Molecular Phylogenetics and Evolution 214 (2026) 108481

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Understanding early stages of speciation: Allopatric divergence, introgression and chromosomal dynamics in the *Erysimum odoratum* species complex

Richard Bačák [a,1], Marek Šlenker [a,1], Barbora Šingliarová [a], Terezie Mandáková [b,c], Katarína Skokanová [a], Ingrid Turisová [d], Peter Turis [d], Janka Smatanová [a], Judita Zozomová-Lihová [a,*]

[a] Institute of Botany, Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Dúbravská cesta 9, SK-845 23 Bratislava, Slovakia
[b] Central European Institute of Technology, Masaryk University, Kamenice 753/5, CZ-625 00 Brno, Czechia
[c] Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czechia
[d] Department of Biology and Environmental Studies, Faculty of Natural Sciences, Matej Bel University in Banská Bystrica, Tajovského 40, SK-974 01 Banská Bystrica, Slovakia

ARTICLE INFO

Keywords:
Carpathians
Dysploidy
Erysimum
Polyploidy
RADseq
Speciation
Target enrichment

ABSTRACT

Accurate species delimitation is essential for understanding biodiversity and evolutionary processes, yet it remains challenging in taxonomically complex groups shaped by recent divergence and reticulate evolution. Nevertheless, such groups offer unique insights into the earliest stages of speciation and its driving forces. The genus *Erysimum* (Brassicaceae), notable for its karyological diversity and high endemism, represents an excellent model for such studies. Here, we investigated the *E. odoratum* complex, encompassing up to 10 recognized species in the Carpathians and western Balkans, to test its monophyletic origin, clarify species boundaries and elucidate the main drivers of diversification. We combined cytotype screening (chromosome counting and flow cytometry), morphometric analysis, and two high-throughput sequencing methods: RADseq, to resolve phylogenetic relationships as well as to detect fine-scale genetic structure and introgression; and target enrichment (Hyb-Seq), to elucidate polyploid origins. Our results demonstrate that the studied complex is polyphyletic, and we focused on the lineage comprising *E. odoratum* s.str. and Carpathian species. Phylogenomic data from the Carpathians contradict traditional taxonomy, which recognized up to four diploid endemics, and instead support a single species, *E. witmannii* with geographically structured genetic variation. Within *E. odoratum* s.str., we identified multiple polyploid cytotypes resulting from independent auto- and allopolyploidization events, although disentangling parental subgenomes, ancestral polymorphisms, and introgression remains difficult. We propose that diversification in this species complex has been driven by a combination of allopatric divergence and reticulate evolution (involving both introgression and allopolyploidy), further shaped by chromosomal dynamics such as dysploidy.

1. Introduction

Species are the fundamental units in many biological fields such as systematics, biogeography, ecology, evolution and conservation biology (de Queiroz, 2005). Accurate species delimitation is therefore essential not only for a correct assessment of biodiversity, but also for any research in these disciplines. Recognizing and delimiting species can be challenging, although an integrative approach is increasingly being taken, in which species boundaries are examined from multiple comparative perspectives (Dayrat, 2005; Oberprieler, 2023; Karbstein et al., 2024). Furthermore, the unified species concept, which defines species as independently evolving metapopulation lineages, eliminates conflicts between alternative species concepts by treating the traditional species defining properties as different lines of evidence relevant for assessing lineage separation (de Queiroz, 2007). This approach is not only more accurate and objective in delineating species, as it integrates

* Corresponding author.
E-mail address: judita.zozomova@savba.sk (J. Zozomová-Lihová).
[1] Equal contribution

https://doi.org/10.1016/j.ympev.2025.108481
Received 9 July 2025; Received in revised form 13 September 2025; Accepted 10 October 2025
Available online 15 October 2025

multiple data types, but also provides insights into the processes that lead to the separation of lineages (de Queiroz, 2007; Wiens, 2007; Oberprieler, 2023; Karbstein et al., 2024). As speciation is a continuous process in which morphological, ecological, and genetic differentiation, as well as reproductive isolation, evolve at different rates, a central challenge is how to integrate and reconcile contradictory results from diverse data sets to avoid under- or overestimating species diversity (de Queiroz, 2007; Dufresnes et al., 2020, 2023; Stankowski and Ravinet, 2021; Karbstein et al., 2024). For instance, strong selection pressure can accelerate morphological or ecological divergence with low genetic differentiation, while genetic drift in peripheral and/or small populations can increase genetic divergence with little or no morphological change (Pinheiro et al., 2018; Oberprieler, 2023). Autopolyploidy or major chromosomal rearrangements leading to instantaneous reproductive isolation may not be accompanied by immediate morphological or ecological divergence (in terms of pollinators, phenology, habitats), resulting in cryptic diversity (Eriksson et al., 2017; Spoelhof et al., 2017; Mairal et al., 2018).

With the advent of high-throughput sequencing methods and genome-scale analyses that can reveal even subtle patterns of genetic structure, genetic divergence is now much easier to detect than in the past, but this also carries the risk of taxonomic over-splitting (Dufresnes et al., 2020; Chambers and Hillis, 2020; Vences et al., 2024). Formalized approaches to species-delimitation applied within a multispecies coalescence model, which can account for confounding processes such as incomplete lineage sorting, offer more certainty and less subjectivity (reviewed by Karbstein et al., 2024; Hausdorf, 2025). Nevertheless, they need to be carefully interpreted in geographically fragmented metapopulations with constrained gene flow and strong intraspecific population structuring (Yang et al., 2019; Mason et al., 2020). Therefore, despite the wealth of available data and tools, delimiting species is not necessarily a trivial task.

In the process of speciation, the genetic divergence of populations is driven by the combined forces of mutation, genetic drift, and differential selection pressures in the genome (Rieseberg and Burke, 2001; Ravinet et al., 2017). Other evolutionary processes that occur frequently in vascular plants include major genomic/chromosomal rearrangements (polyploidy, dysploidy) and hybridization, which generate reticulate evolutionary patterns and lead to taxonomically intricate species groups (Karbstein et al., 2024). These are groups of closely related species or metapopulations that are likely to be at an early stage of divergence with shallow differentiation, harbor ancestral polymorphisms, and are affected by on-going interspecific gene flow and neopolyploidization events. The study of such systems provides a unique opportunity to explore and understand the initial phases of speciation but also poses a challenge for objective and accurate species delimitation and correct assessment of species diversity (Pinheiro et al., 2018). Past taxonomic practice has usually led here to an excess of redundant (synonymous) species names, names of dubious or unknown application, where conclusions were drawn on the basis of sparse plant material, misinterpreted individual phenotypic variants, and failed to consider wider biological evidence (Dayrat, 2005).

The genus *Erysimum* L. is an ideal candidate to validate the traditional, mainly morphology-based taxonomy and to investigate the diversity and complexity of evolutionary and speciation processes. It is one of the largest but poorly explored genera of the Brassicaceae (>200 species, German and Al-Shehbaz, 2008; Kiefer et al., 2014; German et al., 2023), distributed in the Northern Hemisphere, with centres of diversity in Southwest Asia and the Mediterranean Basin. Phylogenetic studies suggest that the current species diversity is the result of a recent and rapid radiation during the late Pliocene to early Pleistocene (Moazzeni et al., 2014; Züst et al., 2020). The genus is rich in endemics (Jalas and Suominen, 1994; Abdelaziz et al., 2014; Moazzeni et al., 2014; Osuna Mascaró, 2020) and exhibits high karyological diversity (Polatschek, 2013; Kiefer et al., 2014). The most comprehensive phylogenetic reconstruction to date (Züst et al., 2020) revealed a high degree of gene discordance, likely due to incomplete lineage sorting, widespread introgression and polyploidy (see e.g., Osuna Mascaró, 2020; Osuna-Mascaró et al., 2022, Osuna Mascaró et al., 2023).

Since the phylogenetic relationships in the genus are still poorly understood, the species are mainly classified into informal species complexes based on similar morphology and growth forms (Ball, 1993; Polatschek, 2013). One of these informal species groups, which is the subject of the present study, is the *E. odoratum* group (Ball, 1993; Jalas and Suominen, 1994). This group comprises diploid to hexaploid biennials growing in xerothermic rocky meadows and shrublands, forest edges, screes and cliffs, predominantly on calcareous substrates (more rarely on serpentine and siliceous ones), with the centre of diversity in Central Europe and the western Balkan Peninsula. Up to ten species have been recognized within the group, although the recognition of some of them is doubtful or their delimitation uncertain. Neither the genetic nor the morphological divergence patterns within this group have yet been studied in detail. As a result, its taxonomic treatment quite differs among different sources (Ball, 1993; Jalas and Suominen, 1994; Michalková, 2002; Oprea, 2005; Marhold, 2011; Polatschek, 2013; Kiefer et al., 2014). The most widespread species is *E. odoratum* Ehrh., which is distributed from northeastern France to eastern Ukraine and spreads south to the northern Balkans (Jalas and Suominen, 1994; Polatschek, 2013). In addition to the widespread tetraploid cytotype (2n = 32), the diploid cytotype (2n = 14) has also been detected in the Western Carpathians (Michalková, 2002; Marhold et al., 2007). Nevertheless, the status of this diploid cytotype is uncertain. Due to presumed morphological differences, some authors considered it as a separate species using not validly published names (*E. vagicum* nom. prov. or *E. carpaticum* nom. prov., see Kliment, 1999; Michalková, 2002), while Polatschek (2013) assigned these diploids to another species, *E. witmannii* Zaw. in his broad concept of this species. Three diploids were reported from the Carpathian range (Jalas and Suominen, 1994; Kliment, 1999; Michalková, 2002): *E. witmannii* (2n = 14, common in the Western Carpathians, scattered in the Eastern and Southern Carpathians), *E. pallidiflorum* Jáv. (2n = 14, restricted to the Western Carpathians and the adjacent Pre-Carpathian region) and *E. transsilvanicum* Schur (2n = 18, common in the Eastern and Southern Carpathians). *Erysimum witmannii* and *E. pallidiflorum* were reported to grow parapatrically but in different habitats (in terms of plant communities, altitude and slope inclination) and morphologically differing in some quantitative traits (Michalková, 1999, 2001, 2002). Both have pale, sulphur-coloured petals in contrast to the bright to golden yellow petals present in *E. odoratum* s.str. (Michalková, 2001). The circumscription of *E. transsilvanicum* (also treated as *E. witmannii* subsp. *transsilvanicum* (Schur) P.W.Ball; Oprea, 2005) and its distinction from *E. witmannii* has never been examined in detail and remains uncertain.

Furthermore, five other species of the *E. odoratum* complex have been reported from the Balkans (Ball, 1993; Jalas and Suominen, 1994): three allopatric species of the montane to (sub)alpine belt, *E. carniolicum* Dolliner (2n = 14, distributed from Slovenia to Montenegro), *E. kuemmerlei* Jáv. (2n = 14, Kosovo, North Macedonia and Albania), *E. pectinatum* Bory & Chaub. (2n = 12, 14; Peloponnese) and two recently described endemics of the southern Dinaric mountains descending to sea level (Polatschek, 2013), *E. vitekii* (2n = 12, southern Bosnia and Herzegovina, southern Croatia, Montenegro, northern Albania) and *E. croaticum* (2n = 42, southern Croatia and Kosovo). The variation in the number of chromosomes found even at the diploid level (2n = 12, 14, 18) and the tetraploid number (2n = 32), which is not twice the known diploid ones, suggest chromosomal rearrangements and high genome dynamics within this species complex.

In the present study, first we aimed at delimiting the *E. odoratum* species complex, testing the hypothesis of its monophyletic origin and resolving the phylogenetic relationships between the recognized species. As the monophyletic origin was refuted, we further focused on the clade comprising the widespread *E. odoratum* s.str. and Carpathian diploids. Using an integrative approach, considering genomic,

morphological, genome size and chromosomal data, we investigated whether the Carpathian endemics, described in the past but of uncertain recognition and delimitation, can be confirmed as separately evolving lineages (species). We also aimed at elucidating the main drivers of evolution within this complex, to find out whether chromosomal evolution (dysploidy and polyploidy) is coupled with speciation and how the polyploids originated (via auto- or allopolyploid pathways). This study provides valuable insights into early phases of speciation and contributes to the debate on appropriate species delimitation within taxonomically critical species complexes.

2. Material and methods

2.1. Plant material

We collected samples (specimens, leaves, seeds) from several populations of each species assigned to the *E. odoratum* group, which were geographically distributed to cover species' ranges. The type localities were also sampled, provided they were sufficiently detailed in the protologues, to ensure correct interpretation of the names. A total of 131 populations of the target complex of *E. odoratum* were sampled. To test the hypothesis of its monophyletic origin, an additional 24 populations of nine other species from the target area were sampled, focusing mainly on diploids to minimize potentially confounding effects of polyploidy and reticulations (outside the target group) on the phylogenetic reconstruction. The additional species sampled included more widespread species (*E. cheiranthoides* L., *E. cuspidatum* (M.Bieb) DC., *E. diffusum* Ehrh.), Balkan endemics (*E. calycinum* Griseb., *E. linariifolium* Tausch, *E. microstylum* Hausskn.) or subendemics extending to the Carpathians, Apennines or Alps (*E. comatum* Pančić, *E. crassistylum* C. Presl, *E. sylvestre* (Crantz) Scop.). All sampled localities are listed in Supplementary Table S1. Geographic distributions of the target species are depicted in Fig. 1, and those of the additional species outside the *E. odoratum* group are shown in Supplementary Fig. S1.

In the Carpathians, we sampled all major mountains with known occurrences or suitable habitats. Unfortunately, the area of Ukraine could not be visited due to the war. However, records of *E. odoratum* and *E. witmannii*/*E. transsilvanicum* are scarce there (Jalas and Suominen, 1994; Polatschek, 2010), reflecting the limited occurrence of alkaline calcareous bedrock in the Ukrainian Carpathians, predominantly restricted to the border area (Chyvchyny Mts., Kobiv and Prokopiv, 2018) of the Maramureş Mts. in Romania. The classification of the pale yellow-flowered populations from the Western Carpathians into *E. witmannii* and *E. pallidiflorum* was morphologically ambiguous and primarily based on the geographic origin of the sampled populations (following Michalková, 1999, 2002, see Supplementary Table S1). The delimitation of *E. witmannii* and *E. transsilvanicum* in the Eastern and Southern Carpathians is unclear, and we could not distinguish between them in the field. All three diploid species, *E. witmannii*, *E. pallidiflorum* and *E. transsilvanicum*, were therefore collectively referred to as *E. witmannii* s.l. Our sampling included the type localities of *E. witmannii* (Pieniny Mts., Poland/Slovakia border) and *E. pallidiflorum* (Naszály, Hungary), and we conducted denser sampling in the Slovenský raj Mts., where both species presumably grow in close proximity (Michalková, 1999). Diploid, bright to golden yellow-flowered populations from the Western Carpathians, which were previously assigned to *E. odoratum* s. str. (Michalková, 2002; Marhold et al., 2007), are referred to here with the provisional name *E. vagicum* nom. prov. to avoid confusion with the tetraploid *E. odoratum* s.str. (from here on without 's.str.'). Accessions of *Malcolmia graeca*, from the genus phylogenetically closest to *Erysimum*

