Genotyping-by-sequencing provides new genetic and taxonomic insights in the critical group of *Centaurea tenorei*

Daniele De Luca[1], Emanuele Del Guacchio[1,2]*, Paola Cennamo[3], Luca Paino[1] and Paolo Caputo[1,2]

[1]Department of Biology, University of Naples Federico II, Naples, Italy, [2]Botanical Garden of Naples, University of Naples Federico II, Naples, Italy, [3]Department of Humanities, University Suor Orsola Benincasa, Naples, Italy

Centaurea L. is one of the most widespread, differentiated, and critical genera of Asteraceae in the Euro-Mediterranean area, with more than 100 currently recognized species inhabiting the region. The controversial *C. tenorei* group, narrowly endemic to the Peninsula of Sorrento (Campania region, southern Italy), includes three weakly differentiated microspecies: *C. tenorei* Guss. ex Lacaita, *C. montaltensis* (Fiori) Peruzzi and *C. lacaitae* Peruzzi. However, their taxonomic distinctiveness and relationships with close or sympatric species are still unclear. In particular, the existence in several localities of individuals with intermediate morphology suggests inadequate taxonomic assessment within the group or hybridization and introgression with other species. In this study we aimed at defining population structure in this complex. With this objective, we sampled the three currently accepted species from their *loci classici* (i.e., the localities in which the taxa were originally described) and from other localities throughout the range, including populations of difficult identification occurring where the ranges of different taxa overlap. We employed a panel of SNPs obtained via genotyping-by-sequencing for investigations on genetic structure, admixture and ploidy inference, the latter also compared with chromosome counts. Our results showed that *Centaurea tenorei* s.l. is consistently tetraploid, contradicting the current taxonomy that was also based on ploidy level. Population structure analyses indicated the presence of four to seven clusters, most of which with clear evidence of admixture. Furthermore, contrarily to what previously supposed, we demonstrated a remarkable contribution of *C. deusta*, more that of *C. cineraria* in the genetic make-up of *C. tenorei*. However, we found a population of *C. cineraria* outside its ecological range, probably driven by climate change, which could be responsible in the future of further hybridization phenomena.

KEYWORDS

Asteraceae, hybridization, mixed ploidy, population structure, SNPs

Introduction

Taxonomy and phylogeny of genus *Centaurea* L. have been intensely investigated in the last years, from both a morphological and molecular point of view (e.g., Font et al., 2002; Garcia-Jacas et al., 2006; Suarez-Santiago et al., 2007; Garcia-Jacas et al., 2009; Hilpold et al., 2011; Hilpold et al., 2014b; López-Alvarado et al., 2014; Ranjbar and Negaresh, 2014; Bona, 2015; Arnelas et al., 2018; Arnelas et al., 2020; Del Guacchio et al., 2020a; Novaković et al., 2022). Various of these contributions show that patterns of descent in this genus do not mirror the current taxonomy and indeed are more often associated to geographical proximity; this has been interpreted as the result of diffuse hybridization/introgression and gene flow between divergent populations, species and even sections in a context of overall low molecular variation or of Incomplete Lineage Sorting (ILS), complicated by convergence in diagnostic morphological traits (Suarez-Santiago et al., 2007; Hilpold et al., 2014a). As a consequence, species sampled in multiple populations are "almost never monophyletic" (Hilpold et al., 2014b, p. 205), incongruence in gene trees is rampant, and phylogeny does not necessarily correspond to systematics (Hilpold et al., 2014b; López-Alvarado et al., 2014; Novaković et al., 2022).

