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; Durović 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 Ančev, 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* Ančev (*C. rivularis* × *C. matthioli*) (Kučera et al., 2005; Ančev 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; Kučera 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

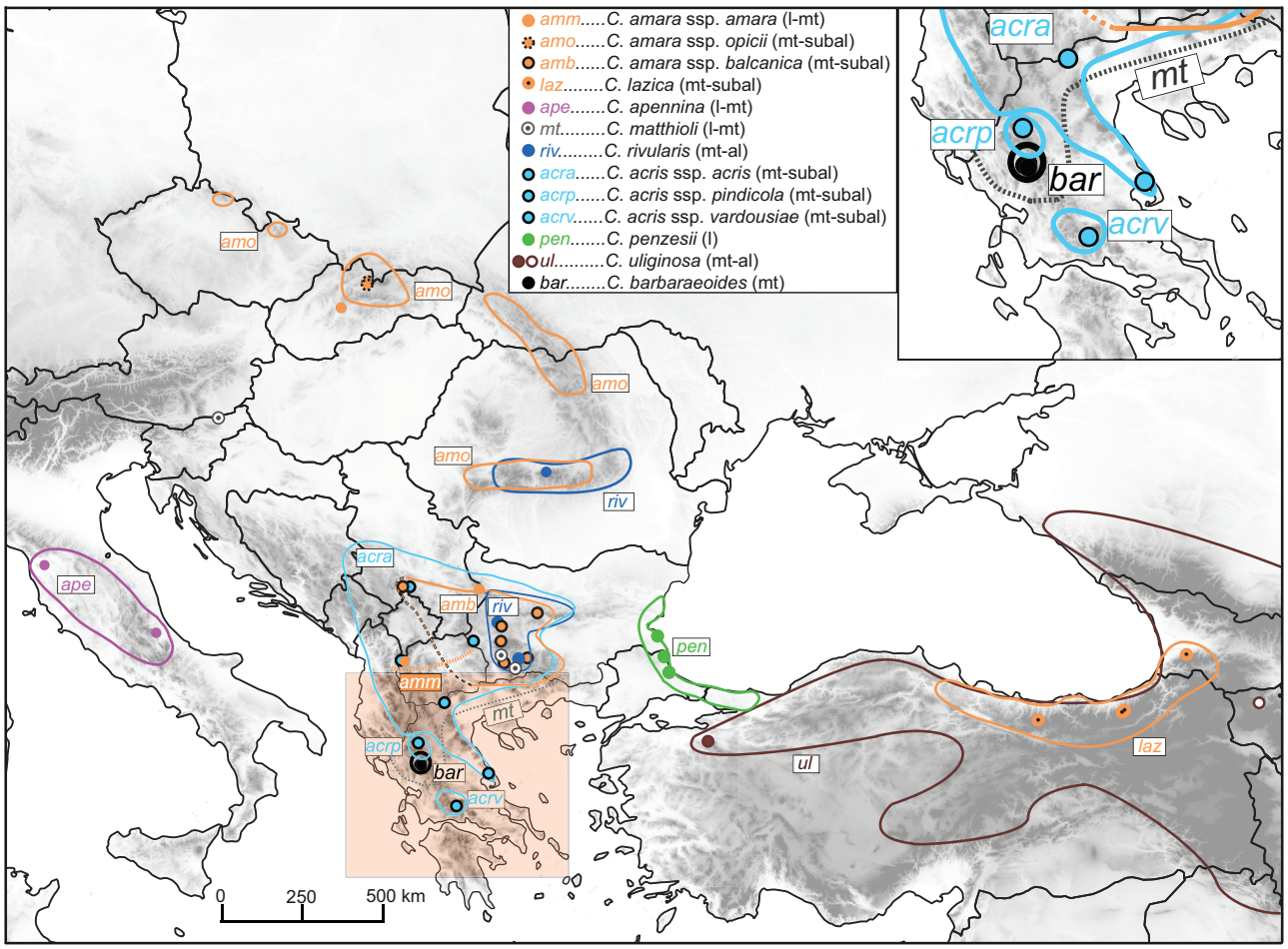

FIGURE 1 | Distribution of the *Cardamine* taxa under study, based on data compiled from floras, herbarium specimens, previous studies, and our own records. The western borders of the area of *Cardamine amara* subsp. *balcanica* remain unknown (marked here by a black–orange dashed line; see Tomović et 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; Tomović 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′,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* (2*C* = 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* (2*C* = 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 65°C) 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 ($\leq$ 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 μg/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 37°C 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 −20°C 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 μg/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 Uludağ 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 Uludağ 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 2*x* 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 Uludağ 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*

