Comenius University in Bratislava

Faculty of Natural Sciences

Slovak Academy of Sciences

Plant Science and Biodiversity Centre

Charles University in Prague

Faculty of Science

Evolution of members of the genus *Cardamine* from the Anatolia-Caucasus region and the Balkan Peninsula

Mgr. Adam Kantor

Dissertation thesis

Bratislava 2023

Prírodovedecká fakulta, Univerzita Komenského v Bratislave

Botanický ústav, Centrum biológie rastlín a biodiverzity,
Slovenská Akadémia Vied

Přírodovědecká fakulta, Univerzita Karlova

Evolúcia zástupcov rodu žerušnica (*Cardamine*, Brassicaceae) z oblasti Anatólie, Kaukazu a Balkánskeho polostrova

Dizertačná práca

Mgr. Adam Kantor

Študijný odbor: Biológia

Študijný program: Botanika

Školitelia: Mgr. Judita Zozomová, PhD. (CBRB SAV, Bratislava),

Mgr. Marek Slovák, PhD. (PřF UK, Praha)

Konzultant: Ing. Jaromír Kučera, PhD. (CBRB SAV, Bratislava).

Bratislava 2023

Faculty of Natural Sciences, Comenius University in Bratislava

Institute of Botany, Plant Science and Biodiversity Centre,
Slovak Academy of Sciences

Faculty of Science, Charles University in Prague

Evolution of members of the genus *Cardamine* from the Anatolia-Caucasus region and the Balkan Peninsula

Dissertation thesis

Mgr. Adam Kantor

Field of study: Biology

Study programme: Botany

Supervisors: Mgr. Judita Zozomová, PhD. (PSBC SAS, Bratislava),

Mgr. Marek Slovák, PhD. (PřF UK, Praha)

Consultant: Ing. Jaromír Kučera, PhD. (PSBC SAS, Bratislava).

Bratislava 2023

Univerzita Komenského v Bratislave
Prírodovedecká fakulta

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Mgr. Adam Kantor
Študijný program: botanika (Jednoodborové štúdium, doktorandské III. st., denná forma)
Študijný odbor: biológia
Typ záverečnej práce: dizertačná
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Evolution of members of the genus Cardamine from the Anatolia-Caucasus region and the Balkan Peninsula
Evolúcia zástupcov rodu žerušnica (Cardamine, Brassicaceae) z oblasti Anatólie, Kaukazu a Balkánskeho polostrova

Anotácia: Cieľom práce je vyriešenie evolučnej histórie hygrofytických druhov z rodu Cardamine z oblasti Anatólie, Kaukazu a Balkánskeho polostrova, tradične klasifikovaných do skupín C. amara, C. pratensis a C. tenera. Autor sa zameria na vysvetlenie evolučných procesov a environmentálnych faktorov, ktoré sa podieľali na diverzifikácii a rozšírení dnešných druhov, ako aj na objasnenie fylogenetických vzťahov medzi druhmi z juhozápadnej Ázie a Balkánu. Pre tento účel bude aplikovaný integratívny prístup, kombinujúci metódu využívajúcu „next generation sequencing" (Hyb-Seq), tradičné molekulárne markery, karyologické analýzy (prietoková cytometria, počítanie chromozómov), metódy multivariačnej morfometriky a analýzy ekologických ník.

Školiteľ: Mgr. Judita Zozomová, PhD.
Katedra: PriF.KBo - Katedra botaniky
PriF vedúci katedry: doc. RNDr. Jana Ščevková, PhD.

Dátum zadania: 24.01.2019

Dátum schválenia: 24.01.2019 prof. RNDr. Karol Mičieta, PhD.
garant študijného programu

.....
študent

.....
školiteľ

Comenius University Bratislava
Faculty of Natural Sciences

THESIS ASSIGNMENT

Name and Surname:	Mgr. Adam Kantor
Study programme:	Botany (Single degree study, Ph.D. III. deg., full time form)
Field of Study:	Biology
Type of Thesis:	Dissertation thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Evolution of members of the genus Cardamine from the Anatolia-Caucasus region and the Balkan Peninsula

Annotation: The thesis aims to resolve evolutionary history of the hygrophytic species of the genus Cardamine in the Anatolia-Caucasus region and the Balkan Peninsula, traditionally classified to the C. amara, C. pratensis and C. tenera groups. The author will focus on explaining which evolutionary processes and environmental factors have shaped the diversification and distribution of the extant species, as well as elucidating the phylogenetic relationships between the representatives from Southwestern Asia and the Balkans. For this purpose, an integrative approach will be used, combining the method based on next generation sequencing (Hyb-Seq), conventional molecular markers, karyological analyses (flow cytometry, chromosome counting), methods of multivariate morphometrics, and ecological niche analyses.

Tutor:	Mgr. Judita Zozomová, PhD.
Department:	PriF.KBo - Katedra botaniky
Head of Department:	doc. RNDr. Jana Ščevková, PhD.
Assigned:	24.01.2019
Approved:	24.01.2019 prof. RNDr. Karol Mičieta, PhD. Guarantor of Study Programme

.....
Student

.....
Tutor

Čestne prehlasujem, že dizertačnú prácu som vypracoval sám, pod odborným vedením svojich školiteľov a s použitím len uvedenej literatúry.

I declare that I have prepared the dissertation thesis solely by myself, under guidance of my supervisors and using only the cited literature.

……

Mgr. Adam Kantor

ACKNOWLEDGEMENT

I would like to thank both of my supervisors, Judita Zozomová and Marek Slovák, for their valuable guidance, friendly approach and sharing of their knowledge and experiences with me. I thank all coauthors of the articles presented in this thesis for their contribution and cooperation, and I want to thank especially Marek Šlenker for his invaluable input in processing of bioinformatic data. I am also grateful for the assistance and supportive attitude of my other colleagues and friends from Department of Botany, Charles University in Prague, and particularly from the Institute of Botany, Slovak Academy of Sciences in Bratislava, where I spent most of the time during my study. I am very thankful for the patience, support and care provided by my family and friends. Special thanks go to my wife, Betka, who stood by me at all times.

My work on studies included in the thesis was supported by the grants provided by the Slovak Research and Development Agency (APVV 17-0616; APVV 21-0044), the Czech Science Foundation (GAČR 19-06632S), the bilateral Slovak-Turkish project TUBITAK no. 475542 ("Evolutionary Legacy of the Turkish Flora and its Influence on Formation of the European Plant Diversity: Evidence from Genome to Ecology") (all four projects were supervised by prof. Karol Marhold), and the programme for PhD students at Slovak Academy of Sciences, Doktogrant APP0271 (Adam Kantor).

COTUTELLE PROGRAMME

The doctoral study was realized under Cotutelle programme, being supervised by two tutors from two scientific institutions: Mgr. Judita Zozomová, PhD. (Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Bratislava, Slovakia) and Mgr. Marek Slovák, PhD. (Department of Botany, Charles University, Prague, Czechia).

ABSTRAKT

Táto práca si kladie za cieľ objasniť evolúciu zástupcov vybraných druhových skupín rodu *Cardamine*, ktorý predstavuje druhovo bohatý rod s celosvetovým rozšírením a veľmi zložitou evolučnou históriou. Skúmané druhy predstavujú vlhkomilné trváce byliny, pričom práca bola zameraná na štúdium ich diverzity na území Balkánskeho polostrova a juhozápadnej Ázie, s dôrazom kladeným najmä na region Anatólie a Kaukazu. Tieto oblasti obsahujú celosvetové centrá biodiverzity (global biodiversity hotspots) a dôležité glaciálne refúgiá, no napriek tomu je len veľmi málo známe o evolučných mechanizmoch a environmentálnych faktoroch, ktoré v týchto oblastiach zohrávali rolu pri diverzifikácii a speciácii rastlín. V tejto práci boli adresované otázky týkajúce sa fylogenézy a taxonómie študovaných skupín, s cieľom stanoviť význam polyploidizácie a hybridizácie v ich evolúcii. Hlavnou aplikovanou metódou bola technika Hyb-Seq, založená na princípe sekvenovania novej generácie (next-generation sequencing), ktorá sa ukázala byť veľmi účinnou a všestrannou metódou pri riešení evolučných otázok. V práci bol použitý integratívny prístup, kombinujúci Hyb-Seq s rôznymi inými karyologickými, molekulárnymi, cytogenetickými technikami a metódou modelovania ekologických ník, čo umožnilo veľmi komplexný pohľad na riešené témy. Práca objasňuje fylogenetické vzťahy medzi študovanými líniami, ako aj medzi taxónmi z oboch študovaných oblastí, pričom zároveň poskytuje dôkaz o historickom biogeografickom prepojení Balkánu s Anatóliou. V kontraste so situáciou v ostatných častiach Európy, výsledky ukázali, že medzidruhová hybridizácia a polyploidizácia ovplyvnili evolučnú históriu študovaných druhov len v malom rozsahu, hoci boli nájdené aj výnimky v podobe dvoch alopolyploidných druhov. Ekologická divergencia a biogeografická izolácia boli identifikované ako jedny z najzásadnejších faktorov podporujúcich speciáciu v študovaných komplexoch. V predchádzajúcich štúdiách bol pozorovaný nesúlad medzi geografickou, morfologickou a ekologickou štruktúrou študovaných druhov, avšak táto nesprávna interpretácia bola podľa všetkého spôsobená relatívne nedávnou a pritom rýchlou diverzifikáciou záujmových skupín, čo dopĺňa aj vysoká vnútrodruhová variabilita pozorovaná u niektorých z nich.

Kľúčové slová: *Cardamine*; Hyb-Seq; alopolyploidia; endemizmus; hygrofytické druhy

ABSTRACT

This thesis aims to elucidate evolution of the members of the selected species groups of the genus *Cardamine*, which is a species-rich genus with cosmopolitan distribution and very complex evolutionary history. The studied species were represented by hygrophytic perennials, with the focus laid on their diversity in the Balkan Peninsula and Southwestern Asia, with the emphasis put on the Anatolica-Caucasus region. These regions harbour global biodiversity hotspots and important glacial refugia, yet, they have been largely understudied in terms of knowledge of evolutionary mechanisms and environmental factors that have played there a role of plants diversification and speciation drivers. In this thesis, questions concerning the phylogeny and taxonomy of the studied groups were addressed, with aim to determine the role of polyploidy and hybridization in their evolution. Hyb-Seq was the majorly applied method, representing a technique based on next-generation sequencing, which has proven to be a very efficient and versatile method for resolving evolutionary questions. An integrative approach was employed, combining Hyb-Seq with a variety of other karyological, molecular, cytogenetic and ecological niche modelling methods, which provided a very complex insight into the addressed topics. This thesis resolves the phylogenetic relationships among the target lineages as well as between the taxa distributed in both studied regions, revealing the evidence of historic Balkan-Anatolian biogeographic links. In contrast to the situation in other parts of Europe, the results showed that the evolutionary history of the studied species was only minorly affected by interspecific hybridization and polyploidization events, even though the exceptions were found, being represented by the two allopolyploid species. The ecological divergence and biogeographic isolation were identified to be among the most crucial factors driving speciation in the target complexes. The reports of conflicts between genetic, morphological and ecological patterns found in previous studies were presumably caused by incorrect interpretation due to the relatively recent and rapid diversification of the studied groups, along with high intraspecific variability observed in part of the species.

Keywords: *Cardamine*; Hyb-Seq; allopolyploidy; endemism; hygrophytic species

CONTENTS

PREFACE 11
1. GENERAL INTRODUCTION 12
1.1 An overview of the genus *Cardamine* 12
1.2 The studied taxa in the context of traditionally recognized species complexes 14
1.3 The focused area of Southwestern Asia and the Balkan Peninsula 18
1.4 Polyploidization and hybridization as the major drivers of speciation in plants and their significance in *Cardamine* 23
2. AIMS OF THE DISSERTATION THESIS 29
3. METHODS 30
3.1 Chromosome counting 30
3.2 Flow cytometry 32
3.3 Morphometric analyses 32
3.4 Ecological niche analyses 33
3.5 Molecular analyses 34
3.5.1 Conventional molecular markers based on single-locus sequencing of chloroplast DNA regions (cpDNA), ITS of nuclear ribosomal (nrDNA) and low-copy nuclear genes 34
3.5.2 Next-generation sequencing (NGS) 37
3.5.3 Hyb-Seq 39
3.6 Genomic in situ hybridization (GISH) 42
4. RESULTS AND DISCUSSION 43
4.1 Allele sorting as a novel approach to resolving the origin of allotetraploids using Hyb-Seq data: A case study of the Balkan mountain endemic *Cardamine barbaraeoides* 44
4.2 Evolution of hygrophytic plant species in the Anatolia-Caucasus region: insights from phylogenomic analyses of *Cardamine* perennials 67
4.3 Balkan-Anatolian biogeographic links and the evolutionary significance of Anatolian mountains evidenced by *Cardamine* (Brassicaceae) 84
5. CONCLUSION 125
5.1 Overall significance of the results of the thesis 129
6. REFERENCES 131

PREFACE

The great species richness of the genus *Cardamine*, along with its worldwide distribution and complicated evolutionary history, makes it a challenging object for studying mechanisms driving speciation, inferring phylogenetic relationships, and reconstructing phylogeographic patterns. Its evolution has been accompanied by relatively recent and rapid diversification and involved multiple polyploidization and hybridization events contributing to its complexity. In the last decades, much attention has been paid to predominantly polyploid species complexes in Europe and Eastern Asia. The present thesis focuses on poorly explored perennial hygrophytic taxa inhabiting the regions of the Balkan Peninsula and Southwestern (SW) Asia, focusing mainly on the Anatolia-Caucasus region within. Both areas harbour biodiversity hotspots of global importance, which are majorly distributed in mountain ranges. Accordingly, majority of the species studied in this dissertation thesis, traditionally classified to the *C. amara*, *C. pratensis* and *C. raphanifolia* species complexes, are ecologically confined to mountains. This thesis provides novel findings concerning their evolutionary history and phylogeny, explaining the processes and environmental factors that shaped its current species diversity and distribution.

Some questions addressed in this thesis arose already in the last two decades, however, methods that would be suitable for answering them were not available yet. Here, we demonstrate the vast significance and utility of using next-generation sequencing based methods and diverse approaches of data procession for resolving complex evolutionary histories. Indeed, the application of Hyb-Seq data allowed us to untangle the problems that remained unresolved in previous studies that used conventional markers based on Sanger sequencing. Additionally, we explored the available methods of analyses of polyploid genomes and introduced a new tool for resolving parentage of allopolyploids. In all studies included in this thesis, the molecular analyses were accompanied by a combination of other karyological, morphometric, cytogenetic or ecological niche modelling methods. Presented results highlight the importance and benefits of using integrative approach, which provided us with very complex insight and understanding of the mechanisms and factors playing role in evolution of the hygrophytic species.

1. GENERAL INTRODUCTION

1.1 An overview of the genus *Cardamine*

Cardamine L. is the largest genus classified in the tribe Cardamineae in the family Brassicaceae (Al-Shehbaz et al. 2006), together with the related genera *Nasturtium* W.T. Aiton, *Barbarea* W.T. Aiton, *Armoracia* G. Gaertn., B. Mey. & Scherb., *Leavenworthia* Torr. and *Rorippa* Scop. (Koch 2003; Beilstein et al. 2008; Couvreur et al. 2010; Huang et al. 2016). While the Brassicaceae family comprises about 4000 taxa in 350 genera according to the most recent checklist (Koch et al. 2018), *Cardamine* by itself is also one of the largest genera, comprising about 280 species (Marhold et al. 2021). The number of recognized taxa strongly depends, however, on the level of taxonomic knowledge and botanical investigations of the given areas. For instance, 31 new species were recently named and described from New Zealand (Heenan 2017). The Plant of the World Online database currently encompasses 263 accepted species in *Cardamine* (as of April 2023; POWO 2023). While there were 31 species listed in Flora Europaea (Jones & Akeroyd 1993), after several detailed taxonomic studies published within one decade, Lihová & Marhold (2006) stated as many as 54 species for the same area, demonstrating the previous underestimation of species diversity and the need of further research of the genus.

Divergence time estimates are complicated in Brassicaceae due to the general lack of reliable fossils within this family (reviewed by Franzke et al. 2016; Huang et al. 2020). For example, the evidence and classification of *Thlaspi primaevum* Becker, the fossil previously used in the phylogenetic divergence time reconstructions (Beilstein et al. 2010; Magallón et al. 2015), was lately disputed (Franzke et al. 2016). Therefore, it was not included in the latest Brassicaceae molecular dating proposed by Huang et al. (2020). According to this study, Brassicaceae diverged in the Oligocene about 30 mya (95 % HPD = 26.8–33.2 Mya), the tribe Cardamineae originated in the Middle Miocene about 17.7 mya (95 % HPD = 12.6–22.1 Mya), while the genus *Cardamine* originated about 12–13 mya (Huang et al. 2020). Consequently, diversification of the genus has been taking place mainly in the Late Miocene, Pliocene, and most of the currently occurring species likely originated during the Pleistocene (Huang et al. 2020). The relatively recent and rapid diversification of the genus has been significantly affected by recurrent polyploidization and interspecific hybridization events (Carlsen et al. 2009; Marhold et al. 2018).

The distribution of the genus is cosmopolitan, except for Antarctica (Al-Shehbaz 1988; Lihová & Marhold 2006). Some of the areas, including the whole Southern Hemisphere, were evidently colonised recurrently by independent events (Carlsen et al. 2009). The centre of origin of the genus is not reliably determined yet. This is mainly due to the insufficiently resolved genus phylogeny, but also because of the global distribution, with several widespread taxa (e.g. *C. bellidifolia* L., *C. hirsuta* L., *C. parviflora* L.) on the one hand, and presence of numerous local endemics throughout the distribution range on the other hand. Based on the number of taxa recorded, the major centres of diversity are found in the European Mediterranean, the Caucasus, Eastern Asia, Himalayas, Central and North America and New Zealand (Lihová & Marhold 2006; Heenan 2017). Representatives of the genus occupy a broad spectrum of terrestrial habitats, from lowlands up to the high alpine zone. Some species have also become unpleasant weeds in cultural fields and gardens, spreading beyond their indigenous distribution area (e.g. *C. hirsuta*, Braithwaite 2003; *C. corymbosa* Hook. f., Kudoh et al. 2007), few of them also obtaining the invasive status (*C. hirsuta*, Matsuhashi et al. 2016; *C. occulta* Hornem, Šlenker et al. 2018).

Cardamine comprises taxa that are either annual, biennial, but predominantly perennial (Al-Shehbaz 1998). They reproduce sexually, by outcrossing, while there are also a few reports of autogamy (reviewed in Lihová & Marhold 2006), including cleistogamy (Braithwaite 2003; Morinaga et al. 2008). However, the vegetative propagation is also common, via rhizome fragmentation, stolons, bulbils, adventitious shoots on leaf blades and adventitious or leafy shoots in the leaf axils (Lihová & Marhold 2006; Tedder et al. 2015; Sun et al. 2020). Apomixis has not been observed in the genus yet, and it is generally very rare in Brassicaceae (Brukhin et al. 2019; Mandáková et al. 2021). Presence of sexual polymorphisms, specifically androdioecy, which is, indeed, unique in Brassicaceae family, has been reported for *C. amara* L. (Tedder et al. 2015). The anatomical structure of fruits and seeds allows the short-distance seed dispersal by means of silique curling (Kimata 1983). Nevertheless, it has been assumed that some of the migration routes must have included extremely long dispersal events (Carlsen et al. 2009). As many of the taxa are confined to water bodies and moist environment, one of the possible dispersal mechanisms is by spreading fragmented vegetative parts of the plant (e.g. rhizomes) by flowing water. Also, the seeds become mucilaginous in wet conditions and can be therefore distributed by their adhesion to animal bodies even for long distances (Al-Shehbaz 1988). Concerning morphology, the genus is characterized by its flat siliques, with a marginate or narrowly winged replum and spirally coiling valves. However, the genus is

morphologically very variable, particularly in leaf morphological traits (Lihová & Marhold 2006; Al-Shehbaz et al. 2010). Leaves are cauline or basal and among different species, all types of compound leaves can be found: trifoliate, palmately compound, pinnate and bipinnately compound, including also simple leaves with the entire to dissected margin (Nikolov et al. 2019). According to Nikolov et al. (2019), Cardamineae represents the tribe with the most diverse leaf morphology in Brassicaceae.

The genus *Cardamine* is karyologically very diverse, with a high proportion of polyploids and different chromosome numbers even within certain species. Several base chromosome numbers were reported (x = 6, 7, 8, 10, 12; Al-Shehbaz 1988), with the most common one being x = 8, and the smallest diploid chromosome number 2n = 16. Some smaller numbers reported in *C. seidlitziana* Albov (2n = 12) and *C. asarifolia* L. (2n = 14) are considered to be erroneous (Kučera et al. 2005; Lihová & Marhold 2006). The highest chromosome number reported is 2n = 256, found in the North American *C. concatenata* (Michx.) O.Schwarz (Kučera et al. 2005). Dysploidy was also proposed in the past (reviewed by Lihová & Marhold 2006), which was recently proved and explained by the process of nested chromosome fusion using comparative chromosome painting (Mandáková et al. 2013). Both aneuploid and dysploid individuals have been repeatedly identified, particularly in the polymorphic *C. pratensis* s. str. (Kučera et al. 2005; Melichárková et al. 2020). The genome sizes of *Cardamine* representatives are among the smallest ones within angiosperms, with *C. impatiens* L. having the value of 1C = 0.21 pg (Leitch et al. 2019).

1.2 The studied taxa in the context of traditionally recognized species complexes

The first infrageneric classification in *Cardamine* was proposed by Schulz (1903), who divided the genus into 12 sections based on morphology. With the arrival of modern methods of cytogenetics and molecular phylogenetics, this classification has been repeatedly proved to be artificial and the monophyly of most of the recognized sections was rejected. Franzke et al. (1998) published the first molecular phylogenetic study of the genus using sequences of rDNA (ITS) and non-coding cpDNA, which gave first insights into the infrageneric relationships. Other phylogenetic studies followed (Sweeney & Price 2000; Bleeker et al. 2002), while Carlsen et al. (2009) presented the most recent genus-wide phylogeny. In their study, 10 infrageneric clades were identified, but relationships among the clades, as well as among the

species within the clades remained unresolved. The recognized clades were mostly congruent with the results of some previous molecular studies focusing on selected species complexes in the genus (e.g. Lihová et al. 2004a; Marhold et al. 2004), as well as with those of the more recent family-wide phylogenetic studies (Huang et al. 2016, 2020). Still, the phylogenetic reconstruction of the genus *Cardamine* is incomplete and complicated by rapid diversification, high incidence of polyploids, and reticulated patterns of evolution (Carlsen et al. 2009).

Over the last three decades, much attention has been paid to three traditionally recognized diploid-polyploid species complexes of *C. amara*, *C. pratensis* and *C. raphanifolia* (Marhold et al. 2018), which comprise a significant part of species recognized in Europe. Altogether they include about 30 taxa, some of which are more widespread in Europe, but most of them represent endemics restricted to Mediterranean regions. The three species complexes comprise perennials growing in humid habitats (i.e., hygrophytes), either flooded or in proximity of standing or running water bodies, from lowlands to the alpine zone (Fig. 1). While they commonly reproduce sexually via seeds, they possess the ability of vegetative propagation by stolons, rhizome fragmentation or leaf vivipary (Marhold & Ančev 1999; Tedder et al. 2015; Sun et al. 2020). Despite the relatively well described morphological delimitation of the three species complexes, phylogenetic studies have shown that this division is partially artificial. While the *C. amara* group remains consistent and apparently monophyletic, the two remaining groups could not be genetically distinguished from each other (Lihová et al. 2004a; Marhold et al. 2004; Carlsen et al. 2009). They formed a single clade, comprising species from both the *C. pratensis* and *C. raphanifolia* groups, including the SW Asian relatives (classified by Khatri (1988) as subsect. Tenerae, also informally referred to as *C. tenera* group; Marhold et al. 2004). According to Huang et al. (2020), these two major clades originated during Pliocene, about 3–4 mya, and most of the extant species evolved in the Pleistocene.

The *C. amara* complex comprises the widespread Eurasian species *C. amara* with five subspecies recognized in Europe (four diploids and one tetraploid), *C. amporitana* Sennen & Pau (tetraploid) with the Catalonian-Apennine range, and *C. lazica* Boiss. & Balansa ex Boiss. (diploid) from the Caucasus and Anatolian region (Jalas & Suominen 1994; Marhold 1992; 1999; Marhold et al. 1996, 2004; Lihová et al. 2000, 2004b). Some controversy, however, has concerned the names *C. lazica* and *C. wiedemanniana* Boiss., as the latter name has either been treated as a synonym, or, by some authors, considered a distinct entity (see e.g. Schulz 1903; Khatri 1988; Lihová et al. 2004b; Kučera et al. 2006). The Mediterranean tetraploid *C. amporitana* (Lihová et al. 2004a) was previously treated at the subspecific level of *C. amara*

(*C. amara* subsp. *olotensis* O. Bolòs; Lihová et al. 2000), but later accepted as a separate species (Lihová et al. 2004b). Its origin (whether auto- or allopolyploid) has been the subject of discussion and is not resolved yet (Lihová et al. 2004a, b; Marhold et al. 2018). The species from the *C. amara* complex have apparently played an important role in the evolutionary history of the genus, as they repeatedly acted as progenitors of several allopolyploids both in Europe and Eastern Asia (e.g. *C. asarifolia*, *C. flexuosa* With., *C. scutata* Thunb., *C. schulzii* Urbanska-Worytkiewicz; Lihová et al. 2006; Mandáková et al. 2013, 2014, 2019; Zozomová-Lihová et al. 2014).

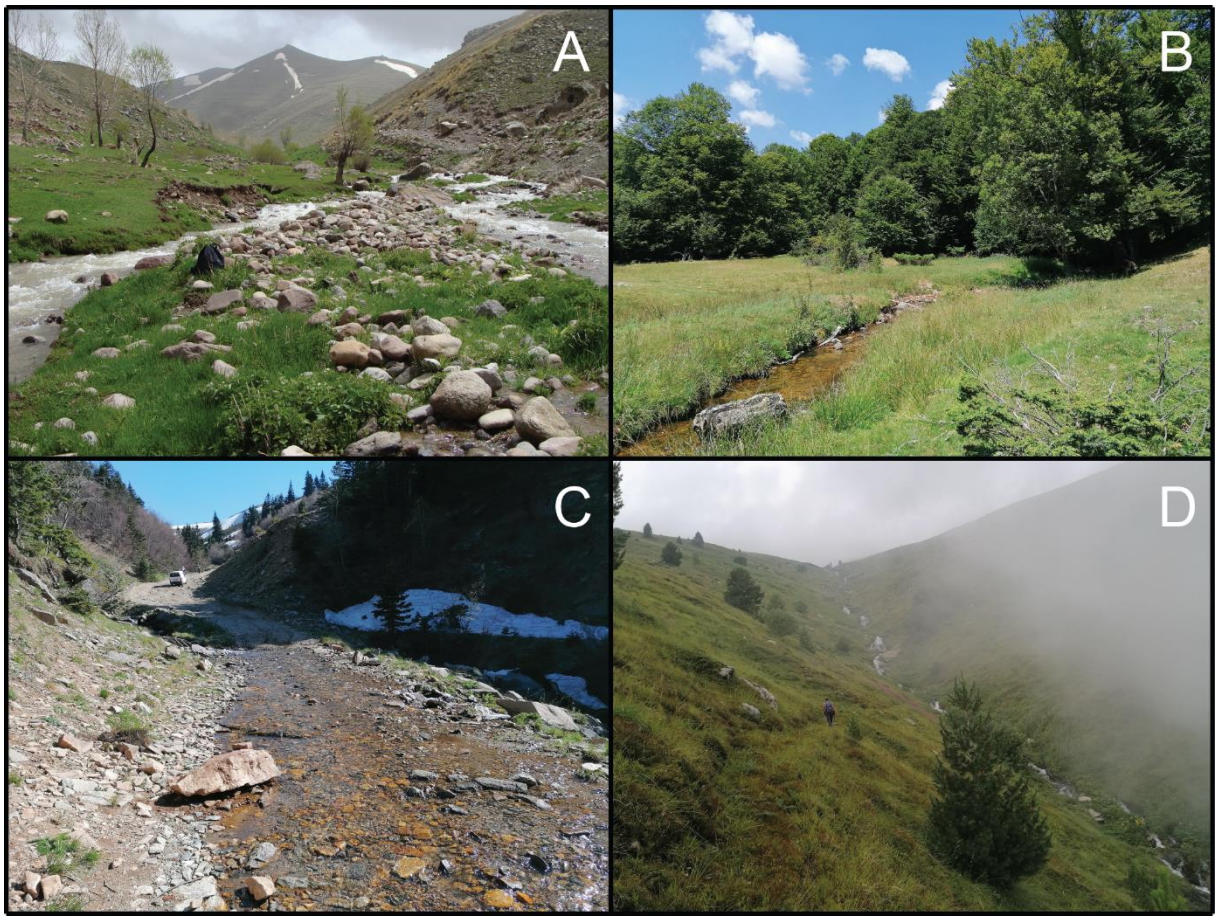


Fig. 1. Examples of habitats of selected species from the studied groups in Iran, Zagros Mts. (A, *C. uliginosa*), Bulgaria, Stara Planina (B, *C. amara* subsp. *balcanica*), Turkey, Uludağ (C, *C. anatolica*), and North Macedonia, Baba Planina (D, *C. acris* subsp. *acris*). Photos taken by Marek Slovák (A) and Adam Kantor (B,C,D).

The traditionally delimited *C. pratensis* species complex is distributed in Europe and extends to Asia, North America, northern Africa, and the Arctic (Jalas & Suominen 1994; Carlsen et al. 2009; Marhold et al. 2018). Its cytotype variability is significantly greater than in

the *C. amara* complex, comprising species and populations with diploid to dodecaploid level (Marhold et al. 2018), including aneuploids and dysploids. There are several well delimited diploid species (e.g. *C. rivularis* Schur, *C. penzesii* Ančev & Marhold*, C. apennina* Lihová & Marhold, *C. crassifolia* Pourr.; Marhold 1995; Marhold & Ančev 1999; Lihová et al. 2003, 2004c), predominantly distributed in Southern Europe. In the central and northern part of Europe, the complex is represented by diploid *C. matthioli* Moretti, its autotetraploid derivate *C. majovskyi* Marhold & Záborský, highly polyploid *C. dentata* Schult.*, C. nymanii* Gand. and *C. pratensis* s. str. (Marhold 1994; Marhold et al. 2018; Melichárková et al. 2020). *Cardamine pratensis* s. str. is the most complex and genetically variable taxon of the group, including diploid to heptaploid populations distributed from lowlands to the alpine level. Various authors attempted to resolve and classify its extraordinarily high chromosomal, genetic, morphological and ecological variation into some segregate taxa (Marhold 1994; Marhold et al. 2018), which, however, turned out to be infeasible. The recent study revealed the existence of three intraspecific genetic lineages just within Central Europe but with diffuse borders and contradicting previously recognized taxa. Multiple polyploidization events both in space and time have been proved, which resulted into a complex reticulate evolutionary history of this species (Melichárková et al. 2020).

The traditionally delimited *C. raphanifolia* group contains four previously recognized subspecies of *C. raphanifolia* Pour. (subsp. *raphanifolia*, subsp. *acris* (Griseb.) O.E. Schulz, subsp. *gallaecica* Laínz and subsp. *barbaraeoides* (Halácsy) Strid; Jalas & Suominen 1994), now treated as separate species (Lihová et al. 2004a; Perný et al. 2004, 2005a, b; Carlsen et al. 2009; Marhold et al. 2018). Whereas *C. raphanifolia* and *C. gallaecica* (Laínz) Rivas Mart. & Izco are Iberian polyploid endemics, *C. acris* Griseb. (diploid) and *C. barbaraeoides* Halácsy (tetraploid) are restricted to the Balkan Peninsula, and all occur in mountainous areas. The taxonomic identity of *C. barbaraeoides*, as well as its distribution, however, has been controversial. It was originally described at the species level (Halácsy 1894), and the following authors treated it either at the species level (Schulz 1903), or more commonly as a subspecies of *C. raphanifolia* (Jones & Akeroyd 1993; Jalas & Suominen 1994) or of *C. amara* (*C. amara* subsp. *barbaraeoides* (Halácsy) Maire & Petitm.; Tan 2002). The species was originally described from the Lakmos Mts. (Southern Pindos, NW Greece), but has commonly been reported also from other parts of the Balkan Peninsula (Strid 1986; Tan 2002; Šlenker et al. 2021). Some previous studies pointed to its close relationship to the *C. amara* group (Lihová et

al. 2004a, b; Perný et al. 2005a), but its circumscription and evolutionary history has not been sufficiently resolved until recently (see chapter 4.1 and Šlenker et al. 2021).

The species-level treatment of *C. acris*, another endemic taxon of the Balkan Peninsula, was proposed by Perný et al. (2004). Based on genetic and morphological differentiation, three subspecies were described. Previous phylogenetic studies have resolved this species close to the species from the *C. pratensis* group, together with three SW Asian entities, *C. tenera* J.G. Gmel. ex C.A. Mey., *C. seidlitziana* and *C. uliginosa* M. Bieb. (Marhold et al. 2004, Lihová et al. 2004a, Huang et al. 2020). These three taxa, sometimes referred to as subsect. Tenerae (Khatri 1988) or *C. tenera* group (Marhold et al. 2004), were known to be distributed in Türkiye and the Caucasus, with extensions to other parts of the Near East (Cullen 1965; Khatri 1988; Marhold et al. 2004). They were considered to be distinguishable by their morphological traits and parapatric occurrence, thought to be growing at different elevations, with *C. tenera* growing primarily in mountain foothills, *C. uliginosa* inhabiting the highest alpine zones and *C. seidlitziana* occurring at middle elevations between the previous two species. Such elevational segregation with parapatric distribution has been observed also within the *C. amara* complex (Lihová et al. 2004b; Zozomová-Lihová et al. 2015), but not among the members of the traditional *C. pratensis* and *C. raphanifolia* groups, which (especially their Mediterranean representatives) grow mostly allopatrically (Marhold et al. 2018). Despite the ecological and morphological differentiation of *C. tenera*, *C. seidlitziana* and *C. uliginosa*, these three entities were genetically intermingled in the previous molecular systematic studies, with no proper evidence supporting their separate taxonomic identity (Lihová et al. 2004a; Marhold et al. 2004; Carlsen et al. 2009), which strongly contrasts with the patterns observed in the European relatives. Unresolved relationships among these three taxa, but also to their relatives from Europe, was suggested to be caused either by incomplete speciation due to repeated range shifts and hybridization events (reticulate evolutionary history), extensive incomplete lineage sorting (retention of ancestral polymorphism in the examined genetic markers), or even combination of both processes (Lihová et al. 2004a; Marhold et al. 2004).

1.3 The focused area of Southwestern Asia and the Balkan Peninsula

The studied geographic area covers the Balkan Peninsula and the area of SW Asia, which includes Anatolia, the Caucasus, major mountain ranges in Iran (Zagros and Alborz), with extension to the Lebanon Mts.

A significant part of the focused area is located in the Eastern Mediterranean or has been under strong influence of this region for an extended period. The Mediterranean Basin is very rich in biodiversity and endemics and represents one of the largest global biodiversity hotspots (Myers et al. 2000; Mittermeier et al. 2011). According to Comes (2004), 80% of European plant endemics are found in the Mediterranean region. The distribution of biodiversity in the Mediterranean, however, is not homogeneous, but several regional hotspots, so called "hotspots-within-hotspots" can be recognized (Médail & Quézel 1997; Cañadas et al. 2014). The centres of plant diversity and endemism are located here mainly on islands and in the mountains (Médail & Quézel 1997; Thompson 2020). Médail & Quézel (1997) identified 10 regional hotspots in the Mediterranean Basin. Furthermore, a finer structure composed of even smaller local hotspots can be recognized within such larger regional units, which, however, can also be important for biodiversity conservation efforts (e.g. Kougioumoutzis et al. 2021). The Mediterranean Basin also played an important role in postglacial colonization of temperate regions in Europe and Asia (Hewitt 2004; Nieto Feliner 2014; Thompson 2020). Médail & Diadema (2009) identified 52 glacial refugia in the Mediterranean region, while 19 of them belong to its eastern part. The contemporary flora of the Mediterranean Basin is thought to have been largely influenced by the interaction of complex geological history, climate changes and human activities (Eastwood 2004; Nieto Feliner 2014; Thompson 2020). The climate of the region is generally characterized by hot and dry summers and mild and rainy winters (Djamali et al. 2012; Nieto Feliner 2014; Thompson 2020). Its current topography has been formed by the collision of the African and Eurasian tectonic plates, which started in the Cenozoic and led to the formation of peninsulas and the uplift of mountain ranges (Hewitt 2011; Nieto Feliner 2014). Additionally, the geological history of the eastern Mediterranean region was influenced by the collision of the Arabian and Eurasian plate in the Miocene (Hewitt 2011; Nieto Feliner 2014). During this epoch the area presently covering the Balkan Peninsula had been largely connected to Anatolia (Hewitt 2011). At the end of the Miocene, drift of the African plate obstructed the western end of the Mediterranean Sea, causing its epochal desiccation, referred to as the Messinian Salinity Crisis, which lasted until the end of the Miocene (Hsü 1972; Krijgsman et al. 2018). And even during the Pleistocene glacial-interglacial cycles, the recurrent marine regressions resulted in the land bridges between the continents (Kerey et al. 2004; Magyari et al. 2008). Due to the proximity of both SW Asia and the Balkan Peninsula to Africa, these areas have represented major land corridors connecting the Holarctic and Palaeotropical Kingdoms (Manafzadeh et al. 2014; Xu et al. 2020).

The Balkan Peninsula is the major biodiversity hotspot in the Mediterranean Basin (Stevanović et al. 2007; Hewitt 2011; Nieto Feliner 2014). The relatively stable and mild paleoenvironmental conditions in this region facilitated the long term survival of taxa, and the region harbours several relic taxa of the Tertiary origin (Tzedakis 2004; Peev & Delcheva 2007). The topographically diverse landscape of the Balkan Peninsula with its mosaic of high mountain ranges, islands and peninsulas is also one of the key factors triggering diversification and speciation, and resulted in high endemism rate (e.g. Peev & Delcheva 2007; Kučera et al. 2010; Olšavská et al. 2016; Španiel et al. 2017; López-González et al. 2021). Also, the proximity of the Balkan Peninsula to SW Asia may have likely contributed to its outstanding species diversity (Španiel & Rešetnik 2022). The Balkan Peninsula served as a central migration corridor for Asian biota during east–west colonization of Europe since the Early Oligocene (Hewitt 2011; Manafzadeh et al. 2014, Nieto Feliner 2014). In addition, this peninsula represents one of the source areas for postglacial colonization of central and northern Europe (Hewitt 2004, 2011; Nieto Feliner 2014; Kougioumoutzis et al. 2021, but see also Gömöry et al. 2020).

The area of SW Asia is highly heterogeneous and covers three phytogeographic regions: the Mediterranean region on Mediterranean Sea coast in the west, the Euro-Siberian in the Caucasus, while the major part is covered by the Irano-Turanian region (Djamali et al. 2012; Noroozi 2020). SW Asia encompasses three global biodiversity hotspots, the Mediterranean Basin, the Caucasus and the Irano-Anatolian (Myers et al. 2000; Mittermeier et al. 2011; Noroozi 2020). Likewise, as in the Mediterranean, a high biodiversity is concentrated here in the mountainous landscapes (Médail & Diadema 2009; Noroozi et al. 2018; Noroozi 2020; Thompson 2020). Mountains are characteristic for their heterogeneity of habitats and buffering effect on severe climate changes, providing favourable conditions for diversification and speciation processes (Noroozi et al. 2018; Muellner-Riehl et al. 2019; Noroozi 2020; Perrigo et al. 2020).

Despite the inclusion of the Caucasus into the Euro-Siberian region (or Circumboreal floristic region, following classification by Takhtajan 1986), its geological history suggests that its flora has been largely isolated since the Cenozoic era and evolved largely independently from the other parts of the Euro-Siberian region (Maharramova et al. 2018). Two major glacial refugia of the Tertiary flora are located in the Caucasus, specifically the Colchis and Hyrcan regions, which are presently covered by temperate rainforests, representing the oldest forests of Western Eurasia (Nakhutsrishvili et al. 2015; Maharramova et al. 2018). These refugia have

played a crucial role in sheltering numerous plants including trees during Pleistocene glaciations (e.g. Krebs et al. 2004; Tarkhnishvili et al. 2012; Maharramova et al. 2015, 2018; Volkova et al. 2020). Because of the mountainous character of the Caucasus and its position between two seas, the climate is mostly humid and relatively warm, therefore allowing the survival even of thermophilous hygrophytes (Denk et al. 2001; Nakhutsrishvili et al. 2015). On the other hand, the Transcaucasian Highlands in the southern part of the Caucasus act as a transitional area between the Euro-Siberian and the Irano-Turanian phytogeographical regions, and the majority of endemics are found there in the arid environment of steppes, open dry forests and semi-deserts (Fayvush & Aleksanyan 2020).

The Irano-Turanian region is one of the largest floristic regions in the world, covering about 30% of the Eurasian surface (Manafzadeh et al. 2017; Moharrek et al. 2019). However, exact delimitation of the biogeographical regions has been discussed and is largely dependent on authors' concept (Cox 2001; Djamali et al. 2012). This is also demonstrated by the variety of definitions, trying to outline the Irano-Turanian region in the past (Manafzadeh et al. 2017). Even though the region covers a large area and harbours a pronounced portion of Old World biodiversity, it has been lacking the greater scientific attention and remained understudied. Numerous national or regional Floras have been published from this region (e.g. Rechinger 1963; Cullen 1965), but the detailed knowledge on phylogeny, taxonomy and distribution of the listed taxa has been largely deficient (Xu et al. 2020). As mentioned above, this region encompasses the Irano-Anatolian biodiversity hotspot of global importance with smaller local hotspots distinguished within it. For example, Noroozi et al. (2018) identified five local hotspots only in Iran. In another study, multiple smaller hotspots were localized in Türkiye (Noroozi et al. 2019). The Irano-Turanian region is known as the centre of origin and source of numerous taxa for the adjacent areas (Manafzadeh et al. 2014, 2017; Noroozi 2020). The region is well defined by its specific continental climate, characterized by dry and hot summers, harsh winters and overall low precipitation (Djamali et al. 2012; Dönmez & Yerli 2018). The biogeographical patterns in the Irano-Turanian region have been influenced by its complex tectonic history, involving the collisions of the Eurasian plate with Indian and Arabian ones in the Paleogene (Manafzadeh et al. 2017; Moharrek et al. 2019). This led to the uplift of present mountain ranges and plateaus, consequently causing rain shadows and aridification in great part of the region (Manafzadeh et al. 2017).

From the studied areas, Anatolia is particularly interesting for its transitional position between Asia and Europe, comprising both Mediterranean and Irano-Turanian floristic

elements (Eastwood 2004; Ansell et al. 2011; Manafzadeh et al. 2017; Thompson 2020). There are several topographic features in Anatolia, which have possibly influenced the biogeographical patterns of extant flora by facilitating or hampering range expansion and leading to allopatric speciation (Bilgin 2011; Manafzadeh et al. 2014; Parolly 2020). Until the end of the Pliocene, the Central Anatolian Plateau, now arid and saline, had been covered by the Central Anatolian Lake system, thus representing barely inhabitable terrain for terrestrial organisms for millions of years (Bilgin 2011; Parolly 2020). The complexity of the Anatolian landscape is also elevated by occurrence of multiple volcanoes in the plateau (Dönmez & Yerli 2018). The overall distribution of Anatolian taxa is influenced by the presence of the significant biogeographical barrier named as the Anatolian Diagonal belt, which limits the range of hundreds of taxa on both of its sides, while also harbouring numerous endemics constrained to the belt itself (Ekim & Güner 1986; Gür 2016). Despite previous attempts to shed more light on distribution patterns associated with the Anatolian Diagonal Belt, the factors and mechanisms behind them are still not fully understood and explained (Manafzadeh et al. 2017; but see Gür 2016). It was hypothesized that the Anatolian Diagonal had played an important role in the postglacial recolonization of Europe, by limiting the migration of taxa at its eastern side (Bilgin 2011). On the other hand, the corridor referred to as the Taurus way, ranging from the Anatolian Diagonal through the Taurus to the mountains of the Balkan peninsula, has likely represented a vital dispersal route, especially for taxa adapted to colder conditions (Ansell et al. 2011; Bilgin 2011; Parolly 2020). Alternatively, the North Anatolian Mountains (Pontic Mountains), located in proximity of the Black Sea, might have played the same role for species adapted to more humid and warmer habitats (Kaya & Çıplak 2017; Özüdoğru & Mummenhoff 2020). While the biogeographical contact between the Balkan Peninsula and SW Asia is currently constrained by the Sea of Marmara, these areas have been recurrently interconnected throughout the Tertiary and Quaternary periods, allowing for the migration between the European and Asian continent (Bilgin 2011; Hewitt 2011; see also Ansell et al. 2011; Sobierajska et al. 2016).

The evolutionary significance of Anatolia has been highlighted in several recent studies, which identified this area as the centre of origin or diversity (Ansell et al. 2011; Koch et al. 2017; Özüdoğru & Mummenhoff 2020; Özüdoğru et al. 2022). For instance, it is also thought to represent the centre of origin of the Brassicaceae family (Franzke et al. 2011; Karl & Koch 2013; Koch et al. 2017; Huang et al. 2020). Moreover, Anatolia is also known as the source of postglacial colonization of Europe (Korkmaz et al. 2014; Ali et al. 2019), or as the dispersal

corridor between Europe and the rest of Asia (Jabbour & Renner 2011; Manafzadeh et al. 2014; Kaya & Çıplak 2017; Moharrek et al. 2019). For example, Özüdoğru & Mummenhoff (2020) resolved the evolutionary history of the disjunctively distributed genus *Bornmuellera* Hausskn., which apparently originated in Anatolia and dispersed into the Balkans before the formation of the Sakarya River Basin in the Plio-Pleistocene transition. Another Brassicaceae representative, *Microthlaspi perfoliatum* (L.) F.K. Mey, survived the Quaternary glaciations in five refugia ranging from the Mediterranean Basin to Central Asia, while the postglacial recolonization of Europe has taken place from Anatolia (Ali et al. 2019). The genus *Haplophyllum* A. Juss. originated in Central Asia, migrated through Anatolia and subsequently diversified in the Mediterranean Basin (Manafzadeh et al. 2014). On contrary, in *Juniperus drupacea* Labill., the disjunct populations in the Peloponnese, Anatolia and Lebanon were found to be genetically and morphologically distinct, suggesting their long-lasting isolation (Sobierajska et al. 2016).

1.4 Polyploidization and hybridization as the major drivers of speciation in plants and their significance in *Cardamine*

Speciation is a process in which populations evolve and differentiate into distinct species. It is usually accompanied by partial or full reproductive isolation among the sets of populations emerging as different species. One of the common classifications of speciation events is based on their geographic modes (Fitzpatrick et al. 2009; Safran & Nosil 2012). In allopatric speciation, limitation of gene flow between populations resulting in speciation is mediated by the presence of geographic barrier. However, in parapatric and sympatric speciation, the physical separation of populations representing newly emerging species is not existent or only partial, i.e. they may occur in adjacent regions with contact zones (in parapatry), or their distribution overlaps (in sympatry). In these scenarios, differentiation leading to speciation (usually accompanied by reproductive isolation) needs to be promoted by other processes, which may involve e.g. adaptation to a new environmental niche (ecological speciation; Nosil 2012), whole-genome duplication (polyploid speciation; Wood et al. 2009), or interspecific hybridization (hybrid speciation; Mallet 2007). These mechanisms may often operate simultaneously and their effects may be complementary. However, outlining the boundaries of species may be often markedly difficult in many groups, due to the fact that not all species are necessarily held together by gene flow and isolated from other species by reproductive barriers (Rieseberg & Willis 2007). Additionally, methods available in modern

plant systematics, despite their rapid advance in recent decades, are not omnipotent and are still not able to unambiguously resolve all problems with inference of phylogenetic relationships, mainly in the groups of closely related species. In plants, the common causes of such problems, which are also associated with taxonomic confusion, include: reticulate evolutionary history, recent and rapid diversification, and presence of incomplete lineage sorting. However, recent phylogenomic approaches based on next-generation sequencing have been found to be particularly efficient in resolving even very complex evolutionary histories (e.g. Larridon et al. 2020; Karbstein et al. 2020; Reichelt et al. 2021).

Speciation may be affected by a wide spectrum of mechanisms and factors, which are the core subject of the field of evolutionary biology. In plants, polyploidization and hybridization are considered to be the most common drivers of diversification and sympatric speciation (Soltis & Soltis 2009; Soltis et al. 2010, 2014). The evolutionary significance of polyploidization (whole-genome duplication, WGD) in angiosperms has been discussed multiple times (Otto & Whitton 2000; Soltis et al. 2003, 2010; Jiao et al. 2011; Nieto Feliner et al. 2020). However, the direct impact of WGD events on diversification rate shift has been questioned (Soltis & Soltis 2009; Wood et al. 2009; Mayrose et al. 2011; Arrigo & Barker 2012; Soltis et al., 2014; Kellogg 2016; Ren et al. 2018; Huang et al. 2020). WGD events impacted angiosperm phylogeny during all stages of their evolutionary history, including the earliest diverging lineages (Soltis et al. 2009; Tank et al. 2015; Landis et al. 2018). Indeed, it is believed that all of the currently recognized lineages have undergone multiple polyploidization events (Soltis et al. 2009; Jiao et al. 2011; Arrigo & Barker 2012; Soltis & Soltis 2016; Landis et al. 2018; Nieto Feliner et al. 2020). From the temporal aspect of their origin and the stage of diploidization, polyploids can be classified into three categories: neopolyploid, mesopolyploid and paleopolyploid (Mandáková et al. 2010). Paleopolyploid events are the most difficult to trace since they were often accompanied by significant chromosomal rearrangements affecting the entire genome architecture (Mandáková & Lysák 2018; Huang et al. 2020).

Based on their origins, polyploids have been traditionally categorized into two groups: autopolyploids that originate by genome doubling within a single species and encompass multiple sets of homologous genomes, and allopolyploids that evolve through interspecific hybridization and possess sets of divergent, nonhomologous genomes (Parisod et al. 2010; Spoelhof et al. 2017). Anyway, it is now widely accepted that chromosome pairing process may be also affected by other factors and the genomes merged in polyploids may, in fact, encompass a gradient of divergence (Parisod et al. 2010). Due to high homology of genomes, polysomic

inheritance and genetic redundancy are the features typically associated with autopolyploids and were proposed as diagnostic traits to distinguish them from allopolyploids (Flagel & Wendel 2009; Parisod et al. 2010; Spoelhof et al. 2017). Traditionally, autopolyploids were thought to be generally rare and less abundant than allopolyploids, which was questioned or even disproved in the latest studies, suggesting that they have been probably overlooked because of their usual morphological similarity to their diploid progenitors (Soltis & Soltis 2009; Segraves 2017; Spoelhof et al. 2017). In fact, Ramsey & Schemske (1998) estimated the rate of autopolyploid formation to be higher than for allopolyploids.

Although there are multiple scenarios leading to the formation of polyploids (reviewed by Ramsey & Schemske 1998; De Storme & Geelen 2013), they always involve either the somatic polyploidization or production of unreduced gametes. Somatic polyploidization (endopolyploidization) is the process of genome doubling within a somatic tissue, induced by incomplete cell-division cycle (Leitch & Dodsworth 2017). In contrast to polyploidization originated by fusion of unreduced gametes mentioned below, endopolyploidization leads to the formation of an individual whose body tissues consist of a mixture of several ploidy level cells. Nevertheless, formation of new polyploid lineages by somatic polyploidization is a rare event (Ramsey & Schemske 1998; Madlung 2013; Tayalé & Parisod 2013). Origin of polyploids is thought to be more frequently caused by formation of unreduced gametes (Parisod et al. 2010; Madlung 2013; Mason & Pires 2015). Unreduced gametes are formed during aberrations in the process of micro- or megasporogenesis and represent gametes with somatic (2n) chromosome number (De Storme & Geelen 2013). Formation of a tetraploid progeny (and subsequently higher polyploids) may take place in one-step, instantaneously, by fusion of two unreduced gametes (bilateral pathway). Alternatively, tetraploid formation may occur through a "triploid bridge", which involves fusion of an unreduced gamete with a reduced one (unilateral pathway), resulting in the production of at least partially fertile triploid individual. By selfing or backcrossing to its diploid progenitor, new tetraploids may be formed (Ramsey & Schemske 1998; Husband 2004; Parisod et al. 2010; Kolář et al. 2017). However, this process may be halted, as triploids were often found to be sterile or have significantly reduced fertility (which is referred to as "triploid block"; Husband 2004; Kolář et al. 2017; Spoelhof et al. 2017). Odd-ploidy level individuals with reduced fertility are often locally surviving with aid of asexual reproduction, including either vegetative propagation or apomixis (Kolář et al. 2017). Ramsey & Schemske (1998) estimated the frequency of formation of unreduced gametes to about 0.5% for non-hybrid systems, while it is known to be markedly higher in hybrids (Zhang et al. 2010;

Tayalé & Parisod 2013). However, the rate of unreduced gametes production is also influenced by environmental stress and disturbances and is apparently lineage-specific, being governed by several genes (Ramsey & Schemske 1998; Parisod et al. 2010; Mason & Pires 2015; Van de Peer et al. 2020).

After the polyploid formation, establishment of novel polyploid cytotypes depends on a very complex interplay of environmental and genetic factors, which has been a subject of studies for decades. Polyploids (especially allopolyploids) harbour greater genetic variation than their diploid progenitors and are characterized by complex and dynamic genomes that may undergo structural rearrangements and differential gene expression patterns (Soltis et al. 2014; Soltis & Soltis 2016; Ren et al. 2018; Nieto Feliner et al. 2020). This potentially results in various morphological and physiological changes, which have been repeatedly studied, both comparing naturally occurring polyploids (e.g. Segraves & Thompson 1999; Chansler et al. 2016; Bomblies 2020; Clo & Kolář 2021) or experimentally induced autopolyploids with their progenitors (e.g. Maherali et al. 2009; Aqafarini et al. 2019; Porturas et al. 2019; Mo et al. 2020). Consequently, these changes may have an impact on interactions of polyploids with pollinators, predators or pests (Segraves & Anneberg 2016; Segraves 2017; Rezende et al. 2020). As a result of the increased genome flexibility, polyploids may become successful colonizers that expand to larger distribution areas and/or adapt to different environmental conditions than their progenitors (e.g. Fawcett et al. 2009; Godfree et al. 2017; Baduel et al. 2018; also reported in *Cardamine*, Shimizu-Inatsugi et al. 2017; Akiyama et al. 2021). Ecological niche diversity favours the establishment and persistence of polyploid cytotypes or hybrids, which may finally result in the evolution of a new polyploid species (Parisod et al. 2010; Madlung 2013; Van de Peer et al. 2020). For example, the ploidy dependent colonization of different niches along the altitudinal gradient has been reported in various plant species groups (Schönswetter et al. 2007a; Mráz et al. 2008; Zozomová-Lihová et al. 2015). Despite numerous studies addressed changes induced by polyploidization as well as factors influencing polyploid establishment, interpreting the actual impact of genome-doubling by itself on adaptive function is often difficult, as it cannot be convincingly separated from other processes accompanying it. Obviously, this is even more complicated in allopolyploids. Moreover, structural and functional changes at the genome level, including its adaptive consequences, may be first manifested many generations after the WGD event (Schranz et al. 2012; Tank et al. 2015; Baduel et al. 2018; Bomblies 2020; Nieto Feliner et al. 2020).

Whole-genome duplication events played an important role also in the evolution of the Brassicaceae family. Three major WGD events were identified based on the genome sequencing of *Arabidopsis thaliana* (L.) Heynh., and labelled as At-α, At-β and At-γ (Franzke et al. 2011). The most recent, At-α-duplication (Schranz et al. 2012), took place at the turn of the Eocene and Oligocene epochs, ca. 50 mya (Huang et al. 2020) and was followed by several mesopolyploidization events within specific tribes of the family (Mandáková et al. 2017; Huang et al. 2020). The fundamental role of recurrent polyploidization events also acting during later phases of Brassicaceae evolution is evidenced by the high number of neopolyploid taxa reported. Warwick & Al-Shehbaz (2006) suggested that 37% of Brassicaceae taxa are of neopolyploid origin, while Hohmann et al. (2015) proposed to increase this number to 43 %.

In *Cardamine*, more than half of the taxa are thought to be neopolyploid (Kučera et al. 2005). Detailed attention has been paid to the investigation of a few polyploids proven to be of allopolyploid origin (*C. silana* Marhold et Perný, *C. asarifolia, C. flexuosa, C. occulta*; Perný et al. 2005a; Lihová et al. 2006; Mandáková et al. 2014, 2019), nevertheless, the origin of most polyploids in *Cardamine* remains unresolved. Some cases of autopolyploidy were documented as well, resulting in the formation of a new species or subspecies. For instance, *C. majovskyi* is an autotetraploid derived from *C. matthioli*, which evolved at least twice independently (Melichárková et al. 2020). Another example is *C. amara* subsp. *austriaca* Marhold, the autotetraploid derived from the nominate subspecies (Zozomová-Lihová et al. 2015). *C. cordifolia* A. Gray, widely distributed in North America, possesses triploid-like chromosome number (2n = 24). Detailed cytogenetical analyses revealed that it is originally a tetraploid that underwent multiple chromosome translocations, significantly reducing its chromosome number (Mandáková et al. 2016).

Another evolutionary significant phenomenon, apart from polyploidization, is homoploid hybridization (hybridization at the same ploidy level). It is a less frequent event, which does not involve genome doubling, but may also lead to speciation, termed as homoploid hybrid speciation (Mallet 2007; Soltis & Soltis 2009; Vallejo-Marín & Hiscock 2016; Nieto Feliner et al. 2017, 2020). The detection of homoploid hybrids is often more challenging since the hybrid or even introgressed individuals might be masked by the overall high phenotypic and ecological variation range of their progenitors (Mallet 2007; Soltis & Soltis 2009). According to Mallet (2005), 25 % of currently known vascular plant species are involved in the hybridization or introgression with their relatives. Similarly, as in polyploidization, hybridization by itself may act as a source of novel genetic structures and adaptive functions,

maintaining complex evolutionary patterns, potentially including subsequent adaptive radiation (Pease et al. 2016; Nieto Feliner et al. 2017, 2020; Schmickl et al. 2017). However, hybridization does not necessarily result in the formation of new hybrid taxa (i.e. speciation), but may result in a variety of possible outcomes, such as occasional gene flow between two related taxa (introgression), origin of hybrid swarms, or even causing local extinction of one or both hybridizing species (Soltis & Soltis 2009; Vallejo-Marín & Hiscock 2016; Nieto Feliner et al. 2020). Due to involved recombination and the effect of transgressive variation, hybrids may acquire extreme phenotypes, which are not in the continuum of the variability observable between their parents, and consequently, may facilitate the colonization of new ecological niches (Mallet 2007).

Interspecific hybridization (with or without ploidy level increase) is also a commonly observed phenomenon in *Cardamine*, which results in diverse patterns and outcomes. These could be sorted into three categories. First, local sterile hybrids, presumably representing F1 generation, have been reported, which persist by vegetative propagation and recurrent origins, but have no evolutionary significance. These originated by both homoploid (*C.* ×*enriquei* Marhold, Lihová et Perný; Marhold et al. 2002) and heteroploid (*C.* ×*ferrarii* Burnat, *C.* ×*paxiana* O.E. Schulz; Lihová et al. 2006, 2007a) hybridization events. Next, partially fertile triploid hybrids were observed (*C.* ×*rhodopaea* Ančev, *C.* ×*insueta* Urbanska-Worytkiewicz; Ančev et al. 2013; Mandáková et al. 2013), possibly participating in backcrossing. In regard of *C.* ×*insueta*, this taxon was also involved in the particularly complex evolutionary scenario found in the formation of the tri-genomic allopolyploid *C. schulzii* (Mandáková et al. 2013). Finally, the last category of *Cardamine* hybrids comprises reports of polyploid hybridization, in which hybrids retained their ability of sexual reproduction, allowing introgression among them and their parents, consequently forming heterogeneous hybrid swarms (*C. pratensis* × *C. raphanifolia*; Lihová et al. 2007b). Other reports of polyploidization and hybridization in the genus are reviewed by Marhold et al. (2018), but the origin of most of the recognized taxa has not yet been resolved.

2. AIMS OF THE DISSERTATION THESIS

In this thesis, two main topic areas outlined below are addressed. For both, we list specific hypotheses that have been tested, and the reference to the respective studies in which they were primarily resolved.

A) Taxonomy and phylogeny of the target species complexes with the focus laid on poorly explored regions of southwestern Asia and Balkan Peninsula

Hypotheses:

- Traditional recognition of three species complexes is in conflict with the phylogenetic patterns and evolutionary history of the studied species (chapters 4.1 and 4.2)

- In depth phylogenomic data and integrative approach will reveal greater species diversity in SW Asia than previously reported (chapters 4.2 and 4.3)

- Anatolian diversity hotspots and refugia have had limited impact on European diversity - 'out of Anatolia' scenario is not supported in the target species (chapters 4.2 and 4.3)

B) Significance of reticulate evolution and polyploid speciation

Hypotheses:

- Evolutionary history of SW Asian species has been significantly and recurrently affected by interspecific gene flow (chapter 4.2)

- Polyploid speciation has played a minor evolutionary role in SW Asia and Balkans in contrast to other European areas (chapters 4.1 and 4.2)

- Mountain ranges have acted not just as refugia but also as melting pots favouring secondary contacts, admixture and hybridization (chapters 4.1, 4.2, 4.3)

3. METHODS

To address the hypotheses stated above, combinations of morphometric, karyological, molecular and cytogenetic methods currently used in the field of plant systematics and phylogenomics were used, along with the ecological niche analysis for assessment of ecological divergence among studied species (Fig. 2). Standard multivariate morphometric methods were applied to study patterns of morphological variation and differentiation. Flow cytometry was used for estimation of the ploidy levels and relative genome sizes and was accompanied by chromosome counting. The evolutionary history and phylogenetic relationships were analysed using next-generation sequencing method Hyb-Seq, while conventional markers based on Sanger sequencing were also used in specific parts. Genomic in situ hybridization was used as a complementary method for verification of the allopolyploid status of one of the studied species.

3.1 Chromosome counting

Direct chromosome counting is a traditionally used approach in plant sciences to unambiguously identify chromosome number and ploidy level of the studied plant species. It also serves as an indispensable standard for ploidy level estimation using flow cytometry (see below). Chromosome number counts were obtained using spread plates of metaphase mitotic cells from root meristems. Root tips were usually taken from cultivated plants originally sampled in the field. Spread plates preparation followed the protocol published by Marhold et al. (2002) or Kučera et al. (2010). Chromosomes were stained with the Giemsa solution and counted from the temporary slides, made by the squashing method.

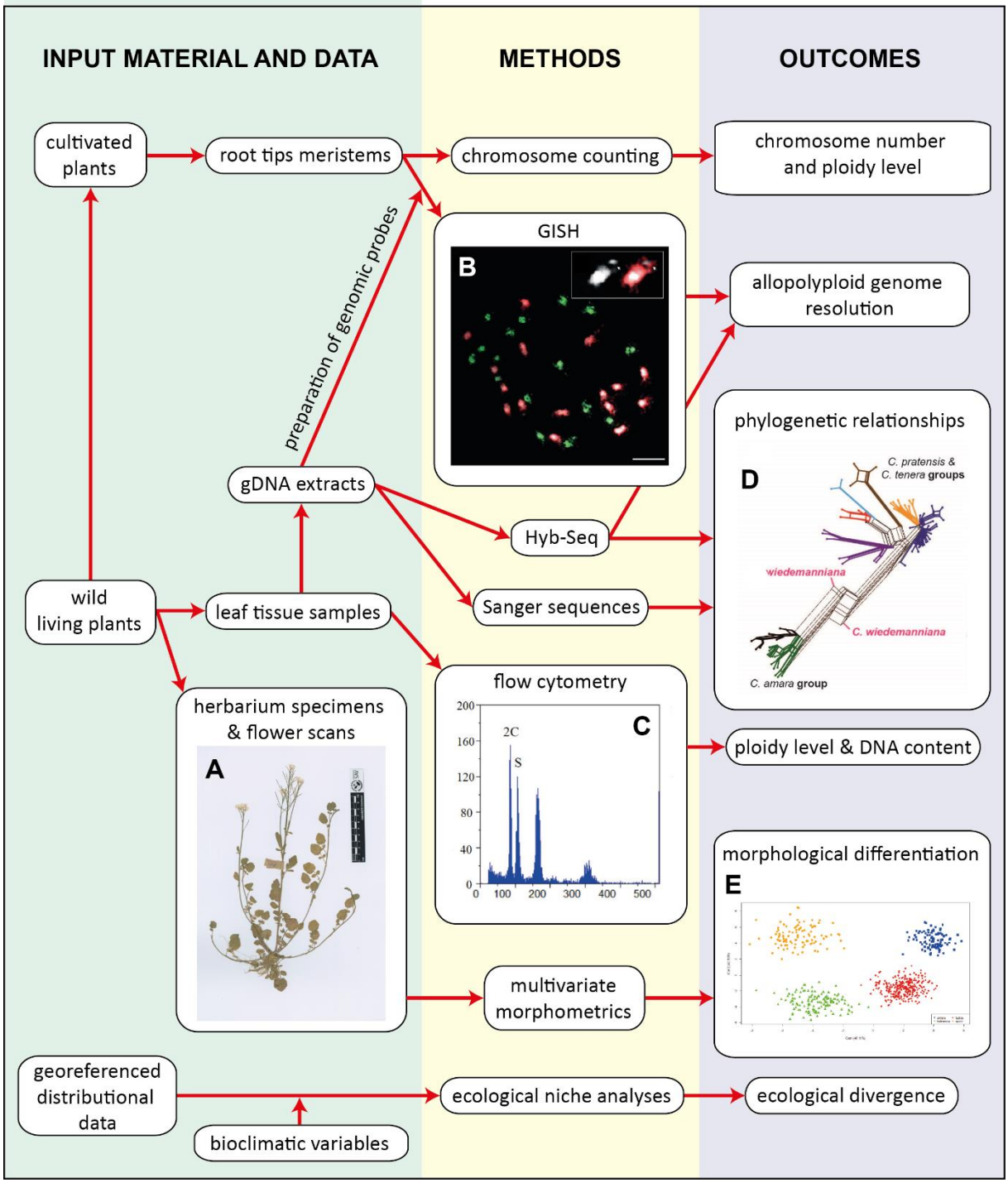


Fig. 2. Scheme of the methodological approaches used in the thesis. From left to right, the sections show: the sources and type of input material or data; the applied methods; and the outcomes (i.e., utilization of the interpreted data) achieved by individual methods. Illustration pictures display examples of: a herbarium specimen of *C. tenera* (A); GISH used in the article presented in the chapter 4.1, done by Terezie Mandáková (B); histogram obtained by flow cytometry (C); SuperQ network based on Hyb-Seq data, used in the article presented in the chapter 4.2 (D); and outcome of the canonical discriminant analysis used for estimation of the extent of morphological differentiation (E).

3.2 Flow cytometry

Flow cytometry is used to determine the ploidy level of analysed plants, as well as to measure their genome size. Both relative and absolute nuclear DNA content was measured, using either DAPI (4',6-diamidino-2-phenylindole; AT-selective), or propidium-iodide (PI; intercalating) fluorochromes, respectively (Doležel et al. 2007). Whereas PI measurements required fresh tissue from living plants, DAPI measurements were done from silica gel-dried leaf tissue, as it provides sufficient resolution and accuracy for relative genome size estimation (Suda & Trávníček 2006). The sample preparation followed the published protocols (Doležel et al. 2007; Marhold et al. 2010) using Otto buffers (Otto 1990). Internal standards with the known and stable genome size values were analysed simultaneously with the measured samples. Fluorescent intensity of stained nuclei was measured using Partec CyFlow ML flow cytometer (Partec GmbH, Münster, Germany) with a UV LED lamp for DAPI measurements, or Partec CyFlow SL cytometer with a green solid-state laser for PI measurements. The resulting histograms were analysed using the Partec FloMax software (v2.7; Partec GmbH Münster, Germany). The relative nuclear DNA content (expressed in arbitrary units) represented the ratio of positions of the G1 peaks of the sample and the standard. Absolute nuclear DNA content was expressed in pg and was calculated from the G1 peaks ratio and the known 2C genome size value of the standard. The coefficients of variation (CV) were calculated for both the standard and sample peaks, and only histograms with CV values below the 5% threshold were used for nuclear DNA content calculations. The selection of proper internal standard was important for getting precise measurements, as the genome size should not differ significantly from the genome size of the analysed sample (Doležel et al. 2007). *Solanum pseudocapsicum* L. (2C = 2.59 pg; Temsch et al. 2010) and *Bellis perennis* L. (2C = 3.38 pg; Schönswetter et al. 2007b) were used as internal standards. The availability of an alternative internal standard during the analysis is recommended (Doležel et al. 2007), as in the case of overlapping peaks of the sample and the standard, the sample needs to be re-measured with another standard.