Hybridization indeed appears to be a significant evolution driver in the genus (Garcia-Jacas et al., 2009; Hilpold et al., 2014b; López-Alvarado et al., 2014; Arnelas et al., 2018), especially at the homoploid level (Font et al., 2009; Pisanu et al., 2011; Mameli et al., 2014). In addition, polyploids can be found, as different cytotypes within the same, often widely distributed species (López-Alvarado et al., 2014). Moreover, they may be involved in the origin of new species (Garcia-Jacas, 1998; Koutecký, 2007; Garcia-Jacas et al., 2009; Mráz et al., 2012; Merle et al., 2022), or even in the appearance of highly invasive hybrids (Blair and Hufbauer, 2010). All these phenomena (i.e., hybridization, ILS, low molecular variation, convergence, intraspecific ploidy variation, polyploid origin of species) generate pervasive incongruence between phylogenetic hypotheses, and poorly supported trees (Hilpold et al., 2014a; López-Alvarado et al., 2014); this in turn greatly complicates taxa delimitation (e.g., Novaković et al., 2022) and phylogenetic inference (e.g., Hilpold et al., 2014b).

One of the taxonomically critical groups with mixed ploidy within *Centaurea* is that of *C. tenorei* Guss. ex Lacaita. According to the latest reassessment (Peruzzi and Scoppola, 2008), the group (basic chromosome number x = 9) is represented by three species, namely *C. lacaitae* Peruzzi (4x; 2n = 36), *C. montaltensis* (Fiori) Peruzzi (4x; 2n = 36), and *C. tenorei* s.str. (2x; 2n = 18), all narrow endemic to the Sorrento Peninsula (Campania, south-western Italy). As indicated in Santangelo et al. (2017), these taxa are remarkably variable and weakly differentiated in terms of morphology, ecology and distribution (Lacaita, 1922; Pignatti and Lausi, 1982; Guarino and Rampone, 2006). As a consequence, taxonomy of the group has greatly varied along time (Santangelo et al., 2017 and references therein), and it is still unsatisfactory (PFI [Portale della Flora d'Italia], 2022), because assignment of some

individuals to one of the three species remains difficult (Santangelo et al., 2017). Interpretation of relationships and boundaries in this group is additionally complicated by the sympatric occurrence of two further taxa in the Peninsula of Sorrento, i.e., *C. deusta* Ten. subsp. *deusta* and *C. cineraria* L. subsp. *cineraria* (both mostly diploids with 2n = 18: Bedini and Peruzzi, 2021), which are said or presumed to form natural hybrids with *C. tenorei* s.l. (Del Guacchio et al., 2020a). The systematic reassessment of this group is not only challenging, but also relevant by a conservational point of view. In fact, it is protected throughout its range at regional level because regarded as endangered but data for taxonomic assessment (and consequent conservation) are deficient and also afflicted by its obscure taxonomy (Rossi et al., 2020).

The study of population structure is a key prerequisite for detection of lineages and/or evolutionary significant units and their phylogenetic potential (Allendorf et al., 2013), as well as for taxonomic recognition of weakly defined taxa or identification of dubious specimens (De Castro et al., 2020; Favret et al., 2021). Moreover, knowledge on population structure is essential for setting conservation priorities and strategies (Wuyun et al., 2015; Kienberg and Becker, 2017; Hatmaker et al., 2018; Cheng et al., 2020). The genus *Centaurea*, however, may be difficult to describe in terms of significant evolutionary units, population structure or phylogeny, especially with traditional molecular techniques (Herrando-Moraira et al., 2019). Higher resolution methods, as those involving the use of a great number of Single Nucleotide Polymorphisms (SNPs) via Next Generation Sequencing (NGS) may be regarded as the elective choice for such problems (e.g., Paetzold et al., 2019; Hyun et al., 2020). However, the different ploidy levels occurring in our chosen system may represent a difficulty. In plants, in facts, inference of population structure is often complicated by the occurrence of mixed-ploidy populations (Kolář, 2021), which occur in about 16% of plant species (Rice et al., 2015). In spite of the significance of ploidy variation in plants, however, the number of software tools available for routine population genomic analyses in polyploids and, especially, in mixed-ploidy systems, is comparatively low in respect to diploids (Dufresne et al., 2014; Kolář, 2021); one of the few examples of such tools for NGS SNPs data is fineRADStructure (Malinsky et al., 2018). On the other hand, mixed-ploidy systems allow the study of possible correlation between ploidy and genetic variability within populations or among single individuals (Soltis et al., 2016). Recently new techniques and accompanying software, are capable of inferring this correlation directly from NGS data (Gompert and Mock, 2017; Weiß et al., 2018; Soraggi et al., 2022).