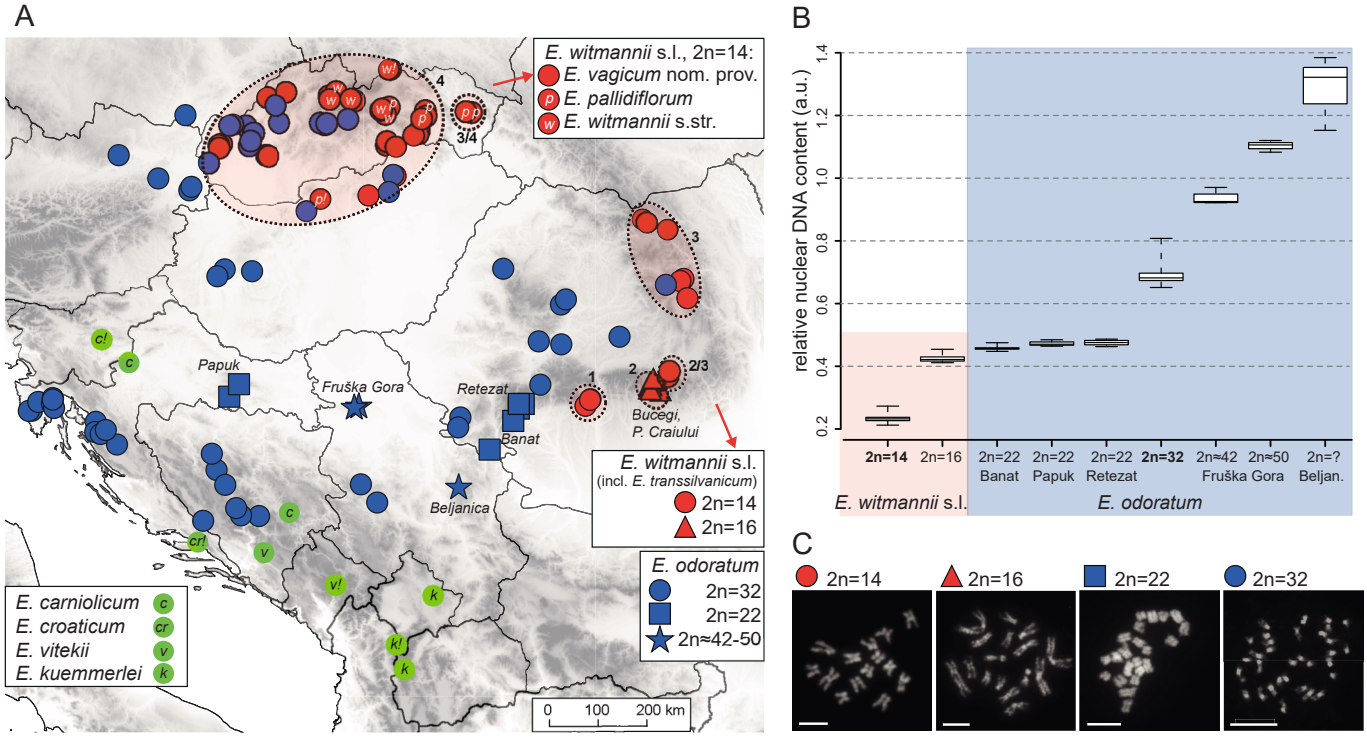

Fig. 1. Sample sites and cytotype variation of the studied *Erysimum* species. **A**. Sample sites of *E. odoratum*, *E. witmannii* s.l., and the Balkan species assigned to the *E. odoratum* species complex (except of *E. pectinatum* from the Peloponnese, see Supplementary Table S1). Encircled populations within *E. witmannii* s.l. and labelled as 1–4, 2/3 and 3/4 correspond to the genetic clusters as resolved by RADseq (Fig. 2). In the Western Carpathians, the population assignment to three putative species of *E. witmannii* s.l. follows Michalková (1999, 2002). Exclamation marks indicate the type localities. For details, see Supplementary Table S1. **B.** Relative genome size variation (sample/standard ratio, in arbitrary units, a.u.) assessed by flow cytometry. The two common cytotypes, 2n = 14 for *E. witmannii* s.l. and 2n = 32 for *E. odoratum*, are marked in bold. **C.** Mitotic chromosomes of selected accessions. *E. witmannii* s.l.: 2n = 14 (S Carpathians, loc. Bist, scale bar 10 μm), 2n = 16 (S Carpathians, loc. Damb, scale bar 10 μm). *E. odoratum*: 2n = 22 (Banat, loc. Ciuc, scale bar 5 μm), 2n = 32 (W Carpathians, loc. Mod, scale bar 5 μm). For accession codes, see Supplementary Table S1.

(Hendriks et al., 2023), were used as outgroups in phylogenetic reconstructions.

2.2. Flow cytometric measurements of nuclear DNA content

Relative nuclear DNA content was estimated by flow cytometry (Doležel et al., 2007a,b) using the CyFlow Space instrument (Partec, Münster, Germany) equipped with a UV-LED excitation source. Usually four individuals were measured per each sampled population, but this number was increased when we observed heterogeneity in the measured values (see Supplementary Table S1). Silica gel-dried leaf samples were co-processed with leaf tissue of *Solanum lycopersicum 'Stupické polní rané'* (2C = 1.96 pg; Doležel et al., 1992) or *Solanum pseudocapsicum* (2C = 2.59 pg; Temsch et al., 2010) as an internal standard following a modified protocol using Otto buffers (Doležel et al., 2007b). Leaf tissue of the sample and the standard were co-chopped in 1 ml of ice-cold Otto I buffer (0.1 M citric acid, 0.5 % Tween 20), filtered through 42-μm nylon mesh, and the filtrate was mixed with 1 ml of a solution containing Otto II buffer (0.4 M $Na_2HPO_4 \cdot 12H_2O$) and DAPI (4 μg/ml) fluorochrome. The fluorescence intensity of 5,000 stained nuclei was measured, and flow cytometric histograms were analysed using FloMax software v. 2.7d (Partec, Münster, Germany). Relative nuclear DNA content (2C value, expressed in arbitrary units; a.u.) was determined based on the ratio of the G0/G1 peak positions between the sample and the internal standard. Coefficients of variation (CV) were calculated for both the sample and the standard, with only histograms exhibiting CV values below 5 % threshold being accepted.

2.3. Chromosome counting

Chromosomes were counted in 43 selected populations (see Supplementary Table S1) chosen to represent the different genome size categories detected within the species studied to ensure that they were correctly assigned to the different cytotypes and ploidies. Mitotic chromosomes were counted from young anthers (collected and fixed either in the field or from plants raised from seeds) or from root tips (harvested from seedlings or cultivated plants). Chromosome spreads were prepared following Mandáková and Lysak (2016), stained with DAPI (2 mg/ml) in Vectashield antifade, and analysed under a Zeiss Axioimager Z2 epifluorescence microscope equipped with a CoolCube camera (MetaSystems).

2.4. RADseq and Hyb-Seq libraries

Two next-generation sequencing methods were applied, double-digest RAD sequencing (ddRADseq, Andrews et al., 2016) and Hyb-Seq, a technique that combines target enrichment with genome skimming (Weitemier et al., 2014). While RADseq, which yields many short loci and a high number of single nucleotide polymorphisms (SNPs) distributed across genomes, is best suited for analysing relationships between recently diverged taxa, detecting fine-scale genetic structure within species, as well as introgression signals, target enrichment generates longer loci that enable allele phasing, gene tree and species tree estimation. The latter is particularly advantageous for phylogenetic inferences in groups affected by reticulation and polyploidy. The combination of these two techniques appears to be efficient in elucidating recent but highly reticulate evolutionary processes (Karbstein et al., 2022). Here, we applied RADseq to each sampled population, analysing one to four individuals per population, while a selection of 34 samples (see Supplementary Table S1) was analysed for Hyb-Seq, specifically to examine the polyploid origin of *E. odoratum*.

Total genomic DNA was extracted from silica gel-dried leaves using the Exgene Plant SV kit (GeneAll, Seoul, Korea) and purified with 0.4x AMPure XP beads (Beckman Coulter, CA, USA). The double-digest RADseq libraries were prepared according to the protocol of Peterson et al. (2012) with some modifications as described in detail in Šlenker et al. (2024). The libraries were sequenced with 150-bp paired-end reads on Illumina systems at Novogene, Cambridge, UK.

The Hyb-Seq libraries were prepared using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, MA, USA) following the manufacturer's protocol. Genomic DNA (250 ng per accession) was fragmented enzymatically, followed by end repair and dA-tailing with a single enzyme mix included in the kit. After the adaptor ligation and clean-up with the QIAquick PCR Purification Kit (Qiagen), size selection was performed by SPRIselect beads (Beckman Coulter) to produce 500–600 bp long fragments. Amplification was performed with eight cycles of PCR, using index primers from NEBNext Multiplex Oligos for Illumina. The amplified libraries were purified with 0.75x AMPure XP beads (Beckman Coulter) and pooled equimolarly (24 accessions/pool). The pooled libraries were enriched by hybridization with biotinylated RNA baits (using an *Erysimum*-specific bait set, see below for probe design and synthesis) at 65 °C for 26 h, using the myBaits kit and following the protocol v. 5.03 (Daicel Arbor Biosciences, MI, USA). The target-enriched libraries were PCR amplified for nine cycles using the KAPA HiFi HotStart mix (Kapa Biosystems, Cape Town, South Africa), purified with the QIAquick PCR Purification Kit, and pooled with an unenriched library fraction in a 2:1 ratio. After the final purification step with 0.9x AMPure XP beads, the libraries were sequenced with 150-bp paired-end reads on an Illumina MiSeq system at BIOCEV, Czechia.

To develop a custom set of Hyb-Seq probes, the genome and genome annotation of *Erysimum cheiranthoides* (ECHE2.0; Züst et al., 2020; available at https://erysimum.org), the chloroplast genome of the same species (NCBI accession number MN207123.1), and the mitochondrial genome of *Arabidopsis thaliana* (NCBI accession number OX261949.1) were used. The probe design followed the protocols outlined by Weitemier et al. (2014) and Schmickl et al. (2016). Specifically, the coordinates of individual exons were extracted from the genome annotation file and used with Bedtools getfasta (Quinlan and Hallm, 2010) to retrieve the corresponding exon sequences. To minimize the enrichment of multi-copy loci, exons with $\geq$ 90 % sequence similarity were identified using pblat (Wang and Kong, 2019) and removed, retaining only unique sequences. Additionally, only exons with a single genomic match were included to avoid targeting duplicated regions. Exons with $\geq$ 90 % similarity to chloroplast or mitochondrial sequences were also excluded. Final filtering retained 1,004 exons from 955 genes, ranging in length from 695 to 1,197 bp. These exons were compiled as target sequences for probe synthesis. In total, 20,000 biotinylated 80-mer RNA probes were synthesized by Daicel Arbor Biosciences (MI, USA).

2.5. RADseq data processing

Illumina sequencing reads were demultiplexed, quality-filtered, and deduplicated using the FASTX-Toolkit v. 0.0.14 (available at: http://hannonlab.cshl.edu/fastx_toolkit/index.html), fastp v. 0.20.1 (Chen, 2023), and BBTools (https://jgi.doe.gov/data-and-tools/bbtools), respectively. Mapping to the genome of *Erysimum cheiranthoides* (GenBank: GCA_011420285.1) was done using BWA 0.7.5a (Li, 2013), and subsequent variant calling and filtration (following a workflow adapted from GATK's best practices; Van der Auwera et al., 2013) was performed by GATK v. 4.4.0.0 (McKenna et al., 2010), to obtain biallelic SNPs with no more than 20 % of missing genotypes. Moreover, samples with missing genotypes above 50 % were discarded. Unlinked RADseq loci were also identified to allow selection of a single random SNP per locus (unlinked SNPs were required for certain analyses, see below). All procedures followed the workflows described in Šlenker et al. (2024) and Šlenker (2024).

Phylogenetic relationships were inferred by constructing a maximum likelihood (ML) tree in RAxML-NG v.0.9.0 (Kozlov et al., 2019) with inputs in PHYLIP format and the heterozygous genotypes represented as IUPAC ambiguities. The GTR model with Felsenstein's method for

ascertainment bias correction were applied. The VCF file was converted to PHYLIP format using the vcf2phylip.py script (Ortiz, 2019), and invariant sites were removed using the ascbias.py script (available at https://github.com/btmartin721/raxml_ascbias), following the recommendations of Leaché et al. (2015). Branch support and the degree of discordance were further assessed using the quartet sampling method (Pease et al., 2018), which enables distinction between weak support and conflicting signal within a phylogenetic tree. Further insight into the overall genetic structure was obtained using a Bayesian clustering approach implemented in STRUCTURE v. 2.3.4 (Pritchard et al., 2000) and a neighbor-net network in SplitsTree4 (Huson and Bryant, 2006). For the STRUCTURE calculations, we generated 100 datasets, each by randomly selecting a single SNP from every RADseq locus containing at least six SNPs, using the vcf_prune.py script (Šlenker, 2024). The calculations were performed and summarized as described in Šlenker et al. (2021). For the neighbor-net (NN) analysis, Nei's genetic distances (Nei, 1972) were calculated in the R package StAMPP (Pembleton et al., 2013) using R 4.4.0 (R Core Team, 2024).

Furthermore, a Bayes factor species delimitation analysis (BFD*, Leaché et al., 2014; Leaché and Bouckaert, 2018) was performed to statistically validate the genetic clusters within the Carpathian diploids (corresponding to *E. witmannii* s.l. and *E. vagicum*). Marginal likelihoods of species trees were derived using the Path Sampling approach with SNAPP v.1.4.2 (Bryant et al., 2012) and BEAST v. 2.5.0 (Bouckaert et al., 2014). The dataset of unlinked SNPs was used, which was reduced to three samples per genetic cluster (two for the cluster from the Vihorlat Mts., see Results). Analyses were run in eight steps for each model, with 1,000,000 MCMC iterations, sampling every 1,000th, and a burn-in cutoff of 10 %. Competing species delimitation models were ranked by comparing their marginal likelihood estimates and their support was assessed by calculating the Bayes factor (Kass and Raftery, 1995), as suggested by Leaché and Bouckaert (2018). Seven alternative species models were explored, either keeping the Carpathian diploids as one unit or splitting them into two to four entities, taking into account the ML tree, NN and STRUCTURE clustering results.

To infer a species tree, we assembled FASTA sequences for each RADseq locus and sample using GATK's FastaAlternateReferenceMaker tool, with heterozygous SNPs encoded using IUPAC ambiguity codes. The RADseq loci that were at least 150 bp long and comprised at least eight phylogenetically informative sites (5,073 loci in total) were used for gene tree construction in RAxML-NG, followed by species tree estimation in ASTRAL-III (Zhang et al., 2018), as described in the Hyb-Seq data processing section (see below). Here, only diploids were included to avoid potential bias due to allopolyploidy (which was addressed in more detail with the Hyb-Seq data, see below).

Potential introgression events between the analyzed species were examined using the ABBA–BABA and related statistics (Durand et al., 2011) implemented in Dsuite (Malinsky et al., 2021). We calculated the *D*, f4-ratio, and f-branch statistics. For this analysis, we used the topology of the species tree constructed above in ASTRAL-III, with the Carpathian diploids sorted into four entities following the phylogenetic and clustering patterns (see Results).

The RADseq reads were also utilized to obtain data from plastomes. The reads were mapped to the plastome of *E. cheiranthoides* (GenBank accession number MN207123.1), processed, and the ML tree was constructed in RAxML-NG as described above.

2.6. Hyb-Seq data processing

The Hyb-Seq reads were processed using HybPiper v. 2.2.0 (Johnson et al., 2016) to extract consensus sequences of targeted exons. Highly variable sequences (indicative of potential paralogs), where the proportion of SNPs exceeded 5 % were excluded from subsequent processing (identified using HybPhaser; https://github.com/LarsNauheimer/HybPhaser). This filtering resulted in a final dataset of 964 sequences (exons/supercontigs). Consensus sequences were aligned using MAFFT v. 7.450 (Katoh and Standley, 2013), and flanking regions and sites with gaps in more than 25 % of sequences were removed using the R package ips (Heibl, 2008) in R 4.4.0 (R Core Team, 2024). ML trees were inferred in RAxML-NG using the best-fitting substitution models as determined by the ModelFinder function of IQ-TREE v.1.6.12 (Chernomor et al., 2016; Kalyaanamoorthy et al., 2017) based on the Bayesian information criterion. Bootstrap analyses were performed using 500 replicates. For the species tree reconstruction, internal branches with bootstrap support $\leq$ 20 % were collapsed using Newick-Utilities v. 1.6 (Junier and Zdobnov, 2010). The species tree was constructed employing a multispecies coalescent model implemented in ASTRAL-III (Zhang et al., 2018), including computation of local posterior probabilities to evaluate branch support (Sayyari and Mirarab, 2016).