3.3 Morphometric analyses

To study the morphological variability and differences among the analysed taxa, methods of standard multivariate morphometrics were used (Marhold 2011). The measured morphological characters followed the characters reported as important in previous taxonomic

studies, floras or determination keys, and were also selected based on field observations. Characters on vegetative organs were measured from the herbarium specimens of the plants sampled in the field and stored in the herbarium SAV. For measuring characters on flowers, the separated floral parts of one flower per individual were attached to plain paper by a transparent adhesive tape, immediately after collection, to preserve its shape after drying. The attached floral parts were then scanned and the measurements were performed using a specialized software (QuickPHOTO Industrial v.2.3, Promicra, Prague, Czechia). The matrices of morphological characters were used for computation of basic descriptive statistics, including minimum and maximum value, arithmetic mean, median and standard deviation. Before morphometric analyses, Spearman (non-parametric) correlation coefficients (Legendre & Legendre 1998) were computed to uncover pairs of highly correlated traits ($\rho > 0.9$) which could distort the computations, especially the discriminant analyses. The datasets were subsequently analysed using hypothesis-generating and hypothesis-testing multivariate methods. The first group encompassed methods enabling first insight into the overall variation in the datasets in general including the principal component analysis (PCA) and cluster analyses. In PCA, the overall patterns of morphological variability among the measured individuals (or populations) along the first two or three axes (principal components), expressing the highest variability, were identified (Sneath & Sokal 1973; Krzanowski 1990). Cluster analyses (Podani 2000) displayed hierarchical clustering based on the matrix of Euclidean distances among the objects, and reveal groups of morphologically similar individuals (or populations). The second group included methods used for testing the patterns observed in the previous steps or between the predefined groups, which were based on the results of molecular analyses, karyology and geographic distribution. The canonical discriminant analyses (CDA) were used to evaluate the extent of morphological differentiation between the pre-defined groups by weighting the characters (Klecka 1980; Krzanowski 1990). By this method, morphological characters with the greatest diagnostical potential were identified. All morphometric analyses were performed in R 4.0.0 software (R Core Team 2022) using a recently developed MorphoTools2 package (Šlenker et al. 2022).

3.4 Ecological niche analyses

In one study included in the thesis (see chapter 4.2), we used ecological niche analyses to determine the extent of ecological differentiation among four species of the genus *Cardamine* distributed in SW Asia (including *C. penzesii*, which is distributed also in eastern parts of

Bulgaria). In this part, we aimed to analyze if and how much do ecological niches of individual species overlap, and we used these findings to estimate if the ecological divergence may have played a role in speciation within the target groups. For this purpose, following procedure was performed. We georeferenced the distributional data of the four target species, which were gathered during revision of herbarium specimens and literature research, as well as based on our own field sampling. For all georeferenced sites, we acquired environmental data represented by 22 environmental variables including information about climate (obtained from WorldClim database; Fick & Hijmans 2017), elevation, slope and aspect of the sites (Jarvis et al. 2008). For each analyzed species, we delineated and visualized the environmental space available for colonization, using the approach used for quantification of environmental niches as described by Broennimann et al. (2012). The PCA was applied to display a correlation matrix of environmental variables in the space defined by first two principal components. In the next step, we analyzed the niche overlap of the species using Schoener's D index, calculating the niche equivalency and niche similarity (Warren et al. 2008), and finally comparing the ecological niche breadths of the four species. Individual analyses were performed in R 4.0.0 software (R Core Team 2022) and its detailed explanations are available in the supplementary material of the article presented in chapter 4.2. This approach was applied in collaboration with Marek Svitok, who did the analyses and produced resulting figures, while I contributed to the preparation of input distributional data and interpretation of results.

3.5 Molecular analyses

3.5.1 Conventional molecular markers based on single-locus sequencing of chloroplast DNA regions (cpDNA), ITS of nuclear ribosomal (nrDNA) and low-copy nuclear genes

Traditional molecular markers based on PCR amplifications of a single DNA locus and subsequent Sanger sequencing were used in the thesis only in a small extent. They were used to verify and optimize the alignments and allele phasing and sorting procedures based on the sequence reads obtained from Hyb-Seq.

Chloroplast DNA (cpDNA) is a circular molecule comprising four partitions, large single copy (LSC), short single copy (SSC) and two inverted regions (IRa and IRb). It is characteristic with its uniparental (in angiosperms mostly maternal) inheritance and relatively

low substitution rate, with the usual length of 120-160 kbp (Wicke et al. 2011). Some of the frequently sequenced regions of cpDNA (used as molecular markers) include coding genes (e.g. *rbc*L, *ndh*F, *mat*K), as well as non-coding introns and spacers (e.g. *trn*L, *atp*B-*rbc*L, *rpo*B-*trn*C; Shaw et al. 2005, 2007; Dong et al. 2012). Generally, the chloroplast genome is highly conserved, which makes the use of cpDNA markers less applicable for studies at the population level and is often used in phylogenetic studies at higher taxonomic levels. However, the non-coding regions (introns and intergenic spacers) are known for the increased rate of both single-site mutations and indels accumulation and are applicable in systematics of closely related species. Due to their properties, cpDNA markers have some specific uses, for example in the studies of phylogeography and gene flow (dispersal by seeds) and in identification of maternal progenitors of hybridogenous taxa. Several cpDNA markers, specifically the *trn*L intron, *trn*L-*trn*F, *rpl*32-*trn*L (Lihová et al. 2004a; Kučera et al. 2010; Lihová et al. 2010), and *rpo*B-*trn*C (Mandáková et al. 2013) spacer regions were previously used in *Cardamine*.

The ITS region, comprising ITS1 and ITS2 internal transcribed spacers and 5.8S gene, is one of the most frequently used molecular markers in plant systematic and phylogenetic studies (Álvarez & Wendel 2003; Nieto Feliner & Rosselló 2007). The region is located between the rDNA genes coding rRNA subunits. In eukaryotic cells, ITS1 is present between the genes coding 18S and 5.8S subunit, while ITS2 is between the genes 5.8S and 26S. Together they form a transcription unit, which is accompanied by other non-coding spacers. These regions are typically found in thousands of tandem repeats. Because of its great abundance in genome, rDNA can be usually easily amplified and sequenced. The ITS spacers, as well as rDNA genes, are heavily influenced by concerted evolution, causing the genetic homogeneity among different repeat units (Álvarez & Wendel 2003). Despite this process, several variants of ITS can usually be identified in the genome, which is commonly observed in hybrids and polyploids. The rate of concerted evolution is not equal for different taxa, and the stage of its completeness influences the variability of the observed intra-individual ITS variants. The intra-individual divergence of ITS sequences has been often marked by the ambiguity codes, which, however, may not have favourable effect in phylogenetic inferences (Nieto Feliner & Roselló 2007). Additionally, the application of ITS region as a molecular marker has several other disadvantages, thoroughly discussed by Álvarez & Wendel (2003) and Nieto Feliner & Rosselló (2007). The quality of phylogenetic reconstruction based on ITS markers may be negatively influenced by incomplete concerted evolution, presence of paralogs, pseudogenes and its specific secondary structure that might have impact on mutation rates, violating the assumptions

of neutrality and independence of characters. On the other hand, amplification of the ITS region, eventually involving rDNA genes, can be made with the use of universal primers (White et al. 1990), and along with its versatile use, these are the factors keeping its popularity as a molecular marker (Álvarez & Wendel 2003; Nieto Feliner & Rosselló 2007). Indeed, it is considered to be one of the barcoding regions in the plant kingdom (Li et al. 2011; see e.g. Chen et al. 2010; Zhao et al. 2018), and it has been repeatedly used also in phylogenetic and evolutionary studies of *Cardamine* (Lihová et al. 2004a, 2006; Marhold et al. 2004; Kučera et al. 2010).

Lastly, also the single or low-copy nuclear genes were implemented as molecular markers and have been recently favoured. They represent specific coding regions, which are present in a single or a few copies in the genome. Unlike cpDNA and ITS, they show biparental inheritance and are usually less affected by concerted evolution (Small et al. 2004). They often include conserved exons and more variable introns, which makes them applicable to a broad range of phylogenetic levels. The typical use of the low-copy nuclear markers is in phylogenetic studies, especially at lower taxonomic levels (e.g. Kučera et al. 2010 in the *Cardamine maritima* species complex), but can be also useful for identification of parental genomes in taxa originated by hybridization and polyploidization (e.g. Lihová et al. 2006 in *C. asarifolia*). However, their use may be complicated because of the lack of universal primers and as they are generally more difficult to amplify in comparison with more abundant regions of the genome (Small et al. 2004). Another disadvantage of nuclear genes is their tendency to occur in gene families, which means that also paralogous or pseudogene copies may persist in the genome. Such non-orthologous gene copies may strongly distort phylogenetic inference. The need of simultaneous use of multiple independent nuclear genes has been emphasized, as the individual gene trees usually do not reflect the actual phylogeny and may be discordant between each other (Zimmer & Wen 2012; Naciri & Linder 2015). The incongruence between the gene trees may be caused by the stochastic segregation of polymorphic ancestor's alleles into its descendants (i.e. incomplete lineage sorting) but also by other processes such as hybridization (Naciri & Linder 2015; also see e.g. Ekenäs et al. 2012; Krak et al. 2013).

The following procedure was implemented to obtain the sequences of either cpDNA, ITS or low-copy nuclear gene regions. Genomic DNA was isolated from the silica gel-dried leaf tissue using commercial DNA extraction kit. PCR amplification of the focused region followed, using the corresponding pair of primers. PCR products were purified and submitted for Sanger sequencing to commercial sequencing centres. If the intra-individual polymorphisms (originating from different alleles, ITS copy variants or homeologs in allopolyploids) were

observed in the sequences obtained by direct sequencing, molecular cloning was used for the separation of different variants present in the PCR products. The detailed protocols of PCR and molecular cloning published by Melichárková et al. (2017, 2019) were followed.

The obtained sequences were trimmed of primers to only contain the sequences of the focused regions, and these were aligned using the Geneious software v. R10 (Kearse et al. 2012). The phylogenetic patterns in the datasets were inferred using a variety of methods. Phylogenetic trees allowed the visualization of relationships among the studied individuals and species, while network-based methods revealed potential conflicts within and among the sequences and were useful to depict reticulate evolutionary patterns. Phylogenetic trees were computed using either maximum-likelihood or Bayesian approach, constructed with use of GARLI v2.01 (Zwickl 2006) or MrBayes v3.2.6 (Huelsenbeck & Ronquist 2001) softwares, respectively. Network-based approaches involved NeighborNet analysis, computed in the software SplitsTree4 v.4.14.4 (Huson & Bryant 2006) and parsimony-based TCS analysis (Clement et al. 2000) performed using the program PopART (Leigh & Bryant 2015).

3.5.2 Next-generation sequencing (NGS)

The molecular markers introduced in previous chapter were all based on Sanger sequencing (Sanger et al. 1977), which had represented the most widely used method of DNA sequencing for about 40 years. Recently, these approaches have started being increasingly substituted by methods based on NGS.

The term "next-generation sequencing" comprises a number of techniques of high-throughput parallel DNA sequencing. It utilizes several sequencing platforms, Illumina being generally the most widely used one. Its procedure has been summarized and explained in details by van Dijk et al. (2014) and Goodwin et al. (2016). Because of its limitation to generation of short reads, the other approaches (e.g. PacBio, Oxford Nanopore, BGISEQ) or their combinations are sometimes preferred too (van Dijk et al. 2014; Li & Harkess 2018). The principle of NGS techniques is somewhat similar to Sanger sequencing, yet, the most crucial difference is that thousands of loci are sequenced simultaneously in NGS, as opposed to obtaining the single-locus sequence of an individual DNA strand in the process of Sanger sequencing (van Dijk et al. 2014). Moreover, pooled samples from multiple individuals may be sequenced by NGS at one time, further increasing the effectivity of this approach.

In relatively short time of its advancement, scientists came up with a remarkable spectrum of NGS-based methods and its modifications, having diverse applications in various fields. For example, it led to development of the field of metagenomics, which focuses on sequencing of genetic material obtained from environmental samples (Thomas et al. 2012). RNA sequencing was found as a novel technique used for analyzing and detailed characterization of a cellular transcriptome (Wang et al. 2009). Regarding specifically the topic of evolutionary history of plants species, NGS-based approaches were found to be extraordinarily effective for phylogenomic and phylogeographic reconstructions, utilizing methods such as target enrichment, genome skimming, Hyb-Seq, restriction site associated DNA sequencing (RADseq) and whole genome resequencing. Target enrichment aims to capture hundreds to thousands of low-copy nuclear loci (exons) from the analyzed genome. Genome skimming represents a low coverage sequencing of the full genome, primarily obtaining the sequences of highly abundant multiple-copy regions such as ribosomal DNA (rDNA) and plastome (cpDNA). Hyb-Seq, being discussed in more details below, has been developed as a versatile method combining the previous two approaches. RADseq reduces the complexity of plant genomes, potentially producing tens of thousands of informative markers in the form of single nucleotide polymorphisms (SNPs; Baird et al. 2008). Its advantage also lies in its applicability in studies of taxa with no or limited knowledge about their phylogenetic position and relationships. Whole genome resequencing consists of assembling of the fragmentated genome into full genome sequence, utilizing the approach of mapping of the unknown sequences onto the available reference genome. Such data potentially provide thousands of nuclear markers applicable in answering evolutionary and ecological questions, providing insight into the evolution of genomic structures, unravelling roles of transposable elements and genomic traits responsible for adaptation, and a vast number of other applications.

In studies included in this thesis, we used Hyb-Seq as a core method employed for addressing the specified aims. We chose specifically this technique because of its high versatility and informativeness (Dodsworth et al. 2019). Compared to RADseq, it provides more details about specific loci, even though it does not cover as large portion of the genome. Importantly, Hyb-Seq works efficiently with the input DNA of much lower quality and quantity in contrast to RADseq, which was utilized by application of target enrichment in studies using samples extracted from herbarium specimens (Hart et al. 2016; Villaverde et al. 2018). In our case, it allowed us to use DNA extracts from silicagel-dried plant tissue samples collected even several years ago. Furthermore, Hyb-Seq already proved to be a powerful tool for

reconstruction of evolutionary histories in *Cardamine* in the previous study by Melichárková et al. (2020). Indeed, obtaining the whole genome sequences of the studied species would be applicable for the same purpose and would enable us to address even more derived and complicated questions, however, such procedure would be significantly more demanding in terms of time and resources costs, as well as data procession complexity.

3.5.3 Hyb-Seq

As mentioned above, low-copy nuclear genes have been preferred in phylogenetic studies because of their biparental inheritance and a high number of loci (Small et al. 2004), potentially serving as multiple independent markers (in contrast to cpDNA and nrDNA). However, due to their relatively costly application involving molecular cloning and lack of universal primers, only sequences from one or a few genes have been usually used in previous studies (Zimmer & Wen 2012). Still, the employment of multiple genes has been recommended (Naciri & Linder 2015) because of the commonly observed gene tree incongruence caused by incomplete lineage sorting, gene duplication (paralogy), gene loss or introgression. As previously mentioned, Hyb-Seq combines target enrichment with genome skimming (Weitemier et al. 2014; Dodsworth et al. 2019), which typically allows it to retrieve hundreds to thousands of exons with flanking intergenic and intronic regions, as well as the multiple-copy regions such as rDNA and cpDNA. Therefore, this method can substitute the use of the traditionally applied low-copy nuclear genes, ITS and cpDNA molecular markers with much higher efficiency, covering a number of genes that would not be technically possible to achieve with Sanger sequencing. Advantages of Hyb-Seq were already demonstrated by a number of studies that applied it for resolving the phylogenetic relationships within diverse genera or species groups with very complex evolutionary histories (e.g. Villaverde et al. 2018, 2020; Carter et al. 2019; Herrando-Moraira et al. 2019; Španiel et al. 2023).

Recent challenges in procession of NGS-generated data, being addressed also in this thesis, concerned mainly their usage for resolving evolutionary histories involving reticulation and polyploidy (Rothfels 2021). Various authors came up with novel insights and approaches that might be utilized in elucidating of evolution of polyploid complexes (e.g. Kamneva et al., 2017; Morales-Briones et al., 2018; Carter et al., 2019). Attempts to resolve species of allopolyploid origin were then usually done via network-based analyses (Morales-Briones et al. 2018; Carter et al. 2019). However, the standard practice of assembling sequencing reads to a

single consensus sequence per locus has been found to be problematic mainly in studies analysing allopolyploid genomes, as in such cases, these apparently comprise a mix of sequences from different homeologs, largely concealing a real variability within the genome. It was suggested that this could be prevented if the reads originating from different subgenomes were treated separately, which can be accomplished by the method of read-backed phasing (Eriksson et al. 2018; Kates et al. 2018). Then, the sorted reads need to be assembled into the haplotype sequences (alleles), and these might be used for inference of parental lineages of the allopolyploid (Rothfels 2021). The aim to realize this procedure has been addressed in several studies (Kamneva et al. 2017; Lautenschlager et al. 2020; Schrinner et al. 2020; Rothfels 2021), however, its practical utilization has not been fully solved. The development of new methodological approaches for processing NGS data of polyploids is intensive and ongoing, as indicated by very recent studies of Sancho et al. (2022) and Freyman et al. (2023).

The Hyb-Seq protocol was first published by Weitemier et al. (2014) and later modified by Schmickl et al. (2016). One of the crucial steps in this method is the design of taxon-specific probes used for capturing the target sequence regions. In recent years, specific baits were developed for Brassicaceae (Nikolov et al. 2019), as well as several other families (e.g. Asteraceae, Mandel et al. 2014; Apocynaceae, Straub et al. 2020; or Cyperaceae, Villaverde et al. 2020), and the universal probes for angiosperms were also introduced (Johnson et al. 2018). For the purpose of designing Hyb-Seq probes, a variety of approaches or tools were developed (Weitemier et al. 2014; Yang & Smith 2014; Chamala et al. 2015; Schmickl et al. 2016; Li et al. 2017). The *Cardamine*-specific baits were designed by Melichárková et al. (2020) following and using the script named Sondovač 0.99 (Schmickl et al. 2016). Sondovač significantly simplifies the procedure of specific probes designing, facilitating the selection of low-copy nuclear genes by comparing transcriptome and genome skim data (Schmickl et al. 2016). In the process of library preparation, genomic DNA was first isolated and fragmented by sonication, aiming for fragments of 500 bp. After adaptor ligation, the fragments were purified and size-selected to contain only fragments of a desired range of lengths. After PCR enrichment, in which also sample-specific index primers were incorporated, the amplified products were cleaned up and the samples were pooled into one library. Size selection was repeated with the pooled library. Part of the library was enriched by hybridization with genus-specific synthesized RNA baits targeting selected nuclear genes (exons). After PCR amplification and purification, the target-enriched library was pooled with the unenriched one, and the final product was

cleaned-up and sequenced with Illumina. The detailed laboratory protocol is described in the first article presented in the Results (chapter 4.1).

The obtained sequences were trimmed of adapters and low-quality bases were discarded using Trimmomatic-0.36 (Bolger et al. 2014). Consensus target sequences were assembled using HybPiper v1.3 (Johnson et al. 2016) and aligned with MAFFT v7.450 (Katoh & Standley 2013). The aligned exons and genes were concatenated using AMAS (Borowiec 2016). Individual gene trees were computed using RAxML-NG v. 0.9.0 (Kozlov et al. 2019) and the species trees were inferred from the gene trees using a multispecies coalescent model in ASTRAL-III (Zhang et al. 2018). The network-based approaches involved the construction of species network using PhyloNet (Wen et al. 2018), supernetwork analysis in SuperQ v.1.1 (Grünewald et al. 2013; Bastkowski et al. 2018), and SNaQ (Solís-Lemus and Ané, 2016). In the studies of allopolyploids, we used read-backed phasing following the procedure described by Kates et al. (2018) and using WhatsHap and GATK (Martin et al. 2016; Schrinner et al. 2020). Three methods were used to elucidate parental lineages that contributed to the origin of allopolyploids: GRAMPA (Thomas et al. 2017); MPAllop (Yan et al. 2022), and AlleleSorting, which was introduced here as a novel approach in the first study of this thesis (see chapter 4.1). The other method of data analysis involved the generation of single-nucleotide polymorphism (SNP) datasets. In these, the Bayesian clustering was employed to reveal genetic clusters in the data using STRUCTURE 2.3.4 (Pritchard et al. 2000).

Plastome sequences obtained by Hyb-Seq were assembled using Fast-Plast v. 1.2.8 (https://github.com/mrmckain/Fast-Plast). The assembled sequences were aligned in MAFFT v. 7.450 (as above) and the alignments were used to infer ML trees in RAxML-NG v. 0.9.0. Separate ML trees were also constructed for the most variable protein-coding genes of plastome (being first annotated by GeSeq; Tillich et al. 2017), which were then used for inference of coalescent tree in ASTRAL-III.

Sequences of rDNA were also retrieved using HybPiper with use of the available reference genome of *C. amara*. Read-backed phasing (explained above) was applied to discover intra-individual nrDNA variants. Sequences of rDNA were used for construction of ML trees in RAxML-NG and for NeighbourNet analysis done in Splitstree v.4.14.4 (Huson & Bryant 2006). The detailed specifications of all mentioned Hyb-Seq data analyses are included in the particular studies presented in Results (chapter 4).

3.6 Genomic in situ hybridization (GISH)

GISH is a cytogenetic method based on chromosome-labelling, representing an exact and efficient approach for analyses at the level of chromosomes or genomes (Raina & Rani 2001; Silva & Souza 2013). The methodology of GISH represents a modification of fluorescent in situ hybridization (FISH; O'Connor 2008), in which fluorescent labelling is applied on individual chromosomes or its parts. In comparison, GISH implements total genomic DNA as a probe, being able to fluorescently detect overall representation of the genome used as a probe in the target DNA (Schwarzacher et al. 1989; Silva & Souza 2013). One of its significant applications is to detect hybrid status of the target entity, while it is also able to distinguish between auto- and allopolyploids. In the latter case, it enables the identification of specific genome components of two or more progenitor species that have contributed to its origin (Silva & Souza 2013). Apart from its use in resolving parentage of hybridogenous species, its application is much more versatile, as it may be used also in karyotype, genome organization, chromosomal morphology and pairing analyses, being useful also in detecting chromatin and chromosomal rearrangements or aberrations (Raina & Rani 2001; Silva & Souza 2013; Younis et al. 2015; Ramzan et al. 2017). In *Cardamine*, this method was already efficiently used in resolving parentage of multiple polyploid species, reporting the evidence of both auto- and allopolyploidy in their evolutionary history (Mandáková et al. 2013, 2014, 2019).

Here, we implemented GISH in the first study (chapter 4.1) to resolve parental lineages of the allotetraploid species, comparing its outcomes with the findings from molecular markers. The preparation of labelled genomic probes via nick-translation was done according to the process explained by Mandáková & Lysak (2016), and the individual steps of GISH protocol followed Mandáková et al. (2013, 2014). This method was applied in collaboration with Terezie Mandáková who did the experimental work, while I participated in the experimental design and interpretation of data.

4. RESULTS AND DISCUSSION

Results are presented via articles that have been published (chapters 4.1, 4.2) or manuscript that has been finalized and submitted to the scientific journal (chapter 4.3) in time of finalization of the present dissertation thesis. Please note that the published and submitted articles inserted into this chapter have their own lists of references, as well as autonomous page and figure numbering. Therefore, the overall page numbering of the thesis is not shown on these pages.

1. Šlenker, M., Kantor, A., Marhold, K., Schmickl, R., Mandáková, T., Lysak, M. A., Perný, M., Caboňová, M., Slovák, M., & Zozomová-Lihová, J. (2021). Allele sorting as a novel approach to resolving the origin of allotetraploids using Hyb-Seq data: A case study of the Balkan mountain endemic *Cardamine barbaraeoides*. *Front. Plant Sci.*, 12, 659275. doi: 10.3389/fpls.2021.659275

2. Kantor, A., Kučera, J., Šlenker, M., Breidy, J., Dönmez, A. A., Marhold, K., Slovák, M., Svitok, M., & Zozomová-Lihová, J. (2023). Evolution of hygrophytic plant species in the Anatolia-Caucasus region: insights from phylogenomic analyses of *Cardamine* perennials. *Ann. Bot.*, 131, 585–600. doi: 10.1093/aob/mcad008

3. Kantor, A., Šlenker, M., Kučera, J., Marhold, K., Dönmez, A. A., Yüzbaşıoğlu, S., & Zozomová-Lihová, J. Balkan-Anatolian biogeographic links and the evolutionary significance of Anatolian mountains evidenced by *Cardamine* (Brassicaceae). Manuscript was submitted to *Taxon* in April 2023, and is currently in a review process.

4.1 Allele sorting as a novel approach to resolving the origin of allotetraploids using Hyb-Seq data: A case study of the Balkan mountain endemic *Cardamine barbaraeoides*

Frontiers in Plant Science 12: 659275, 2021

doi: 10.3389/fpls.2021.659275

Šlenker, M.*, Kantor, A.*, Marhold, K., Schmickl, R., Mandáková, T., Lysak, M. A., Perný, M., Caboňová, M., Slovák, M., & Zozomová-Lihová, J.

*these authors have contributed equally to this work

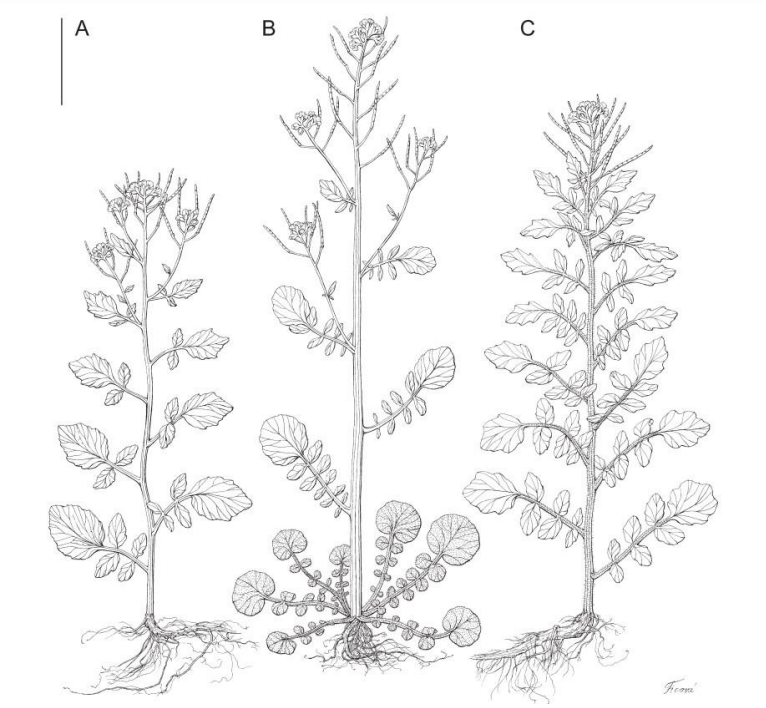


Fig. 3. Drawing of allotetraploid *C. barbaraeoides* (A) and the species that represent extant closest relatives of the lineages that contributed to its origin: *C. acris* subsp. *acris* (B) and *C. amara* subsp. *balcanica* (C). Scale bar represents 5 cm. Figure adopted from Šlenker et al. (2021). Drawn by Jana Ficová.

My contribution: I was involved in collection of the plant material; and in the laboratory work (DNA extraction, PCR, DNA cloning in case of ITS and low-copy nuclear gene (CHS) markers; Hyb-Seq library preparation). I performed the flow-cytometric measurements of both relative genome size and absolute DNA content; participated in procession and analysis of the molecular and flow-cytometric data; and contributed to the preparation of the manuscript.

Supplementary materials are available at:
https://www.frontiersin.org/articles/10.3389/fpls.2021.659275/full#supplementary-material

frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 April 2021
doi: 10.3389/fpls.2021.659275

Allele Sorting as a Novel Approach to Resolving the Origin of Allotetraploids Using Hyb-Seq Data: A Case Study of the Balkan Mountain Endemic Cardamine barbaraeoides

Marek Šlenker[1,2†], Adam Kantor[1†], Karol Marhold[1,2], Roswitha Schmickl[2,3], Terezie Mandáková[4,5], Martin A. Lysak[4,6], Marián Perný[7], Michaela Caboňová[1], Marek Slovák[1,2] and Judita Zozomová-Lihová[1*]

[1] Institute of Botany, Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Bratislava, Slovakia, [2] Department of Botany, Faculty of Science, Charles University, Prague, Czechia, [3] Institute of Botany, The Czech Academy of Sciences, Průhonice, Czechia, [4] Central European Institute of Technology, Masaryk University, Brno, Czechia, [5] Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czechia, [6] National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czechia, [7] Independent Researcher, Žibritov, Slovakia

OPEN ACCESS

Edited by:
Božo Frajman,
University of Innsbruck, Austria

Reviewed by:
Salvatore Tomasello,
University of Göttingen, Germany
Rie Shimizu-Inatsugi,
University of Zurich, Switzerland

*Correspondence:
Judita Zozomová-Lihová
judita.zozomova@savba.sk

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: 27 January 2021
Accepted: 10 March 2021
Published: 28 April 2021

Citation:
Šlenker M, Kantor A, Marhold K, Schmickl R, Mandáková T, Lysak MA, Perný M, Caboňová M, Slovák M and Zozomová-Lihová J (2021) Allele Sorting as a Novel Approach to Resolving the Origin of Allotetraploids Using Hyb-Seq Data: A Case Study of the Balkan Mountain Endemic Cardamine barbaraeoides. Front. Plant Sci. 12:659275. doi: 10.3389/fpls.2021.659275

Mountains of the Balkan Peninsula are significant biodiversity hotspots with great species richness and a large proportion of narrow endemics. Processes that have driven the evolution of the rich Balkan mountain flora, however, are still insufficiently explored and understood. Here we focus on a group of Cardamine (Brassicaceae) perennials growing in wet, mainly mountainous habitats. It comprises several Mediterranean endemics, including those restricted to the Balkan Peninsula. We used target enrichment with genome skimming (Hyb-Seq) to infer their phylogenetic relationships, and, along with genomic in situ hybridization (GISH), to resolve the origin of tetraploid Cardamine barbaraeoides endemic to the Southern Pindos Mts. (Greece). We also explored the challenges of phylogenomic analyses of polyploid species and developed a new approach of allele sorting into homeologs that allows identifying subgenomes inherited from different progenitors. We obtained a robust phylogenetic reconstruction for diploids based on 1,168 low-copy nuclear genes, which suggested both allopatric and ecological speciation events. In addition, cases of plastid–nuclear discordance, in agreement with divergent nuclear ribosomal DNA (nrDNA) copy variants in some species, indicated traces of interspecific gene flow. Our results also support biogeographic links between the Balkan and Anatolian–Caucasus regions and illustrate the contribution of the latter region to high Balkan biodiversity. An allopolyploid origin was inferred for C. barbaraeoides, which highlights the role of mountains in the Balkan Peninsula both as refugia and melting pots favoring species contacts and polyploid evolution in response to Pleistocene climate-induced range dynamics. Overall, our study demonstrates the importance of a thorough phylogenomic approach when studying

the evolution of recently diverged species complexes affected by reticulation events at both diploid and polyploid levels. We emphasize the significance of retrieving allelic and homeologous variation from nuclear genes, as well as multiple nrDNA copy variants from genome skim data.

Keywords: allopolyploidy, Balkan endemism, genomic in situ hybridization, Hyb-Seq, nrDNA, Pindhos Mts., read-backed phasing, target enrichment

INTRODUCTION

The Mediterranean Basin is one of Earth's major biodiversity centers (Myers et al., 2000) harboring several regional hotspots with increased levels of species richness and endemism (Médail and Quézel, 1997; Thompson, 2020). Processes that have given rise to such biodiversity hotspots at a finer scale are complex and reflect interactions of climatic, geological, and biogeographic history of the Mediterranean region (Hewitt, 2011; Nieto Feliner, 2014; Thompson, 2020). Areas of high endemism are concentrated particularly on islands and in mountains, which provide favorable conditions for both speciation and long-term population persistence (Médail and Quézel, 1997; Stevanović et al., 2007; Panitsa et al., 2018; Thompson, 2020). Complex mountainous landscape has a buffering effect on climate change and enables species to survive periods of climatic fluctuations through minor range shifts (Médail and Diadema, 2009; Harrison and Noss, 2017; Muellner-Riehl et al., 2019). Mountains, however, are not just reservoirs, but also cradles of diversity. Great habitat diversity over short geographic distances and high topographic complexity of the mountains creates opportunities in which both adaptive and nonadaptive speciation may occur (Harrison and Noss, 2017; Perrigo et al., 2020). These factors also favored the evolution of narrow endemism in the Mediterranean (Thompson, 2020). In addition, range or niche shifts in response to geological and climatic events may bring vicariant taxa into contact and cause hybridization, with or without a ploidy level increase (Nieto Feliner, 2014). Although hybridization and polyploidization are recognized as significant processes for plant evolution and speciation (Soltis and Soltis, 2009; Soltis et al., 2014), their frequency and contribution to the high species diversity and endemism in the Mediterranean are still poorly understood (Marques et al., 2018; Thompson, 2020).

Here, we focus on the mainland area of the central and southern Balkan Peninsula, which is one of the regional biodiversity hotspots with a large proportion of narrow endemics (Stevanović et al., 2007; Georghiou and Delipetrou, 2010; Tomović et al., 2014). Despite extensive botanical explorations and well-described endemism patterns in this area, speciation processes that have driven the evolution of the rich mountain flora are still not sufficiently explored. Mainly allopatric speciation often accompanied by reticulate and polyploid evolution has been suggested in recent studies (López-Vinyallonga et al., 2015; Olšavská et al., 2016; Đurović et al., 2017; Španiel et al., 2017). High species diversity in this area may also be connected with adjacent Anatolia, which is recognized as a center of lineage diversification in several plant genera and a possible source for the colonization of the Balkan Peninsula (e.g., Ansell et al., 2011; Surina et al., 2014; Caković et al., 2015; Koch et al., 2017). Plant migration via two dispersal corridors, the North Anatolian Mountains or the Taurus Mountains, has been proposed, which was enhanced by land bridges that existed since the Messinian salinity crisis until the Pliocene–Pleistocene transition (Bilgin, 2011; Kaya and Çiplak, 2017; Özüdoğru and Mummenhoff, 2020).

Cardamine L. (Brassicaceae) is a worldwide distributed and species-rich genus (>200 spp.), which has one of its diversity centers located in the European Mediterranean (Marhold et al., 2004, 2018; Lihová and Marhold, 2006; Carlsen et al., 2009; Kučera et al., 2010). The target group of species studied here comprises approximately 30 taxa, both at species and subspecies levels, and includes a few widespread taxa distributed across Europe, several endemics confined to Southern Europe, and also some species from SW Asia (mainly the Anatolian and Caucasus regions). They have commonly been delimited as three related diploid–polyploid species complexes: the Cardamine amara, Cardamine pratensis, and Cardamine raphanifolia groups (Lihová et al., 2004a; Marhold et al., 2004, 2018). In contrast to this traditional, morphology-based delimitation, phylogenetic reconstructions suggested the existence of only two complexes resolved as respective monophyletic clades, one comprising the C. amara complex and the other the remaining species (Marhold et al., 2004; Carlsen et al., 2009). The crown group ages of both clades have been dated back to the Pliocene (approximately 3–4 Mya), and divergence of the extant species likely occurred during the Pleistocene (Huang et al., 2020). Most of the species diversity of these complexes is concentrated in Mediterranean mountains, which host several diploid and polyploid endemics (Marhold et al., 2018). Polyploid origins have been resolved or hypothesized in only a few cases (Lihová et al., 2004a, 2006; Perný et al., 2005a), and even at the diploid level, species relationships within the complexes have remained poorly understood (Lihová et al., 2004a; Marhold et al., 2004). In the Balkan Peninsula, diploid endemics prevail, and these include Cardamine penzesii Ančev et Marhold, Cardamine rivularis Schur, C. amara subsp. balcanica Ančev, Marhold et Kit Tan, and Cardamine acris Griseb. with three subspecies recognized. In addition, tetraploid populations from the Pindos Mts. in northwestern Greece have been reported and attributed to Cardamine barbaraeoides Halácsy. It is a species with an uncertain circumscription and unknown polyploid origin (Marhold et al., 2018).

High-throughput DNA sequencing has brought excellent opportunities to improve phylogenetic inferences, particularly when facing difficult evolutionary cases, such as rapid radiations or recent speciation characterized by low genetic divergence and presence of incomplete lineage sorting (ILS) often complicated

by hybridization and polyploidy (Schmickl et al., 2016; Nikolov et al., 2019; Karbstein et al., 2020; Larridon et al., 2020). Disentangling reticulate and polyploid evolution, however, has been a difficult task, and phylogenomic studies on polyploids have lagged behind (Oxelman et al., 2017; Rothfels, 2021). Recent advances in this respect (see, e.g., Kamneva et al., 2017; Morales-Briones et al., 2018; Carter et al., 2019; Brandrud et al., 2020) have opened up new perspectives on analyses of polyploid species complexes. Approaches that account simultaneously for ILS and reticulation have been developed and improved (Oberprieler et al., 2017; Wen et al., 2018; Cao et al., 2019). Those network methods can provide significant insights into the evolution of polyploids based on multilocus sequence data (e.g., Kamneva et al., 2017; Morales-Briones et al., 2018). Still, standard practice when assembling sequencing reads is to generate a single consensus sequence per locus and individual, which represents a strong violation for allopolyploid genomes. The outcome of such consensus assembly is a mix of sequences retrieved from different homeologs (parental subgenomes) and chimeric sequences. Therefore, the crucial steps to resolve in polyploid phylogenetics are to separate sequencing reads originating from different subgenomes, assemble haplotype (allele) sequences, assign them to the subgenomes, and trace the parental origin of these subgenomes by multilabeled species tree or network inference methods (Rothfels, 2021). A few recent studies have explored different ways how to accomplish these steps, either via mapping and categorization of the sequence reads to the reference diploid genomes (Page et al., 2013; Grover et al., 2015), developing bioinformatics pipelines for amplicon sequences of polyploids from long-read sequencing platforms (Rothfels et al., 2017), or via the assembly of haplotype sequences by read-backed phasing (Eriksson et al., 2018; Kates et al., 2018). Nevertheless, the assignment of alleles to parental subgenomes has been critical and difficult to achieve readily for hundreds of loci typically recovered by target enrichment techniques. Some statistical methods for this task are under development and appear promising (Freyman et al., 2020; Lautenschlager et al., 2020), but may also be computationally intensive.

In this article, we employ target enrichment with genome skimming (Hyb-Seq) using genus-specific probes to capture hundreds of orthologous low-copy nuclear loci (target exons with flanking intronic and intergenic regions), along with obtaining the complete plastid genome and high-copy nuclear ribosomal DNA (Weitemier et al., 2014; Schmickl et al., 2016). Here we develop a novel computational approach to sort alleles obtained from polyploids into parental subgenomes, utilizing genetic distances among alleles, and employ it to reconstruct the origin and parentage of tetraploid C. barbaraeoides. We complement this phylogenomic approach with genomic in situ hybridization (GISH, Silva and Souza, 2013). In detail, we aimed to (1) resolve phylogenetic relationships among Balkan Cardamine species and determine major factors affecting endemism patterns in mountains of the Balkan Peninsula; (2) reconstruct the origin of tetraploid C. barbaraeoides from the Pindos Mts. in Greece to shed light on the evolution of mountain endemic flora through polyploidy; and (3) identify challenges of phylogenomic analyses of polyploid species, where we focus on resolving heterozygous and homeologous sequence variation and its sorting into parental subgenomes.

MATERIALS AND METHODS

Study Species and Sampling

The target species complexes of Cardamine comprise rhizomatous perennials with an allogamous or mixed mating system, capable of vegetative propagation (Lövkvist, 1956; Marhold and Ančev, 1999; Tedder et al., 2015). They grow in wet habitats from lowlands up to the alpine belt, in or nearby running or standing water, usually along river and stream banks, in springs, wet meadows and pastures, in flood-plain to montane forests. Morphologically, they are characterized by pinnate basal leaves, pinnate to pinnatisect stem leaves, and white, pale pink to purple flowers arranged in racemes (e.g., Marhold et al., 1996; Marhold and Ancev, 1999; Lihová et al., 2004b; Perný et al., 2004). In the Balkan Peninsula, they include mostly endemics (C. amara subsp. balcanica, C. acris subsp. acris, subsp. vardousiae Perný et Marhold, subsp. pindicola Perný et Marhold, C. barbaraeoides, C. penzesii, C. rivularis) or more widespread European taxa reaching their southeastern distribution margins there [Cardamine matthioli Moretti, C. amara subsp. amara, subsp. opicii (J. Presl et C. Presl) Čelak; **Figure 1**]. Apart from tetraploid records for C. barbaraeoides (Perný et al., 2005a; Lihová and Marhold, 2006], the other Balkan taxa are known to be diploid, with exceptional triploid plants reported for C. rivularis and C. rhodopaea Ancev (C. rivularis C. matthioli) (Kucera et al., 2005; Ancev et al., 2013; Melichárková et al., 2020). Only diploid representatives have so far been reported from the adjacent Anatolian–Caucasus region (Marhold et al., 2004; Kuceřa et al., 2005). In the Apennines, on the contrary, one diploid (Cardamine apennina Lihová et Marhold) and two polyploids (Cardamine silana Marhold et Perný, Cardamine amporitana Sennen et Pau, both presumably allopolyploids, Perný et al., 2005a, and unpubl. results) occur (Marhold et al., 2018). The three species, C. amara, C. amporitana, and Cardamine lazica Boiss. et Balansa ex Buser (the last one being referred to as Cardamine wiedemanniana Boiss. in our previous studies; see, e.g., Lihová et al., 2004a), have been regarded as members of the C. amara complex, whereas the other species have been attributed to either the C. pratensis or C. raphanifolia groups. The position of C. barbaraeoides remained uncertain and was commonly classified either as C. amara subsp. barbaraeoides (Halácsy) Maire et Petitm. (Tan, 2002) or as C. raphanifolia subsp. barbaraeoides (Halácsy) Strid (Strid, 1986; Jones and Akeroyd, 1993).

Here, we included all taxa occurring in the Balkan Peninsula, plus diploids from adjacent areas, C. apennina from the Apennines, and C. lazica and Cardamine uliginosa M.Bieb. from the Anatolian–Caucasus region (**Figure 1**, **Supplementary Data Sheet 1**). C. uliginosa is a highly polymorphic and widespread species (**Figure 1**) described from the Caucasus, but probably being polyphyletic and pending further detailed studies (Marhold et al., 2004; study under progress). Two geographically distant

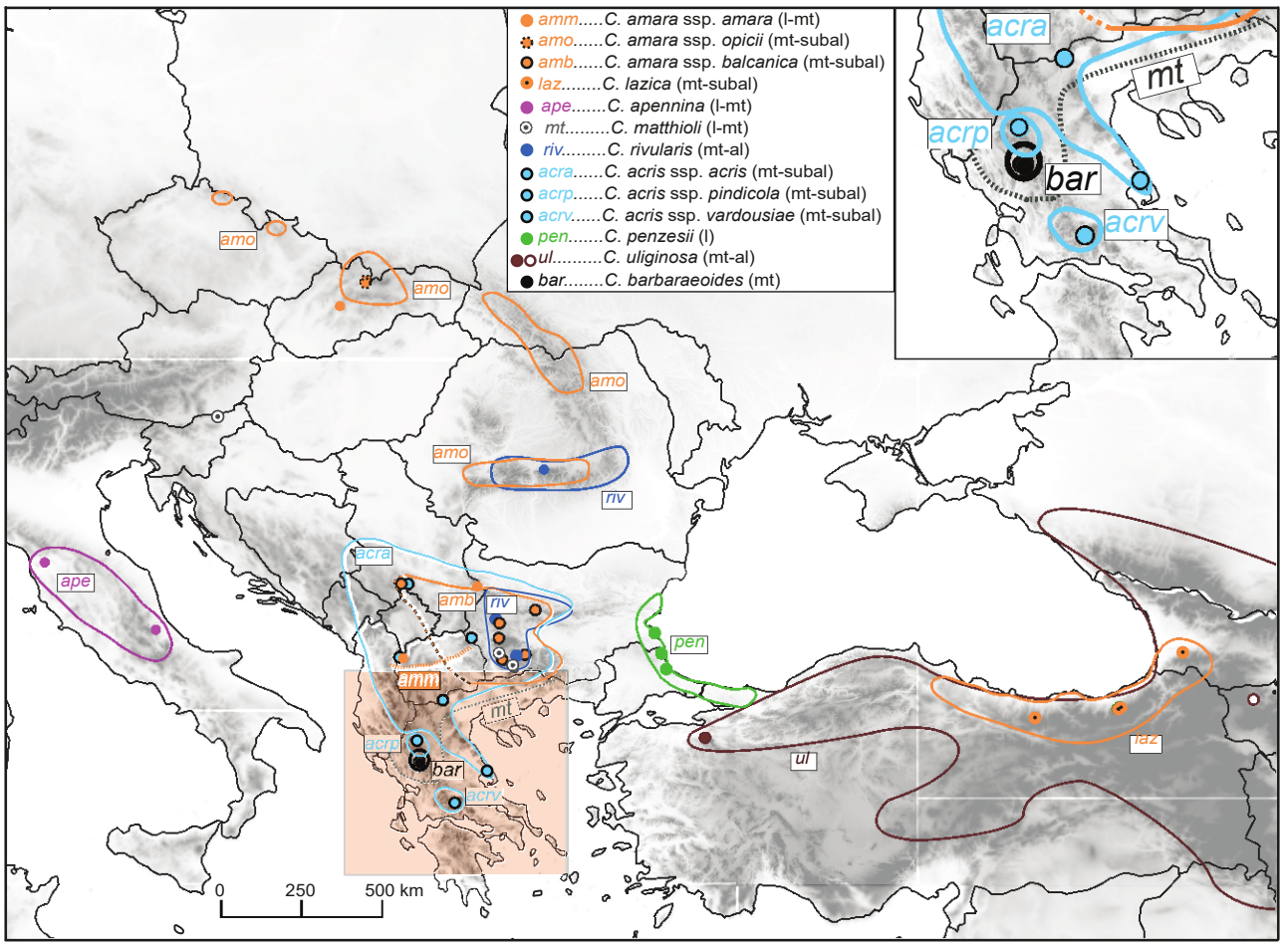


FIGURE 1 | Distribution of the Cardamine taxa under study, based on data compiled from floras, herbarium specimens, previous studies, and our own records. The western borders of the area of Cardamine amara subsp. balcanica remain unknown (marked here by a black–orange dashed line; see Tomovic´et al., 2009); this taxon has been thoroughly studied so far only in its Bulgarian range (Marhold et al., 1996). Cardamine matthioli and C. amara subsp. amara are widespread taxa in Europe (Jalas and Suominen, 1994), but their precise distribution in the Balkan Peninsula remains unclear, with the southernmost records reported from central and northeastern Greece (gray dotted line; Marhold and Tan, 2000) and North Macedonia (orange dotted line; Jalas and Suominen, 1994; Tomovic´et al., 2009, and this study), respectively. The area of Cardamine uliginosa extends further to the south and southeast, reaching the mountains of Iran and Lebanon. The occurrence of the taxa along the elevational gradient is indicated in brackets as follows: l, lowland; mt, montane; subal, subalpine; al, alpine belt. Circles indicate our sample sites; see Supplementary Data Sheet 1 for details on the populations sampled.

accessions attributed to this species were included here, one from the Uludağ Mts. in NW Turkey, here referred to as C. cf. uliginosa, and the other from the Caucasus (Armenia). Altogether, we sampled 46 populations representing nine species (13 taxa), which were used for ploidy level and genome size measurements by flow cytometry (307 accessions), polymerase chain reaction (PCR) amplification of the nuclear ribosomal DNA (nrDNA) ITS region (48 accessions), and Hyb-Seq analyses (22 accessions) capturing target nuclear genes, plastid DNA, and nrDNA. The tetraploid C. barbaraeoides and a selection of potential parental candidates were used in GISH experiments. In addition, four diploids representing phylogenetically divergent lineages (following the genus phylogeny, Carlsen et al., 2009) were included as outgroups. The list of the populations sampled and accessions analyzed is given in **Supplementary Data Sheet 1**.

Chromosome Counting and Flow Cytometry

Chromosomes of C. barbaraeoides were counted from mitotic metaphase plates observed in cells of young, actively growing root tips obtained from cultivated plants. Chromosome spreads were prepared following Marhold et al. (2002) using the Giemsa stain, or following Mandáková and Lysak (2016a) using the DAPI (4^0,6-diamidino-2-phenylindole) fluorochrome. For the other sampled species and subspecies, chromosome number records were available from previous studies (Kučera et al., 2005), in some cases even from the here sampled localities (see **Supplementary Data Sheet 1** for details).

Flow cytometry was applied here to measure nuclear DNA content of the sampled accessions (Doležel et al., 2007). These measurements were performed to confirm that the ploidy level

of the analyzed populations and accessions is uniform and agrees with the known records, as well as to determine genome size differences between the species. Both absolute and relative nuclear DNA content was measured, using the DNA-intercalating fluorochrome [propidium iodide (PI)], and the AT-selective DAPI fluorochrome, respectively (Doležel et al., 2007). For PI measurements, we used fresh leaf tissue from cultivated plants, whereas for DAPI measurements we used silica gel–dried tissue (Suda and Trávníček, 2006a,b). Each individual was analyzed separately (for precise relative or absolute nuclear DNA content values), or up to three individuals were pooled (for ploidy level inference only; see **Supplementary Data Sheet 1**). Sample preparation followed the protocols described by Marhold et al. (2010). Fluorescence of the stained nuclei was measured using Partec CyFlow flow cytometers (Partec GmbH, Münster, Germany), either with a UV LED lamp for DAPI measurements or a green solid-state laser for PI measurements. Relative nuclear DNA content (2C value given in arbitrary units) was calculated as the ratio between the positions of the G1 peaks of the sample and the standard. Absolute nuclear DNA content (2C value given in pg) was calculated from the ratio of the respective G1 peaks and the known 2C value of the standard. Solanum pseudocapsicum (2C = 2.59 pg; Temsch et al., 2010) was used as the primary internal standard. In cases when peak overlaps between the sample and standard were observed or expected, Bellis perennis (2C = 3.38 pg; Schönswetter et al., 2007) was used as the secondary standard (see **Supplementary Data Sheet 1**).

PCR Amplification, Molecular Cloning, and Sanger Sequencing

Polymerase chain reaction amplification, molecular cloning, and Sanger sequencing of the ITS region of nrDNA were employed to explore the diversity of ITS variants within and between individuals, both diploid and tetraploid, as well as to compare this approach with the accuracy and efficiency of retrieving different ITS variants from high-throughput genomic reads. In addition, PCR amplification, molecular cloning, and Sanger sequencing of chalcone-synthase (CHS) was performed for tetraploid accessions only (**Supplementary Data Sheet 1**). CHS is a single-copy nuclear gene of high phylogenetic resolution, used previously to infer polyploid origins and phylogeny of Cardamine species (Lihová et al., 2006; Kučera et al., 2010). It was included among the target genes in the Hyb-Seq approach, and therefore, the sequenced CHS clones were used to verify and optimize the assembly of allele sequences by read-backed phasing and the procedure of allele sorting into parental homeologs in tetraploid accessions (see below).

Genomic DNA (gDNA) was isolated from silica gel–dried leaves using the DNeasy Plant Mini Kit (Qiagen, Germany) or GeneAll Exgene Plant SV mini kit (GeneAll Biotechnology Co., LTD., South Korea). ITS amplifications and molecular cloning followed the protocols specified in Melichárková et al. (2017, 2019). Exon 2 of CHS was amplified with the primers CHSF2 and CHSR1 (Lihová et al., 2006) and cloned following Melichárková et al. (2017). The PCR reaction mix contained also 3% dimethyl sulfoxide to suppress PCR-mediated recombination events. Multiple clones per sample were sequenced (see **Supplementary Data Sheet 1** for details). The sequencing was carried out at Eurofins Genomics Company (Konstanz, Germany).

Hyb-Seq Library Preparation

Sequencing libraries were prepared using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, MA, United States) following the manufacturer's protocol. gDNA (400 ng per accession) was fragmented with a Covaris M220 sonicator (Woburn, MA, United States) to a target fragment size of 500 bp. Adaptor-ligated DNA fragments were purified with the QIAquick PCR Purification Kit (Qiagen) and size-selected using SPRIselect beads (Beckman Coulter, MA, United States) to a 500- to 600-bp size range. PCR enrichment with eight cycles was performed using index primers from NEBNext® Multiplex Oligos for Illumina®. The amplified libraries were cleaned up with AMPure XP beads (Beckman Coulter), measured with a Qubit 2.0 fluorometer (ThermoFisher Scientific, MA, United States), and pooled equimolarly (24 accessions/pool). The pooled library was size-selected using SPRIselect beads as above and measured again with the Qubit 2.0. An aliquot containing 250 ng was enriched by hybridization with synthesized RNA baits (26 h at 65C) using the MYbaits® kit, following the protocol v. 3.02 (Arbor Biosciences, MI, United States). The target-enriched library was amplified by PCR with nine cycles using the KAPA HiFi HotStart mix (Kapa Biosystems, Wilmington, MA, United States) and purified with the QIAquick PCR Purification Kit. Enriched and unenriched library aliquots were pooled in a ratio 2:1, finally purified with AMPure XP beads, and submitted for sequencing with 150-bp paired end reads on an Illumina MiSeq system at BIOCEV, Czechia.

The design of the Cardamine-specific target enrichment probes is described in detail in Melichárková et al. (2020). In brief, we used genome skim data of Cardamine parviflora (NCBI accession no.: SRR11977919) omitting plastid and mitochondrial reads, which were matched against unique transcripts of C. amara (SRR11977918), utilizing the workflow of the Sondovač 0.99 script[1] (Schmickl et al., 2016). Genome skim hits were assembled into larger contigs, which were filtered for length and uniqueness, and compiled as probe sequences for bait synthesis. In total, 14,464 120-mer biotinylated RNA baits, capturing 2,246 exons from 1,235 genes, were synthesized by MYcroarray (now Arbor Biosciences).

Hyb-Seq Data Processing and Phylogenomic Analyses

Demultiplexed reads were trimmed of adapters and low-quality bases using Trimmomatic v. 0.36 (Bolger et al., 2014). Read ends with quality below Q20 were discarded, and the remaining part of the read was trimmed if average quality in a 4-bp sliding window was below Q15. Finally, any reads trimmed to less than 50 bp were discarded. PCR duplicates were removed using the Clumpify command of BBTools[2].

[1]https://github.com/V-Z/sondovac

[2]https://jgi.doe.gov/data-and-tools/bbtools

Consensus target sequences were assembled using HybPiper version 1.3 (Johnson et al., 2016) utilizing BWA v. 0.7.13 (Li and Durbin, 2009), SPAdes v. 3.13 (Bankevich et al., 2012), and Exonerate v. 2.2 (Slater and Birney, 2005). HybPiper generates a single consensus sequence per individual, with potentially heterozygous bases called as the nucleotide with the highest read frequency. "Supercontigs" (targeted exons and flanking sequences) were recovered using the script intronerate.py. Recovered consensus supercontig sequences were aligned using MAFFT v. 7.313 (Katoh and Standley, 2013). Flanks and sites with gaps in more than 25% of sequences were removed using the ips R package (Heibl, 2008 onward) in R 3.3.2 (R Core Team, 2019). Alignments were inspected visually, and misassemblies were removed. In addition to using the consensus supercontig sequences, the allele sequences were inferred with read-backed phasing (described in detail below in Extracting Allele Sequences and Identifying Homeologs Inherited From Different Parents) using WhatsHap (Martin et al., 2016). Both consensus and allele data sets were used in further analyses.

The recovered sequences of the target nuclear genes were analyzed using the following workflow. First, we performed phylogenomic analyses of diploid taxa only (with both the consensus and allele sequence alignments), to provide a robust phylogenetic framework, using both concatenation of assembled genes and species tree inference under the multispecies coalescent model. As next, we analyzed diploids together with the tetraploid C. barbaraeoides. Considering that the tetraploid genome consists of two subgenomes that may be more or less differentiated, and thus potentially conveys conflicting phylogenetic information, we used here multiple approaches. To gain initial insights into the tetraploid genome, we used consensus supercontig sequences and applied methods that can detect and visualize conflict caused by potential discordance between consensus supercontigs retrieved from independent genes. In allopolyploids, the consensus sequences may comprise different homeologs or even consist of artificial, chimeric sequences. The analyses included supernetwork and species network calculations based on the gene trees obtained from the assembled consensus sequences, as well as single-nucleotide polymorphisms (SNPs) calling followed by Bayesian clustering of the SNP datasets. Finally, when the conflict between the subgenomes of the tetraploid became apparent, we derived allele sequences of the exons by read-backed phasing also from the tetraploids (see below in Extracting Allele Sequences and Identifying Homeologs Inherited From Different Parents). Up to four different alleles obtained from the exons of tetraploid C. barbaraeoides were sorted into two distinct homeologs based on allelic divergence (computing interallelic distances, see below) using an optimized threshold value. The resulting allele alignments were submitted to coalescent-based species tree inference.

Phylogenetic trees were constructed using RAxML-NG v. 0.9.0 (Kozlov et al., 2019). The best-fit model of substitution for each gene, exon, or partitioning scheme was estimated using the IQ-TREE's ModelFinder function (Chernomor et al., 2016; Kalyaanamoorthy et al., 2017) under the Bayesian information criterion. Branch support of the best ML trees was estimated by 500 bootstrap (BS) replicates. The quartet sampling method (Pease et al., 2018), which can distinguish strong conflict from weak signal, was applied to assess branch support of the trees generated from the concatenated alignments. The concatenation of the aligned exons and genes was performed by AMAS (Borowiec, 2016). Species trees were inferred from individual gene trees under a multispecies coalescent model using ASTRAL-III (Zhang et al., 2018). PhyloNet was employed to infer a species network evaluating reticulate evolutionary relationships in individual gene trees. The network was inferred with a single reticulation node using the InferNetwork_MP method in 10 runs, each with two optimal networks returned (Wen et al., 2018). SuperQ v.1.1 (Grünewald et al., 2013; Bastkowski et al., 2018) decomposed gene trees into quartets, and inferred a supernetwork selecting the JOptimizer scaling and Gurobi optimizer. The trees used as input data for species tree reconstruction and both network analyses had contracted branches with low support (20%) by Newick-Utilities v. 1.6 (Junier and Zdobnov, 2010). Bayesian clustering of SNP data was performed to infer homogeneous genetic clusters with STRUCTURE 2.3.4 (Pritchard et al., 2000). Input datasets were generated by the snipStrup pipeline [available online at: https://github.com/MarekSlenker/snipStrup; described in detail in Melichárková et al. (2020)]. This pipeline uses target sequences (those used for probe synthesis) as a reference and calls variants with respect to ploidy. To ensure that no linkage existed between sites, 500 datasets were produced by drawing a single random SNP site from each gene containing at least 10 SNPs across the samples. Each dataset was run for each K = 1–10 (user-defined number of clusters), with a burn-in length of 100,000 generations and data collection for an additional 900,000 generations, setting the admixture model and correlated allele frequencies. The results of 500 datasets were averaged using the program CLUMPP (Jakobsson and Rosenberg, 2007) and drawn with Distruct (Rosenberg, 2004). The approach of Evanno et al. (2005) was used to determine the optimal K value.

Extracting Allele Sequences and Identifying Homeologs Inherited From Different Parents

Allele sequences were derived using the scripts and following the workflow available online at: https://github.com/mossmatters/phyloscripts/tree/master/alleles_workflow, described in detail by Kates et al. (2018), only using the latest versions of GATK and WhatsHap (Martin et al., 2016; Schrinner et al., 2020) enabling to call and phase variants in polyploids. If the phased sequences were divided into multiple blocks, only the longest phase block for each individual was retained, and the remaining interallelic variant sites were masked by using Ns on those positions.

The alleles obtained from the tetraploid C. barbaraeoides were sorted into two distinct homeologs as follows. The first step was to find two pairs of alleles, in which the alleles are closest to each other within the pairs while more distant between the pairs. Interallelic distances were estimated from the branch lengths of the corresponding exon or gene ML trees (computed by cophenetic function of package stats, R Core Team, 2019). The optimal threshold for unequivocal allele sorting was set to

4 (for more details about searching for the optimal threshold value, see **Supplementary Text 1**). This means that if an average distance between alleles within the proposed two pairs was more than four-time shorter than the average distance between alleles within any other possible arrangement, these pairs of alleles were considered unequivocally different and attributable to different homeologs (see also **Supplementary Text 2**). If the allele sorting did not pass the desired threshold, two options were followed. Either the interallelic SNPs were masked by using Ns on those positions (such unsorted, masked exons were used for further concatenation into gene alignments, see below) or the sample was removed from the alignment (for exon-based analyses). As next, the allele pairs were attributed to different homeologs and labeled by calculating their distances to the alleles of all diploid species. The allele pair that was closer to C. amara (proposed as the maternal parent according to the plastome phylogeny, see below) was marked as homeolog "A", and the other pair as homeolog "B". Gene alignments were also assembled, in which the phased alleles of the respective exons were concatenated to genes to obtain longer alignments with potentially stronger phylogenetic signal. The concatenated exons included those with successfully sorted alleles into "A" and "B" homeologs and those for which allele sorting was equivocal, with masked interallelic SNPs. After exon concatenation, the allele sorting into two homeologs was verified for each gene, with the same threshold as set for the exons above, to confirm unambiguity or to remove the equivocal sample from the gene alignments. Both exon-based and gene-based alignments were used for species tree inference in ASTRAL-III. The labeled homeologs, representing the two subgenomes within C. barbaraeoides, were treated as independent accessions. The scripts used are available online at: http://github.com/MarekSlenker/AlleleSorting.

Gene Genealogy Interrogation Analyses

To explore the significance of phylogenetic placements of the A and B homeologs of C. barbaraeoides, we performed alternative topology testing using the gene genealogy interrogation (GGI) analyses (Arcila et al., 2017). This approach accounts for gene tree estimation error and evaluates the relative support for specific alternative hypotheses. First, the hypotheses to be tested are defined by performing constrained ML gene tree searches with enforced monophyly of the examined clades in RAxML. Here we considered three different topologies for both A and B homeologs, following the results of PhyloNet analyses and exon- and gene-based species trees inferred from phased sequences (see Results for details). The topology test was then performed for each nuclear gene or exon (i.e., considering both exon- and gene-based phased datasets) by statistically comparing the site likelihood scores obtained for each constrained tree in RAxML using the approximately unbiased (AU) topology test implemented in CONSEL (Shimodaira and Hasegawa, 2001; Shimodaira, 2002). The AU test performs simultaneous comparisons of multiple trees and estimates a P value for each topology. The trees are then ranked according to the P values, and the results are visualized as plots of the cumulative number of constrained gene trees and their AU test P values for each topology.

Analyses of nrDNA Sequence Data

nrDNA sequences obtained from molecular cloning were aligned in Geneious v. R10 (Kearse et al., 2012). Sequences of nrDNA were also recovered from Hyb-Seq data in HybPiper using C. amara (AY260579.1) and C. pratensis (KF987809.1) reference sequences, as specified above for the target nuclear loci, but omitting the "supercontig" option. The sequences were aligned using MAFFT v. 7.450 (Katoh and Standley, 2013), and only the ITS region was extracted and kept for further analyses to allow for direct comparison with the cloned data. The sequences recovered from HybPiper were also proceeded further to read-backed phasing to retrieve multiple nrDNA variants, as described above. Here were generated four nrDNA datasets as follows: (1) alignment obtained from molecular cloning; (2) consensus assembly with base calling following the majority rule criterion, as produced by HybPiper; (3) ambiguous assembly with intraindividual SNPs replaced by IUPAC codes produced by bcftools consensus command; and (4) "multiallelic" (read-backed phasing) alignment, where multiple nrDNA variants were retrieved for each sample. Maximum likelihood (ML) trees were inferred with RAxML-NG as above.

Analyses of Chloroplast Genome Data

Chloroplast DNA sequences were assembled using Fast-Plast v. 1.2.8 (available online at: https://github.com/mrmckain/Fast-Plast) with default settings. This pipeline utilizes Trimmomatic v. 0.39 (Bolger et al., 2014) for initial read cleaning, Bowtie 2 v. 2.3.5.1 (Langmead and Salzberg, 2012) to extract chloroplast reads using a database of reference plastomes, SPAdes v. 3.13 (Bankevich et al., 2012), and afin (available online at: https://github.com/mrmckain/Fast-Plast/tree/master/afin) for de novo sequence assembly. For two accessions, for which the plastome assembly failed in Fast-Plast, chloroplast DNA sequences were assembled in HybPiper using the C. amara (KY562580.1) reference sequence. The obtained plastome sequences, comprising the large single copy (LSC), the small single copy (SSC), and one copy of the inverted repeats (IRb), were aligned using MAFFT v. 7.450 (Katoh and Standley, 2013). Gene annotation (protein coding, tRNA and rRNA genes) was performed with GeSeq (Tillich et al., 2017). Two chloroplast DNA (cpDNA) alignments were generated and used for phylogenetic tree reconstructions, one comprising the complete sequences of the LSC, SSC, and IRb regions, including intergenic spacers, and the other consisting of the concatenated sequences of annotated genes only. ML trees were inferred in RAxML-NG as above. Although it has been widely assumed that plastid genes are inherited as a single locus, favoring their concatenation before phylogenetic analyses, some recent studies have indicated that these genes may not be as tightly linked as expected and may experience different evolutionary histories. Therefore, the application of multispecies coalescent methods to account for potential discordance between gene trees has been advocated also for plastome genes (Gonçalves et al., 2019; Walker et al., 2019). Following this research, we extracted the most variable protein-coding genes (42 genes, those > 350 bp long with > 10 variable positions in the alignment), for which separate ML gene trees

were constructed in RAxML-NG. The obtained ML gene trees were then used for species tree inference in ASTRAL-III.

Genomic in situ Hybridization

Genomic in situ hybridization was performed in C. barbaraeoides to identify its parental chromosome complements. GISH probes were prepared from total gDNA of eight diploid taxa, C. acris subsp. acris, C. amara subsp. amara, subsp. balcanica, C. lazica, C. matthioli, C. penzesii, C. rivularis, and C. uliginosa (see **Supplementary Data Sheet 1**), which were used in different combinations. Mitotic chromosome spreads of C. barbaraeoides were prepared as described above, following Mandáková and Lysak (2016a). To remove RNA and cytoplasm, the preparations were treated with 100 mg/mL RNase (AppliChem) in 2 sodium saline citrate (20 sodium saline citrate: 3 M sodium chloride, 300 mM trisodium citrate, pH 7.0) for 60 min, and 0.1 mg/mL pepsin (Sigma) in 0.01 M HCl at 37C for 5 min, and then postfixed in 4% formaldehyde in 2 sodium saline citrate for 10 min, washed in 2 sodium saline citrate twice for 5 min, dehydrated in an ethanol series (70%, 80%, and 96%, 2 min each), and air-dried. gDNA of the diploids was extracted from silica gel–dried leaves using the DNeasy Plant Mini Kit (Qiagen). Isolated gDNA was labeled with either biotin-dUTP or digoxigenin-dUTP via nick translation according to Mandáková and Lysak (2016b). Individual labeled probes were stored at 20C until use. The GISH protocol followed Mandáková et al. (2013, 2014). The immunodetection of hapten-labeled probes was performed as follows: biotin-dUTP was detected by avidin–Texas red (Vector Laboratories) and amplified by goat anti-avidin–biotin (Vector Laboratories) and avidin–Texas red; digoxigenin-dUTP was detected by mouse antidigoxigenin (Jackson ImmunoResearch) and goat anti-mouse–Alexa Fluor 488 (Invitrogen). After immunodetection, chromosomes were counterstained with DAPI (2 mg/mL) in Vectashield (Vector Laboratories). Painted chromosome figures were photographed using an Axioimager Z2 epifluorescence microscope (Zeiss) equipped with CoolCube CCD camera (MetaSystems). Images were acquired separately for the three fluorochromes using appropriate excitation and emission filters (AHF Analysentechnik). The three monochromatic images were pseudocolored, merged, and cropped using Photoshop CS (Adobe Systems) and Image J (National Institutes of Health) software.