In this study, using a NGS approach, we aim at (a) describing genetic structure in populations of *C. tenorei* group, (b) testing the occurrence of admixture within the *C. tenorei* species group or between these latter and *C. cineraria* or *C. deusta*, (c) detect a possible correlation between ploidy and populations \ taxa, (d) understanding whether genetic variation among our three taxa of interest is indeed sufficiently discontinuous to grant their taxonomic recognition.

Materials and methods

Selection of taxa, sample collection and DNA extraction

We included in this study the populations of the three microspecies of the group (see codes in Table 1 and Figure 1), i.e., *C. lacaitae* [lacCA], *C. montaltensis* [monAV], and *C. tenorei* [tenMO] from the localities in which the taxa were originally described and are currently found. Some populations, difficult to ascribe to a single taxon, were included as well: some atypical individuals of *C. cineraria* from Valico di Chiunzi with glabrescent leaves [cixVA]; a single presumed hybrid between the former taxon

and *C. deusta* [hybVA]; several populations showing, at various degrees, intermediate features between *C. tenorei* s.l. and *C. cineraria* (cf. Lacaita, 1922; Santangelo et al., 2017; Del Guacchio et al., 2020a): Santa Maria del Castello [intSM], Vico Equense [intVI], and again Valico di Chiunzi [intVA]. We also included some populations of two taxa known or much suspected to hybridize with *C. tenorei* s.l. (Del Guacchio et al., 2020a): two populations of *C. deusta* subsp. *deusta* from Mt. Molare [deuMO] and Mt. Finestra [deuFI], and three populations of *C. cineraria* subsp. *cineraria*, gathered from Minori [cinMI], Marina della Lobra [cinML], and Capri Island [cinCA].

For comparison purposes, we employed one population of the other subspecies of *C. deusta*, i.e., *C. deusta* subsp. *leucolepis* (DC.)

TABLE 1 List of taxa sampled in the present study, with indications of localities, geographic coordinates, voucher specimen preserved at NAP and number of individuals collected.

Taxon	Locality (province, region)	Coordinates	Population voucher specimen	Individuals	Code
C. lacaitae Peruzzi	Capo d'Orso (Salerno, Campania) *	40.636 N 14.678 E	NAP0001741	9	lacCA
C. montaltensis (Fiori) Peruzzi	Mt. Avvocata (Salerno, Campania) *	40.653 N 14.669 E	NAP0001750	16	monAV
C. tenorei Guss. ex Lacaita	Mt. Molare (Naples, Campania) *	40.648 N 14.501 E	NAP0001746	7	tenMO
Possible intermediates between *C. cineraria* and *C. tenorei*	Vico Equense (Naples, Campania)	40.677 N 14.436 E	NAP0001744	7	intVI
	Santa Maria del Castello (Naples, Italy)	40.639 N 14.479 E	NAP0001737	6	intSM
	Valico di Chiunzi, (Salerno. Italy)	40.718 N 14.618 E	NAP0001743	5	intVA
C. cineraria L. subsp. *cineraria*	Capri (Naples, Campania)	40.553 N 14.243 E	NAP0001745	5	cinCA
	Marina della Lobra (Naples, Campania)	40.608 N 14.333 E	NAP0001748	5	cinML
	Minori (Salerno, Campania)	40.650 N 14.627 E	NAP0001738	7	cinMI
Centaurea cf. *cineraria*	Valico di Chiunzi (Salerno, Campania)	40.717 N 14.631 E	NAP0001751	7	cixVA
C. deusta Ten. subsp. *deusta*	Mt. Finestra (Salerno, Campania)	40.692 N 14.669 E	NAP0001739	5	deuFI
	Mt. Molare (Naples, Campania)	40.648 N 14.501 E	NAP0001740	5	deuMO
C. deusta subsp. *leucolepis* (DC.) Del Guacchio, Cennamo & P.Caputo	Bagnoli (Naples, Campania) *	40.815 N 14.162 E	NAP0001742	3	leuBA
Possible *C. deusta* × *C.* cf. *cineraria*	Valico di Chiunzi (Salerno, Campania)	40.719 N 14.616 E	—	1	hybVA
C. aeolica Guss. ex Lojac.	Stromboli Island (Messina, Sicily) *	38.796 N 15.237 E	NAP0001791	1	aeoST
C. ambigua Guss. subsp. *ambigua*	Piano delle Cinque Miglia (L'Aquila, Abruzzo)	41.888 N 14.000 E	NAP0001749	2	ambPC
C. pandataria (Fiori & Bég.) Bég.	Ventotene Island (Latina, Latium) *	40.796 N 13.428 E	NAP0001747	5	panVE