The supercontig sequences of the polyploid accessions of *E. odoratum* were further processed for read-backed phasing to infer allele sequences, as described in detail in Šlenker et al. (2021). We applied four different methods to identify homeologous diploid subgenomes and the most likely parental species or lineages: PhyloSD (Sancho et al., 2022), EPA-ng (Barbera et al., 2019), AlleleSorting (Šlenker et al., 2021), and GRAMPA (Thomas et al., 2017). In the PhyloSD approach, we followed the pipeline outlined by Sancho et al. (2022) with some modifications. Since the pipeline requires a single representative for each diploid genome, we calculated the species tree for each gene tree using ASTRAL-III. Due to unacceptable data loss, we did not discard incongruent diploid skeletons (unlike in Sancho et al., 2022), but rather applied stricter criteria in the "*Bootstrapping Refinement*" step, keeping only the homeologs that were confirmed by at least 20 % of bootstrap replicates. Only the major homeolog-types (those with at least 12–15 % representation in the polyploid genome) were further processed with the "*Subgenome Assignment*" algorithm. The principal coordinate analysis combined with a superimposed minimum spanning tree did not indicate multiple homeologs referring to the same subgenome; therefore, ASTRAL-III was used to infer the final subgenomic tree. In the AlleleSorting approach (applicable to the tetraploids only), alleles were sorted into two homeologs based on sequence divergence, labelled to attribute them to different subgenomes (Šlenker et al., 2021), and treated as independent accessions in the coalescent based species tree inference in ASTRAL-III. EPA-ng (Barbera et al., 2019), a reimplementation of the evolutionary placement algorithm (EPA), performs maximum likelihood-based placement of allelic sequences onto a reference phylogenetic tree. Only placements with a likelihood weight ratio greater than 0.9 were considered significant and subsequently used for species tree inference in ASTRAL-III. Finally, GRAMPA (Gene-tree Reconciliation Algorithm with MUL-trees for Polyploid Analysis) uses an algorithm for counting gene duplications and losses to identify polyploidy events, distinguishing between allo- and autopolyploid, and place them on a phylogeny (Thomas et al., 2017).

2.7. Morphometrics

A selection of 70 populations, representing *E. odoratum* and the Carpathian diploids (*E. witmannii* s.l., *E. vagicum* nom. prov.), was used for a detailed morphological assessment using multivariate morphometrics (Supplementary Table S1). The morphometric samples consisted mostly of 20 individuals per population collected in the field and preserved as herbarium specimens, as well as separately detached and pressed floral parts (one flower per individual, spread out and glued to a sheet of paper with transparent tape to preserve shape and size). In addition, seeds were sown from 21 populations and the reared plants were kept in a common garden experiment to observe morphological traits under the same conditions and to produce a parallel morphometric dataset to the samples collected in the field. When the cultivated individuals reached the flowering stage in the second year of cultivation, one flower per individual was detached and pressed as described above, as well as three cauline leaves from the lower, middle and upper part of

the stem (to avoid shrivelling of the leaves at a later phenological stage), and the plants were further cultivated until the siliques approached the mature stage. The plants were then harvested and preserved as specimens for morphometric measurements. In this way, both floral and fruit characters were measured on the same individuals.

We measured or recorded 32 morphological characters on herbarium specimens, either sampled in the field or cultivated in an experimental garden. In addition, ratio characters were derived; in these cases, one of the original characters used for ratio calculation was removed from the final data matrix. The list of all characters is provided in Supplementary Table S2. Size of floral parts was measured using ImageJ v.1.53e software (Schneider et al., 2012). Trichomes were observed using a stereomicroscope (Olympus SZ61) and QuickPHOTO Micro v.3.2 software. Non-parametric Spearman rank correlations were first calculated to detect and avoid pairs of highly correlated traits (applying the threshold of $\rho > 0.85$) that could bias discriminant analyses; in such cases one character from the highly correlated pair was excluded from the final matrices. Since lower cauline leaves are sometimes shed or shrivelled at the flowering stage, two datasets were assembled: matrix 1 that included all sampled individuals but omitting the characters (and ratios) on lower cauline leaves (1,139 individuals x 26 characters), matrix 2 that included only specimens with retained lower cauline leaves and comprised all characters (929 individuals x 34 characters). In addition, a population-level dataset (matrix 3, 66 population samples x 35 characters) and a dataset of cultivated specimens (matrix 4, 233 individuals x 39 characters) were generated. Data subsets from matrices 1, 2 and 4 were also assembled for specific partial analyses (see Results). We performed cluster analyses, principal component analyses (PCA) and canonical discriminant analyses (CDA) using the MorphoTools2 package (Šlenker et al., 2022) in R v.4.0.0 (R Core Team, 2020).

3. Results

3.1. Chromosome numbers and genome size variation

For the populations of *E. witmannii* s.l. (Carpathians) and *E. vagicum* nom. prov. (Western Carpathians) we confirmed the diploid cytotype with 2n = 14 and assessed the relative DNA content (relative genome size, RGS) ranging from 0.21 to 0.27 a.u. (0.23 ± 0.008, sample/standard ratio, with *Solanum lycopersicum* as the primary standard). A few DNA triploids were also detected in some populations of *E. witmannii* s.l. (Supplementary Table S1). In contrast, for the populations from the Munţii Bucegi and Piatra Craiului Mts. in the Southern Carpathians, we found 2n = 16 chromosomes and the RGS of 0.41–0.45 a.u. (Fig. 1).

In the vast majority of the sampled populations of *E. odoratum* we found the tetraploid cytotype with 2n = 32 and the relative DNA content ranging from 0.65 to 0.81 a.u. (0.69 ± 0.02). In some regions, however, we recorded some different cytotypes as follows (Fig. 1). Populations from the Papuk Mts. (northeastern Croatia, populations Malis, Bacin), Banat Mts. (southwestern Romania, pop. Ciuc), and Retezat Mts. (southwestern Romania, pop. Ret, Sco, Cam) had 2n = 22 with RGS = 0.46–0.49 a.u. Populations from the Fruška Gora Mts. (northern Serbia) had 2n≈42 (pop. Glav) and 2n≈50 (pop. Vgrad; the precise number was difficult to determine in both cases due to fragile chromosomes, requiring further cytogenetic analyses) with the RGS of 0.92–0.97 a.u. (pop. Glav) and 1.08–1.12 a.u. (pop. Vgrad), probably representing hexaploids, alike the population from the Homoljske planine Mts. in eastern Serbia (pop. Beljanica, RGS = 1.15–1.38 a.u., chromosomes not counted).

As for the Balkan species of the complex, we found two cytotypes in both *E. carniolicum* and *E. pectinatum*, 2n = 12 and 2n = 16, but with different RGS values. In *E. pectinatum*, the 2n = 12 cytotype had RGS = 0.23 a.u. and that with 2n = 16 had RGS = 0.41–0.43 a.u. In contrast, *E. carniolicum* exhibited higher RGS value: 0.43–0.51 a.u. for the cytotype with 2n = 12 and 0.62–0.65 a.u. for 2n = 16. In both *E. vitekii* and *E. kuemmerlei* we found so far only the cytotype with 2n = 12 and RGS = 0.21–0.23 a.u. *Erysimum croaticum* was confirmed as hexaploid with 2n≈38–40 and RGS = 0.88–0.95 a.u. (Supplementary Table S1).

3.2. Phylogenetic relationships, species delimitation and introgression signals inferred from RADseq data

Sequencing of the RADseq libraries yielded 1,024,119 to 22,815,642 paired reads per sample, with an average of 4,652,345.09 reads retained after quality control and deduplication. Of these, 40.42 % to 97.43 % were successfully mapped to the reference genome, representing an average mapping rate of 89.9 %. The final VCF file comprised 694,358 SNPs, distributed across 18,870 presumably unlinked genomic regions (RADseq loci).

The maximum-likelihood (ML) tree showed that the species attributed to the *E. odoratum* complex do not form a monophyletic lineage but are split into three clades (Fig. 2A, Supplementary Fig. S2): 1. hexaploid *E. croaticum* is resolved as sister to Balkan perennial diploids *E. linariifolium* and *E. microstylum* with high support (bootstrap support, BS = 98 %; quartet concordance, QC = 0.5; quartet differential, QD = 0.33); 2. *E. carniolicum, E. vitekii* and *E. kuemmerlei* form a weakly supported clade (BS = 72 %, QC = 0.18, QD = 0) and together with the branch of *E. pectinatum* (receiving, however, only BS = 32 %, QC = 0.048, QD = 0) being sister to Alpine-Balkan (*E. sylvestre*) and Carpathian-Balkan (*E. comatum*) perennial diploids; and 3. *E. odoratum, E. witmannii* s.l. and *E. vagicum* nom. prov. form a monophyletic clade (BS = 100 %, QC = 0.69, QD = 0.27), with the latter two placed in a nested monophyletic subclade (BS = 100 %, QC = 0.86, QD = 0.12), resulting in paraphyletic *E. odoratum*. Clustering within *E. odoratum* largely follows geographic patterns, with the minor cytotype 2n = 22 found in three nested positions (each receiving BS = 100 %, QC = 0.83–0.98, QD = 0.3–0.71) among the populations of the widespread 2n = 32 cytotype. The hexaploids were resolved at the base of the clade, in congruence with their marginal positions in the neighbor-net (NN, Fig. 2B). Within the clade of the Carpathian diploids (see also Supplementary Fig. S2A), i.e., comprising *E. witmannii* s.l. and *E. vagicum* nom. prov., four clades (or clusters, nr. 1–4) were consistently delimited in ML (each with BS = 100 % and QC = 0.43–1), NN as well as in STUCTURE analyses (at K = 4, Fig. 2C). Furthermore, two other clades (clusters) were found in intermediate positions in ML and NN (denoted as 2/3, 3/4, with BS = 100 % and QC = 0.98–1; but their placement in the tree receiving very low QC and QD, Supplementary Fig. S2A) and genetically admixed in STRUCTURE. All six clades can be geographically well delimited as follows: 1. Southern Carpathian populations with 2n = 14 (Buila-Vânturariţa Mts.), 2. Southern Carpathian populations with 2n = 16 (Bucegi, Piatra Craiului), 3. Eastern Carpathian populations (Haşmaş, Maramureş, Ştânişoarei Mts.), 2/3. Eastern Carpathian populations from the Postăvarul Mts. (being geographically close to clade 2), 4. Western Carpathian populations, and 3/4. Eastern Carpathian populations from the Vihorlat Mts. (being geographically close to clade 4). The three diploid taxa recognized within the Western Carpathians, i.e., the pale yellow-flowered *E. witmannii* s.str. and *E. pallidiflorum* s.str., and the bright yellow-flowered *E. vagicum* nom. prov., did not form monophyletic clades; in contrast, all the diploid populations clustered by mountain ranges (Supplementary Fig. S2A).

Bayes factor delimitation analyses (BFD*) of the Carpathian diploids based on RADseq data (10,044 SNPs) computed in SNAPP supported the model with four entities, i.e. splitting them into the Southern Carpathian with 2n = 14 (corresponding to clade 1, see Fig. 2), Southern Carpathian with 2n = 16 (clade 2), Eastern Carpathian (including the Vihorlat samples, i.e. comprising three clades: 3, 3/4 and 2/3), and the Western

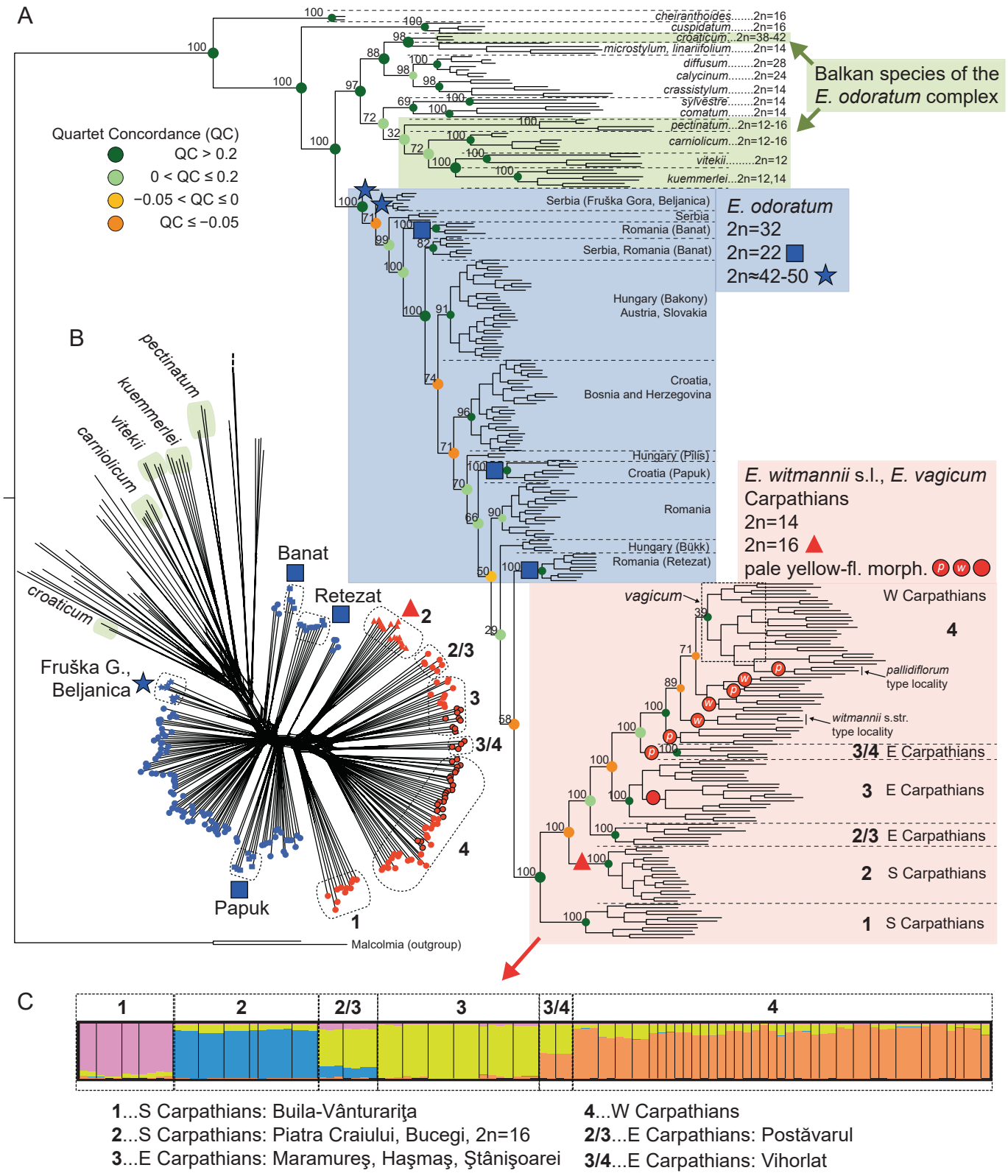

Fig. 2. Phylogenetic reconstruction in the *Erysimum odoratum* complex and related species based on RADseq data. **A.** Maximum-likelihood tree inferred in RAxML. Branch support (major branches only) is indicated by bootstrap and quartet concordance values. Species assigned to the study complex are highlighted in colours. Symbols indicate minor cytotypes or pale yellow-flowered morphotypes, in the Western Carpathians assigned to *E. pallidiflorum* (with the letter 'p' in the circle) or *E. witmannii* s.str. ('w' in the circle). The chromosome numbers of the species of the *E. odoratum* complex follow Supplementary Table S1, those outside the complex correspond to the data by Polatschek (1997, 2013). **B.** Neighbor-net (NN) analysis based on Nei's genetic distance. Symbols and clade numbering as in A. **C.** Bayesian clustering based on unlinked SNPs of Carpathian diploids at K = 4. Vertical lines separate different populations. The coloring in the graph indicates the sample assignment to the genetic clusters. The cluster numbering as in A and B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Results of species delimitation in the clade of Carpathian diploids, comprising *Erysimum witmannii* s.l. and *E. vagicum* nom. prov., using Bayesian factor delimitation analysis (BFD*) in SNAPP based on unlinked SNPs of RADseq data. The model of a single entity comprising all diploid populations from the Carpathians (model 0) is tested against six other models, in which populations are divided into two to four entities based on geographic and cytotype patterns and genetic clustering (Figs. 1, 2). W. Carp.: Western Carpathians; SE Carp.: Southeastern Carpathians; E. Carp.: Eastern Carpathians; S. Carp.: Southern Carpathians. The models also include a different assignment of the genetically intermediate population from the Vihorlat Mts. (biogeographically part of the Eastern Carpathians, but geographically close to the Western Carpathians). The model with the highest support is highlighted in bold. BF: Bayes factor; MLE: marginal likelihood estimate.