RESULTS

Chromosome Numbers and Genome Size Variation

Chromosome counting revealed the tetraploid level with 2n = 32 chromosomes in C. barbaraeoides, determined in two populations. Flow cytometry confirmed the tetraploid level in all five sampled populations (27 individuals in total; **Supplementary Data Sheet 1**). Ploidy level screening within the other studied species showed consistent results, supporting a single, diploid level. Only few exceptions were identified, such as one apparently triploid individual of C. acris and population C018 of C. acris with increased genome size values not attributable to any ploidy level with certainty (**Supplementary Data Sheet 1**). The diploid species displayed a wide range of 2C values, and most of the species differed from each other in their nuclear DNA content (**Supplementary Data Sheet 1**, **Figure 2**). Populations of C. cf. uliginosa from the Uludag Mts. (UD, northwestern Turkey) and the Caucasus Mts. (AM, Armenia) showed markedly different values (in accordance with their genetic divergence, see below) and were kept as two separate entities. The smallest genome sizes were observed in C. amara and C. lazica, whereas the largest ones in C. acris, C. rivularis, and C. cf. uliginosa from the Uludag Mts., being more than twice as big as in C. amara. In accordance with the tetraploid level, the largest nuclear DNA content was measured in C. barbaraeoides, but when recalculated to the meiotically reduced genome (corresponding to the 2x level), it showed an intermediate value placed among the diploids (**Figure 2**).

Hyb-Seq Data

The sequencing process yielded, on average, 1.28 million reads per sample. Adapter trimming, quality filtering and deduplication resulted in an average loss of 1.06% of reads. Of the remaining reads, 54.11% on average were mapped to the target nuclear gene sequences, which ensured mean coverage of more than 97 reads per base. Mean coverage of the plastid genome fluctuated widely among samples, from 13.5 to 96.23 reads per base (43.56 on average). The same was true for the ITS region of nrDNA, but the mean coverage of all samples was more than 70 reads per base. Of the 2,246 exons from 1,235 genes, targeted by the designed RNA baits, 1,858 (82.72%) consensus sequences were assembled in all 22 samples. More than 98% of sequences, that is, 1,829 supercontigs representing 1,168 genes, passed inspection and were used for further analyses. The length of the exon alignments ranged from 63 to 3,548 bp (709 bp on average), whereas the gene length ranged from 72 to 8,458 bp, with a mean of 1,111 bases. The concatenated alignment of all genes was 1,297,401 bp long.

Phylogenomic Analyses of Diploids Based on Target Nuclear Loci

Maximum likelihood analysis of the diploid taxa, based on the concatenated dataset of all 1,829 loci (consensus supercontigs) from 1,168 nuclear genes, resulted in a tree with two major well-supported clades (**Figure 3A**, **Supplementary Figure 1**). One clade comprised accessions of C. amara and C. lazica in a sister position, supported by high BS as well as quartet concordance (QC) values. The other major clade exhibited a topology with strong to moderate support (QC = 0.42–1, BS = 69%–100%) and comprised three subclades as follows: (1) C. acris resolved in a sister position to C. cf. uliginosa from the Uludag Mts.; (2) C. penzesii together with the accession of C. uliginosa from the Caucasus; (3) C. apennina and C. matthioli in a sister position, together with C. rivularis. Because the two geographically distant accessions of C. uliginosa

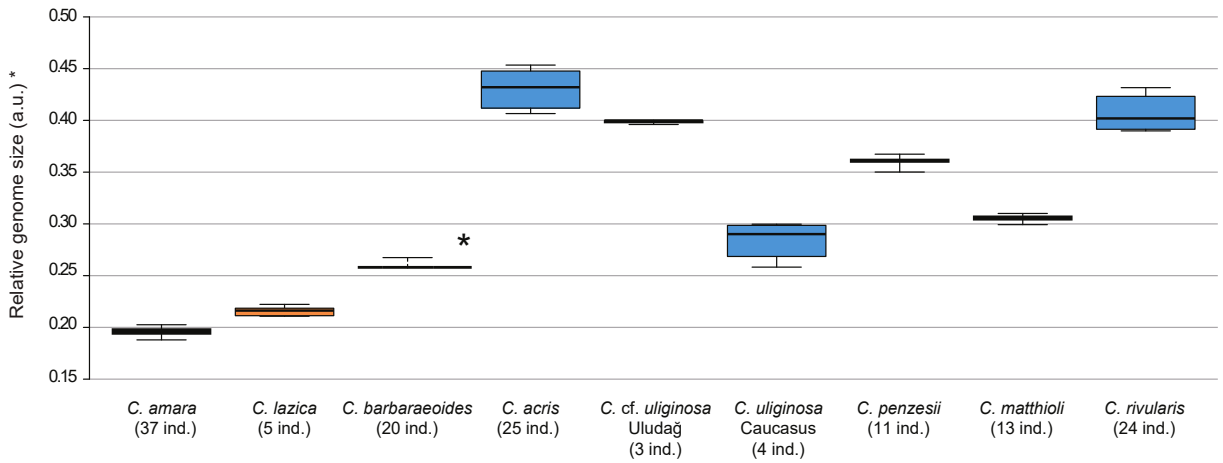


FIGURE 2 | Genome size variation of the Cardamine species under study, based on flow cytometric analyses. Relative nuclear DNA content inferred from DAPI measurements is presented, given as the ratio of the sample and standard G1 peaks (2C values in arbitrary units, a.u.). In the tetraploid Cardamine barbaraeoides (marked by asterisk), however, DNA content of the meiotically reduced genome (corresponding to the 2x level) is assessed and presented. Population C018 of Cardamine acris was omitted because of its divergent DNA content and unclear ploidy level (see Supplementary Data Sheet 1). Boxplot graphs show the 25th and 75th percentiles (boxes), median values (vertical lines within boxes), and minimum to maximum values (whiskers). Orange color is used for species of the Cardamine amara group, blue for the other diploids, and black for the tetraploid C. barbaraeoides. The number of analyzed individuals per species is indicated. See Supplementary Data Sheet 1 for more details and population-level values.

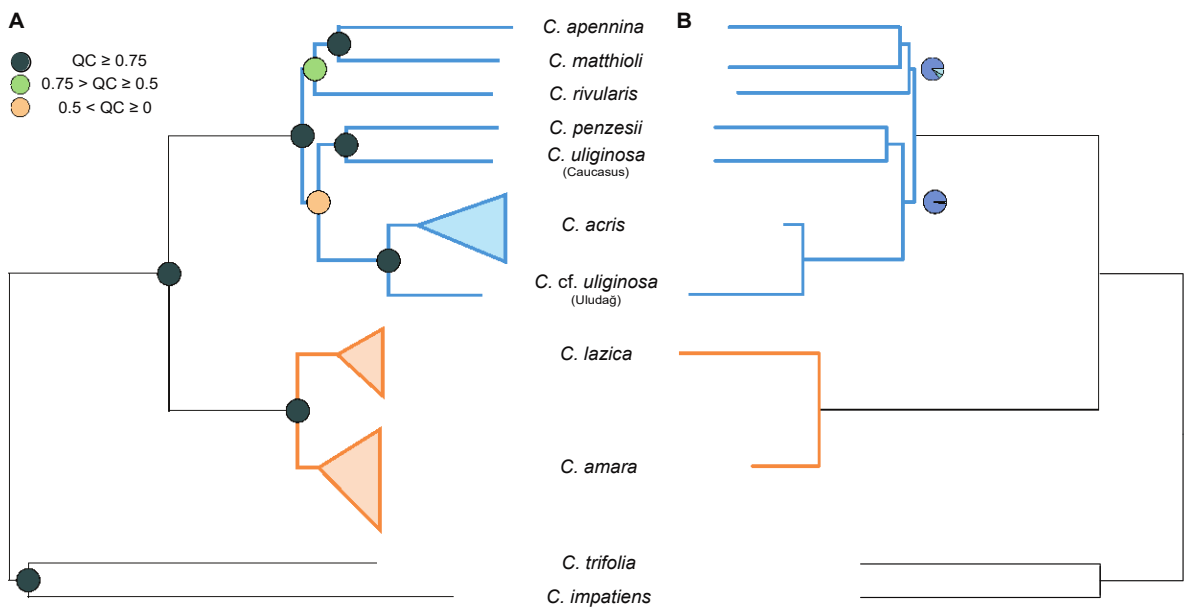


FIGURE 3 | Phylogenetic trees inferred from the complete dataset of 1,168 nuclear genes, based on consensus supercontig sequences of diploid Cardamine accessions. Orange color is used for the clades and branches of the Cardamine amara group, whereas blue is used for the remaining diploids resolved in the sister position. (A) Maximum likelihood tree inferred in RAxML-NG from the concatenated genes. Multiple individuals per species are shown collapsed (see Supplementary Figure 1 for the fully labeled version of the tree including bootstrap support). Branch support is indicated by quartet concordance (QC) values (colored circles in the nodes). (B) Species tree inferred in ASTRAL-III. Branch support is indicated by pie charts, depicting three local posterior probabilities for the given branch (dark blue for the main topology as resolved here and light blue for the alternative ones; not shown for the fully supported branches when the local posterior probability for the present topology equals 1).

(Caucasus vs. Uludağ) appeared clearly differentiated in all datasets (including nrDNA and cpDNA data, see below), they were treated as two distinct entities in all multispecies coalescent methods. The species trees inferred using ASTRAL from 1,168 ML gene trees, based either on consensus sequences (**Figure 3B**) or phased allele sequences (results not shown), showed identical topologies and branch support. These trees were also fully congruent with the ML tree of the concatenated

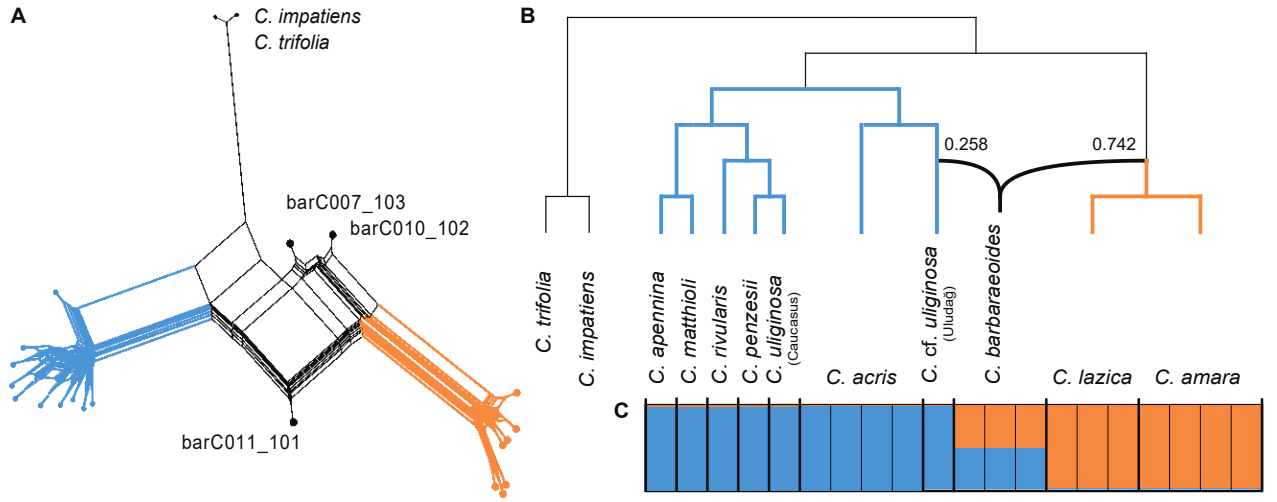


FIGURE 4 | Phylogenetic analyses indicating the hybrid (allopolyploid) origin of the tetraploid Cardamine barbaraeoides. Orange color is used for the accessions from the Cardamine amara group (C. amara and Cardamine lazica), blue for the remaining diploids, and black for the tetraploid C. barbaraeoides. (A) Supernetwork representation of quartets generated in SuperQ, which was derived from 1,168 nuclear gene trees estimated in RAxML and based on consensus supercontig sequences. (B) Species network inferred in PhyloNet from the same set of 1,168 nuclear gene trees. Inheritance probabilities are shown along the branches indicating the origin of C. barbaraeoides. (C) Bayesian clustering of SNP datasets in STRUCTURE at optimal K = 2, obtained from variant calling in a selection of 947 most informative genes. The coloring in the graph indicates the sample assignment to the two genetic clusters. Thick vertical lines separate different species.

dataset. Two branches that received lower QC values in the ML tree, congruently, showed slightly decreased local posterior probabilities in the species trees.

The Tetraploid Genome of C. barbaraeoides: Insights From Target Nuclear Loci

Displaying Conflict: Network Analyses Based on Consensus Sequences and Bayesian Clustering of SNP Variation

The SuperQ network derived from 1,168 ML gene trees based on consensus sequences displayed two well-differentiated groups of diploid taxa (corresponding to the two major clades as resolved above) and strong conflict in the placement of the tetraploid accessions (**Figure 4A**). The species network analysis (PhyloNet) based on the same set of ML gene trees suggested a hybrid origin of C. barbaraeoides as well, with one ancestral lineage from the clade of the C. amara group (comprising C. amara and C. lazica) indicating a greater inheritance probability (74.2%) and the other pointing to the C. cf. uliginosa accession from the Uludağ Mts. (25.8%), which was sister to C. acris (**Figure 4B**). Interestingly, some of the repeated PhyloNet runs indicated a reticulation event also for C. penzesii, involving C. cf. uliginosa from the Uludağ and the Caucasus as the two ancestors (**Supplementary Figure 2**).

Single-nucleotide polymorphisms calling utilized 947 genes, which harbored at least 10 SNPs across the samples. STRUCTURE analyses of 500 SNP datasets (each with one SNP randomly drawn per gene) identified the optimal genetic partitioning at K = 2, with the same two clusters of diploid taxa as identified in the trees above, whereas significant genetic admixture was observed in the tetraploid C. barbaraeoides (**Figure 4C**). Thus, all these analyses showed strong conflict in the consensus supercontig sequences of the tetraploid and suggested an allopolyploid origin of C. barbaraeoides, its progenitors being derived from the two major clades of diploids.

Identification of Parental Progenitors: Gene Tree and Species Tree Reconstructions Based on Phased Allele Sequences

Read-backed phasing yielded two alleles per exon for diploids and four alleles for tetraploids. In diploids, the level of heterozygosity varied widely from 10.28% to 51.34% (34.01% on average). Allele phasing in the tetraploid C. barbaraeoides yielded similar results among the samples. Homozygous (10.02% on average), fully heterozygous (13.5%), and partially heterozygous exons with two different alleles in the ratio 1:3 (8.4%) were relatively rare, while partially heterozygous loci with two different alleles in the ratio 2:2 (21.54%) and especially those with three different alleles (46.53%) were much more frequent (**Supplementary Figure 3** and **Supplementary Table 1**). The complete set of 1,829 targeted exons of C. barbaraeoides, each phased to four alleles, was further processed to allele sorting.

The optimized threshold for allele sorting invalidated 47.64% sequences of C. barbaraeoides, which could not be sorted unequivocally. They definitely regarded the homozygous exons and partially heterozygous one (those with the alleles in the ratio 1:3) and part of the other heterozygous exons (**Supplementary Table 1**). Alleles from all three samples of C. barbaraeoides were successfully attributed to the A and B homeologs only in 612 exons (33.46%), but on the other hand, more than 70% of exons (1,287) kept at least one sample with successfully sorted alleles

and thus held at least partial information available for coalescent-based tree reconstruction. At the gene level (with concatenated exons), attempts to sort the alleles into two different homeologs succeeded in 38.13% of sequences. Alleles from all three samples of C. barbaraeoides were successfully attributed to A and B homeologs in 274 genes (23.46%), and those from at least one sample were present in 621 genes (53.17%).

Subsequently, for species tree inferences in ASTRAL, we assembled multiple datasets that were derived from phased exon- and gene-based alignments. For exons, they included the following: No. 1, a dataset comprising all 1,829 exons with zero to three tetraploid accessions retained for each exon (i.e., a dataset with missing accessions allowed); No. 2, a dataset comprising 974 exons each with at least two tetraploid accessions (a dataset allowing at most one accession missing); and No. 3, a dataset comprising 612 exons, in which all three tetraploid accessions were retained for each exon. The species trees inferred from all three datasets recovered the same topology and differed only in some branch support values (**Figure 5A**, **Supplementary Figures 4A–C**). As for the diploid taxa, the topology was largely congruent with that of the trees derived from the diploid sequence data only (**Figure 3**, see above), differing only in the placement of the species pair C. penzesii–C. uliginosa from the Caucasus. The position of this species pair, however, received a relatively low QC value in the tree of diploids (**Figure 3A**). The A homeolog of C. barbaraeoides was resolved in a sister position to the C. amara clade, comprising C. amara and C. lazica. The B homeolog of C. barbaraeoides was placed in a sister position to the clade consisting of C. acris and C. cf. uliginosa from the Uludağ (**Figure 5A**).

Similarly, as for the exons, three datasets of phased gene-based alignments were assembled: No. 1, a dataset comprising all 1,168 genes with zero to three tetraploid accessions retained for each gene; No. 2, a dataset comprising 441 genes each with at least two tetraploid accessions; and No. 3, a dataset comprising 274 genes, in which all three tetraploid accessions were retained for each gene. The species trees recovered the same topology for all three datasets, with differences only in branch support (**Figure 5B**, **Supplementary Figures 4D–F**), and were almost identical to those inferred from exon-based data. The only difference was in the placement of the A homeolog of C. barbaraeoides, which was resolved here in a sister position to C. amara (and not to the whole C. amara clade as above in exon-based trees).

When computing distances between the alleles retrieved from C. barbaraeoides and successfully sorted into A and B homeologs and the alleles of each diploid species, it becomes apparent that the A homeolog is closest to C. amara alleles, tightly followed by those of C. lazica, whereas the B homeolog is closest, almost equally, to the alleles of C. cf. uliginosa from the Uludağ and those of C. acris (**Supplementary Figure 5**).

Alternative Topology Testing: GGI Analyses

Topology tests based on the GGI analyses yielded robust and highly congruent results both from the exon- and gene-based datasets, when considering the set of trees in which alleles from all three accessions of C. barbaraeoides were present (i.e., successfully phased and sorted, 612 exons or 274 genes). The GGI results clearly favored the topology in which C. barbaraeoides homeolog A was resolved in a sister position to the clade of the C. amara group (**Figure 5**). This topology was significantly supported by a greater number of genes and exons than the alternative topologies ($P < 0.05$) and agrees also with the exon-based ASTRAL species tree. Two alternative topologies, i.e., with C. barbaraeoides homeolog A being sister to either C. amara (as seen on the gene-based species tree, **Figure 5B**) or C. lazica, received much less support. As for the placement of the B homeolog of C. barbaraeoides, the GGI analyses favored the topology in which C. barbaraeoides was placed in a sister position to the clade comprising C. acris and C. cf. uliginosa from the Uludağ, in accordance with the ASTRAL species trees. The second topology, with C. barbaraeoides being sister to C. acris, was significantly supported by a much smaller number of trees, followed by the third topology (C. barbaraeoides sister to C. cf. uliginosa from the Uludağ, suggested by PhyloNet) with only negligible support (**Figure 5**).

Slightly different and also equivocal GGI results in some cases were obtained when including also the exons or genes, in which one or two accessions of C. barbaraeoides were missing (i.e., one individual kept at minimum) because of failed allele sorting (1,287 exons or 621 gene in total). In those datasets, the two topologies with C. barbaraeoides homeolog B being sister either to C. acris or to the clade of C. acris and C. cf. uliginosa from the Uludağ received similar support, and none of them could be strongly favored over the other (**Supplementary Figure 4**). The placement of homeolog A in the dataset of 1,287 exons also remained equivocal, with similar support given for its sister position to either C. amara or to the clade of the C. amara group (comprising also C. lazica). In the dataset of 621 genes, the same topology for C. barbaraeoides homeolog A was favored as in the dataset of 274 genes (**Supplementary Figure 4**).

Analyses of nrDNA Polymorphisms Obtained From Molecular Cloning and Genome Skim Data

The ITS alignment obtained from molecular cloning was 623 bp long and comprised 180 sequences from 48 ingroup individuals. It contained 209 variable sites (33.5%) and 99 parsimony-informative sites (15.9%). High intraspecific and even intraindividual diversity of the ITS variants (ribotypes) was revealed in the diploid taxa (**Supplementary Data Sheet 1**). Nevertheless, the ribotypes observed within individuals and within species were mostly similar and clustered together, with the exceptions of rare divergent ribotypes found in a single accession of C. acris (C015-107) and C. penzesii (DEM7) (**Supplementary Figure 6**). In accordance with the data from the target nuclear loci, genetic differentiation was observed within C. uliginosa; ribotypes from the Uludağ samples were nested within the diversity of C. acris, whereas those from the Caucasus appeared closest to C. penzesii or C. matthioli (**Figure 6**, **Supplementary Figure 6**). In the tetraploid C. barbaraeoides, the vast majority (approximately 78%) of ITS sequences were placed within the C. amara clade. Three ribotypes (i.e., 4.6%) of C. barbaraeoides (found in three different accessions),

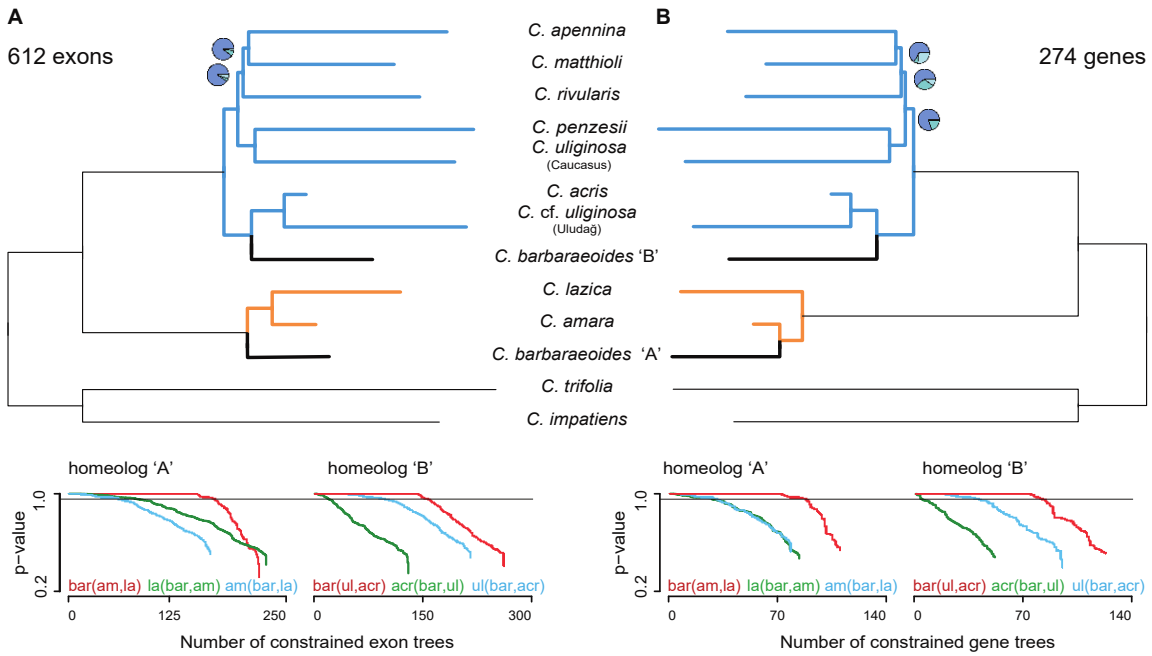


FIGURE 5 | Phylogenetic analyses based on phased allele sequences of 612 exons (A) or 274 genes (B), for which alleles of all three accessions of Cardamine barbaraeoides were successfully phased and sorted into A and B homeologs. Species trees were inferred in ASTRAL-III. Branch support is indicated by pie charts, depicting three local posterior probabilities for the given branch (not shown for the fully supported branches). Tree branches are colored according to the group membership; branches in black show C. barbaraeoides. The plots below the trees show GGI analyses, testing different phylogenetic placements of the A and B homeologs of C. barbaraeoides. The plots depict the cumulative number of constrained gene trees, which support the given topology, and their P values obtained from the approximately unbiased (AU) tests. Curves above the dashed lines indicate the number of trees that support the given topology significantly better (P 0.05) than the alternative ones. The tested topologies were as follows: homeolog A, red: C. barbaraeoides "A" being sister to the clade of Cardamine amara and Cardamine lazica, green: C. lazica being sister to the clade of C. amara and C. barbaraeoides "A", blue: C. amara being sister to the clade of C. barbaraeoides "A" and C. lazica; homeolog B, red: C. barbaraeoides "B" being sister to the clade of Cardamine acris and Cardamine cf. uliginosa from the Uludag Mts., green: C. acris being sister to the clade of C. barbaraeoides "B" and C. cf. uliginosa from the Uludag Mts., blue: C. cf. uliginosa from the Uludag Mts. being sister to the clade of C. barbaraeoides "B" and C. acris. Species trees and GGI plots from the alternative datasets allowing for missing tetraploid accessions are presented in Supplementary Figure 4.

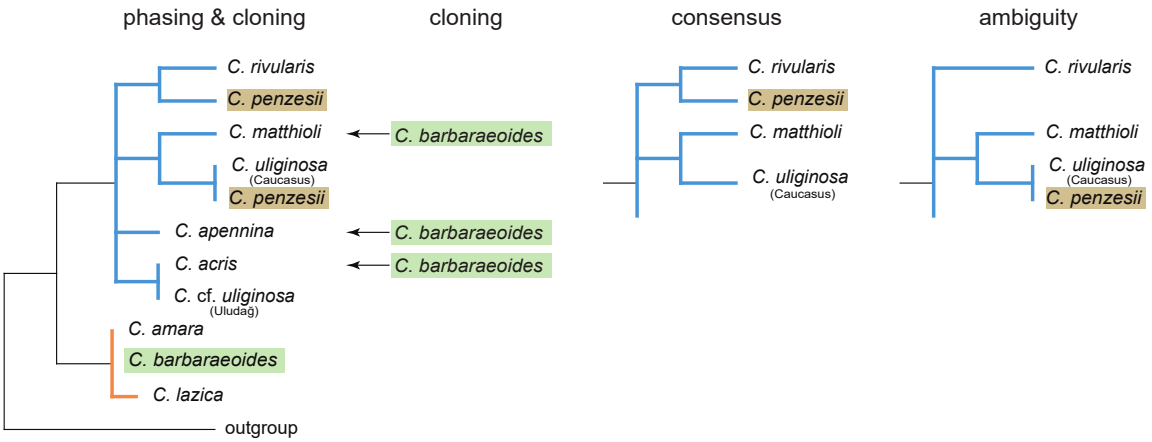


FIGURE 6 | Schematic representation of maximum likelihood (ML) trees inferred from four datasets of nrDNA (ITS1-5.8S-ITS2 region) and their topological differences. Cloning: dataset obtained from molecular cloning and Sanger sequencing; phasing: dataset obtained from read-backed phasing of Hyb-Seq reads, with multiple nrDNA variants retrieved per sample; consensus: consensus assembly (with the majority rule base calls) from Hyb-Seq reads; ambiguity: ambiguous assembly from Hyb-Seq reads with intraindividual SNPs replaced by IUPAC codes. See Supplementary Figure 6 for the complete ML trees obtained from these datasets.

however, were clearly divergent and clustered closest to C. acris, C. matthioli, or C. apennina (**Figure 6**, **Supplementary Figure 6**). The rest of the ribotypes (17.4%) showed recombinant patterns between the two major clades (not included in the ML tree).

The ITS alignment from the consensus assembly of the reads mapping to the reference sequence comprised 20 ingroup sequences with 53 variable (8.5%) and 30 parsimony-informative sites (4.8%). The alignment of the ambiguous assembly contained 68 ambiguous bases and 43 variable (6.9%) and 28 parsimony-informative sites (4.5%). Read-backed phasing of the assembled ITS sequences resulted in 1 to 4 ITS variants per individual, and the alignment comprised 47 different ingroup sequences with 77 variable (12.4%) and 57 parsimony-informative sites (9.2%). The topologies of the ML trees obtained from the consensus and ambiguous datasets were largely congruent (**Figure 6**, **Supplementary Figure 6**), except of the position of C. penzesii. In the consensus dataset, C. penzesii was resolved as sister to C. rivularis, whereas in the ambiguous dataset it was placed sister to C. uliginosa from the Caucasus. The former topology agreed with the position of all but one ribotype resolved in C. penzesii by molecular cloning, whereas the latter corresponded to the position of one divergent ribotype revealed in this species. In both the consensus and ambiguous datasets, the tetraploid C. barbaraeoides was placed within the C. amara clade. Phasing revealed the presence of divergent nrDNA variants in both C. acris (accession C015-107) and C. penzesii (DEM7) that were placed outside of the respective species-specific clades, being in congruence with the cloned data (**Figure 6**, **Supplementary Figure 6**). In the tetraploid C. barbaraeoides, by contrast, with phasing using GATK and WhatsHap tools as described above, we were able to retrieve only nrDNA variants corresponding to the C. amara sequence types. The rare ribotypes clustering with C. acris, C. matthioli, or C. apennina as found by cloning could not be successfully extracted, although visual exploration of the genomic data (using IGV; Robinson et al., 2011) confirmed the presence of a low proportion SNPs (approximately 10%) suggesting that these rare sequence variants are indeed present in the genome of C. barbaraeoides.

Analyses of Chloroplast Genome Data

The alignment of the complete LSC, SSC, and IRb regions was 128,344 bp long. The alignment of the concatenated annotated genes was 96,838 bp long and included 74 protein-coding genes and 31 tRNA and four rRNA genes. The ML trees inferred from the two alignments showed high congruence (**Supplementary Figure 7**). Topological differences were found only in clades that displayed very short branches and low BS support. Two major clades with high BS were retrieved in both ML trees, which corresponded to those resolved by nuclear genes. One comprised C. amara and C. lazica in a sister position, which were successively sister to C. barbaraeoides. The other major clades in the ML trees comprised two well-differentiated and supported subclades. One subclade consisted of C. acris (three out of four accessions) in a sister position to C. cf. uliginosa from the Uludağ, in concordance with the topology retrieved from nuclear genes. The other subclade comprised C. apennina, C. penzesii, one accession of C. acris (C015), C. rivularis, and C. matthioli. Except of the last two species, resolved in a well-supported sister position, the relationships within this subclade received only low support and differed between the two cpDNA datasets.

The ASTRAL species tree based on 42 most variable protein-coding chloroplast genes (**Figure 7A**) showed high congruence with the ML trees inferred from the concatenated alignments (**Supplementary Figure 7A**). Two internal branches, which determined the positions of C. penzesii and C. uliginosa from the Caucasus, received low local PP values that imply low support for the given topology. This agrees with the topological differences between the ML trees from the concatenated data and thus suggests that the placement of these two species is uncertain.

Genomic in situ Hybridization

DAPI staining of mitotic chromosomes in C. barbaraeoides revealed 16 bigger (L) chromosomes with more extensive pericentromeric heterochromatin that were readily discernible from the other 16 smaller (S) chromosomes (**Figure 8A**). The L chromosomes carried terminal heterochromatin knobs (**Figure 8B**), which were previously reported in the C. pratensis group (Mandáková et al., 2013). The gDNA probes of the three tested accessions from the C. amara clade (C. amara subsp. amara, subsp. balcanica, and C. lazica) hybridized on 16S chromosomes of C. barbaraeoides; the signal strengths of all three probes were comparable. The gDNA probes of the other five accessions tested (C. acris subsp. acris, C. matthioli, C. penzesii, C. rivularis, and C. uliginosa from the Caucasus) hybridized on 16L chromosomes of C. barbaraeoides. Although quantification of hybridization signals is problematic, we observed stronger fluorescence of the gDNA probes of C. acris compared to the other probes tested (**Figure 8A**, **Supplementary Figure 8**). Thus, GISH data suggest that C. barbaraeoides is an allotetraploid that originated via hybridization between members or recent ancestors of the C. amara clade and the other major clade, where C. acris appeared to be the most likely parental candidate. These two major groups of species differ in their genome size (see above in Chromosome Numbers and Genome Size Variation and **Figure 2**). This difference is reflected by bigger chromosomes, more pericentromeric heterochromatin, and terminal heterochromatic knobs within the C. acris–like subgenome, in contrast to smaller chromosomes within the C. amara–like subgenome of C. barbaraeoides (**Figure 8A**).

DISCUSSION

Evolutionary Relationships and Polyploid Speciation in Balkan Cardamine: Evidence From Phylogenomic and Cytogenetic Data

Uncovering phylogenetic relationships within recently diverged plant groups can be challenging even at the diploid level. Persistence of ancestral polymorphisms, low genetic divergence between species, and both past and contemporary interspecific gene flow hamper robust phylogenetic inferences (Naciri and Linder, 2015; see, e.g., Blanco-Pastor et al., 2012; Krak et al., 2013;

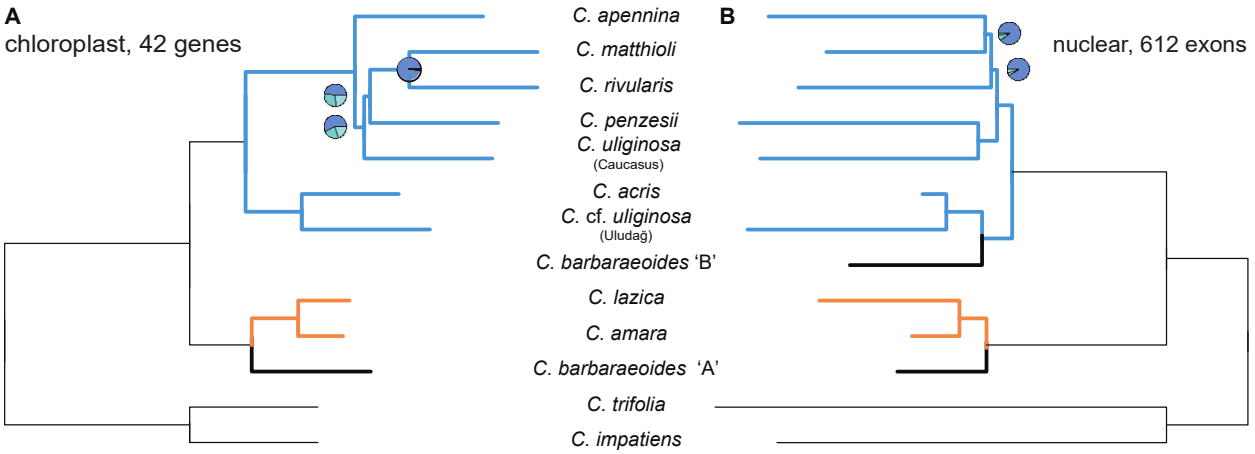


FIGURE 7 | Species trees inferred in ASTRAL-III from the dataset of 42 protein-coding most informative chloroplast genes (A) and 612 nuclear exons [(B), as in Figure 5A]. Branch support is shown by pie charts, depicting three local posterior probabilities for the given branch (not shown for the fully supported branches). Tree branches are colored according to the group membership; branches in black show Cardamine barbaraeoides.

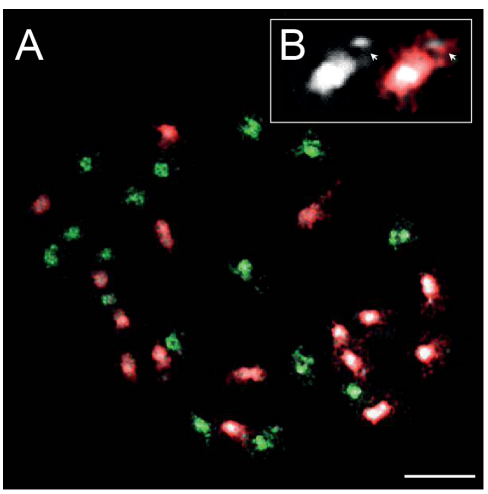


FIGURE 8 | Genomic in situ hybridization (GISH) on mitotic chromosomes in the allotetraploid Cardamine barbaraeoides (2n = 32). (A) GISH with total genomic DNA of Cardamine acris subsp. acris (red fluorescence) and Cardamine amara subsp. balcanica (green fluorescence) revealed two subgenomes contributed by ancestors of the two diploid species. (B) Close-up view of a heterochromatic knob-bearing chromosome (the terminal knob marked by arrowheads). Chromosomes were counterstained by 4^0,6-diamidino-2-phenylindole (DAPI); scale bar, 10 mm.

Konowalik et al., 2015). In the Cardamine species complexes studied here, previously applied ITS and noncoding cpDNA Sanger sequences showed a low level of sequence polymorphism, as well as conflicting phylogenetic signal (Lihová et al., 2004a; Marhold et al., 2004). AFLP fingerprinting proved to be efficient at delimiting and describing species, in concordance with morphological, ploidy level, and distribution patterns (Lihová et al., 2003, 2004b), but performed poorly in phylogenetic inference (Marhold et al., 2004).

In the present study, we applied a target enrichment approach, recently shown to provide high resolution also at low phylogenetic levels between the closest relatives (Villaverde et al., 2018; Carter et al., 2019; Tomasello et al., 2020). Indeed, using custom, genus-specific probes, we were able to retrieve sequences from more than 1,000 nuclear genes from each sample. At the diploid level, the topologies of the ML tree obtained from concatenation of targeted loci and the coalescent-based species tree were fully congruent, which suggests a low degree of ILS, in accordance with high support in the species tree (**Figure 3**). Recently, it has been emphasized that allele phasing should be preferred to the use of consensus sequences (or "contig", sensu Andermann et al., 2019) ignoring heterozygous positions and allelic variation, as it can improve phylogenetic inference and yield a more accurate tree estimate especially in recently diverged species (Andermann et al., 2019; Tomasello et al., 2020; but see Kates et al., 2018). Using either consensus or phased allele sequences, here we obtained the same species tree topology, which additionally supports the robustness of our data. Some topological conflicts, however, appeared between the nuclear-and plastome-derived phylogenetic trees (**Figure 7**). This plastid– nuclear discordance among the diploids, described in more detail below, can identify traces of interspecific gene flow and thus shed further light onto the evolutionary histories of Cardamine species. With more extensive sampling in the future, including all representatives of the studied species complex across Europe, this approach has great potential to infer their phylogeny comprehensively. Here we provide our first insights from the perspective of Balkan species.

In accordance with previous phylogenetic studies (Lihová et al., 2004a; Marhold et al., 2004; Carlsen et al., 2009), C. amara was supported here as a distinct phylogenetic lineage separated from the taxa classified within the other two species complexes (C. pratensis and C. raphanifolia groups). C. amara is a widespread and polymorphic Eurasian species

consisting of several subspecies in Europe (Lihová et al., 2004a). Cardamine lazica, a species from the Pontic mountains and western Caucasus, was identified here as a sister species to C. amara, as already suggested by AFLPs, but not sufficiently resolved previously by Sanger sequence data (under the name C. wiedemanniana, Lihová et al., 2004a; Marhold et al., 2004). Furthermore, the present data supported the monophyly of C. acris, the most widely distributed Balkan endemic with extensive morphological and genetic variation and three subspecies recognized (Perný et al., 2004). One accession of C. acris (C015-107) was misplaced in the plastome tree (**Supplementary Figure 7**), which is most likely a sign of interspecific hybridization or introgression, a scenario supported also by a mixture of divergent nrDNA variants found in this individual (**Supplementary Figure 6**). Furthermore, both nuclear and plastid data congruently revealed a sister species relationship between C. acris and the population from the Uludağ Mts. in northwestern Turkey, classified as C. cf. uliginosa. On the other hand, the population of C. uliginosa from the Caucasus was genetically divergent. C. uliginosa grows across Anatolia (but with very scarce records from its western parts except for the Uludağ Mts., J. Kučera, pers. comm.) and the Caucasus, extending further to the south and south-east, reaching the mountains of Iran and Lebanon. Previous studies have already indicated that it is a highly polymorphic species (Marhold et al., 2004), maybe even a complex of (cryptic) species.

The relationships between the other four species, C. apennina, C. matthioli, C. penzesii, and C. rivularis, all traditionally classified within the C. pratensis group, showed conflicting patterns between nuclear and plastid trees but also low support (**Figure 7**). Both ongoing and past gene flows, the latter probably facilitated by range shifts in glacial-interglacial periods, have been inferred to occur between C. matthioli and C. rivularis in Bulgaria (Ančev et al., 2013; Melichárková et al., 2020), which may explain their close position in the plastid tree, in contrast to the nuclear tree. The lowland species C. penzesii was resolved as sister to C. uliginosa from the Caucasus in nuclear trees, whereas the positions of both species remained uncertain in plastome trees. PhyloNet analyses of nuclear loci, which account for both ILS and interspecific gene flow (Wen et al., 2018), as well as the presence of divergent nrDNA variants, suggested a reticulate evolutionary history of C. penzesii (**Supplementary Figures 2**, **6**).

Our Hyb-Seq and GISH results provide strong evidence that C. barbaraeoides, a stenoendemic of the Southern Pindos Mts., is of allotetraploid origin. Interestingly, the phylogenetic placements of its homeologs do not favor a very recent (i.e., postglacial) origin, as might have been suspected from its narrow range within the area occupied by C. acris. Alleles retrieved from two subgenomes appeared differentiated from those observed in present-day diploids, suggesting that the parental species of C. barbaraeoides were most likely the common ancestors of C. amara and C. lazica on one side (the maternal one, as inferred from cpDNA) and of C. acris and a western Anatolian taxon (so far attributed to C. uliginosa) on the other (**Figures 5**, **7**). A possible alternative scenario is that extensive genomic changes in the polyploid in response to a "genomic shock," including nonhomologous recombination, have significantly altered and differentiated the polyploid genome from its diploid progenitors (Nieto Feliner and Rosselló, 2012; Madlung and Wendel, 2013). Still, the former hypothesis of an older allopolyploidization event and a relict character of this species is favored also by the plastome tree, which confirmed the same phylogenetic placement of C. barbaraeoides as was revealed for the "A" homeolog in the nuclear species trees. The phylogenetic patterns also agreed with the strength of GISH signal, where both C. amara and C. lazica probes hybridized well on S chromosomes of C. barbaraeoides, whereas C. acris probes hybridized stronger on L chromosomes (Uludağ accessions were not available) than the other diploids analyzed (**Supplementary Figure 8**).

Based on a recently published tribe-wide dated phylogeny (Huang et al., 2020), we can infer early- to mid-Pleistocene divergence among the diploids analyzed here within both major clades, which suggests also the approximate age of this allopolyploid. The highly restricted occurrence of C. barbaraeoides at only a few sites within the small range of the Lakmos Mts. (S Pindos) remains intriguing. We may speculate whether the present occurrence is only a remnant of a previously wider range, or whether the allopolyploidization event took place within the current area and the species never expanded much beyond it. The former hypothesis appears more plausible when we reject the species' very recent origin and consider also evidence that Mediterranean mountains have experienced significant changes in vegetation, habitat availability, and diversity during Quaternary climatic oscillations (Médail and Diadema, 2009; Nieto Feliner, 2014).

Drivers of Speciation Within the Cardamine Species Complexes: The Role of Mountains of the Balkan Peninsula and Adjacent Biogeographic Regions in Shaping Diversity and Endemism Patterns

The Cardamine taxa under study exhibit parapatric to allopatric distributions, and all occupy similar wet habitats, partly with different elevational preferences (**Figure 1**). From the presented phylogenetic reconstructions, we can infer that they likely evolved via both allopatric and ecological speciation processes, which have also been affected by interspecific gene flow. The studied species complexes comprise numerous endemics not only in the Balkan Peninsula but also across the other parts of the Mediterranean (Marhold et al., 2018). The prevalence of endemics in the Mediterranean, along with the commonly observed geographically structured genetic variation within several species (Lihová et al., 2003; Perný et al., 2005a,b; Melichárková et al., 2020), suggests that geographic isolation played an important evolutionary role. These patterns may reflect past range fragmentation in response to Pleistocene climatic oscillations (Nieto Feliner, 2014), as well as spatially restricted gene flow and species dispersal, which may be the two principal causes, acting in concert, of the current high endemism rate in these species complexes. Lowland-alpine

species pairs, such as C. penzesii and C. uliginosa, also show signs of ecological speciation as a result of adaptation to habitats at high elevations, typically with higher precipitation and solar radiation input, and lower temperatures, as proven for C. amara subsp. amara and subsp. austriaca in the Eastern Alps (Zozomová-Lihová et al., 2015). Ecological niche analyses in four species of the C. pratensis complex growing from lowlands up to the alpine belt in Central and Southeastern Europe (including C. rivularis and C. matthioli studied here; Melichárková et al., 2020) found niche shifts and niche breadth differences, but still considerable niche overlaps among species, representing both sympatric and allopatric cases. It appears that divergent ecological requirements may play a certain role in the evolution of these species complexes but probably do not constitute a strong constraint that would significantly hamper range expansion and explain the high incidence of endemics.

With the present results, we provide additional support for the prominent role of Mediterranean mountains both as cradles and reservoirs of species and genetic diversity and, more specifically, for the contribution of polyploid speciation to the origin of biodiversity hotspots. Indeed, the Southern Pindos range, the area of C. barbaraeoides, is recognized as an important center of endemism and also a refugial area (Stevanović et al., 2007; Médail and Diadema, 2009; Kougioumoutzis et al., 2021). Quaternary climatic oscillations have led to species range shifts, repeated range fragmentation, and reduction followed by expansion, and these processes have facilitated contacts between previously isolated lineages and brought opportunities for hybridization (Nieto Feliner, 2014; Marques et al., 2018; see, e.g., Blanco-Pastor et al., 2012; Maguilla et al., 2017; Zozomová-Lihová et al., 2020). The great ecological and topographic heterogeneity of Mediterranean mountains has likely favored not only hybridization events, but also the establishment and persistence of newly formed allopolyploids. Several examples of polyploid endemics confined to some mountains of the Balkan Peninsula (e.g., Cires et al., 2014; Olšavská et al., 2016; Španiel et al., 2017; López-González et al., 2021) suggest that allopolyploid speciation may significantly contribute to the diversity of the Balkan endemic mountain flora, but this topic is still poorly explored, and further studies are needed.

Our present study revealed cases in which Balkan taxa have their phylogenetically closest counterparts in the Anatolian or Caucasus regions, in support of the known biogeographic links between these areas (Strid, 1986; Bilgin, 2011; Thompson, 2020). Indeed, the Anatolian phytogeographic element is well represented in the Greek mountain flora, and this is particularly true for species distributed in the Uludağ Mts. in northwestern Turkey (Strid, 1986). The Aegean Sea, the Sea of Marmara, and the Thracian Plain are significant barriers to mountain species dispersal between the Balkan Peninsula and Anatolia at present (Ansell et al., 2011; Bilgin, 2011). However, they may have been penetrated especially in colder periods at the Pliocene–Pleistocene transition and during Pleistocene glaciations (Strid, 1986; Ansell et al., 2011). One of the common phylogeographic patterns recognized in Anatolia suggests a genetic break within Anatolia, differentiating populations in western Anatolia and the Balkan Peninsula from those in eastern Anatolia (Bilgin, 2011). This pattern resembles the present case of high affinity between C. acris and the population from the Uludağ Mts.; however, more detailed studies of C. uliginosa across its distribution range are needed. Furthermore, closer evolutionary relationships and traces of hybridization between C. penzesii from flood-plain forests near the Black Sea coast and high-mountain C. uliginosa from the Caucasus seem to support the Northern Anatolian dispersal corridor (Kaya and Çiplak, 2017; Özüdoğru and Mummenhoff, 2020). Northern Anatolia may have provided sites ecologically suitable for both lowland and mountain population survival in close proximity and allowed for allopatric, as well as ecological speciation (Kučera et al., 2006; Roces-Díaz et al., 2018).

Resolving Allopolyploid Origins From Hyb-Seq Data and Potential of nrDNA Polymorphisms for Detecting Reticulate Evolution

The employment of low-copy nuclear genes in phylogenetic studies, especially when polyploids are involved, is crucial. Nuclear genes show biparental inheritance and typically retain evidence of a reticulate history (e.g., Brysting et al., 2007; Rousseau-Gueutin et al., 2009; Brassac and Blattner, 2015). Still, it is known that individual gene trees may show discordant histories that do not match the true evolutionary history, because of various processes related to the complexity of nuclear genomes, such as high allelic variation and ILS, nonhomologous recombination, gene duplication, and gene loss (Maddison, 1997; Small et al., 2004; Degnan and Rosenberg, 2009). Therefore, the use of multiple unlinked loci has been strongly advised (Naciri and Linder, 2015). Target enrichment techniques that may capture hundreds of unlinked orthologous loci are promising in resolving the origins and evolutionary histories of polyploid species with much greater confidence (Kamneva et al., 2017). Assembly of short sequence reads, however, remains a challenge for allopolyploid genomes, because a mixture of reads belonging to both homologous and homeologous loci is obtained (Kyriakidou et al., 2018). Most phylogenetic studies have used consensus assembly (e.g., Crowl et al., 2017; Morales-Briones et al., 2018; Carter et al., 2019), that is, a single majority sequence at a given locus. This means, however, that sequences from different homeologs (parental subgenomes), as well as chimeric sequences, may be retrieved. Allopolyploid speciation is then commonly inferred by network analyses, which account for both ILS and hybridization (Crowl et al., 2017; Morales-Briones et al., 2018; Carter et al., 2019).

In the present study, we employed the network analyses based on the consensus sequences, which, in congruence with the SNP data analyses, identified conflicting signal within the data and suggested allotetraploid origin of C. barbaraeoides. Nevertheless, as a significant step further, we proceeded to allele assembly and sorting. Some approaches or tools for assembling allele sequences and distinguishing among homeologs have recently been proposed for polyploids (Page et al., 2013; Kamneva

et al., 2017; Rothfels et al., 2017; Schrinner et al., 2020; Rothfels, 2021). Several previous studies used parallel amplicon sequencing to analyze polyploid species, but capturing only a low number of loci (up to 12 loci) and with manual homeolog identification and sorting (Brassac and Blattner, 2015; Rothfels et al., 2017; Dauphin et al., 2018; see also Eriksson et al., 2018, specifically for target enrichment data). Here we propose a novel approach in which hundreds of loci obtained from target enrichment techniques can be analyzed simultaneously and allele sorting does not require manual inspection and labeling. We inferred phased alleles based on available tools and developed a bioinformatics procedure to sort them into homeologs. Allele sorting is based on calculating distances between alleles, obtained from branch lengths of corresponding gene trees, first between alleles from a given polyploid (to identify allele pairs) and then from its diploid relatives. Homeolog labeling is based on allele pair distances to the suspected maternal species, as identified by plastome analyses. The phylogenetic positions of the obtained homeologs, representing two parental subgenomes in the polyploid, are then explored by a species tree inference. This approach is most straightforward when the maternal species is at least approximately determined, but could be applied even if this information is unknown, in the case of missing cpDNA data, a possibly extinct or an unsampled maternal parent. Under such scenarios, one of the most closely related species, a possible progenitor of the investigated polyploid, could be identified from the network analyses inferred from the consensus sequences and subsequently used for homeolog labeling.

Two shortcomings may potentially limit the efficiency of our approach. One is specific to the target loci and/or species studied. Successful allele sorting in polyploids, namely, depends on both parental genome divergence and the informativeness (phylogenetic signal) of target loci. Alleles from some genes may not be unequivocally sorted into homeologs, because of low phylogenetic signal and low sequence divergence. Still, when employing a large set of target loci during sequence capture and including also more variable flanking intronic and intergenic regions (as is achieved via the Hyb-Seq approach; Weitemier et al., 2014), sufficient data and resolution can be obtained. Here we demonstrate that with several reduced datasets, allowing either missing accessions or loci, we obtained the same topologies of the species trees, and the same allopolyploid scenario was inferred. The second obstacle is related to the short length of sequence reads obtained from the Illumina platform, which throws down a challenge to allele phasing software. The shorter length of sequence reads more often causes sequence splitting into multiple phase blocks. Variant sites are phased with other sites within the given block but cannot be phased with respect to variants in the other blocks because of insufficient read data between the blocks (see Kates et al., 2018). If multiple phase blocks occurred, phased alleles were retained only in the largest phase block, and the remaining intraindividual variants were masked (12.92% of SNPs per sample in average). Concatenation of exon sequences to genes has a dual (partially contradicting) effect. The sequence length has a positive effect on the resolution of the phylogenetic tree. On the other hand, concatenation involved both sorted and unsorted (with masked interallelic SNPs) exons, which means that interallelic variation was partly homogenized. To investigate the impact of this issue on phylogenetic reconstruction, we compared two datasets that differed in the length of the loci used and the amount of masked SNP variants: shorter exon-based and longer gene-based datasets. Only a single topological difference was observed between the species trees inferred from these datasets, inspected in more detail by running GGI topology tests (Arcila et al., 2017) and suggesting that this issue may be worth considering. Overall, we demonstrate that allele phasing and distinguishing homeologous copies are crucial for determining the origin of polyploids and for resolving reticulate evolution of polyploid complexes. The here proposed approach works so far for suspected allotetraploids, but future developments will focus on resolving genomes of higher ploidy levels that may be composed of more than two subgenomes (such as Cardamine occulta and Cardamine schulzii, both identified as trigenomic allopolyploids by advanced cytogenetic techniques; Mandáková et al., 2013, 2019), as well as autopolyploids.

Genome skimming, performed as part of the Hyb-Seq approach (Weitemier et al., 2014), allowed us to assemble also the high-copy nrDNA with sufficient coverage and to compare it with the variation obtained by molecular cloning. Molecular cloning found substantial intragenomic variation in most species studied, in agreement with the commonly observed patterns that concerted evolution acting in nrDNA may be incomplete (Álvarez and Wendel, 2003; Weitemier et al., 2015). Because high-throughput sequencing recovers reads from all potential repeat variants within and among nrDNA loci, we explored different possibilities how to deal with such intraindividual polymorphisms. We compared the two most commonly used coding schemes for such polymorphisms, the consensus (majority) one and the ambiguous one (Vargas et al., 2017; Fonseca and Lohmann, 2019), with phasing that enables to extract different sequence variants comparable to those obtained through cloning. Indeed, as we revealed in the cases of C. penzesii and C. acris, phasing may be an efficient way to recover phylogenetically relevant intragenomic nrDNA variation, suggesting a reticulate history, which replaces laborious cloning and PCR amplifications. On the other hand, really rare variants, such as those in C. barbaraeoides that have apparently remained as traces from its paternal progenitor, may be difficult to obtain from genome skim data and require improvements in bioinformatics tools. By contrast, with the amplicon sequencing approach, Tkach et al. (2019) were able to detect extremely rare (present in as few as 0.2% reads) ITS2 variants that indicated ancient hybridization events. Therefore, although genome skim data are easy to obtain and provide huge amounts of data from both organellar and nuclear DNA high-copy fractions, they should be considered with caution especially in groups with reticulate evolutionary histories (e.g., Vargas et al., 2017; del Valle et al., 2019; Chen et al., 2020). With the recently increasing efforts to develop target enrichment probes specific to relatively narrow focus groups (e.g., Schmickl et al., 2016; Vatanparast et al., 2018; Nikolov et al., 2019), this approach will become available for a wider spectrum of

taxa, and genome skimming may become a useful complement to, phylogenetically more robust, datasets of hundreds of independent nuclear loci.

CONCLUSION

Our study demonstrates the importance of a thorough phylogenomic approach when studying the evolution of recently diverged species complexes affected by reticulation events at both the diploid and polyploid levels. We emphasize the significance of retrieving allelic and homeologous variation from nuclear genes, as well as divergent nrDNA copy variants from high-throughput genomic data. Along with the employment of multiple analysis methods, they all, in concert, allow to resolve the origins of polyploids, detect cases of interspecific gene flow, and explain plastid– nuclear phylogenetic discordance. We suggest that despite recent advances in phylogenomic data analyses, significant improvements are needed especially in processing and analyzing sequence data from polyploid and hybrid genomes. With the present results, we also illustrate the prominent role of Mediterranean mountains as biodiversity hotspots, favoring long-term survival and speciation in allopatry, but also acting as melting pots that promote secondary contacts between species, hybridization, and polyploid evolution.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, PRJNA687126; https://www.ncbi.nlm.nih.gov/genbank/, MW476310–MW476485, MW480861–MW480862, and MW435615–MW435620.

AUTHOR CONTRIBUTIONS

JZ-L, KM, and MŠe conceived and designed study. MŠe, MP, KM, MSo, AK, and JZ-L collected plant material. AK, JZ-L, TM, MC, and MP performed laboratory work. AK, MŠe, JZ-L, and TM analyzed the data. MŠe performed bioinformatics scripting. RS contributed to bait development and Hyb-Seq protocol optimization. JZ-L and MŠe wrote the manuscript with contributions from KM and MSo. All authors have discussed, read, and commented on the manuscript.

FUNDING

This work was supported by research grants from the Slovak Research and Development Agency (APVV; grant APVV-17-0616), the Czech Science Foundation (grant GAČR 19-06632S), and the CEITEC 2020 project (grant LQ1601).

ACKNOWLEDGMENTS

We thank Jaromír Kučera (Plant Science and Biodiversity Centre SAS, Bratislava) for providing samples from Turkey and Georgia, and for collecting Cardamine amara subsp. opicii. We also thank Mária Šedivá (Institute of Chemistry SAS, Bratislava) for giving us access to the sonicator. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.659275/full#supplementary-material

REFERENCES

Álvarez, I., and Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29, 417–434. doi: 10.1016/s1055-7903(03)00208-2

Ančev, M., Yurukova-Grancharova, P., Ignatova, P., Goranova, V., Stoyanov, S., Yankova-Tsvetkova, E., et al. (2013). Cardamine rhodopaea (Brassicaceae), a triploid hybrid from the West Rhodope Mts: morphology, distribution, relationships and origin. Phytol. Balcan. 19, 323–338.

Andermann, T., Fernandes, A. M., Olsson, U., Töpel, M., Pfeil, B., Oxelman, B., et al. (2019). Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. Syst. Biol. 68, 32–46. doi: 10.1093/sysbio/syy039

Ansell, S. W., Stenøien, H. K., Grundmann, M., Russell, S. J., Koch, M. A., Schneider, H., et al. (2011). The importance of Anatolian mountains as the cradle of global diversity in Arabis alpina, a key arctic–alpine species. Ann. Bot. (Oxford) 108, 241–252. doi: 10.1093/aob/mcr134

Arcila, D., Ortí, G., Vari, R., Armbruster, J. W., Stiassny, M. L. J., Ko, K. D., et al. (2017). Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. Nat. Ecol. Evol. 1:0020. doi: 10.1038/s41559-016-0020

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. doi: 10.1089/cmb.2012.0021

Bastkowski, S., Mapleson, D., Spillner, A., Wu, T., Balvociute, M., and Moulton, V. (2018). SPECTRE: a suite of phylogenetic tools for reticulate evolution. Bioinformatics 34, 1056–1057. doi: 10.1093/bioinformatics/btx740

Bilgin, R. (2011). Back to the suture: the distribution of intraspecific genetic diversity in and around Anatolia. Int. J. Molec. Sci. 12, 4080–4103. doi: 10.3390/ijms12064080

Blanco-Pastor, J. L., Vargas, P., and Pfeil, B. E. (2012). Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean Linaria. PLoS One 7:e39089. doi: 10.1371/journal.pone.0039089

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4:e1660. doi: 10.7717/peerj.1660

Brandrud, M. K., Baar, J., Lorenzo, M. T., Athanasiadis, A., Bateman, R. M., Chase, M. W., et al. (2020). Phylogenomic relationships of diploids and the origins of allotetraploids in Dactylorhiza (Orchidaceae). Syst. Biol. 69, 91–109. doi: 10.1093/sysbio/syz035

Brassac, J., and Blattner, F. R. (2015). Species-level phylogeny and polyploid relationships in Hordeum (Poaceae) inferred by next-generation sequencing and in silico cloning of multiple nuclear loci. Syst. Biol. 64, 792–808. doi: 10.1093/sysbio/syv035

Brysting, A. K., Oxelman, B., Huber, K. T., Moulton, V., and Brochmann, C. (2007). Untangling complex histories of genome mergings in high polyploids. Syst. Biol. 56, 467–476. doi: 10.1080/10635150701424553

Caković, D., Stešević, D., Schönswetter, P., and Frajman, B. (2015). How many taxa? Spatiotemporal evolution and taxonomy of Amphoricarpos (Asteraceae, Carduoideae) on the Balkan Peninsula. Org. Divers. Evol. 15, 429–445. doi: 10.1007/s13127-015-0218-6

Cao, Z., Liu, X., Ogilvie, H. A., Yan, Z., and Nakhleh, L. (2019). Practical Aspects of Phylogenetic Network Analysis Using PhyloNet. bioRxiv [Preprint]. Available online at: https://doi.org/10.1101/746362 (Accessed January 7, 2021).

Carlsen, T., Bleeker, W., Hurka, H., Elven, R., and Brochmann, C. (2009). Biogeography and phylogeny of Cardamine (Brassicaceae). Ann. Missouri Bot. Gard. 96, 215–236. doi: 10.2307/40389931

Carter, K. A., Liston, A., Bassil, N. V., Alice, L. A., Bushakra, J. M., Sutherland, B. L., et al. (2019). Target capture sequencing unravels Rubus evolution. Front. Plant Sci. 10:1615. doi: 10.3389/fpls.2019.01615

Chen, H., German, D. A., Al-Shehbaz, I. A., Yue, J., and Sun, H. (2020). Phylogeny of Euclidieae (Brassicaceae) based on plastome and nuclear ribosomal DNA data. Mol. Phylogenet. Evol. 153:106940. doi: 10.1016/j.ympev.2020.10 6940

Chernomor, O., von Haeseler, A., and Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol. 65, 997–1008. doi: 10.1093/sysbio/syw037

Cires, E., Baltisberger, M., Cuesta, C., Vargas, P., and Prieto, J. A. F. (2014). Allopolyploid origin of the Balkan endemic Ranunculus wettsteinii (Ranunculaceae) inferred from nuclear and plastid DNA sequences. Org. Divers. Evol. 14, 1–10. doi: 10.1007/s13127-013-0150-6

Crowl, A. A., Myers, C., and Cellinese, N. (2017). Embracing discordance: phylogenomic analyses provide evidence for allopolyploidy leading to cryptic diversity in a Mediterranean Campanula (Campanulaceae) clade. Evolution 71, 913–922. doi: 10.1111/evo.13203

Dauphin, B., Grant, J. R., Farrar, D. R., and Rothfels, C. J. (2018). Rapid allopolyploid radiation of moonwort ferns (Botrychium; Ophioglossaceae) revealed by PacBio sequencing of homologous and homeologous nuclear regions. Mol. Phylogenet. Evol. 120, 342–353. doi: 10.1016/j.ympev.2017.11.025

Degnan, J. H., and Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340. doi: 10.1016/j.tree.2009.01.009

del Valle, J. C., Casimiro-Soriguer, I., Buide, M. L., Narbona, E., and Whittall, J. B. (2019). Whole plastome sequencing within Silene section Psammophilae reveals mainland hybridization and divergence with the Balearic Island populations. Front. Plant Sci. 10:1466. doi: 10.3389/fpls.2019.01466

Doležel, J., Greilhuber, J., and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. Nat. Protoc. 2, 2233–2244. doi: 10.1038/nprot.2007.310

Durović, S., Schönswetter, P., Niketić, M., Tomović, G., and Frajman, B. (2017). Disentangling relationships among the members of the Silene saxifraga alliance (Caryophyllaceae): phylogenetic structure is geographically rather than taxonomically segregated. Taxon 66, 343–364. doi: 10.12705/662.4

Eriksson, J. S., de Sousa, F., Bertrand, Y. J. K., Antonelli, A., Oxelman, B., and Pfeil, B. E. (2018). Allele phasing is critical to revealing a shared allopolyploid origin of Medicago arborea and M. strasseri (Fabaceae). BMC Evol. Biol. 18:9. doi: 10.1186/s12862-018-1127-z

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Fonseca, L. H. M., and Lohmann, L. G. (2019). Exploring the potential of nuclear and mitochondrial sequencing data generated through genome-skimming for plant phylogenetics: a case study from a clade of neotropical lianas. J. Syst. Evol. 58, 18–32. doi: 10.1111/jse.12533

Freyman, W. A., Johnson, M. G., and Rothfels, C. J. (2020). Homologizer: phylogenetic phasing of gene copies into polyploid subgenomes. bioRxiv [Preprint] doi: 10.1101/2020.10.22.351486

Georghiou, K., and Delipetrou, P. (2010). Patterns and traits of the endemic plants of Greece. Bot. J. Linn. Soc. 162, 130–422. doi: 10.1111/j.1095-8339.2010.01025.x

Gonçalves, D. J. P., Simpson, B. B., Ortiz, E. M., Shimizu, G. H., and Jansen, R. K. (2019). Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. Mol. Phylogenet. Evol. 138, 219–232. doi: 10.1016/j.ympev.2019.05.022

Grover, C. E., Gallagher, J. P., Jareczek, J. J., Page, J. T., Udall, J. A., Gore, M. A., et al. (2015). Re-evaluating the phylogeny of allopolyploid Gossypium L. Mol. Phylogenet. Evol. 92, 45–52. doi: 10.1016/j.ympev.2015.05.023

Grünewald, S., Spillner, A., Bastkowski, S., Bögershausen, A., and Moulton, V. (2013). SuperQ: computing super networks from quartets. IEEE/ACM Trans. Comput. Biol. Bioinform. 10, 151–160. doi: 10.1109/TCBB.2013.8

Harrison, S., and Noss, R. (2017). Endemism hotspots are linked to stable climatic refugia. Ann. Bot. (Oxford) 119, 207–214. doi: 10.1093/aob/mcw248

Heibl, C. (2008). PHYLOCH: R Language Tree Plotting Tools and Interfaces to Diverse Phylogenetic Software Packages. Available online at: http://www.christophheibl.de/Rpackages.html (accessed November 17, 2020).

Hewitt, G. M. (2011). "Mediterranean peninsulas: the evolution of hotspots," in Biodiversity Hotspots, eds F. E. Zachos, and J. C. Habel (Berlin: Springer), 123–147. doi: 10.1007/978-3-642-20992-5_7

Huang, X. C., German, D. A., and Koch, M. A. (2020). Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. Ann. Bot. (Oxford) 125, 29–47. doi: 10.1093/aob/mcz123

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

Jalas, J., and Suominen, J. (1994). Atlas Florae Europaeae 10. Helsinki: The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo.

Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., et al. (2016). HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl. Plant Sci. 4:1600016. doi: 10.3732/apps.1600016

Jones, B. M. G., and Akeroyd, J. R. (1993). "Cardamine," in Flora Europaea 1, Psilotaceae to Platanaceae. 2nd Edn, eds T. G. Tutin, V. H. Heywood, N. A. Burges, D. H. Valentine, S. M. Walters, and D. A. Webb (Cambridge: Cambridge University Press), 346–352.

Junier, T., and Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics 26, 1669–1670. doi: 10.1093/bioinformatics/btq243

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., and Jermiin, L. S. (2017). Model finder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. doi: 10.1038/nmeth.4285

Kamneva, O. K., Syring, J., Liston, A., and Rosenberg, N. A. (2017). Evaluating allopolyploid origins in strawberries (Fragaria) using haplotypes generated from target capture sequencing. BMC Evol. Biol. 17:180. doi: 10.1186/s12862-017-1019-7

Karbstein, K., Tomasello, S., Hodač, L., Dunkel, F. G., Daubert, M., and Hörandl, E. (2020). Phylogenomics supported by geometric morphometrics reveals delimitation of sexual species within the polyploid apomictic Ranunculus auricomus complex (Ranunculaceae). Taxon 69, 1191–1220. doi: 10.1002/tax.12365

Kates, H. R., Johnson, M. G., Gardner, E. M., Zerega, N. J. C., and Wickett, N. J. (2018). Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of Artocarpus. Amer. J. Bot. 105, 404–416. doi: 10.1002/ajb2.1068

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. doi: 10.1093/molbev/mst010

Kaya, S., and Çiplak, B. (2017). Phylogeography and taxonomy of the Psorodonotus caucasicus (Orthoptera, Tettigoniidae) group: independent double invasion of the Balkans from the Caucasus. Syst. Entomol. 42, 118–133. doi: 10.1111/syen.12197

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software

platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Koch, M. A., Karl, R., and German, D. A. (2017). Underexplored biodiversity of Eastern Mediterranean biota: systematics and evolutionary history of the genus Aubrieta (Brassicaceae). Ann. Bot. (Oxford) 119, 39–57. doi: 10.1093/aob/mcw204

Konowalik, K., Wagner, F., Tomasello, S., Vogt, R., and Oberprieler, C. (2015). Detecting reticulate relationships among diploid Leucanthemum Mill. (Compositae, Anthemideae) taxa using multilocus species tree reconstruction methods and AFLP fingerprinting. Mol. Phylogenet. Evol. 92, 308–328. doi: 10.1016/j.ympev.2015.06.003

Kougioumoutzis, K., Kokkoris, I. P., Panitsa, M., Kallimanis, A., Strid, A., and Dimopoulos, P. (2021). Plant endemism centres and biodiversity hotspots in Greece. Biology 10:72. doi: 10.3390/biology10020072

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35, 4453–4455. doi: 10.1093/bioinformatics/btz305

Krak, K., Caklová, P., Chrtek, J., and Fehrer, J. (2013). Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (Hieracium, Asteraceae). Heredity 110, 138–151. doi: 10.1038/hdy.2012.100

Kučera, J., Lihová, J., and Marhold, K. (2006). Taxonomy and phylogeography of Cardamine impatiens and C. pectinata (Brassicaceae). Bot. J. Linn. Soc. 152, 169–195. doi: 10.1111/j.1095-8339.2006.00559.x

Kučera, J., Marhold, K., and Lihová, J. (2010). Cardamine maritima group (Brassicaceae) in the amphi-Adriatic area: a hotspot of species diversity revealed by DNA sequences and morphological variation. Taxon 59, 148–164. doi: 10.2307/27757059

Kučera, J., Valko, I., and Marhold, K. (2005). On-line database of the chromosome numbers of the genus Cardamine (Brassicaceae). Biologia (Bratislava) 60, 473–476.