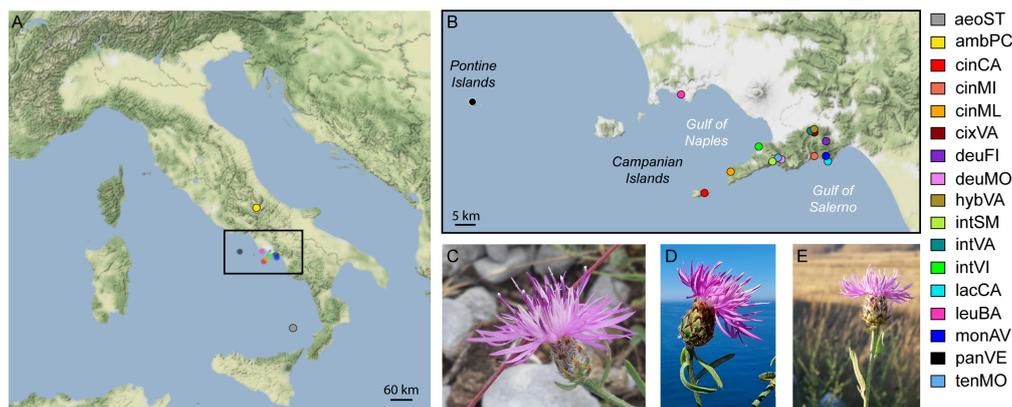An asterisk indicates localities cited in the protologues of the species.

FIGURE 1
Distribution of sampling localities. **(A)** sampling points; **(B)** zoom of sampling points in Campania region (Italy); **(C)** flowering head of *C. montaltensis* (Mt. Avvocata, Salerno Province); **(D)** flowering head of *C. lacaitae* (Vico Equense, Naples Metropolitan City); **(E)** flowering head of *C. ambigua* subsp. *ambigua* (Piano delle Cinque Miglia, Rocca Pia, Aquila province). Abbreviations in legend are as in Table 1.

Del Guacchio, Cennamo & P.Caputo (= *C. corensis* Vals. & Filigh.) [leuBA], occurring in the same region but not sympatric with our group of interest. We also added two individuals of *C. ambigua* Guss. subsp. *ambigua* [ambPC] because, even if *C. ambigua* s.l. is completely allopatric, it is often included in the taxonomic aggregate of *C. tenorei* (Santangelo et al., 2017), and indeed regarded by Guadagno (1932) as an ancestor of it. As outgroups, we employed a population of the allopatric taxon *C. pandataria* (Fiori & Bég.) Bég. [panVE] (for the correct authorship of the name and its relationship with *C. cineraria*, see Del Guacchio et al., 2022 but cf. also Turland et al., 2018, art. 46.2), and an individual of *C. aeolica* Guss. ex Lojac. [aeoSTR], very close to the former species (Del Guacchio et al., 2022 and references therein).

