoris* has been demonstrated by Han et al. (2015).

5 Conclusion

Capsella is an evolutionary young taxon that originated in the western part of the Irano-Turanian floristic region, probably in the Kazakh and adjacent Central Asia steppe biomes. Its history is highly influenced by the Pleistocene/Holocene climatic macrocycles which caused range shifts, range contractions and expansions as well as range splits of the steppe biomes and consequently its genetic pool. This biogeographic dynamic is a key factor in understanding differentiation and speciation processes in time and space. The hybridisation leading to the allotetraploid *C. bursa-pastoris* occurred in the Late Pleistocene in the southwestern part of the East European Plain. Hybridisation events in post-glacial times are placed again into the southwestern part of the East European Plain and in the Mediterranean (Fig. 7). The evolutionary history and historical biogeography of the genus *Capsella* reflect the climate-landscape history of the Eurasian steppe belt.

6 References

Akhmetyev MA, Dodonov AE, Sotnikova MV, Spasskaya II, Kremenetzky KV and Klimanov VA. 1999. Kazakhstan and Central Asia – Plains and Foothills, in: Climate and Environment Changes During the Last 65 Million Years; Cenozoic: From Paleocene to Holocene. GEOS, Moscow, pp. 165–191 (in Russian).

Arkhipov SA, Volkova VS, Zolnikov ID, Zykina VS, Krukover AA and Kulkova LA. 2005. West Siberia. In Velichko AA and Nechaev VP. (Eds). Climate and Environment Changes in Russia, pp. 67–88. Geological Society of America, Special Paper 382. Boulder, Colorado USA

Bachmann JA, Tedder A, Laenen B, Fracassetti M, Désamoré A, Lafon-Placette C, Steige KA et al. 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 PMC, Meijer N et al. 2020. Cenozoic evolution of the steppe-desert biome in Central Asia. *Science Advances* **6**, eabb8227.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, Heled J et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* **15**, e1006650.

Brandvain Y, Slotte T, Hazzouri KM, Wright SI and Coop G. 2013. Genomic identification of founding haplotypes reveals the history of the selfing species *Capsella rubella*. *PLoS Genetics* **9**, e1003754.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA and 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.

Buono D, Khan G, von Hagen KB, Kosachev PA, Mayland-Quellhorst E, Mosyakin SL and Albach DC. 2021. Comparative phylogeography of *Veronica spicata* and *V. longifolia* (Plantaginaceae) across Europe: integrating hybridization and polyploidy in phylogeography. *Frontiers in Plant Science* **11**, 588354.

Chater AO. 1964. *Capsella* Medicus. In: Flora Europaea, vol. 1, First Edition. Cambridge Univ. Press, New York.

Chater AO. 1993. *Capsella* Medicus. In: Flora Europaea, vol. 1, Second Edition. Cambridge Univ. Press, New York.

Chifman J and 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 SI and 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 TLP, Franzke A, Al-Shehbaz IA, Bakker FT, Koch MA and Mummenhoff K. 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Molecular Biology and Evolution* **27**, 55–71.

Douglas GM, Gos GG, Steige KA, Salcedo A, Holm K, Josephs EB, Arunkumar R et al. 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 et al. 2020. Evolution of self-incompatibility in the Brassicaceae: Lessons from a textbook example of natural selection. *Evolutionary Applications* **13**, 1279–1297.

Earl DA and von Holdt BM. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359–361.

Eaton D and Overcast I. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* **36**, 2592–2594.

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

Evanno G, Regnaut S and Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* **14,** 2611–2620.

Excoffier L and Lischer HEL. 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.

Fernández-Mazuecos M, Mellers G, Vigalondo B, Sáez L et al. 2018. Resolving recent plant radiations: power and robustness of genotyping-by-sequencing. *Systematic Biology* **67**, 250–268.

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

Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H and Wright SI. 2009. Recent speciation associated with the evolution of selfing in *Capsella*. *Proceedings of the National Academy of Sciences USA* **106**, 5241–5245.

Franzke A, Hurka H, Janssen D, Neuffer B, Friesen N, Markov M and 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.

Frichot E and François O. 2015. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* **6**, 925–929.

Friesen N, German DA, Hurka H, Herden T, Oyuntsetseg B and 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 SV, Shmakov AI, Herden T, Oyuntsetseg B and Hurka H. 2020. *Allium* species of section *Rhizomatosa*, early members of the Central Asian steppe vegetation. *Flora* **263**, 151536.

Geltman DA and Ebel AL. 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 VP. 1989. History of flora and vegetation of the Russian Plain during the Pleistocene. Nauka Press, Moscow, pp. 183 (In Russian)

Grichuk VP. 1992a. Last interglacial (about 120,000 yr B.P.): Vegetation. In: Frenzel B, Pésci M, Velichko AA. (Eds). Atlas of Paleoclimates and Paleoenvironments of the Northern Hemisphere, pp. 11. G. Fischer, Budapest, Stuttgart.