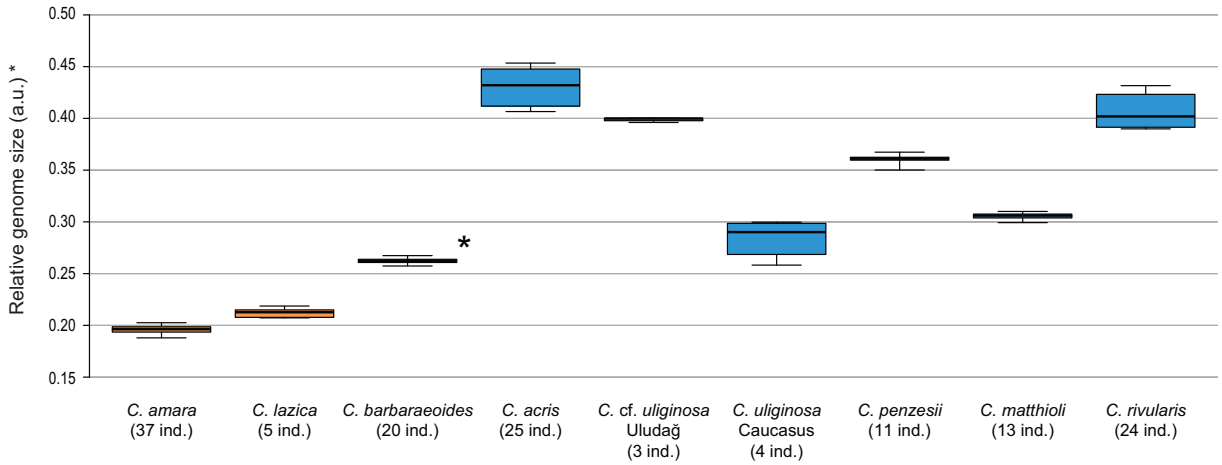

FIGURE 2 | Genome size variation of the *Cardamine* species under study, based on flow cytometric analyses. Relative nuclear DNA content inferred from DAPI measurements is presented, given as the ratio of the sample and standard G1 peaks (2C values in arbitrary units, a.u.). In the tetraploid *Cardamine barbaraeoides* (marked by asterisk), however, DNA content of the meiotically reduced genome (corresponding to the 2*x* 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.