We field-collected leaf material from 96 samples of the above-mentioned taxa (Table 1, Figure 1). The only exception was *C. aeolica*, for which we employed material from a plant cultivated at the Botanic Garden of Naples and originated from seeds collected in the wild. We collected 4–7 individuals per population (see Table 1), with the exception of hybVA (only a single individual found) and some outgroups (*C. aeolica* and *C. ambigua*). We gathered a larger number of samples for monAV because this population extends broadly along an altitudinal cline (600–1000 m a.s.l.), and we aimed at checking whether any genetic difference might occur within it. The taxa were identified mainly by using Peruzzi and Scoppola (2008) and Pignatti (2018). Voucher specimens are deposited at the Herbarium of the Botanic Garden of Naples (NAP). Samples were taken during flowering time (late spring-summer) and leaf material was stored at −80°C until DNA extraction. DNA extraction was carried out using the GeneAll® Exgene™ Plant SV mini kit (GeneAll Biotechnology, Seoul, Korea) following manufacturer's protocol. To obtain high-quality DNA, frozen leaves were grinded in liquid nitrogen. Integrity of extracted DNA was evaluated on 0.8% gel electrophoresis and concentration was determined with Qubit® dsDNA HS Fluorometer assay (Invitrogen, ThermoFisher Scientific, Waltham, MA, USA).

Genomic library preparation and SNP discovery and genotyping

We performed genotyping-by-sequencing (GBS) for *de novo* SNP detection. Libraries were prepared using MsII restriction enzyme (mean insert size of about 200 bp; recognition sequence and cut site, ↑↓: CAYNN↑↓NNRTG) and sequenced at LGC Biosearch Technologies (Berlin, Germany) on a 2 × 150 bp Illumina NextSeq® 550 platform (Illumina, San Diego, CA, USA). Library groups were demultiplexed into samples using the Illumina bcl2fastq v2.17.1.14 software (available at https://support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html) according to their inline barcodes and verification of restriction site (no mismatches or Ns were allowed in the inline barcodes, but Ns were allowed in the restriction site). Sequencing adapters were clipped from all reads, reads with final length < 20 bp were discarded and reads with 5' ends not matching the restriction site were removed as well. Reads were further filtered by quality, trimming at 3' end to get a Phred quality score ≥ 20 over a window of ten bases and discarding reads with a final length < 20 bp. Forward and reverse reads were assembled in contigs using BBMerge v34.48 (Bushnell et al., 2017). Unique (non-redundant) combined reads were obtained with CD-HIT-EST v4.6.1 (Li and Godzik, 2006) allowing up to 5% differences and excluding singletons and clusters containing < 20 reads. In absence of a reference genome for any of the taxa of interest, alignment of quality trimmed reads was carried out using a cluster reference in BWA-MEM v0.7.12 (Li and Durbin, 2009). SNP discovery and genotyping were performed with Freebayes v1.0.2-16 (Garrison and Marth, 2012) at the following parameters: min-base quality = 10, min-supporting-allele-qsum = 10, read-mismatch-limit = 3, min-coverage = 5, no-indels, min-alternate-count = 4, exclude-unobserved-genotypes, genotype-qualities, no-mnps, no-complex, mismatch-base-quality-threshold = 10). This analysis was initially run twice, separately setting ploidy to 2 and 4 according to

chromosome numbers known from literature for some of the investigated species. SNPs were then filtered under the following rules: variants had to exceed 8 reads per locus, genotypes had to be observed in at least 66% of samples, and the minimum allele frequency across samples should exceeded 5%. After results of ploidy estimation using NGS data (see below), we also generated a dataset with triploid counts.