Kyriakidou, M., Tai, H. H., Anglin, N. L., Ellis, D., and Strömvik, M. V. S. (2018). Current strategies of polyploid plant genome sequence assembly. Front. Plant Sci. 9:1660. doi: 10.3389/fpls.2018.01660

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. doi: 10.1038/nmeth.1923

Larridon, I., Villaverde, T., Zuntini, A. R., Pokorny, L., Brewer, G. E., Epitawalage, N., et al. (2020). Tackling rapid radiations with targeted sequencing. Front. Plant Sci. 10:1655. doi: 10.3389/fpls.2019.01655

Lautenschlager, U., Wagner, F., and Oberprieler, C. (2020). AllCoPol: inferring allele co-ancestry in polyploids. BMC Bioinform. 21:441. doi: 10.1186/s12859-020-03750-9

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Lihová, J., Fuertes Aguilar, J., Marhold, K., and Nieto Feliner, G. (2004a). Origin of the disjunct tetraploid Cardamine amporitana (Brassicaceae) assessed with nuclear and chloroplast DNA sequence data. Amer. J. Bot. 91, 1231–1242. doi: 10.3732/ajb.91.8.1231

Lihová, J., and Marhold, K. (2006). "Phylogenetic and diversity patterns in Cardamine (Brassicaceae) - a genus with conspicuous polyploid and reticulate evolution," in Plant Genome: Biodiversity and Evolution, Vol. 1C: Phanerogams (Angiosperms – Dicotyledons), eds A. K. Sharma and A. Sharma (Enfield: Science Publishers, Inc), 149–186.

Lihová, J., Shimizu, K. K., and Marhold, K. (2006). Allopolyploid origin of Cardamine asarifolia (Brassicaceae): incongruence between plastid and nuclear ribosomal DNA sequences solved by a single-copy nuclear gene. Mol. Phylogenet. Evol. 39, 759–786. doi: 10.1016/j.ympev.2006. 01.027

Lihová, J., Tribsch, A., and Marhold, K. (2003). The Cardamine pratensis (Brassicaceae) group in the Iberian Peninsula: taxonomy, polyploidy and distribution. Taxon 52, 783–802. doi: 10.2307/3647352

Lihová, J., Tribsch, A., and Stuessy, T. F. (2004b). Cardamine apennina: a new endemic diploid species of the C. pratensis group (Brassicaceae) from Italy. Plant Syst. Evol. 245, 69–92. doi: 10.1007/s00606-003-0119-6

López-González, N., Bobo-Pinilla, J., Padilla-García, N., Loureiro, J., Castro, S., Rojas-Andrés, B. M., et al. (2021). Genetic similarities versus morphological resemblance: unraveling a polyploid complex in a Mediterranean biodiversity hotspot. Mol. Phylogenet. Evol. 155:107006. doi: 10.1016/j.ympev.2020.107006

López-Vinyallonga, S., López-Pujol, J., Constantinidis, T., Susanna, A., and Garcia-Jacas, N. (2015). Mountains and refuges: genetic structure and evolutionary history in closely related, endemic Centaurea in continental Greece. Mol. Phylogenet. Evol. 92, 243–254. doi: 10.1016/j.ympev.2015.06.018

Lövkvist, B. (1956). The Cardamine pratensis complex. Outlines of its cytogenetics and taxonomy. Symb. Bot. Upsal. 14/2, 1–131.

Maddison, W. P. (1997). Gene trees in species trees. Syst. Biol. 46, 523–536. doi: 10.1093/sysbio/46.3.523

Madlung, A., and Wendel, J. F. (2013). Genetic and epigenetic aspects of polyploid evolution in plants. Cytogenet. Genome Res. 140, 270–285. doi: 10.1159/000351430

Maguilla, E., Escudero, M., Hipp, A. L., and Luceño, M. (2017). Allopatric speciation despite historical gene flow: divergence and hybridization in Carex furva and C. lucennoiberica (Cyperaceae) inferred from plastid and nuclear RAD-seq data. Mol. Ecol. 26, 5646–5662. doi: 10.1111/mec.14253

Mandáková, T., Kovařík, A., Zozomová-Lihová, J., Shimizu-Inatsugi, R., Shimizu, K. K., Mummenhoff, K., et al. (2013). The more the merrier: recent hybridization and polyploidy in Cardamine. Plant Cell 25, 3280–3295. doi: 10.1105/tpc.113.114405

Mandáková, T., and Lysak, M. A. (2016a). Chromosome preparation for cytogenetic analyses in Arabidopsis. Curr. Protoc. Plant Biol. 1, 43–51. doi: 10.1002/cppb.20009

Mandáková, T., and Lysak, M. A. (2016b). Painting of Arabidopsis chromosomes with chromosome-specific BAC clones. Curr. Protoc. Plant Biol. 1, 359–371. doi: 10.1002/cppb.20022

Mandáková, T., Marhold, K., and Lysak, M. A. (2014). The widespread crucifer species Cardamine flexuosa is an allotetraploid with a conserved subgenomic structure. New Phytol. 201, 982–992. doi: 10.1111/nph.12567

Mandáková, T., Zozomová-Lihová, J., Kudoh, H., Zhao, Y., Lysak, M. A., 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. (Oxford) 124, 209–220. doi: 10.1093/aob/mcz019

Marhold, K., and Ančev, M. E. (1999). Cardamine penzesii, a rediscovered taxon of the C. pratensis group (Cruciferae). Ann. Bot. Fenn. 36, 171–180.

Marhold, K., Ančev, M. E., and Tan, K. (1996). A new subspecies of Cardamine amara (Brassicaceae) from Bulgaria and Greece. Ann. Bot. Fenn. 33, 199–204.

Marhold, K., Kudoh, H., Pak, J. H., Watanabe, K., Španiel, S., and Lihová, J. (2010). Cytotype diversity and genome size variation in eastern Asian polyploid Cardamine (Brassicaceae) species. Ann. Bot. (Oxford) 105, 249–264. doi: 10.1093/aob/mcp282

Marhold, K., Lihová, J., Perný, M., and Bleeker, W. (2004). Comparative ITS and AFLP analysis of diploid Cardamine (Brassicaceae) taxa from closely related polyploid complexes. Ann. Bot. (Oxford) 93, 507–520. doi: 10.1093/aob/mch073

Marhold, K., Lihová, J., Perný, M., Grupe, R., and Neuffer, B. (2002). Natural hybridization in Cardamine (Brassicaceae) in the Pyrenees: evidence from morphological and molecular data. Bot. J. Linn. Soc. 139, 275–294. doi: 10.1046/j.1095-8339.2002.00066.x

Marhold, K., Šlenker, M., and Zozomová-Lihová, J. (2018). Polyploidy and hybridization in the Mediterranean and neighbouring areas towards the north: examples from the genus Cardamine (Brassicaceae). Biol. Serb. 40, 47–59. doi: 10.5281/zenodo.1406320

Marhold, K., and Tan, K. (2000). The distribution of Cardamine matthioli (Brassicaceae) in Greece. Thaiszia J. Bot. 9 (1999), 109–112.

Marques, I., Loureiro, J., Draper, D., Castro, O., and Castro, S. (2018). How much do we know about the frequency of hybridisation and polyploidy in the Mediterranean region? Plant Biol. 20 (Suppl. 1) 21–37. doi: 10.1111/plb.12639

Martin, M., Patterson, M., Garg, S., Fischer, S., Pisanti, N., Klau, G. W., et al. (2016). WhatsHap: Fast and Accurate Read-Based Phasing. bioRxiv [Preprint]. Available online at: https://doi.org/10.1101/085050 (Accessed January 7, 2021).

Médail, F., and Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. J. Biogeogr. 36, 1333–1345. doi: 10.1111/j.1365-2699.2008.02051.x

Médail, F., and Quézel, P. (1997). Hot-spots analysis for conservation of plant biodiversity in the Mediterranean Basin. Ann. Missouri. Bot. Gard. 84, 112–127. doi: 10.2307/2399957

Melichárková, A., Šlenker, M., Zozomová-Lihová, J., Skokanová, K., Šingliarová, B., Kačmárová, T., et al. (2020). So closely related and yet so different: Strong contrasts between the evolutionary histories of species of the Cardamine pratensis polyploid complex in Central Europe. Front. Plant Sci. 11:588856. doi: 10.3389/fpls.2020.588856

Melichárková, A., Španiel, S., Brišková, D., Marhold, K., and Zozomová-Lihová, J. (2017). Unravelling allopolyploid origins in the Alyssum montanum–A. repens species complex (Brassicaceae): low-copy nuclear gene data complement plastid DNA sequences and AFLPs. Bot. J. Linn. Soc. 184, 485–502. doi: 10.1093/botlinnean/box039

Melichárková, A., Španiel, S., Marhold, K., Hurdu, B. I., Drescher, A., and Zozomová-Lihová, J. (2019). Diversification and independent polyploid origins in the disjunct species Alyssum repens from the Southeastern Alps and the Carpathians. Amer. J. Bot. 106, 1499–1518. doi: 10.1002/ajb2.1370

Morales-Briones, D. F., Liston, A., and Tank, D. C. (2018). Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus Lachemilla (Rosaceae). New Phytol. 218, 1668–1684. doi: 10.1111/nph.15099

Muellner-Riehl, A. N., Schnitzler, J., Kissling, W. D., Mosbrugger, V., Rijsdijk, K. F., Seijmonsbergen, A. C., et al. (2019). Origins of global mountain plant biodiversity: testing the 'mountain-geobiodiversity hypothesis'. J. Biogeogr. 46, 2826–2838. doi: 10.1111/jbi.13715

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature 403, 853–858. doi: 10.1038/35002501

Naciri, Y., and Linder, H. P. (2015). Species delimitation and relationships: The dance of the seven veils. Taxon 64, 3–16. doi: 10.12705/641.24

Nieto Feliner, G. (2014). Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. Perspect. Plant Ecol. Evol. Syst. 16, 265–278. doi: 10.1016/j.ppees.2014.07.002

Nieto Feliner, G., and Rosselló, J. A. (2012). "Concerted evolution of multigene families and homoeologous recombination," in Plant Genome Diversity Vol. 1, eds J. Wendel, J. Greilhuber, J. Doležel, and I. Leitch (Vienna: Springer), 171–193. doi: 10.1007/978-3-7091-1130-7_12

Nikolov, L. A., Shushkov, P., Nevado, B., Gan, X., Al-Shehbaz, I. A., Filatov, D., et al. (2019). Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. New Phytol. 222, 1638–1651. doi: 10.1111/nph.15732

Oberprieler, C., Wagner, F., Tomasello, S., and Konowalik, K. (2017). A permutation approach for inferring species networks from gene trees in polyploid complexes by minimising deep coalescences. Methods Ecol. Evol. 8, 835–849. doi: 10.1111/2041-210X.12694

Olšavská, K., Slovák, M., Marhold, K., Štubňová, E., and Kučera, J. (2016). On the origins of Balkan endemics: the complex evolutionary history of the Cyanus napulifer group (Asteraceae). Ann. Bot. (Oxford) 118, 1071–1088. doi: 10.1093/aob/mcw142

Oxelman, B., Brysting, A. K., Jones, G. R., Marcussen, T., Oberprieler, C., and Pfeil, B. E. (2017). Phylogenetics of allopolyploids. Ann. Rev. Ecol. Evol. Syst. 48, 543–557.

Özüdoğru, B., and Mummenhoff, K. (2020). Phylogenetic and biogeographical history confirm the Anatolian origin of Bornmuellera (Brassicaceae) and clade divergence between Anatolia and the Balkans in the Plio-Pleistocene transition. Turkish J. Bot. 44, 593–603. doi: 10.3906/bot-2007-42

Page, J. T., Gingle, A. R., and Udall, J. A. (2013). PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms. G3 (Bethesda) 3, 517–525. doi: 10.1534/g3.112.005298

Panitsa, M., Kagiampaki, A., and Kougioumoutzis, K. (2018). "Plant diversity and biogeography of the Aegean Archipelago: a new synthesis," in Biogeography and Biodiversity of the Aegean. In Honour of Prof. Moysis Mylonas, eds M. Moysis, P. Pafilis, A. Parmakelis, N. Poulakakis, S. Sfenthourakis, and K. Triantis (Nicosia: Broken Hill Publishers, Ltd.), 269–278.

Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., and Smith, S. A. (2018). Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. Amer. J. Bot. 105, 385–403. doi: 10.1002/ajb2.1016

Perný, M., Tribsch, A., and Anchev, M. E. (2004). Infraspecific differentiation in the Balkan diploid Cardamine acris (Brassicaceae): molecular and morphological evidence. Folia Geobot. 39, 405–429. doi: 10.1007/BF02803211

Perný, M., Tribsch, A., Stuessy, T. F., and Marhold, K. (2005a). Allopolyploid origin of Cardamine silana (Brassicaceae) from Calabria (Southern Italy): karyological, morphological and molecular evidence. Bot. J. Linn. Soc. 148, 101–116. doi: 10.1111/j.1095-8339.2005.00389.x

Perný, M., Tribsch, A., Stuessy, T. F., and Marhold, K. (2005b). Taxonomy and cytogeography of Cardamine raphanifolia and C. gallaecica (Brassicaceae) in the Iberian Peninsula. Plant Syst. Evol. 254, 69–91. doi: 10.1007/s00606-005-0317-5

Perrigo, A., Hoorn, C., and Antonelli, A. (2020). Why mountains matter for biodiversity. J. Biogeogr. 47, 315–325. doi: 10.1111/jbi.13731

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

R Core Team (2019). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. doi: 10.1038/nbt.1754

Roces-Díaz, J. V., Jiménez-Alfaro, B., Chytrý, M., Díaz-Varela, E. R., and Álvarez-Álvarez, P. (2018). Glacial refugia and mid-Holocene expansion delineate the current distribution of Castanea sativa in Europe. Palaeogeogr. Palaeoclimatol. Palaeoecol. 491, 152–160. doi: 10.1016/j.palaeo.2017.12.004

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes 4, 137–138. doi: 10.1046/j.1471-8286.2003.00566.x

Rothfels, C. J. (2021). Polyploid phylogenetics. New Phytol. 230, 66–72. doi: 10.1111/nph.17105

Rothfels, C. J., Pryer, K. M., and Li, F. W. (2017). Next-generation polyploid phylogenetics: rapid resolution of hybrid polyploid complexes using PacBio single-molecule sequencing. New Phytol. 213, 413–429. doi: 10.1111/nph.14111

Rousseau-Gueutin, M., Gaston, A., Aïnouche, A., Aïnouche, M. L., Olbricht, K., Staudt, G., et al. (2009). Tracking the evolutionary history of polyploidy in Fragaria L. (strawberry): New insights from phylogenetic analyses of low-copy nuclear genes. Mol. Phylogenet. Evol. 51, 515–530. doi: 10.1016/j.ympev.2008.12.024

Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S. C. K., et al. (2016). Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African Oxalis (Oxalidaceae). Mol. Ecol. Resour. 16, 1124–1135. doi: 10.1111/1755-0998.12487

Schönswetter, P., Suda, J., Popp, M., Weiss-Schneeweiss, H., and Brochmann, C. (2007). Circumpolar phylogeography of Juncus biglumis (Juncaceae) inferred from AFLP fingerprints, cpDNA sequences, nuclear DNA content and chromosome numbers. Molec. Phylogenet. Evol. 42, 92–103. doi: 10.1016/j.ympev.2006.06.016

Schrinner, S. D., Serra Mari, R., Ebler, J., Rautiainen, M., Seillier, L., Reimer, J. J., et al. (2020). Haplotype threading: accurate polyploid phasing from long reads. Genome Biol. 21:252. doi: 10.1186/s13059-020-02158-1

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508. doi: 10.1080/10635150290069913

Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247. doi: 10.1093/bioinformatics/17.12.1246

Silva, G. S., and Souza, M. M. (2013). Genomic in situ hybridization in plants. Genet. Mol. Res. 12, 2953–2965. doi: 10.4238/2013.August.12.11

Slater, G. S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinform. 6:31. doi: 10.1186/1471-2105-6-31

Small, R. L., Cronn, R. C., and Wendel, J. F. (2004). Use of nuclear genes for phylogeny reconstruction in plants. Aus. Syst. Bot. 17, 145–170. doi: 10.1071/SB03015

Soltis, D. E., Visger, C. J., and Soltis, P. S. (2014). The polyploidy revolution then…and now: Stebbins revisited. Amer. J. Bot. 101, 1057–1078. doi: 10.3732/ajb.1400178

Soltis, P. S., and Soltis, D. E. (2009). The role of hybridization in plant speciation. Ann. Rev. Plant Biol. 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Španiel, S., Marhold, K., and Zozomová-Lihová, J. (2017). The polyploid Alyssum montanum-A. repens complex in the Balkans: a hotspot of species and genetic diversity. Plant Syst. Evol. 303, 1443-1465. doi: 10.1007/s00606-017-1470-3

Stevanović, V., Tan, K., and Petrova, A. (2007). Mapping the endemic flora of the Balkans – a progress report. Bocconea 21, 131–137.

Strid, A. (1986). "Cardamine L.," in Mountain Flora of Greece 1, ed. A. Strid (Cambridge: Cambridge University Press), 256–261.

Suda, J., and Trávníček, P. (2006a). Estimation of relative nuclear DNA content in dehydrated plant tissues by flow cytometry. Curr. Protoc. Cytometry 38, 7.30.1–7.30.14. doi: 10.1002/0471142956.cy0730s38

Suda, J., and Trávníček, P. (2006b). Reliable DNA ploidy determination in dehydrated tissues of vascular plants by DAPI flow cytometry–new prospects for plant research. Cytometry 69A, 273–280. doi: 10.1002/cyto.a.20253

Surina, B., Pfanzelt, S., Einzmann, H. J. R., and Albach, D. C. (2014). Bridging the Alps and the Middle East: evolution, phylogeny and systematics of the genus Wulfenia (Plantaginaceae). Taxon 63, 843–858. doi: 10.12705/634.18

Tan, K. (2002). "Cardamine L.," in Flora Hellenica 2, eds A. Strid and K. Tan (Ruggell: A. R. G. Gantner Verlag K. G.), 178–184.

Tedder, A., Helling, M., Pannell, J. R., Shimizu-Inatsugi, R., Kawagoe, T., van Campen, J., et al. (2015). Female sterility associated with increased clonal propagation suggests a unique combination of androdioecy and asexual reproduction in populations of Cardamine amara (Brassicaceae). Ann. Bot. (Oxford) 115, 763–776. doi: 10.1093/aob/mcv006

Temsch, E. M., Greilhuber, J., and Krisai, R. (2010). Genome size in liverworts. Preslia 82, 63–80.

Thompson, J. D. (2020). Plant Evolution in the Mediterranean: Insights for conservation. 2nd Edn. New York: Oxford University Press. doi: 10.1093/oso/9780198835141.001.0001

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45, W6-W11. doi: 10.1093/nar/gkx391

Tkach, N., Röser, M., Suchan, T., Cieślak, E., Schönswetter, P., and Ronikier, M. (2019). Contrasting evolutionary origins of two mountain endemics: Saxifraga wahlenbergii (Western Carpathians) and S. styriaca (Eastern Alps). BMC Evol. Biol. 19:18. doi: 10.1186/s12862-019-1355-x

Tomasello, S., Karbstein, K., Hodač, L., Paetzold, C., and Hörandl, E. (2020). Phylogenomics unravels Quaternary vicariance and allopatric speciation patterns in temperate-montane plant species: a case study on the Ranunculus auricomus species complex. Molec. Ecol. 29, 2031–2049. doi: 10.1111/mec.15458

Tomović, G., Lakušić, D., Ranđelović, V., and Marhold, K. (2009). Cardamine amara (Brassicaceae) in Serbia and Republic of Macedonia. Biologia (Bratislava) 64, 1095–1099. doi: 10.2478/s11756-009-0182-8

Tomović, G., Niketić, M., Lakušić, D., Ranđelović, V., and Stevanović, V. (2014). Balkan endemic plants in Central Serbia and Kosovo regions: distribution patterns, ecological characteristics, and centres of diversity. Bot. J. Linn. Soc. 176, 173–202. doi: 10.1111/boj.12197

Vargas, O. M., Ortiz, E. M., and Simpson, B. B. (2017). Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-andean diversification (Asteraceae: Astereae: Diplostephium). New Phytol. 214, 1736–1750. doi: 10.1111/nph.14530

Vatanparast, M., Powell, A., Doyle, J. J., and Egan, A. N. (2018). Targeting legume loci: a comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. Appl. Plant Sci. 6:e1036. doi: 10.1002/aps3.1036

Villaverde, T., Pokorny, L., Olsson, S., Rincón-Barrado, M., Johnson, M. G., Gardner, E. M., et al. (2018). Bridging the micro-and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. New Phytol. 220, 636–650. doi: 10.1111/nph.15312

Walker, J. F., Walker-Hale, N., Vargas, O. M., Larson, D. A., and Stull, G. W. (2019). Characterizing gene tree conflict in plastome-inferred phylogenies. PeerJ 7:e7747. doi: 10.7717/peerj.7747

Weitemier, K., Straub, S. C., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., et al. (2014). Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. Appl. Plant Sci. 2:1400042. doi: 10.3732/apps.1400042

Weitemier, K., Straub, S. C. K., Fishbein, M., and Liston, A. (2015). Intragenomic polymorphisms among high-copy loci: a genus-wide study of nuclear ribosomal DNA in Asclepias (Apocynaceae). PeerJ 3:e718. doi: 10.7717/peerj.718

Wen, D., Yu, Y., Zhu, J., and Nakhleh, L. (2018). Inferring phylogenetic networks using PhyloNet. Syst. Biol. 67, 735–740. doi: 10.1093/sysbio/syy015

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinform. 19:153. doi: 10.1186/s12859-018-2129-y

Zozomová-Lihová, J., Malánová-Krásná, I., Vít, P., Urfus, T., Senko, D., Svitok, M., et al. (2015). Cytotype distribution patterns, ecological differentiation, and genetic structure in a diploid-tetraploid contact zone of Cardamine amara. Amer. J. Bot. 102, 1380–1395. doi: 10.3732/ajb.1500052

Zozomová-Lihová, J., Melichárková, A., Svitok, M., and Španiel, S. (2020). Pleistocene range disruption and postglacial expansion with secondary contacts explain the genetic and cytotype structure in the western Balkan endemic Alyssum austrodalmaticum (Brassicaceae). Plant Syst. Evol. 306:47. doi: 10.1007/s00606-020-01677-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4.2 Evolution of hygrophytic plant species in the Anatolia-Caucasus region: insights from phylogenomic analyses of *Cardamine* perennials

Annals of Botany 131: 585–600, 2023

doi: 10.1093/aob/mcad008

Kantor, A.*, Kučera, J.*, Šlenker, M.*, Breidy, J., Dönmez, A. A., Marhold, K., Slovák, M., Svitok, M., & Zozomová-Lihová, J.

*these authors have contributed equally to this work

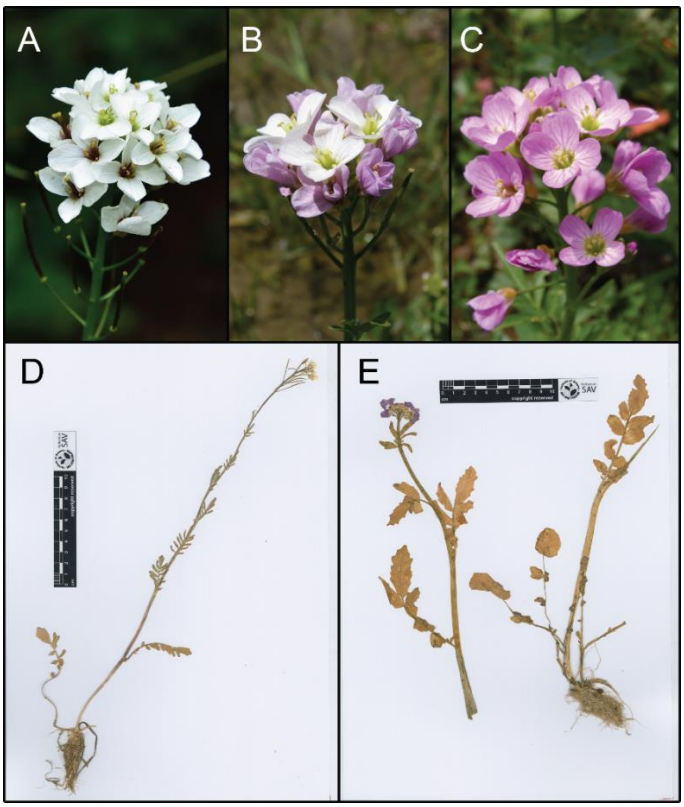


Fig. 4. Photographs and scanned herbarium specimens of *C. uliginosa* (including morphotypes formerly referred to as *C. seidlitziana*, C, E) demonstrating its polymorphism in flower colour (A, B, C) and habitus (D, E). Photos of inflorescences (A, B, C) were taken by Jaromír Kučera.

My contribution: I participated in the laboratory work (DNA extraction, Hyb-Seq library preparation, chromosome counting); measured the morphological traits from herbarium specimens and scanned flowers material, then performed the morphometric analyses in R software. I was involved in preparation of the distributional data for the niche analysis; I contributed to the interpretation of the results and to the preparation of the manuscript.

Supplementary materials are available at:
https://academic.oup.com/aob/article/131/4/585/6993035

Annals of Botany **131**: 585–600, 2023
https://doi.org/10.1093/aob/mcad008, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Founded 1887

Evolution of hygrophytic plant species in the Anatolia–Caucasus region: insights from phylogenomic analyses of *Cardamine* perennials

Adam Kantor[1,2,†], Jaromír Kučera[1,†], Marek Šlenker[1,†], Joêlle Breidy[3], Ali A. Dönmez[4], Karol Marhold[1,2], Marek Slovák[1,2], Marek Svitok[5,6] and Judita Zozomová-Lihová[1,*]

[1]Institute of Botany, Plant Science and Biodiversity Centre, Slovak Academy of Sciences, 845 23 Bratislava, Slovakia, [2]Department of Botany, Faculty of Science, Charles University, 128 01 Prague, Czechia, [3]National Genebank, Lebanese Agricultural Research Institute, Zahle 1801, Lebanon, [4]Botany Section, Department of Biology, Faculty of Science, Hacettepe University, 06800 Beytepe-Ankara, Turkey, [5]Department of Biology and General Ecology, Faculty of Ecology and Environmental Sciences, Technical University in Zvolen, 960 01 Zvolen, Slovakia, and [6]Department of Ecosystem Biology, Faculty of Science, University of South Bohemia, 370 05 České Budějovice, Czechia
[*]For correspondence. E-mail judita.zozomova@savba.sk
[†]Equal contribution.

Received: 10 October 2022 Returned for revision: 21 December 2022 Editorial decision: 6 January 2023 Accepted: 10 January 2023
Electronically published: 19 January 2023

• **Background and Aims** Southwestern Asia is a significant centre of biodiversity and a cradle of diversification for many plant groups, especially xerophytic elements. In contrast, little is known about the evolution and diversification of its hygrophytic flora. To fill this gap, we focus on *Cardamine* (Brassicaceae) species that grow in wetlands over a wide altitudinal range. We aimed to elucidate their evolution, assess the extent of presumed historical gene flow between species, and draw inferences about intraspecific structure.
• **Methods** We applied the phylogenomic Hyb-Seq approach, ecological niche analyses and multivariate morphometrics to a total of 85 *Cardamine* populations from the target region of Anatolia–Caucasus, usually treated as four to six species, and supplemented them with close relatives from Europe.
• **Key Results** Five diploids are recognized in the focus area, three of which occur in regions adjacent to the Black and/or Caspian Sea (*C. penzesii*, *C. tenera*, *C. lazica*), one species widely distributed from the Caucasus to Lebanon and Iran (*C. uliginosa*), and one western Anatolian entity (provisionally *C.* cf. *uliginosa*). Phylogenomic data suggest recent speciation during the Pleistocene, likely driven by both geographic separation (allopatry) and ecological divergence. With the exception of a single hybrid (allotetraploid) speciation event proven for *C. wiedemanniana*, an endemic of southern Turkey, no significant traces of past or present interspecific gene flow were observed. Genetic variation within the studied species is spatially structured, suggesting reduced gene flow due to geographic and ecological barriers, but also glacial survival in different refugia.
• **Conclusions** This study highlights the importance of the refugial regions of the Black and Caspian Seas for both harbouring and generating hygrophytic species diversity in Southwestern Asia. It also supports the significance of evolutionary links between Anatolia and the Balkan Peninsula. Reticulation and polyploidization played a minor evolutionary role here in contrast to the European relatives.

Key words: Allopolyploidy, Anatolia, Caucasus, *Cardamine*, ecological niche, endemism, Hyb-Seq, hygrophytic flora, phylogenomics.

INTRODUCTION

Southwestern (SW) Asia is a significant centre of biodiversity in the Northern Hemisphere, located at the junction of three biogeographic regions, the Mediterranean, Euro-Siberian and Irano-Turanian (Mittermeier *et al.*, 2011; Noroozi, 2020). Thanks to its complex topography and geology, diverse climate and a strategical crossroad position, it has acted as both a cradle of diversification and a migration corridor (Manafzadeh et al., 2014, 2017; Noroozi et al., 2018, 2019; Noroozi, 2020). Many vascular plant genera and even families (including Brassicaceae; Franzke *et al.*, 2011) are thought to have originated and diversified in this area, as also supported by recent phylogenetic studies (e.g. Koch *et al.*, 2017; Özüdoğru and Mummenhoff, 2020; Özüdoğru *et al.*, 2020). The area has a mountainous character, with the Anatolian mountain ranges spanning the western, central and southern parts, the Caucasus delimiting SW Asia in the north, and the Zagros and Alborz Mountains in the east (Noroozi, 2020). The impact of Pleistocene glaciations on mountain flora in SW Asia was generally weaker than in Europe, with extensive glaciation occurring only in the western and central parts of the Greater Caucasus and in the eastern parts of the North Anatolian Mountains (Gobejishvili *et al.*, 2011; Sarikaya and Çiner, 2015). Multiple glacial refugia have been proposed from fossil records, palaeoclimatic modelling and concordant phylogeographic patterns, located mainly in the regions adjacent to the Caspian and Black Seas, as well as in

eastern and western Anatolia (Bilgin, 2011; Tarkhnishvili *et al.*, 2012; Nakhutsrishvili *et al.*, 2015; Maharramova *et al.*, 2018; Parvizi et al., 2018, 2019). An important role in plant colonization has been played here by the mountains of the Anatolian Diagonal, extending diagonally across central and eastern Turkey, which have been recognized as a dispersal barrier as well as a north–south migration corridor for Euro-Siberian mountain species (Ansell *et al.*, 2011; Bilgin, 2011; Kaya and Çıplak, 2017; Parolly, 2020).

Most of the SW Asian landscape belongs to the Irano-Turanian floristic region, which is characterized by dry and hot summers, harsh winters and overall low precipitation (Djamali *et al.*, 2012; Dönmez and Yerli, 2018). Although xerophilous vegetation is predominant in the region, and many lineages adapted to dry environments show high rates of net diversification here (e.g. Moharrek *et al.*, 2019; Rudov *et al.*, 2020), there are still some wetland areas with favourable conditions for meso- and hygrophilous vegetation. They are scattered and patchily distributed, mainly in montane to alpine zones, e.g. on the south-facing slopes of the Taurus Mountains, on the Inner Anatolian volcanoes, and in the Zagros and Alborz Mountains (Parolly, 2004; Noroozi, 2020; Naqinezhad *et al.*, 2021). The most significant for hygrophytes, however, are the Colchis and Hyrcan regions, which are located disjunctively along the coasts of the Black and Caspian Seas, respectively, and are both characterized by a mild and humid climate that has prevailed since the late Tertiary (Denk *et al.*, 2001; Kikvidze and Ohsawa, 2001; Nakhutsrishvili *et al.*, 2015). Despite the global significance of wetlands, we have limited knowledge of the evolutionary history of hygrophytic flora in SW Asia. Wetlands harbour high biodiversity and provide several important ecosystem services (Millennium Ecosystem Assessment, 2005). They are highly dependent on water supply and are therefore sensitive to anthropogenic disturbance and climate changes, which can negatively affect their structure and functioning (Erwin, 2009; López-Merino *et al.*, 2011; Khelifa *et al.*, 2022). Wetland degradation and loss is, indeed, a worldwide phenomenon (Davidson, 2014). Knowledge of the species and genetic diversity, ecology and evolution of hygrophytes, especially in less-explored regions, is critical for effective wetland conservation and management. In the predominantly arid environment of SW Asia, evolutionary studies of wetland lineages characterized by disjunct and patchy distributions can also provide important insights into the spatiotemporal aspects of speciation and diversification processes and add to the biogeographic history of this area, which has been reconstructed mainly from xerophytic elements.

An excellent resource for studying the evolution of hygrophytes in SW Asia, especially in the Anatolia–Caucasus region, are species of *Cardamine* (Brassicaceae). *Cardamine* is a species-rich genus (~280 species; Marhold *et al.*, 2021 onwards) that originated in the middle to late Miocene (Huang *et al.*, 2020), with several centres of diversity formed around the world (Carlsen *et al.*, 2009). *Cardamine* species occur in diverse habitats, but most of them have adapted to moist sites. In Europe, three polyploid species complexes, the *C. amara*, *C. pratensis* and *C. raphanifolia* groups, have been traditionally recognized, comprising more than 20 taxa, all growing in wetland habitats (Lihová and Marhold, 2006; Marhold *et al.*, 2018). They have been found to form two related phylogenetic clades (Lihová *et al.*, 2004; Marhold *et al.*, 2004; Šlenker *et al.*, 2021; see also Carlsen *et al.*, 2009) that diverged around the Miocene–Pliocene transition (Huang *et al.*, 2020), and include also five poorly known species distributed in the adjacent Anatolia–Caucasus region, which are the focus of the present study: *C. lazica* Boiss. & Balansa, *C. wiedemanniana* Boiss., *C. uliginosa* M. Bieb., *C. seidlitziana* Albov, and *C. tenera* J.G.Gmel. ex C.A.Mey (Fig. 1). Reticulation and polyploidization have been identified as significant evolutionary processes in these species complexes in Europe (Marhold *et al.*, 2018) but, in contrast, only diploid chromosome numbers have been recorded from the Caucasus and Anatolia (12 records in total; Marhold *et al.*, 2021 onwards). Available literature data suggest that taxonomic confusion surrounds several of these species (e.g. dubious distinction between *C. lazica* and *C. wiedemanniana* and between *C. seidlitziana* and *C. acris* Griseb. from the Balkan Peninsula; see Supplementary Data Text S1 for a detailed taxonomic overview). Previous phylogenetic reconstructions also showed discrepancies between their genetic patterns on the one hand and ecological and morphological differentiation on the other (Lihová *et al.*, 2004; Marhold *et al.*, 2004). It was hypothesized that small-scale and altitudinal shifts during the Pleistocene may have brought *C. uliginosa*, *C. seidlitziana* and *C. tenera*, which presumably grow parapatrically along an altitudinal gradient, into contact and interspecific gene flow blurred their genetic distinction (Marhold *et al.*, 2004). In addition, a recent study revealed that the (sub)alpine *C. uliginosa* is closely related to *C. penzesii*, a species growing in floodplain forests along the SW Black Sea coast, and suggested reticulation traces in the evolutionary history of *C. penzesii* in which *C. uliginosa* was likely involved (Šlenker *et al.*, 2021).

Phylogenomic approaches using high-throughput DNA sequencing allow the analysis of hundreds to thousands of nuclear loci and provide orders of magnitude more genetic data than were available in plant phylogenetic studies in the past. They appear to be particularly efficient when dealing with conflicting or poorly resolved relationships due to either recent divergence or hybridization events (e.g. Morales-Briones *et al.*, 2018; Escudero *et al.*, 2020; Karbstein *et al.*, 2020; Reichelt *et al.*, 2021). Here, we apply the Hyb-Seq technique (Weitemier *et al.*, 2014), a sequence capture approach that reduces genomic complexity by enriching low-copy nuclear target loci (exons with flanking intronic and intergenic regions), in combination with genome skimming. The latter allows assembly of high-copy DNA regions, such as organellar DNA or ribosomal DNA, which can provide additional evolutionary insights, especially if interspecific hybridization is assumed (Morales-Briones *et al.*, 2018; Rose *et al.*, 2021; Yoo *et al.*, 2021). In this study, we combine phylogenomic Hyb-Seq data with morphometric and ecological niche analyses to elucidate the evolutionary history of the above-mentioned *Cardamine* species as representatives of the hygrophytic flora in the Anatolia–Caucasus region. We aimed to (1) determine which of the *Cardamine* species listed above can be recognized in the Anatolia–Caucasus region using an integrative approach that considers genetic, ecological and morphological divergence patterns, and to infer relationships between species; (2) search for traces of past or present interspecific gene flow while examining three alternative hypotheses: whether gene flow has hindered genetic divergence and thus speciation, disrupted established species boundaries in contact

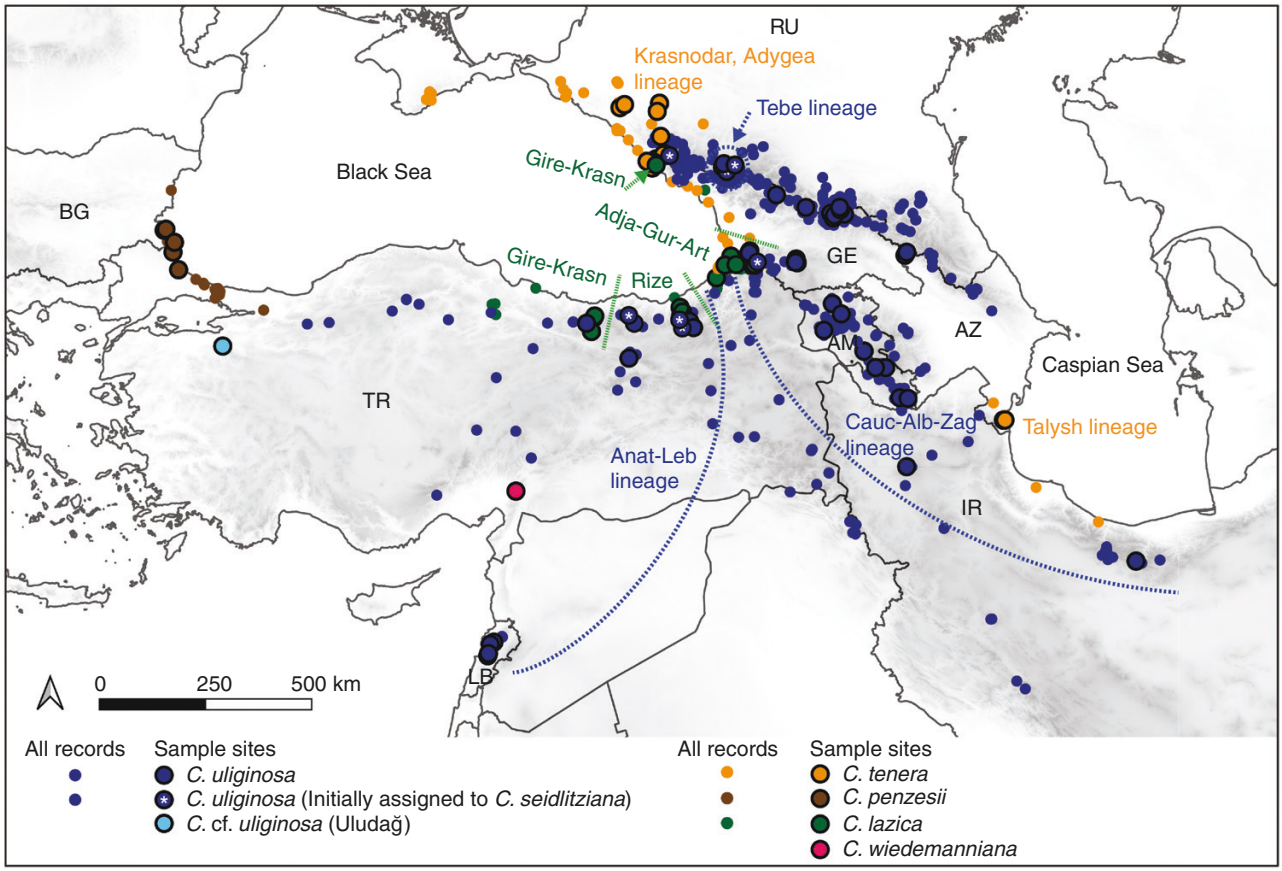


FIG. 1. Map showing occurrence records and sample sites for the target species in SW Asia: *C. uliginosa*, *C. tenera*, *C. lazica*, *C. wiedemanniana* and *C. penzesii*. Details of the sampled populations, including European relatives, are provided in Supplementary Data Table S1. Phylogeographic lineages recognized within *C. uliginosa*, *C. lazica* and *C. tenera* are indicated (abbreviations following Fig. 2), depicting also their approximate boundaries (dashed lines). Countries are identified by their two-letter ISO 3166-1 alpha-2 codes.

zones, or promoted speciation (e.g. through hybridogenous origins); and (3) examine intraspecific genetic structure to shed light on the phylogeographic history of these wetland species in both coastal and inland mountainous regions.

MATERIALS AND METHODS

Plant sampling

We sampled all *Cardamine* species from the target phylogenetic clades that occur in the Anatolia–Caucasus region: *C. uliginosa*, *C. seidlitziana*, *C. tenera*, *C. lazica*, *C. wiedemanniana* and *C. penzesii* (Fig. 1, Supplementary Data Table S1). The first three species, together with *C. acris* (endemic to the Balkan Peninsula), have also been referred to as the *C. tenera* group (Marhold *et al.*, 2004), and we use this designation here as well to facilitate references to the respective clade. Samples were obtained from multiple populations covering the species ranges (Fig. 1). When possible, we also included populations from the type localities or at least from the region or mountain range from which the species were described. The initial field identification of *C. seidlitziana* followed the morphology of the type material and diagnostic characters reported for this species (petal colour and leaflet shape; Khatri, 1988; Supplementary Data Text S1), although in some cases it was inconclusive. The population from the Uludağ Mountains, northwestern (NW) Turkey, usually attributed to either *C. uliginosa* or *C. tenera* (Cullen, 1965) but recently found to be distinct from these two species (Šlenker *et al.*, 2021), was also included and is referred to provisionally as *C.* cf. *uliginosa*. In total, we sampled 85 populations, of which 56 were included in Hyb-Seq analyses and 59 in morphometric analyses. This sampling in the Anatolia–Caucasus region was supplemented by 38 populations of the diploid relatives from Europe (Supplementary Data Table S1): *C. acris*, species of the *C. pratensis* group (*C. rivularis*, *C. matthioli*, *C. pratensis* s.str., *C. apennina*) and *C. amara*. For details on the distribution of these European species, see Lihová *et al.* (2004) and Šlenker et al. (2021). Three phylogenetically distinct diploids, *C. hirsuta*, *C. impatiens* and *C. trifolia*, were used as outgroups (Supplementary Data Table S1).

Chromosome number determination

Since the species groups studied here are polyploid complexes, we first determined the ploidy levels of analysed populations. For this purpose, we counted chromosomes in 40 populations from the Anatolia–Caucasus region (Supplementary Data Table S1). Squash preparations were made from the root

tips of cultivated plants according to the protocol described by Marhold *et al.* (2002).

Hyb-Seq library preparation, data processing and phylogenomic analyses

The Hyb-Seq technique has been applied to infer genetic divergence patterns between and within species, to clarify relationships between species, and to identify potential reticulation events. Both phylogenetic trees and networks were generated based on targeted nuclear loci, as well as plastome and nrDNA sequences obtained by genome skimming, as described below and with further details provided in Supplementary Data Text S2.

For target enrichment, we used *Cardamine*-specific probes designed as described in detail in Melichárková *et al.* (2020). In total, 14 464 120-mer biotinylated RNA baits, capturing 2246 exons from 1235 genes, were synthesized (myBaits® kit, Arbor Biosciences, MI, USA) and used to hybridize with target DNA. The Hyb-Seq libraries were prepared following the protocol described in detail in Šlenker et al. (2021). The libraries were sequenced with 150-bp paired-end reads on an Illumina MiSeq system at BIOCEV, Czechia.

Illumina reads were trimmed of adapters and low-quality bases in Trimmomatic v. 0.36 (Bolger *et al.*, 2014). PCR duplicates were removed in BBTools (https://jgi.doe.gov/data-and-tools/bbtools) using the Clumpify tool (see Supplementary Data Text S2A for more details). Consensus target sequences were assembled in HybPiper v. 1.3 (Johnson *et al.*, 2016), which utilizes BWA v. 0.7.13 (Li and Durbin, 2009), SPAdes v. 3.13 (Bankevich *et al.*, 2012) and Exonerate v. 2.2 (Slater and Birney, 2005). Targeted exons and flanking sequences ('supercontigs') were recovered using the intronerate.py script. HybPiper generates a single consensus sequence per individual, with heterozygous bases called as the nucleotide with the highest read frequency. The use of phased allele sequences, however, is crucial when analysing polyploids (Rothfels, 2021). In the case of allopolyploids, different homeologues or even chimeric sequences may be present in the assembled consensus sequences. Therefore, allele sequences were also inferred by read-backed phasing (Supplementary Data Text S2B) using the scripts and workflow available online at https://github.com/mossmatters/phyloscripts/tree/master/alleles_workflow, described in detail by Kates *et al.* (2018), but using the later versions of GATK and WhatsHap, which can handle polyploid genomes (Martin *et al.*, 2016; Schrinner *et al.*, 2020). When the phased sequences were divided into multiple blocks (Kates *et al.*, 2018), only the longest phase block was retained and the remaining interallelic variant sites were masked by inserting Ns at those positions. Both the consensus and allele data sets were used for downstream phylogenomic analyses (Supplementary Data Text S2C). Sequences were aligned using MAFFT v. 7.450 (Katoh and Standley, 2013); flanks and sites with gaps in >25% of sequences were removed using the R package ips (Heibl, 2008 onward) in R 4.0.0 (R Core Team, 2020). Consensus exon sequences were concatenated to genes, but phased exon sequences were not, as they represent unphasable blocks. Concatenation of aligned sequences was performed using AMAS (Borowiec, 2016). Further details are given in Supplementary Data Text S2C.

First, we analysed the diploid accessions (i.e. we omitted *C. wiedemanniana*, which turned out to be tetraploid; see Results) using the consensus sequences. Maximum likelihood (ML) trees were inferred from each of the assembled genes, as well as from their concatenation, the former followed by species tree reconstruction. The best-fit substitution models were assessed using the IQ-TREE's ModelFinder function (Chernomor *et al.*, 2016; Kalyaanamoorthy *et al.*, 2017) under the Bayesian information criterion. The ML trees were generated in RAxML-NG v. 0.9.0 (Kozlov *et al.*, 2019) and branch support was estimated by bootstrap replicates. Internal branches of low support (≤20 %) were collapsed in the gene trees used for species tree and network calculations, using Newick Utilities v. 1.6 (Junier and Zdobnov, 2010). Branch support of the concatenated ML tree was also assessed using the quartet sampling method (Pease *et al.*, 2018; https://www.github.com/fephyfofum/quartetsampling). The species tree was reconstructed under a multispecies coalescent model using ASTRAL-III (Zhang *et al.*, 2018), computing also local posterior probabilities to assess branch support (Sayyari and Mirarab, 2016). To estimate divergence times, we employed a rapid relaxed-clock dating method based on the relative rate framework, RelTime (Tamura et al., 2012, 2018), implemented in MEGA11 (Tamura *et al.*, 2021). The method was applied to the ML tree inferred from the concatenated genes. We used three secondary calibration points taken from the tribe-wide dated phylogeny (Huang *et al.*, 2020), setting the crown ages of the *C. amara* clade, the *C. tenera* plus *C. pratensis* clade, and the node of their divergence (their 95 % highest posterior density values, with prior normal distribution). Confidence intervals for the obtained divergence time estimates were computed in MEGA11 following Tao *et al.* (2020). More details and parameter settings are given in Supplementary Data Text S2C.

Next, we applied several approaches to resolve the polyploid origin of *C. wiedemanniana*. First, we extracted SNPs from Hyb-Seq reads mapped to the target exon sequences using the snipStrup pipeline [available online at https://github.com/MarekSlenker/snipStrup; described in detail in Melichárková *et al.* (2020)], which includes ploidy-aware SNP calling, and analysed them in STRUCTURE (Pritchard *et al.*, 2000) to detect potential admixture patterns. Next, we ran two network analyses based on the gene trees generated from consensus sequences: SNaQ implemented in PhyloNetworks, which is a pseudolikelihood method for generating explicit phylogenetic networks that infers reticulation events while accounting for incomplete lineage sorting (ILS) (Solís-Lemus and Ané, 2016); and an implicit, split network in SuperQ (Grünewald *et al.*, 2013; Bastkowski *et al.*, 2018). Then we proceeded with the phased exon sequences (allele data sets) and used three methods to identify the most likely parental species or lineages involved in the origin of *C. wiedemanniana*: GRAMPA (Thomas *et al.*, 2017); MPAllopp implemented in PhyloNet (Yan *et al.*, 2022); and AlleleSorting (Šlenker *et al.*, 2021). GRAMPA uses an algorithm for counting gene duplications and losses, adapted to work with multi-labelled trees. This makes it possible to distinguish between allo- and autopolyploids and to place polyploidy events on a phylogeny. MPAllopp is a newly developed maximum parsimony method that infers phylogenetic

networks while accounting for both ILS and polyploidy. In the AlleleSorting approach, alleles are sorted into two homeologues based on sequence divergence and labelled as 'A' or 'B' (https://github.com/MarekSlenker/AlleleSorting). The labelled homeologues, representing two subgenomes of the tetraploid, are then treated as independent accessions in the coalescent-based species tree inference in ASTRAL-III. The relatedness of the A and B homeologues to the alleles of the diploid species was also assessed by calculating pairwise distances in R 4.0.0 ((R Core Team, 2020) using the dist.dna function (Paradis and Schliep, 2019). An overview of the input and output data of the employed tools and details of the parameter settings is given in Supplementary Data Text S2.

Plastome sequences were extracted from Hyb-Seq reads and assembled using Fast-Plast v. 1.2.8 (available online at https://github.com/mrmckain/Fast-Plast) with default settings. This pipeline utilizes Trimmomatic v. 0.39 (Bolger *et al.*, 2014) for initial read cleaning, Bowtie 2 v. 2.3.5.1 (Langmead and Salzberg, 2012) to extract plastome reads using a database of reference plastomes, SPAdes v. 3.13 (Bankevich *et al.*, 2012) and afin (available online at https://github.com/mrmckain/Fast-Plast/tree/master/afin) for *de novo* sequence assembly. The plastome sequences obtained, comprising the large single copy (LSC), the small single copy (SSC) and one copy of the inverted repeats (IRb), were aligned using MAFFT v. 7.450 and an ML tree was inferred in RAxML-NG.

Sequences of nrDNA were obtained from the Hyb-Seq data using HybPiper and the reference sequence of *C. amara* (AY260579.1), as described above for the target nuclear loci, but omitting the supercontig option. The sequences were aligned using MAFFT v. 7.450, and the ITS region (ITS1-5.8S-ITS2) was extracted and retained for further analysis. The sequences recovered from HybPiper were subjected to read-backed phasing as described above to find multiple nrDNA variants per accession, if present. Maximum likelihood trees were constructed in RAxML-NG, and NeighbourNet graphs were generated in SplitsTree4 v. 4.14.4 (Huson and Bryant, 2006).

Modelling of environmental niches

To prove the assumed ecological differentiation among the studied species in the Anatolia–Caucasus region (see Introduction and Supplementary Data Text S1) and to test whether it played a role in speciation, we used 22 environmental variables to define and compare the environmental niches of the species (Supplementary Data Table S2). The environmental data were gathered from a total of 589 sites: 43 for *C. lazica*, 457 for *C. uliginosa*, 62 for *C. tenera* and 27 for *C. penzesii* (Fig. 1). Neither *C. wiedemanniana* nor the records from the Uludağ Mountains (*C.* cf. *uliginosa*) were included, since only very few sites have been confirmed for these entities.

We operationally delineated the environmental space available for colonization as the conditions occurring in a convex hull encompassing all occurrence sites plus a 200-km wide buffer zone. The environmental niches were evaluated using the kernel smoother approach proposed by Broennimann *et al.* (2012). The overlap of niches between the species was quantified using Schoener's *D* index (Schoener, 1968). The observed values of the index were tested for niche equivalency and niche similarity (Warren *et al.*, 2008) using randomization procedures outlined by Broennimann *et al.* (2012).

Finally, we compared the environmental niche breadths of the four species. The breadths were calculated as the dispersion of occurrence sites from their spatial medians in the ordination space of principal component analysis (PCA) based on the correlation matrix of environmental data. A distance-based test of homogeneity of multivariate dispersions (Anderson, 2006) was employed to compare environmental niche breadths among the species. Details of the environmental niche analysis are given in Supplementary Data Text S3. The analyses were performed in R 4.0.0 ((R Core Team, 2020) using the libraries ecospat (Broennimann *et al.*, 2022) and vegan (Oksanen *et al.*, 2019).

Morphometric analyses

Multivariate morphometric analyses aimed to investigate the morphological differentiation of the species from the Anatolia–Caucasus region, including their distinction from European relatives. We also specifically sought to identify populations that should correspond to the morphotype of *C. seidlitziana*. A total of 29 vegetative and floral characters were measured or scored on plants sampled in the field at the flowering stage and preserved as herbarium specimens (Table 1). Floral parts (one well-developed flower per individual) were attached by adhesive tape to a sheet of paper, dried and scanned. They were measured by QuickPHOTO Industrial v. 2.3 (Promicra, Prague, Czechia). Ratio characters were also derived, and in such cases at least one of the original characters was omitted from the morphometric analyses (see Table 1 for details). Measurements were also made on basal leaves, but these are not preserved in some taxa (especially *C. lazica* and *C. amara*) or individuals. Therefore, we assembled two data matrices: a matrix 'with basal leaves' that comprised all traits, including those measured on basal leaves, but on a limited number of individuals (31 characters × 1107 specimens); and a matrix 'without basal leaves' that omitted traits on basal leaves but included all sampled individuals (25 characters × 2058 specimens). A series of different morphometric analyses were performed to gain deeper insights into patterns of variation and differentiation (Klecka, 1980; Krzanowski, 2000; Marhold, 2011) and to identify the characters that best distinguish the species from each other: cluster analyses; standardized PCAs based on the correlation matrix; and canonical discriminant analyses (CDAs). The analyses were run at both the individual and the population level, and also for multiple data subsets comprising only selected populations or species. Spearman correlation coefficients were calculated prior to the analyses to detect and eliminate highly correlated characters that could bias the discriminant analyses. The analyses were performed in R 4.0.0 software ((R Core Team, 2020) using the MorphoTools2 package (Šlenker *et al.*, 2022). In addition, the surf function from the labdsv package (Roberts, 2019) was used to add surface contours for selected quantitative characters onto ordination graphs.

TABLE 1. *List of morphological characters measured or scored and included in the morphometric analyses*

Vegetative characters	
WS	Width of stem at the base (mm)
LSL*	Height of stem from its base to the base of the uppermost stem leaf (cm)
NB	Number of stem branches (including the main stem) bearing inflorescences
NB2	Number of lateral stem branches not bearing inflorescences; usually in lower part of the main stem
NL	Number of stem leaves
NLR	Degree of congestion of leaves beneath the inflorescence on the main stem, expressed as the number of leaves reaching the base of the uppermost leaf
LC1	Length of basal leaf (cm)
LC2†	Length of middle stem leaf (cm)
LC3*	Length of the uppermost leaf (cm)
NFB	Number of pairs of lateral leaflets of the basal leaf
LTB*	Length of terminal leaflet of the basal leaf (cm)
WTB*	Width of terminal leaflet of the basal leaf (cm)
LLB*	Length of first (distal) lateral leaflet of the basal leaf (cm)
WLB*	Width of first (distal) lateral leaflet of the basal leaf (cm)
NFS†	Number of pairs of lateral leaflets of the middle stem leaf
LTS*†	Length of terminal leaflet of the middle stem leaf (cm)
WTS*†	Width of terminal leaflet of the middle stem leaf (cm)
LLS*†	Length of first (distal) lateral leaflet of the middle stem leaf (cm)
WLS*†	Width of first (distal) lateral leaflet of the middle stem leaf (cm)
RHIZ	Growth orientation of the rhizome: (0) vertical (at 0–45° angle to the vertical axis) or (1) creeping (at 45–90° angle to the vertical axis)
BASE	Stem base: (0) erect or (1) ascending
STEM	Stem above the stem base: (0) erect or (1) ascending
HAIR	Stem base: (0) glabrous, (1) sparsely hairy, or (2) densely hairy
Floral characters	
LP	Petal length (mm)
WP*	Petal width (mm)
LS	Sepal length (mm)
LFL	Length of longer filaments (mm)
LFS*	Length of shorter filaments (mm)
COL	Petals: (0) white or (1) pink, violet or white with a pink tone
Ratio characters	
NL/LSL, LC2/LSL, LC3/LSL, LC3/LC2, WTB/LTB, WLB/LLB, LLB/LTB, WLB/WTB, WTS/LTS, WLS/LLS, LLS/LTS, WLS/WTS, WP/LP, LFS/LFL	

*Characters used only for computing ratios.

†The middle stem leaf is the leaf closest to the midpoint of the leafy part of the main stem (LSL/2 point).

RESULTS

Chromosome numbers

The diploid level with $2n = 16$ was determined here for *C. lazica*, *C. tenera* and *C. uliginosa* (including tentative *C. seidlitziana* populations). In contrast, the only population of *C. wiedemanniana* found and sampled from the Central Taurus (Amanos Mountains, southern Turkey) was found to be tetraploid with $2n = 32$ (Supplementary Data Table S1).

Hyb-Seq results

Sequence data summary. Sequencing of the Hyb-Seq libraries yielded an average of 2 798 470 reads per individual. On average, 50.65 % of the reads mapped to the target nuclear gene sequences, resulting in an average depth of coverage of 98.67 reads per base. Of the 2246 exons from 1235 genes, targeted by the designed RNA baits, 1763 consensus exon (supercontig) sequences (concatenated to 1125 genes) were obtained and assembled in all 78 samples analysed. The length of the exon alignments varied from 60 to 3672 bp (736 bp on average); that of the gene alignments varied from 60 to 8565 bp (1154 bp on average). The concatenated alignment of all nuclear genes was 1 298 288 bp long. Plastomes, assembled from the Hyb-Seq sequencing data, had an average depth of 13.01–129.36 reads per base (57.06 on average). The alignment of the plastomes was 127 736 bp long. The average depth of coverage of the ITS region of nrDNA ranged from 44.15 to 588.43 reads per base

(224.87 on average). Sequencing data are stored at the SRA database of NCBI (BioProject PRJNA830631).

Phylogenetic inferences in the diploids. The ML analysis of the diploid species, based on the concatenated dataset of all 1125 target nuclear genes, generated a tree with two main and well-supported clades (Fig. 2A). One clade comprised *C. amara* and *C. lazica* in a sister position, representing the *C. amara* group. The other main clade included members of the *C. pratensis* and *C. tenera* groups. It consisted of three subclades, each of which had strong support (bootstrap support [BS] = 100%, quartet concordance [QC] > 0.5), as follows: *C. acris* together with *C.* cf. *uliginosa* from the Uludağ Mountains (subclade acr-Ulu); then the members of the *C. pratensis* groups with the exception of *C. penzesii* (subclade prat); and finally *C. penzesii*, *C. tenera*, *C. uliginosa* and *C. seidlitziana* (subclade pen-ul-ten). However, the relationships among these three subclades were only weakly supported (BS = 45%, QC = 0.054, quartet differential [QD] = 0.77). Samples tentatively assigned to *C. seidlitziana* (marked with asterisks in Fig. 2) were scattered in various positions within the clade of *C. uliginosa*. They were generally clustered with the geographically closest accessions from higher altitudes representing the 'typical' morphotype of *C. uliginosa*. Within the clade of *C. uliginosa*, three geographically structured groups with strong to moderate support could be delineated: (1) accessions from the Caucasus (Russia, Georgia, Armenia) and the Zagros and Alborz Mountains in Iran (referred to as the Cauc-Alb-Zag lineage); (2) accessions from the Teberda region in

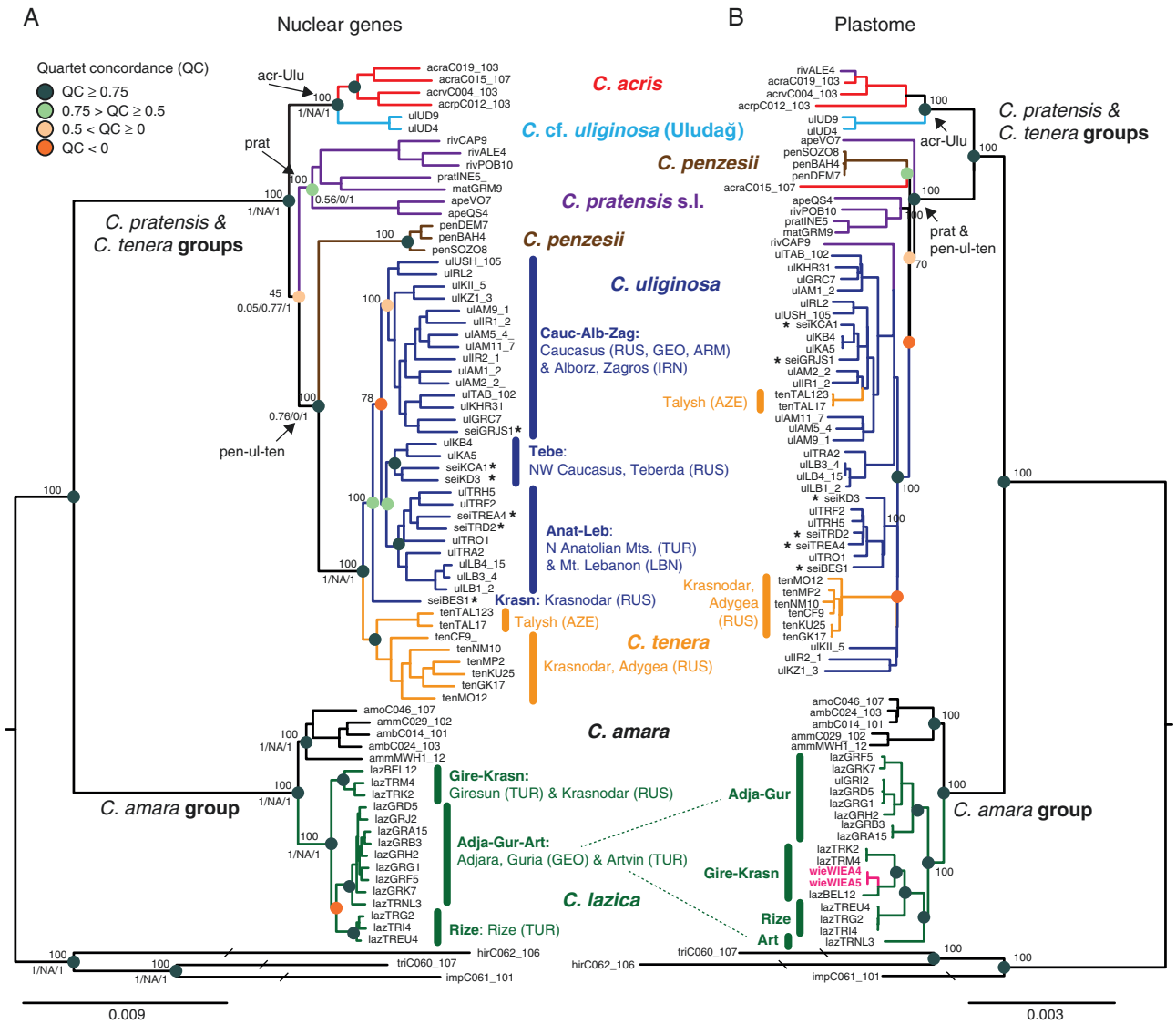


FIG. 2. (A) Maximum likelihood tree (RAxML-NG) based on the concatenated sequences of all 1125 target nuclear genes (consensus sequences) in diploid *Cardamine* samples. Branch support is shown with bootstrap values (BS > 50 %) and quartet sampling scores (QC/QD/QI) for the main branches; coloured circles in the nodes indicate QC value intervals. (B) Maximum likelihood tree (RAxML-NG) of the complete plastome data (LSC, SSC, IRb) with bootstrap values shown above the main branches. Note that, in contrast to (A), the tetraploid *C. wiedemanniana* ('wie' in pink) is also included here. Vertical bars indicate intraspecific groups designated by abbreviations used in the text. Country abbreviations in brackets follow ISO 3166-1 alpha-3 codes. Asterisks mark accessions tentatively determined as *C. seidlitziana* (see Supplementary Data Table S1 for locality details). Branches of the outgroup taxa are shortened.

Russia in the NW Caucasus (Tebe lineage); and (3) accessions from the North Anatolian Mountains and Lebanon (Anat-Leb lineage) (Fig. 1). The single accession collected in Krasnodarskii krai, Russia (the NW edge of the range) was placed separately. Geographic clustering was also observed within the clade of *C. tenera*: the accessions from the Caspian Sea region (Talysh Mountains) were differentiated from those from the Black Sea region (Krasnodar and Adygea, Russia). Finally, geographic patterns of strong support (high BS and QC) are also evident within *C. lazica*, with three clades in the North Anatolian Mountains structured from west to east (Figs 1 and 2A): (1) the westernmost Giresun region, which clusters with the sample from the northern Black Sea coast in the Krasnodarskii krai (Gire-Krasn lineage); (2) the Rize region (Rize lineage); and (3) the easternmost Artvin region, which clusters with the adjacent Georgian regions of Adjara and Guria (Adja-Gur-Art lineage).

The species tree inferred in ASTRAL from 1125 ML gene trees (Supplementary Data Fig. S1; samples initially assigned to *C. seidlitziana* are included here in *C. uliginosa*) was highly congruent with the ML tree of the concatenated dataset. The same three subclades were resolved within the clade of the *C. pratensis* and *C. tenera* groups, but their branching was equivocal, with low local posterior probabilities and very short branches.

The estimate of divergence times (Supplementary Data Fig. S2) suggested that the clade of the *C. pratensis* and *C. tenera* groups originated in the late Pliocene and their split into three subclades occurred almost simultaneously, with considerable temporal overlap in the late Pliocene to early Pleistocene. The divergence between *C. acris* and *C.* cf. *uliginosa* from Uludağ, and that between *C. uliginosa* and *C. tenera*, could be dated to the early Pleistocene. This was followed by intraspecific diversification that occurred at the end of the early Pleistocene. *Cardamine amara* and *C. lazica* probably separated from each other in the early Pleistocene, and diversification within *C. lazica* began in the early to middle Pleistocene.