Model	No. of entities	MLE	BF
Carpathians	1	−116509.21	0
W Carp. (incl. Vihorlat Mts.), SE Carp.	2	−114876.83	3264.77
W Carp., SE Carp. (incl. Vihorlat Mts.)	2	−116651.15	283.88
W Carp. (incl. Vihorlat Mts.), E Carp., S Carp.	3	−114547.38	3923.65
W Carp., E Carp. (incl. Vihorlat Mts.), S Carp.	3	−114710.79	3596.84
W Carp. (incl. Vihorlat Mts.), E Carp., S Carp. (2n = 14), S Carp. (2n = 16)	4	−112242.82	8532.78
W Carp., E Carp. (incl. Vihorlat Mts.), S Carp. (2n = 14), S Carp. (2n = 16)	**4**	**−112105.57**	**8807.28**

Carpathian entity (clade 4). The model with the Vihorlat samples (clade 3/4) assigned to the Western Carpathian entity (4) received only a slightly lower support than the one with the highest score. The models with fewer entities, i.e. one, two or three, received much lower support (Table 1).

The ASTRAL-III species tree (Fig. 3A), derived from 5,073 RADseq loci for diploids only (i.e. omitting *E. diffusum, E. croaticum* and *E. odoratum*), also showed non-monophyly of the *E. odoratum* complex. In accordance with the ML tree (Fig. 2A), the three western Balkan species of the complex, *E. carniolicum, E. vitekii* and *E. kuemmerlei*, were found to be closely related (albeit with low local posterior probability, LPP = 0.6) and in a sister position to the perennials *E. sylvestre* and *E. comatum* (with LPP = 0.55). On the other hand, *E. pectinatum* from Peloponnese was resolved in a different clade (LPP = 0.92) compared to the ML tree (see Fig. 2A with BS = 32 %). The Carpathian diploids (*E. witmannii* s.l. and *E. vagicum* nom. prov.) were supported as a lineage distinct from the Balkan species (LPP = 1).

ABBA-BABA tests computed for diploids only revealed significant introgression signals between several species both within and across the clades (Supplementary Table S3, Fig. 3B). As for the Balkan members of the *E. odoratum* complex, the f-branch statistics suggested gene flow between *E. carniolicum* and *E. sylvestre*, between the species pair *E. vitekii* and *E. kuemmerlei* and *E. linariifolium*, as well as between *E. carniolicum* and *E. vitekii* (Fig. 3B). In all cases, these taxa show parapatric to sympatric distribution patterns (Fig. 1, Supplementary Fig. S1). Nevertheless, these gene flow inferences may also be biased by the weak statistical support of (and within) this clade (Figs. 2A, 3A). Among the Carpathian diploids of the complex (*E. witmannii* s.l. and *E. vagicum* nom. prov.), we detected the strongest signal of introgression between the Southern Carpathian populations with 2n = 14 and the internal branch comprising the Eastern and Western Carpathian populations (both 2n = 14), between the Southern Carpathian diploids with 2n = 16 and the Eastern Carpathian ones (2n = 14; Fig. 3B), as well as with other species outside the species complex, in particular with the parapatric *E. comatum* (Supplementary Table S3A). ABBA-BABA tests that also included polyploid *E. odoratum* (Supplementary Table S3B, Supplementary Fig. S3) revealed signals of introgression mainly with *E. comatum* and *E. witmannii* s.l. (all lineages with 2n = 14, incl. *E. vagicum* nom. prov.), both of which grow in parapatry with *E. odoratum* (Supplementary Fig. S1).

3.3. Phylogenetic relationships and the origin of polyploids reconstructed from Hyb-Seq data

Sequencing of the Hyb-Seq libraries resulted in, on average, 721,535.18 quality-filtered reads per individual, out of which 55.77 %, on average, were mapped to the 1,004 targeted nuclear sequences. After excluding sequences with a variability exceeding 5 %, we successfully assembled 964 exons (supercontigs) ranging from 695 to 1,412 bp, with an average length of 1,072.14 bp.

The ASTRAL-III species tree inferred from the gene trees based on consensus sequences (Fig. 4A) showed the topology largely congruent with the RADseq-based trees (Figs. 2A, 3A), with a few discrepancies as described below. As in the RADseq tree, the monophyly of the *E. odoratum* species complex was refuted, as its members were split into three different clades. Three western Balkan representatives of the complex, *E. carniolicum, E. vitekii* and *E. kuemmerlei*, were resolved in a well-supported clade (LPP = 1) but including also the Alpine-Balkan perennial diploid *E. sylvestre*. The branch of the Carpathian-Balkan perennial *E. comatum* attached to this clade as well but with a lower support (LPP = 0.78). The southern Balkan member of the species complex, *E. pectinatum* from the Peloponnese, alike the species tree of RADseq loci (Fig. 3A) but in contrast to the ML tree of RADseq data, was resolved as sister (LPP = 0.66) to the clade of Balkan diploid perennials *E. linariifolium, E. microstylum* and *E. crassistylum*. Finally, the Carpathian diploids (*E. witmannii* s.l. and *E. vagicum* nom. prov.) and the polyploid *E. odoratum* formed a distinct clade (LPP = 1.0).

The PhyloSD and EPA-ng algorithms showed very similar results in the assignment of alleles identified in the polyploids of *E. odoratum*. Both analyses unambiguously classified between 25 % and 40 % of the alleles (on average 34.5 %, see Supplementary Table S4). The highest proportions of classified alleles, ranging from 33–40 % (in the hexaploids from the Fruška Gora Mts.) to 60–65 % (in the cytotype with 2n = 22 from the Papuk Mts.), were assigned to the 'witmannii' type (Supplementary Table S4, Fig. 4B). In the ASTRAL trees, these witmannii-like homeologs were resolved in the sister position to the branch of *E. witmannii* s.l. (Fig. 4C), supporting this species as one of the parental species of polyploid *E. odoratum*. The other major-type alleles, those reaching a proportion of more than 12–15 % up to 24 %, were assigned either to the 'cuspidatum' or the 'comatum' type, and placed in the sister positions to the branches of these two species in the ASTRAL trees (Fig. 4B,C). Nevertheless, the boundary between the major and minor homeolog-types was not clear, as 2–11 % of the homeologs (depending on the accession) were assigned to a broader range of diploids (Fig. 4B, Supplementary Table S4).

In the AlleleSorting procedure, the optimal threshold for unequivocal allele sorting in *E. odoratum* ('between homeolog distance', for details see Šlenker et al., 2021 and https://github.com/MarekSlenker/AlleleSorting; applicable to the tetraploids only, therefore, omitting the hexaploids here), was found to be 2.5. On average 8.92 % of exon sequences passed the threshold and their alleles could be unambiguously sorted into homeolog A (alleles closer to the clade of *E. witmannii* s.l., one of the suspected progenitors based on the topology of the species tree inferred from the consensus sequences and the PhyloSD and EPA-ng results described above) and homeolog B. In the ASTRAL species tree, the A homeologs of all analyzed accessions were resolved in the sister position to *E. witmannii* s.l., while the positions of the B homeologs differed as follows (Fig. 4C, Supplementary Fig. S4). The B homeolog of the common tetraploids (2n = 32, represented by the accessions Kavlar and Breit) was placed as sister to *E. cuspidatum*, while that of the cytotype with 2n = 22 from the Banat Mts. (accession Ciuc) was resolved as sister to *E. comatum;* both these placements are congruent with the PhyloSD and EPA-ng results and support independent allopolyploidization events for these two cytotypes (Fig. 4C). In contrast, the B homeologs of the accessions from the Retezat (accessions Ret, Sco, 2n = 22) and Papuk Mts. (accession Malis, 2n = 22) Mts. were placed as sister to the clade comprising the A homeolog and *E. witmannii* s.l., favouring

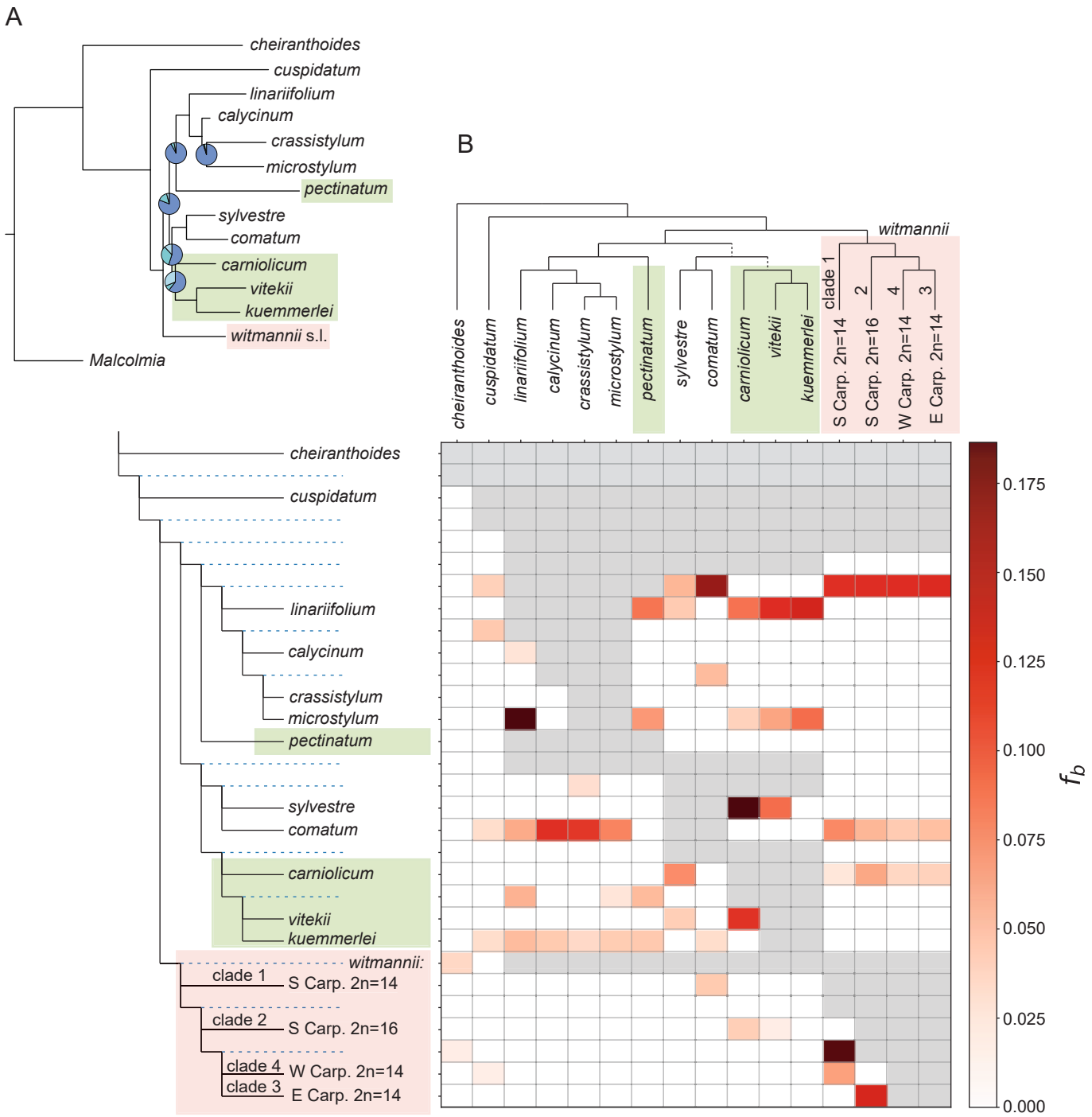

Fig. 3. Phylogenetic and gene flow inferences in the *Erysimum odoratum* complex and related species (diploids only) based on RADseq data. **A.** Species tree inferred in ASTRAL-III from 5,073 RADseq loci at least 150 bp long and comprising at least eight phylogenetically informative sites. Species of the *E. odoratum* complex are highlighted in colours. Pie charts depict three local posterior probabilities (LPP) for the given branches (not shown for the fully supported branches when the local LPP equals 1). **B.** Results of f-branch analysis computed in Dsuite. The split of *E. witmannii* s.l. into four lineages follows the phylogenetic and clustering patterns resolved congruently by maximum-likelihood, neighbor-net and STRUCTURE analyses (clades/clusters 1–4 as in Fig. 2), omitting the genetically admixed populations (clades denoted as 2/3 and 3/4). Coloured squares indicate a significant excess of SNPs sharing between the respective branches caused by gene flow. White and grey squares represent non-significant values or trios incompatible with the phylogenetic tree. Note that two branches marked by dashed lines received low support in the species tree.

autopolyploid origins (Supplementary Fig. S4). Autopolyploidy was indicated for the accession from the Papuk Mts. also by the PhyloSD and EPA-ng approach, but not for the accessions from the Retezat Mts. (Fig. 4C).

Using the GRAMPA algorithm (Fig. 4C, Supplementary Fig. S5), the cytotype from the Papuk Mts. (accession Malis, 2n = 22) was resolved as an autopolyploid of *E. witmannii* s.l., while all other accessions of *E. odoratum* were identified as allopolyploids, with *E. witmannii* s.l. being one of the parents, while the others were either *E. cuspidatum* (for the common tetraploid cytotype 2n = 32 as well as the hexaploids from the Fruška Gora Mts.) or *E. comatum* (for the cytotype with 2n = 22 from the Banat Mts. and the Retezat Mts.).

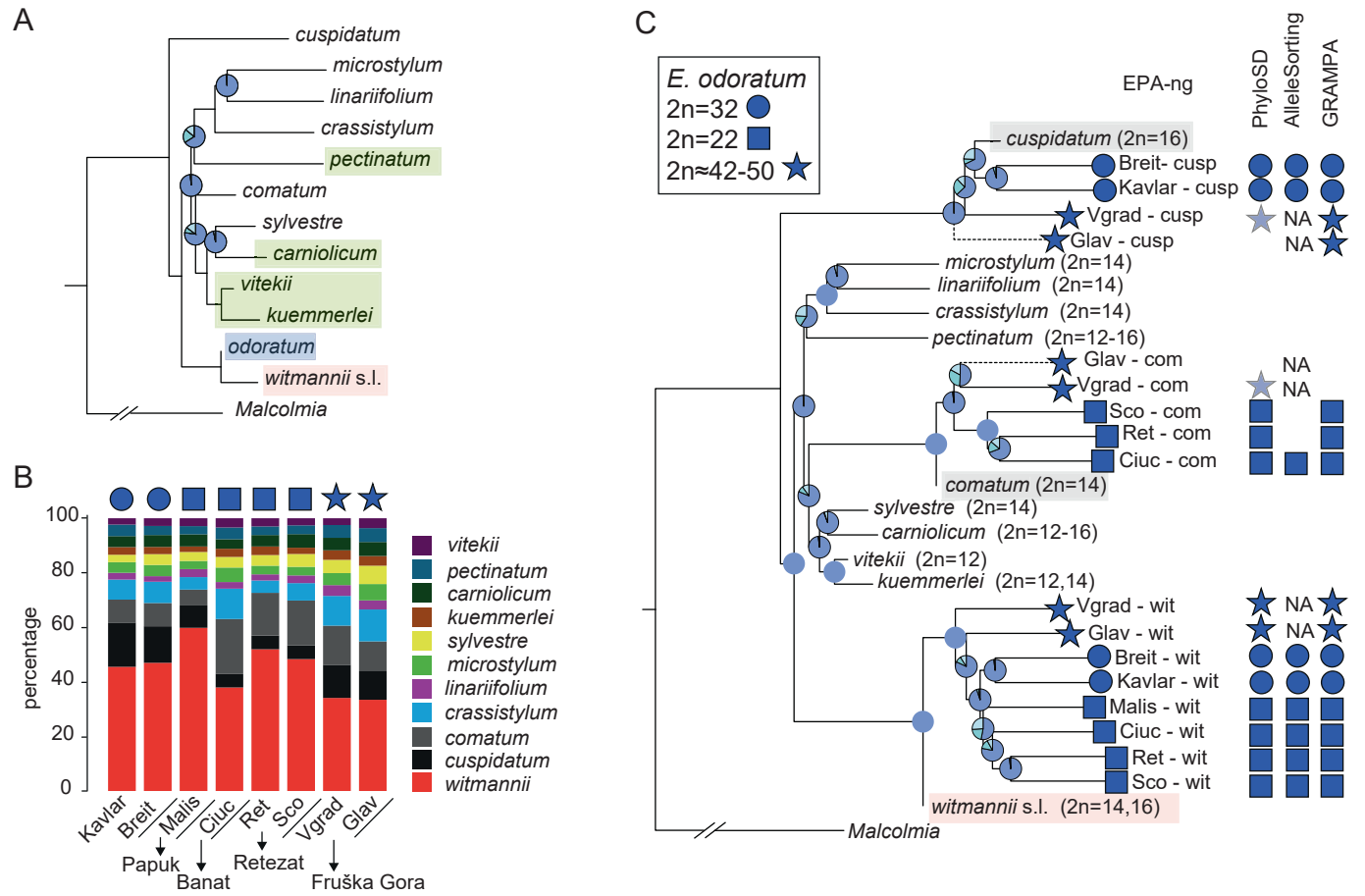

Fig. 4. Phylogenetic reconstruction in the *Erysimum odoratum* complex and related species based on Hyb-Seq data. **A**. Species tree inferred in ASTRAL-III from 964 gene trees based on consensus sequences. Species of the *E. odoratum* complex are highlighted in colours. **B**. The bar diagrams show the proportion of different homeolog-types identified in the polyploid accessions of *E. odoratum* using the modified PhyloSD pipeline, after the "Bootstrapping Refinement" algorithm. Here, the phased exon sequences of polyploids are classified into homeolog-types according to their position with respect to the skeleton of diploid species. **C**. ASTRAL-III species tree summarizing gene trees generated using the phylogenetic placement algorithm implemented in EPA-ng. The homeologs are labeled as 'cusp', 'com', 'wit' according to their phylogenetic positions. Homeolog placements using the PhyloSD, AlleleSorting (not applicable for hexaploids, NA) and GRAMPA tools (see Supplementary Figs. S4 and S5 for details) are indicated by the symbols depicted on the right side of the EPA-ng tree. Only the major homeolog-types with the frequency above 12% (but those below 15% indicated by dashed branches in EPA-ng and light blue symbols in PhyloSD) were processed. Pie charts in both A and C depict three local posterior probabilities (LPP) for the given branches (not shown for the fully supported branches when the local LPP equals 1). The indicated chromosome numbers follow Supplementary Table S1 and Polatschek (1997, 2013). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Overall, these results suggest multiple allopolyploid origins of *E. odoratum*, with probably only the populations from the Papuk Mts. (pop. Malis, 2n = 22) to be of an autopolyploid origin (Fig. 4C).