Grichuk VP. 1992b. Maximum cooling of the last glaciation (about 20,000 to 18,000 yr B.P.): Main types of vegetation (ecosystems). In: Frenzel B, Pésci M, Velichko AA. (Eds). Atlas of Paleoclimates and Paleoenvironments of the Northern Hemisphere, pp. 57. G. Fischer, Budapest, Stuttgart.

German DA, Chen WL, Smirnov SV, Liu B, Kutzev MG, Wang J, Shmakov AI et al. 2012. Plant genera and species new to China recently found in northwest Xinjiang. *Nordic Journal of Botany* **30**, 61–69.

Guo X, Liu J, Hao G, Zhang L, Mao K, Wang X, Zhang D et al. 2017 Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics* **18,** 176.

Guo Y-L, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, Weigel D and Schierup MH. 2009. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-

incompatibility and an extreme bottleneck. *Proceedings of the National Academy of Sciences USA* **106**, 5246–5251.

Han T-S, Wu Q, Hou X-H, Li Z-W, Zou Y-P, Ge S and 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.

Hantemirova EV, Belyaev AY, Korchagina OS and laletina IA. 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.

Hijmans RJ, Cameron SE, Parra JL, Jones PG and Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965–1978.

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

Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q et al. 2016. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Molecular Biology and Evolution* **33**, 394–412.

Hurka H and Neuffer B. 1997. Evolutionary processes in the genus *Capsella* (Brassicaceae). P*lant Systematics and Evolution* **206**, 295–316.

Hurka H, Friesen N, Bernhardt K-G, Neuffer B, Smirnov SV, Shmakov AI and Blattner FR. 2019. The Eurasian steppe belt: Status quo, origin and evolutionary history. *Turczaninowia* **22**, 5–71.

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

Jakobsson M and Rosenberg NA. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23,** 1801–1806.

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

Jia D-R and Bartish IV. 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.

Jombart T 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405.

Kasianov AS, Klepikova AV, Kulakovskiy IV, Gerasimov ES, Fedotova AV, Besedina EG, Kondrashov AS et al. 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 MV. 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 MA, German DA, Kiefer M and Franzke A. 2018. Database taxonomics as key to modern plant biology. *Trends Plant Science* **23**, 4–6.

Kryvokhyzha D, Salcedo A, Eriksson MC, Duan T, Tawari N, Chen J et al. 2019a. Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid *Capsella bursa-pastoris* (Brassicaceae). *PLoS Genetics* **15**, e1007949.

Kryvokhyza D, Kilesi P, Duan T, Orsucci M, Wright SI, 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.

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

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

Leroy SAG, Arpe K and Mikolajewicz U. 2011. Vegetation context and climatic limits of the Early Pleistocene hominin dispersal in Europe. *Quaternary Science Reviews* **30**, 1558–1463.

Magyari EK, Chapman JC, Gaydarska B, Marinova E, Deli T, Huntley JP, Allen JRM and 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.

Mai DH. 1995. Tertiäre Vegetationsgeschichte Europas: Methoden und Ergebnisse. Gustav Fischer Verlag Jena, Stuttgart.

Manafzadeh S, Salvo G and 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 and 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 TM, Laenen B and Slotte T. 2020. Population genomics of transitions to selfing in Brassicaceae model systems. In: Dutheil JY (Ed.), Statistical Population Genomics, Methods in Molecular Biology, vol. 2090. Humana, New York, pp. 269–287.

Meyer M, Stenzel U and Hofreiter M. 2008. Parallel tagged sequencing on the 454 platform. *Nature Protocols* **3**, 267–278.

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

Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M and Anderson, RP. 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 and Wendel JF. 2020. Genomics of evolutionary novelty in hybrids and polyploids. *Frontiers in Genetics* **11**, 792.

Omelchenko DO, Makarenko MS, Kasianov AS, Schelkunov MI, Logacheva MD and Penin AA. 2020. Assembly and analysis of the complete mitochondrial genome of *Capsella bursa-pastoris*. *Plants* **9**, 469.

Phillips SJ, Anderson RP and Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231–259.

Poland JA, Brown PJ, Sorrells ME and 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.

Popescu S-M, Biltekin D, Winter H, Suc J-P, Melinte-Dobrinescu MC., Klotz S, Rabineau M et al. 2010. Pliocene and Lower Pleistocene vegetation and climate changes at the European scale: Long pollen records and climatostratigraphy. *Quaternary International* **219**, 152–167.

Popov SV, Shcherba IG, Ilyina LB, Nevesskaya LA, Paramonova NP, Khondkarian SO et al. 2006. Late Miocene to Pliocene plalaeogeography of the Paratethys and its relation to the Mediterranean. *Palaeogeography, Plaeoclimatology, Palaeoecology* **238**, 91–106.