Ploidy estimation from NGS data and chromosome counts

To estimate the ploidy of single individuals, we analyzed the GBS data using the software nQuire (Weiß et al., 2018). This software models the distribution of base frequencies of SNPs from BAM files using a Gaussian Mixture Model, compares it with the expected distribution in diploids, triploids and tetraploids, and uses a maximum likelihood approach to select the most plausible ploidy model. We used a minimum frequency of 0.2 for segregating bases (-f option), a minimum coverage of 10 (-c option) and a minimum mapping quality of 30 (-q option). Despite the stringent filters used, we also ran the "denoising" subcommand to reduce background noise in base frequency histograms due to mismappings, as suggested by the software manual. This subcommand uses a Gaussian Mixture Model with Uniform noise component (GMMU) (Weiß et al., 2018) to assess the extent of this uniform noise and scales it down to easily detect peaks in the histogram of base frequencies. The fixed model with the smallest $\Delta\log L$ is chosen as the true ploidy level. After the results of this analysis, we detected some triploids and, therefore, recalled SNPs in FreeBayes setting ploidy also to 3.

We compared such estimated ploidy values with chromosome counts for the following populations: aeoST, cinML, intSM, intVA, leuBA, lacCA, monAV, panVE, and tenMO. For this purpose, we employed cypselae deposited in the seed bank of the Botanic Garden of Naples or collected in field. Chromosomal observations were made on root tips obtained from the germination of cypselae in 1% agar Petri dishes. After the appearance of root tips, plates were stored for 12-24 hours at 4°C. Then, seeds with attached root tip were immersed in 0.4% colchicine for 4 hours and subsequently fixed in Carnoy fixative solution for 1 hour. After hydrolysis in 1N HCl at 60° C for 7 min, the tips were stained with Schiff's fuchsin-sulfite reagent (Sigma-Aldrich by Merck, St. Louis, MO, USA). Root tips were then soaked in 45% acetic acid, macerated and squashed. Metaphasic plates were observed, whenever possible, from root tips of 3-5 different individuals from seeds of at least 3 different mother-plants, using a Nikon Eclipse Ci-L microscope (Nikon, Tokyo, Japan).

Data filtering and population structure analyses

The final dataset containing mixed-ploidy individuals was further subject to filtering by missingness and linkage to reduce their impact of subsequent analyses. Specifically, we first calculated the missingness per locus, then ordered the loci by increasing missingness and retained, when multiple SNPs occurred in the same contig, only the one with less missing information across most of individuals. After this first selection, we filtered the dataset by missingness per individual, removing all the samples with missing data < 50%, and, per locus, removing all loci with ≥ 30% of missing data. To assess if the selected loci were neutral, in absence of a pipeline/software capable of handling data with mixed ploidy, we manually created an input file for BayeScan v2.1 (Foll and Gaggiotti, 2008) making all samples of the same ploidy (tetraploid) by adding one or two columns of missing data to triploids and diploids, respectively. The analysis was run with prior odds set to 100 and the other values as default. Convergence of runs and loci under selection were assessed following the tutorial available at http://evomics.org/wp-content/uploads/2016/01/BayeScan_BayeScEnv_exercises.pdf.

As first exploratory analysis of structure in our data, we carried out a principal component analysis (PCA) in adegenet v2.1.5 (Jombart, 2008; Jombart and Ahmed, 2011). After creating a 'genlight' object following the instructions in the package manual, we used the function 'glPca' to implement the analysis accommodating different levels of ploidy across samples in the data. We also used the function 'find.clusters' to group our samples in clusters based on their genetic similarity. This function first transforms the data using PCA (after specifying the number of retained PCs) and then choses the optimal number of clusters using the Bayesian information criterion (BIC). Estimates of allele frequency distribution and of expected (He) and observed heterozygosity (Ho) per locus were conducted in the same package using the functions 'gl.Mean', 'gl.He', and 'gl.Ho' respectively. Putative significant differences between He and Ho were evaluated with the Bartlett's test of homogeneity of variances.

To assess population structure and detect recent shared ancestry among population samples, we used fineRADstructure (Malinsky et al., 2018), a model-based Bayesian clustering approach that works with mixed-ploidy datasets from unmapped SNPs. In this approach, co-ancestry is determined calculating the number of SNPs shared by individuals (the "co-ancestry proportion"), and then assigning individuals with similar co-ancestry values to the same genetic cluster. We ran the analysis following the instructions provided on the software webpage (https://www.milan-malinsky.org/fineradstructure), also ordering loci according to linkage disequilibrium (script 'sampleLD.R') as suggested by the author for unmapped loci.