3.4. Plastome phylogeny

Phylogenetic reconstructions based on plastome data retrieved from Hyb-Seq and RADseq reads were fully congruent with each other, but showed strong discordance with the results of the nuclear data. Three major clades could be delimited in the plastome trees, but each included accessions from different unrelated species, and in turn conspecific accessions were often split into different clades (Supplementary Fig. S6). Overall, the topology of the plastome tree lacked any taxonomic or geographic structure.

3.5. Morphometrics

3.5.1. Morphological differentiation between the diploids and tetraploids

Cluster analysis and PCA based on population means of 66 field-sampled and morphometrically analysed populations (matrix 3) indicated morphological distinction between *E. odoratum* (tetra- and hexaploids) and the diploids comprising *E. witmannii* s.l. and *E. vagicum* nom. prov., albeit with overlaps (Supplementary Fig. S7A, B), which was further explored using discriminant analyses (CDA) based on individual plants (matrix 1 and 2). CDA with two predefined groups, tetraploids (*E. odoratum*, including the common cytotype 2n = 32 and the cytotype 2n = 22 from the Papuk Mts., but excluding the hexaploid populations and that from the Retezat Mts., see below) and diploids (*E. witmannii* s.l. and *E. vagicum*) showed morphological differentiation between these two groupings, but still with some overlap (Fig. 5A). The characters with the highest values of total canonical structure (Supplementary Table S5), contributing best to this distinction, were as follows: number of leaves, length of trichome rays, petal blade/petal length ratio (all three with the higher values in the tetraploids), the leaf length/stem height ratio and petal length (the higher values in the diploids, Supplementary Table S6). CDA of the same two predefined groups but based on the cultivated plants (matrix 4) showed a similar pattern and supported the same diagnostic morphological characters, except of the leaf length/stem height ratio (Supplementary Fig. S7D, Supplementary Table S6). The leaf length/stem height ratio, nevertheless, shows many

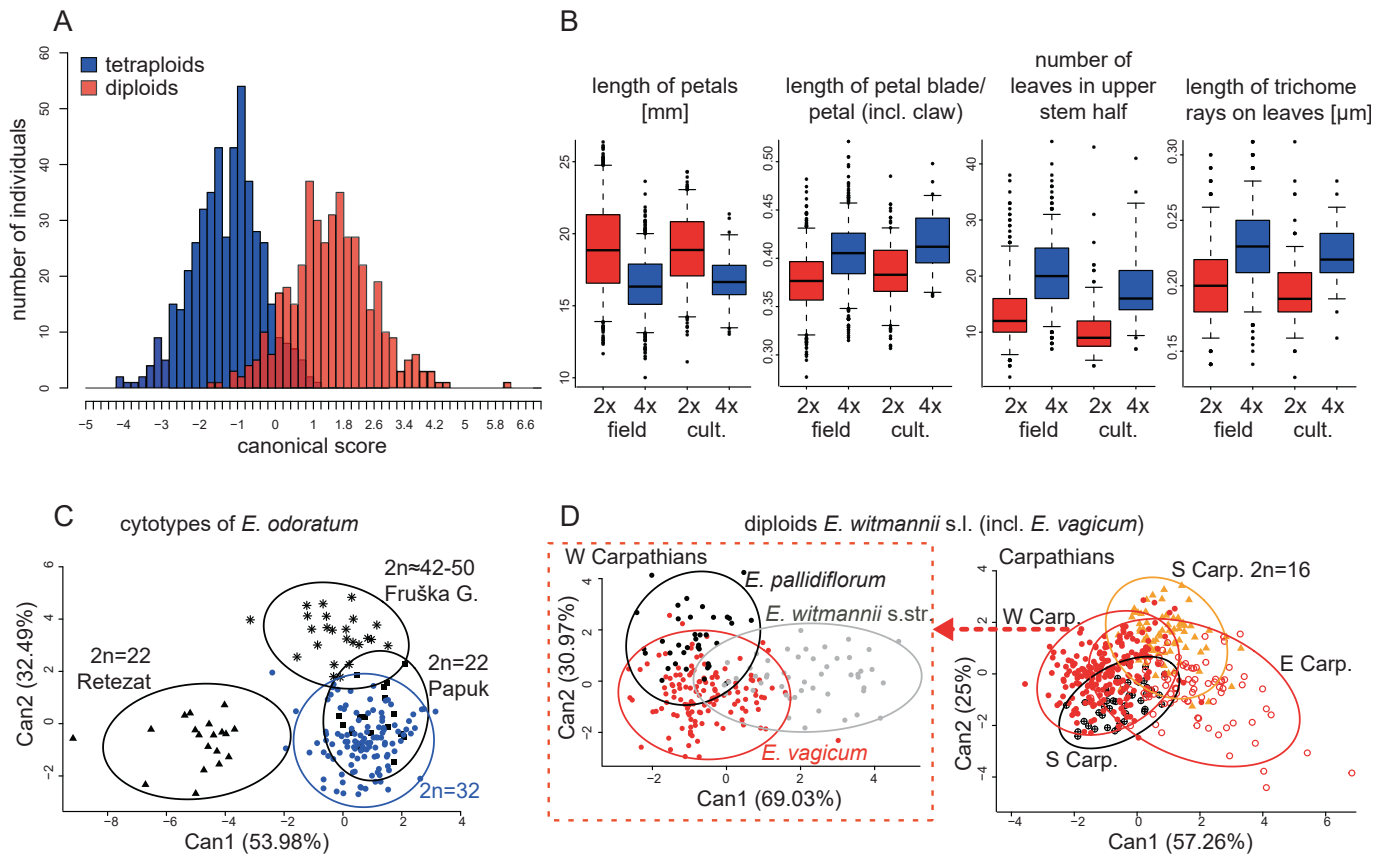

Fig. 5. Morphometric analyses of the studied *Erysimum* species. **A.** Canonical discriminant analysis (CDA) based on 886 field-collected herbarium specimens and 34 characters of matrix 2 with two predefined groups: tetraploids of *E. odoratum* (including the cytotypes 2n = 32 and 2n = 22) and diploids of *E. witmannii* s.l. (incl. *E. vagicum* nom. prov.). **B.** Variation in selected morphological characters shown in box-and-whiskers plots that best discriminate between the tetraploids and diploids (as shown in A), including field-collected specimens (field) and those grown from seeds and cultivated in an experimental garden (cult.). **C.** CDA based on 174 field-collected herbarium specimens and 34 characters of matrix 2, including only the polyploids of *E. odoratum*, with four predefined groups as depicted. Prediction ellipses define the regions where any new independent observations from the respective group will fall with the probability of 95 %. **D.** CDA based on 414 field-collected herbarium specimens and 34 characters of matrix 2, including all the diploids from the Carpathians, with four predefined groups following geography and RADseq results (genetically intermediate populations from the Vihorlat Mts. are classified here into the Western Carpathian group). The Western Carpathian populations (215 field-collected specimens) are also analysed in a separate CDA (red-framed graph) with three predefined groups following traditional taxonomic treatment: *E. witmannii* s.str., *E. pallidiflorum* and *E. vagicum* nom. prov. Values of total canonical structure of all four CDA are listed in Supplementary Table S5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

outlying values of field-sampled individuals, decreasing its reliability. While the length of trichome rays, length of petals, and petal blade/petal length ratio remain fairly stable when the field-sampled and cultivated plants are compared (Fig. 5B), the number of stem leaves is markedly shifted to lower numbers in the cultivated plants, but the differences between the diploids and tetraploids (and thus the diagnostic value of this character) are retained (Fig. 5B).

3.5.2. *Morphological differentiation between the polyploid cytotypes of E. odoratum*

CDA focusing on *E. odoratum* only, with four predefined groups corresponding to different cytotypes – the common 2n = 32 cytotype (selection of eight populations to keep the group sizes balanced), the cytotype 2n = 22 from the Papuk Mts., the cytotype 2n = 22 from the Retezat Mts., and hexaploids from the Fruška Gora Mts. – showed that the Retezat population and the hexaploids are differentiated from the common tetraploid cytotype (2n = 32) along the first and second canonical axes, respectively, while the 2n = 22 cytotype from the Papuk Mts. is not (Fig. 5C). Total canonical structure values suggest that the population from the Retezat Mts. is differentiated by having fewer stem leaves, higher leaf length/stem height ratio, the stem leaves are more densely hairy and have deeper teeth on margin when compared with the common tetraploids (Supplementary Tables S5 and S6). The hexaploids from the Fruška Gora Mts. also have higher leaf length/stem height ratio, but differ by fewer trichomes on leaves, longer trichome rays, longer petal blade in ratio to the total petal length and narrower petals (Supplementary Table S6). When including also the diploids (CDA with four predefined groups, Supplementary Fig. S7C), it appears that both the Retezat and Fruška Gora populations are in some characters intermediate between the diploids and the common tetraploids (e.g., in the number of leaves, and the leaf length/stem height ratio, most correlated with the first axis), but are also shifted along the second canonical axis in opposite directions, in respect of the more hairy and dentate leaves, wider stigma (Retezat) and longer leaves and longer trichome rays (Fruška Gora).

3.5.3. *Morphological differentiation in the Carpathian diploids*

To examine the differentiation patterns within the diploids, first we computed CDA based on field-collected samples with four pre-defined groups, following their genetic clustering and being consistent with the geographic origins: Western Carpathian populations (incl. geographically adjacent and genetically intermediate populations from the Vihorlat Mts.), Eastern Carpathian populations in Romania, and the Southern Carpathian populations split into two groups geographically

and cytotypically (2n = 14 vs 2n = 16). The results (Fig. 5D) show large overlap between the four groups, although some tendencies can be recognized. The Eastern Carpathian populations (2n = 14) and the Southern Carpathian ones with 2n = 16 are shifted from the remaining two groups along the first canonical axis, due to the larger floral parts (especially pronounced in the cytotype 2n = 16), narrower and fewer stem leaves, and the higher leaf length/stem height ratio (especially in the Eastern Carpathian populations) (Supplementary Tables S5 and S6). An alternative analysis with the populations from the Vihorlat Mts. classified into the Eastern Carpathian group showed only negligible differences (not shown).

PCA analysis of cultivated diploid individuals showed extensive overlap between the four groups, while CDA indicated that the Romanian (Eastern and Southern Carpathian) accessions are shifted from the Western Carpathian (incl. Vihorlat Mts.) along the first axis, due to having longer silicules, larger siliqua/stem angle, and higher leaf length/stem height ratio (Supplementary Fig. S7E, F).

CDA focusing on the Western Carpathians (incl. Vihorlat Mts.) with three predefined groups corresponding to the traditionally recognized diploid taxa (albeit not supported by the present RADseq data, see above) – *E. witmannii* s.str., *E. pallidiflorum* (both having pale yellow petals) and *E. vagicum* nom. prov. (bright yellow petals) – showed that the former two are best differentiated from each other, but both show much overlap with *E. vagicum* (Fig. 5D, Supplementary Table S6). The analysis shows that *E. witmannii* s.str. differs from *E. pallidiflorum* by larger floral parts and longer silicules and pedicels (the latter two characters shown by the analyses based on cultivated plants, Supplementary Fig. S7F), and, in turn, *E. pallidiflorum* has more dentate leaf margins (in respect of the number and depth of teeth).

4. Discussion

4.1. Polyphyly of the Erysimum odoratum species complex and evidence of multiple introgression events

Erysimum is one of the largest genera in the Brassicaceae family (German and Al-Shehbaz, 2008; Al-Shehbaz, 2012), yet it is surprisingly understudied. The low resolution of conventional phylogenetic markers (Abdelaziz et al., 2014; Moazzeni et al., 2014; Czarna et al., 2016) and the high gene discordance observed in transcriptome data (Züst et al., 2020) – probably a consequence of recent and rapid radiation, incomplete lineage sorting, and reticulate evolution – have hindered deeper evolutionary insights, making it a taxonomically and phylogenetically challenging genus (Osuna Mascaró, 2020). In the absence of a robust phylogenetic framework, it remains uncertain whether the species complexes delimited in the genus primarily on the basis of morphological traits, growth habits and biogeographic patterns (Ball, 1993; Jalas and Suominen, 1994; Polatschek, 2013) represent distinct evolutionary lineages. Recently, phylogenetic analyses have not supported the recognition of the *E. nevadense* group from the southeastern Iberian Peninsula (Abdelaziz et al., 2014). In the present study, it was unsurprisingly found that the *E. odoratum* complex is polyphyletic, too. Morphology alone hardly allows reliable inferences about evolutionary relationships, especially when recent and reticulate speciation events are involved (e.g., Frajman et al., 2016; Đurović et al., 2017; Karbstein et al., 2020; Vasile et al., 2024). The two high-throughput sequencing methods (RADseq, Hyb-Seq) applied here yielded largely concordant phylogenetic patterns, but also some discrepancies that can be attributed to interspecific gene flow (see below). The combination of these two techniques proved to be beneficial: RADseq has been shown to be efficient to study genetic differentiation in recently diverged clades (Boucher et al., 2016; Karbstein et al., 2020, 2022; Ott et al., 2022; Šlenker et al., 2024) as well as to assess introgression (Buono et al., 2021; Carnicero et al., 2023; Yi et al., 2023; Slovák et al., 2023), while target enrichment (Hyb-Seq) enables the clarification of polyploid origins (Morales-Briones et al., 2022; Ren et al., 2024; Španiel et al., 2023).

Despite morphological resemblance of the species assigned to the *E. odoratum* complex, they were divided into at least three phylogenetic clades, two Balkan and one widespread clade comprising polyploid *E. odoratum* and nested Carpathian diploids. Of the Balkan representatives of this species complex, *E. vitekii* and *E. kuemmerlei*, which grow allopatrically in the western Balkans (Polatschek, 2013, and our unpublished data), were resolved as well-supported sister species in both the RADseq and Hyb-Seq trees (Figs. 2A, 3A, 4A). They were also found to be closely related to three other parapatrically distributed species, namely *E. carniolicum* (another member of the target species complex growing in the western Balkans), *E. comatum* and *E. sylvestre* (the latter two not considered part of this complex), which occur in the Balkans and extend to the Southern Carpathians and the Eastern Alps, respectively (Jalas and Suominen, 1994; Polatschek, 2013). However, topological differences between the trees constructed from RADseq and Hyb-Seq data and the low support, especially in the RADseq trees (Figs. 2A, 3A), did not allow precise phylogenetic inferences within this clade. The low quartet concordance (QC = 0.18) and quartet differential (QD = 0) values for the position of *E. carniolicum* (Supplementary Fig. S2B) indicate that this weak support is due to conflict caused by introgression rather than incomplete lineage sorting (see Pease et al., 2018), which was also supported by f-branch analyses suggesting introgression with *E. sylvestre* (Fig. 3B). A similar introgression scenario could apply to another member of the complex, *E. pectinatum*. This endemic of the Peloponnese, morphologically similar to *E. kuemmerlei* and even synonymized in the past (Josifović et al., 1972; Vangjeli, 2017), appeared to be genetically distinct, which supports its separate taxonomic treatment. However, its precise phylogenetic placement remained uncertain: it was resolved either as a sister lineage to the clade of *E. vitekii, E. kuemmerlei* and *E. carniolicum* (in the ML tree of RADseq) or to the clade comprising other Balkan species – *E. microstylum, E. linariifolium,* and *E. crassistylum* (in the species trees of both Hyb-Seq and RADseq data). Low statistical support (Figs. 2A, 4A) in combination with discordant skew (QD = 0, Supplementary Fig. S2B) favours hybridization events, as also indicated by the f-branch statistics of the RADseq data (Fig. 3B).