Popova S, Utescher T, Gromyko DV, Bruch AA, Henrot A-J and Mosbrugger V. 2017. Cenozoic vegetation gradients in the mid- and higher latitudes of Central Eurasia and climatic implications. *Palaeogeography Palaeoclimatology Palaeoecolgy* **467**, 69–82.

Pritchard JK, Stephens M and Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

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

Rambaut A, Drummond AJ, Xie D, Baele G and Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67**, 901–904.

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

Seidl A, Tremetsberger K, Pfanzelt S, Blattner FR, Neuffer B, Friesen F, Hurka H et al. 2021.The phylogeographic history of *Kraschenninnikovia* reflects the development of dry steppes and semi-deserts in Eurasia. *Scientific Reports* **11**, 6645.

Seregin A, Anackov G and 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 SA, Ebel AL and Buko TE. 2011. Supplement to the flora of Kemerovo region since 2001 till 2010. *Turczaninowia* **14**, 65–74.

Soltis PS and Soltis DE. 2020. Plant genomes: markers of evolutionary history and drivers of evolutionary change. *Plants People Planet* **3**, 74–82.

Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus S, Guo Y-L, Steige K et al. 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature Genetics* **45**, 831–835.

Spoelhof JP, Keeffe R and MCDaniel SF. 2020. Does reproductive assurance explain the insidence of polyploidy in plants and animals? *New Phytologist* **227**, 14–21.

St Onge KR, Källman T, Slotte T, Lascoux M and Palmé AE. 2011. Contrasting demographic history and population structure in *Capsella rubella* and *Capsella grandiflora*, two closely related species with different mating systems. *Molecular Ecology* **20**, 3306–3320.

Stange M, Sanchez-Villagra MR, Salzburger W and 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.

Swofford DL. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Svitoch AA. 2016. Regressive periods of the Great Caspian. *Water Resources and the Regime of Water Bodies* **43**, 134–148.

Suc J-P, Popescu S-M, Fauquette S, Bessedik M, Jiménez-Moreno G, Bachirir Taoufiq N, Zheng Z et al. 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.

Takhtajan A. 1986. Floristic Regions of the World. Univ. California Press, Berkeley.

Tarasov PE, Volkova VS, Webb T, Guiot J, Andreev AA, Bezusko LG, Bezusko TV et al. 2000. Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern Eurasia. *Journal of Biogeography* **27**, 609–620.

Turrill WB. 1929. The plant life of the Balkan Peninsula: a phytogeographical study. Clarendon Press, Oxford.

Velichko AA. 1999. General features of the landscape and climate of northern Eurasia in the Cenozoic, in: Climate and environment changes during the last 65 million years; Cenozoic: from Paleocene to Holocene. GEOS, Moscow, pp. 219–233.

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

Velichko AA, Akhlestina EF, Borisova OK, Gribchnko YuN, Zhidovinov NYa, Zelikson EM, Iosifova YuI et al. 2005. East European Plains. In Velichko AA. and Nechaev VP (Eds). Climate and Environment Changes in Russia, pp. 31–66. Geological Society of America, Special Paper 382. Boulder, Colorado USA.

Velichko AA and Isayeva LL. 1992. Maximum cooling of the last glaciation (about 20,000 to 18,000m yr B.P.): Landscape types. In Frenzel B, Pésci M and Velichko AA (Eds). Atlas of Paleoclimates and Paleoenvironments of the Northern Hemisphere, pp. 59. G. Fischer, Budapest, Stuttgart.

Velichko AA, Faustova MA, Pisareva VV, Gribchenk YuN, Sudakova NG and Lavrientev NV. 2011a. Glaciations of the East European Plain: Distribution and chronology. Development in *Quaternary Science* **15**, 337–359.

Velichko AA, Timireva SN, Kremenetski KV, MacDonald GM and Smith LC. 2011b. West Siberian Plain as a late glacial desert. *Quaternary International* **237**, 45–53.

Volkova PA, Herden T and 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 DA, Wolf EM, Kiefer M, Rigault P, Huang X-C, Kiefer C et al. 2020. Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. *Nature Communications* **11,** 3795

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

West RG, Lambert CA and Sparks B. 1964. Interglacial deposits at Ilford, Essex. *Philosophical Transactions of the Royal Society B* **247**, 185–212.

Woźniak NJ, Kappel C, Marona C, Altschmied L, Neuffer B and 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 AV, Mottley J, Lewis R and Brandham PE. 2000. Nuclear DNA amounts in Roses. *Annals of Botany* **85**, 557–561.

Žerdoner Čalasan A, Seregin AP, Hurka H, Hofford NP and 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.

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