To evaluate the genetic make-up of triploid individuals discarded from previous analyses for their high degree of missingness, we selected a panel of 267 SNPs with locus missingness ≤ 0.03% and, retaining all the original 96 individuals (diploids, triploids, and tetraploids), we carried out an analysis in STRUCTURE v2.3.4 (Pritchard et al., 2000). This software was chosen over others commonly used in population assignation and admixture detection (e.g., ADMIXTURE (Alexander et al., 2009), fastSTRUCTURE (Raj et al., 2014)) because of its capability and robustness at handling datasets with mixed-ploidy populations (Stift et al., 2019). To estimate the number of genetic clusters (K), STRUCTURE was run with K varying from 2 to 17 for 100,000 generations, the first 50,000 of which were discarded as burn-in. Each value of K was evaluated

using ten independent Markov Chain Monte Carlo (MCMC) replicates. To choose the best value of K, we looked at the plot showing the log probability of the data [ln Pr(X|K)] against a range of K values as provided by STRUCTURE HARVESTER (Earl and vonHoldt, 2012) and selected the K after which the trend of ln Pr (X|K) reaches a plateau (Pritchard et al., 2000, Pritchard et al., 2010; Gilbert et al., 2012). We chose this method over that of Evanno (Evanno et al., 2005) because it has been demonstrated that the latter is strongly biased towards K=2 (Janes et al., 2017). The results of the ten MCMC iterations of each K obtained from STRUCTURE HARVESTER were combined and analysed in CLUMPAK (Kopelman et al., 2015) using LargeKGreedy algorithm (Jakobsson and Rosenberg, 2007) and visualised in the same webserver using DISTRUCT (Rosenberg, 2004).

Genetic differentiation among population groups identified by previous analyses was calculated using the Fixation index (F_{ST}) (Weir and Cockerham, 1984) in the R package StAMPP (Pembleton et al., 2013). Population samples constituted by 1-2 individuals (i.e., aeoST, ambPC, HybVA) were removed from the analysis. Significance of pairwise F_{ST} values was determined after 1000 replications.

Results

Ploidy estimation and chromosome count

We obtained from SNP data analysis (through nQuire software) the following estimation for each taxon (see Table 2 for the single

populations, complete data in Supplementary Table 1): *C. aeolica* (2n = 36), *C. ambigua* subsp. *ambigua* (2n = 27, 36), *C. deusta* subsp. *deusta* (2n = 18), *C. deusta* subsp. *leucolepis* (2n = 36), *C. lacaitae* (2n = 27, 36), *C. montaltensis* (2n = 27, 36), *C. pandataria* (2n = 18), *C. tenorei* (2n = 27, 36). Intermediate populations (intSM, intVA, intVI) resulted tetraploid, or rarely triploids (see Table 2 for further details). Metaphasic chromosome counts were obtained for 9 populations (Table 2) and returned the same results of NGS estimation, except for *C. aeolica*, which was counted as diploid (2n = 18).

Overview of GBS data

Illumina sequencing returned ~2.7 million reads per sample, for a total of 205,140 loci (~13× coverage). Of these loci, 130,411 were polymorphic and contained 466,078 SNPs. After filtering procedures described in Materials and Methods, we obtained two datasets: one, including 7,473 loci and 75 individuals and utilized for PCA, fineRADstructure, and F_{ST} inferences (Supplementary Table 2); a second one, with 267 loci and all 96 individuals for STRUCTURE analysis (Supplementary Table 3). From the first dataset, one individual of *C. ambigua*, two individuals of *C. deusta* (from Mt. Molare) and one intermediate individual (from intVI), as well as all the triploid samples were discarded because afflicted by high individual missingness (mean = 0.68). BayeScan analysis identified only 37 loci out of 7473 under positive selection (qval < 0.05) (Supplementary Table 4), which were kept in the final dataset

TABLE 2 Ploidy of taxa under investigation estimated by chromosome counting and NGS data.