Introgression signals were detected in several other cases, both within and between clades, providing evidence of pervasive interspecific gene flow in the genus (Fig. 3, Supplementary Fig. S3, Supplementary Table S3). Similarly, recent studies of Iberian endemics indicated very weak pre-zygotic barriers between species, which were consistent with the strong evidence of hybridization and introgression, even across ploidy levels (Osuna Mascaró et al., 2023). Cytonuclear discordance observed for the Iberian species (Osuna Mascaró et al., 2023), as well as in the present study could also reflect past hybridization events, along with the retention of ancestral variation across different species (incomplete lineage sorting), a feature typical of recent and rapid speciation (Lee-Yaw et al., 2019; Meng et al., 2021).

The four Balkan members of the complex discussed above, *E. vitekii, E. kuemmerlei, E. carniolicum*, and *E. pectinatum*, were confirmed to be diploid with 2n = 12 and 2n = 16 (in contrast to the previously reported 2n = 12 and 2n = 14, Polatschek, 1997, 2013) but with highly variable genome size, indicative of complex chromosomal and repetitive DNA dynamics and/or genome duplication followed by rapid diploidization (Bennetzen and Wang, 2014; Mandáková and Lysak, 2018; Wang et al., 2021), pending further research. Hexaploid level was corroborated for *E. croaticum,* a recently described species that was thought to be close to *E. carniolicum* and *E. vitekii* (Polatschek, 2013) but found to be only distantly related to these species. While the Balkan species assigned to the *E. odoratum* complex will be studied in more detail elsewhere (study in progress), we have focused here on the core of this complex – Carpathian *E. witmannii* s.l. and *E. odoratum* s.str., which are discussed in detail below.

4.2. Evolution in the Carpathians: Diploids harbouring great and geographically structured genetic variation but in conflict with the traditional taxonomy

Diploid populations from the Carpathians assigned to the *E. odoratum* species complex formed a monophyletic lineage comprising several clades with allopatric distribution (Fig. 2). However, this well-defined genetic structure is at odds with traditional taxonomic treatments that delimited up to four species in the Carpathians (Jalas and Suominen, 1994; Michalková, 2002; Oprea, 2005). While the circumscription of *E. transsilvanicum*, described from the Southern Carpathians (Schur, 1866), has remained uncertain due to the unclear morphological or geographic delimitation of this putative taxon, particularly in relation to the sympatric *E. witmannii* (Jalas and Suominen, 1994), the other three entities – *E. witmannii* s.str., *E. pallidiflorum* and *E. vagicum* nom. prov. – were studied in more detail in the Western Carpathians (incl. Vihorlat Mts.) from morphological, ecological and chorological aspects (Kliment, 1999; Michalková, 1999, 2002). This enabled us to classify the populations sampled from the Western Carpathians to these three putative species, and to test their genetic and morphological differentiation. We found that the genetic clustering (Fig. 2, Supplementary Fig. S2A) strictly followed geographic origin (mountain ranges), which contradicted their taxonomic classifications. Recent divergence and ongoing homogenizing gene flow may lead to clustering by sampling location rather than taxonomic entity, as recently observed in the genus *Euphrasia* (Ding et al., 2025), but we assume that this is not the case for the *Erysimum* studied here, as argued below. Morphologically, we found some tendencies of differentiation between these three putative species in terms of flower and fruit size and leaf margin, both in the field-sampled and cultivated plants (Fig. 5D, Supplementary Fig. S7E,F), but with much overlap in contrast to Michalková (1999) who reported non-overlapping differences in petal length and width. We assume that the recognition of these three entities in the Western Carpathians was artificial and caused by the limited sampling from a broad morphological and ecological range, characteristic of the Western Carpathian diploid populations, and probably also obscured by the variation in petal colors. Pale yellow-flowered morphotypes likely arose (or were selected for) repeatedly in both the Western and Eastern Carpathians (Fig. 2, Supplementary Fig. S2A) and do not merit separate taxonomic treatment.

Species delimitation analyses calculated under the multispecies coalescence model (BFD*, Table 1) favoured the scenario of four entities in the Carpathians, which essentially follow four major, allopatrically distributed clades (Figs. 1, 2): two in the Southern Carpathians (corresponding to clades nr. 1 and 2, which differ in chromosome number), one in the Eastern Carpathians (polyphyletic, consisting of clade nr. 3, and two smaller clades nr. 2/3 and 3/4 of geographically more distant populations), and one in the Western Carpathians (clade nr. 4). However, we found no morphological support for these four entities (Fig. 5D), suggesting that the observed genetic divergence represents an intraspecific phylogeographic structure shaped by the isolation of populations in glacial refugia followed by gradual postglacial colonization, and the presence of migration barriers in the heterogeneous landscape of the Carpathians (Mráz and Ronikier, 2016, and see below). It is acknowledged that species delimitation inferences can be biased in the presence of some general evolutionary processes such as interspecific gene flow, isolation by distance, gene flow barriers, or selection, either over- or underestimating the number of species, and that other components of the integrative approach beyond genetics need to be considered for proper species delimitation (Leaché and Bouckaert, 2018; Yang et al., 2019; Mason et al., 2020; Karbstein et al., 2024).

Geographically structured genetic variation, as observed here in *Erysimum*, is a common feature of Carpathian plant species (e.g. Ronikier et al., 2008; Băcilă et al., 2015; Šrámková-Fuxová et al., 2017; Melichárková et al., 2019; Stachurska-Swakoń et al., 2020), which reflects the high topographic complexity and geological heterogeneity of this mountain range and may also be related to the past range dynamics, the presence of multiple local glacial refugia and different recolonization routes (Ronikier, 2011; Mráz and Ronikier, 2016). The present genetic structure supports some of the known intra-Carpathian biogeographic boundaries (Ronikier, 2011; Mráz and Ronikier, 2016), such as between the Western and Eastern Carpathians, between the Eastern and Southern Carpathians or within the Southern Carpathians along deep transverse valleys (e.g. the tectonic valley of the river Olt, which separates the populations from the Buila-Vânturariţa Mts., Figs. 1, 2). The divergence of the populations from the Buila-Vânturariţa Mts. is probably also reinforced by their isolated location, as they represent the westernmost confirmed occurrence and a distribution gap is suspected to the east in the adjacent Fagaraş Mts. (due to the scarcity of suitable habitats and the prevailing crystalline bedrock, personal observ.). Thus, the observed genetic patterns suggest the existence of multiple glacial refugia for *E. witmannii* across the Carpathians and a gradual, stepwise range expansion (following the stepping stone colonization model, Kimura and Weiss, 1964), similar to other Carpathian plants (Ronikier, 2011). The presence of genetically admixed populations in the Postăvarul and Vihorlat Mts. (clades 2/3 and 3/4, Figs. 1, 2), located in the border areas, suggest the existence of contact zones between the different recolonization routes, further supporting this migration model.

The introgression tests also showed evidence of gene flow between the allopatric lineages in the Carpathians (Fig. 3B), which was supported by very low QD scores (Supplementary Fig. S2A) measuring skew in discordance (Pease et al., 2018). The strongest signal of introgression was observed between the lineages of the same cytotype (2n = 14), but to a lesser extent also between the two cytotypes (2n = 14 and 2n = 16) despite the expected reproductive barrier due to gamete incompatibilities. The origin of the geographically restricted cytotype with 2n = 16 is intriguing particularly due to its large genome size, which is almost twice that of 2n = 14, suggesting a complex genome restructuring (Mayrose and Lysak, 2020) alike in the Balkan species (see above), which remains to be investigated in the future.

4.3. Paraphyletic Erysimum odoratum s.str. displays multiple polyploid origins, morphological and chromosomal variation

Here we confirmed the tetraploid chromosome number (2n = 32) as the dominant cytotype of *E. odoratum* (Michalková, 2002; Marhold et al., 2007; Polatschek, 2013) in the study area, which extends from the Carpathians to the southernmost distribution limit in the northern Balkans (Fig. 1). Additionally, we also identified previously unknown cytotypes occurring in the mountains of northeastern Croatia (Papuk Mts.), southwestern Romania (Retezat and Banat Mts.), and Serbia (Fruška Gora and Beljanica Mts.). On the other hand, our genetic data show that the diploid cytotype reported from the Western Carpathians (provisionally also under the names *E. vagicum* nom. prov. or *E. carpaticum* nom. prov., Kliment, 1999; Michalková, 2002) should be excluded from *E. odoratum*, and belongs to *E. witmannii* s.l. The cytotype recorded in the Papuk, Retezat and Banat Mts. has an unusual chromosome number of 2n = 22, which might suggest a triploid level. However, initial cytogenomic analyses of accessions from the Banat Mts. have indicated a tetraploid genomic composition (Mandáková, unpubl.), suggesting chromosome number reduction from regular tetraploids (post-polyploid descending dysploidy, Mandáková and Lysak, 2018; Mayrose and Lysak 2020). Post-polyploid structural chromosomal changes are a key component of the complex process of post-polyploid diploidization, which gradually converts a polyploid genome back to a functionally diploid state and can contribute significantly to speciation events and adaptive radiation (Mandáková et al., 2016; Farhat et al., 2023; Zuo et al., 2022). However, these chromosomal rearrangements usually occur in the long term (e.g. Jiang et al., 2025) and are rarely observed at the intraspecific level (see Mas de Xaxars et al., 2015; Melichárková et al., 2020; Scarlett et al., 2023). Whether the populations from the Banat, Retezat and Papuk Mts. harbouring 2n = 22 are

characterized by the same genomic rearrangements remains unclear, but phylogenomic evidence from RADseq and Hyb-Seq data favours at least two separate origins (see below).

The paraphyly of *E. odoratum* (Fig. 2) may seem intriguing and taxonomically challenging at first glance, but some non-allopatric modes of speciation, such as progenitor-derivative (budding) or polyploid speciation, can lead to paraphyletic species (Crawford, 2010; Hörandl and Stuessy, 2010; Otero et al., 2022). Paraphyletic species are in fact not rare (Rieseberg and Brouillet, 1994) and can be seen as a natural transitional stage in evolution (Hörandl and Stuessy, 2010). Based on the polyploid status of *E. odoratum* and a nested position of the diploid *E. witmannii* s.l. (Fig. 2), we hypothesized that the paraphyly of *E. odoratum* is due to an allopolyploid origin, with *E. witmannii* s.l. being one probable ancestor. We tested this hypothesis using phased exon sequences from Hyb-Seq target genes and with different approaches suitable to identify parental subgenomes and their phylogenetic positions with respect to the extant diploids, following the principle described by Rothfels (2021) and implemented in several tools (Thomas et al., 2017; Barbera et al., 2019; Šlenker et al., 2021; Sancho et al., 2022). The results differed slightly between approaches but consistently yielded two outcomes: 1) strong evidence that *E. witmannii* s.l. is one of the parents for all cytotypes, and 2) indication of multiple independent polyploid origins, likely involving different ancestors. Allopolyploid origins with *E. cuspidatum* as the second parental species were inferred for the common tetraploid (2n = 32) cytotype, while *E. comatum* was suggested as the second parental species for the 2n = 22 cytotype from the Banat and Retezat Mts. The genomes of both these diploids, *E. cuspidatum* and *E. comatum*, together with that of *E. witmannii*, are possibly present also in the hexaploids from the Fruška Gora Mts. In contrast, the populations from the Papuk Mts. (2n = 22) appeared to be of an autopolyploid origin. However, the inference of a polyploid origin may be obscured by traces of interspecific gene flow found in *Erysimum* genomes (Abdelaziz et al., 2014, Osuna Mascaró, 2020; Osuna-Mascaró et al., 2022, Osuna Mascaró et al., 2023, and present data), as well as by a high degree of ancestral variation due to recent speciation (Moazzeni et al., 2014; Züst et al., 2020). Assignments of the alleles both using the PhyloSD and EPA-ng methods (Supplementary Table S4, Fig. 4B) pointed to a broad spectrum of diploids, with no clear threshold for distinguishing between major (parental) and minor (arising from gene flow and ancestral variation) homeologs (see Sancho et al., 2022). Furthermore, we found that some cytotypes can be differentiated morphologically, such as the cytotype from the Retezat Mts. and that from the Fruška Gora Mts. (Fig. 5C), and also observed a phenological shift in the population from the Banat Mts. (pop. Ciuc, 2n = 22). This population started flowering approximately 4 weeks earlier than other populations of *E. odoratum*, both in the field and in common garden experiments (personal observ.), which implies an additional prezygotic barrier restricting gene flow from the adjacent populations (Fig. 1). Variation in flowering time across populations has also been recorded in Iberian *Erysimum*, which likely affected introgression patterns (Osuna Mascaró et al., 2023).

4.4. *Erysimum as a model for the study of early stages of speciation with various evolutionary drivers*

The geographic structuring of genetic variation in *Erysimum* occurs at multiple levels, and, together with the high proportion of narrow endemics (Jalas & Suominen, 1994), suggests that ecogeographic separation is one of the key drivers of speciation. Geographic patterns are evident not only in previously published phylogenies at the genus level (Moazzeni et al., 2014; Züst et al., 2020) but also here within the *E. odoratum* species complex in its traditional circumscription, particularly in the phylogenetically distinct positions of the Carpathian *E. witmannii* s.l. and western Balkan endemics, as well as at the species level. Beyond the inherent effect of geographic distance, there are also biogeographical barriers in the mountainous landscapes of the Carpathians (Mráz and Ronikier, 2016) and the northwestern Balkans (Španiel and Rešetnik, 2022), which appear to limit gene flow, allowing genetic drift and selection to shape evolutionary trajectories (Butlin, 2024) both in *E. odoratum* and *E. witmannii* s.l. Additionally, karyotype evolution has led to cytotype variation within both species, potentially establishing intrinsic barriers to gene flow in sympatric or geographically proximate populations. These barriers may further drive population differentiation and contribute to speciation (Rieseberg, 2001; Kennedy et al., 2006; Soltis et al., 2007; Peirson et al., 2012). Furthermore, phylogenomic data indicate that *E. odoratum* may be an assemblage of independently evolved polyploids, however, distinguishing between remnants of parental genomes, shared ancestral variation, and introgression remained challenging. Taken together, the complex patterns of variation observed in both *E. odoratum* and *E. witmannii* s.l. appear to result from a broad spectrum of evolutionary processes. We suggest that these taxa – each composed of multiple lineages that have evolved through allopatric divergence as well as reticulate events (introgression or allopolyploidy), further shaped by chromosomal evolution – are in early stages of the process of speciation (de Queiroz, 2007; Butlin and Faria, 2024). From a taxonomic point of view, however, we refrain from simplistic taxonomic splitting based on genetic, geographic or cytotype differentiation, as it would not capture the observed more complex, network-like patterns. Therefore, we argue for a broad species concept for both *E. odoratum*, encompassing all polyploid cytotypes, and *E. witmannii*, comprising all Carpathian diploid populations, subsuming also the previously reported taxa/names *E. pallidiflorum*, *E. transsilvanicum*, *E. vagicum* nom. prov., and *E. carpaticum* nom. prov. under *E. witmannii*, which has nomenclatural priority (Turland et al., 2025).

In conclusion, this study enhances our understanding of the evolutionary processes that have shaped this complex genus of crucifers. Our findings suggest that the morphology-based infrageneric division into species complexes is probably largely artificial and in need of revision. However, defining intrageneric entities – whether formally as sections or informally as species complexes – may be complicated by reticulation events that probably also involved distantly related taxa. Here, we also highlight the challenges of clarifying taxonomy and evolution in recently radiated plant lineages affected by multiple evolutionary drivers including polyploidy, introgression and karyotype rearrangements.