The ML tree based on plastome sequences displayed the same two main clades as the nuclear ML tree (Fig. 2B). Within the clade of the *C. amara* group, the internal structure was also highly congruent with the nuclear tree, including geographic clustering within *C. lazica*. Within the other main clade of the plastid tree (comprising the *C. pratensis* and *C. tenera* groups), only two subclades (compared with three of the nuclear tree) were resolved: *C. acris* clustered with *C.* cf. *uliginosa* from the Uludağ Mountains (acr-Ulu subclade) and the remaining species clustered together (prat and pen-ul-ten subclade). Thus, members of the *C. pratensis* groups could not be well separated from those of the *C. tenera* group. Samples of *C. tenera* were found to be nested within *C. uliginosa*, with two geographically structured subclades (Caspian and Black Sea regions). Some patterns of geographic clustering were also present within *C. uliginosa*, but only with respect to some smaller, terminal clades. Accessions initially attributed to *C. seidlitziana* were scattered within the clade of *C. uliginosa*, similar to the nuclear tree. Few misplaced accessions of *C. acris* and *C. rivularis* were also observed that contradicted the position of the other conspecific samples.

Resolving the origin of the tetraploid Cardamine wiedemanniana

Both the supernetwork (SuperQ) analysis and Bayesian STRUCTURE clustering suggest an allopolyploid origin of *C. wiedemanniana*, with ancestors from the two main clades of diploids one clade being the *C. amara* group and the other clade comprising the *C. pratensis* and *C. tenera* groups (Supplementary Data Fig. S3). In the ML tree of plastomes (Fig. 2B), the accessions of *C. wiedemanniana* were placed within the clade of *C. lazica*, which identified its maternal ancestor. The species network analyses SNaQ and MPAllopp and the GRAMPA analysis supported the allopolyploid origin, but slightly differed in the identification of the ancestral species or lineages (Fig. 3A). Both MPAllopp and GRAMPA confirmed *C. lazica* as the maternal ancestor, whereas SNaQ pointed to the lineage including both *C. amara* and *C. lazica* (BS = 95.7 %). The paternal ancestor was inferred to be either *C. tenera* (GRAMPA), *C. uliginosa* (SNaQ, BS = 98.9 %), or the lineage including both these sister species (MPAllopp). The results of all three analyses remained the same whether all genes/exons or only those with a length of at least 800 bp were considered (see below).

A more detailed examination of the allotetraploid scenarios was performed in AlleleSorting. Allele phasing in *C. wiedemanniana* resulted in 9.02 % homozygous exons, 11.37 % fully heterozygous and 79.61 % partially heterozygous (with two or three different alleles) exons. The optimal threshold for unequivocal allele sorting was found to be four (for details of threshold evaluation see Šlenker *et al.*, 2021). A total of 55.73 % exon sequences passed the threshold and their alleles could be unambiguously sorted into homeologue A (alleles closer to the *C. lazica* plus *C. amara* clade, i.e. the maternal progenitor clade) and homeologue B. To construct a species tree in ASTRAL, we assembled several exon sets, testing the effect of exon length on the phylogenetic placement of the A and B homeologues and on their relatedness (genetic distance) to the alleles present in the diploid species. The length of the included exons did not affect the phylogenetic placement of the A homeologue. It was resolved in the sister position to *C. lazica*, as also shown in PhyloNet and GRAMPA analyses, and closest to the Gire-Krasn lineage, which agrees with the plastid phylogeny (Figs 2B and 3B, Supplementary Data Fig. S4). Some variation, however, was observed in the position of the B homeologue. In the exon sets that contained exons at least 600 bp long (915 exons) or 700 bp long (657 exons), the B homeologue was placed in a sister position to the clade comprising *C. uliginosa* and *C. tenera*. When only the longer exons were considered, i.e. those that were at least 800 bp (502 exons), 900 bp (409 exons), 1000 bp (355 exons) or 1100 bp (304 exons) long, the B homeologue was consistently resolved as sister to *C. uliginosa*, and its Anat-Leb lineage (Fig. 3B, Supplementary Data Fig. S4). When pairwise interallelic distances were calculated, we found that, with increasing exon length, the proportion of phased alleles in *C. wiedemanniana* that were closest to the identified parental species was higher. In contrast, when considering only short exons (up to 300 or 600 bp), a certain proportion of B alleles in *C. wiedemanniana* resembled a wide range of different diploids (Supplementary Data Fig. S5). Thus, the fact that the B homeologue in exon sets containing exons at least 600 or 700 bp long was placed in a sister position to the *C. uliginosa* plus *C. tenera* clade was most likely due to the ambiguous phylogenetic signal of short sequences. Removal of exons with low information value improved phylogenetic resolution.

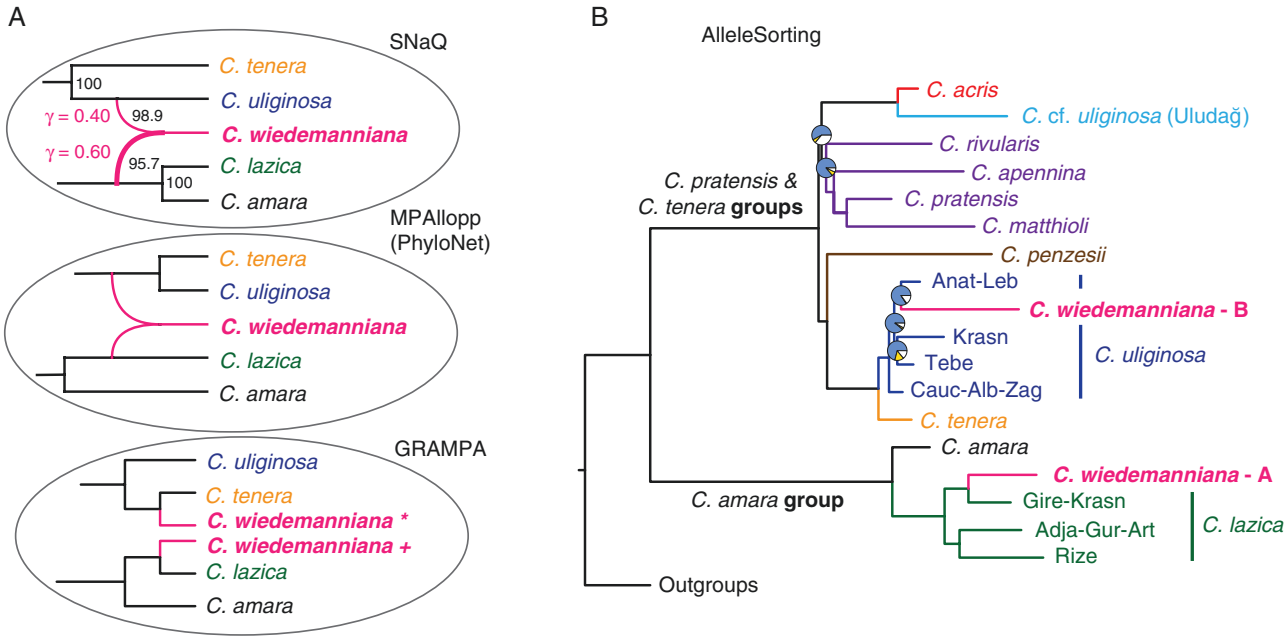


FIG. 3. Phylogenetic networks and trees examining the allotetraploid origin of *C. wiedemanniana*. (A) Networks inferred by SNaQ (based on 1125 consensus sequences of nuclear genes; includes inheritance probabilities γ in pink and bootstrap support in black letters), MPAllopp in PhyloNet and a multi-label species tree reconstructed in GRAMPA (both based on 1763 exon trees inferred from phased allele sequences). Only the respective clades with the inferred progenitors are shown (see Supplementary Data Fig. S3 for the complete trees). (B) Species tree inferred in ASTRAL-III based on 502 exon trees, those generated from phased allele sequences of exons at least 800 bp long, where allele pairs in *C. wiedemanniana* were sorted using AlleleSorting tool into A (corresponding to the maternal parent based on plastome data) and B homeologues. Phylogeographic lineages detected in *C. uliginosa* and *C. lazica* (Fig. 2A) were defined here as separate entities. Branch support is indicated by pie charts, depicting three local posterior probabilities for the given branch (not shown for the fully supported branches). Alternative species trees based on different sets of exons, with or without splitting *C. uliginosa* and *C. lazica* into phylogeographic units, are shown in Supplementary Data Fig. S4.

Patterns in nrDNA variation

The analyses of nrDNA sequences showed the same two main clades (ML tree not shown; see Supplementary Data Fig. S6 for the NeighborNet) as obtained in the phylogenies inferred from the target nuclear genes and plastomes. Phasing of nrDNA sequences in tetraploid *C. wiedemanniana* revealed divergent nrDNA copy variants, further supporting its allotetraploid origin. Among the diploids, phased nrDNA sequences grouped largely according to species, but high variation in nrDNA copy variants was found in *C. uliginosa* and *C. tenera* (the latter nested within the former, as also seen in the plastid phylogeny). Two diploid species, *C. apennina* and *C. penzesii*, showed some intraspecific divergence in their nrDNA sequences; in *C. penzesii*, each of the three sequenced individuals had one nrDNA variant that was placed close to the *C. acris* and *C. pratensis* group, and the other within the cluster of *C. uliginosa*.

Environmental niches

The analysis of niche overlap revealed that the environmental niches of *C. lazica* and *C. tenera* were equivalent, with a Schoener's *D* value of 0.66. In contrast, the overlaps in other species pairs were low ($D = 0.0002–0.029$), indicating a high degree of dissimilarity between their niches ($P > 0.05$ for niche similarity tests; Table 2). While the niches of *C. lazica* and *C. tenera* are typical of higher precipitations and lower precipitation seasonality, *C. penzesii* and *C. uliginosa* prefer lower humidity. The latter two species are separated along the temperature gradient – *C. penzesii* grows in warmer conditions at lower altitudes, whereas *C. uliginosa* grows in colder conditions in the mountains (Fig. 4; Supplementary Data Table S2, Supplementary Data Fig. S7).

The species showed significantly heterogeneous niche breadths ($F_m = 12.4$, $P < 0.001$). The environmental niche of *C. penzesii* was significantly narrower than the niches of the other species (all Tukey contrasts $P < 0.001$). The niche breadths of *C. lazica*, *C. tenera* and *C. uliginosa* were statistically comparable among each other ($P > 0.05$).

Morphometrics

Spearman correlations between the measured morphological traits did not exceed the commonly used threshold values of 0.90–0.95. The highest values were at most 0.8 and therefore all characters were retained in the morphometric analyses. A set of CDAs showed that *C. lazica* is readily distinguishable from the other diploid species from the Anatolia–Caucasus region (mainly by fewer leaves and leaflets and smaller floral parts) as well as from *C. amara* (by longer stem leaves and different shape of leaflets; Supplementary Data Fig. S8, Supplementary Data Table S3, CDA4–6). A CDA with three *a priori* defined groups within *C. lazica* (phylogeographic lineages as revealed by Hyb-Seq data) showed some tendency towards their morphological

TABLE 2. *Results of niche equivalency and niche similarity tests of four* Cardamine *species. Observed niche overlaps (Schoener's* D*), 95 % ranges of simulated* D *values under the null hypothesis (lower and upper limits) and associated probabilities (*P*) are displayed*

		Equivalency			Similarity		
Species contrast	***D***	Lower limit	Upper limit	***P***	Lower limit	Upper limit	***P***
C. lazica–C. penzesii	< 0.001	0.204	0.804	< 0.001	0.000	0.033	0.511
C. lazica–C. tenera	0.660	0.612	0.909	0.089	–	–	–
C. lazica–C. uliginosa	0.126	0.454	0.720	< 0.001	0.000	0.745	0.567
C. penzesii–C. tenera	0.007	0.107	0.795	< 0.001	0.000	0.058	0.246
C. penzesii–C. uliginosa	0.001	0.280	0.724	< 0.001	0.000	0.055	0.491
C. tenera–C. uliginosa	0.029	0.544	0.770	< 0.001	0.000	0.696	0.477

differentiation, albeit with overlap (Supplementary Data Fig. S8, CDA7).

Cluster analyses in the *C. tenera* group, which also included *C. penzesii*, revealed two main clusters (Fig. 5A). Cluster 1 comprised populations of *C. uliginosa* (including samples tentatively assigned to *C. seidlitziana*) and *C. penzesii*. Cluster 2 comprised *C. tenera*, *C. acris* and the population from the Uludağ Mountains. CDA with *C. uliginosa* and *C. penzesii* (cluster 1) as two groups showed clear differentiation between these two species, mainly caused by branched stems, larger petals, erect stem and tuberous rhizome in the latter species (Fig. 5B, CDA1, Supplementary Data Table S3). A CDA with three groups of cluster 2, i.e. *C. tenera*, *C. acris* and the population from the Uludağ Mountains, also revealed a high degree of morphological differentiation with minimal overlap (Fig. 5D, CDA2). *Cardamine tenera* and *C. acris* were differentiated along the first canonical axis, with the population from the Uludağ Mountains occupying an intermediate position along the first axis but separated along the second axis. A combination of characters mainly related to the size and shape of leaves, leaflets, petals and sepals contributed to their separation (Supplementary Data Table S3)

The CDA of *C. uliginosa* (including the samples tentatively assigned to *C. seidlitziana*) with three predefined groups corresponding to the lineages suggested by the Hyb-Seq data showed some tendency towards their morphological differentiation, but with many overlaps (Fig. 5C, CDA3). The assignment of the studied populations to *C. seidlitziana* was often ambiguous. Petal colour proved to be an unreliable trait for identification, as even within the typical *C. uliginosa* (i.e. populations from the highest altitudes and with narrow leaflets) there is a large variation: ~60 % of the individuals had at least tinted petals, but not infrequently had also completely pink to deep violet petals. Another character highlighted in *C. seidlitziana*, the shape of terminal leaflets, varied widely both among and within multiple populations of *C. uliginosa* (ratio values WTS/LTS, WLS/WTS, WTB/LTB, WLB/WTB) and showed a largely continuous variation. Mapping of the WLS/WTS trait in CDA (Fig. 5C, CDA3) revealed that wider terminal leaflets are found mainly in some Anatolian and Lebanese populations, whereas the narrowest are found in the Teberda region.

Both PCA (Supplementary Data Fig. S8, PCA) and CDA (not shown) of *C. wiedemanniana* and its inferred parental species (*C. uliginosa* and *C. lazica*) showed that this tetraploid is morphologically closer to *C. uliginosa* (placed at its variation margin) than to *C. lazica*.

DISCUSSION

Evolution of hygrophytic Cardamine *species in the Anatolia–Caucasus region is characterized by recent speciation rather than extensive reticulation*

The occurrence of plant lineages adapted to wetland habitats is fairly limited in SW Asia due to the dominance of the Irano-Turanian biogeographic region with a dry continental climate, which favours the spread and diversification of xerophytic flora (Noroozi, 2020). The distribution and diversification of hygrophytes here is largely associated with the more favourable Caucasus region and the refugial temperate rainforest areas of the Colchis and Hyrcan adjacent to the Black and Caspian Seas (Biltekin *et al.*, 2015; Nakhutsrishvili *et al.*, 2015). The present study fully supports these patterns. Three of the studied species, *C. lazica*, *C. tenera* and *C. penzesii*, occur exclusively in the areas close to the seas, from lowlands to the montane zone, while only the (sub)alpine *C. uliginosa* has spread throughout the inland areas and grows across several mountain ranges. Altitudinal differentiation was reported between the three species *C. uliginosa*, *C. seidlitziana* and *C. tenera*, but the confusing morphological patterns, lack of genetic differentiation (inferred from Sanger sequences and amplified fragment length polymorphism markers) on the one hand, and high intragenomic ITS sequence variation on the other hand, appeared puzzling (Lihová *et al.*, 2004; Marhold *et al.*, 2004). It has been suggested that extensive interspecific gene flow may have occurred in these species, presumably enhanced by small-scale and altitudinal range shifts in response to climate fluctuations during the Pleistocene (Hewitt, 2011; Nieto Feliner, 2011; and see Marhold *et al.*, 2004). Nevertheless, the present results allow us to reject the hypothesis of extensive reticulate evolution. Using an integrative approach (Dayrat, 2005; de Queiroz, 2007) that included phylogenomic data and morphological and ecological niche analyses, we prove the distinction of *C. tenera* but refuse the recognition of *C. seidlitziana* as a separate entity.

The morphological delimitation of *C. seidlitziana* turned out to be vague, but not with respect to *C. acris* from the Balkans, as assumed (Khatri, 1988; see also Supplementary Data Text S1), but with respect to *C. uliginosa*. The diagnostic characters reported for *C. seidlitziana* (petal colour, leaflet shape) vary widely in *C. uliginosa*. For some traits, we found clinal patterns of variation, commonly seen in broadly

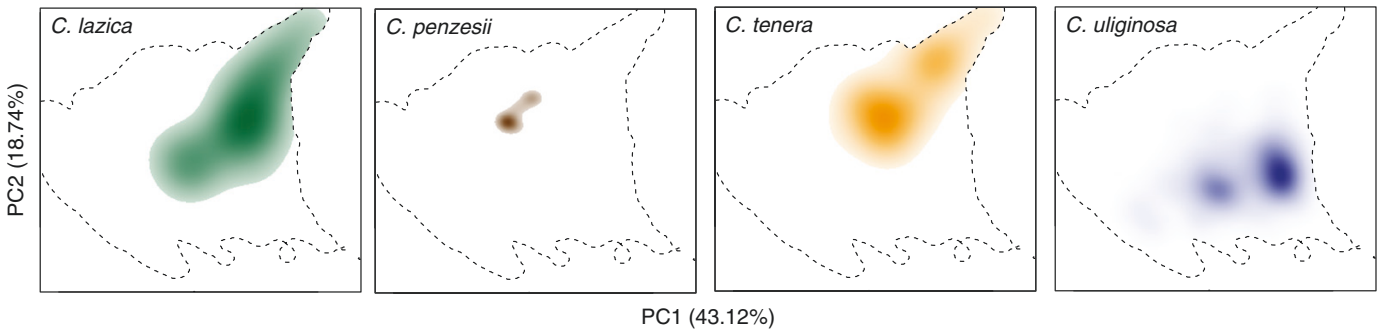


FIG. 4. Environmental niche differentiation of four *Cardamine* species displayed in principal component space as kernel densities with shading proportional to the occurrence density. The dashed lines indicate the available environmental space. Variance explained by the components is given in parentheses. For details of environmental variables and principal component space, see Supplementary Data Table S2 and Supplementary Data Fig. S7.

distributed species as a response to environmental gradients (e.g. Etterson *et al.*, 2008). Genetic clustering patterns in the *C. uliginosa* clade largely followed geographic origins, implying unrestricted gene flow between adjacent lower-altitude (initially attributed to *C. seidlitziana*) and upper-altitude populations (typical *C. uliginosa*). Given all of these patterns, we propose to merge the traditionally recognized *C. seidlitziana* with *C. uliginosa*. On the contrary, the lack of genetic divergence between *C. tenera* and *C. uliginosa* reported in the earlier phylogenetic studies (Lihová *et al.*, 2004; Marhold *et al.*, 2004) can be clearly attributed to recent speciation. The Hyb-Seq nuclear data resolved *C. tenera* and *C. uliginosa* as sister species that likely diverged in the early Pleistocene, and their distinction is supported by ecological and morphological differentiation. In the plastome and nrDNA data, two or three *tenera*-specific sequence clusters were nested within the larger sequence variation of *C. uliginosa*, most likely reflecting incomplete lineage sorting. Speciation of *C. uliginosa* and *C. tenera*, as well as their closest relative, *C. penzesii*, likely involved ecological adaptations, but was also reinforced by geographic partitioning. Environmental niche modelling showed that these species, growing in allo- or parapatry, occupy significantly distinct niches. Recently, a reticulate evolutionary history was proposed for *C. penzesii* (Šlenker *et al.*, 2021), but the present data do not support this. The eastern edge of its range is very close to the western one of *C. uliginosa*, so some interspecific gene flow may have occurred in the past, particularly in response to climatic changes during the Pleistocene. Divergent nrDNA copy variants within *C. penzesii* could support this, but otherwise the nuclear genes targeted here do not provide convincing evidence for genome admixture in *C. penzesii*. Genome-wide screening for admixture patterns may help to investigate this hypothesis in more detail in the future (e.g. Martin *et al.*, 2013).

The taxonomically uncertain population from the Uludağ Mountains, NW Anatolia (provisionally *C.* cf. *uliginosa*), likely diverged from the Balkan *C. acris* in the early Pleistocene. Their common ancestor may have had a larger and more continuous range during the late Pliocene, when there were land connections between the Balkans and western Anatolia and sea levels were lower (Chobanov *et al.*, 2016), but has become fragmented later on. Currently, these sister lineages grow in allopatry, with the Aegean Sea, Sea of Marmara and Thracian plain representing significant geographic barriers to gene flow or dispersal. The phylogenetic relationships observed here are in good agreement with known biogeographic and evolutionary links between Anatolia and the Balkans (e.g. Ansell *et al.*, 2011; Bilgin, 2011; Kaya and Çiplak, 2017; Özüdoğru and Mummenhoff, 2020).

Allopolyploid speciation in southern Anatolia: the rare case of Cardamine wiedemanniana

Our results clearly show that *C. lazica* and *C. wiedemanniana*, which were previously confused (Khatri, 1988; Supplementary Data Text S1), are two different species. We have demonstrated that *C. wiedemanniana* has an allopolyploid origin and have identified intraspecific lineages within *C. lazica* and *C. uliginosa* that gave rise to this tetraploid. Based on the time estimates for intraspecific diversification within the diploids, we can infer that *C. wiedemanniana* arose in recent times, probably no earlier than the middle Pleistocene. Its distribution is very limited, so far confirmed only from the northern part of the Amanos region in southern Turkey, which has a Mediterranean climate but relatively high annual precipitation (Parolly, 2020). The occurrence of hygrophytic species in the southern and central Anatolian regions is indeed very rare, and among the related species only *C. uliginosa* has been recorded from a few scattered sites in neighbouring areas.

The polyploid origin of *C. wiedemanniana* has been addressed here using several approaches based on either consensus or phased allele sequences. The study of polyploid genomes using short sequence reads remains a challenge, although several tools have been developed recently (Thomas *et al.*, 2017; Lautenschlager *et al.*, 2020; Rothfels, 2021; Šlenker *et al.*, 2021). Here, we examined several commonly used phylogenetic network methods such as PhyloNet, GRAMPA and SNaQ (e.g. Morales-Briones *et al.*, 2018; Amarasinghe *et al.*, 2021; Guo *et al.*, 2021; Hodel *et al.*, 2022) and a newly developed AlleleSorting tool (Šlenker *et al.*, 2021), which are based on different assumptions and algorithms (Solís-Lemus and Ané, 2016; Thomas *et al.*, 2017; Yan *et al.*, 2022), and found that they differed slightly in identifying ancestral lineages. Based on this experience, we recommend exploring and applying

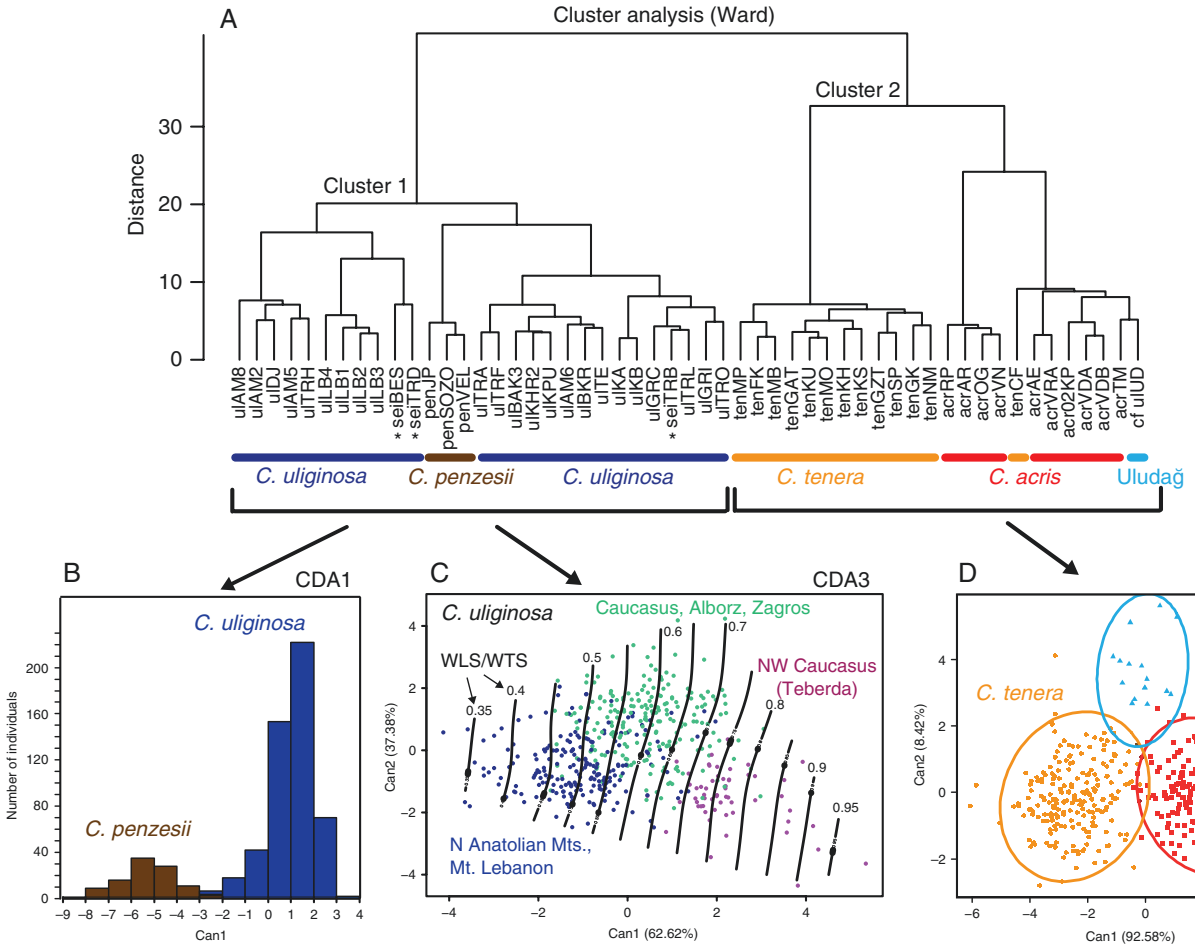


FIG. 5. Morphometric analyses of the *Cardamine* species based on the characters of the 'with basal leaves' matrix. (A) Cluster analysis (Ward's method) of 53 populations of the *C. tenera* group plus *C. penzesii*. Populations initially attributed to *C. seidlitziana* are included in *C. uliginosa* and indicated by asterisks (see Supplementary Data Table S1 for locality details). (B) CDA1, with two predefined groups, *C. uliginosa* (510 individuals) and *C. penzesii* (103 individuals), which formed cluster 1 in (A). (C) CDA3, based on 482 individuals of *C. uliginosa*, and three predefined groups corresponding to phylogeographic groups as resolved by the Hyb-Seq data (Fig. 2A). The character WLS/WTS (Table 1) was fitted and plotted on the ordination diagram. (D) CDA2, with three predefined groups, *C. acris* (259 individuals), *C. tenera* (219 individuals) and *C.* cf. *uliginosa* from Uludağ (16 individuals), which formed cluster 2 in (A). Prediction ellipses define the regions where any new independent observations from the respective taxa will fall with 95 % probability. Can1 and Can2 stand for canonical axis 1 and 2. The values of the total canonical structure, expressing the correlations of the characters with the ordination axes, is given in Supplementary Data Table S3.

multiple approaches with different settings and data filtering when studying polyploid origins.

Spatial genetic divergence argues for long-term population isolation and the existence of multiple glacial refugia

Genetic variation within *C. lazica*, *C. tenera* and *C. uliginosa* is spatially structured, suggesting reduced gene flow due to existing geographic and ecological factors, such as long distances, landscape barriers or environmental dissimilarities (Wang *et al.*, 2013), but may also reflect their phylogeographic history, glacial survival in different refugia, and postglacial colonization routes (Avise, 2009).

The two lineages distinguished in *C. tenera*, estimated to have diverged in the early Pleistocene, correspond to the geographic disjunction of this species. The distribution range of this species, inhabiting lowland to lower montane sites, is largely confined to the Colchis and Hyrcan regions, which have been known as refugial areas for mesophytic trees and shrubs since the late Tertiary (Denk *et al.*, 2001). The intervening Caucasian mountains represent a significant dispersal barrier for hygrophytic lowland species; therefore, the genetic structure found in *C. tenera* was indeed expected and suggests long-term population separation in disjunct refugia along the Black Sea and Caspian Sea coasts (Maharramova *et al.*, 2015; Nakhutsrishvili *et al.*, 2015). Environmental niche analyses have shown that *C. lazica* grows in very similar habitats, but its range is restricted to the Colchis region, where it occurs in the sea-facing mountain valleys with high precipitation rates throughout the year (Nakhutsrishvili *et al.*, 2015; Parolly, 2020). In this species, we found strong genetic differentiation even within a relatively small area of northeastern Anatolia, structured in a west–east direction and dated back to the early to middle Pleistocene. This pattern suggests that mountain ridges rising from sea level to peaks above 3000

m.a.s.l. (Parolly, 2020) have been significant barriers to gene flow in the west–east direction. Hence, *C. lazica* likely experienced only small-scale, altitudinal and north–south shifts within the valleys in response to Pleistocene climatic oscillations. These results clearly support the concept of refugia within refugia (Gómez and Lunt, 2007; Nieto Feliner, 2014), which revises previous simplifications commonly associated with large refugial areas.

Finally, three geographically structured lineages have been identified in *C. uliginosa*, estimated to have originated in the early to middle Pleistocene. *Cardamine uliginosa* is the only species that spreads over large inland mountainous areas in SW Asia and may have been more severely affected by Pleistocene glaciations. Glacier development was significant in the western and central Caucasus, and eastern North Anatolian Mountains, which was related to high humidity in these regions (Gobejishvili *et al.*, 2011; Sarikaya and Çiner, 2015; Tielidze, 2017). Otherwise, Pleistocene glaciations were mostly restricted to the highest mountain peaks, and the late Pleistocene snowline was much higher in SW Asia than in Europe (Kehl, 2009; Sarikaya and Çiner, 2015; Tielidze, 2017). This left a wide altitudinal range unglaciated and provided a diversity of habitats that could have been colonized by species from higher altitudes (Djamali *et al.*, 2012). Indeed, suitable ecological conditions and widespread occurrence have been hypothesized for alpine *Arabis*, *Noccaea* and *Cousinia* species in Anatolia during the glacial periods (Ansell *et al.*, 2011; Djamali *et al.*, 2012; Özüdoğru *et al.*, 2020). However, because the glacial periods were generally characterized by dry climate and the expansion of steppe vegetation (Webb and Bartlein, 1992; Biltekin *et al.*, 2015; Noroozi, 2020), it remains unclear whether this scenario also applies to alpine hygrophytes such as *C. uliginosa*. Within the two widespread intraspecific lineages of *C. uliginosa*, even a finer geographic structure is evident. For example, two clusters can be delineated in northeastern Anatolia, one of which includes populations from the Lebanon Mountains, supporting the Anatolian Diagonal as an important migration corridor between these areas (Bilgin, 2011; Parolly, 2020; e.g. Ansell *et al.*, 2011; Kaya and Çiplak, 2017). These genetic divergence patterns also suggest the existence of multiple refugia within the two main lineages of *C. uliginosa*. The third lineage points to the Teberda region (NW Caucasus), which was heavily glaciated (Gobejishvili *et al.*, 2011), with the highest peaks at 4000 m and populations of *C. uliginosa* recorded at altitudes up to 2800 m. It is likely that populations of *C. uliginosa* descended there to lower altitudes and survived in refugia on the northern foothills, keeping this lineage isolated from adjacent populations in the central Greater and Lesser Caucasus. Species distribution modelling, which can predict areas of suitable habitats in both the present and past, can be an efficient approach to complementing inferences based on genetic data, to prove, revise or refine various phylogeographic scenarios (e.g. Alvarado-Serrano and Knowles, 2014). Nevertheless, the available and predominantly used bioclimatic variables may not accurately capture some specific microclimatic or topographic conditions, such as those of wetland sites. In fact, we have not been able to obtain reasonable predictions of species distribution based on bioclimatic variables (results not shown).

Conclusions

Here we support the recognition of five wetland *Cardamine* species from the Anatolia–Caucasus target region, which could be delimited based on the concordant phylogenomic, morphological and ecological divergence patterns. The hygrophytic character restricted them to mild and humid refugial areas near the Black and Caspian Seas (*C. tenera*, *C. lazica* and *C. penzesii*), to a very limited extent near the Mediterranean Sea (*C. wiedemanniana*) or to the (sub)alpine belt over large inland areas (*C. uliginosa*). An isolated NW Anatolian population (provisionally *C.* cf. *uliginosa*) shares a common ancestor with the Balkan *C. acris*. We have identified three phylogenetic lineages in the traditionally recognized *C. pratensis* and *C. tenera* groups that evolved and diversified relatively rapidly and recently, probably in the late Pliocene to early Pleistocene. Speciation in the Anatolia–Caucasus region was driven mainly by geographic separation (allopatry) and ecological divergence. The initial hypothesis of extensive reticulate evolution, which was suspected based on previous phylogenetic studies and known patterns in the European relatives, was rejected. With the exception of an allotetraploid stenoendemic *C. wiedemanniana*, interspecific gene flow and polyploidization have remained limited here. Spatially structured genetic variation within the studied species suggests limited gene flow due to existing geographic and ecological barriers, but also favours glacial survival in several refugia.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/aob and consist of the following. Text S1: taxonomic overview of the studied *Cardamine* species from the Anatolia–Caucasus region. Text S2: Hyb-Seq data processing: overview of the employed tools, input data, details of the parameter settings, and output data. Text S3: modelling of environmental niches. Table S1: list of *Cardamine* species and populations analysed in the present study, including NCBI accession codes and chromosome number records. Table S2: mean values and ranges of environmental variables at georeferenced sites of *C. lazica*, *C. penzesii*, *C. tenera* and *C. uliginosa*. Table S3: results of PCAs and CDAs based on morphometric data of the studied species. Figure S1: species tree inferred for diploid *Cardamine* in ASTRAL-III from the target nuclear genes. Figure S2: chronogram based on the ML tree of the concatenated sequences of the target nuclear genes in diploid *Cardamine* with divergence time estimation using the RelTime method. Figure S3: phylogenetic analyses examining the allotetraploid origin of *C. wiedemanniana*. Figure S4: species trees inferred in ASTRAL-III from different sets of exon trees generated from phased allele sequences of both diploids and tetraploid *C. wiedemanniana*, the latter sorted into homeologues using the AlleleSorting tool. Figure S5: genetic relatedness of alleles retrieved from *C. wiedemanniana* to the alleles observed in the diploids, depicted in pie charts. Figure S6: NeighbourNet graph of nrDNA sequences assembled from Hyb-Seq data with multiple nrDNA sequence variants per accession obtained by read-back phasing. Figure S7: environmental niche differentiation of *C. lazica*, *C. penzesii*, *C. tenera* and *C. uliginosa* displayed in principal component space. Figure S8: CDAs and PCAs

showing morphological differentiation of *C. lazica* from related species.

FUNDING

This work was supported by the Slovak Research and Development Agency (APVV-21-0044) and the Turkish-Slovak joint research project JRP [SAS—TUBITAK no. 475542 (ELTtoEFDi)]. M.Sv. was supported by the Operational Programme Integrated Infrastructure (OPII), funded by the ERDF (ITMS 313011T721).

ACKNOWLEDGEMENTS

We thank several botanists who accompanied us during field trips or provided us with samples: Marián Perný, Martin Kolník (both formerly from the Institute of Botany, SAS, Bratislava, Slovakia), Anna Bérešová, Milan Valachovič (Plant Science and Biodiversity Centre, SAS, Bratislava, Slovakia), Hiba Dokmak (Lebanese Agricultural Research Institute, Zahleh, Lebanon), Filip Kolář (Charles University, Prague), Vladimir Onipchenko (Moscow State University, Russia) and Boris S. Tuniev (Sochi National Park, Sochi, Russia). We also thank Michaela Caboňová (Plant Science and Biodiversity Centre, SAS, Bratislava, Slovakia) for initial laboratory assistance. Computational resources were supplied by the project e-Infrastruktura CZ (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

AUTHOR CONTRIBUTIONS

J.Z.-L. supervised the research and wrote the paper with contribution of M.Š. and M.Sv. A.K., M.Š., M.Sv., J.K. and J.Z.-L. generated, analysed and interpreted data. K.M., J.K., M.Sl., A.A.D., J.B. and J.Z.-L. participated in plant sampling. All authors participated in interpretation and discussion of results and contributed with comments and suggestions that improved the final draft of the manuscript.

LITERATURE CITED

Alvarado-Serrano DF, **Knowles LL. 2014**. Ecological niche models in phylogeographic studies: applications, advances, precautions. *Molecular Ecology Resources* **14**: 233–248.

Amarasinghe P, **Pham P**, **Stone RD**, **Cellinese N. 2021**. Discordance in a South African *Memelycon* clade (Melastomataceae): evidence for reticulate evolution. *International Journal of Plant Sciences* **182**: 6828–6694.

Anderson MJ. 2006. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**: 245–253.

Ansell SW, **Stenøien HK**, **Grundmann M**, ***et al***. **2011**. The importance of Anatolian mountains as the cradle of global diversity in *Arabis alpina*, a key arctic–alpine species. *Annals of Botany* **108**: 241–252.

Avise JC. 2009. Phylogeography: retrospect and prospect. *Journal of Biogeography* **36**: 3–15.

Bankevich A, **Nurk S**, **Antipov D**, ***et al***. **2012**. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.

Bastkowski S, **Mapleson D**, **Spillner A**, **Wu T**, **Balvociute M**, **Moulton V. 2018**. SPECTRE: a suite of phylogenetic tools for reticulate evolution. *Bioinformatics* **34**: 1056–1057.

Bilgin R. 2011. Back to the suture: the distribution of intraspecific genetic diversity in and around Anatolia. *International Journal of Molecular Sciences* **12**: 4080–4103.

Biltekin D, **Popescu SM**, **Suc JP**, ***et al***. **2015**. Anatolia: a long-time plant refuge area documented by pollen records over the last 23 million years. *Review of Palaeobotany and Palynology* **215**: 1–22.

Bolger AM, **Lohse M**, **Usadel B. 2014**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4**: e1660.

Broennimann O, **Fitzpatrick MC**, **Pearman PB**, ***et al***. **2012**. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* **21**: 481–497.

Broennimann O, **Di Cola V**, **Guisan A. 2022**. *ecospat: spatial ecology miscellaneous methods. R package version 3.2.1*. https://CRAN.R-project.org/package=ecospat

Carlsen T, **Bleeker W**, **Hurka H**, **Elven R**, **Brochmann C. 2009**. Biogeography and phylogeny of *Cardamine* (Brassicaceae). *Annals of the Missouri Botanical Garden* **96**: 215–236.

Chernomor O, **von Haeseler A**, **Minh BQ. 2016**. Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology* **65**: 997–1008.

Chobanov DP, **Kaya S**, **Grzywacz B**, **Warchałowska-Śliwa E**, **Çıplak B. 2016**. The Anatolio-Balkan phylogeographic fault: a snapshot from the genus *Isophya* (Orthoptera, Tettigoniidae). *Zoologica Scripta* **46**: 165–179.

Cullen J. 1965. *Cardamine* L. In: **Davis PH**, **Cullen J**, **Coode MJE**, eds. *Flora of Turkey and the East Aegean islands, Vol. 1*. Edinburgh: Edinburgh University Press, 438–444.

Davidson NC. 2014. How much wetland has the world lost? Long-term and recent trends in global wetland area. *Marine and Freshwater Research* **65**: 934–941.

Dayrat B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* **85**: 407–415.

Denk T, **Frotzler N**, **Davitashvili N. 2001**. Vegetational patterns and distribution of relict taxa in humid temperate forests and wetlands of Georgia (Transcaucasia). *Biological Journal of the Linnean Society* **72**: 287–332.

Djamali M, **Brewer S**, **Breckle SW**, **Jackson ST. 2012**. Climatic determinism in phytogeographic regionalization: a test from the Irano-Turanian region, SW and Central Asia. *Flora* **207**: 237–249.

Dönmez A, **Yerli SV. 2018**. Biodiversity in Turkey. In: **Pullaiah T**, ed. *Global Biodiversity. Volume 2: selected countries in Europe*. Oakville: Apple Academic Press, 397–442.

Erwin KL. 2009. Wetlands and global climate change: the role of wetland restoration in a changing world. *Wetlands Ecology and Management* **17**: 71–84.

Escudero M, **Nieto Feliner G**, **Pokorny L**, **Spalink D**, **Viruel J. 2020**. Editorial: Phylogenomic approaches to deal with particularly challenging plant lineages. *Frontiers in Plant Science* **11**: 591762.

Etterson JR, **Delf DE**, **Craig TP**, **Ando Y**, **Ohgushi T. 2008**. Parallel patterns of clinal variation in *Solidago altissima* in its native range in central USA and its invasive range in Japan. *Botany* **86**: 91–97.

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

Gobejishvili R, **Lomidze N**, **Tielidze L. 2011**. Late Pleistocene (Würmian) glaciations of the Caucasus. In: **Ehlers J**, **Gibbard PL**, **Hughes PD**, eds. *Quaternary glaciations: extent and chronology*. Amsterdam: Elsevier, 141–147.

Gómez A, **Lunt DH. 2007**. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: **Weiss S**, **Ferrand N**, eds. *Phylogeography of southern European refugia*. Dordrecht: Springer, 155–188.

Grünewald S, **Spillner A**, **Bastkowski S**, **Bögershausen A**, **Moulton V. 2013**. SuperQ: computing supernetworks from quartets. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **10**: 151–160.

Guo X, **Mandáková T**, **Trachtová K**, **Özüdoğru B**, **Liu J**, **Lysak MA. 2021**. Linked by ancestral bonds: multiple whole-genome duplications and reticulate evolution in a Brassicaceae tribe. *Molecular Biology and Evolution* **38**: 1695–1714.

Heibl C. 2008. *PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software packages*. http://www.christophheibl.de/Rpackages.html (23 February 2022).

Hewitt GM. 2011. Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**: 617–638.

Hodel RGJ, **Zimmer EA**, **Liu BB**, **Wen J. 2022**. Synthesis of nuclear and chloroplast data combined with network analyses supports the polyploid origin of the apple tribe and the hybrid origin of the Maleae–Gillenieae clade. *Frontiers in Plant Science* **12**: 820997.

Huang XC, **German DA**, **Koch MA. 2020**. Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. *Annals of Botany* **125**: 29–47.

Huson DH, **Bryant D. 2006**. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.

Johnson MG, **Gardner EM**, **Liu Y**, ***et al***. **2016**. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* **4**: 1600016.

Junier T, **Zdobnov EM. 2010**. The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics* **26**: 1669–1670.

Kalyaanamoorthy S, **Minh BQ**, **Wong TK**, **von Haeseler A**, **Jermiin LS. 2017**. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

Karbstein K, **Tomasello S**, **Hodač L**, **Dunkel FG**, **Daubert M**, **Hörandl E. 2020**. Phylogenomics supported by geometric morphometrics reveals delimitation of sexual species within the polyploid apomictic *Ranunculus auricomus* complex (Ranunculaceae). *Taxon* **69**: 1191–1220.

Kates HR, **Johnson MG**, **Gardner EM**, **Zerega NJC**, **Wickett NJ. 2018**. Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus*. *American Journal of Botany* **105**: 404–416.

Katoh K, **Standley DM. 2013**. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kaya S, **Çiplak B. 2017**. Phylogeography and taxonomy of the *Psorodonotus caucasicus* (Orthoptera, Tettigoniidae) group: independent double invasion of the Balkans from the Caucasus. *Systematic Entomology* **42**: 118–133.

Kehl M. 2009. Quaternary climate change in Iran – the state of knowledge. *Erdkunde* **63**: 1–17.

Khatri KS. 1988. Studies in the Caucasian species of *Cardamine* L. (Cruciferae). *Phyton* **28**: 55–80.

Khelifa R, **Mahdjoub H**, **Samways MJ. 2022**. Combined climatic and anthropogenic stress threaten resilience of important wetland sites in an arid region. *Science of the Total Environment* **806**: 150806.

Kikvidze Z, **Ohsawa M. 2001**. Richness of Colchic vegetation: comparison between refugia of south-western and East Asia. *BMC Ecology* **1**: 6.

Klecka WR. 1980. *Discriminant analysis. Quantitative applications in the social sciences, Vol. 19*. Beverly Hills: Sage Publications.

Koch MA, **Karl R**, **German DA. 2017**. Underexplored biodiversity of Eastern Mediterranean biota: systematics and evolutionary history of the genus *Aubrieta* (Brassicaceae). *Annals of Botany* **119**: 39–57.

Kozlov AM, **Darriba D**, **Flouri T**, **Morel B**, **Stamatakis A. 2019**. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Krzanowski WJ. 2000. *Principles of multivariate analysis: a user's perspective*. Oxford: Oxford University Press.

Langmead B, **Salzberg SL. 2012**. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Lautenschlager U, **Wagner F**, **Oberprieler C. 2020**. ALLCOPOL: inferring allele co-ancestry in polyploids. *BMC Bioinformatics* **21**: 441.

Li H, **Durbin R. 2009**. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Lihová J, **Marhold K. 2006**. Phylogenetic and diversity patterns in *Cardamine* (Brassicaceae) – a genus with conspicuous polyploid and reticulate evolution. In: **Sharma AK**, **Sharma A**, eds. *Plant genome: biodiversity and evolution, Vol. 1C: Phanerogams (Angiosperms – Dicotyledons)*. Enfield: Science Publishers, 149–186.

Lihová J, **Fuertes Aguilar J**, **Marhold K**, **Nieto Feliner G. 2004**. Origin of the disjunct tetraploid *Cardamine amporitana* (Brassicaceae) assessed with nuclear and chloroplast DNA sequence data. *American Journal of Botany* **91**: 1231–1242.

López-Merino L, **Martínez Cortizas A**, **López-Sáez JA. 2011**. Human-induced changes on wetlands: a study case from NW Iberia. *Quaternary Science Reviews* **30**: 2745–2754.

Maharramova EH, **Safarov HM**, **Kozlowski G**, **Borsch T**, **Muller LA. 2015**. Analysis of nuclear microsatellites reveals limited differentiation between Colchic and Hyrcanian populations of the wind-pollinated relict tree Zelkova carpinifolia (Ulmaceae). *American Journal of Botany* **102**: 119–128.

Maharramova E, **Huseynova I**, **Kolbaia S**, **Gruenstaeudl M**, **Borsch T**, **Muller LAH. 2018**. Phylogeography and population genetics of the riparian relict tree *Pterocarya fraxinifolia* (Juglandaceae) in the South Caucasus. *Systematics and Biodiversity* **16**: 14–27.

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.

Manafzadeh S, **Staedler YM**, **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* **92**: 1365–1388.

Marhold K. 2011. Multivariate morphometrics and its application to monography at specific and infraspecific levels. In: **Stuessy TF**, **Lack HW**, eds. *Monographic plant systematics: fundamental assessment of plant biodiversity*. Ruggell: A. R. G. Gantner: 73–99.

Marhold K, **Lihová J**, **Perný M**, **Grupe R**, **Neuffer B. 2002**. Natural hybridization in *Cardamine* (Brassicaceae) in the Pyrenees: evidence from morphological and molecular data. *Botanical Journal of the Linnean Society* **139**: 275–294.

Marhold K, **Lihová J**, **Perný M**, **Bleeker W. 2004**. Comparative ITS and AFLP analysis of diploid *Cardamine* (Brassicaceae) taxa from closely related polyploid complexes. *Annals of Botany* **93**: 507–520.

Marhold K, **Šlenker M**, **Zozomová-Lihová J. 2018**. Polyploidy and hybridization in the Mediterranean and neighbouring areas towards the north: examples from the genus *Cardamine* (Brassicaceae). *Biologia Serbica* **40**: 47–59.

Marhold K, **Kempa M**, **Kučera J**, ***et al.*** **2021 onwards.** Datababase of names, chromosome numbers, ploidy levels and genome sizes of the tribe Cardamineae. https://cardamine.sav.sk (1 March 2022).

Martin SH, **Dasmahapatra KK**, **Nadeau NJ**, ***et al***. **2013**. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**: 1817–1828.

Martin M, **Patterson M**, **Garg S**, ***et al***. **2016**. WhatsHap: fast and accurate read-based phasing. *BioRxiv*. doi: 10.1101/085050. 14 November 2016, preprint: not peer reviewed.

Melichárková A, **Šlenker M**, **Zozomová-Lihová J**, ***et al***. **2020**. So closely related and yet so different: strong contrasts between the evolutionary histories of species of the *Cardamine pratensis* polyploid complex in Central Europe. *Frontiers in Plant Science* **11**: 588856.

Millennium Ecosystem Assessment. 2005. *Ecosystems and human well-being: wetlands and water synthesis*. Washington, DC: World Resources Institute.

Mittermeier RA, **Turner WR**, **Larsen FW**, **Brooks TM**, **Gascon C. 2011**. Global biodiversity conservation: the critical role of hotspots. In: **Zachos F**, **Habel J**, eds. *Biodiversity hotspots*. Berlin: Springer, 3–22.

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

Morales-Briones DF, **Liston A**, **Tank DC. 2018**. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist* **218**: 1668–1684.

Nakhutsrishvili G, **Zazanashvili N**, **Batsatsashvili K**, **Montalvo Mancheno CS. 2015**. Colchic and Hyrcanian forests of the Caucasus: similarities, differences and conservation status. *Flora Mediterranea* **25**: 185–192.

Naqinezhad A, **Nowak A**, **Świerszcz S**, ***et al***. **2021**. Syntaxonomy and biogeography of the Irano-Turanian mires and springs. *Applied Vegetation Science* **24**: e12571.

Nieto Feliner G. 2011. Southern European glacial refugia: a tale of tales. *Taxon* **60**: 365–372.

Nieto Feliner G. 2014. Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics* **16**: 265–278.

Noroozi J (ed.). 2020 *Plant biogeography and vegetation of high mountains of Central and South-West Asia*. Cham: Springer Nature.

Noroozi J, **Talebi A**, **Doostmohammadi M**, **Rumpf SB**, **Linder HP**, **Schneeweiss GM. 2018**. Hotspots within a global biodiversity hotspot – areas of endemism are associated with high mountain ranges. *Scientific Reports* **8**: 10345.

Noroozi J, **Talebi A**, **Doostmohammadi M**, **Manafzadeh S**, **Asgarpour Z**, **Schneeweiss GM. 2019**. Endemic diversity and distribution of the Iranian vascular flora across phytogeographical regions, biodiversity hotspots and areas of endemism. *Scientific Reports* **9**: 12991.

Oksanen J, **Blanchet FG**, **Friendly M *et al.* 2019**. *vegan: Community Ecology Package. R package version 2.5-6*. https://CRAN.R-project.org/package=vegan

Özüdoğru B, **Mummenhoff K. 2020**. Phylogenetic and biogeographical history confirm the Anatolian origin of *Bornmuellera* (Brassicaceae) and clade divergence between Anatolia and the Balkans in the Plio-Pleistocene transition. *Turkish Journal of Botany* **44**: 593–603.

Özüdoğru B, **Karacaoğlu C**, **Akaydın G**, **Erik S**, **Mummenhoff K**, **Sağlam IK. 2020**. Ecological specialization promotes diversity and diversification in the Eastern Mediterranean genus *Ricotia* (Brassicaceae). *Journal of Systematics and Evolution* **60**: 331–343.

Paradis E, **Schliep K. 2019**. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528.

Parolly G. 2004. The high mountain vegetation of Turkey – a state of the art report, including a first annotated conspectus of the major syntaxa. *Turkish Journal of Botany* **28**: 39–63.

Parolly G. 2020. The Anatolian high-mountain ranges – plant diversity between two seas. In: **Noroozi J**, ed. *Plant biogeography and vegetation of high mountains of Central and South-West Asia*. Cham: Springer Nature, 215–286.

Parvizi E, **Naderloo R**, **Keikhosravi A**, **Solhjouy-Fard S**, **Schubart CD. 2018**. Multiple *Pleistocene refugia* and repeated phylogeographic breaks in the southern Caspian Sea region: Insights from the freshwater crab *Potamon ibericum*. *Journal of Biogeography* **45**: 1234–1245.

Parvizi E, **Keikhosravi A**, **Naderloo R**, **Solhjouy-Fard S**, **Sheibak F**, **Schubart CD. 2019**. Phylogeography of *Potamon ibericum* (Brachyura: Potamidae) identifies Quaternary glacial refugia within the Caucasus biodiversity hot spot. *Ecology and Evolution* **9**: 4749–4759.

Pease JB, **Brown JW**, **Walker JF**, **Hinchliff CE**, **Smith SA. 2018**. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal of Botany* **105**: 385–403.

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

de Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* **56**: 879–886.

R Core Team. 2020. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Reichelt N, **Wen J**, **Pätzold C**, **Appelhans MS. 2021**. Target enrichment improves phylogenetic resolution in the genus *Zanthoxylum* (Rutaceae) and indicates both incomplete lineage sorting and hybridization events. *Annals of Botany* **128**: 497–510.

Roberts DW. 2019. *labdsv: Ordination and multivariate analysis for ecology. R package version 2.0-1*. https://CRAN.R-project.org/package=labdsv

Rose JP, **Toledo CAP**, **Lemmon EM**, **Lemmon AR**, **Sytsma KJ. 2021**. Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus *Polemonium* (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Systematic Biology* **70**: 162–180.

Rothfels CJ. 2021. Polyploid phylogenetics. *New Phytologist* **230**: 66–72.

Rudov A, **Mashkour M**, **Djamali M**, **Akhani H. 2020**. A review of C_4 plants in Southwest Asia: an ecological, geographical and taxonomical analysis of a region with high diversity of C_4 eudicots. *Frontiers in Plant Science* **11**: 546518.

Sarikaya MA, **Çiner A. 2015**. Late Pleistocene glaciations and paleoclimate of Turkey. *Bulletin of the Mineral Research and Exploration* **151**: 107–127.

Sayyari E, **Mirarab S. 2016**. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* **33**: 1654–1668.

Schoener TW. 1968. The *Anolis* lizards of Bimini: resource partitioning in a complex fauna. *Ecology* **49**: 704–726.

Schrinner SD, **Serra Mari R**, **Ebler J**, ***et al*. 2020**. Haplotype threading: accurate polyploid phasing from long reads. *Genome Biology* **21**: 252.

Slater GS, **Birney E. 2005**. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**: 31.

Šlenker M, **Kantor A**, **Marhold K**, ***et al*. 2021**. Allele sorting as a novel approach to resolving the origin of allotetraploids using Hyb-Seq data: a case study of the Balkan mountain endemic *Cardamine barbaraeoides*. *Frontiers in Plant Science* **12**: 659275.

Šlenker M, **Koutecký P**, **Marhold K. 2022**. MorphoTools2: an R package for multivariate morphometric analysis. *Bioinformatics* **38**: 2954–2955.

Solís-Lemus C, **Ané C. 2016**. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics* **12**: e1005896.

Tamura K, **Battistuzzi FU**, **Billing-Ross P**, **Murillo O**, **Filipski A**, **Kumar S. 2012**. Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences of the USA* **109**: 19333–19338.

Tamura K, **Tao Q**, **Kumar S. 2018**. Theoretical foundation of the RelTime method for estimating divergence times from variable evolutionary rates. *Molecular Biology and Evolution* **35**: 1770–1782.

Tamura K, **Stecher G**, **Kumar S. 2021**. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* **38**: 3022–3027.

Tao Q, **Tamura K**, **Mello B**, **Kumar S. 2020**. Reliable confidence intervals for RelTime estimates of evolutionary divergence times. *Molecular Biology and Evolution* **37**: 280–290.

Tarkhnishvili D, **Gavashelishvili A**, **Mumladze L. 2012**. Palaeoclimatic models help to understand current distribution of Caucasian forest species. *Biological Journal of the Linnean Society* **105**: 231–248.

Thomas GWC, **Ather SH**, **Hahn MW. 2017**. Gene-Tree reconciliation with MUL-trees to resolve polyploidy events. *Systematic Biology* **66**: 1007–1018.

Tielidze L. 2017. Late Pleistocene and Holocene glaciation. In: **Tielidze L**, ed. *Glaciers of Georgia*. Cham: Springer, 129–167.

Wang IJ, **Glor RE**, **Losos JB. 2013**. Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters* **16**: 175–182.

Warren DL, **Glor RE**, **Turelli M. 2008**. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62**: 2868–2883.

Webb T, **Bartlein PJ. 1992**. Global changes during the last 3 million years: climatic controls and biotic responses. *Annual Review of Ecology and Systematics* **23**: 141–173.

Weitemier K, **Straub SC**, **Cronn RC**, ***et al*. 2014**. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* **2**: 1400042.

Yan Z, **Cao Z**, **Liu Y**, **Ogilvie HA**, **Nakhleh L. 2022**. Maximum parsimony inference of phylogenetic networks in the presence of polyploid complexes. *Systematic Biology* **71**: 706–720.

Yoo MJ, **Lee BY**, **Kim S**, **Lim CE. 2021**. Phylogenomics with Hyb-Seq unravels Korean *Hosta* evolution. *Frontiers in Plant Science* **12**: 645735.

Zhang C, **Rabiee M**, **Sayyari E**, **Mirarab S. 2018**. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**: 153.

4.3 Balkan-Anatolian biogeographic links and the evolutionary significance of Anatolian mountains evidenced by *Cardamine* (Brassicaceae)

Submitted to Taxon in April 2023

Kantor, A., Šlenker, M., Kučera, J., Marhold, K., Dönmez, A. A., Yüzbaşıoğlu, S., & Zozomová-Lihová, J.

Fig. 5. Scanned herbarium specimen of the holotype of *C. anatolica*, a newly described hygrophytic species from W Anatolia circumscribed in the present thesis.

My contribution: I participated in the plant material collection; in the laboratory work (DNA extraction); measured the morphological traits from herbarium specimens and scanned flowers material, then performed the morphometric analyses in R software. As part of this study, I did the flow-cytometric measurements of relative genome size of all available material from SW Asian *Cardamine* species + samples of *C. acris*. I extracted the distribution data; wrote the first version of the manuscript and further worked on improving of the final version.

Balkan-Anatolian biogeographic links and the evolutionary significance of Anatolian mountains evidenced by *Cardamine* (Brassicaceae)

Running title: *Cardamine* in Anatolian mountains

Adam Kantor[1,2], Marek Šlenker[1], Jaromír Kučera[1], Karol Marhold[1,2], Ali A. Dönmez[3], Sırrı Yüzbaşıoğlu[4], Judita Zozomová-Lihová[1]

[1]*Institute of Botany, Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Bratislava, Slovakia*
[2]*Department of Botany, Faculty of Science, Charles University, Prague, Czechia*
[3]*Botany Section, Department of Biology, Faculty of Science, Hacettepe University, Beytepe-Ankara, Türkiye*
[4]*Department of Botany, Faculty of Science, Istanbul University, Vezneciler, Istanbul, Türkiye*
Address for correspondence: Judita Zozomová-Lihová, judita.zozomova@savba.sk

Abstract

Anatolia is a significant center of biodiversity and endemism with diversity hotspots located mainly in mountain ranges. Its complex geological history and heterogeneous topography have generated natural barriers to gene flow that favor speciation, and migration corridors that accentuate its transitional biogeographic position. While more attention has been paid to the predominant Irano-Turanian and Mediterranean xerophytic elements, the evolution of species adapted to wet habitats with limited occurrence is understudied in this area. Here, we investigated *Cardamine* representatives in northern Anatolia with the aims of resolving the taxonomically uncertain populations from northwestern Anatolia (the Uludağ, previously assigned to either *C. uliginosa* or *C. tenera*) and elucidating the genetic structure of (sub)alpine *C. uliginosa* recorded mainly from the North Anatolian Mts. (Pontic Mts.). Using a combination of phylogenomic (Hyb-Seq), morphometric, and flow cytometric analyses, we support a distinct position of the northwestern Anatolian populations, described here as a new species *Cardamine anatolica*. Apart from the Uludağ, a few other sites were found in the montane to subalpine belts in the Marmara and Aegean regions. A sister phylogenetic position of *C. anatolica* to *C. acris*, a widespread and polymorphic Balkan species, supports the existence of biogeographic links between these areas and favors a vicariance scenario. We revealed a pronounced intraspecific diversification of *C. uliginosa* with geographic structuring and admixture in the Pontic Mts., which highlights this area as a significant hotspot of biodiversity not only at the species level but also at the level of genetic variation. Due to the common misinterpretation of the species treated here, we revise their distributional data, provide details on their morphological differentiation, and present an identification key. The study highlights the evolutionary importance of Anatolian mountains, which promote speciation, favor accumulation of diversity, and serve as a meeting place of colonization routes.

Keywords *Cardamine anatolica, Cardamine uliginosa*, endemics, Hyb-Seq, morphometrics

INTRODUCTION

Anatolia represents an area of global importance in terms of its high biodiversity, endemism, and transitional biogeographic position, historically fulfilling the role of interconnection between Europe, Asia, and Africa. Three global biodiversity hotspots are partially located in this area (Caucasian, Irano-Anatolian, and Mediterranean; Mittermeier & al., 2011; Noroozi & al., 2019, 2022), and three phytogeographic regions meet there (Euro-Siberian, Irano-Turanian and Mediterranean; Ekim & Güner, 2000; Parolly, 2020). About 10,150 seed plant species have been reported from Türkiye, of which nearly one-third are endemics (Dönmez & Yerli, 2018). Several centers of endemism have been identified here, and all of them include high mountains (Noroozi & al., 2019), which is in accordance with general patterns of increased species richness in mountains due to their high topographic complexity, environmental heterogeneity, and buffering capacity in response to climatic fluctuations (Körner, 2004; Perrigo & al., 2020). Anatolia has a complex geological history and contains several natural barriers to gene flow that can promote speciation. One of these is the semi-arid Central Anatolian Plateau, which was covered by the brackish Central Anatolian Lake System in the Pliocene and largely inhabited by the Irano-Turanian elements, providing a biogeographic barrier to species with higher freshwater requirements for millions of years (Kosswig, 1955; Bilgin, 2011; Kürschner & Parolly, 2012; Parolly, 2020). The Anatolian Diagonal mountain chain has been considered another geographic barrier that limited contact between western and eastern lineages of numerous groups (Bilgin, 2011; Gür, 2016; Manafzadeh & al., 2017). The Anatolian Diagonal has also been accentuated as a major migration corridor, being part of the Taurus colonization route (Ansell & al., 2011; Bilgin, 2011; Kaya & Çıplak, 2017; Parolly, 2020), while the North Anatolian Mountains (Pontic Mts.) may have played a similar role in northern Türkiye (Kaya & Çıplak, 2017; Özüdoğru & Mummenhoff, 2020). Both the Anatolian Diagonal and the Pontic Mts. intersect two phytogeographic regions and two biodiversity hotspots and represent important areas for plant diversity in Anatolia (Noroozi & al., 2019; Parolly, 2020). Anatolia has also played a significant role in shaping European biodiversity, being recognized as both a center of diversification and a refuge area for many plant and animal lineages that subsequently spread to the Balkans (and other parts of Europe; Médail & Diadema, 2009; Bilgin, 2011; Thompson, 2020). The significance of a biogeographic link between Europe and Asia via Anatolia is also evidenced by the occurrence of Balkan elements in the extant flora of Anatolia (Ekim & Güner, 2000; Parolly, 2020) and, conversely, Anatolian elements in the flora of the Balkans (Strid, 1986; Stevanović, 1996).

A review of phylogeographic studies in Anatolia (Bilgin, 2011) shows that one of the commonly observed patterns has placed a suture zone (a genetic break or a contact zone) within Anatolia, splitting it into the western and eastern parts (see also Barker & al., 2017). The western Anatolian populations were genetically clustering with the Balkan ones, while the eastern populations were distinct or exhibited closer affinities with the Caucasus and Caspian Sea regions. This pattern could be attributed to the biogeographic barriers mentioned above, as well as physical connections between the Balkans and western Anatolia that occurred in the past (from the Miocene to the Late Pliocene and during recurrent marine regressions with land bridges formed in the course of the Pleistocene glaciations; Kerey & al., 2004; Magyari & al., 2008). These events may have facilitated bilateral gene flow in the past between areas now separated by the Aegean Sea, the Sea of Marmara, and the Dardanelles and Bosporus straits. Western Anatolia is indeed a biogeographically remarkable region, as it represents the area of meeting and gradual transition of

three phytogeographical regions (Ekim & Güner, 2000), and the mountain ranges here are relatively low and isolated compared to the other Anatolian regions (Parolly, 2020). The Uludağ is the highest mountain (Mount Uludağ, 2543 m) among the West Anatolian Exclaves, and the only one reaching the (sub-)alpine zone in the Marmara region (NW Anatolia). The other nearest mountain chains with peaks above 2,000 m a.s.l. are located towards southwestern Anatolia (Murat Dağı, or Bozdağ and Aydın Dağı of the western Taurus) and northern Anatolia (Köroğlu Dağları and other components of the Pontic mountains) (Parolly, 2020). In this context, the Uludağ range could be an important stepping-stone for the migration of (sub)alpine species between Anatolia and the Balkans. On the other hand, its isolated position could also promote genetic diversification of local populations, potentially leading to speciation. The latter scenario is supported by the fact that the Uludağ harbours high species diversity and has been identified as one of the centres of endemism in Türkiye (Noroozi & al., 2019). Even recently, new vascular plant taxa have been described that are restricted to this mountain or its close proximity (Yılmaz & al., 2003; Bağci & al., 2009; Daşkin & al., 2009; Özbek & al., 2011).

In our recent phylogenomic study, while focusing on two sister lineages of hygrophytic *Cardamine* (Brassicaceae) species in southwestern Asia, we revealed a divergent position of the population from the Uludağ, previously classified as either *C. uliginosa* or *C. tenera* (Kantor & al., 2023, see also Cullen, 1965). We proved that *C. tenera* is a lowland to montane species with a disjunct distribution restricted to the southern Caspian and northeastern Black Sea coast regions, but does not occur in Türkiye. *C. uliginosa* was confirmed as the widespread, predominantly (sub)alpine species growing in Anatolia and the Caucasus, extending into Lebanon and Iran. Within Anatolia, it grows abundantly in the Pontic Mts., less frequently and scattered in the Taurus Mts., but it probably does not reach westernmost Anatolia (Kantor & al., 2023). The Uludağ population, in contrast to its previous taxonomic classifications, showed affinity with *C. acris*, a highly polymorphic species comprising three subspecies growing in the montane to alpine belt of the Balkan Peninsula. However, the taxonomic status, geographic distribution, and circumscription of this uncertain entity, found so far in the Uludağ, remained unresolved. Molecular dating analyses suggested a recent divergence between the above-mentioned taxa, referred to as the *C. tenera* group, estimated to the late Pleistocene to early Pleistocene. Extensive intraspecific genetic variation was observed in particular within the widespread *C. uliginosa*, which was spatially structured and suggested both limited gene flow across large distances and postglacial recolonization from multiple glacial refugia. Two widespread intraspecific lineages were identified, a Caucasian reaching to Iran, and an Anatolian extending into Lebanon (Kantor & al., 2023). Chromosome counting revealed that both *C. uliginosa* and *C. tenera* are diploid (2n = 16; Kantor & al., 2023), whereas no data on ploidy level are available for the population from the Uludağ.

In the present study, we follow up on the previous findings described above, sample and examine *Cardamine* populations from both the western and northern Anatolian mountain regions in much greater detail. Using a combination of phylogenomic Hyb-Seq data, genome size estimates, and multivariate morphometrics, we aimed to: 1) circumscribe the northwestern Anatolian entity chorologically, morphologically, and phylogenetically, with respect to its relatives in Anatolia and the Balkans, which is described here as the new species, *Cardamine anatolica*; 2) discuss and highlight the evolutionary importance of western Anatolia and the role of biogeographic links between Anatolia and the Balkan Peninsula affecting species diversity on

both sides; 3) reveal fine-scale intraspecific genetic structure of *C. uliginosa* in the Pontic Mts. and shed light into the phylogeographic patterns within this mountain range.