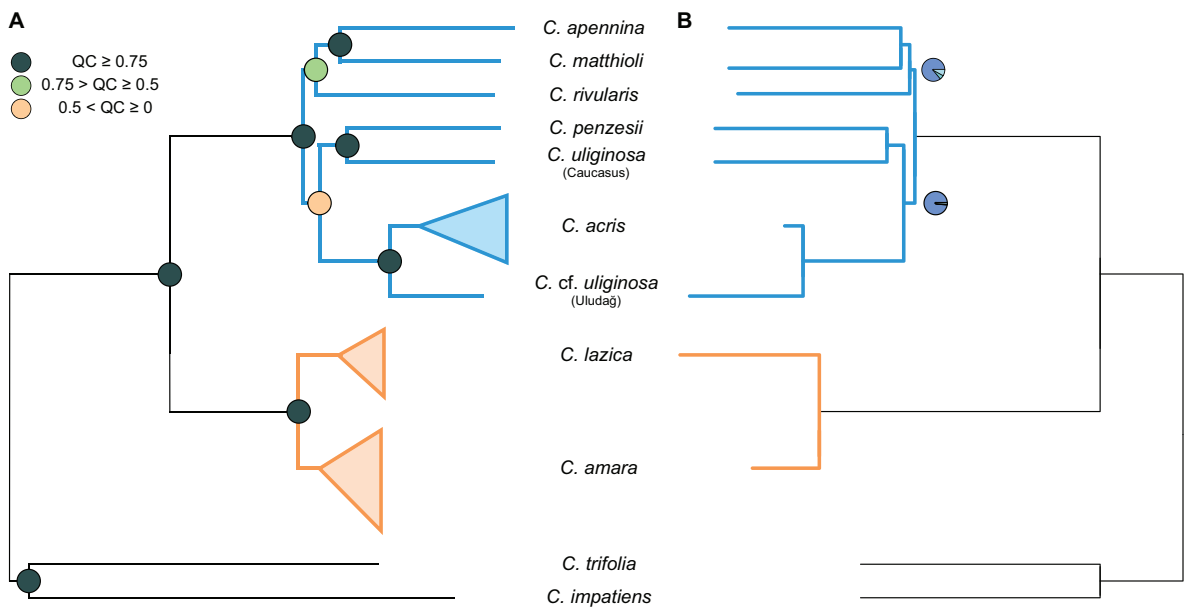

FIGURE 3 | Phylogenetic trees inferred from the complete dataset of 1,168 nuclear genes, based on consensus supercontig sequences of diploid *Cardamine* accessions. Orange color is used for the clades and branches of the *Cardamine amara* group, whereas blue is used for the remaining diploids resolved 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

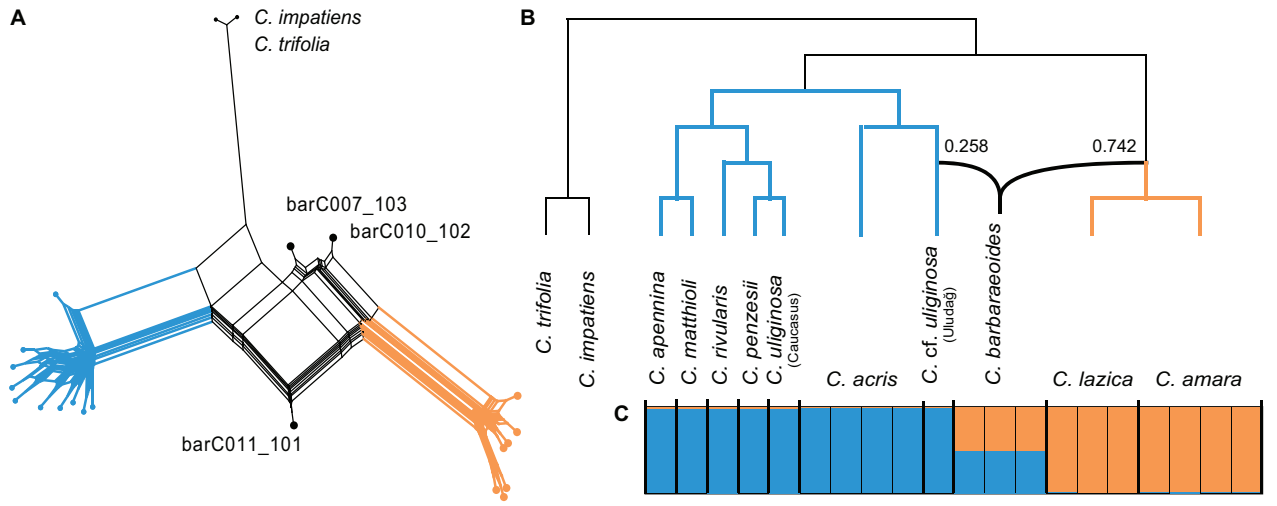

FIGURE 4 | Phylogenetic analyses indicating the hybrid (allopolyploid) origin of the tetraploid *Cardamine barbaraeoides*. Orange color is used for the accessions from the *Cardamine amara* group (*C. amara* and *Cardamine lazica*), blue for the remaining diploids, and black for the tetraploid *C. barbaraeoides*. **(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),