CRediT authorship contribution statement

Richard Bačák: Writing – review & editing, Visualization, Investigation, Formal analysis. **Marek Šlenker:** Writing – review & editing, Visualization, Validation, Supervision, Investigation, Formal analysis, Data curation. **Barbora Šingliarová:** Writing – review & editing, Validation, Supervision, Investigation, Formal analysis. **Terezie Mandáková:** Writing – review & editing, Visualization, Validation, Investigation, Formal analysis. **Katarína Skokanová:** Writing – review & editing, Investigation. **Ingrid Turisová:** Writing – review & editing, Investigation. **Peter Turis:** Writing – review & editing, Investigation. **Janka Smatanová:** Writing – review & editing, Investigation. **Judita Zozomová-Lihová:** Writing – original draft, Visualization, Validation, Supervision, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the EU NextGenerationEU through the

Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00494 (to M.Š.). Computational resources were provided by the e-INFRA CZ project (ID: 90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic. We wish to thank to colleagues who contributed to sampling or accompanied us in the field: V. Cetlová, A. Kantor, H. Majerová, K. Marhold, E. Víziová, S. Španiel (all from the Plant Science and Biodiversity Centre, Bratislava, Slovakia), B. Hegyi, I.-D. Becze (both from Cheile Bicazului-Hăşmaş National Park, Izvoru Mureşului, Romania), and E. Víziová also for lab assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2025.108481.

Data availability

The Illumina sequencing reads generated for this study are available from NCBI Sequence Read Archive under BioProject PRJNA1284580, with Biosamples from SAMN49725651 to SAMN49725931. The code for data processing can be found in the following GitHub repository: https://github.com/MarekSlenker/Code-Availability/tree/main/Bacak_et_al_2025_Mol_Phylogenet_Evol.

References

Abdelaziz, M., Muñoz-Pajares, A.J., Lorite, J., Herrador, M.B., Perfectti, F., Gómez, J.M., 2014. Phylogenetic relationships of *Erysimum* (Brassicaceae) from the Baetic mountains (SE Iberian Peninsula). An. Jard. Bot. Madrid 71, e005. https://doi.org/10.3989/ajbm.2377.

Al-Shehbaz, I.A., 2012. A generic and tribal synopsis of the Brassicaceae. Taxon 61, 931–954. https://doi.org/10.1002/tax.615002.

Andrews, K., Good, J., Miller, M., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. 17, 81–92. https://doi.org/10.1038/nrg.2015.28.

Băcilă, I., Şuteu, D., Coldea, G., 2015. Genetic divergence and phylogeography of the alpine plant taxon *Onobrychis transsilvanica* (Fabaceae). Botany-Botanique 93, 257–266. https://doi.org/10.1139/cjb-2014-017.

Ball, P.W., 1993. *Erysimum* L., in: Tutin, T.G., Burges, N.A., Chater, A.O., Edmondson, J. R., Heywood, V.H., Moore, D.M., Valentine, D.H., Walters, S.M., Webb, D.A. (Eds.), Flora Europaea, vol. 1: *Psilotaceae to Platanaceae*, second ed. Cambridge University Press, Cambridge, pp. 325–335.

Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T., Stamatakis, A., 2019. EPA-ng: massively parallel evolutionary placement of genetic sequences. Syst. Biol. 68, 365–369. https://doi.org/10.1093/sysbio/syy054.

Bennetzen, J.L., Wang, H., 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu. Rev. Plant Biol. 65, 505–530. https://doi.org/10.1146/annurev-arplant-050213-035811.

Boucher, F.C., Casazza, G., Szövényi, P., Conti, E., 2016. Sequence capture using RAD probes clarifies phylogenetic relationships and species boundaries in *Primula* sect, Auricula. Mol. Phylogenet. Evol. 104, 60–72. https://doi.org/10.1016/j.ympev.2016.08.003.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537. https://doi.org/10.1371/journal.pcbi.1003537.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932. https://doi.org/10.1093/molbev/mss086.

Buono, D., Khan, G., von Hagen, K.B., Kosachev, P.A., Mayland-Quellhorst, E., Mosyakin, S.L., Albach, D.C., 2021. Comparative phylogeography of *Veronica spicata* and *V. longifolia* (Plantaginaceae) across Europe: Integrating hybridization and polyploidy in phylogeography. Front. Plant Sci. 11, 588354. https://doi.org/10.3389/fpls.2020.588354.

Butlin, R.K., 2024. Perspectives on speciation. Evol. J. Linn. Soc. 3, kzae035. https://doi.org/10.1093/evolinnean/kzae035.

Butlin, R.K., Faria, R., 2024. Local adaptation and reproductive isolation: when does speciation start? Evol. J. Linn. Soc. 3, kzae003. https://doi.org/10.1093/evolinnean/kzae003.

Carnicero, P., Kröll, J., Schönswetter, P., 2023. Homoploid hybrids are common but evolutionary dead ends, whereas polyploidy is not linked to hybridization in a group of Pyrenean saxifrages. Mol. Phylogenet. Evol. 180, 107703. https://doi.org/10.1016/j.ympev.2023.107703.

Chambers, E.A., Hillis, D.M., 2020. The multispecies coalescent over-splits species in the case of geographically widespread taxa. Syst. Biol. 69, 184–193. https://doi.org/10.1093/sysbio/syz042.

Chen, S., 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta 2, e107. https://doi.org/10.1002/imt2.107.

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

Crawford, D.J., 2010. Progenitor-derivative species pairs and plant speciation. Taxon 59, 1413–1423. https://doi.org/10.1002/tax.595008.

Czarna, A., Gawrońska, B., Nowińska, R., Morozowska, M., Kosiński, P., 2016. Morphological and molecular variation in selected species of *Erysimum* (Brassicaceae) from Central Europe and their taxonomic significance. Flora 222, 68–85. https://doi.org/10.1016/j.flora.2016.03.008.

Dayrat, B., 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85, 407–415. https://doi.org/10.1111/j.1095-8312.2005.00503.x.

de Queiroz, K., 2005. Ernst Mayr and the modern concept of species. Proc. Natl. Acad. Sci. 102, 6600–6607. https://www.pnas.org/doi/pdf/10.1073/pnas.0502030102.

de Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886. https://doi.org/10.1080/10635150701701083.

Ding, Y., Metherell, C., Huang, W., Hollingsworth, P.M., Twyford, A.D., 2025. Genome-wide differentiation by geography not species in taxonomically complex eyebrights (*Euphrasia*). Evolution 79, 483–492. https://doi.org/10.1093/evolut/qpae185.

Doležel, J., Sgorbati, S., Lucretti, S., 1992. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. Physiol. Plant. 85, 625–631. https://doi.org/10.1111/j.1399-3054.1992.tb04764.x.

Doležel, J., Greilhuber, J., Suda, J., 2007a. Flow Cytometry with plant cells, analysis of genes, chromosomes and genomes. WILEY-CH, Weinhem.

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

Dufresnes, C., Pribille, M., Alard, B., Gonçalves, H., Amat, F., Crochet, P.-A., Dubey, S., Perrin, N., et al., 2020. Integrating hybrid zone analyses in species delimitation: lessons from two anuran radiations of the Western Mediterranean. Heredity 124, 423–438. https://doi.org/10.1038/s41437-020-0294-z.

Dufresnes, C., Poyarkov, N., Jablonski, D., 2023. Acknowledging more biodiversity without more species. Proc. Natl. Acad. Sci. 120, e2302424120. https://doi.org/10.1073/pnas.2302424120.

Durand, E.Y., Patterson, N., Reich, D., Slatkin, M., 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28, 2239–2252. https://doi.org/10.1093/molbev/msr048.

Eriksson, J.S., Blanco-Pastor, J.L., Sousa, F., Bertrand, Y.J.K., Pfeil, B.E., 2017. A cryptic species produced by autopolyploidy and subsequent introgression involving *Medicago prostrata* (Fabaceae). Mol. Phylogenet. Evol. 107, 367–381. https://doi.org/10.1016/j.ympev.2016.11.020.

Farhat, P., Mandáková, T., Divíšek, J., Kudoh, H., German, D.A., Lysak, M.A., 2023. The evolution of the hypotetraploid *Catolobus pendulus* genome – the poorly known sister species of *Capsella*. Front. Plant Sci. 14, 1165140. https://doi.org/10.3389/fpls.2023.1165140.

Frajman, B., Rešetnik, I., Niketić, M., Ehrendorfer, F., Schönswetter, P., 2016. Patterns of rapid diversification in heteroploid *Knautia* sect. *Trichera* (Caprifoliaceae, Dipsacoideae), one of the most intricate taxa of the European flora. BMC Evol. Biol. 16, 204. https://link.springer.com/article/10.1186/s12862-016-0773-2.

German, D.A., Al-Shehbaz, I.A., 2008. Five additional tribes (Aphragmeae, Biscutelleae, Calepineae, Conringieae, and Erysimeae) in the Brassicaceae (Cruciferae). Harv. Pap. Bot. 13 (1), 165–170. https://doi.org/10.3100/1043-4534(2008)13[165:FATABC]2.0.CO;2.

German, D.A., Hendriks, K.P., Koch, M.A., Lens, F., Lysak, M.A., Bailey, C.D., Mummenhoff, K., Al-Shehbaz, I.A., 2023. An updated classification of the Brassicaceae (Cruciferae). PhytoKeys 220, 127–144. https://doi.org/10.3897/phytokeys.220.97724.

Hausdorf, B., 2025. Species delimitation using genomic data: Options and limitations. Mol. Ecol. 34, e17717. https://doi.org/10.1111/mec.17717.

Heibl, C., 2008. PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software packages. http://www.christophheibl.de/Rpackages.html (accessed June 2025).

Hendriks, K.P., Kiefer, C., Al-Shehbaz, I.A., Bailey, C.D., van Huysduynen, A.H., Nikolov, L.A., Nauheimer, L., et al., 2023. Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. Curr. Biol. 33, 4052–4068. https://doi.org/10.1016/j.cub.2023.08.026.

Hörandl, E., Stuessy, T.F., 2010. Paraphyletic groups as natural units of biological classification. Taxon 59, 1641–1653. https://doi.org/10.1002/tax.596001.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267. https://doi.org/10.1093/molbev/msj030.

Jalas, J., Suominen, J., 1994. Atlas Florae Europaeae. Distribution of Vascular Plants in Europe: 10 Cruciferae (Sisymbrium to Aubrieta), first ed. The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinky.

Jiang, X., Hu, Q., Mei, D., Li, X., Xiang, L., Al-Shehbaz, I.A., Song, X., Liu, J., Lysak, M.A., Sun, P., 2025. Chromosome fusions shaped karyotype evolution and evolutionary relationships in the model family Brassicaceae. Nat. Commun. 16, 4631. https://doi.org/10.1038/s41467-025-59640-2.

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

Josifović, A.M., Stjepanović, L., Gajić, M., Kojić, M., Diklić, N., 1972. Flore of Serbia III. Academie Serbe des sciences et des arts, Beograd.

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

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.

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

Karbstein, K., Tomasello, S., Hodač, L., Wagner, N., Marinček, P., Barke, B.H., Paetzold, C., Hörandl, E., 2022. Untying Gordian knots: unraveling reticulate polyploid plant evolution by genomic data using the large *Ranunculus auricomus* species complex. New Phytol. 235, 2081–2098. https://doi.org/10.1111/nph.18284.

Karbstein, K., Kösters, L., Hodač, L., Hofmann, M., Hörandl, E., Tomasello, S., Wagner, N. D., et al., 2024. Species delimitation 4.0: integrative taxonomy meets artificial intelligence. Trends Ecol. Evol. 39, 771–784. https://doi.org/10.1016/j.tree.2023.11.002.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795. https://doi.org/10.2307/2291091.

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

Kennedy, B.F., Sabara, H.A., Haydon, D., Husband, B.C., 2006. Pollinator-mediated assortative mating in mixed ploidy populations of *Chamerion angustifolium* (Onagraceae). Oecologia 150, 398–408. https://doi.org/10.1007/s00442-006-0536-7.

Kiefer, M., Schmickl, R., German, D.A., Mandáková, T., Lysak, M.A., Al-Shehbaz, I.A., Franzke, A., Mummenhoff, K., Stamatakis, A., Koch, M.A., 2014. BrassiBase: Introduction to novel knowledge database on Brassicaceae evolution. Plant Cell Physiol. 55, e3. https://doi.org/10.1093/pcp/pct158.

Kimura, M., Weiss, G.H., 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49, 561–576.

Kliment, J., 1999, Komentovaný prehľad vyšších rastlín flóry Slovenska, uvádzaných vliteratúre ako endemické taxóny. Slovenská botanická spoločnosť pri SAV aBotanická záhrada UK, Bratislava.

Kobiv, Y., Prokopiv, A., 2018. Bili Skeli limestone cliffs as an important hotspot of plant diversity in the Chyvchyny Mountains (Ukrainian Carpathians). Contribuţii Botanice LIII 19–26. https://doi.org/10.24193/Contrib.Bot.53.2.

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

Leaché, A.D., Bouckaert, R.R., 2018. Species trees and species delimitation with SNAPP: A tutorial and worked example. Workshop on Population and Speciation Genomics. https://www.beast2.org/tutorials/.

Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63, 534–542. https://doi.org/10.1093/sysbio/syu018.

Leaché, A.D., Banbury, B.L., Felsenstein, J., De Oca, A.N., Stamatakis, A., 2015. Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. Syst. Biol. 64, 1032–1047. https://doi.org/10.1093/sysbio/syv053.

Lee-Yaw, J., Grassa, C.J., Joly, S., Andrew, R.L., Rieseberg, L.H., 2019. An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). New Phytol. 221, 515–526. https://doi.org/10.1111/nph.15386.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [preprint]. https://doi.org/10.48550/arXiv.1303.3997.

Mairal, M., Šurinová, M., Castro, S., Münzbergová, Z., 2018. Unmasking cryptic biodiversity in polyploids: origin and diversification of *Aster amellus* aggregate. Ann. Bot. 122, 1047–1059. https://doi.org/10.1093/aob/mcy149.

Malinsky, M., Matschiner, M., Svardal, H., 2021. Dsuite - Fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21, 584–595. https://doi.org/10.1111/1755-0998.13265.

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

Mandáková, T., Lysak, M.A., 2018. Post-polyploid diploidization and diversification through dysploid changes. Curr. Opin. Plant Biol. 42, 55–65. https://doi.org/10.1016/j.pbi.2018.03.001.

Mandáková, T., Gloss, A.D., Whiteman, N.K., Lysak, M.A., 2016. How diploidization turned a tetraploid into a pseudotriploid. Am. J. Bot. 103, 1187–1196. https://doi.org/10.3732/ajb.1500452.

Marhold, K., 2011+. Brassicaceae. In: Euro+Med Plantbase - the information resource for Euro-Mediterranean plant diversity. Published at http://www.europlusmed.org (accessed June 2025).

Marhold, K., Mártonfi, P., Mereďa, P., Mráz, P., 2007. Chromosome number survey of the ferns and flowering plants of Slovakia. VEDA, vydavateľstvo SAV, Bratislava.

Mas de Xaxars, G., Garnatje, T., Pellicer, J., Siljak-Yakovlev, S., Vallès, J., Garcia, S., 2015. Impact of dysploidy and polyploidy on the diversification of high mountain *Artemisia* (Asteraceae) and allies. Alp. Bot. 126, 35–48. https://doi.org/10.1007/s00035-015-0159-x.

Mason, N.A., Fletcher, N.K., Gill, B.A., Funk, W.C., Zamudio, K.R., 2020. Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. Syst. Biodivers. 18, 269–280. https://doi.org/10.1080/14772000.2020.1730475.

Mayrose, I., Lysak, M.A., 2020. The evolution of chromosome numbers: Mechanistic models and experimental approaches. Genome Biol. Evol. 13, evaa220. https://doi.org/10.1093/gbe/evaa220.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20 (9), 1297–1303. https://doi.org/10.1101/gr.107524.110.

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

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

Meng, K.-K., Chen, S.-F., Xu, K.-W., Zhou, R.-C., Li, M.-W., Dhamala, M.K., Liao, W.-B., Fan, Q., 2021. Phylogenomic analyses based on genome-skimming data reveal cyto-nuclear discordance in the evolutionary history of *Cotoneaster* (Rosaceae). Mol. Phylogenet. Evol. 158, 107083. https://doi.org/10.1016/j.ympev.2021.107083.