MATERIALS AND METHODS

Plant sampling

We sampled all species of the target *Cardamine tenera* group occurring in Anatolia and adjacent areas (Caucasus, Balkans), focusing on western and northern Anatolia: *C. anatolica*, described here as a new species, *C. uliginosa* M.Bieb., *C. tenera* J.G.Gmel. ex C.A.Mey, *C. penzesii* Ančev & Marhold, and *C. acris* Griseb. A total of six populations of *C. anatolica* were found and sampled in northwestern Anatolia in the course of the present study. For Hyb-Seq analyses, we utilized 49 previously analyzed samples of these five species and supplemented them with 16 samples collected and analyzed for the present study, originating from western and northern Anatolia. Two Balkan members of the related *C. pratensis* group were also included, *C. matthioli* Moretti ex Comolli and *C. rivularis* Schur. *C. lazica* Boiss. & Balansa ex Boiss. was used as an outgroup, which is a species distributed in the regions adjacent to the eastern Black Sea coast and belonging to a different phylogenetic lineage (Kantor & al., 2023). Flow cytometric (FCM) and morphometric analyses focused on the comparison of *C. anatolica* with its closest relatives: *C. uliginosa*, *C. tenera* (excluded from FCM due to unsuccessful measurements of available material), *C. penzesii*, and *C. acris.* An overview of the material studied, including the number of populations and individuals analyzed, is given in Appendix 1, and the sample sites in Anatolia are shown in Fig. 1. Voucher specimens are deposited in herbaria SAV and HUB. To characterize the distribution of all three target species growing in Türkiye (*C. anatolica, C. uliginosa,* and *C. penzesii*), we also examined and revised herbarium specimens deposited in several major herbaria (see Table S1 for a complete list of revised specimens).

Hyb-Seq library preparation, data processing and analysis

The Hyb-Seq technique that combines target enrichment and genome skimming (Weitemier & al., 2014) was applied here to reconstruct phylogenetic relationships between the species, especially regarding the position of *C. anatolica*, and to study genetic structure of *C. uliginosa* in Anatolia. For target enrichment, we used *Cardamine*-specific probes that capture 2,246 exons from 1,235 genes, designed as described in detail in Melichárková & al. (2020). The Hyb-Seq libraries were prepared using the NEBNext®Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, MA, United States) as described in detail in Šlenker & al. (2021). The libraries were sequenced with 150-bp paired end reads on an Illumina MiSeq system at BIOCEV, Czechia. Illumina reads were processed using HybPiper v. 1.3 (Johnson & al., 2016) to assemble sequences of the targeted exons and flanking regions, which were aligned using MAFFT v. 7.450 (Katoh & Standley, 2013) and concatenated to genes by AMAS (Borowiec, 2016), as described in detail in Šlenker & al. (2021). HybPhaser (available online at: https://github.com/LarsNauheimer/HybPhaser) was used to identify highly variable exons (those with a SNP proportion > 0.02), which were omitted from downstream computations as potential paralogs. We generated a maximum-likelihood (ML) tree (RAxML-NG v.0.9.0, Kozlov & al., 2019) based on a concatenated alignment of all targeted genes, as well as a species tree using a summary multispecies coalescent approach (ASTRAL-III, Zhang & al., 2018) based on individual ML gene trees. Branch support of the concatenated ML tree was assessed by bootstrap replicates

and using the quartet sampling method (Pease & al., 2018, https://www.github.com/fephyfofum/quartetsampling).

In addition, to detect genetic structure within *C. uliginosa* in more detail, we extracted SNPs from Hyb-Seq reads mapped to the target exon sequences using the snipStrup pipeline [available online at: https://github.com/MarekSlenker/snipStrup; described in detail in Melichárková & al. (2020)] and analyzed them in STRUCTURE (Pritchard & al., 2000). To ensure that no linkage existed between sites, 500 datasets were produced by drawing a single random SNP site from each gene containing at least seven SNPs across the samples of *C. uliginosa*. Each dataset was run for each K = 1–10 (user-defined number of clusters), with a burn-in length of 100,000 generations and data collection for an additional 900,000 generations, setting the admixture model and correlated allele frequencies. The results of 500 datasets were averaged using the program CLUMPP (Jakobsson & Rosenberg, 2007) and drawn with distruct (Rosenberg, 2004). The approach of Evanno & al. (2005) was used to determine the optimal K value.

Chloroplast DNA sequences were extracted and assembled from Hyb-Seq reads using Fast-Plast v. 1.2.8 (available at: https://github.com/mrmckain/Fast-Plast) as described in detail in Šlenker & al. (2021). The recovered plastome sequences consisting of the large single copy (LSC), the small single copy (SSC), and one copy of the inverted repeats (IRb) were aligned in MAFFT v. 7.450 and a ML tree was inferred in RAxML-NG.

Flow cytometry

Nuclear DNA content of the studied species was measured by flow cytometry using the AT-selective DAPI fluorochrome (Doležel & al., 2007). The measured values were used to determine the ploidy level of the analyzed populations and samples, and to investigate and test the relative genome size differences (2C values) between the studied species. Measurements were performed using silica gel-dried leaf tissue following the protocol published by Marhold & al. (2010). Fluorescence intensity was measured using a Partec CyFlow ML flow cytometer (Partec GmbH Münster, Germany) with a UV LED lamp and the resulting histograms were analyzed in Partec FloMax v2.7 (Partec GmbH). Relative genome size (2C values in arbitrary units) was expressed as the ratio between the G0/G1 peaks of the measured sample and the internal standard, either *Solanum pseudocapsicum* (2C = 2.59 pg; Temsch & al., 2010) or *Bellis perennis* (2C = 3.38 pg; Schönswetter & al., 2007). The coefficients of variation (CV) were calculated for both standard and sample peaks, and only histograms with CV values below the 5% threshold were accepted for nuclear DNA content calculations. The significance of genome size differences between species was tested using the Kruskal-Wallis H test followed by Mann-Whitney pairwise comparisons with Bonferroni correction of P values using the stats package in R 4.0.0 (R Core Team, 2020). In addition, the chromosome number in *C. anatolica* was determined from mitotic metaphase plates in cells of root tips (protocol described in Marhold & al., 2002). The chromosome numbers of the other species analyzed here were already known from previous studies (summarized in Kantor & al., 2023).

Morphometric analyses

Methods of multivariate morphometric analyses were applied to reveal the extent of morphological differentiation of *C. anatolica* from its closest relatives and to identify the most important characters for its distinction. They were also used to extract diagnostic characters for all species studied and to help compile the identification key. A total of 22 vegetative and six floral

characters were measured or scored on herbarium specimens and floral parts attached to a sheet of paper with transparent adhesive tape. Size of floral parts was measured using ImageJ 1.53e software (Schneider & al., 2012). Fourteen ratio characters were also derived; in these cases, one of the original characters used for ratio calculation was removed from the final data matrix. The list of all characters is provided in Table 1. The final data matrix comprised 608 individuals and 32 characters. Data subsets were also assembled for specific partial analyses (see below). Prior to the analyses, Spearman correlation coefficients were calculated to detect pairs of highly correlated characters ($\rho > 0.9$) that might bias the results of discriminant analyses. We performed a series of canonical discriminant analyses (CDA) to gain deeper insights into the differentiation patterns; CDA1, which included all five species: *C. anatolica*, *C. uliginosa*, *C. tenera*, *C. penzesii*, and *C. acris*; CDA2, comprising *C. anatolica* and two species to which the populations from the Uludağ were previously assigned: *C. uliginosa* and *C. tenera*; CDA3, consisting of *C. anatolica* and the other two species, the geographically closest *C. penzesii* and the genetically closest *C. acris* from the Balkans; CDA4-CDA7, each comprising two species, *C. anatolica* and one of the four relatives. Analyses were performed using the MorphoTools2 package (Šlenker & al., 2022) in R 4.0.0 (R Core Team, 2020).

RESULTS

Phylogenomic analyses

Mapping of the Hyb-Seq reads in HybPiper recovered 1786 supercontig sequences (targeted nuclear exons and flanking regions) present in all 67 accessions. Of these, 65 exons were identified as potentially paralogous in HybPhaser and excluded, resulting in the final dataset of 1721 exons from 1104 genes used for further analyses. Gene alignments were from 93 to 6500 bp long (1029 bp on average), and the concatenated alignment of all nuclear genes was 1 136 046 bp long. The plastome alignment was 127 713 bp long. Sequencing data are available at the SRA database of NCBI (BioProjects PRJNA687126 and PRJNA830631).

The ML tree of the concatenated dataset of all nuclear genes and the species tree inferred in ASTRAL (Fig. 2A,B) yielded congruent topologies with three main clades. They corresponded to the *C. pratensis* group (represented here by *C. matthioli* and *C. rivularis*), the *C. tenera* group (*C. uliginosa* and *C. tenera* with *C. penzesii* in a sister position), and *C. acris* with *C. anatolica*. *C. anatolica* was resolved in a sister position to *C. acris*, and the clade comprising these two species received high support (LPP = 1, BS = 100%, QC = 1). The clade of *C. uliginosa* showed pronounced geographic structuring. All accessions of *C. uliginosa* from Anatolia formed a single clade (BS = 64%, QC = 0.2), which was divided into four geographically structured subclades (Fig. 1, 2A), designated as Anat A to Anat D. The westernmost subclade Anat A comprised also accessions from Lebanon, while the easternmost subclade Anat D included adjacent accessions from SW Georgia (the Adjara and Guria regions). The remaining accessions from the Caucasus, Zagros, and Alborz Mts. formed three subclades (two designated as Cauc-Alb-Zag I and II, one as Teberda, following Kantor & al., 2023).

Bayesian STRUCTURE analyses based on SNP variation of 957 genes (those which had at least seven SNPs) in *C. uliginosa* showed optimal genetic partitioning into two clusters. One of the clusters corresponded to the Anatolian subclades Anat A and Anat B (14 populations), while

the other cluster coincided with the Cauc-Alb-Zag I clade (11 populations), which is mainly distributed in the Lesser Caucasus, Zagros, and Alborz Mts. (Fig. 2C). Considerable genetic admixture was found in the populations of the remaining four subclades distributed in the contact zone of the two Bayesian clusters: Anat C (two populations), Anat D (five populations), Teberda (NW Caucasus, four populations), and Cauc-Alb-Zag II (two populations from the central Greater Caucasus, Fig. 2A,C). At K = 3 (Suppl. Fig. S1), the Teberda clade formed a separate third cluster, and accessions from the Anat D subclade showed genetic admixture from all three Bayesian clusters.

In the ML tree based on plastome sequences (Suppl. Fig. S2), *C. anatolica* was in a well-supported clade (BS = 100%) in a sister position to *C. acris*, in congruence with the nuclear trees. Only one accession of *C. acris* was in a distant position compared to the other conspecific samples, and one accession of *C. rivularis* was misplaced among the *C. acris* accessions. The intraspecific structure in the clade of *C. uliginosa* (which included nested *C. tenera*) showed geographic patterns that only partially matched those observed in the nuclear tree. All Anatolian accessions (except the SUR accession from the Anat A clade) formed two well-supported clades (both with BS = 100%) in a sister position (BS < 50%), designated as Anat I and Anat II, the latter comprising also the accessions from Lebanon. Two geographically adjacent Georgian samples (those from the Adjara and Guria regions) were resolved here among the Caucasian samples, in contrast to their position in the nuclear tree.

Flow cytometry and chromosome counting

Here, we found that *Cardamine anatolica* is diploid, as demonstrated by chromosome counting in population AKD (Akdağ, 2n = 16) and flow cytometric measurements of nuclear DNA content in all six populations of this species (64 individuals in total, see Appendix 1). In accordance with previous findings (Marhold & Ančev, 1999; Perný & al., 2004; Kantor & al., 2023), flow cytometric analyses confirmed the diploid level in all analyzed samples of *C. acris* (222 ind.), *C. penzesii* (36 ind.), and *C. uliginosa* (136 ind.). However, the studied species differed significantly in the relative genome size values (Kruskal-Wallis H test, $P < 0.001$; followed by Mann-Whitney pairwise comparisons with Bonferroni correction of P values, $P < 0.001$ for each pair of species), which ranged from 0.27 to 0.49 (Fig. 3, Table S2). The highest values were observed in *C. acris*, which overlapped only marginally with those measured in *C. anatolica*. The lowest relative genome size values were found in *C. uliginosa*.

Multivariate morphometrics

Spearman correlations revealed no highly correlated characters ($\rho > 0.9$), with the highest value of $\rho = 0.79$ found for the character pair WTS/LTS and WLS/LLS. Therefore, all 32 characters (see Table 1) were retained in canonical discriminant analyses (CDA). In CDA1 comprising all five target species, *C. anatolica* and *C. acris* were differentiated from *C. uliginosa* and *C. penzesii* along the first canonical axis, *C. uliginosa* and *C. penzesii* were in turn separated from each other along the second axis, and *C. tenera* was differentiated from all four species along the third axis (Fig. 4A). The characters that contribute most to species differentiation are highlighted in Table S3. CDA2 and CDA3, each based on a subset consisting of three species and examining the differentiation of *C. anatolica* in more detail, show that this species is well separated

in the space represented by the first two canonical axes, with no or minimal overlap (Fig. 4B,C). CDA2 demonstrates that *C. anatolica* can be morphologically distinguished from both *C. tenera* and *C. uliginosa*, to which populations of *C. anatolica* were historically assigned. CDA3 illustrates that *C. anatolica* is morphologically distinct from relatives that are closest either genetically (*C. acris*) or geographically (*C. penzesii*). A series of CDA4-CDA7, each comprising *C. anatolica* and one of the four relatives (*C. acris*, *C. penzesii*, *C. tenera*, *C. uliginosa*; Suppl. Fig. 3A–D, Table S3), showed congruent results and, together with CDA2 and CDA3, identified the diagnostic morphological characters for the target species (Table 2). In summary, the most significant characters that can be used to define *C. anatolica* and distinguish it from the relatives include: 1) exclusively white petals; 2) basal rosette with short leaves, the longest of which is 2.8–9.9 cm long; and 3) shape of lateral leaflets of the middle stem leaf being oblong to broadly elliptical, much narrower than the terminal leaflets (the lateral leaflet width reaching 18–43% of the terminal leaflet width).

Geographic distribution

Herbarium and field research revealed the existence of nine sites that unequivocally belong to *C. anatolica* (Appendix 1, 2). They are all restricted to a relatively small area in northwestern Türkiye, extending from Kütahya province in the south to Yalova and Kocaeli provinces in the north (Fig. 1). The present data suggest that the species grows allopatrically from both *C. penzesii* and *C. uliginosa*. *C. penzesii*, a lowland species typically growing in floodplain forests along the Black Sea coast, occurs mainly near and west of the Bosporus Strait, with only a few records in Kocaeli province, where it is found very close to the northern range margin of *C. anatolica*. The distribution of the (sub)alpine *C. uliginosa* in Türkiye is concentrated in the Pontic Mts. with scattered occurrences in the East Anatolian highlands and the Taurus Mts. The westernmost record is from Bolu province (Fig. 1, Table S1).

DISCUSSION

Cardamine anatolica: a new species recognized in northwestern Anatolia

The distinct position of the northwestern Anatolian entity, previously classified as either *C. uliginosa* or *C. tenera* and recorded from the Uludağ, was first noted by Šlenker & al. (2021) and then examined by Kantor & al. (2023) in a broader context of related species from Southwestern Asia. Nevertheless, until the present study, only a single population had been found and analyzed and the entity was provisionally designated as *C.* cf. *uliginosa*. Based on more detailed herbarium and field research in the present study, we have revealed the existence of nine localities in northwestern Anatolia (Fig. 1, Appendix 1, 2) and sampled populations from six of them, which can be assigned to this entity and are described here as the new species *C. anatolica* (Fig. 5, Appendix 1). Both phylogenomic and morphometric analyses proved that *C. anatolica* is distinct from *C. uliginosa* and *C. tenera*, as well as from the geographically closest species, *C. penzesii*. Phylogenetic trees based on nuclear genes and plastome data were congruent and placed *C. anatolica* in a separate and well-supported clade, in a sister position to the Balkan *C. acris*. The relative genome size of *C. anatolica* was significantly different from all other relatives analyzed (*C. acris*, *C. penzesii*, and *C. uliginosa*), and in agreement with phylogenetic relationships, the genome size values of *C. anatolica* were closest to and overlapped slightly with those of *C. acris*.

All these species, also referred to as the *C. tenera* group, are morphologically variable, which is particularly true of the widely distributed *C. uliginosa* (see Kantor & al., 2023). They are phenotypically similar to each other, so that a combination of several characters and examination of multiple specimens per population is usually needed to reliably determine them. Using a series of canonical discriminant analyses, we have shown here that *C. anatolica* can be morphologically distinguished from its relatives, including the sister species *C. acris*. In the taxonomic treatment below, we highlight the diagnostic characters and present an identification key to the studied species in Türkiye and adjacent areas.

Based on the current herbarium and field survey, the distribution of *C. anatolica* is confined to wetland habitats of montane to (sub)alpine regions (725–2060 m a.s.l.) in northwestern Türkiye. Its distribution range appears to be restricted to the West Anatolian Exclaves and their close vicinity, rarely descending to altitudes below 1000 m. Nevertheless, we expect that some more localities of *C. anatolica* than those listed here (Table S1) may be found in the future, as this taxon has been overlooked in the past due to its misinterpretation with *C. tenera* or *C. uliginosa* and the taxonomic confusion that surrounded these species until recently (see Marhold & Ančev, 1999; Kantor & al., 2023 for details). Moreover, western Anatolia appears to be still largely unexplored in terms of taxonomic research, as evidenced by the recent descriptions of several new taxa in this area (e.g., Yüzbaşıoğlu & al., 2015; Yüzbaşıoğlu, 2017; Hamzaoğlu & al., 2022; Yildirim & al., 2022).

Biogeographic links between Anatolia and the Balkan Peninsula

Both the Balkan Peninsula and Anatolia are significant centers of biodiversity and endemism. Their largely mountainous landscapes, complex topography, and remarkable tectonic history favored the accumulation of diversity as well as speciation processes (Médail & Diadema, 2009; Bilgin, 2011; Nieto Feliner, 2014; Parolly, 2020; Španiel & Rešetnik, 2022). The Anatolian and Balkan peninsulas are now separated by the Aegean Sea, the Sea of Marmara, the Black Sea, and the straits of the Dardanelles and the Bosporus, but the land-sea configuration was highly variable in the past. Land bridges existed from the Messinian salinity crisis in the late Miocene to the Plio-Pleistocene transition and in the late Pleistocene (Elmas, 2003; Chobanov & al., 2017; Kaya & Çıplak, 2017). As a result, biogeographic and evolutionary links between Anatolia and the Balkans have been repeatedly observed (e.g., Ansell & al., 2011; Bilgin, 2011; Kaya & Çıplak, 2017; Özüdoğru & Mummenhoff, 2020; Carnicero & al., 2021), which are also evidenced here by the phylogenetic lineage of *C. acris* and *C. anatolica*. Divergence time estimates suggested that this lineage originated in the late Pliocene (Kantor & al., 2023), and we assume that it had a broader range encompassing both the Balkans and western Anatolia. The restricted gene flow between populations in the Balkans and western Anatolia, caused by the barriers of the intervening seas and the Thracian lowland area, probably led to genetic differentiation and allopatric speciation, dated to the early Pleistocene (Kantor & al., 2023). The existence of this western Anatolian-Balkan phylogenetic lineage, as well as the vicariance of *C. anatolica* and *C. acris,* support phylogeographic scenarios and suture zones inferred in Türkiye from concordant genetic patterns within several taxa (Bilgin, 2011). A similar pattern of Anatolian-Balkan divergence has also been observed in a variety of organisms, such as snakes (Ursenbacher & al., 2008), newts (Wielstra & Arntzen, 2011), trees (*Quercus cerris*, Bagnoli & al., 2016; *Juniperus drupacea*, Sobierajska & al., 2016), and ferns (*Cryptogramma crispa*, Metzgar & al., 2016).

While *C. acris* is widespread in the (sub)alpine belt and grows in wet habitats in several mountain ranges throughout the Balkan Peninsula (Perný & al., 2004; Kantor & al., unpubl.), *C. anatolica* is restricted to a few sites of the West Anatolian Exclaves probably due to the scarcity of suitable high alpine habitats for hygrophilous plants in western Anatolia (Parolly, 2020). The mountains adjacent to the Sea of Marmara and the northern Aegean Sea mostly do not reach more than 1500 m and are characterized by a Mediterranean climate (Parolly, 2020). Most records and sites of *C. anatolica* in fact refer to the Uludağ, which may have played an important role in the persistence of this species, especially during the warmer interglacial periods. Indeed, the Uludağ is a hotspot of diversity in western Anatolia (Noroozi & al., 2019), hosting 1309 vascular plant taxa (Daşkın & Kaynak, 2011) with a high proportion of species that are either endemic to this mountain range or have a limited distribution that extends to adjacent mountains reaching lower altitudes (Yılmaz & al., 2003; Bağci & al., 2009; Özbek & al., 2011; Yüzbaşıoğlu & al., 2015). One of these endemics, *Berteroa physocarpa*, occurs in the Uludağ (at 2200 m) and Kartepe Mts. (1600 m) in Kocaeli province (Yüzbaşıoğlu & al., 2017), which coincides with the occurrence of the northernmost sites of *C. anatolica*.

Two dispersal corridors connecting the Caucasus and the Balkans have been proposed for the colonization of mountainous species in Anatolia, referred to as the Taurus way and the North Anatolian route, which meet in the Marmara region (Bilgin, 2011; Kaya & Çıplak, 2017; Özüdoğru & Mummenhoff, 2020). The distribution range of *C. anatolica* apparently does not extend southwards to the western Taurus; in general, the occurrence of hygrophytic *Cardamine* in the Eastern and Central Taurus in southern Türkiye is very rare (*C. uliginosa*, *C. wiedemanniana* Boiss.), with no records known at all from the Western Taurus (Table S1, Kantor & al., 2023). On the other hand, the distribution of *C. uliginosa* in Anatolia largely follows the North Anatolian colonization route, but this species does not reach the Marmara region (the westernmost records are from Bolu province, Fig. 1), and there are no records to suggest an overlap between the ranges of ecologically similar *C. anatolica* and *C. uliginosa*. Geographically, the closest to *C. anatolica* is *C. penzesii*, which occurs near the southwestern Black Sea coast (their closest populations are only 20 km apart, Fig. 1), but it is a lowland species growing mainly in floodplain forests at 0-240 m a.s.l. (Marhold & Ančev, 1999; Kantor & al., 2023).

Accumulation of genetic diversity and admixture in northeastern Anatolia

The Pontic Mts. in northeastern Anatolia (part of the Colchis region) are a significant hotspot of plant diversity and endemism. They intersect two biogeographical regions and have served as both a refugium and a migration corridor between Anatolian mountain ranges and the Caucasus. Their high altitudinal amplitude, ranging from the sea level to the high alpine zone with numerous peaks above 3000 m, heterogeneous topography, and diverse climate with high rainfall on the northern, sea-facing slopes in contrast to low precipitation on the southern ones, result in a large habitat diversity and support high species diversity (Noroozi & al., 2019; Parolly, 2020). Pleistocene glaciers formed in this area (Sarikaya & Çiner, 2015) affected mainly alpine species that constitute a large proportion of its present-day flora. Consequently, high genetic diversity and complex patterns are expected for mountainous flora of northeastern Anatolia and indeed they are evidenced here by *C. uliginosa*. This is the most widespread, genetically and morphologically complex species among the *Cardamine* hygrophytes in Southwestern Asia. The centers of its distribution and genetic diversity are in northern Anatolia and the Caucasus (Kantor & al., 2023 and present results), which provide favourable ecological conditions with numerous peaks reaching into the (sub)alpine zone and high humidity from the adjacent Black and Caspian Seas

even during climatic oscillations (Parolly, 2020). We found that its genetic variation in northeastern Anatolia is highly structured geographically, with four adjacent but allopatric clades (Anat A – Anat D). The easternmost clades (Anat C and D), extending into neighbouring regions of Georgia also showed significant admixture between the major Anatolian and Caucasian gene pools (Fig. 2, Suppl. Fig. S1). We hypothesize that the observed geographic patterns, accumulation of genetic diversity, and genetic admixture reflect survival of the species in multiple glacial refugia (consistent with the refugia-within-refugia concept, Gómez & Lunt, 2007; Nieto Feliner, 2014), partial genetic differentiation due to restricted gene flow, and small-scale range shifts during glacial-interglacial periods that brought different gene pools into contact. Similarly, concentration of genetic diversity with strong geographic structuring within northeastern Anatolia has also been observed in *C. lazica*, a hygrophytic species with the niches shifted towards higher precipitation and lower precipitation seasonality than *C. uliginosa* (Kantor & al., 2023). The sparse records of *C. uliginosa* in the Eastern Taurus Mts. (extending as far south as to the Mt. Lebanon range) suggest that southward migration via Anatolian Diagonal did occur but probably remained limited due to the prevailing dry continental climate and scarcity of suitable wetland habitats (Parolly, 2020).

Conclusion

The close phylogenetic relationship of *C. anatolica*, an endemic of northwestern Anatolia described here, with the Balkan species *C. acris* argues for the existence of biogeographic links between these two areas. In contrast to most published studies that highlight Anatolia as an important refugium and center of diversification (e.g., Font & al., 2009; Özüdoğru & al., 2015; Özüdoğru & Mummenhoff, 2020) and describe migration routes predominantly from Anatolia to Europe ("out of Anatolia scenarios", e.g., Rokas & al., 2003; Ansell & al., 2011), however, the present data for *Cardamine* favour a vicariance scenario. More thorough studies of organisms inhabiting the Thrace, the Marmara region, and adjacent areas on both the Anatolian and Balkan sides are needed for a better understanding of phylogeographic patterns and evolutionary processes in this important transitional zone between Europe and Asia. To obtain a complete picture of the biogeographic history of the region, research should also focus on species with specific ecological requirements, as in the case of hygrophytic *Cardamine* perennials, which have limited occurrence in the Mediterranean and Irano-Turanian phytogeographical regions characterized by dry climate and dominance of xerophilous vegetation. Furthermore, the present results emphasize that the Pontic Mts. in northeastern Anatolia are a diversity hotspot, not only at the level of species diversity and endemism (Noroozi & al., 2019, 2022; Parolly, 2020), but also at the level of genetic diversity (e.g., Veith & al., 2003), which, however, remains largely unexplored. We illustrate here that this area can support long-term species persistence and diversity accumulation in situ, and also serves as a meeting point for lineages spreading from adjacent areas.

TAXONOMIC TREATMENT

Identification key to the species of the *Cardamine tenera* group in Türkiye and adjacent areas

Note. Due to the high intraspecific and intrapopulation variation of several morphological characters in the species treated here, it is recommended to examine multiple specimens per population for reliable species determination. For definitions of the terms 'stem length', 'middle stem leaf', and 'one side of the leaflet margin', see the notes below Table 1. In the identification

key, the values of the quantitative characters correspond to the 5th–95th percentiles; in the species description below, they are expressed by the (1st–)5th–95th(–99th) percentiles.

1. Petals white, 5.6–8.3 mm wide; stems with numerous (2–8) branches (median 4); leaflets on stem leaves usually long and relatively narrow: terminal leaflet of the middle stem leaf narrowly to broadly elliptical, 1.5–5.5 times longer than wide; basal leaves with 4–7 pairs of lateral leaflets; additional leaves sometimes arise from adventitious buds on rosette leaves, cauline leaf axils and in inflorescences; rhizomes usually with several globular tubers; growing in lowland flood-plain forests (in altitudes 0–240 m) ..... ***C. penzesii***

1. Petals white, pink or violet, 3.5–7.9 mm wide; stems unbranched or seldom with up to 8 branches (median 0); leaflets on stem leaves variable in shape, wider than long, as wide as long, or longer than wide, but in that case not as conspicuously as in 1.a: terminal leaflet of the middle stem leaf oblong to broadly circular, slightly wider than long to up to 3 times longer than wide; basal leaves with 2–7 pairs of lateral leaflets; not forming adventitious leaves arising from buds on basal rosettes, cauline leaf axils or in inflorescence; rhizomes without tubers; mostly growing in montane regions, from foothills to the alpine zone (in altitudes 30–2800 m).....2.

2. The uppermost stem leaves 1.8–7.4 cm long, reaching 9–34% of the stem length; middle stem leaves 4.4–10.4 cm long, reaching 18–44% of stem length, with the terminal leaflet 1.8–4.7 cm long and the uppermost lateral leaflets 0.8–2.5 cm long; growing in foothill regions (altitudes 30–850 m) ..... ***C. tenera***

2. The uppermost stem leaves 0.7–3.9 cm long, reaching 2–18% of the stem length; middle stem leaves 1.3–10.8 cm long, reaching 6–29% of the stem length, with the terminal leaflet 0.7–3.8 cm long and the uppermost lateral leaflets 0.3–2.4 cm long; growing in higher altitudes (400–2800 m, mostly above 1000 m).....3.

3. Stems unbranched or rarely with maximum of 2 branches (median 0); middle stem leaves with 2–7 pairs of lateral leaflets; terminal leaflet of the middle stem leaf narrowly elliptical to circular (0.2–1.9 cm wide, the width reaching 20–100% of its length), with 0–5 teeth or lobes on one side; terminal leaflets usually not conspicuously enlarged compared to the lateral ones: terminal leaflets of the basal and middle stem leaves are of the same size or up to two times longer and three times wider than the uppermost lateral leaflets..... ***C. uliginosa***

3. Stems unbranched or with up to 8 branches (median 1); middle stem leaves with 1–4 pairs of lateral leaflets; terminal leaflet of the middle stem leaf broadly elliptical to broadly circular (0.6–3.5 cm wide, the width reaching 55–120% of its length), with 2–7 teeth or lobes on one side; terminal leaflets usually conspicuously enlarged compared to the lateral ones: terminal leaflets of the middle stem leaves are 1.5–3 times longer and 2–5.5 times wider than the uppermost lateral leaflets; terminal leaflets of the basal leaves are 1.5–4 times longer and 1.5–4 times wider than the uppermost lateral leaflets.....4.

4. Petals white, pink, violet or white with a pink tone, 7.4–11.8 mm long; basal leaves up to 3.5–17.6 cm long; uppermost lateral leaflets of the middle stem leaves elliptical to circular**,** as wide as long or up to 2.5 times longer than wide (its width reaching 22–50% of the width of the terminal leaflet)..... ***C. acris***

4. Petals white, 9.3–13.5 mm long; basal leaves up to 2.8–9.9 cm long; uppermost lateral leaflets

of the middle stem leaves oblong to broadly elliptical, usually conspicuously narrow, always longer than wide, up to 3.5 times longer than wide (its width reaching 18–43% of the width of the terminal leaflet)..... ***C. anatolica***

Cardamine anatolica Jar.Kučera, Kantor, Dönmez, Yüzb., Marhold, Slovák, Šlenker, Lihová, **sp. nov. –** Holotype: Türkiye. Bursa, Osmangazi distr., Uludağ, ski resort, 40.1026989° N, 29.1535106° E, 2059 m, 15 Jun 2010, *J. Kučera, M. Slovák & A. Bérešová* (SAV barcode SAV0017449 [https://ibot.sav.sk/herbarium/object/SAV0017449]); Isotypes deposited in HUB and ISTE.

Diagnosis: Morphologically similar to *C. acris*, *C. penzesii*, *C. tenera* and *C. uliginosa* but differs by very narrow lateral leaflets on middle stem leaves compared to much wider terminal leaflets, and by mostly having shorter basal leaves. From *C. acris*, *C. tenera* and *C. uliginosa* it differs also by having exclusively white and longer petals. Compared to *C. tenera* it has shorter middle and uppermost stem leaves.

Description: Perennial herb, rhizome erect or rarely ascending, stem erect or rarely ascending only at base, (16–)19–41(–45) cm tall up to the base of the uppermost stem leaf, (1.1–)1.4–4.1(–5.2) mm wide at base, simple or with up to 7(–9) inflorescence-bearing branches, glabrous and pruinose. Basal leaves forming a rosette, youngest rosette leaves hairy, later becoming glabrous on the blade, with persisting trichomes on margin or completely glabrous, pinnate, (2.3–)2.8–9.9(–12.3) cm long, with (2–)3–6 pairs of lateral leaflets, their terminal leaflet conspicuously larger than the lateral ones, almost circular to transversely broadly elliptical, (0.7–)0.8–2.6(–3.0) cm long and (0.6–)0.9–2.7(–3.1) cm wide, (1.2–)1.4–3.8(–4.1) times longer and (1.2–)1.7–4.1(–4.7) times wider than the uppermost lateral leaflets; lateral leaflets sessile, broadly elliptical to broadly circular, uppermost lateral leaflets (0.3–)0.4–1.5 cm long, (0.3–)0.4–1.3(–1.6) cm wide. Stem leaves 5–9(–12), usually distributed along the entire stem length with more leaves in the upper half of the stem, pinnate or pinnatisect, similar to the rosette leaves, glabrous or hairy on the margins; middle stem leaves (1.2–)1.3–8.9(–10.1) cm long, with 1–4(–5) pairs of sessile lateral leaflets or segments; terminal leaflet elliptic to broadly circular, with 2–7 teeth or lobes on one side of the leaflet margin, (0.6–)0.7–3.3(–3.6) cm long and (0.5–)0.6–2.6(–3.2) cm wide, (1.3–)1.4–2.9(–3.7) times longer and 2.3–5.4(–6.3) times wider than the uppermost lateral leaflets; lateral leaflets usually conspicuously narrow, oblong to broadly elliptical, uppermost lateral leaflets 0.3–1.7(–1.9) cm long and (0.1–)0.2–1.1(–1.2) cm wide. Uppermost leaves (0.5–)0.7–3.9(–4.8) cm long, simple or tripartite, rarely pinnate with two pairs of lateral leaflets, margin of terminal leaflet entire or with up to 3(–4) teeth or lobes on one side. Inflorescence racemose, peduncles glabrous. Sepals lanceolate to ovate, (3.3–)3.4–4.9(–5.3) mm long. Petals white, narrowly obovate, (9.1–)9.3–13.5(–14.4) mm long and (4.0–)4.3–6.5(–7.1) mm wide, with claw forming 1/3–1/5 of the petal length, apex retuse, truncate or rounded. Stamens 6, tetradynamous, shorter filaments (2.6–)2.7–4.8(–5.1) mm long, longer filaments (4.4–)5.0–7.2(–7.7) mm long; anthers yellow. Stigma conspicuous, enlarged. Siliques (1.5–)1.9–3.2(–3.3) cm long (excluding peduncle and style) and up to 2 mm wide, linear. Seeds brown, broadly elliptic to almost circular.

Chromosome number: 2n = 2x = 16

Distribution and ecology: The currently known distribution range of *Cardamine anatolica* is limited to northwestern Türkiye, specifically to the provinces of Bursa, Kocaeli, Kütahya and Yalova (parts of the Marmara and Aegean regions, Fig. 1; Appendix 1, 2). Most records are from the Uludağ, but occurrences have also been documented in the wider surroundings of Mt. Kartepe (Samanlı Mts.). The species grows in the montane to (sub)alpine belts (725–2060 m) and is ecologically confined to wet sites near streams, springs or in wet meadows (Fig. 5E,F).

AUTHOR CONTRIBUTIONS

JZL designed the research with the contribution of AK, MŠ, JK, and KM; all authors collected samples; AK, JK, and JZL generated data; AK, JK, AAD, and SY extracted distribution data; AK and MŠ performed data analyses; AK and JZL wrote the manuscript with the contribution of MŠ, JK, and KM; all authors contributed to data interpretation, commented and approved the manuscript.

ACKNOWLEDGEMENTS

We thank Lenka Martonfiová (Košice, Slovakia) for counting chromosomes of *C. anatolica*; and Zübeyde Uğurlu Aydın (Ankara, Türkiye) for help with georeferencing the examined specimens of *Cardamine* from the herbaria ANK, EGE, HUB, ISTE. We are also thankful to curators of the cited herbaria who provided us access to their collections; and to all botanists that helped us with sample collection or accompanied us in the field: Tatiana Kačmárová, Martin Kolník, Marián Perný (all formerly Bratislava, Slovakia), Anna Bérešová, Viktor Kučera, Katarína Skokanová, Marek Slovák, Milan Valachovič (all Bratislava, Slovakia), Paolo Bartolić, Filip Kolář, Nélida Padilla-García (all Prague, Czechia), Minčo Ančev, Rossen Vassilev, Elena Georgieva (all Sofia, Bulgaria), Boris S. Tuniev (Sochi, Russian Federation), Joêlle Breidy, Hiba Dokmak (both Zahle, Lebanon), Erich Hübl and Gerald M. Schneeweiss (both Vienna, Austria).

The work was supported by the Turkish-Slovak joint research project JRP [SAS – TUBITAK no. 475542 (ELTtoEFDi)] and the Slovak Research and Development Agency [APVV-21-0044]. Computational resources were provided by the e-INFRA CZ project (ID:90140), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

LITERATURE CITED

Ansell, S.W., Stenøien, H.K., Grundmann, M., Russell, S.J., Koch, M.A., Schneider, H. & Vogel, J.C. 2011. *Ann. Bot.* 108: 241–252. https://doi.org/10.1093/aob/mcr134

Bağci, Y., Savran, A. & Başköse, I. 2009. *Ornithogalum nurdaniae* (Liliaceae), a New Species from North-West Anatolia, Turkey. *Turk. J. Bot*. 33: 163–167. https://doi.org/10.3906/bot-0811-7

Bagnoli, F., Tsuda, Y., Fineschi, S., Bruschi, P., Magri, D., Zhelev, P., Paule, L., Simeone, M. C., González-Martínez, S.C. & Vendramin, G.G. 2016. Combining molecular and fossil data to infer demographic history of *Quercus cerris*: insights on European eastern glacial refugia. *J. Biogeogr.* 43: 679–690. https://doi.org/10.1111/jbi.12673

Barker, B.S., Andonian, K., Swope, S.M., Luster, D.G. & Dlugosch, K.M. 2017. Population genomic analyses reveal a history of range expansion and trait evolution across the native and invaded range of yellow starthistle (*Centaurea solstitialis*). *Mol. Ecol.* 26: 1131–1147. https://doi.org/10.1111/mec.13998

Bilgin, R. 2011. Back to the suture: the distribution of intraspecific genetic diversity in and around Anatolia. *Int. J. Mol. Sci.* 12: 4080–4103. https://doi.org/10.3390/ijms12064080

Borowiec, M.L. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660. https://doi.org/10.7717/peerj.1660

Carnicero, P., Garcia-Jacas, N., Sáez, L., Constantinidis, T., & Galbani-Casals, M. 2021. Disentangling relationships among eastern Mediterranean *Cymbalaria* including description of a novel species from the southern Peloponnese (Greece). *Plant Syst Evol* 307:13. https://doi.org/10.1007/s00606-020-01730-3

Chobanov, D.P., Kaya, S., Grzywacz, B., Warchałowska-Śliwa, E. & Çıplak, B. 2017. The Anatolio-Balkan phylogeographic fault: a snapshot from the genus *Isophya* (Orthoptera, Tettigoniidae). *Zool. Scr.* 46: 165–179. https://doi.org/10.1111/zsc.12194

Cullen, J. 1965. *Cardamine* L. Pp. 438–444 in: Davis, P.H., Cullen, J. & Coode, M.J.E. (eds.), *Flora of Turkey and the East Aegean islands, Volume 1*. Edinburgh: University Press.

Daşkın, R. & Kaynak, G. 2011. Conservation status of five endemic species distributed in Northwest Turkey. *Phytol. Balc.* 17: 213–219.

Daşkin, R., Yılmaz, Ö. & Kaynak, G. 2009. *Prometheum chrysanthum* subsp. *uludaghense* (Crassulaceae), a new taxon from Turkey. *Ann. Bot. Fenn*. 46: 456–458.

Doležel, J., Greilhuber, J. & Suda, J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* 2: 2233–2244. https://doi.org/10.1038/nprot.2007.310

Dönmez, A. & Yerli, S.V. 2018. Biodiversity in Turkey. Pp. 397–442 in: Pullaiah, T. (ed.), *Global Biodiversity 2, Selected countries in Europe*. Oakville: Apple Academic Press.

Ekim, T. & Güner, A. 2000. Introduction: The floristic richness of Turkey. *Curtis's Bot. Mag.* 17: 48–59. https://doi.org/10.1111/1467-8748.00247

Elmas, A. 2003. Late Cenozoic tectonics and stratigraphy of northwestern Anatolia: the effects of the North Anatolian Fault of the region. *Int. J. Earth Sci.* 92: 380–396. https://doi.org/10.1007/s00531-003-0322-2

Evanno, G., Regnaut, S. & Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 26112620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Font, M., Garcia-Jacas, N., Vilatersana, R., Roquet, C. & Susanna, A. 2009. Evolution and biogeography of *Centaurea* section *Acrocentron* inferred from nuclear and plastid DNA sequence analyses. *Ann. Bot.* 103: 985–997. https://doi.org/10.1093/aob/mcp022

Gómez, A. & Lunt, D.H. 2007. Refugia within Refugia: Patterns of Phylogeographic Concordance in the Iberian Peninsula. Pp. 155–158 in: Weiss, J. & Ferrand, N. (eds.), *Phylogeography of Southern European Refugia*. Dordrecht: Springer.

Gür, H. 2016. The Anatolian diagonal revisited: Testing the ecological basis of a biogeographic boundary. *Zool. Middle East* 62: 189–199. https://doi.org/10.1080/09397140.2016.1226544

Hamzaoğlu, E., Kanoğlu, S.S. & Aksoy, N. 2022. *Dianthus nezahatiae* (Caryophyllaceae), a new species from Northwest Turkey. *Acta Bot. Croat.* 81: 206–212. https://doi.org/10.37427/botcro-2022-018

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

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J.C. & Wickett, N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* 4: 1600016. https://doi.org/10.3732/apps.1600016

Kantor, A., Kučera, J., Šlenker, M., Breidy, J., Dönmez, A.A., Marhold, K., Slovák, M., Svitok, M. & Zozomová-Lihová, J. In press. Evolution of hygrophytic plant species in the Anatolia-Caucasus region: Insights from phylogenomic analyses of *Cardamine* perennials. *Ann. Bot.* https://doi.org/10.1093/aob/mcad008

Katoh, K. & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30: 772–780. https://doi.org/10.1093/molbev/mst010

Kaya, S. & Çıplak, B. 2017. Phylogeography and taxonomy of the *Psorodonotus caucasicus* (Orthoptera, Tettigoniidae) group: independent double invasion of the Balkans from the Caucasus. *Syst. Entomol.* 42: 118–133. https://doi.org/10.1111/syen.12197

Kerey, I.E., Meriç, E., Tunoğlu, C., Kelling, G., Brenner, R.L. & Doğan, A.U. 2004. Black Sea–Marmara Sea Quaternary connections: new data from the Bosphorus, İstanbul, Turkey. *Palaeogeogr. Palaeocl.* 204: 277–295. https://doi.org/10.1016/S0031-0182(03)00731-4

Körner, C. 2004. Mountain biodiversity, its causes and function. *Ambio* 33: 11–17. https://doi.org/10.1007/0044-7447-33.sp13.11

Kosswig, C. 1955. Zoogeography of the Near East. *Syst. Zool.* 4: 49–73. https://doi.org/10.2307/2411949

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35: 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Kürschner, H. & Parolly, G. 2012. The Central Anatolian Steppe. Pp. 149–171 in: Werger, M.J.A. & van Staalduinen, M.A. (eds.), *Eurasian Steppes. Ecological Problems and Livelihoods in a Changing World*. Dordrecht, Heidelberg, New York, London: Springer. https://doi.org/10.1007/978-94-007-3886-7_4

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. *J. Biogeogr.* 35: 865–883. https://doi.org/10.1111/j.1365-2699.2007.01849.x

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. *Biol. Rev.* 92: 1365–1388. https://doi.org/10.1111/brv.12287

Marhold, K. & Ančev, M.E. 1999. *Cardamine penzesii*, a rediscovered taxon of the *C. pratensis* group (Cruciferae). *Ann. Bot. Fenn*. 36: 171–180.

Marhold, K., Lihová, J., Perný, M., Grupe, R. & Neuffer, B. 2002. Natural hybridization in *Cardamine* (Brassicaceae) in the Pyrenees: evidence from morphological and molecular data. *Bot. J. Linn. Soc.* 139: 275–294. https://doi.org/10.1046/j.1095-8339.2002.00066.x

Marhold, K., Kudoh, H., Pak, J.H., Watanabe, K., Španiel, S. & Lihová, J. 2010. Cytotype diversity and genome size variation in eastern Asian polyploid *Cardamine* (Brassicaceae) species. *Ann. Bot.* 105: 249–264. https://doi.org/10.1093/aob/mcp282

Médail, F. & Diadema, K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *J. Biogeogr.* 36: 1333–1345. https://doi.org/10.1111/j.1365-2699.2008.02051.x

Melichárková, A., Šlenker, M., Zozomová-Lihová, J., Skokanová, K., Šingliarová, B., Kačmárová, T., Caboňová, M., Kempa, M., Šrámková, G., Mandáková, T., Lysák, M. A., Svitok, M., Mártonfiová, L. & Marhold, K. 2020. So closely related and yet so different: Strong contrasts between the evolutionary histories of species of the *Cardamine pratensis* polyploid complex in Central Europe. *Front. Plant Sci.* 11: 588856. https://doi.org/10.3389/fpls.2020.588856

Metzgar, J., Stamey, M. & Ickert-Bond, S. 2016. Genetic differentiation and polyploid formation within the *Cryptogramma crispa* complex (Polypodiales: Pteridaceae). *Turk. J. Bot*. 40: 231–240. https://doi.org/10.3906/bot-1501-54

Mittermeier, R.A., Turner, W.R., Larsen, F.W., Brooks, T.M. & Gascon, C. 2011. Global biodiversity conservation: the critical role of hotspots. Pp. 3–22 in: Zachos, F. & Habel, J. (eds.), *Biodiversity hotspots*. Berlin, Heidelberg: Springer.

Nieto Feliner, G. 2014. Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspect. Plant Ecol.* 16: 265–278. https://doi.org/10.1016/j.ppees.2014.07.002

Noroozi, J., Zare, G., Sherafati, M., Mahmoodi, M., Moser, D., Asgarpour, Z. & Schneeweiss, G.M. 2019. Patterns of endemism in Turkey, the meeting point of three global biodiversity hotspots, based on three diverse families of vascular plants. *Front. Ecol. Evol.* 7: 159. https://doi.org/10.3389/fevo.2019.00159

Noroozi, J., Minaei, M., Khalvati, S., Kaveh, A., Nafisi, H., Nazari, B., Zare, G., Vitek, E., Hülber, K. & Schneeweiss, G.M. 2022. Hotspots of (sub)alpine plants in the Irano-Anatolian global biodiversity hotspot are insufficiently protected. Divers. Distrib. 29: 244–253. https://doi.org/10.1111/ddi.13656

Özbek, M.U., Vural, M. & Daşkin, R. 2011. A new species of the genus *Cota* (Asteraceae) from Uludağ, Turkey. *Turk. J. Bot.* 35: 331–336. https://doi.org/10.3906/bot-1002-27

Özüdoğru, B. & Mummenhoff, K. 2020. Phylogenetic and biogeographical history confirm the Anatolian origin of *Bornmuellera* (Brassicaceae) and clade divergence between Anatolia and the Balkans in the Plio-Pleistocene transition. *Turk. J. Bot.* 44: 593–603. https://doi.org/10.3906/bot-2007-42

Özüdoğru, B., Akaydın, G., Erik, S., Al-Shehbaz, I.A. & Mummenhoff, K. 2015. Phylogeny, diversification and biogeographic implications of the eastern Mediterranean endemic genus *Ricotia* (Brassicaceae). *Taxon* 64: 727–740. https://doi.org/10.12705/644.5

Parolly, G. 2020. The Anatolian High-Mountain Ranges – Plant Diversity Between Two Seas. Pp. 215–286 in: Noroozi, J. (ed.), *Plant Biogeography and Vegetation of High Mountains of Central and South-West Asia*. Cham: Springer Nature. https://doi.org/10.1007/978-3-030-45212-4_7

Pease, J.B., Brown, J.W., Walker, J.F., Hinchliff, C.E. & Smith, S.A. 2018. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am. J. Bot.* 105: 385–403. https://doi.org/10.1002/ajb2.1016

Perný, M., Tribsch, A. & Anchev, M.E. 2004. Infraspecific differentiation in the Balkan diploid *Cardamine acris* (Brassicaceae): molecular and morphological evidence. *Folia Geobot.* 39: 405–429. https://doi.org/10.1007/BF02803211

Perrigo, A., Hoorn, C. & Antonelli, A. 2020. Why mountains matter for biodiversity. *J. Biogeogr.* 47: 315–325. https://doi.org/10.1111/jbi.13731

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

R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rokas, A., Atkinson, R.J., Webster, L.M.I., Csóka, G. & Stone G.N. 2003. Out of Anatolia: longitudinal gradients in genetic diversity support an eastern origin for a circum-Mediterranean oak gallwasp *Andricus quercustozae*. *Mol. Ecol*. 12: 2153–2174. https://doi.org/10.1046/j.1365-294X.2003.01894.x

Rosenberg, N.A. 2004. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* 4: 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Sarikaya, M.A. & Çiner, A. 2015. Late Pleistocene glaciations and paleoclimate of Turkey. *Bull. Min. Res. Exp*. 151: 107–127.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9: 671–675. https://doi.org/10.1038/nmeth.2089

Schönswetter, P., Suda, J., Popp, M., Weiss-Schneeweiss, H. & Brochmann, C. 2007. Circumpolar phylogeography of *Juncus biglumis* (Juncaceae) inferred from AFLP fingerprints, cpDNA sequences, nuclear DNA content and chromosome numbers. *Mol. Phylogenet. Evol.* 42: 92–103. https://doi.org/10.1016/j.ympev.2006.06.016

Sobierajska, K., Boratyńska, K., Jasińska, A., Dering, M., Ok, T., Douaihy, B., Dagher-Kharrat, M.B., Romo, Á. & Boratyński, A. 2016. Effect of the Aegean Sea barrier between Europe and Asia on differentiation in *Juniperus drupacea* (Cupressaceae). *Bot. J. Linn. Soc.* 180: 365–385. https://doi.org/10.1111/boj.12377

Stevanović, V. 1996. Analysis of the Central European and Mediterranean orophytic element on the mountains of the W. and Central Balkan Peninsula, with special reference to endemics. *Bocconea* 5: 77–97.

Strid, A. 1986. The mountain flora of Greece with special reference to the Anatolian element. *Proc. R. Soc. Edinb.* 89: 59–68. https://doi.org/10.1017/S0269727000008903

Šlenker, M., Kantor, A., Marhold, K., Schmickl, R., Mandáková, T., Lysak, M.A., Perný, M., Caboňová, M., Slovák, M. & Zozomová-Lihová, J. 2021. Allele sorting as a novel approach to resolving the origin of allotetraploids using Hyb-Seq data: A case study of the Balkan mountain endemic *Cardamine barbaraeoides*. *Front. Plant Sci.* 12: 659275. https://doi.org/10.3389/fpls.2021.659275

Šlenker, M., Koutecký, P. & Marhold, K. 2022. MorphoTools2: an R package for multivariate morphometric analysis. *Bioinformatics* 38: 2954–2955. https://doi.org/10.1093/bioinformatics/btac173

Španiel, S. & Rešetnik, I. 2022. Plant phylogeography of the Balkan Peninsula: spatiotemporal patterns and processes. *Plant Syst. Evol*. 308: 38. https://doi.org/10.1007/s00606-022-01831-1

Temsch, E.M., Greilhuber, J. & Krisai, R. 2010. Genome size in liverworts. *Preslia* 82: 63–80.

Thiers, B.M. 2023. Index Herbariorum. https://sweetgum.nybg.org/science/ih/ (accessed 8 Mar 2023).

Thompson, J.D. 2020. *Plant Evolution in the Mediterranean: Insights for conservation, 2nd Edition.* New York: Oxford University Press. https://doi.org/10.1093/oso/9780198835141.001.0001

Ursenbacher, S., Schweiger, S., Tomović, L., Crnobrnja-Isailović, J., Fumagalli, L. & Mayer, W. 2008. Molecular phylogeography of the nose-horned viper (*Vipera ammodytes*, Linnaeus (1758)): Evidence for high genetic diversity and multiple refugia in the Balkan peninsula. *Mol. Phylogenet. Evol*. 46: 1116–1128. https://doi.org/10.1016/j.ympev.2007.11.002

Veith, M., Schmidtler, J.F., Kosuch, J., Baran, I. & Seitz, A. 2003. Palaeoclimatic changes explain Anatolian mountain frog evolution: a test for alternating vicariance and dispersal events. *Mol. Ecol*. 12: 185–199. https://doi.org/10.1046/j.1365-294X.2003.01714.x

Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A. & Liston, A. 2014. Combining target enrichment and genome skimming for plant phylogenomics. *Appl. Plant Sci.* 2: 1400042. https://doi.org/10.3732/apps.1400042

Wielstra, B. & Arntzen, J.W. 2011. Unraveling the rapid radiation of crested newts (*Triturus cristatus* superspecies) using complete mitogenomic sequences. *BMC Evol. Biol.* 11: 162. https://doi.org/10.1186/1471-2148-11-162

Yildirim, H., Özdöl, T., Can, I.D., Erdem, S., Oğur, E. & Özüdoğru, B. 2022. *Knautia goecmenii* (Caprifoliaceae): an unexpected caespitose suffrutescent new species from the Western Anatolia, Turkey. *Phytotaxa* 531: 97–110. https://doi.org/10.11646/phytotaxa.531.2.2

Yılmaz, Ö., Kaynak, G. & Vural, M. 2003. A new taxon of *Linum* (Linaceae) from NW Anatolia, Turkey. *Ann. Bot. Fenn.* 40: 147–150.

Yüzbaşıoğlu, S. 2017. *Crocus tuna-ekimii* (Iridaceae), a new species from Turkey. *Phytotaxa* 314: 110–116. https://doi.org/10.11646/phytotaxa.314.1.10

Yüzbaşıoğlu, S., Koch, M.A. & Al-Shehbaz, I.A.A. 2015. Proof of a knowledge database concept. *Aubrieta ekimii* (Brassicaceae), a new species from NW Anatolia (Turkey): morphological and molecular support. *Plant Syst. Evol.* 301: 2043–2055. https://doi.org/10.1007/s00606-015-1212-3

Yüzbaşıoğlu, S., Al-Shehbaz, I.A.A., Yüzbaşıoğlu, E. & Dalyan, E. 2017. *Berteroa physocarpa* (Brassicaceae), a new species from NW Turkey based on morphological and molecular data. *Phytotaxa* 305: 87–96. https://doi.org/10.11646/phytotaxa.305.2.2

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform*. 19: 153. https://doi.org/10.1186/s12859-018-2129-y

Table 1. *List of morphological characters measured or scored and included in morphometric analyses.*

Vegetative characters	
WS	width of stem at the base [mm]
LSL*‡	stem length [cm]
NB	number of stem branches (including the main stem) bearing inflorescences
NB2	number of lateral stem branches not bearing inflorescences; usually in lower part of the main stem
NL	number of stem leaves
NLR	degree of congestion of leaves beneath the inflorescence on the main stem, expressed by the number of leaves reaching the base of the uppermost leaf
LC1	length of the longest basal leaf [cm]
LC2†	length of the middle stem leaf [cm]
LC3*	length of the uppermost leaf [cm]
NFB	number of pairs of lateral leaflets of the basal leaf
LTB*	length of terminal leaflet of the basal leaf [cm]
WTB*	width of terminal leaflet of the basal leaf [cm]
LLB*	length of the uppermost lateral leaflet of the basal leaf [cm]
WLB*	width of the uppermost lateral leaflet of the basal leaf [cm]
NFS†	number of pairs of lateral leaflets of the middle stem leaf
LTS*†	length of terminal leaflet of the middle stem leaf [cm]
WTS*†	width of terminal leaflet of the middle stem leaf [cm]
LLS*†	length of the uppermost lateral leaflet of the middle stem leaf [cm]
WLS*†	width of the uppermost lateral leaflet of the middle stem leaf [cm]
NTL2†$	number of teeth/lobes on one side of the margin of the terminal leaflet of the middle stem leaf
NTL3$	number of teeth/lobes on one side of the margin of the terminal leaflet of the uppermost leaf
RHIZ	growth orientation of the rhizome: (0) vertical (at 0-45° angle to the vertical axis) or (1) creeping (at 45-90° angle to the vertical axis)
BASE	stem base: (0) erect or (1) ascending
STEM	stem above the stem base: (0) erect or (1) ascending
Floral characters	
LP	petal length [mm]
WP*	petal width [mm]
LS	sepal length [mm]
LFL	length of longer filaments [mm]
LFS*	length of shorter filaments [mm]
COL	petals: (0) white or (1) pink, violet or white with a pink tone
Ratio characters	
NL/LSL, LC2/LSL, LC3/LSL, LC3/LC2, WTB/LTB, WLB/LLB, LLB/LTB, WLB/WTB, WTS/LTS, WLS/LLS, LLS/LTS, WLS/WTS, WP/LP, LFS/LFL	

* characters used only for computing ratios
‡ stem length = height of stem from its base to the base of the uppermost stem leaf
† middle stem leaf = the leaf closest to the midpoint of the leafy part of the main stem (LSL/2 point)
$ one side of the leaflet margin = one half of the leaflet margin delineated by the midvein (one side from the base to the apex of the leaflet)

Table 2. *Diagnostic characters of the studied Cardamine species. Values of morphological characters are expressed by the 5th–95th percentiles.*

	C. anatolica	***C. acris***	***C. penzesii***	***C. tenera***	***C. uliginosa***
Stem base	mostly erect	erect or ascending	mostly erect	mostly ascending	ascending
Stem branches	0–7	0–8	2–8	0–2	0–2
Basal leaves	2.8–9.9 cm long, with 3–6 pairs of lateral leaflets	3.5–17.6 cm long, with 2–6 pairs of lateral leaflets	4.9–16.3 cm long, with 4–7 pairs of lateral leaflets	5.6–12.6 cm long, with 2–5 pairs of lateral leaflets	3.2–15.6 cm long, with 2–7 pairs of lateral leaflets
– Terminal leaflet	1.5–4 times longer and 1.5–4 times wider than the first lateral leaflets	1.5–3 times longer and 2–3.5 times wider than the first lateral leaflets	up to 2 times longer and 1.5–3 times wider than the first lateral leaflets	up to 2.5 times longer and 1.5–3.5 times wider than the first lateral leaflets	up to 2 times longer and up to 3 times wider than the first lateral leaflets
Middle stem leaves	1.3–8.9 cm long (6–29% of stem length), with 1–4 pairs of lateral leaflets	2.3–10.8 cm long (8–23% of stem length), with 2–4 pairs of lateral leaflets	3.2–9.3 cm long (11–23% of stem length), with 3–5 pairs of lateral leaflets	4.4–10.4 cm long (18–44% of stem length), with 2–5 pairs of lateral leaflets	1.9–8.2 cm long (9–29% of stem length), with 2–7 pairs of lateral leaflets
– Terminal leaflet	elliptic to broadly circular; 1.5–3 times longer and 2.5–5.5 times wider than the first lateral leaflets; with 2–7 teeth or lobes on one side of the leaf margin	broadly elliptic to broadly circular; 1.5–3 times longer and 2–4.5 times wider than the first lateral leaflets; with 3–7 teeth or lobes on one side of the leaf margin	narrowly to broadly elliptic; up to 2 times longer and up to 3 times wider than the first lateral leaflets; with 1–3 teeth or lobes on one side of the leaf margin	elliptic to circular; 1.5–2.5 times longer and 2–4 times wider than the first lateral leaflets; with 3–7 teeth or lobes on one side of the leaf margin	narrowly to broadly elliptic; up to 2 times longer and up to 3 times wider than the first lateral leaflets; with 0–5 teeth or lobes on one side of the leaf margin
– Lateral leaflet	oblong to broadly elliptic	elliptic to circular	narrowly elliptic to elliptic	oblong to broadly elliptic	oblong to broadly elliptic
Uppermost stem leaves	0.7–3.9 cm long (2–13% of stem length)	1–3.7 cm long (2–15% of stem length)	1.4–3.8 cm long (5–12% of stem length)	1.8–7.4 cm long (9–34% of stem length)	0.8–3.5 cm long (3–18% of stem length)
Rhizome	mostly vertical	vertical or creeping	vertical, short, strongly tuberous	mostly creeping	creeping
Petals	white; 9.3–13.5 mm long, 4.3–6.5 mm wide	white, pink, violet or white with a pink tone; 7.4–11.8 mm long, 3.5–7.4 mm wide	white; 10.2–13.4 mm long, 5.6–8.3 mm wide	white, pink, violet or white with a pink tone; 8.1–11.9 mm long, 4.8–7.9 mm wide	white, pink, violet or white with a pink tone; 8.1–11.6 mm long, 4.5–7.5 mm wide

FIGURE LEGENDS

Fig. 1. Distribution and sample sites of the studied *Cardamine* species in northern Türkiye and neighbouring regions. Localities of the sampled populations are provided in Appendix 1. Distribution data (all records) are derived from herbarium records, field observations, literature and database data, as listed in detail in Table S1. Blue dotted outlines show four intraspecific lineages of *C. uliginosa* from Anatolia (Anat A – Anat D) recognized in the maximum-likelihood tree based on the concatenated nuclear genes (see Fig. 2A). The arrow indicates the placement of Lebanese populations within the lineage Anat A(-Leb). The inset map shows distribution of *C. acris* in the Balkan Peninsula (Kantor & al., unpubl. data).

Fig. 2. Phylogenetic inferences based on the Hyb-Seq target nuclear genes. (A) Maximum-likelihood tree (RAxML-NG) inferred from the concatenated sequences of all 1104 nuclear genes. Branch support is shown with bootstrap values (BS > 50%) and quartet sampling scores (QC/QD/QI) for the main branches; coloured circles in the nodes indicate QC value intervals. Accession codes follow Appendix 1. The position of *C. anatolica* described here as a new species is highlighted. Clades resolved within *C. uliginosa* are marked with horizontal bars and labelled. Bar colours refer to Bayesian genetic clusters obtained in STRUCTURE at K = 2 (green and red, two-coloured bars indicate admixture). (B) Species tree inferred in ASTRAL-III based on 1104 gene trees. Branch support is indicated by pie charts, depicting three local posterior probabilities for the given branch. (C) Distribution of the analyzed accessions of *C. uliginosa* and their assignment to Bayesian STRUCTURE clusters at K = 2. The STRUCTURE analysis was based on SNP data of 957 nuclear genes, see text for detail.

Fig. 3. Variation in relative genome size in the studied *Cardamine* species as inferred from flow cytometric measurements using DAPI fluorochrome. Relative genome size values (2C values in arbitrary units, a. u.) represent the ratio of G1 peaks of the sample and standard (*Solanum pseudocapsicum*; 2C = 2.59 pg). Lower and upper outlines of boxes show 25th and 75th percentiles, vertical lines inside are median values, and whiskers range from the minimum to maximum values. The number of analyzed individuals for each species is indicated (for more details, see Appendix 1, Table S2).

Fig. 4. Morphometric analyses of the studied *Cardamine* species. (A) CDA1, canonical discriminant analysis based on 608 individuals and 32 morphological characters with the studied species predefined as five groups; diagrams with the canonical axes 1 and 2 (left diagram) and axes 2 and 3 (right diagram) are shown; (B) CDA2 based on 348 individuals and 32 morphological characters and three predefined groups: *C. anatolica, C. tenera*, and *C. uliginosa*; (C) CDA3 based on 316 individuals and 32 morphological characters and three predefined groups: *C. anatolica*, *C. acris*, and *C. penzesii*. Values of total canonical structure are listed in Table S3.

Fig. 5. Habitus of *Cardamine anatolica* (A), with details of its inflorescence (B), middle stem leaf (C), and leaves forming a basal rosette (D). Habitat of *C. anatolica*, population DEL in Yalova prov., Delmece pasture, 725 m a.s.l. (E); and population ALA in Bursa prov., eastern part of the Uludağ, 1770 m a.s.l. (F).

Appendix 1. List of the sampled *Cardamine* populations analyzed in the present study and the corresponding voucher specimens: *C. acris* (29 populations), *C. anatolica* (6 populations), *C. penzesii* (9 populations), *C. tenera* (7 populations), *C. uliginosa* (39 populations), *C. matthioli* (1 population), *C. rivularis* (3 populations), and *C. lazica* (2 populations) as an outgroup. Details for each population are presented in the following order: population code; country of origin; locality; latitude (N); longitude (E); collector name(s) (herbarium acronym); Biosample accession code(s) from the NCBI database for samples included in Hyb-Seq analyses; number of individuals analyzed by flow cytometry (denoted by "FCM" and a number); number of individuals included in morphometric analyses (denoted by "morph" and a number). Biosample accessions used in the present study are part of BioProjects PRJNA687126 (Šlenker & al., 2021) and PRJNA830631 (Kantor & al., 2023 and the here generated data, marked by asterisks).

Appendix 2. List of examined herbarium specimens of *Cardamine anatolica*. Herbarium acronyms follow Thiers (2023).

Supplementary material

Supplementary Table S1. Distribution records of *Cardamine anatolica*, *C. penzesii* and *C. uliginosa* in Türkiye based on herbarium records, field observations, literature and database data. In the column 'Source', acronyms of herbaria follow Index Herbariorum (Thiers, 2023).

Supplementary Table S2. Summary of the relative genome size values of the studied *Cardamine* species and the results of Mann-Whitney pairwise comparison tests with Bonferroni correction of P values. Relative genome size is expressed as the sample/standard G1 peak ratio (standard: *Solanum pseudocapsicum*; 2C = 2.59 pg). n, number of measured individuals.

Supplementary Table S3. Results of canonical discriminant analyses (CDA) of the studied *Cardamine* species. The total canonical structure coefficients are listed for CDA1-CDA7 (Fig. 4, Suppl. Fig. S3), expressing the correlations of the characters with the canonical axes (Can1-Can4). Characters with the highest values are highlighted in bold. Character abbreviations follow Table 1. CDA1: CDA with all studied species predefined as five groups, *C. acris*, *C. anatolica*, *C. penzesii*, *C. tenera*, and *C. uliginosa* (Fig. 4A); CDA2: CDA with three predefined groups, *C. anatolica*, *C. tenera*, and *C. uliginosa* (Fig. 4B); CDA3: CDA with three predefined groups, *C. anatolica*, *C. acris*, and *C. penzesii* (Fig. 4C). CDA4-7: each with two predefined groups: *C. anatolica* and *C. acris* (CDA4, Suppl. Fig. S3A); *C. anatolica* and *C. penzesii* (CDA5, Suppl. Fig. S3B); *C. anatolica* and *C. tenera* (CDA6, Suppl. Fig. S3C); *C. anatolica* and *C. uliginosa* (CDA7, Suppl. Fig. S3D).

Supplementary Fig. S1. Distribution of the analyzed accessions of *Cardamine uliginosa* and their assignment to Bayesian STRUCTURE clusters at K = 3. The STRUCTURE analysis was based on SNP data of 957 target nuclear genes obtained by Hyb-Seq.

Supplementary Fig. S2. Maximum-likelihood tree (RAxML-NG) of the complete plastome data (LSC, SSC, IRb) with bootstrap values shown above the main branches. The position of *Cardamine anatolica* described here as a new species is highlighted. Clades of *C. uliginosa* are marked with thick horizontal bars; those referring to Anatolian accessions are in green and labelled

(Anat I, Anat II). Three accessions in green letters were placed within Anatolian clades in the nuclear ML tree, while here they cluster within the Caucasian clades.

Supplementary Fig. S3. Morphological differentiation of *Cardamine anatolica* from the closest relatives explored with canonical discriminant analyses CDA4-CDA7, each based on 32 morphological characters and individual plants, with two predefined groups. Values of total canonical structure are listed in Table S3.