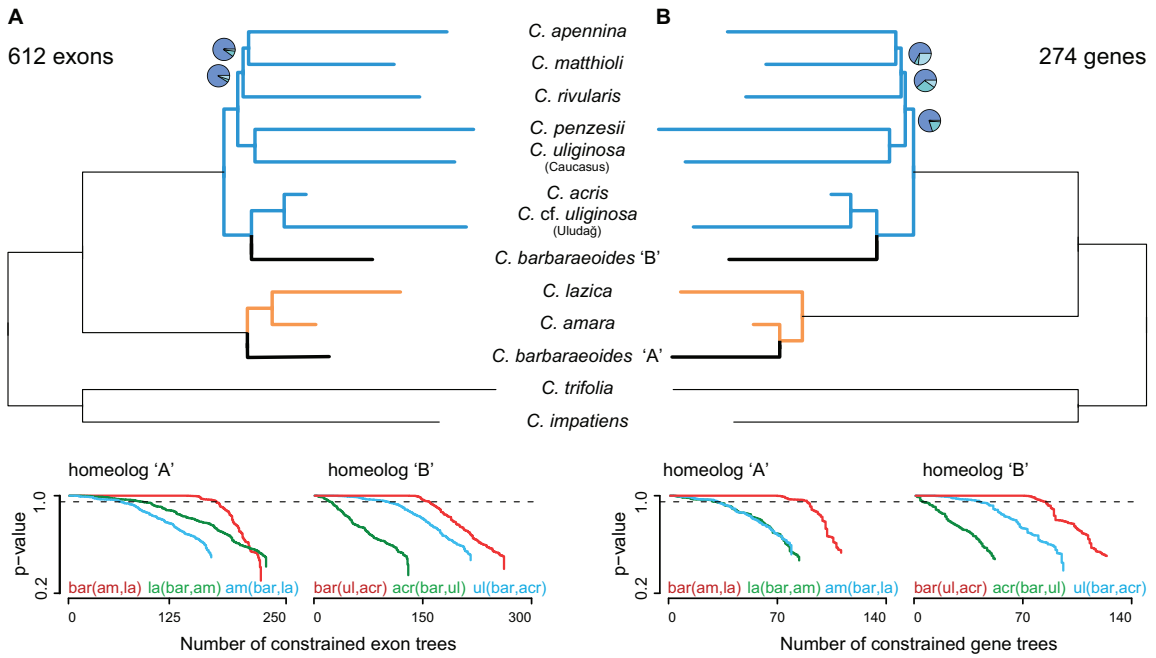

FIGURE 5 | Phylogenetic analyses based on phased allele sequences of 612 exons **(A)** or 274 genes **(B)**, for which alleles of all three accessions of *Cardamine barbaraeoides* were successfully phased and sorted into A and B homeologs. Species trees were inferred in ASTRAL-III. Branch support is 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 ($P \leq 0.05$) 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 Uludağ Mts., green: *C. acris* being sister to the clade of *C. barbaraeoides* "B" and *C.* cf. *uliginosa* from the Uludağ Mts., blue: *C.* cf. *uliginosa* from the Uludağ 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**.