Michalková, E., 1999. *Erysimum pallidiflorum* Jáv. (Brassicaceae) na Slovensku. Ochrana prírody 17, 39–47.

Michalková, E., 2001. Poznámky k rodu *Erysimum* (Brassicaceae) na Slovensku. Bulletin Slovenskej Botanickej Spoločnosti 23, 95–100.

Michalková, E., 2002. *Erysimum* L., in: Goliašová, K., Šípošová, H. (Eds.), Flóra Slovenska V/4. VEDA, vydavateľstvo SAV, Bratislava, pp. 182–224.

Moazzeni, H., Zarre, S., Pfeil, B.E., Bertrand, Y.J.K., German, D.A., Al-Shehbaz, I.A., Mummenhoff, K., Oxelman, B., 2014. Phylogenetic perspectives on diversification and character evolution in the species-rich genus *Erysimum* (Erysimeae; Brassicaeae) based on a densely sampled ITS approach. Bot. J. Linn. Soc. 175, 497–522. https://doi.org/10.1111/boj.12184.

Morales-Briones, D.F., Gehrke, B., Huang, C.-H., Liston, A., Ma, H., Marx, H.E., Tank, D. C., Yang, Y., 2022. Analysis of paralogs in target enrichment data pinpoints multiple ancient polyploidy events in *Alchemilla* s.l. (Rosaceae). Syst. Biol. 71, 190–207. https://doi.org/10.1093/sysbio/syab032.

Mráz, P., Ronikier, M., 2016. Biogeography of the Carpathians: Evolutionary and spatial facets of biodiversity. Biol. J. Linn. Soc. 119, 528–559. https://doi.org/10.1111/bij.12918.

Nei, M., 1972. Genetic distance between populations. Am. Nat. 106, 283–292. https://www.jstor.org/stable/2459777.

Oberprieler, C., 2023. The Wettstein tesseract: A tool for conceptualising species-rank decisions and illustrating speciation trajectories. Taxon 72, 1–7. https://onlinelibrary.wiley.com/doi/10.1002/tax.12825.

Oprea, A., 2005. Lista critică a plantelor vasculare din România. Editura universităţii „Alexandru Ioan Cuza“, Iaşi.

Ortiz, E.M., 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. Zenodo. https://doi.org/10.5281/zenodo.2540861.

Osuna Mascaró, C.I., 2020. Hybridization as an evolutionary driver for speciation: A case in the Southern European Erysimum species. PhD thesis. Universidad de Granada, Granada.

Osuna-Mascaró, C., de Casas, R.R., Berbel, M., Gómez, J.M., Perfectti, F., 2022. Lack of ITS sequence homogenization in *Erysimum* species (Brassicaceae) with different ploidy levels. Sci. Rep. 12, 16907. https://doi.org/10.1038/s41598-022-20194-8.

Osuna Mascaró, C., de Casas, R.R., Gómez, J.M., Loureiro, J., Castro, S., Landis, J.B., Hopkins, R., Perfectti, F., 2023. Hybridization and introgression are prevalent in Southern European *Erysimum* (Brassicaceae) species. Ann. Bot. 131, 171–183. https://doi.org/10.1093/aob/mcac048.

Otero, A., Vargas, P., Fernández-Mazuecos, M., Jiménez-Mejías, P., Valcárcel, V., Villa-Machío, I., Hipp, A.L., 2022. A snapshot od progenitor-derivative speciation in Iberodes (Boraginaceae). Mol. Ecol. 31, 3192–3209. https://doi.org/10.1111/mec.16459.

Ott, T., Schall, M., Vogt, R., Oberprieler, C., 2022. The warps and wefts of a polyploidy complex: Integrative species delimitation of the diploid *Leucanthemum* (Compositae, Anthemideae) representatives. Plants 11, 1878. https://doi.org/10.3390/plants11141878.

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

Peirson, J.A., Reznicek, A.A., Semple, J.C., 2012. Polyploidy, infraspecific cytotype variation, and speciation in Goldenrods: The cytogeography of *Solidago* subsect. *Humiles* (Asteraceae) in North America. Taxon 61, 197–210. https://doi.org/10.1002/tax.611014.

Pembleton, L.W., Cogan, N.O.I., Forster, J.W., 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol. Ecol. Resour. 13, 946–952. https://doi.org/10.1111/1755-0998.12129.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One 7, e37135. https://doi.org/10.1371/journal.pone.0037135.

Pinheiro, F., Dantas-Queiroz, M.V., Palma-Silva, C., 2018. Plant species complexes as models to understand speciation and evolution: A review of South American studies. Crit. Rev. Plant Sci. 37, 54–80. https://doi.org/10.1080/07352689.2018.1471565.
Polatschek, A., 1997. *Erysimum* (Brassicaceae): Chromosomenzählungen griechischer Arten. Linzer biologische Beiträge 29, 545–553.
Polatschek, A., 2010. Revision der Gattung *Erysimum* (Cruciferae). Teil 1: Russland, die Nachfolgestaaten der USSR (excl. Georgien, Armenien, Azerbaidzan), China, Indien, Pakistan, Japan und Korea. Annalen des Naturhistorischen Museums in Wien, Serie B 111, 181–275.
Polatschek, A., 2013. Revision der Gattung *Erysimum* (Cruciferae). Teil 5: Nord-, West-, Zentraleuropa, Rumänien und westliche Balkan-Halbisel bis Albanien. Annalen des Naturhistorischen Museums in Wien, Serie B 115, 75–218.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959. https://doi.org/10.1093/genetics/155.2.945.
Quinlan, A.R., Hallm, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafaljović, M., Noor, M.A.F., Mehlig, B., Westram, A.M., 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. 30, 1450–1477. https://doi.org/10.1111/jeb.13047.
R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (accessed Mar 2025).
R Core Team, 2024. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (accessed Mar 2025).
Ren, C., Wang, L., Nie, Z.-L., Tang, M., Johnson, G., Tan, H.-T., Xia, N.-H., Wen, J., Yang, Q.-E., 2024. Complex but clear allopolyploid pattern of subtribe Tussilagininae (Asteraceae: Senecioneae) revealed by robust phylogenomic evidence, with development of a novel homeolog-sorting pipeline. Syst. Biol. 73, 941–963. https://doi.org/10.1093/sysbio/syae046.
Rieseberg, L.H., 2001. Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16, 351–358.
Rieseberg, L.H., Brouillet, L., 1994. Are many plant species paraphyletic? Taxon 43, 21–32.
Rieseberg, L.H., Burke, J.M., 2001. The biological reality of species: gene flow, selection, and collective evolution. Taxon 50, 47–67. https://doi.org/10.2307/1224511.
Ronikier, M., 2011. Biogeography of high-mountain plants in the Carpathians: An emerging phylogeographical perspective. Taxon 60, 373–389. https://doi.org/10.1002/tax.602008.
Ronikier, M., Cieślak, E., Korbecka, G., 2008. High genetic differentiation in the alpine plant *Campanula alpina* Jacq. (Campanulaceae): evidence for glacial survival in several Carpathian regions and long-term isolation between the Carpathians and the Alps. Mol. Ecol. 17, 1763–1775. https://doi.org/10.1111/j.1365-294X.2008.03664.x.
Rothfels, C.J., 2021. Polyploid phylogenetics. New Phytol. 230, 66–72. https://doi.org/10.1111/nph.17105.
Sancho, R., Inda, L.A., Díaz-Pérez, A., Des Marais, D.L., Gordon, S., Vogel, J.P., Lusinska, J., Hasterok, R., Contreras-Moreira, B., Catalán, P., 2022. Tracking the ancestry of known and 'ghost' homeologous subgenomes in model grass *Brachypodium* polyploids. Plant J. 109, 1535–1558. https://doi.org/10.1111/tpj.15650.
Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668. https://doi.org/10.1093/molbev/msw079.
Scarlett, V.T., Lovell, J.T., Shao, M., Phillips, J., Shu, S., Lusinska, J., Goodstein, D.M., Jenkins, J., Grimwood, J., Barry, K., et al., 2023. Multiple origins, one evolutionary trajectory: gradual evolution characterizes distinct lineages of allotetraploid *Brachypodium*. Genetics 223, iyac146. https://doi.org/10.1093/genetics/iyac146.
Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S.C.K., Cronn, R.C., Dreyer, L.L., Suda, J., 2016. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African *Oxalis* (Oxalidaceae). Mol. Ecol. Resour. 16, 1124–1135. https://doi.org/10.1111/1755-0998.12487.
Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675. https://doi.org/10.1038/nmeth.2089.
Schur, P.J.F., 1866. Enumeratio plantarum Transsilvaniae:exhibens: stirpes phanerogamas sponte crescentes atque frequentius cultas, cryptogamas vasculares, characeas, etiam muscos hepaticasque. Sumptibus C. Graeser, Vindobonae.
Šlenker, M., 2024. VCF_prune (1.1.0) [software]. Zenodo. https://doi.org/10.5281/zenodo.10522817.
Šlenker, M., Kantor, A., Marhold, K., Schmickl, R., Mandáková, T., Lysak, M.A., Perný, M., Caboňová, M., Slovák, M., Zozomová-Lihová, J., 2021. Allele sorting as a novel approach to resolving the origin of allotetraploids using Hyb-Seq data: A case study of the Balkan mountain endemic *Cardamine barbaraeoides*. Front. Plant Sci. 12, 659275. https://doi.org/10.3389/fpls.2021.659275.
Šlenker, M., Koutecký, P., Marhold, K., 2022. MorphoTools2: an R package for multivariate morphometric analysis. Bioinformatics 38, 2954–2955. https://doi.org/10.1093/bioinformatics/btac173.
Šlenker, M., Kantor, A., Senko, D., Mártonfiová, L., Šrámková, G., Cetlová, V., Dönmez, A., Yüzbaşıoğlu, S., Zozomová-Lihová, J., 2024. Genome-wide data uncover cryptic diversity with multiple reticulation events in the Balkan-Anatolian *Cardamine* (Brassicaceae) species complex. Mol. Ecol. 33, e17564. https://doi.org/10.1111/mec.17564.
Slovák, M., Melichárková, A., Gbúrová Štubňová, E., Kučera, J., Mandáková, T., Smyčka, J., Lavergne, S., Passalacqua, N.G., Vďačný, P., Paun, O., 2023. Pervasive introgression during rapid diversification of the European mountain genus *Soldanella* (L.) (Primulaceae). Syst. Biol. 72, 491–504. https://academic.oup.com/sysbio/article/72/3/491/6798865.
Soltis, D.E., Soltis, P.S., Schemske, D.W., Hancock, J.F., Thompson, J.N., Husband, B.C., Judd, W.S., 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon 56, 13–30. https://doi.org/10.2307/25065732.
Španiel, S., Rešetnik, I., 2022. Plant phylogeography of the Balkan Peninsula: Spatiotemporal patterns and processes. Plant Syst. Evol. 308, 38. https://doi.org/10.1007/s0060 6-022-01831-1.
Španiel, S., Šlenker, M., Melichárková, A., Caboňová, M., Šandalová, M., Zeisek, V., Marhold, K., Zozomová-Lihová, J., 2023. Phylogenetic challenges in a recently diversified and polyploid-rich *Alyssum* (Brassicaceae) lineage: low divergence, reticulation, and parallel polyploid speciation. Evolution 77, 1226–1244. https://doi.org/10.1093/evolut/qpad035.
Spoelhof, J.P., Soltis, P.S., Soltis, D.E., 2017. Pure polyploidy: Closing the gaps in autopolyploid research. J. Syst. Evol. 55, 340–352. https://doi.org/10.1111/jse.12253.
Šrámková-Fuxová, G., Záveská, E., Kolář, F., Lučanová, M., Španiel, S., Marhold, K., 2017. Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction. Bot. J. Linn. Soc. 185, 321–342. https://doi.org/10.1093/botlinnean/box064.
Stachurska-Swakoń, A., Cieślak, E., Ronikier, M., Nowak, J., Kaczmarczyk, A., 2020. Genetic structure of *Doronicum austriacum* (*Asteraceae*) in the Carpathians and adjacent areas: toward a comparative phylogeographical analysis of tall-herb species. Plant Syst. Evol. 306, 14. https://doi.org/10.1007/s00606-020-01652-0.
Stankowski, S., Ravinet, M., 2021. Defining the speciation continuum. Evolution 75, 1256–1273. https://doi.org/10.1111/evo.14215.
Temsch, E., Greilhuber, J., Krisai, R., 2010. Genome size in liverworts. Preslia 82, 63–80.
Thomas, G.W.C., Ather, S.H., Hahn, M.W., 2017. Gene-Tree reconciliation with MUL-trees to resolve polyploidy events. Syst. Biol. 66, 1007–1018. https://doi.org/10.1093/sysbio/syx044.
Turland, N.J., Wiersema, J.H., Barrie, F.R., Gandhi, K.N., Gravendyck, J., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Klopper, R.R., Knapp, S., Kusber, W.-H., Li, D.-Z., May, T.W., Monro, A.M., Prado, J., Price, M.J., Smith, G.F., Zamora Señoret, J.C., 2025. International Code of Nomenclature for algae, fungi, and plants (Madrid Code). Regnum Vegetabile 162. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226839479.001.0001.
Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43 (1), 11-10. https://doi.org/10.1002/0471250953.bi1110s43.
Vangjeli, J., 2017. Flora Albanica Atlas, Vol. 1. Pteridophyta – Apiaceae. Koeltz Botanical Books, Schmitten-Obereifenberg.
Vasile, M.-A., Böhnert, T., O'Leary, N., Weigend, M., 2024. Is it all *Phacelia pinnatifida*? Molecular delimitations and taxonomic revision of *Phacelia* sect. *Glandulosae* in South America based on ddRADseq. Taxon 73, 1437–1459. https://doi.org/10.1002/tax.13265.
Vences, M., Miralles, A., Dufresnes, C., 2024. Next-generation species delimitation and taxonomy: Implications for biogeography. J. Biogeogr. 51, 1709–1722. https://doi.org/10.1111/jbi.14807.
Wang, M., Kong, L., 2019. pblat: a multithread blat algorithm speeding up aligning sequences to genomes. BMC Bioinf. 20, 28. https://doi.org/10.1186/s12859-019-2597-8.
Wang, X., Morton, J.A., Pellicer, J., Leitch, I.J., Leitch, A.R., 2021. Genome downsizing after polyploidy: mechanisms, rates and selection pressures. Plant J. 107, 1003–1015. https://doi.org/10.1111/tpj.15363.
Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A., Liston, A., 2014. Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. Applications in Plant Sciences 2, 1400042. https://doi.org/10.3732/apps.1400042.
Wiens, J.J., 2007. Species delimitation: New approaches for discovering diversity. Syst. Biol. 56, 875–878. https://doi.org/10.1080/10635150701748506.
Yang, L., Kong, H., Huang, J.-P., Kang, M., 2019. Different species or genetically divergent populations? Integrative species delimitation of the *Primulina hochiensis* complex from isolated karst habitats. Mol. Phylogenet. Evol. 132, 219–231. https://doi.org/10.1016/j.ympev.2018.12.011.
Yi, H., Dong, S., Yang, L., Wang, J., Kidner, C., Kang, M., 2023. Genome-wide data reveal cryptic diversity and hybridization in a group of tree ferns. Mol. Phylogenet. Evol. 184, 107801. https://doi.org/10.1016/j.ympev.2023.107801.
Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19 (Suppl. 6), 153. https://doi.org/10.1186/s12859-018-2129-y.
Zuo, S., Guo, X., Mandáková, T., Edginton, M., Al-Shehbaz, I.A., Lysak, M.A., 2022. Genome diploidization associates with cladogenesis, trait disparity, and plastid gene evolution. Plant Physiol. 190, 403–420. https://doi.org/10.1093/plphys/kiac268.
Züst, T., Strickler, S.R., Powell, A.F., Mabry, M.E., An, H., Mirzaei, M., York, T., Holland, C.K., Kumar, P., Erb, M., Petschenka, G., Gómez, J.-M., Perfectti, F., Müller, C., Pires, J.C., Mueller, L.A., Jander, G., 2020. Independent evolution of ancestral and novel defenses in a genus of toxic plants (*Erysimum*, Brassicaceae). eLife 9, e51712. https://doi.org/10.7554/eLife.51712.
Đurović, S., Schönswetter, P., Niketić, M., Tomović, G., 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. https://doi.org/10.12705/662.4.