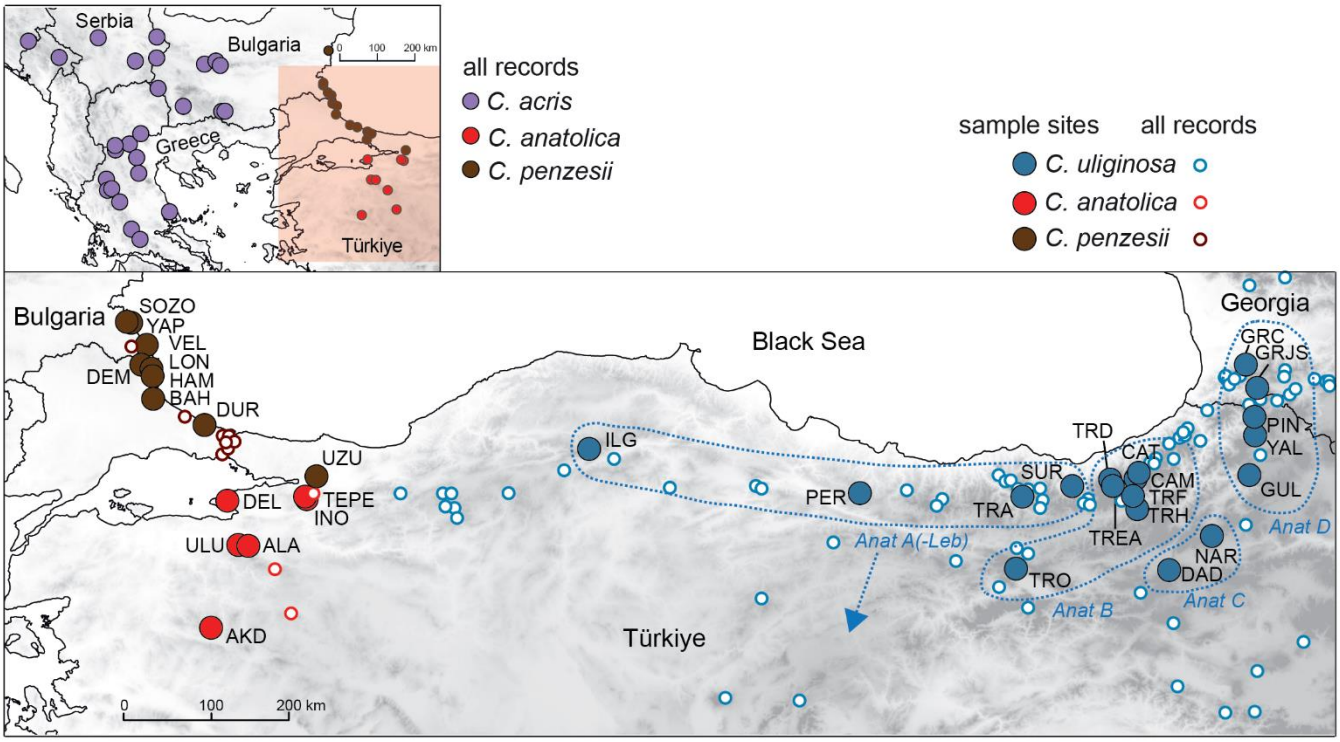


Fig. 1. Distribution and sample sites of the studied *Cardamine* species in northern Türkiye and neighbouring regions. Localities of the sampled populations are provided in Appendix 1. Distribution data (all records) are derived from herbarium records, field observations, literature and database data, as listed in detail in Table S1. Blue dotted outlines show four intraspecific lineages of *C. uliginosa* from Anatolia (Anat A – Anat D) recognized in the maximum-likelihood tree based on the concatenated nuclear genes (see Fig. 2A). The arrow indicates the placement of Lebanese populations within the lineage Anat A(-Leb). The inset map shows distribution of *C. acris* in the Balkan Peninsula (Kantor & al., unpubl. data).

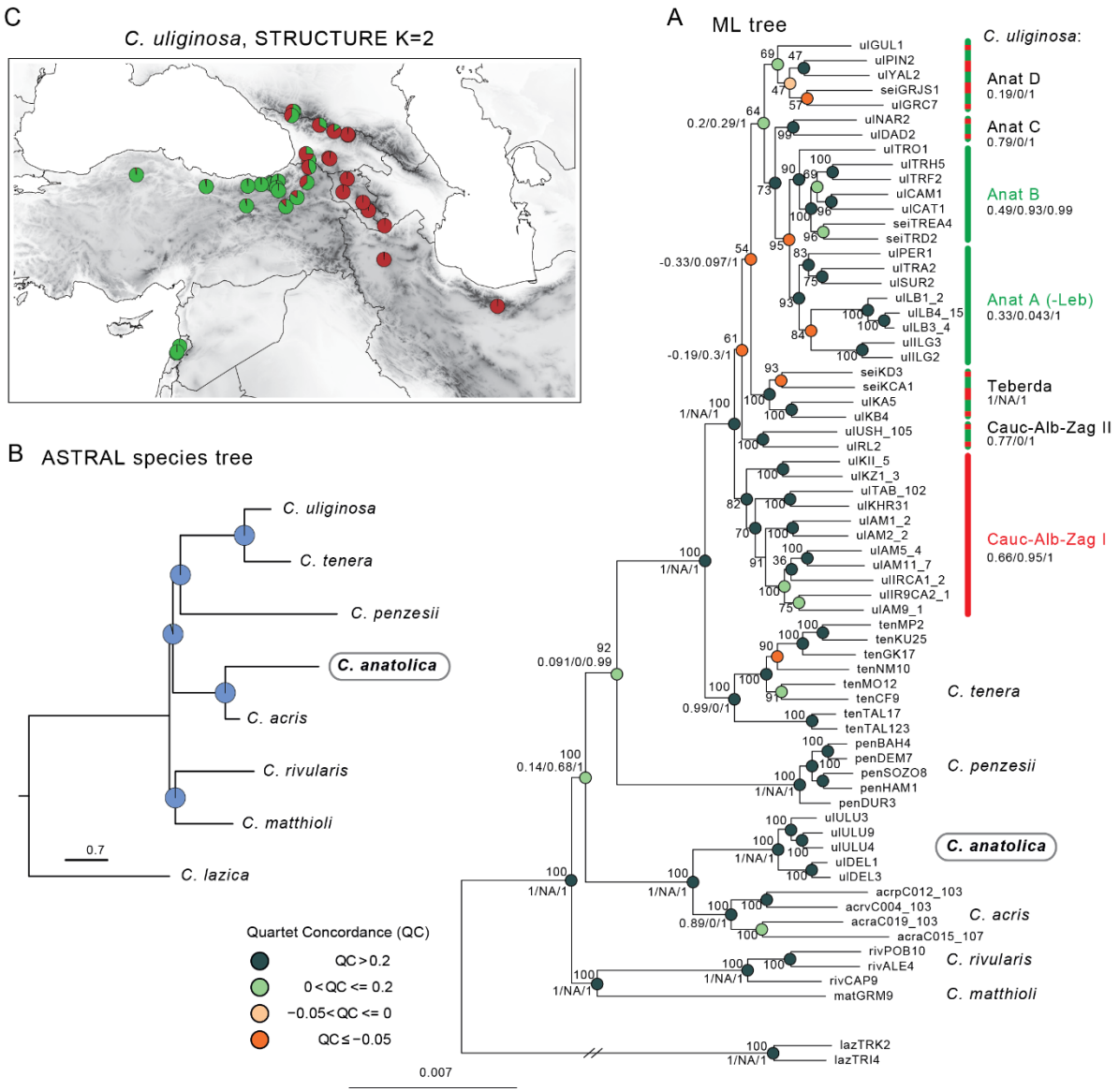


Fig. 2. Phylogenetic inferences based on the Hyb-Seq target nuclear genes. (A) Maximum-likelihood tree (RAxML-NG) inferred from the concatenated sequences of all 1104 nuclear genes. Branch support is shown with bootstrap values (BS > 50%) and quartet sampling scores (QC/QD/QI) for the main branches; coloured circles in the nodes indicate QC value intervals. Accession codes follow Appendix 1. The position of *C. anatolica* described here as a new species is highlighted. Clades resolved within *C. uliginosa* are marked with horizontal bars and labelled. Bar colours refer to Bayesian genetic clusters obtained in STRUCTURE at K = 2 (green and red, two-coloured bars indicate admixture). (B) Species tree inferred in ASTRAL-III based on 1104 gene trees. Branch support is indicated by pie charts, depicting three local posterior probabilities for the given branch. (C) Distribution of the analyzed accessions of *C. uliginosa* and their assignment to Bayesian STRUCTURE clusters at K = 2. The STRUCTURE analysis was based on SNP data of 957 nuclear genes, see text for detail.

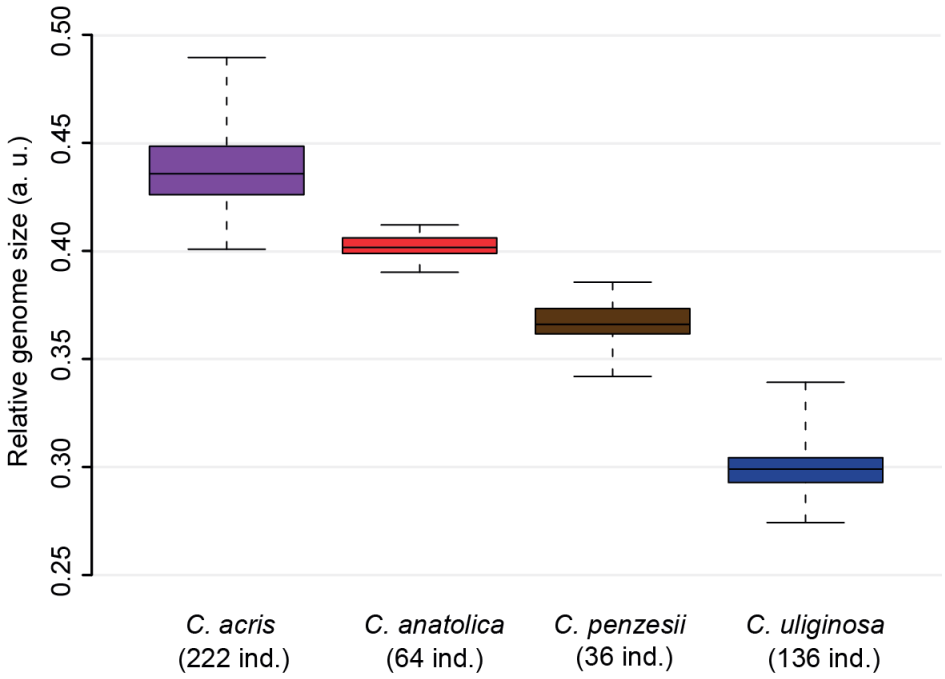


Fig. 3. Variation in relative genome size in the studied *Cardamine* species as inferred from flow cytometric measurements using DAPI fluorochrome. Relative genome size values (2C values in arbitrary units, a. u.) represent the ratio of G1 peaks of the sample and standard (*Solanum pseudocapsicum*; 2C = 2.59 pg). Lower and upper outlines of boxes show 25th and 75th percentiles, vertical lines inside are median values, and whiskers range from the minimum to maximum values. The number of analyzed individuals for each species is indicated (for more details, see Appendix 1, Table S2).

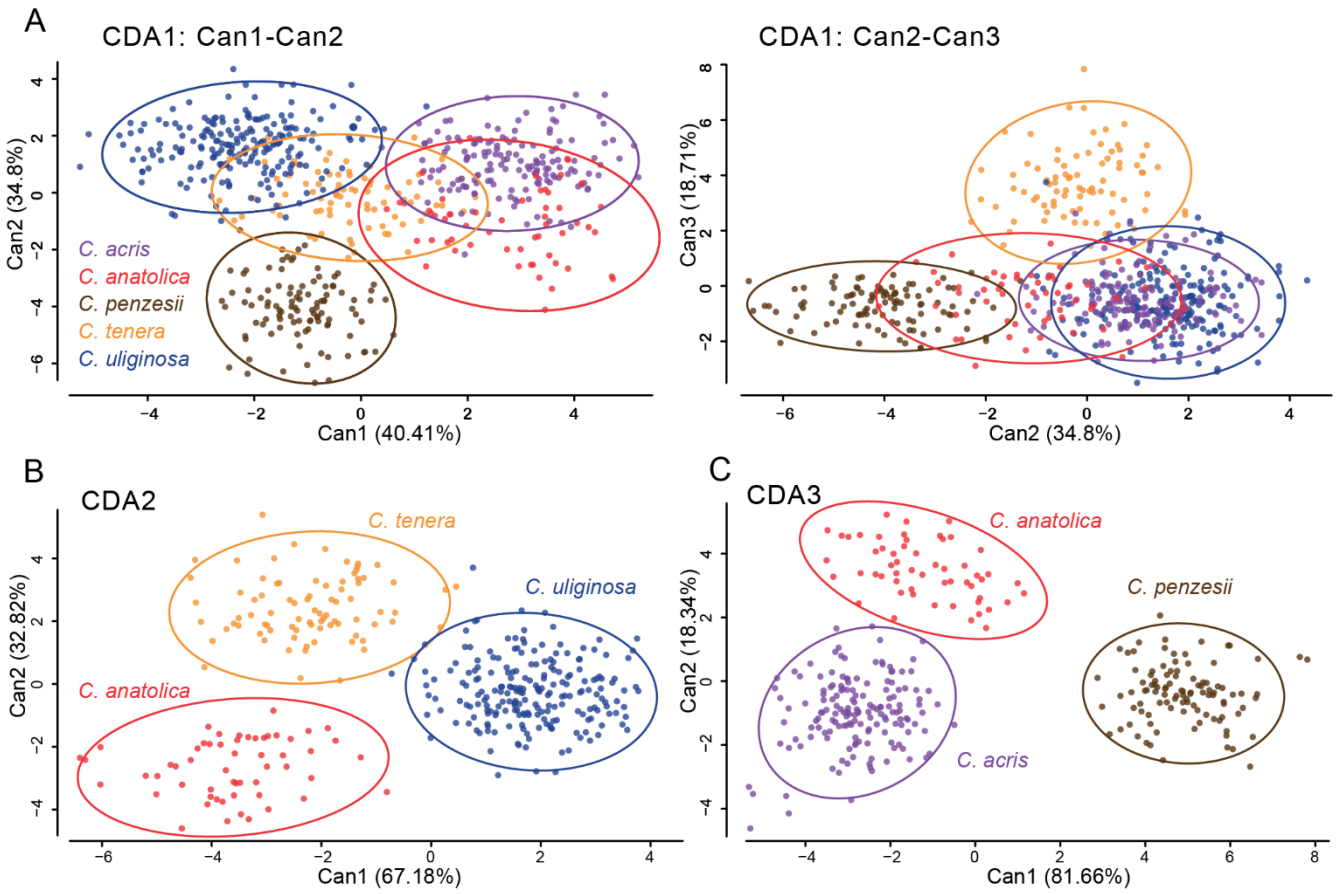


Fig. 4. Morphometric analyses of the studied *Cardamine* species. (A) CDA1, canonical discriminant analysis based on 608 individuals and 32 morphological characters with the studied species predefined as five groups; diagrams with the canonical axes 1 and 2 (left diagram) and axes 2 and 3 (right diagram) are shown; (B) CDA2 based on 348 individuals and 32 morphological characters and three predefined groups: *C. anatolica, C. tenera*, and *C. uliginosa*; (C) CDA3 based on 316 individuals and 32 morphological characters and three predefined groups: *C. anatolica*, *C. acris*, and *C. penzesii*. Values of total canonical structure are listed in Table S3.

Fig. 5. Habitus of *Cardamine anatolica* (A), with details of its inflorescence (B), middle stem leaf (C), and leaves forming a basal rosette (D). Habitat of *C. anatolica*, population DEL in Yalova prov., Delmece pasture, 725 m a.s.l. (E); and population ALA in Bursa prov., eastern part of the Uludağ, 1770 m a.s.l. (F).

Appendix 1. List of the sampled *Cardamine* populations analyzed in the present study and the corresponding voucher specimens: *C. acris* (29 populations), *C. anatolica* (6 populations), *C. penzesii* (9 populations), *C. tenera* (7 populations), *C. uliginosa* (39 populations), *C. matthioli* (1 population), *C. rivularis* (3 populations), and *C. lazica* (2 populations) as an outgroup. Details for each population are presented in the following order: population code; country of origin; locality; latitude (N); longitude (E); collector name(s) (herbarium acronym); Biosample accession code(s) from the NCBI database for samples included in Hyb-Seq analyses; number of individuals analyzed by flow cytometry (denoted by "FCM" and a number); number of individuals included in morphometric analyses (denoted by "morph" and a number). Biosample accessions used in the present study are part of BioProjects PRJNA687126 (Šlenker & al., 2021) and PRJNA830631 (Kantor & al., 2023 and the here generated data, marked by asterisks).

Cardamine acris Griseb.; BAB, Serbia, Stara planina Mts., Babin zub peak, stream, 43.3825, 22.601056, *Skokanová, K., Kačmárová, T. & Šlenker, M. s.n.* (SAV), –, FCM 8, –; C002, Greece, Central Macedonia, Pella Region, Voras Mts., Voras Ski Resort, stream along the road above the resort, 40.913046, 21.809686, *Šlenker, M. s.n.* (SAV), –, FCM 8, –; C003, Greece, Thessaly Region, Volos, W from Chania, concrete ditch along the road to Chania, 39.393399, 23.042808, *Šlenker, M. s.n.* (SAV), –, FCM 8, –; C004, Greece, Sterea Ellas, Fokida, Vardousia Mts., Athanasios Diakos, 38.69886, 22.142021, *Perný, M. & Kučera, V. s.n.* (SAV), *Šlenker, M. s.n.* (SAV), SAMN17256855, FCM 10, morph 43; C012, Greece, N Pindhos, Grevena, Vasilitsa Ski Resort, stream below road, 40.05725, 21.081888, *Šlenker, M. s.n.* (SAV), SAMN17256856, FCM 10, –; C013, Greece, Western Macedonia, Florina Region, Varnous Mts., Pisoderi Ski Resort, close to the ski resort, wet places below a ski-lift, 40.772174, 21.266527, *Šlenker, M. s.n.* (SAV), –, FCM 8, –; C015, Serbia, Kopaonik mountains, by the path to Metođe, 43.308559, 20.846734, *Šlenker, M. s.n.* (SAV), SAMN17256853, FCM 10, –; C016, Montenegro, Berane, Vranjak, Zekova Glava, 42.844094, 19.650309, *Šlenker, M. s.n.* (SAV), –, FCM 8, –; C019, Bulgaria, Kyustendil, Osogovska planina, above Osogovo chalet, 42.199947, 22.614689, *Perný, M. s.n.* (SAV), *Šlenker, M. s.n.* (SAV), SAMN17256854, FCM 5, morph 22; C023, Bulgaria, Blagoevgrad, Bansko, Pirin, Near Demjanica chalet, 41.74585, 23.467073, *Šlenker, M. s.n.* (SAV), *Bartolić, P. & Padilla-García, N. s.n.* (SAV), –, FCM 19, –; C025, Bulgaria, Smolyan, Rodopy Mts., Smolyanski ezera area, 41.62025, 24.68079, *Šlenker, M. s.n.* (SAV), –, FCM 10, –; C026, Bulgaria, Lovech, Stara planina, Tetevenska planina, by the path to Vežen hut, 42.776703, 24.389592, *Šlenker, M. s.n.* (SAV), –, FCM 2, –; C151, North Macedonia, Vardar, Ski center Kozuf, 41.183857, 22.202842, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 9, –; C152, North Macedonia, Pelagonia, National Park Pelister, under Golemo Ezero lake, 40.978188, 21.221972, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 6, –; C154, Greece, Epirus, Katara Pass, 39.795492, 21.223945, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 9, –; C156, Greece, Thessaly, Pyrra, below Mantra peak, 39.553711, 21.425723, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 8, –; C157, Greece, West Macedonia, Velventos, 40.219676, 22.094289, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 10, –; C158, Greece, Central Macedonia, Ano Seli, 40.607367, 21.946245, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 9, –; C159, Bulgaria, Plovdiv, Stara Planina, yellow tourist trail 4 km N of Klisura, 42.72835, 24.457522, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 9, –; C161, Bulgaria, Plovdiv, Etropolska Planina, 2 km SW of Zlatiski Prokhod, 42.733078, 24.062361, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 11, –; C163, Bulgaria, Plovdiv, Ruj

Planina, 1 km NEE of Mount Ruj, 42.865635, 22.585854, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 10, –; C240, Bosnia and Herzegovina, Republika Srbska, Foča-Tjeniste, NP Sutjeska - Prijevor, stream 2 km NW of Maglić peak, 43.291111, 18.715212, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 10, –; C245, Serbia, Central Serbia province, Pčinja district, streams around Planinarski dom Preka voda, 42.787939, 21.954076, *Šlenker, M. & Kantor, A. s.n.* (SAV), –, FCM 9, –; GOL, Bulgaria, Smolyan, Rhodopy Mts., Golyam Perelik, 41.615917, 24.633583, *Skokanová, K., Kačmárová, T. & Šlenker, M. s.n.* (SAV), –, FCM 9, –; KTP, Greece, Ioannina, N Pindhos, Katara Pass above Metsovo, 39.781052, 21.201579, *Perný, M. & Kučera, V. s.n.* (SAV), *Perný, M. s.n.* (SAV), –, –, morph 13; OSA, Bulgaria, Kyustendil, Osogovska planina, Gramadite, 42.193111, 22.604306, *Skokanová, K., Kačmárová, T. & Šlenker, M. s.n.* (SAV), –, FCM 7, –; TFM, Greece, Sterea Ellas, Evritania, Timfristos Mts., Karpenisi Ski Resort, 38.944726, 21.808016, *Perný, M. & Kučera, V. s.n.* (SAV), –, –, morph 31; VRA, Greece, Central Macedonia, Pella Region, Voras Mts., Voras Ski Resort, 40.913046, 21.809686, *Perný, M. & Kučera, V. s.n.* (SAV), –, –, morph 23; VZN, Bulgaria, Lovech, Stara planina, Tetevenska planina, NW of Mt. Vezhen, 42.760631, 24.396833, *Perný, M. & Georgieva, E. s.n.* (SAV), –, –, morph 25. ***Cardamine anatolica*** Jar.Kučera et al.; AKD, Türkiye, Kütahya province, Akdağ, stream in pine forest, 39.2796942, 28.7825035, *Slovák, M., Kučera, J., Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), –, FCM 9, –; ALA, Türkiye, Bursa province, E of Uludağ Nature Park, Kestel, Alacam village, road between Alacam and lakes region, stream side, 40.094000, 29.2808889, *Kantor, A., Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), –, FCM 10, –; DEL, Türkiye, Yalova province, Çınarcık, Delmece pasture, stream side, 40.553167, 29.006444, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), *Kantor, A., Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482344*, SAMN33482345*, FCM 13, morph 20; INO, Türkiye, Kocaeli province, Tepecik village, İnönü plateau, wet meadow, 40.571278, 30.011528, *Kantor, A., Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), –, FCM 10, –; TEPE, Türkiye, Kocaeli province, Tepecik village, road between Tepecik village and İnönü plateau, stream side, 40.594722, 30.005333, *Kantor, A., Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), –, FCM 10, morph 20; ULU, Türkiye, Bursa Province, Uludağ, ski resort, 40.102699, 29.153511, *Kučera, J., Slovák, M. & Bérešová s.n.* (SAV 0017449, HUB, ISTE), *A., Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN17256861, SAMN27735366, SAMN33482353*, FCM 12, morph 16. ***Cardamine lazica*** Boiss. & Balansa ex Boiss.; TRI, Türkiye, Rize, Doğ Karadeniz Dağlari, Cimil, 40.7553581, 40.6851133, *Kučera, J. & Kolník, M. s.n.* (SAV, HUB), SAMN27735409, –, –; TRK, Türkiye, Giresun, Giresun Dağlari, Güdül, 40.6394531, 38.4563044, *Kučera, J. & Kolník, M. s.n.* (SAV, HUB), SAMN27735408, –, –. ***Cardamine matthioli*** Moretti ex Comolli; GRM, Bulgaria, Blagoevgrad Province, W Rhodopi Mts., Metsa valley, close to Grmen village, 41.584778, 23.787750, *Šlenker, M. et al. s.n.* (SAV), SAMN17256857, –, –. ***Cardamine penzesii*** Ančev & Marhold; BAH, Türkiye, Instanbul, Bahçeköy, 41.548611, 28.046944, *Marhold, K. s.n.* (SAV), SAMN27735406, FCM 4, –; DEM, Türkiye, Kırklareli, between Demirköy and İğneada, close to Avcılar, 41.876389, 27.909444, *Marhold, K. s.n.* (SAV), SAMN17256860, FCM 4, –; DUR, Türkiye, Durusu entrance, 41.295861, 28.668917, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482339*, FCM 8, –; HAM, Türkiye, Longoz forest, road side of Hamam Village, 41.822472, 27.964639, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482340*, FCM 7, –; LON, Türkiye, Kırklareli, İğneada, entrance to Longoz forest, 41.870583, 27.934917,

Dönmez, A.A. & Yüzbaşıoğlu, S. s.n. (SAV, HUB), –, FCM 9, –; SOZO, Bulgaria, Burgas Province, Primorsko, flood-plain forest next to Ropotamo river, near the bridge of the road Primorsko-Tsarevo, 42.302481, 27.728755, *Marhold, K. & Ančev, M. s.n.* (SAV), *Marhold, K. & Vassilev R. s.n.* (SAV), *Zozomová-Lihová, J. & Marhold, K. s.n.* (SAV), SAMN27735405, FCM 3, morph 33; UZU, Türkiye, Kocaeli, Kartepe, Uzuntarla, Ormanya nature park, 40.733928, 30.164378, *Dönmez, A.A., Yüzbaşıoğlu, S. & Kantor, A. s.n.* (SAV, HUB), –, FCM 1, –; VEL, Bulgaria, Burgas Province, Natural Park Strandzha, Sinemorets, Veleka river, ca. 2 km W of the bridge of the road Sinemorets-Rezovo, 42.060176, 27.943757, *Marhold, K. & Vassilev, R. s.n.* (SAV), –, –, morph 37; YAP, Bulgaria, Burgas Province, ca. 2 km N of the road Yasna Polyana - Primorsko (ca. 6 km of Yasna Polyana), 42.285256, 27.684185, *Marhold, K. & Ančev, M. s.n.* (SAV), –, –, morph 33. ***Cardamine rivularis*** Schur; ALE, Bulgaria, Sofia Province, Vitosha Mts., Aleko cottage, 42.586667, 23.290056, *Šlenker, M. et al. s.n.* (SAV), SAMN27735403, –, –; CAP, Romania, Argeș county, Făgăraş Mts., Lacul Capra, 45.595361, 24.634583, *Kolář, F. s.n.* (SAV), SAMN27735402, –, –; POB, Bulgaria, Rhodopi Mts., Pobit Kamak, 41.830556, 23.870278, *Šlenker, M. et al. s.n.* (SAV), SAMN17256858, –, –. ***Cardamine tenera*** J.G.Gmel. ex C.A.Mey; CF, Russia, Krasnodarskii krai, Sochinskii rajon, Sochi, part Adler, 43.426100, 39.924200, *Perný, M. & Tuniev, B.S.*, SAMN27735395, –, morph 4; GK, Russia, Krasnodarskii krai, E of Goryachii Klyuch, 44.5686869, 39.228525, *Perný, M. s.n.* (SAV), SAMN27735394, –, morph 7; KU, Russia, Republic of Adygea, Kurdzhipskaya near Maikop, S of the village, 44.4480878, 40.0482931, *Perný, M. s.n.* (SAV), SAMN27735393, –, morph 18; MO, Russia, Krasnodarskii krai, Sochinskii rajon, Mzymta valley, Monastyr, 43.5890756, 40.0140367, *Perný, M. & Tuniev, B.S. s.n.* (SAV), SAMN27735392, –, morph 24; MP, Russia, Republic of Adygea, Maikop, the left bank of the Belaya river, 44.5926389, 40.1060567, *Perný, M. s.n.* (SAV), SAMN27735391, –, morph 14; NM, Russia, Krasnodarskii krai, Sochinskii rajon, Sochi, part Novaya Macesta, 43.5478811, 39.7947614, *Perný, M.*, SAMN27735390, –, morph 15; TAL, Azerbaijan, Talysh Mts., SW of Lenkoran, W of Hirkan, 38.671760, 48.768225, *Zozomová-Lihová, J. s.n.* (SAV), SAMN27735388, SAMN27735389, –, –. ***Cardamine uliginosa*** M.Bieb.; AM1, Armenia, Lori distr., Vanadzor town, Pushkin pass, 40.9093833, 44.4396667, *Kučera, J. & Slovák, M. s.n.* (SAV), SAMN17256862, FCM 10, –; AM11, Armenia, Vayots Dzor distr., along Herher river, under Herher water reservoir, 39.6893167, 45.5222333, *Kučera, J. & Slovák, M. s.n.* (SAV), SAMN27735387, FCM 10, –; AM2, Armenia, Aragatsotn distr., near Amberd fortress, 40.405650, 44.2277333, *Kučera, J. & Slovák, M. s.n.* (SAV), SAMN27735386, FCM 10, morph 14; AM5, Armenia, Gegharkunik distr., Sulema pass, 40.0059333, 45.234750, *Kučera, J. & Slovák, M. s.n.* (SAV), SAMN27735385, FCM 10, morph 13; AM9, Armenia, Syunik distr., above Shiskert village, 39.0928833, 46.334400, *Kučera, J. & Slovák, M. s.n.* (SAV), SAMN27735384, FCM 9, –; CAM, Türkiye, Rize, Çamlıhemşin, Ortaklar-Verçenik road junction, 40.798083, 40.907472, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482341*, FCM 10, –; CAT, Türkiye, Rize, Çamlıhemşin, Çat-Verçenik pasture road, 40.821444, 40.937944, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482342*, FCM 1, –; DAD, Türkiye, Erzurum, Tekman road, Dadaş village, Çan pasture, 39.844333, 41.331583, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482343*, FCM 10, –; GRC, Georgia, Guria, Bakhmaro, 41.8664008, 42.3553725, *Kučera, J. s.n.* (SAV), SAMN27735383, –, morph 28; GRJS, Georgia, Adjara, Chulskij rajon, Goderzi pereval, 41.6304167, 42.4943667, *Kučera, J. s.n.* (SAV),

SAMN27735400, –, –; GUL, Türkiye, Erzurum, Şenkaya, Gülveren village, 40.773806, 42.402306, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482346*, FCM 10, –; ILG, Türkiye, Çankırı, Ilgaz, Çankırı side of Ilgaz Pass, 41.063583, 33.759722, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482347*, SAMN33482348*, FCM 9, –; IR1, Iran, Zagros Mts., East Azerbaijan province, Kandovan village, 37.7547222, 46.302500, *Slovák, M. s.n.* (SAV), SAMN27735381, –, –; IR2, Iran, Mazandaran Province, Alborz Mts., Polour village, 35.860000, 52.0541667, *Slovák, M. s.n.* (SAV), SAMN27735382, –, –; KA, Russia, Karachaevo-Cherkessian Republic, Teberda, Mala Khatipara, 43.441250, 41.691751, *Kučera, J. s.n.* (SAV), SAMN27735380, –, morph 33; KB, Russia, Karachaevo-Cherkessian Republic, Teberda, small lake near Mala Khatipara, 43.4458614, 41.6814994, *Kučera, J. s.n.* (SAV), SAMN27735379, –, –; KCA, Russia, Karachaevo-Cherkessian Republic, Dombay, 43.2908772, 41.6253319, *Kučera, J. s.n.* (SAV), SAMN27735399, –, –; KD, Russia, Karachaevo-Cherkessian Republic, Teberda, near Teberda river, 43.4364097, 41.7361364, *Kučera, J. s.n.* (SAV), SAMN27735398, –, –; KHR, Georgia, Samtskhe-Javakheti, S of Bakuriani, Tskhratskaro pass, 41.7030975, 43.5077306, *Zozomová-Lihová, J. s.n.* (SAV), SAMN27735378, –, morph 9; KII, Georgia, Mtskheta-Mtianeti, Kazbegi, N of Cross pass, 42.562205, 44.492040, *Hübl, E. s.n.* (SAV), SAMN27735377, –, –; KPU, Georgia, Mtskheta-Mtianeti, Cross pass (Djvris ugeltekhili), 42.5095139, 44.460835, *Marhold, K. s.n.* (SAV), –, –, morph 23; KZ1, Georgia, Mtskheta-Mtianeti, Mna valley mouth, 42.590000, 44.467500, *Valachovič, M. s.n.* (SAV), SAMN27735376, –, –; LB1, Lebanon, North Gov., near Harissa village, 34.198950, 35.952200, *Kučera, J., Slovák, M. & Bérešová, A. s.n.* (SAV), SAMN27735375, –, –; LB3, Lebanon, Mount Lebanon Gov., Mchaymcheh village, 33.9179333, 35.7993167, *Kučera, J., Slovák, M. & Bérešová, A. s.n.* (SAV), SAMN27735374, –, –; LB4, Lebanon, Mount Lebanon Gov., Faqra village, Qanat Bakish loc., 33.9627333, 35.8184833, *Kučera, J., Slovák, M. & Bérešová, A. s.n.* (SAV), SAMN27735373, –, –; NAR, Türkiye, Erzurum, Narman, 3 km from Çimenlli village to Narman, 40.191611, 41.896500, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482349*, FCM 9, –; PER, Türkiye, Ordu, Fatsa, Aybastı, Perşembe pasture, 40.618333, 37.298583, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482350*, FCM 10, –; PIN, Türkiye, Artvin, Şavşat, above Pınarlı village, 41.358278, 42.495972, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482351*, FCM 9, –; RL, Georgia, Racha-Lechkhumi, from Shovi to Mamisoni pass, 42.7060111, 43.7751039, *Schneeweiss, G.M. et al. s.n.* (SAV), SAMN27735372, –, –; SUR, Türkiye; Trabzon, Sürmene, Köprübaşı, Ağaçbaşı pasture, 40.695083, 40.08425, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482352*, FCM 10, –; TAB, Georgia, Samtskhe-Javakheti, Bakuriani, from Tkhratskaro pass to Tabatskuri lake, 41.6794994, 43.5589281, *Kolář, F. s.n.* (SAV), SAMN27735371, –, –; TRA, Türkiye, Trabzon, Kalkanli Dağlari, near Zigana Geç., 40.6378422, 39.4043397, *Kučera, J. & Kolník, M. s.n.* (SAV, HUB), SAMN27735370, –, morph 26; TRD, Türkiye, Rize, Doğ Karadeniz Dağlari, Ikizdere - Camcaruş, 40.7437433, 40.5972342, *Kučera, J. & Kolník, M. s.n.* (SAV, HUB), SAMN27735397, –, morph 11; TREA, Türkiye, Rize, Doğ Karadeniz Dağlari, Camlik, 40.7096583, 40.6416133, *Kučera, J. & Kolník, M. s.n.* (SAV, HUB), SAMN27735396, –, –; TRF, Türkiye, Erzurum, Doğ Karadeniz Dağlari, Ovit Dagi - Il Simiri, 40.5744228, 40.8621069, *Kučera, J. & Kolník, M. s.n.* (SAV, HUB), SAMN27735369, –, morph 26; TRH, Türkiye, Erzurum, Doğ Karadeniz Dağlari, 5 km N of Ispir, 40.4571667, 40.9586667, *Kučera, J. & Kolník, M. s.n.* (SAV, HUB), SAMN27735368, –, morph 7; TRO,

Türkiye, Erzincan, Otlukbeli Dağlari, Ahmetli, 39.8714647, 39.3403428, *Kučera, J. & Kolník, M. s.n.* (SAV, HUB), SAMN27735367, –, morph 20; USH, Georgia, Samegrelo and Zemo Svaneti, Ushguli, Enguri river valley, 42.936900, 43.036150, *Kolář, F. s.n.* (SAV), SAMN27735365, –, – ; YAL, Türkiye, Artvin, Şavşat, Yalnızçam Mountains, 1 km from Çam Pass to Şavşat, 41.205833, 42.491972, *Dönmez, A.A. & Yüzbaşıoğlu, S. s.n.* (SAV, HUB), SAMN33482354*, FCM 9, –.

Appendix 2. List of examined herbarium specimens of *Cardamine anatolica*. Herbarium acronyms follow Thiers (2023).

Cardamine anatolica Jar.Kučera et al.; Türkiye, Bursa province, Osmangazi distr., Bithynie: in monte Olympi (Keschischdagh) [Uludağ], 1750 m, 31 May 1899, *Bornmüller, J., Iter Anatolicum Tertium 1899 no. 4067* (P barcode P00557269 [image!]); Türkiye, Bursa province, Osmangazi distr., Banks of stream below hotel, Uludag, 1600-1800 m, 20 Jun 1956, *Moore, H.E. 7314* (E barcode E00380504 [image!]); Türkiye, Bursa province, Osmangazi distr., Uludağ, 1750 m, 27 Jul 1968, *Quezel & Pamukçuoğlu s.n.* (HUB 08053!); Türkiye, Bursa province, Osmangazi distr., Uludağ, 1750 m, 23 Aug 1971, *Baytop, A. & Baytop, T. s.n.* (ISTE 20873!); Türkiye, Kocaeli province, Kartepe distr., Kuzuyayla, 1350 m, 20 May 1973, *Özhatay, N. & Özhatay, E. s.n.* (ISTE 24611!); Türkiye, Kütahya province, Domaniç distr., İnegöl yolu, 1350 m, 11 May 1982, *Alpınar, K. s.n.* (ISTE 48663!); Türkiye, Kütahya province, Kütahya distr., Pehlivantaşı'na doğru tepenin güneyi, 1600 m, 8 Jun 1981, *Görk, G. G-72101* (EGE 18523!); Türkiye, Kütahya province, Simav distr., Kiçir to Akdag, forest road on N side of Akdag, stream in *Pinus* forest, c. 1700 m, 19 Jun 1965, *Coode, M.J.E. & Jones, B.M.G. 2741* (E barcode E00380497 [image!]).

Supplementary Table S1. Distribution records of *Cardamine anatolica*, *C. penzesii* and *C. uliginosa* in Türkiye based on herbarium records, field observations, literature and database data. In the column 'Source', acronyms of herbaria follow Index Herbariorum (Thiers, 2023).

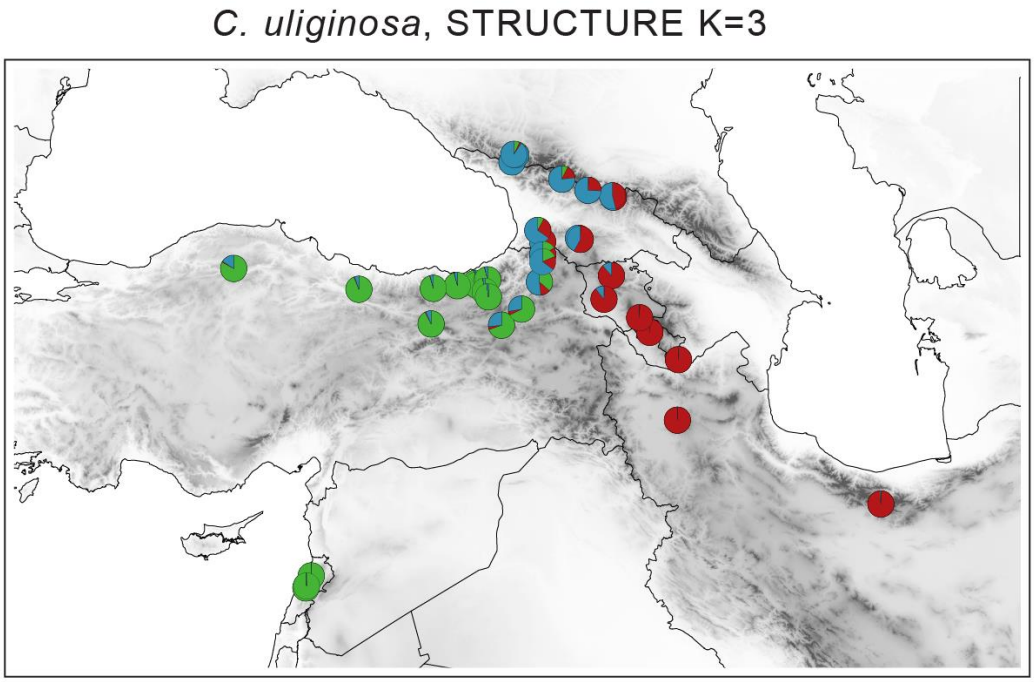
Due to the large dimension of this supplementary table, it was not inserted in the presented dissertation thesis and will be provided by the author upon request.

Supplementary Table S2. Summary of the relative genome size values of the studied *Cardamine* species and the results of Mann-Whitney pairwise comparison tests with Bonferroni correction of P values. Relative genome size is expressed as the sample/standard G1 peak ratio (standard: *Solanum pseudocapsicum*; 2C = 2.59 pg). n, number of measured individuals.

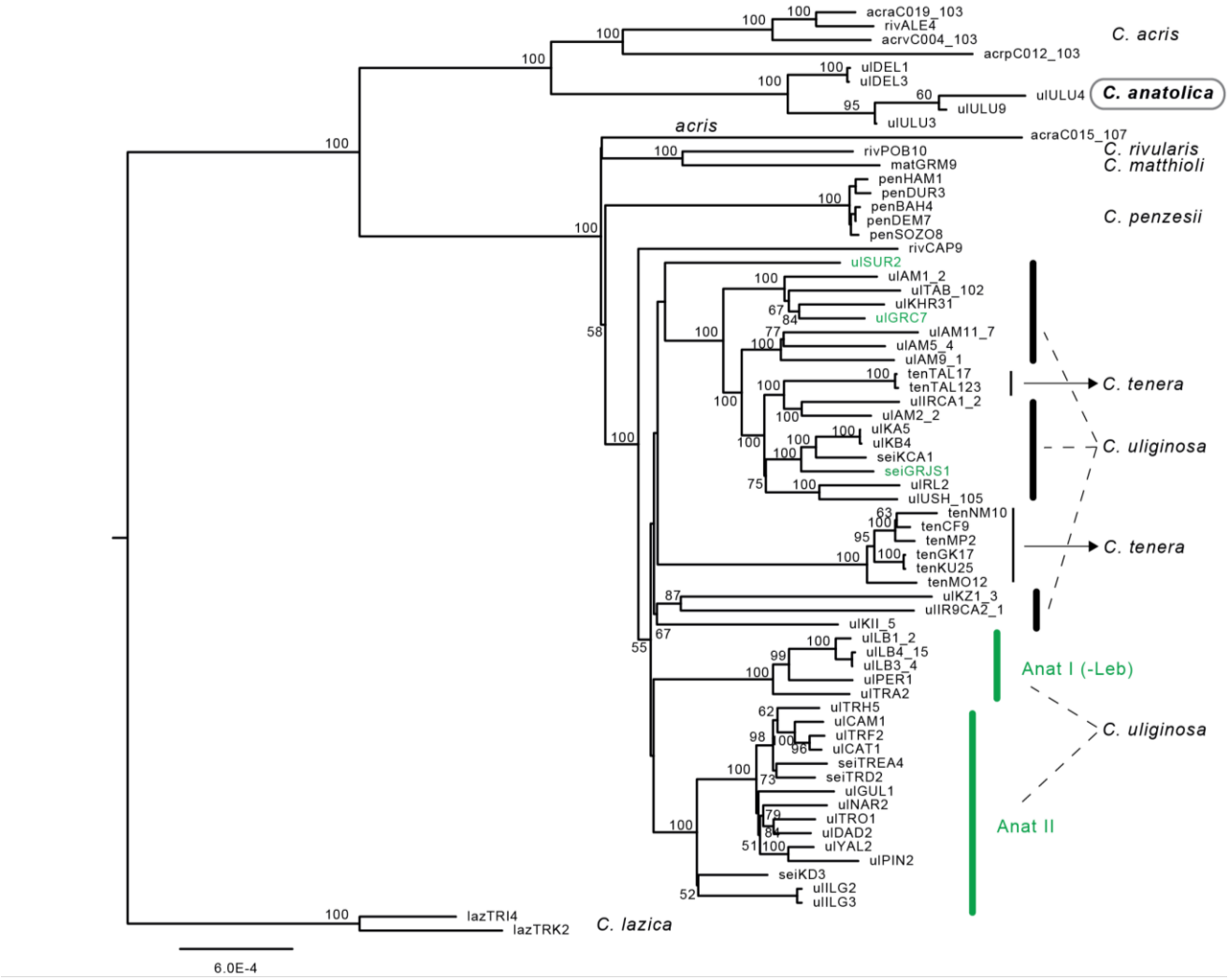
		Relative genome size (2C)		Mann-Whitney pairwise comparison (P-value)		
	n	min–max	mean ± SD	C. anatolica	C. penzesii	C. uliginosa
C. acris	222	0.401–0.490	0.438 ± 0.017	< 0.001	< 0.001	< 0.001
C. anatolica	64	0.390–0.412	0.402 ± 0.005	–	< 0.001	< 0.001
C. penzesii	36	0.342–0.385	0.367 ± 0.009	–	–	< 0.001
C. uliginosa	136	0.274–0.339	0.300 ± 0.012	–	–	–

Supplementary Table S3. Results of canonical discriminant analyses (CDA) of the studied *Cardamine* species. The total canonical structure coefficients are listed for CDA1-CDA7 (Fig. 4, Suppl. Fig. S3), expressing the correlations of the characters with the canonical axes (Can1-Can4). Characters with the highest values are highlighted in bold. Character abbreviations follow Table 1. CDA1: CDA with all studied species predefined as five groups, *C. acris*, *C. anatolica*, *C. penzesii*, *C. tenera*, and *C. uliginosa* (Fig. 4A); CDA2: CDA with three predefined groups, *C. anatolica*, *C. tenera*, and *C. uliginosa* (Fig. 4B); CDA3: CDA with three predefined groups, *C. anatolica*, *C. acris*, and *C. penzesii* (Fig. 4C). CDA4-7: each with two predefined groups: *C. anatolica* and *C. acris* (CDA4, Suppl. Fig. S3A); *C. anatolica* and *C. penzesii* (CDA5, Suppl. Fig. S3B); *C. anatolica* and *C. tenera* (CDA6, Suppl. Fig. S3C); *C. anatolica* and *C. uliginosa* (CDA7, Suppl. Fig. S3D).

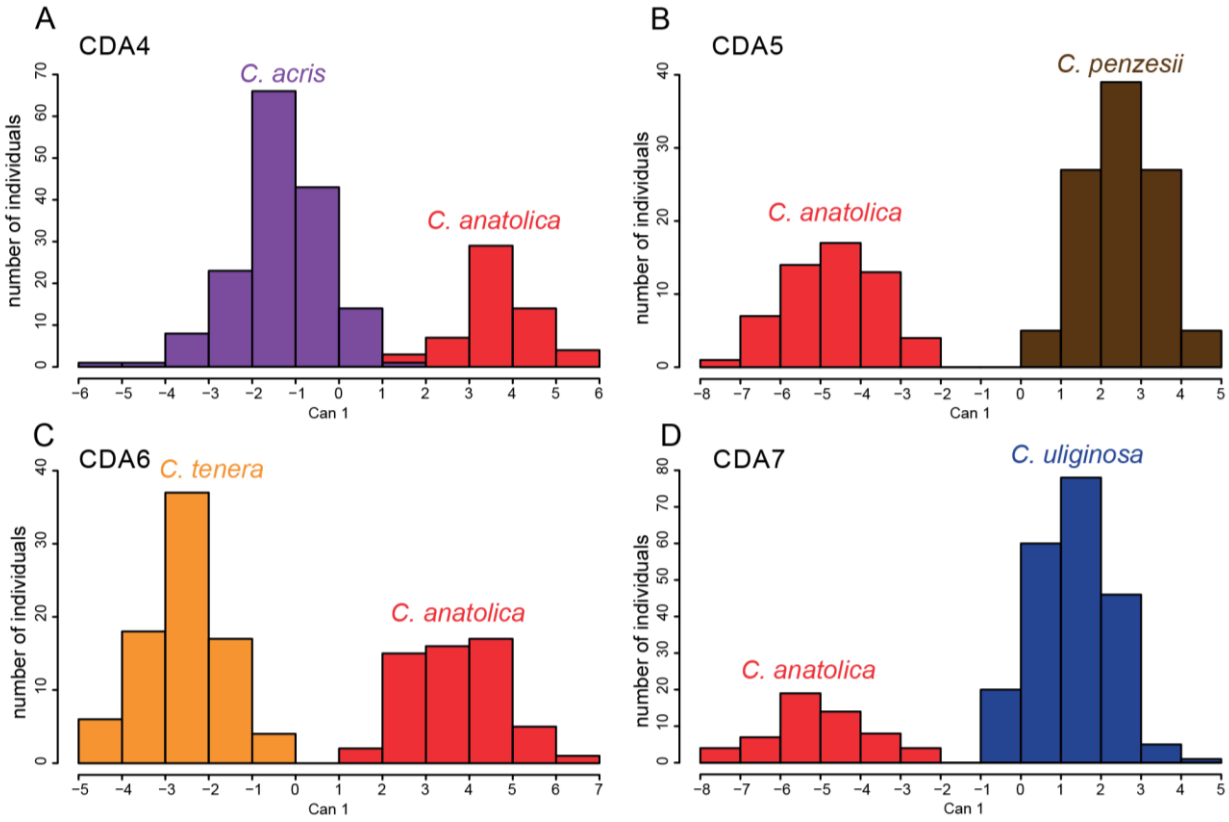
	CDA1			CDA2		CDA3		CDA4	CDA5	CDA6	CDA7
	Can1	Can2	Can3	Can1	Can2	Can1	Can2	Can1	Can1	Can1	Can1
Eigenvalues	4.7602	4.0994	2.2038	5.3115	2.5944	12.0453	2.7044	4.7886	11.9565	9.6175	6.5319
Eigenvalues as Percent	40.41	34.8	18.71	67.18	32.82	81.66	18.34	100	100	100	100
Cumulative Percentage of Eigenvalues	40.41	75.22	93.93	67.18	100	81.66	100	100	100	100	100
WS	0.27471023	0.086928665	-0.14385339	0.072534903	-0.037759906	-0.241837995	-0.27776888	-0.321945837	0.12630912	0.03140355	0.048501997
NB	0.27265811	-0.481779006	-0.241819662	-0.35755279	-0.325189072	0.277155069	-0.111293404	-0.064541424	0.41349171	0.46324505	-0.486631419
NB2	0.09726161	0.093530052	0.164240496	-0.066374336	0.229255165	-0.188906897	-0.174096342	-0.211853938	0.02467321	-0.2757644	0.076521979
NL	-0.07115122	-0.107100506	-0.296959359	0.177945784	-0.248820381	0.220958463	-0.016496732	0.026481538	0.23920326	0.40164771	0.029171596
NLR	-0.31699312	-0.068397617	0.006611389	0.167071003	0.053123598	0.379457093	0.050578108	0.15661895	0.2575479	-0.1524811	0.180385271
LC1	0.08837806	-0.06017453	0.075809899	-0.006349456	0.28493688	0.007492208	-0.314289597	-0.346631007	0.35763553	-0.48793736	0.147039452
LC2	0.1343257	-0.129921341	0.352104372	-0.285563313	0.470409437	0.021165906	-0.223827881	-0.230914016	0.27447382	-0.52965355	-0.010111457
NFB	-0.11097786	-0.330723218	-0.297718326	0.133382399	-0.220764942	0.499865276	-0.050975411	0.049912639	0.5389419	0.38980579	0.00323096
NFS	-0.44801393	0.149247803	-0.080702942	0.464299896	0.093132368	0.261016494	-0.205764316	-0.167160573	0.45901298	-0.35873639	0.460488027
NTL2	0.69031438	0.238839565	0.314178407	-0.66477757	0.062326735	-0.80845846	0.147300898	-0.004854896	-0.837678	0.13018232	-0.616768833
NTL3	0.20596504	0.106501682	0.207632905	-0.198884485	0.193291701	-0.286369756	-0.07142049	-0.126654027	-0.21733204	-0.21034039	-0.083415361
RHIZ	-0.31912828	0.590174851	0.135571093	0.499960028	0.234108162	-0.405200415	-0.088855777	-0.169978152	-0.35106796	-0.41576449	0.621408587
BASE	-0.29772202	0.308568882	0.129365445	0.358967893	0.277034319	-0.085365058	-0.146455045	-0.1818799	0.07753614	-0.44933222	0.493759088
STEM	-0.20380542	0.088146719	0.163156154	0.065890772	0.160008793	0.090460905	0.041723243	0.076269814	0.02947231	-0.25106843	0.158646798
LP	-0.0574216	-0.56464099	-0.024969865	-0.418161417	-0.252394871	0.53723867	0.386234064	0.546958303	0.10272755	0.44122723	-0.533357976
LS	-0.21593242	0.026621367	-0.502646727	0.337066622	-0.568019712	0.182574498	0.3190725	0.354577148	-0.21492529	0.71913362	0.003584556
LFL	-0.2991879	0.153388225	0.076495219	0.077806784	-0.106518932	0.05463284	0.469055311	0.474963302	-0.47431199	0.12914019	0.014594564
COL	0.08166186	0.537609747	-0.133888149	0.460044391	0.090721763	-0.603427278	-0.50965915	-0.665952949	-0.03635448	-0.35747105	0.480228019
NL/LSL	-0.42511708	0.101117239	-0.105051062	0.306957022	-0.158980857	0.284627584	0.322350675	0.434545772	-0.08772016	0.2074446	0.194006882
LC2/LSL	-0.18680103	-0.018046444	0.694487722	-0.260619251	0.647237314	0.18519589	0.007835269	0.044482428	0.14777428	-0.68157304	0.156469849
LC3/LSL	-0.18214771	-0.026016513	0.712209284	-0.271665614	0.655840073	0.212484809	0.003696955	0.043738643	0.20532974	-0.69276233	0.156472872
LC3/LC2	-0.19891791	-0.068952162	0.328249045	-0.091108755	0.357342097	0.233227283	0.022772092	0.076725857	0.18167752	-0.51039527	0.114849831
WTB/LTB	0.29611585	0.312725556	0.012224845	-0.121198128	-0.132045665	-0.603801017	0.114927828	0.004019958	-0.56104696	0.23664815	-0.184071143
WLB/LLB	0.63505007	0.282216258	-0.081422448	-0.31800506	-0.269029456	-0.766337242	0.000636634	-0.226347398	-0.68759302	0.49149666	-0.427977691
LLB/LTB	-0.7381956	-0.15545992	-0.156340008	0.602676533	0.042959934	0.798801988	-0.045960703	0.183733034	0.72435428	-0.29567325	0.585398088
WLB/WTB	-0.48910097	-0.155677205	-0.203777691	0.425060403	-0.033420413	0.610621321	-0.095663664	0.026097658	0.56277374	-0.08487683	0.388738506
WTS/LTS	0.68129615	0.27523372	0.068592646	-0.484630289	-0.167179784	-0.810885219	0.092568369	-0.104267436	-0.7733545	0.45886771	-0.513943696
WLS/LLS	0.57024671	0.33204281	0.007810212	-0.23545679	0.00118532	-0.673372637	-0.216104445	-0.423285527	-0.52662736	0.06200244	-0.225558667
LLS/LTS	-0.67777951	-0.002623267	-0.275312407	0.696113269	-0.073048057	0.615669645	-0.106314248	0.011191287	0.63619006	-0.15600789	0.617326664
WLS/WTS	-0.56562277	0.05129671	-0.255357954	0.609101509	-0.001894509	0.558449248	-0.311453066	-0.447017497	0.62038419	-0.49636037	0.537291774
WP/LP	-0.47019765	0.104703949	0.185507344	0.444247134	0.535545602	0.29460878	-0.521718314	-0.544902989	0.71854449	-0.79105089	0.703292694
LFS/LFL	-0.24055421	0.124603308	0.112984438	0.151194931	0.144275784	0.057054964	0.018434437	0.034513234	0.02743496	-0.27137953	0.213200587



Supplementary Fig. S1. Distribution of the analyzed accessions of *Cardamine uliginosa* and their assignment to Bayesian STRUCTURE clusters at K = 3. The STRUCTURE analysis was based on SNP data of 957 target nuclear genes obtained by Hyb-Seq.



Supplementary Fig. S2. Maximum-likelihood tree (RAxML-NG) of the complete plastome data (LSC, SSC, IRb) with bootstrap values shown above the main branches. The position of *Cardamine anatolica* described here as a new species is highlighted. Clades of *C. uliginosa* are marked with thick horizontal bars; those referring to Anatolian accessions are in green and labelled (Anat I, Anat II). Three accessions in green letters were placed within Anatolian clades in the nuclear ML tree, while here they cluster within the Caucasian clades.



Supplementary Fig. S3. Morphological differentiation of *Cardamine anatolica* from the closest relatives explored with canonical discriminant analyses CDA4-CDA7, each based on 32 morphological characters and individual plants, with two predefined groups. Values of total canonical structure are listed in Table S3.

5. CONCLUSION

In the first study focusing predominantly on the European part of species diversity (chapter 4.1), we got initial insight about the phylogenetic relationships among species from the Balkan Peninsula and SW Asia, supporting the idea of biogeographic links between these two regions. Using integrative approach involving Hyb-Seq, we resolved the evolutionary history of allotetraploid *C. barbaraeoides*, a stenoendemic species of the Southern Pindos Mts. in Greece. We discussed the challenges of phylogenomic analyses of polyploid genomes and applied a novel approach of allele sorting, by which we elucidated the parental lineages that led to origin of *C. barbaraeoides* (Fig. 3, 6). In this regard, we explained and emphasized the evolutionary role of the Balkan mountain ranges, which were found to serve both as melting pots mediating species contacts and polyploid speciation, as well as refugia sheltering the species during turbulent Pleistocene climate changes.

In the second study (chapter 4.2), major attention was paid to the SW Asian taxa. We resolved phylogenetic relationships among them, providing a taxonomic re-evaluation of this previously puzzling groups of species. We found that five diploid species should be recognized in the region: *C. lazica*, *C. penzesii*, *C. tenera*, *C. uliginosa*, and one western Anatolian entity representing newly discovered, in that time yet undescribed species (within the first two studies provisionally referred to as *C.* cf. *uliginosa* from Uludağ). In the process, we dismissed the autonomous taxonomic status of populations formerly referred to as *C. seidlitziana*, which were found to be part of the genetically and morphologically polymorphic *C. uliginosa* (Fig. 4). Rejecting the initial hypothesis that all SW Asian taxa of the studied groups are exclusively diploid, we discovered that another taxon *C. wiedemanniana* is, in fact, an allotetraploid. The name *C. wiedemanniana* has been occasionally used in synonymy to *C. lazica*, however, it was here found to represent an independent taxon stenoendemic to the Southern Anatolia. We resolved its origin (Fig. 6) using the approach introduced in the previous study. Except the case of *C. wiedemanniana*, we found no significant evidence of recent or historic interspecific hybridization events among the diploids. Corresponding to the spatially distributed genetic variability observed in multiple species, we proposed that speciation in this area was predominantly driven by geographic separation and ecological divergence. We discussed the roles of individual ecological and geographic features within Anatolia-Caucasus region which may have acted as gene flow barriers. On the other hand, we confirmed that the terrain topography and proximity of the seas both contributed to establishment of several important glacial refugia within SW Asia.

The third study (chapter 4.3) elucidated the significance of Anatolia and its mountain ranges in evolution of hygrophytic *Cardamine* species, additionally highlighting the evidences of Balkan-Anatolian biogeographic links. For this purpose, we used and thoroughly explored the W Anatolian entity which was in previous two studies referred to as *C.* cf. *uliginosa*. We described it as a new species *C. anatolica* (Fig. 5), having found that it was well delimited chorologically, genetically, as well as morphologically. Moreover, we confirmed the previous suggestions that this species is a closest relative to the Balkan *C. acris* (Fig. 6). We presumed that these two species originated as the result of vicariance scenario with the common ancestor being historically distributed in the larger area comprising both the Balkans and Anatolia. Separation and divergence of the lineages may have been promoted by biogeographic factors that are known to have recurrently established and halted the contact between these two regions. Additionally, we explored the genetic diversity of polymorphic *C. uliginosa* within Türkiye in more details. This allowed us to reveal the role of Anatolian mountains as important biogeographic elements promoting long-term persistence in multiple refugia, accumulation of genetic diversity, speciation, as well as admixture among intraspecific lineages.

In summary, the results presented in this thesis consistently showed that all studied taxa are part of two major clades, of which both comprise species from both regions of the Balkan Peninsula and SW Asia. While the first clade corresponds to the traditionally recognized *C. amara* species complex, the second clade further splits into three lineages, which are clearly not attributable to the traditional classification of the taxa into *C. pratensis* and *C. raphanifolia* species complexes (eventually including the *C. tenera* group if SW Asian taxa were taken into consideration). The discordance between traditional classification and the observed phylogenetic patterns are demonstrated in Fig. 6. At the beginning, it was presumed that the use of high-throughput phylogenomic approach in combination with wide selection of other methods will be able to reveal greater species diversity in previously understudied SW Asia, as such phenomenon has been observed in European Mediterranean and in other plant species groups. However, it applied to this case only partially – we discovered a new species (*C. anatolica*) and elucidated the independent position of another one (*C. wiedemanniana*), on the other hand, we rejected the existence of *C. seidlitziana* as a separate taxonomic entity, merging its populations with *C. uliginosa*. While it is clear that the evolution of the species from the Balkan Peninsula and SW Asia is interconnected, our findings do not evidence that the European species would be majorly derived from the ancestors originally inhabiting the Anatolia-Caucasus region (‘out of Anatolia’) nor vice-versa. Instead, we presumed that the

extant species may represent the outcome of range fragmentation of common ancestors that may have been distributed in larger areas, eventually covering respective parts of both continents. The sister position of the species *C. acris* and *C. anatolica* serves as a good demonstration of this theory at the lineage-level and provides a direct evidence of the Balkan-Anatolian biogeographic links. However, it seems that these two species have not played a significant role in the evolution of overall species diversity in Europe nor Asia, as their distribution remained restricted to relatively small parts of the continents on both sides.

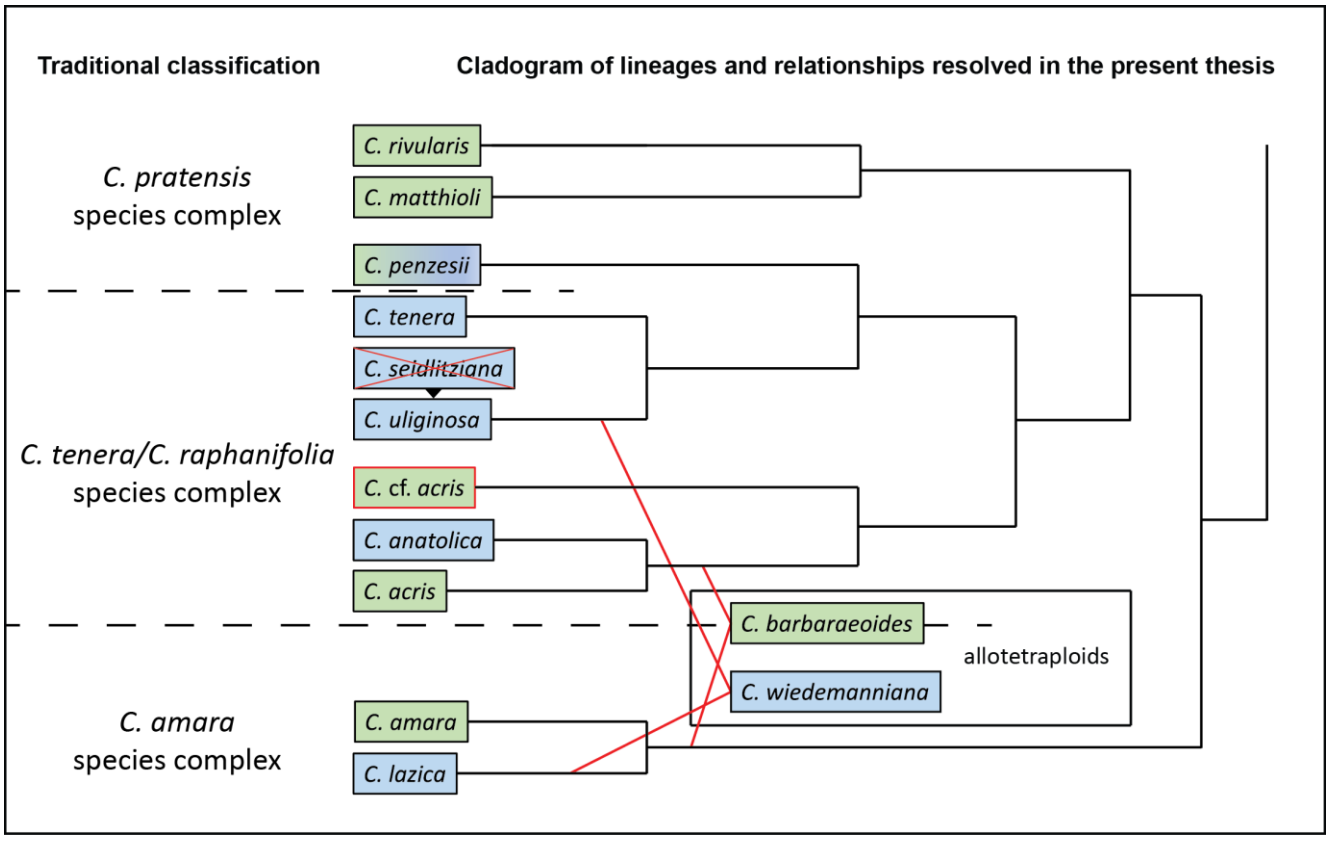


Fig. 6. The schematic summary showing the traditional classification (in the left part) and actual relationships of the studied species from the Balkan Peninsula (green) and Southwestern Asia (blue) resolved in the present thesis, demonstrated by the cladogram (in the right part). Recognition of *C. seidlitziana* (crossed) as a separate entity was rejected. Red lines indicate the lineages that were identified as parental in the origin of the two allotetraploid species. The taxonomically untreated entity in the red rectangle has not been studied in this thesis, as it was discovered only very recently and will be the subject of the following study (Kantor et al., in prep.).

Contrary to the initial hypothesis that the evolutionary history of SW Asian taxa was significantly and recurrently affected by introgression, we found very little evidence of historic or recent interspecific gene flow. Apparently, hybridization associated with polyploidization played a role in the origin of *C. wiedemanniana*, but the other species seem to be reproductively

well isolated, which is presumably secured mainly by ecological and geographic barriers. These findings suggest that the problems with inference of phylogenetic relationships within SW Asian taxa in previous molecular studies were not caused by extensive reticulation, but rather by their recent and relatively rapid diversification, often resulting in high intraspecific variability (especially in *C. uliginosa*). Similarly, the role of polyploidy as a speciation mechanism was found to be very minor in the studied groups both in SW Asia and the Balkans, which is in a contrast with situation in the rest of Europe, mainly in Southern Mediterranean. Indeed, we found one polyploid taxon in Türkiye (*C. wiedemannana*), rejecting the initial theory that there is none, and one polyploid taxon in the Balkan Peninsula (*C. barbaraeoides*). Noteworthily, both of these represent stenoendemic species with only a few, small and closely distributed populations. Due to the fact that they apparently were not able to colonize larger areas and different or wider selection of niches, they were suspected to have no significant impact on the evolution of the target species groups. In the Balkans, however, the recent ongoing studies of *C. acris* (Kantor et al., in preparation) revealed that there are few more, previously unknown polyploid populations in the Balkans that belong to the *C. acris* group (incl. *C. anatolica*). It is clear that these represent independent entities that have not been yet taxonomically treated (see *C.* cf. *acris* in Fig. 6). That said, the number of recognized species within the studied groups and proportion of polyploids is comparable between SW Asia and the Balkan Peninsula: the former comprises five diploids (*C. anatolica*, *C. lazica*, *C. penzesii*, *C. tenera*, *C. uliginosa*) and one tetraploid (*C. wiedemanniana*), while in the latter, these lineages are represented also by five diploid species (*C. acris* including three subspecies, *C. amara* subsp. *balcanica* Marhold, Ančev & Kit Tan, *C. matthioli*, *C. penzesii*, *C. rivularis*), one tetraploid (*C. barbaraeoides*), and another undescribed polyploid entity within *C. acris* group. However, when other parts of Europe are taken into consideration, the species diversity of the target groups (comprising about 30 taxa) and proportion of polyploids are much higher than those observed in SW Asia. We hypothesize that this disproportion may be caused by the fact that Europe (and especially Mediterranean) might have provided more abundant and more diverse habitats that could serve as refugia, migration routes and melting pots for hygrophytic species, contributing to both persistence and contacts among them, possibly promoting higher speciation rates. In contrast, most of the area of Southwestern Asia is covered by the Irano-Turanian phytogeographic region, which is characterized by dry climate being suitable mainly for xerophytes. Consequently, most of the studied *Cardamine* species were found to be limited there to the humid regions adjacent to the seas. The common phenomenon for both European Mediterranean and Southwestern Asia is, that the mountains ranges in both regions possess a

major evolutionary significance, harbouring major refugia and biodiversity hotspots. We found this to be applicable also in the case of studied *Cardamine*, as we explored the role of mountains in various evolutionary scenarios including: survival in multiple glacial refugia, associated with intraspecific diversification resulting in spatially distributed genetic variability (mainly in *C. lazica*, *C. tenera*), but also establishment of admixture among different intraspecific lineages (in *C. uliginosa*); limitation of the interspecific gene flow caused by natural topographic barriers (suggested to have contributed to the origin of *C. anatolica*); and on the other hand, promotion of allopolyploid speciation (in *C. barbaraeoides*).

5.1 Overall significance of the results of the thesis

The results presented in the three studies included in this thesis provide novel insights into multiple topics covering diverse fields of plant science and molecular phylogenetics.