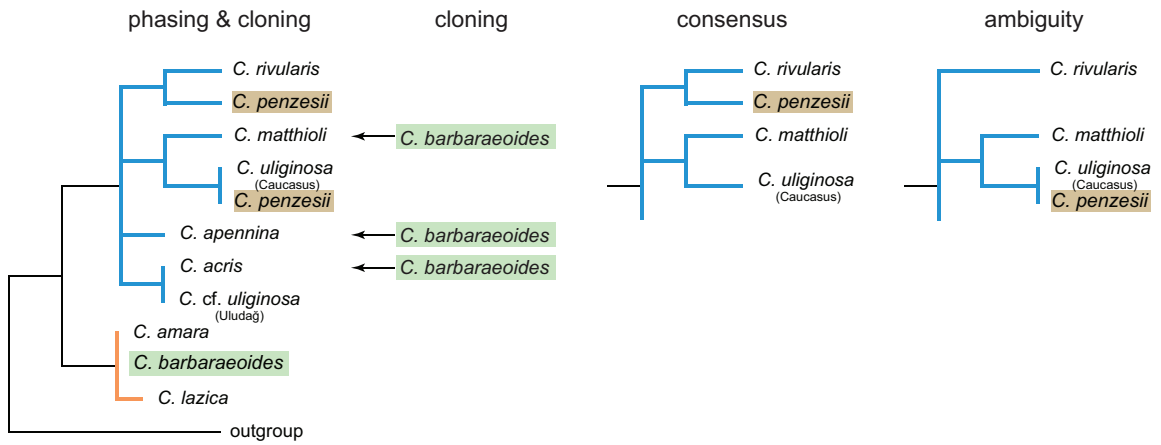

FIGURE 6 | Schematic representation of maximum likelihood (ML) trees inferred from four datasets of nrDNA (ITS1-5.8S-ITS2 region) and their topological differences. Cloning: dataset obtained from molecular cloning and Sanger sequencing; phasing: dataset obtained from read-backed phasing of Hyb-Seq reads, 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;

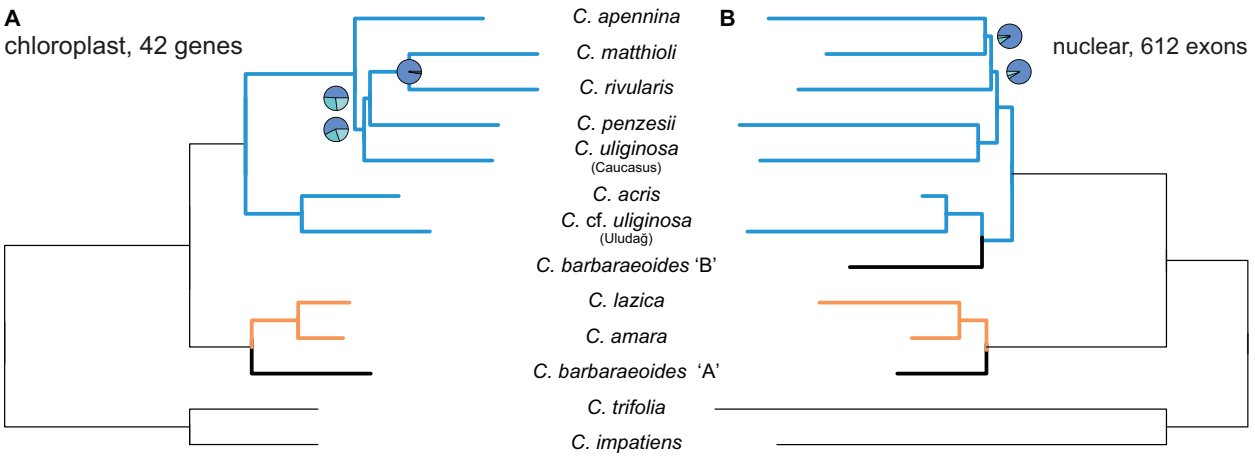

FIGURE 7 | Species trees inferred in ASTRAL-III from the dataset of 42 protein-coding most informative chloroplast genes **(A)** and 612 nuclear exons [**(B)**, as in **Figure 5A**]. Branch support is shown by pie charts, depicting three local posterior probabilities for the given branch (not shown for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

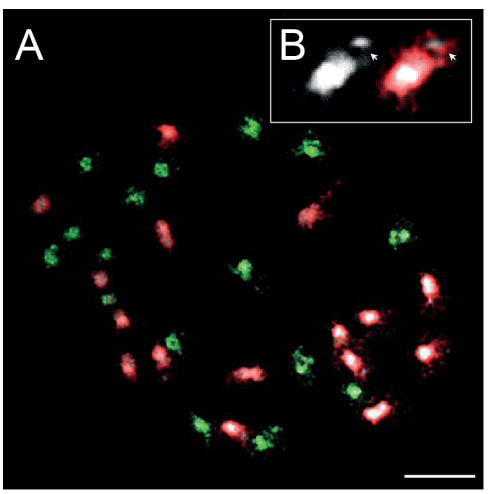

FIGURE 8 | Genomic *in situ* hybridization (GISH) on mitotic chromosomes in the allotetraploid *Cardamine barbaraeoides* (2*n* = 32). **(A)** GISH with total genomic DNA of *Cardamine acris* subsp. *acris* (red fluorescence) and *Cardamine amara* subsp. *balcanica* (green fluorescence) revealed two subgenomes 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′,6-diamidino-2-phenylindole (DAPI); scale bar, 10 μm.

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.2020.106940

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.