In the field of plant evolution, this thesis provides new information about the evolutionary processes that shaped plant species diversity in regions which have been historically largely understudied, even though they are known to represent significant biodiversity hotspots and harbour important refugia and intercontinental colonization routes. Additionally, the groups focused here comprised species with very specific ecological demands, as all of them were hygrophytes, most of them being limited to the regions of montane to subalpine belts. For instance, this specific combination of environmental requirements may be especially challenging for plant species in SW Asia, which is known to be providing suitable conditions mainly for xerophytes. Indeed, even less is known about the evolution of hygrophytes in the regions with predominantly arid climate. In this regard, our findings provide valuable information about the roles of specific ecological factors and biogeographic features within the Balkan Peninsula and Caucasus-Anatolia region, which may have played an important role also in evolution of other organisms. Moreover, knowledge of evolutionary patterns and distribution of the hygrophytic species in rather limiting environment may be helpful for understanding and better prediction of impacts of climate changes on the evolution of hygrophytic and montane flora in general. As most of the studied species represent entities with rather restricted distribution, studying them also provides new insights into the patterns of endemism in the focus regions.

Our molecular data-based interpretations demonstrate the power of NGS based methods (in this case Hyb-Seq) for resolving complex evolutionary histories of groups of closely related

species, which, in this case, involved recent speciation, remarkable intraspecific diversification, as well as events of hybridization and polyploidization. Aspiring to reconstruct the evolutionary history of the allopolyploid species, we reviewed the available methods of utilizing NGS data for resolving the origin of polyploids, and we introduced allele sorting, a new generally applicable tool for processing of data from polyploid genomes.

To test the addressed hypotheses, we used diverse combinations of karyological, molecular, and morphometric methods, along with the cytogenetic technique GISH and the ecological niche analysis in individual studies. This demonstrated the significance and utility of integrative approach for getting complementary answers for the stated questions, which may provide irreplaceable verification of credibility of the evolutionary interpretations. In this case, it enabled us to elucidate several problems which remained unresolved for a long time since previous studies.

In the end, we provided taxonomical revision of the studied species groups in the Balkan Peninsula and SW Asia. This involved description of a new species from W Anatolia, as well as an update of circumscription of established species. We gathered and revised distributional data of several species occurring in Türkiye and in the end, compiled the identification key to the SW Asian species that were found to be often misinterpreted. These information may be valuable for local botanists or field workers and will be applicable during preparation of the floras of the individual countries. As many of the studied species represent endemics, knowledge of their distribution and ecological demands may assist in effective planning of nature protection management.

6. REFERENCES

Akiyama, R., Sun, J., Hatakeyama, M., Lischer, H.E.L., Briskine, R.V., Hay, A., Gan, X., Tsiantis, M., Kudoh, H., Kanaoka, M.M., Sese, J., Shimizu, K.K. & Shimizu-Inatsugi, R. (2021). Fine-scale empirical data on niche divergence and homeolog expression patterns in an allopolyploid and its diploid progenitor species. *New Phytol.* 229, 3587–3601. doi: 10.1111/nph.17101

Al-Shehbaz, I.A. (1988). The genera of Arabideae (Cruciferae; Brassicaceae) in the southeastern United States. *J. Arnold Arboretum* 69, 85–166. doi: 10.5962/bhl.part.2391

Al-Shehbaz, I.A., Beilstein, M.A. & Kellogg, E. A. (2006). Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Syst. Evol*. 259, 89–120. doi: 10.1007/s00606-006-0415-z

Al-Shehbaz, I.A., Marhold, K. & Lihová, J. (2010). *Cardamine* Linnaeus. In: Flora of North America Editorial Committee (Eds.), *Flora of North America: North of Mexico, Volume 7* (pp. 464–484). New York & Oxford: Oxford University Press.

Ali, T., Muñoz-Fuentes, V., Buch, A., Çelik, A., Dutbayev, A., Gabrielyan, I., Glynou, K., Kachour, L., Khaliq, I., Kitner, M., Nigrelli, L., Ploch, S., Runge, F., Solovyeva, I., Schmuker, A., Vakhrusheva, L., Xia, X., Maciá-Vicente, J., Nowak, C. & Thines, M. (2019). Out of Transcaucasia: Origin of Western and Central Palearctic populations of *Microthlaspi perfoliatum*. *Flora* 253, 127–141. doi: 10.1016/j.flora.2019.02.012

Álvarez, I. & Wendel, J.F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434. doi: 10.1016/s1055-7903(03)00208-2

Ančev, M., Yurukova-Grancharova, P., Ignatova, P., Goranova, V., Stoyanov, S., Yankova-Tsvetkova, E. & Neykov, N. (2013). *Cardamine* ×*rhodopaea* (Brassicaceae), a triploid hybrid from the West Rhodope Mts: Morphology, distribution, relationships and origin. *Phytol. Balcan.* 19, 323–338.

Ansell, S.W., Stenøien, H.K., Grundmann, M., Russell, S.J., Koch, M.A., Schneider, H. & Vogel, J.C. (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. doi: 10.1093/aob/mcr134

Aqafarini, A., Lotfi, M., Norouzi, M. & Karimzadeh, G. (2019): Induction of tetraploidy in garden cress: morphological and cytological changes. *Plant Cell Tiss. Org.* 137, 627–635. doi: 10.1007/s11240-019-01596-5

Arrigo, N. & Barker, M.S. (2012). Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin. Plant Biol.* 15, 140–146. doi: 10.1016/j.pbi.2012.03.010

Baduel, P., Bray, S., Vallejo-Marin, M., Kolář, F. & Yant, L. (2018). The 'polyploidhop': shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front. Ecol. Evol.* 6, 117. doi.org/10.3389/fevo.2018.00117

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3, e3376. doi: 10.1371/journal.pone.0003376

Bastkowski, S., Mapleson, D., Spillner, A., Wu, T., Balvociute, M. & Moulton, V. (2018). SPECTRE: a suite of phylogenetic tools for reticulate evolution. *Bioinformatics* 34, 1056–1057. doi: 10.1093/bioinformatics/btx740

Beilstein, M.A., Al-Shehbaz, I.A., Mathews, S. & Kellogg, E.A. (2008). Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: tribes and trichomes revisited. *Am. J. Bot.* 95, 1307–1327. doi: 10.3732/ajb.0800065

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*. *P. Natl. Acad. Sci. USA* 107, 18724–18728. doi: 10.1073/pnas.0909766107

Bilgin, R. (2011). Back to the suture: the distribution of intraspecific genetic diversity in and around Anatolia. *Int. J. Mol. Sci.* 12, 4080–4103. doi: 10.3390/ijms12064080

Bleeker, W., Franzke, A., Pollmann, K., Brown, A.H.D. & Hurka, H. (2002). Phylogeny and biogeography of southern hemisphere high-mountain *Cardamine* species. *Aust. Syst. Bot.* 15, 575–581. doi: 10.1071/SB01026

Bolger, A.M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bomblies, K. (2020). When everything changes at once: finding a new normal after genome duplication. *Proc. Royal Soc. B* 287, 20202154. doi: 10.1098/rspb.2020.2154

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ.* 4, e1660. doi: 10.7717/peerj.1660

Braithwaite, M. (2003). A crucifer with 'no flowers to speak of' - cleistogamy and prolifery in *Cardamine corymbosa* (New Zealand bitter-cress). *BSBI News* 92, 14–16.

Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., Petitpierre, B., Pellissier, L., Yoccoz, N.G., Thuiller, W., Fortin, M.J., Randin, C., Zimmermann, N.E., Graham, C.H. & Guisan, A. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Glob. Ecol. Biogeogr.* 21, 481–497. doi: 10.1111/j.1466-8238.2011.00698.x

Brukhin, V., Osadtchiy, J.V., Florez-Rueda, A.M., Smetanin, D., Bakin, E., Nobre, M.S. & Grossniklaus, U. (2019). The *Boechera* genus as a resource for apomixis research. *Front. Plant Sci.* 10, 392. doi: 10.3389/fpls.2019.00392

Cañadas, E.M., Fenu, G., Peñas, J., Lorite, J., Mattana, E. & Bacchetta, G. (2014). Hotspots within hotspots: Endemic plant richness, environmental drivers, and implications for conservation. *Biol. Conserv.* 170, 282–291. doi: 10.1016/j.biocon.2013.12.007

Carlsen, T., Bleeker, W., Hurka, H., Elven, R. & Brochmann, C. (2009). Biogeography and phylogeny of *Cardamine* (Brassicaceae). *Ann. Mo. Bot. Gard.* 96, 215–236. doi: 10.2307/40389931

Carter, K.A., Liston, A., Bassil, N.V., Alice, L.A., Bushakra, J.M., Sutherland, B.L., Mockler, T.C., Bryant, D.W. & Hummer, K. E. (2019). Target capture sequencing unravels *Rubus* evolution. *Front. Plant Sci.* 10, 1615. doi: 10.3389/fpls.2019.01615

Chamala, S., García, N., Godden, G.T., Krishnakumar, V., Jordon-Thaden, I.E., De Smet, R., Barbazuk, B., Soltis, D.E. & Soltis, P.S. (2015). MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. *Appl. Plant Sci.* 3, 1400115. doi: 10.3732/apps.1400115

Chansler, M.T., Ferguson, C.J., Fehlberg, S.D. & Prather, L.A. (2016): The role of polyploidy in shaping morphological diversity in natural populations of *Phlox amabilis*. *Am. J. Bot.* 103, 1546–1558. doi: 10.3732/ajb.1600183

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y. & Leon, C. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* 5, e8613. doi: 10.1371/journal.pone.0008613

Clement, M., Posada, D. & Crandall, K.A. (2000). TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1659. doi: 10.1046/j.1365-294x.2000.01020.x

Clo, J. & Kolář, F. (2021). Short- and long-term consequences of genome doubling: a meta-analysis. *Am. J. Bot.* 108, 2315–2322. doi: 10.1002/ajb2.1759

Comes, H.P. (2004). The Mediterranean region - a hotspot for plant biogeographic research. *New Phytol.* 164, 11–14. doi: 10.1111/j.1469-8137.2004.01194.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). *Mol. Biol. Evol.* 27, 55–71. doi: 10.1093/molbev/msp202

Cox, B. (2001). The biogeographic regions reconsidered. *J. Biogeogr.* 28, 511–523. doi: 10.1046/j.1365-2699.2001.00566.x

Cullen, J. (1965). *Cardamine* L. In P. H. Davis, J. Cullen & M. J. E. Coode (Eds.), *Flora of Turkey and the East Aegean islands, Volume 1* (pp. 438–444). Edinburgh: University Press.

De Storme, N. & Geelen, D. (2013). Sexual polyploidization in plants – cytological mechanisms and molecular regulation. *New Phytol.* 198, 670–684. doi: 10.1111/nph.12184

Denk, T., Frotzler, N. & Davitashvili, N. (2001). Vegetational patterns and distribution of relict taxa in humid temperate forests and wetlands of Georgia (Transcaucasia). *Biol. J. Linn. Soc.* 72, 287–332. doi: 10.1006/bijl.2000.0502

Djamali, M., Brewer, S., Breckle, S.W. & Jackson, S.T. (2012). Climatic determinism in phytogeographic regionalization: A test from the Irano-Turanian region, SW and Central Asia. *Flora* 207, 237–249. doi: 10.1016/j.flora.2012.01.009

Dodsworth, S., Pokorny, L., Johnson, M.G., Kim, J.T., Maurin, O., Wickett, N.J., Forest, F. & Baker, W.J. (2019). Hyb-Seq for flowering plant systematics. *Trends Plant Sci.* 24, 887–891. doi: 10.1016/j.tplants.2019.07.011

Doležel, J., Greilhuber, J. & Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* 2, 2233–2244.

Dong, W., Liu, J., Yu, J., Wang, L. & Zhou, S. (2012). Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* 7, e35071. doi: 10.1371/journal.pone.0035071

Dönmez, A. & Yerli, S.V. (2018). Biodiversity in Turkey. In T. Pullaiah (Ed.), *Global Biodiversity 2, Selected countries in Europe* (pp. 397-442). Apple Academic Press.

Eastwood, W.J. (2004). East Mediterranean vegetation and climate change. In H. I. Griffiths, B. Kryštufek & J. M. Reed (Eds.), *Balkan Biodiversity* (pp. 25–48). Dordrecht: Springer. doi: 10.1007/978-1-4020-2854-0_3

Ekenäs, C., Heidari, N. & Andreasen, K. (2012). *Arnica* (Asteraceae) phylogeny revisited using *RPB*2: Complex patterns and multipled-paralogues. *Mol. Phylogenet. Evol.* 64, 261–270. doi: 10.1016/j.ympev.2012.02.032

Ekim, T. & Güner, A. (1986). The Anatolian Diagonal: fact or fiction? *Proc. Royal Soc. B* 89, 69–77. doi: 10.1017/S0269727000008915

Eriksson, J.S., de Sousa, F., Bertrand, Y.J.K., Antonelli, A., Oxelman, B. & Pfeil, B.E. (2018). Allele phasing is critical to revealing a shared allopolyploid origin of *Medicago arborea* and *M. strasseri* (Fabaceae). *BMC Ecol. Evol.* 18, 9. doi: 10.1186/s12862-018-1127-z

Fawcett, J. A., Maere, S. & Van de Peer, Y. (2009). Plants with doubled genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *P. Natl. Acad. Sci. USA* 106, 5737–5742. doi: 10.1073/pnas.0900906106

Fayvush, G. M. & Aleksanyan, A. S. (2020). The Transcaucasian Highlands. In J. Noroozi (Ed.), *Plant biogeography and vegetation of high mountains of Central and South-West Asia* (pp. 287–313). Springer.

Fick, S.E. & Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi: 10.1002/joc.5086

Fitzpatrick, B.M., Fordyce, J.A. & Gavrilets, S. (2009). Pattern, process and geographic modes of speciation. *J. Evol. Biol.* 22, 2342–2347. doi: 10.1111/j.1420-9101.2009.01833.x

Flagel, L.E. & Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* 183, 557–564. doi: 10.1111/j.1469-8137.2009.02923.x

Franzke, A., Pollmann, K., Bleeker, W., Kohrt, R. & Hurka, H. (1998). Molecular systematics of *Cardamine* and allied genera (Brassicaceae): ITS and non-coding chloroplast DNA. *Folia Geobot.* 33, 225–240. doi: 10.1007/BF03216204

Franzke, A., Lysák, M.A., Al-Shehbaz, I.A., Koch, M.A. & Mummenhoff, K. (2011). Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci.* 16, 108–116. doi: 10.1016/j.tplants.2010.11.005

Franzke, A., Koch, M.A. & Mummenhoff, K. (2016). Turnip time travels: age estimates in Brassicaceae. *Trends Plant Sci.* 21, 554–561. doi: 10.1016/j.tplants.2016.01.024

Freyman, W.A., Johnson, M.G. & Rothfels, C.J. (2023). homologizer: Phylogenetic phasing of gene copies into polyploid subgenomes. *Methods Ecol. Evol.* 14, 12301244. doi: 10.1111/2041-210X.14072

Godfree, R.C., Marshall, D.J., Young, A.G., Miller, C.H. & Mathews, S. (2017). Empirical evidence of fixed and homeostatic patterns of polyploid advantage in a keystone grass exposed to drought and heat stress. *R. Soc. Open Sci.* 4, 170934. doi: 10.1098/rsos.170934

Gömöry, D., Zhelev, P. & Brus, R. (2020). The Balkans: a genetic hotspot but not a universal colonization source for trees. *Plant Syst. Evol.* 306, 5. doi: 10.1007/s00606-020-01647-x

Goodwin, S., McPherson, J.D. & McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 331–351. doi: 10.1038/nrg.2016.49

Grünewald, S., Spillner, A., Bastkowski, S., Bögershausen, A. & Moulton, V. (2013). SuperQ: computing supernetworks from quartets. *IEEE ACM T. Comput. Bi.* 10, 151–160. doi: 10.1109/TCBB.2013.8

Gür, H. (2016). The Anatolian diagonal revisited: Testing the ecological basis of a biogeographic boundary. *Zool. Middle East* 62, 189–199. doi: 10.1080/09397140.2016.1226544

Halácsy, E. (1894). Beitrag zur Flora von Epirus. *Denkschriften der Kaiserlichen Akademie der Wissenschaften* 61, 217–268.

Hart, M.L., Forrest, L.L., Nicholls, J.A. & Kidner, C.A. (2016). Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* 32, 1081–1092. doi: 10.12705/655.9

Heenan, P.B. (2017). A taxonomic revision of *Cardamine* L. (Brassicaceae) in New Zealand. *Phytotaxa* 330, 1. doi: 10.11646/phytotaxa.330.1.1

Herrando-Moraira, S., the Cardueae Radiations Group, Calleja, J.A., Galbany-Casals, M., Garcia-Jacas, N., Liu, J.Q., López-Alvarado, J., López-Pujol, J., Mandel, J.R., Massó, S., Montes-Moreno, N., Roquet, C., Sáez, L., Sennikov, A., Susanna, A. & Vilatersana, R. (2019). Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with Hyb-Seq data: A new subtribal classification and a temporal diversification framework. *Mol. Phylogenet. Evol.* 137, 313–332. doi: 10.1016/j.ympev.2019.05.001

Hewitt, G.M. (2004). Genetic consequences of climatic oscillationsin the Quaternary. *Philos. T. Roy. Soc. B.* 359, 183–195. doi: 10.1098/rstb.2003.1388

Hewitt, G.M. (2011). Mediterranean peninsulas: The evolution of hotspots. In F. E. Zachos & J. C. Habel (Eds.), *Biodiversity Hotspots* (pp. 123–147). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-20992-5_7

Hohmann, N., Wolf, E. M., Lysák, M. A. & Koch, M. A. (2015). A time-calibrated roadmap of Brassicaceae species radiation and evolutionary history. *Plant Cell* 27, 2770–2784. doi: 10.1105/tpc.15.00482

Hsü, K.J. (1972). When the Mediterranean dried up. *Sci. Am.* 227, 27–36.

Huang, C.H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M.A., Al-Shehbaz, I., Edger, P.P., Pires, J.C., Tan, D.Y., Zhong, Y., & Ma, H. (2016). Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol. Biol. Evol.* 33, 394–412. doi: 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. *Ann. Bot.* 125, 29–47. doi: 10.1093/aob/mcz123

Huelsenbeck, J.P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

Husband, B.C. (2004). The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biol. J. Linn. Soc.* 82, 537–546. doi: 10.1111/j.1095-8312.2004.00339.x

Huson, D.H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030

Jabbour, F. & Renner, S.S. (2011). *Consolida* and *Aconitella* are an annual clade of *Delphinium* (Ranunculaceae) that diversified in the Mediterranean basin and the Irano-Turanian region. *Taxon 60*, 1029–1040. doi: 10.1002/tax.604007

Jalas, J. & Suominen, J. (1994). *Atlas Florae Europaeae 10*. Helsinki: The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo.

Jarvis, A., Reuter, H.I., Nelson, A. & Guevara, E. (2008). Hole-filled seamless SRTM data V4. International Centre for Tropical Agriculture (CIAT). Available at: http://srtm.csi.cgiar.org.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J. & dePamphilis, C.W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N.J.C. & Wickett, N.J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* 4, 1600016. doi: 10.3732/apps.1600016

Johnson, M.G., Pokorny, L., Dodsworth, S., Botigué, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L., Epitawalage, N., Forest, F., Kim J.T., Leebens-Mack, J.H., Leitch, I.J., Maurin, O., Soltis, D.E., Soltis, P.S., Wong, G.K., Baker, W.J. & Wickett, N.J. (2018). A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering. *Syst. Biol*. 68, 594–606. doi: 10.1093/sysbio/syy086

Jones, B.M.G. & Akeroyd, J.R. (1993). *Cardamine*. In T.G. Tutin, V.H. Heywood, N.A. Burges, D.H. Valentine, S.M. Walters & D.A. Webb (Eds.), *Flora Europaea 1, Psilotaceae to Platanaceae. 2nd edition* (pp. 346–352). Cambridge: Cambridge University Press.

Kamneva, O.K., Syring, J., Liston, A. & Rosenberg, N.A. (2017). Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. *BMC Evol. Biol.* 17, 180. doi: 10.1186/s12862-017-1019-7

Karbstein, K., Tomasello, S., Hodač, L., Wagner, N., Marinček, P., Barke, B.H., Paetzold, C. & Hörandl, E. (2022). Untying Gordian knots: unraveling reticulate polyploid plant evolution by genomic data using the large *Ranunculus auricomus* species complex. *New Phytol.* 235, 2081–2098. doi: 10.1111/nph.18284

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. doi: 10.1093/aob/mct165

Kates, H.R., Johnson, M.G., Gardner, E.M., Zerega, N.J.C. & Wickett, N.J. (2018). Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus*. *Am. J. Bot.* 105, 404–416. doi: 10.1002/ajb2.1068

Katoh, K. & Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kaya, S. & Çiplak, B. (2017). Phylogeography and taxonomy of the *Psorodonotus caucasicus* (Orthoptera, Tettigoniidae) group: independent double invasion of the Balkans from the Caucasus. *Syst. Entomol.* 42, 118–133. doi: 10.1111/syen.12197

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kellogg, E.A. (2016). Has the connection between polyploidy and diversification actually been tested? *Curr. Opin. Plant Biol.* 30, 25–32. doi: 10.1016/j.pbi.2016.01.002

Kerey, I.E., Meriç, E., Tunoğlu, C., Kelling, G., Brenner, R.L. & Doğan, A.U. (2004). Black Sea–Marmara Sea Quaternary connections: new data from the Bosphorus, İstanbul, Turkey. *Palaeogeogr. Palaeocl.* 204, 277–295. doi: 10.1016/S0031-0182(03)00731-4

Khatri, K.S. (1988). Studies in the Caucasian species of *Cardamine* L. (Cruciferae). *Phyton* 28, 55–80.

Kimata, M. (1983). Comparative studies on the reproductive systems of *Cardamine flexuosa*, *C. impatiens*, *C. scutata*, and *C. lyrata*, Cruciferae. *Bot. Mag.* 96, 299–312. doi: 10.1007/BF02488175

Klecka, W.R. (1980). *Discriminant analysis (Quantitative applications in the social sciences), no. 19*. Beverly Hills: Sage Publications, Inc.

Koch, M. (2003). Molecular phylogenetics, evolution and population biology in Brassicaceae. In A.K. Sharma & A. Sharma (Eds.), *Plant genome, biodiversity and evolution* (pp. 1–35). Enfield, New Hampshire: Science Publishers.

Koch, M., Karl, R. & German, D.A. (2017). Underexplored biodiversity of Eastern Mediterranean biota: systematics andevolutionary history of the genus *Aubrieta* (Brassicaceae). *Ann. Bot.* 119, 39–57. doi: 10.1093/aob/mcw204

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. doi: 10.1016/j.tplants.2017.10.005

Kolář, F., Čertner, M., Suda, J., Schönswetter, P. & Husband, B.C. (2017). Mixed-ploidy species: progress and opportunities in polyploid research. *Trends Plant Sci.* 22, 1041–1055. doi: 10.1016/j.tplants.2017.09.011

Korkmaz, E.M., Lunt, D.H., Çıplak, B., Değerli, N. & Başıbüyük, H.H. (2014). The contribution of Anatolia to European phylogeography: the centre of origin of the meadow grasshopper, *Chorthippus parallelus*. *J. Biogeogr.* 41, 1793–1805. doi: 10.1111/jbi.12332

Kougioumoutzis, K., Kokkoris, I.P., Panitsa, M., Kallimanis, A., Strid, A. & Dimopoulos, P. (2021). Plant endemism centres and biodiversity hotspots in Greece. *Biology* 10, 72. doi: 10.3390/biology10020072

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305

Krak, K., Caklová, P., Chrtek, J. & Fehrer, J. (2013). Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (*Hieracium*, Asteraceae). *Heredity* 110, 138–151. doi: 10.1038/hdy.2012.100

Krebs, P., Conedera, M., Pradella, M., Torriani, D., Felber, M. & Tinner, W. (2004). Quaternary refugia of the sweet chestnut (*Castanea sativa* Mill.): an extended palynological approach. *Veg. Hist. Archaeobot.* 13, 145–160. doi: 10.1007/s00334-004-0041-z

Krijgsman, W., Capella, W., Simon, D., Hilgen, F.J., Kouwenhoven, T.J., Meijer, P.T., Sierro, F.J., Tulbure, M.A., Berg, B.C.J., Schee, M. & Flecker, R. (2018). The Gibraltar corridor: watergate of the Messinian Salinity Crisis. *Mar. Geol.* 403, 238–246. doi: 10.1016/j.margeo.2018.06.008

Krzanowski, W.J. (1990). *Principles of multivariate analysis: A user 's perspective*. OUP Oxford.

Kučera, J., Valko, I. & Marhold, K. (2005). On-line database of the chromosome numbers of the genus *Cardamine* (Brassicaceae). *Biologia (Bratislava)* 60, 473–476.

Kučera, J., Lihová, J. & Marhold, K. (2006). Taxonomy and phylogeography of *Cardamine impatiens* and *C. pectinata* (Brassicaceae). *Bot. J. Linn. Soc.* 152, 169–195. doi: 10.1111/j.1095-8339.2006.00559.x

Kučera, J., Marhold, K. & Lihová, J. (2010). *Cardamine maritima* group (Brassicaceae) in the amphi-Adriatic area: A hotspot of species diversity revealed by DNA sequences and morphological variation. *Taxon* 59, 148–164. doi: 10.2307/27757059

Kudoh, H., Nakayama, M., Lihová, J. & Marhold, K. (2007). Does invasion involve alternation of germination requirements? A comparative study between native and introduced strains of an

annual Brassicaceae, *Cardamine hirsuta*. *Ecol. Res.* 22, 869–875. doi: 10.1007/s11284-007-0417-5

Landis, J.B., Soltis, D.E., Li, Z., Marx, H.E., Barker, M.S., Tank, D.C. & Soltis, P.S. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* 105, 348–363. doi: 10.1002/ajb2.1060

Larridon, I., Villaverde, T., Zuntini, A.R., Pokorny, L., Brewer, G.E., Epitawalage, N., Fairlie, I., Hahn, M., Kim, J., Maguilla, E., Maurin, O., Xanthos, M., Hipp, A.L., Forest, F. & Baker, W.J. (2019). Tackling Rapid Radiations With Targeted Sequencing. *Front. Plant Sci.* 10, 1655. doi: 10.3389/fpls.2019.01655

Lautenschlager, U., Wagner, F. & Oberprieler, C. (2020). AllCoPol: inferring allele co-ancestry in polyploids. *BMC Bioinform.* 21, 441. doi: 10.1186/s12859-020-03750-9

Legendre, P. & Legendre, L. (1998). *Numerical ecology, Edition 2*. Amsterdam: Elsevier.

Leigh, J.W. & Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210X.12410

Leitch, I.J. & Dodsworth, S. (2017). Endopolyploidy in plants. In J. Wiley & Sons (Eds.), *eLS* (pp. 1–10). Chichester: John Wiley & Sons, Ltd. doi: 10.1002/9780470015902.a0020097.pub2

Leitch, I.J., Johnston, E., Pellicer, J., Hidalgo, O. & Bennett, M.D. (2019). Plant DNA C-values database (release 7.1, Apr 2019). Retrieved from https://cvalues.science.kew.org/ on May 2023).

Li, D.Z., Gao, L.M., Li, H.T., Wang, H., Ge, X.J., Liu, J.Q., Chen, Z.D., Zhou, S.L., Chen, S.L., Yang, J.B., Fu, C.X., Zeng, C.X., Yan, H.F., Zhu, Y.J., Sun, Y.S., Chen, S.Y., Zhao, L., Wang, K., Yang, T. & Duan, G.W. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *P. Natl. Acad. Sci. USA* 108, 19641–19646. doi: 10.1073/pnas.1104551108

Li, F.W. & Harkess, A. (2018). A guide to sequence your favorite plant genomes. *Appl. Plant Sci.* 6, e1030. doi: 10.1002/aps3.1030

Li, X., Hao, B., Pan, D. & Schneeweiss, G.M. (2017). Marker Development for Phylogenomics: The Case of Orobanchaceae, a Plant Family with Contrasting Nutritional Modes. *Front. Plant Sci.* 8, 1973. doi: 10.3389/fpls.2017.01973

Lihová, J. & Marhold, K. (2006). Phylogenetic and diversity patterns in *Cardamine* (Brassicaceae) - a genus with conspicuous polyploid and reticulate evolution. In A.K. Sharma & A. Sharma (Eds.), *Plant genome: Biodiversity and evolution, Volume 1C, Phanerogams* (pp. 149–186). Enfield: Science Publishers, Inc.

Lihová, J., Marhold, K. & Neuffer, B. (2000). Taxonomy of *Cardamine amara* (Cruciferae) in the Iberian Peninsula. *Taxon* 49, 747–763. doi: 10.2307/1223975

Lihová, J., Tribsch, A. & Marhold, K. (2003): The *Cardamine pratensis* group (Brassicaceae) in the Iberian Peninsula: taxonomy, polyploidy and distribution. *Taxon* 52, 783–802. doi: 10.2307/3647352

Lihová, J., Marhold, K., Tribsch, A. & Stuessy, T.F. (2004a): Morphometric and AFLP reevaluation of tetraploid *Cardamine amara* (Brassicaceae) in the Mediterranean. *Syst. Bot.* 2*,* 134–146. doi: 10.1600/036364404772974059

Lihová, J., Fuertes Aguilar, J., Marhold, K. & Nieto Feliner, G. (2004b). Origin of the disjunct tetraploid *Cardamine amporitana* (Brassicaceae) assessed with nuclear and chloroplast DNA sequence data. *Am. J. Bot.* 91, 1231–1242. doi: 10.3732/ajb.91.8.1231

Lihová J., Tribsch A. & Stuessy T.F. (2004c): *Cardamine apennina*: a new endemic diploid species of the *C. pratensis* group (Brassicaceae) from Italy. *Plant Syst. Evol.* 245, 69–92. doi: 10.1007/s00606-003-0119-6

Lihová, J., Shimizu, K.K. & Marhold, K. (2006). Allopolyploid origin of *Cardamine asarifolia* (Brassicaceae): incongruence between plastid and nuclear ribosomal DNA sequences solved by a single-copy nuclear gene. *Mol. Phylogenet. Evol.* 39, 759–786. doi: 10.1016/j.ympev.2006.01.027

Lihová, J., Kučera, J., Perný, M. & Marhold, K. (2007a). Hybridization between two polyploid *Cardamine* (Brassicaceae) species in north-western Spain: discordance between morphological and genetic variation patterns. *Ann. Bot.* 99, 1083–1096. doi: 10.1093/aob/mcm056

Lihová, J., Kochjarová, J. & Marhold, K. (2007b). Hybridization between polyploids *Cardamine enneaphyllos* and *C. glanduligera* (Brassicaceae) in the West Carpathians: evidence from morphology, pollen fertility and PCR-RFLP patterns. *Preslia* 79, 101–125.

Lihová, J., Kudoh, H. & Marhold, K. (2010). Genetic structure and phylogeography of a temperate-boreal herb, *Cardamine scutata* (Brassicaceae), in northeast Asia inferred from AFLPs and cpDNA haplotypes. *Am. J. Bot.* 97, 1058–1070. doi: 10.3732/ajb.0900361

López-González, N., Bobo-Pinilla, J., Padilla-García, N., Loureiro, J., Castro, S., Rojas-Andrés, B.M. & Martínez-Ortega, M.M. (2021). Genetic similarities versus morphological resemblance: Unraveling a polyploid complex in a Mediterranean biodiversity hotspot. *Mol. Phylogenet. Evol.* 155, 107006. doi: 10.1016/j.ympev.2020.107006

Madlung, A. (2013). Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* 110, 99–104. doi: 10.1038/hdy.2012.79

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L. & Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the earlyrise of flowering plant phylogenetic diversity. *New Phytol.* 207, 437–453. doi: 10.1111/nph.13264

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. *J. Biogeogr.* 35, 865–883. doi: 10.1111/j.1365-2699.2007.01849.x

Maharramova, E.H., Safarov, H.M., Kozlowski, G., Borsch, T. & Muller, L.A. (2015). Analysis of nuclear microsatellites reveals limited differentiation between Colchic and Hyrcanian populations of the wind-pollinated relict tree *Zelkova carpinifolia* (Ulmaceae). *Am. J. Bot.* 102, 119–128. doi: 10.3732/ajb.1400370

Maharramova, E., Huseynova, I., Kolbaia, S., Gruenstaeudl, M., Borsch, T. & Muller, L.A.H. (2018). Phylogeography and population genetics of the riparian relict tree *Pterocarya fraxinifolia* (Juglandaceae) in the South Caucasus. *Syst. Biodivers.* 16, 14–27. doi: 10.1080/14772000.2017.1333540

Maherali, H., Walden, A.E. & Husband, B.C. (2009). Genome duplication and the evolution of physiological responses to water stress. *New Phytol.* 184, 721–731. doi: 10.1111/j.1469-8137.2009.02997.x

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237. doi: 10.1016/j.tree.2005.02.010

Mallet, J. (2007). Hybrid speciation. *Nature* 446, 279–283. doi: 10.1038/nature05706

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. *J. Biogeogr.* 41, 366–379. doi: 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. *Biol. Rev.* 92, 1365–1388. doi: 10.1111/brv.12287

Mandáková, T. & Lysak, M.A. (2016). Painting of Arabidopsis chromosomes with chromosome-specific BAC clones. *Curr. Protoc. Plant Biol.* 1, 359–371. doi: 10.1002/cppb.20022

Mandáková, T. & Lysak, M. A. (2018). Post-polyploid diploidization and diversification through dysploid changes. *Curr. Opin. Plant Biol.* 42, 55–65. doi: 10.1016/j.pbi.2018.03.001

Mandáková, T., Joly, S., Krzywinski, M., Mummenhoff, K. & Lysak, M.A. (2010). Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. *Plant Cell* 22, 22772290. doi: 10.1105/tpc.110.074526

Mandáková, T., Kovařík, A., Zozomová-Lihová, J., Shimizu-Inatsugi, R., Shimizu, K.K., Mummenhoff, K., Marhold, K. & Lysák, M.A. (2013). The more the merrier: recent hybridization and polyploidy in *Cardamine*. *Plant Cell* 25, 3280–3295. doi: 10.1105/tpc.113.114405

Mandáková, T., Marhold, K. & Lysák, M.A. (2014). The widespread crucifer species *Cardamine flexuosa* is an allotetraploid with a conserved subgenomic structure. *New Phytol.* 201, 982–992. doi: 10.1111/nph.12567

Mandáková, T., Gloss, A.D., Whiteman, N.K. & Lysák, M.A. (2016). How diploidization turned a tetraploid into a pseudotriploid. *Am. J. Bot.* 103, 1187–1196. doi: 10.3732/ajb.1500452

Mandáková, T., Li, Z., Barker, M.S. & Lysák, M.A. (2017). Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *Plant J.* 91, 3–21. doi: 10.1111/tpj.13553

Mandáková, T., Zozomová-Lihová, J., Kudoh, H., Zhao, Y., Lysák, M.A. & 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. doi: 10.1093/aob/mcz019

Mandáková, T., Ashby, K., Price, B.J., Windham, M.D., Carman, J.G. & Lysák, M.A. (2021). Genome structure and apomixis in *Phoenicaulis* (Brassicaceae; Boechereae). *J. Syst. Evol.* 59, 83–92. doi: 10.1111/jse.12555

Mandel, J.R., Dikow, R.B., Funk, V.A., Masalia, R.R., Staton, S.E., Kozik, A., Michelmore, R.W., Rieseberg, L.H. & Burke, J.M. (2014). A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Appl. Plant Sci.* 2, 1300085. doi: 10.3732/apps.1300085

Marhold, K. (1992). A multivariate morphometric study of the *Cardamine amara* group (Cruciferae) in the Carpathian and Sudeten mountains. *Bot. J. Linn. Soc.* 110, 121–135. doi: 10.1111/j.1095-8339.1992.tb00286.x

Marhold, K. (1994). Taxonomy of the genus *Cardamine* L. (Cruciferae) in the Carpathians and Pannonia. I. *Cardamine pratensis* group. *Folia Geobot. Phytotax.* 29, 335–374. doi: 10.1007/BF02882912

Marhold, K. (1995). *Cardamine rivularis* auct. non Schur in the Eastern Alps. *Carinthia II* 53, 101–102.

Marhold K. (1999). Taxonomic evaluation of the tetraploid populations of *Cardamine amara* (Brassicaceae) from the Eastern Alps and adjacent areas. *Bot. Helv.* 109, 67–84.

Marhold, K. (2011). Multivariate morphometrics and its application to monography at specific and infraspecific levels. In T.F. Stuessy & H.W. Lack (Eds.), *Monographic plant systematics: fundamental assessment of plant biodiversity* (pp. 73–99). Ruggell: A. R. G. Gantner Verlag K. G.

Marhold, K. & Ančev, M.E. (1999). *Cardamine penzesii*, a rediscovered taxon of the *C. pratensis* group (Cruciferae). *Ann. Bot. Fenn.* 36, 171–180.

Marhold, K., Ančev, M.E. & Tan, K. (1996). A new subspecies of *Cardamine amara* (Brassicaceae) from Bulgaria and Greece. *Ann. Bot. Fenn*. 33, 199–204.

Marhold, K., Lihová, J., Perný, M., Grupe, R. & Neuffer, B. (2002). Natural hybridization in *Cardamine* (Brassicaceae) in the Pyrenees: evidence from morphological and molecular data. *Bot. J. Linn. Soc.* 139, 275–294. doi: 10.1046/j.1095-8339.2002.00066.x

Marhold, K., Lihová, J., Perný, M. & Bleeker, W. (2004). Comparative ITS and AFLP analysis of diploid *Cardamine* (Brassicaceae) taxa from closely related polyploid complexes. *Ann. Bot.* 93, 507–520. doi: 10.1093/aob/mch073

Marhold, K., Kudoh, H., Pak, J., Watanabe, K., Španiel, S. & Lihová, J. (2010). Cytotype diversity and genome size variation in eastern Asian polyploid *Cardamine* (Brassicaceae) species. *Ann. Bot.* 105, 249–264. doi: 10.1093/aob/mcp282

Marhold, K., Šlenker, M. & Zozomová-Lihová, J. (2018). Polyploidy and hybridization in the Mediterranean and neighbouring northern areas: examples from the genus *Cardamine* (Brassicaceae). *Biologia Serbica* 40, 47–59. doi: 10.5281/zenodo.1424339

Marhold, K., Kempa, M. & Kučera, J. (2021 onwards). Datababase of names, chromosome numbers, ploidy levels and genome sizes of the tribe Cardamineae. https://cardamine.sav.sk (1 March 2022).

Martin, M., Patterson, M., Garg, S., Fischer, S.O., Pisanti, N., Klau, G.W., Schöenhuth, A., Marschall, T. (2016). WhatsHap: fast and accurate read-based phasing. *bioRxiv*. doi: 10.1101/085050

Mason, A.S. & Pires, J.C. (2015). Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends Genet.* 31, 5–10. doi: 10.1016/j.tig.2014.09.011

Matsuhashi, S., Kudoh, H., Maki, M., Cartolano, M., Tsiantis, M., Itagaki, T. & Sakai, S. (2016). Invasion history of *Cardamine hirsuta* in Japan inferred from genetic analyses of herbarium specimens and current populations. *Biol. Invasions* 18, 1939–1951. doi: 10.1007/s10530-016-1139-9

Mayrose, I., Zhan, S.H., Rothfels, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H. & Otto, S.P. (2011). Recently formed polyploid plants diversify at lower rates. *Science* 333, 1257. doi: 10.1126/science.1207205

Médail, F. & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *J. Biogeogr.* 36, 1333–1345. doi: 10.1111/j.1365-2699.2008.02051.x

Médail, F. & Quézel, P. (1997). Hot-spots analysis for conservation of plant biodiversity in the Mediterranean Basin. *Ann. Mo. Bot. Gard.* 84, 112–127. doi: 10.2307/2399957

Melichárková, A., Španiel, S., Brišková, D., Marhold, K. & Zozomová-Lihová, J. (2017). Unravelling allopolyploid origins in the *Alyssum montanum–A. repens* species complex (Brassicaceae): low-copy nuclear gene data complement plastid DNA sequences and AFLPs. *Bot. J. Linn. Soc.* 184, 485–502. doi: 10.1093/botlinnean/box039

Melichárková, A., Španiel, S., Marhold, K., Hurdu, B.I., Drescher, A. & Zozomová-Lihová, J. (2019). Diversification and independent polyploid origins in the disjunct species *Alyssum repens* from the Southeastern Alps and the Carpathians. *Am. J. Bot.* 106, 1499–1518. doi: 10.1002/ajb2.1370

Melichárková, A., Šlenker, M., Zozomová-Lihová, J., Skokanová, K., Šingliarová, B., Kačmárová, T., Caboňová, M., Kempa, M., Šrámková, G., Mandáková, T., Lysák, M. A., Svitok, M., Mártonfiová, L. & Marhold, K. (2020). So closely related and yet so different: Strong contrasts between the evolutionary histories of species of the *Cardamine pratensis* polyploid complex in Central Europe. *Front. Plant Sci.* 11, 588856. doi: 10.3389/fpls.2020.588856

Mittermeier, R. A., Turner, W.R., Larsen, F.W., Brooks, T.M. & Gascon, C. (2011). Global biodiversity conservation: the critical role of hotspots. In F. Zachos & J. Habel (Eds.), *Biodiversity Hotspots* (pp. 3–22). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-20992-5_1

Mo, L., Chen, J., Chen, F., Xu, Q., Tong, Z., Huang, H., Dong, R., Lou, X., & Lin, E. (2020). Induction and characterization of polyploids from seeds of *Rhododendron fortunei* Lindl. *J. Integr. Agr.* 19, 2016–2026. doi: 10.1016/S2095-3119(20)63210-5

Moharrek, F., Sanmartín, I., Kazempour-Osaloo, S. & Nieto Feliner, G. (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. doi: 10.3389/fgene.2018.00698

Morales-Briones, D.F., Liston, A. & Tank, D.C. (2018). Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytol.* 218, 1668–1684. doi: 10.1111/nph.15099

Morinaga, S.I., Nagano, A.J., Miyazaki, S., Kubo, M., Demura, T., Fukuda, H., Sakai, S. & Hasebe, M. (2008). Ecogenomics of cleistogamous and chasmogamous flowering: genome-wide gene expression patterns from cross-species microarray analysis in *Cardamine kokaiensis* (Brassicaceae). *J. Ecol.* 96, 1086–1097. doi: 10.1111/j.1365-2745.2008.01392.x

Mráz, P., Šingliarová, B., Urfus, T., & Krahulec, F. (2008). Cytogeography of *Pilosella officinarum* (Compositae): Altitudinal and longitudinal differences in ploidy level distribution in the Czech Republic and Slovakia and the general pattern in Europe. *Ann. Bot.* 101, 59–71. doi: 10.1093/aob/mcm282

Muellner-Riehl, A.N., Schnitzler, J., Kissling, W.D., Mosbrugger, V., Rijsdijk, K.F., Seijmonsbergen, A.C., Versteegh, H. & Favre, A. (2019). Origins of global mountain plant biodiversity: Testing the 'mountain-geobiodiversity hypothesis'. *J. Biogeogr.* 46, 2826–2838. doi: 10.1111/jbi.13715

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A. & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858. doi: 10.1038/35002501

Naciri, Y., & Linder, H.P. (2015). Species delimitation and relationships: The dance of the seven veils. *Taxon* 64, 3–16. doi: 10.12705/641.24

Nakhutsrishvili, G., Zazanashvili, N., Batsatsashvili, K. & Montalvo Mancheno, C. S. (2015). Colchic and Hyrcanian forests of the Caucasus: similarities, differences and conservation status. *Fl. Medit.* 25, 185–192. doi: 10.7320/FlMedit25SI.185

Nieto Feliner, G. (2014). Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspect. Plant Ecol.* 16, 265–278. doi: 10.1016/j.ppees.2014.07.002

Nieto Feliner, G., & Rosselló, J.A. (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol. Phylogenet. Evol.* 44, 911–919. doi: 10.1016/j.ympev.2007.01.013

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

Nieto Feliner, G., Álvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F., Piñeiro, R., Riina, R., Rosselló, J.A., Soltis, P.S. & Villa-Machío, I. (2017). Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity* 118, 513–516. doi: 10.1038/hdy.2017.7

Nikolov, L.A., Shushkov, P., Nevado, B., Gan, X., Al-Shehbaz, I.A., Filatov, D., Donovan Bailey, C. & Tsiantis, M. (2019). Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytol.* 222, 1638–1651. doi: 10.1111/nph.15732

Noroozi, J. (2020). Introduction. In J. Noroozi (Ed.), *Plant biogeography and vegetation of high mountains of Central and South-West Asia* (pp. vii–xi). Springer.

Noroozi, J., Talebi, A., Doostmohammadi, M., Rumpf, S.B., Linder, H.P. & Schneeweiss, G.M. (2018). Hotspots within a global biodiversity hotspot - areas of endemism are associated with high mountain ranges. *Sci. Rep.* 8, 10345. doi: 10.1038/s41598-018-28504-9

Noroozi, J., Zare, G., Sherafati, M., Mahmoodi, M., Moser, D., Asgarpour, Z., & Schneeweiss, G.M. (2019). Patterns of endemism in Turkey, the meeting point of three global biodiversity hotspots, based on three diverse families of vascular plants. *Front. Ecol. Evol.* 7, 159. doi: 10.3389/fevo.2019.00159

Nosil, P. (2012). What is ecological speciation? In P. Nosil (Ed.), *Ecological speciation* (pp. 3–21). New York: Oxford University Press Inc.

O'Connor, C. (2008). Fluorescence *In Situ* Hybridization (FISH). *Nature Education* 1, 171.

Olšavská, K., Slovák, M., Marhold, K., Štubňová, E. & Kučera, J. (2016). On the origins of Balkan endemics: the complex evolutionary history of the *Cyanus napulifer* group (Asteraceae). *Ann. Bot.* 118, 1071–1088. doi: 10.1093/aob/mcw142

Otto, F. (1990). DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. *Method. Cell Biol.* 33, 105–110. doi: 10.1016/s0091-679x(08)60516-6

Otto, S.P. & Whitton, J. (2000). Polyploid incidence and evolution. *Ann. Rev. Genet.* 34, 401–437. doi: 10.1146/annurev.genet.34.1.401

Özüdoğru, B. & Mummenhoff, K. (2020). Phylogenetic and biogeographical history confirm the Anatolian origin of *Bornmuellera* (Brassicaceae) and clade divergence between Anatolia and the Balkans in the Plio-Pleistocene transition. *Turk. J. Bot.* 44, 593–603. doi: 10.3906/bot-2007-42

Özüdoğru, B., Karacaoğlu, Ç., Akaydın, G., Erik, S., Mummenhoff, K. & Sağlam, I. K. (2022). Ecological specialization promotes diversity and diversification in the East Mediterranean genus *Ricotia* (Brassicaceae). *J. Syst. Evol.* 60, 331–343. doi: 10.1111/jse.12749

Parisod, C., Holderegger, R. & Brochmann, C. (2010). Evolutionary consequences of autopolyploidy. *New Phytol.* 186, 5–17. doi: 10.1111/j.1469-8137.2009.03142.x

Parolly, G. 2020. The Anatolian High-Mountain Ranges – Plant Diversity Between Two Seas. Pp. 215–286 in: Noroozi, J. (Ed.), *Plant Biogeography and Vegetation of High Mountains of Central and South-West Asia*. Cham: Springer Nature. doi: 10.1007/978-3-030-45212-4_7

Pease, J.B., Haak, D.C., Hahn, M.W. & Moyle, L.C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol.* 14, e1002379. doi: 10.1371/journal.pbio.1002379

Peev, D. & Delcheva, M. (2007). Florogeographical elements and evolutionary trends in the Bulgarian flora. In V. Fet & A. Popov (Eds.), *Biogeography and ecology of Bulgaria* (pp. 571–579). Dordrecht: Springer. doi: 10.1007/978-1-4020-5781-6_21

Perný, M., Tribsch, A. & Anchev, M.E. (2004). Infraspecific differentiation in the Balkan diploid *Cardamine acris* (Brassicaceae): molecular and morphological evidence. *Folia Geobot.* 39, 405–429. doi: 10.1007/BF02803211

Perný, M., Tribsch, A., Stuessy, T.F. & Marhold, K. (2005a). Allopolyploid origin of *Cardamine silana* (Brassicaceae) from Calabria (Southern Italy): karyological, morphological and molecular evidence. *Bot. J. Linn. Soc.* 148, 101–116. doi: 10.1111/j.1095-8339.2005.00389.x

Perný, M., Tribsch, A., Stuessy, T.F. & Marhold, K. (2005b). Taxonomy and cytogeography of *Cardamine raphanifolia* and *C. gallaecica* (Brassicaceae) in the Iberian Peninsula. *Plant Syst. Evol.* 254, 69–91. doi: 10.1007/s00606-005-0317-5

Perrigo, A., Hoorn, C. & Antonelli, A. (2020). Why mountains matter for biodiversity. *J. Biogeogr.* 47, 315–325. doi: 10.1111/jbi.13731

Podani, J. (2000). *Introduction to the exploration of multivariate biological data*. Leiden: Backhuys Publishers.

Porturas, L.D., Anneberg, T.J., Curé, A.E., Wang, S., Althoff, D.M. & Segraves, K.A. (2019). A meta-analysis of whole genome duplication and the effects on flowering traits in plants. *Am. J. Bot.* 106, 469–476. doi: 10.1002/ajb2.1258

POWO. (2023). Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Retrieved from: http://www.plantsoftheworldonline.org/ (April 2023).

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

R Core Team (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raina, S.N. & Rani, V. (2001). GISH technology in plant genome research. *Methods Cell Sci.* 23, 83–104. doi: 10.1023/A:1013197705523

Ramsey, J. & Schemske, D.W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467

Ramzan, F., Younis, A. & Lim, K.B. (2017). Application of Genomic In Situ Hybridization in horticultural science. *Int. J. Genomics* 2017: 7561909. doi: 10.1155/2017/7561909

Rechinger, K. H. (1963). *Flora Iranica: Flora des iranischen Hochlandes und der umrahmenden Gebirge*. Graz: Akademische Druck- und Verlagsanstalt.

Reichelt, N., Wen, J., Pätzold, C. & Appelhans, M.S. (2021). Target enrichment improves phylogenetic resolution in the genus *Zanthoxylum* (Rutaceae) and indicates both incomplete lineage sorting and hybridization events. *Ann. Bot.* 128, 497–510. doi: 10.1093/aob/mcab092

Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., Ma, H. & Qi, J. (2018). Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant* 11, 414–428. doi: 10.1016/j.molp.2018.01.002

Rezende, L., Suzigan, J., Amorim, F.W. & Moraes, A.P. (2020): Can plant hybridization and polyploidy lead to pollinator shift? *Acta Bot. Bras.* 34, 229–242. doi: 10.1590/0102-33062020abb0025

Rieseberg, L.H. & Willis, J.H. (2007). Plant speciation. *Science* 317, 910–914. doi: 10.1126/science.1137729

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* 4, 137–138. doi: 10.1046/j.1471-8286.2003.00566.x

Rothfels, C.J. (2021). Polyploid phylogenetics. *New Phytol.* 230, 66–72. doi: 10.1111/nph.17105

Safran, R.J. & Nosil, P. (2012). Speciation: The Origin of New Species. *Nature Education* 3, 17.

Sancho, R., Inda, L.A., Díaz-Pérez, A., Des Marais, D.L., Gordon, S., Vogel, J.P., Lusinska, J., Hasterok, R., Contreras-Moreira, B. & Catalán, P. (2022). Tracking the ancestry of known and 'ghost' homeologous subgenomes in model grass *Brachypodium* polyploids. *Plant J.* 109, 1535–1558. doi: 10.1111/tpj.15650

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *P. Natl. Acad. Sci. USA* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463

Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S.C.K., Cronn, R.C., Dreyer, L.L. & Suda, J. (2016). Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). *Molec. Ecol. Resour.* 16, 1124–1135. doi: 10.1111/1755-0998.12487

Schmickl, R., Marburger, S., Bray, S. & Yant, L. (2017). Hybrids and horizontal transfer: introgression allows adaptive allele discovery. *J. Exp. Bot.* 68, 5453–5470. doi: 10.1093/jxb/erx297

Schönswetter, P., Lachmayer, M., Lettner, C., Prehsler, D., Rechnitzer, S., Reich, D.S., Sonnleitner, M., Wagner, I., Hülber, K., Schneeweiss, G.M., Trávníček, P. & Suda, J. (2007a). Sympatric diploid and hexaploid cytotypes of *Senecio carniolicus* (Asteraceae) in the Eastern Alps are separated along an altitudinal gradient. *J. Plant Res.* 120, 721–725. doi: 10.1007/s10265-007-0108-x

Schönswetter, P., Suda, J., Popp, M., Weiss-Schneeweiss, H. & Brochmann, C. (2007b). Circumpolar phylogeography of *Juncus biglumis* (Juncaceae) inferred from AFLP fingerprints, cpDNA sequences, nuclear DNA content and chromosome numbers. *Mol. Phylogenet. Evol.* 42, 92–103. doi: 10.1016/j.ympev.2006.06.016

Schranz, M.E., Mohammadin, S. & Edger, P.P. (2012). Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Curr. Opin. Plant Biol.* 15, 147–153. doi: 10.1016/j.pbi.2012.03.011

Schrinner, S.D., Mari, R.S., Ebler, J., Rautiainen, M., Seillier, L., Reimer, J.J., Usadel, B., Marschall, T. & Klau, G.W. (2020). Haplotype threading: accurate polyploid phasing from long reads. *Genome Biol.* 21, 252. doi: 10.1186/s13059-020-02158-1

Schulz, O.E. (1903). Monographie der Gattung *Cardamine*. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie* 32, 280–623.

Schwarzacher, T., Leitch, A.R., Bennett, M.D. & Heslop-Harrison, J.S. (1989). *In situ* localization of parental genomes in a wide hybrid. *Ann. Bot.* 64, 315–324.

Segraves, K.A. (2017). The effects of genome duplications in a community context. *New Phytol.* 215, 57–69. doi: 10.1111/nph.14564

Segraves, K.A. & Anneberg, T.J. (2016). Species interactions and plant polyploidy. *Am. J. Bot.* 103, 1326–1335. doi: 10.3732/ajb.1500529

Segraves, K.A. & Thompson, J.N. (1999). Plant polyploidy and pollination: floral traits and insect visits to diploid and tetraploid *Heuchera grossulariifolia*. *Evolution* 53, 1114–1127. doi: 10.1111/j.1558-5646.1999.tb04526.x

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E. & Small, R.L. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142–166. doi: 10.3732/ajb.92.1.142

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. doi: 10.3732/ajb.94.3.275

Shimizu-Inatsugi, R., Terada, A., Hirose, K., Kudoh, H., Sese, J. & Shimizu, K.K. (2017). Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. *Mol. Ecol.* 26, 193–207. doi: 10.1111/mec.13738

Silva, G.S. & Souza, M.M. (2013). Genomic *in situ* hybridization in plants. *Genet. Mol. Res.* 12, 2953–2965. doi: 10.4238/2013.August.12.11

Small, R.L., Cronn, R.C. & Wendel, J.F. (2004). Use of nuclear genes for phylogeny reconstruction in plants. *Aust. Syst. Bot.* 17, 145–170. doi: 10.1071/SB03015

Sneath, P.H.A. & Sokal, R.R. (1973). *Numerical taxonomy. The principles and practice of numerical classification*. San Francisco, CA: Freeman.

Sobierajska, K., Boratyńska, K., Jasińska, A., Dering, M., Ok, T., Douaihy, B., Dagher-Kharrat, M.B., Romo, A. & Boratyński, A. (2016). Effect of the Aegean Sea barrier between Europe and Asia on differentiation in *Juniperus drupacea* (Cupressaceae). *Bot. J. Linn. Soc.* 180, 365–385. doi: 10.1111/boj.12377

Solís-Lemus, C. & Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet.* 12, e1005896. doi: 10.1371/journal.pgen.1005896

Soltis, P.S. & Soltis, D.E. (2009). The role of hybridization in plant speciation. *Ann. Rev. Plant Biol.* 60, 561–588. doi: 10.1146/annurev.arplant.043008.092039

Soltis, P.S. & Soltis, D.E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* 30, 159–165. doi: 10.1016/j.pbi.2016.03.015

Soltis, D.E., Soltis, P.S. & Tate, J.A. (2003). Advances in the study of polyploidy since Plant speciation. *New Phytol.* 161, 173–191. doi: 10.1046/j.1469-8137.2003.00948.x

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., Wall, P.K. & Soltis, P.S. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348. doi: 10.3732/ajb.0800079

Soltis, D.E., Buggs, R.J.A., Doyle, J.J. & Soltis, P.S. (2010). What we still don't know about polyploidy. *Taxon* 59, 1387–1403. doi: 10.1002/tax.595006

Soltis, D.E., Visger, C.J. & Soltis, P.S. (2014). The polyploidy revolution then…and now: Stebbins revisited. *Am. J. Bot.* 101, 1057–1078. doi: 10.3732/ajb.1400178

Spoelhof, J.P., Soltis, P.S. & Soltis, D.E. (2017). Pure polyploidy: Closing the gaps in autopolyploid research. *J. Syst. Evol.* 55, 340–352. doi: 10.1111/jse.12253

Stevanović, V., Tan, K. & Petrova, A. (2007). Mapping the endemic flora of the Balkans - a progress report. *Bocconea* 21, 131–137.

Straub, S.C.K., Boutte, J., Fishbein, M. & Livshultz, T. (2020). Enabling evolutionary studies at multiple scales in Apocynaceae through Hyb-Seq. *Appl. Plant Sci.* 8, e11400. doi: 10.1002/aps3.11400

Strid, A. (1986). *Cardamine* L. In A. Strid (Ed.), *Mountain flora of Greece 1* (pp. 256–261). Cambridge, London, New York, New Rochelle, Melbourne, Sydney: Cambridge University Press.

Suda, J. & Trávníček, P. (2006). Reliable DNA ploidy determination in dehydrated tissues of vascular plants by DAPI flow cytometry–new prospects for plant research. *Cytometry* 69A, 273–280. doi: 10.1002/cyto.a.20253

Sun, J., Shimizu-Inatsugi, R., Hofhuis, H., Shimizu, K., Hay, A., Shimizu, K.K. & Sese, J. (2020). A recently formed triploid *Cardamine insueta* inherits leaf vivipary and submergence tolerance traits of parents. *Front. Genet.* 11, 567262. doi: 10.3389/fgene.2020.567262

Sweeney, P.W. & Price, R.A. (2000). Polyphyly of the genus *Dentaria* (Brassicaceae): Evidence from *trnL* intron and *ndhF* sequence data. *Syst. Bot.* 25, 468–478. doi: 10.2307/2666690

Šlenker, M., Zozomová-Lihová, J., Mandáková, T., Kudoh, H., Zhao, Y., Soejima, A., Yahara, T., Skokanová, K., Španiel, S. & Marhold, K. (2018). Morphology and genome size of the widespread weed *Cardamine occulta*: how it differs from cleistogamic *C. kokaiensis* and other closely related taxa in Europe and Asia. *Biol. J. Linn. Soc.* 187, 456–482. doi: 10.1093/botlinnean/boy030

Šlenker, M., Perný, M., Zozomová-Lihová, J. & Marhold, K. (2021). Taxonomic position and circumscription of *Cardamine barbaraeoides* (Brassicaceae), a systematically challenging taxon from the Balkan Peninsula. *Phytotaxa* 502, 111–132. doi: 10.11646/phytotaxa.502.2.1

Šlenker, M., Koutecký, P. & Marhold, K. (2022). MorphoTools2: an R package for multivariate morphometric analysis. *Bioinformatics* 38, 2954–2955. doi: 10.1093/bioinformatics/btac173

Španiel, S., & Rešetnik, I. (2022). Plant phylogeography of the Balkan Peninsula: spatiotemporal patterns and processes. *Plant Syst. Evol.* 308, 38. doi: 10.1007/s00606-022-01831-1

Španiel, S., Marhold, K., & Zozomová-Lihová, J. (2017). The polyploid *Alyssum montanum-A. repens* complex in the Balkans: a hotspot of species and genetic diversity. *Plant Syst. Evol.* 303, 1443–1465. doi: 10.1007/s00606-017-1470-3

Španiel, S., Šlenker, M., Melichárková, A., Caboňová, M., Šandalová, M., Zeisek, V., Marhold, K. & Zozomová-Lihová, J. (2023). Phylogenetic challenges in a recently diversified and polyploid-rich *Alyssum* (Brassicaceae) lineage: low divergence, reticulation, and parallel polyploid speciation. *Evolution* 77, 1226–1244. doi: 10.1093/evolut/qpad035

Takhtajan, A. (1986). *Floristic Regions of the World*. Berkeley: University of California Press.

Tarkhnishvili, D., Gavashelishvili, A. & Mumladze, L. (2012). Palaeoclimatic models help to understand current distribution of Caucasian forest species. *Biol. J. Linn. Soc.* 105, 231–248. doi: 10.1111/j.1095-8312.2011.01788.x

Tayalé, A. & Parisod, C. (2013). Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenet. Genome Res.* 140, 79–96. doi: 10.1159/000351318

Tan, K. (2002). *Cardamine* L. In A. Strid & K. Tan (Eds.), *Flora Hellenica 2* (pp. 178–184). Ruggell: A. R. G. Gantner Verlag K. G.

Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B. & Harmon, L.J. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* 207, 454–467. doi: 10.1111/nph.13491

Tedder, A., Helling, M., Pannell, J.R., Shimizu-Inatsugi, R., Kawagoe, T., van Campen, J., Sese, J., & Shimizu, K.K. (2015). Female sterility associated with increased clonal propagation suggests a unique combination of androdioecy and asexual reproduction in populations of *Cardamine amara* (Brassicaceae). *Ann. Bot.* 115, 763–776. doi: 10.1093/aob/mcv006

Temsch, E.M., Greilhuber, J., & Krisai, R. (2010). Genome size in liverworts. *Preslia* 82, 63–80.

Thomas, G.W.C., Ather, S.H. & Hahn, M.W. (2017). Gene-Tree Reconciliation with MUL-Trees to Resolve Polyploidy Events. *Syst. Biol.* 66, 10071018. doi: 10.1093/sysbio/syx044

Thomas, T., Gilbert, J. & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation* 2, 3. doi: 10.1186/2042-5783-2-3

Thompson, J.D. (2020). *Plant Evolution in the Mediterranean: Insights for conservation. 2nd Edition*. New York: Oxford University Press. doi: 10.1093/oso/9780198835141.001.0001

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R. & Greiner, S. (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391

Tzedakis, P.C. (2004). The Balkans as prime glacial refugial territory of European temperate trees. In H. I. Griffiths, B. Kryštufek & J. M. Reed (Eds.), *Balkan Biodiversity* (pp. 49–68). Dordrecht: Springer. doi: 10.1007/978-1-4020-2854-0_4

Vallejo-Marín, M., & Hiscock, S.J. (2016). Hybridization and hybrid speciation under global change. *New Phytol.* 211, 1170–1187. doi: 10.1111/nph.14004

Van de Peer, Y., Ashman, T., Soltis, P.S. & Soltis, D.E. (2020). Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell*, koaa015, 1–16. doi: 10.1093/plcell/koaa015

van Dijk, E.L., Auger, H., Jaszczyszyn, Y. & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426. doi: 10.1016/j.tig.2014.07.001

Villaverde, T., Pokorny, L., Olsson, S., Rincón-Barrado, M., Johnson, M.G., Gardner, E.M., Wickett, N.J., Molero, J., Riina, R. & Sanmartín, I. (2018). Bridging the micro-and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytol.* 220, 636–650. doi: 10.1111/nph.15312

Villaverde, T., Jiménez-Mejías, P., Luceno, M., Waterway, M.J., Kim, S., Lee, B., Rincón-Barrado, M., Hahn, M., Maguilla, E., Roalson, E.H., Hipp, A.L. & The Global Carex Group. (2020). *Bot. J. Linn. Soc.* 194, 141–163. doi: 10.1093/botlinnean/boaa042

Volkova, P., Laczkó, L., Demina, O., Schanzer, I. & Sramkó, G. (2020). Out of Colchis: The colonization of Europe by *Primula vulgaris* Huds. (Primulaceae). *Acta Soc. Bot. Pol.* 89, 89313. doi: 10.5586/asbp.89313

Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

Warren, D.L., Glor, R.E. & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution* 62, 2868–2883. doi: 10.1111/j.1558-5646.2008.00482.x

Warwick S.I. & Al-Shehbaz, I.A. (2006). Brassicaceae: chromosome number index and database on CD-Rom. *Plant Syst. Evol.* 259, 237–248. doi: 10.1007/s00606-006-0421-1

Weitemier, K., Straub, S.C., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A. & Liston, A. (2014). Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Appl. Plant Sci.* 2, 1400042. doi: 10.3732/apps.1400042

Wen, D., Yu, Y., Zhu, J. & Nakhleh, L. (2018). Inferring phylogenetic networks using PhyloNet. *Syst. Biol.* 67, 735–740. doi: 10.1093/sysbio/syy015

White, T.J., Bruns, T., Lee, S. & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M.A. Innis, D.H. Gelfand, J.J. Sninsky & T.J. White (Eds.), *PCR protocols: a guide to methods and applications* (pp. 315–322). Academic press, Inc.

Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Müller, K.F. & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B. & Rieseberg, L.H. (2009). The frequency of polyploid speciation in vascular plants. *PNAS USA* 106, 13875–13879. doi: 10.1073/pnas.0811575106

Xu, X., Naqinezhad, A., Ghazanfar, S.A., Fragman-Sapir, O., Oganesian, M., Kharrat, M.B.D., Taifour, H., Filimban, F.Z., Matchutadze, I., Shavvon, R.S., Abdullah, M.T., Al-Khulaidi, A. W., Al-Issai, G., Al-Newani, H.R.H., Asswad, N.G., Chepinoga, V.V., Eliades, N.H., Elkordy, A., Farzaliyev, V., Liu, Y., Niu, S., Özcan, T., Sadat, H.A., Seyfullayev, F. & Ma, K. (2020). Mapping Asia plants: Current status on floristic information in Southwest Asia. *Global Ecol. Conser.* 24, e01257. doi: 10.1016/j.gecco.2020.e01257

Yan, Z., Cao, Z., Liu, Y., Ogilvie, H.A. & Nakhleh, L. (2022). Maximum parsimony inference of phylogenetic networks in the presence of polyploid complexes. *Syst. Biol.* 71, 706–720. doi: 10.1093/sysbio/syab081

Yang, Y. & Smith, S.A. (2014). Orthology Inference in Nonmodel Organisms Using Transcriptomes and Low-Coverage Genomes: Improving Accuracy and Matrix Occupancy for Phylogenomics. *Mol. Biol. Evol.* 31, 3081–3092. doi: 10.1093/molbev/msu245

Younis, A., Ramzan, F., Hwang, Y.J. & Lim, K.B. (2015). FISH and GISH: molecular cytogenetic tools and their applications in ornamental plants. *Plant Cell Rep.* 34, 1477–1488. doi: 10.1007/s00299-015-1828-3

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* 19, 153. doi: 10.1186/s12859-018-2129-y

Zhang, L., Liu, D., Zheng, Y., Yan, Z., Dai, S., Li, Y., Jiang, Q., Ye, Y. & Yen, Y. (2010). Frequent occurrence of unreduced gametes in *Triticum turgidum–Aegilops tauschii* hybrids. *Euphytica* 172, 285–294. doi: 10.1007/s10681-009-0081-7

Zhao, L.L., Feng, S.J., Tian, J.Y., Wei, A.Z. & Yang, T.X. (2018). Internal transcribed spacer 2 (ITS2) barcodes: A useful tool for identifying Chinese *Zanthoxylum*. *Appl. Plant Sci.* 6, e01157. doi: 10.1002/aps3.1157

Zimmer, E.A. & Wen, J. (2012). Using nuclear gene data for plant phylogenetics: Progress and prospects. *Mol. Phylogenet. Evol.* 65, 774–785. doi: 10.1016/j.ympev.2012.07.015

Zozomová-Lihová, J., Krak, K., Mandáková, T., Shimizu, K.K., Španiel, S., Vít, P. & Lysák, M.A. (2014). Multiple hybridization events in *Cardamine* (Brassicaceae) during the last 150 years: revisiting a textbook example of neoallopolyploidy. *Ann. Bot.* 113, 817–830. doi: 10.1093/aob/mcu012

Zozomová-Lihová, J., Malánová-Krásná, I., Vít, P., Urfus, T., Senko, D., Svitok, M., Kempa, M., & Marhold K. (2015). Cytotype distribution patterns, ecological differentiation, and genetic structure in a diploid–tetraploid contact zone of *Cardamine amara*. *Am. J. Bot.* 102, 1380–1395. doi: 10.3732/ajb.1500052

Zwickl, D.J. (2006). *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion* [PhD. thesis; The University of Texas, Austin]. Retrieved from http://hdl.handle.net/2152/2666