Taxon	Population	Count	Ploidy from NGS	Literature
C. lacaitae	lacCA	2n = 36	tetraploid (triploid)	2n = 36
C. montaltensis	monAV	2n = 36	tetraploid (triploid)	2n = 36
C. tenorei	tenMO	2n = 36	tetraploid (triploid)	2n = 18
Possible intermediates between *C. cineraria* and *C. tenorei*	intSM	2n = 27	tetraploid (triploid)	–
	intVA	2n = 36	tetraploid (triploid)	–
	intVI	–	tetraploid (rarely triploid)	–
C. cineraria subsp. *cineraria*	cinCA	–	diploid	2n = 18, 36
	cinMI	2n = 18	diploid (rarely tetraploid)	
	cinML	–	diploid	
Centaurea cf. *cineraria*	cixVA	–	diploid	–
C. deusta subsp. *deusta*	deuFI	–	diploid	2n = 18
	deuMO		diploid	
C. deusta subsp. *leucolepis*	leuBA	2n = 36	tetraploid	2n = 36
C. deusta hybrid	hybVA	–	diploid	–
C. aeolica	aeoST	2n = 18	tetraploid	2n = 18
C. ambigua subsp. *ambigua*	ambPC	–	triploid/tetraploid	2n = 36
C. pandataria	panVE	2n = 18	diploid	2n = 18

Literature data are retrieved from Bedini and Peruzzi (2021). No viable root tip was obtained for *C. deusta*.

for subsequent analyses, after verifying that the results of an exploratory PCA were completely overlapping to those of the full dataset.

Mean observed heterozygosity was 0.120, significantly lower (p < 0.001) than expected heterozygosity (= 0.285) (Supplementary Figures 1 and 2). Most of loci contained alleles with frequency between 0.2 and 0.9, and only a limited portion of them was fixed (Supplementary Figure 3).

Population structure and genetic differentiation

The PCA explained the 39.9% of the variance along the three main axes (PC1 = 25.2; PC2 = 7.8%; PC3 = 6.9%) and showed that all the operational taxonomic units (OTUs) from the same population clustered together (Figure 2), with the partial exception of two samples from cixVA. *Centaurea pandataria* was distant from all the other populations, and therefore can be considered as the best outgroup for our system. Halfway between this group and the others, the OTUs relative to *C. aeolica* and *C. ambigua* subsp. *ambigua* are localized. The three microspecies of the *C. tenorei* group are very close and can be distinguished only in one projection (PC1/PC3); in all cases, they are very close to the intermediate populations (intSM, intVA, intVI), which effectively are halfway between them and the *C. cineraria* group (Figure 2). No difference was found among individuals of monAV. OTUs representing *C. deusta* are consistently distinct from the others.

The co-ancestry matrix resulting from the fineRADstructure analysis (Figure 3) showed a complex pattern of gene flow in *C. tenorei* s.l. For instance, *C. montaltensis* had high co-ancestry with the neighboring *C. deusta* (subsp. *deusta* and subsp. *leucolepis*), as well as with *C. tenorei* and *C. lacaitae* and, to a lower extent, with

the intermediate populations. These latter ones had comparable levels of ancestry with *C. tenorei* s.l. and *C. cineraria*; *C. cineraria* showed high co-ancestry with *C. pandataria* and some individuals from intVI and intSM. Within *C. cineraria*, the sampled populations resulted highly structured: we observed four distinct sub-clusters, corresponding to cinCA, cinML (these two very close each to the other), cinMI, and cixVA (Figure 3). The two subspecies of *C. deusta* have, as expected, high co-ancestry, as well as *C. aeolica* and *C. pandataria*. The relationships of *C. ambigua* are not clear: the highest co-ancestry is here found with *C. tenorei* s.l. and *C. pandataria*.

STRUCTURE analysis indicated K = 7 as the optimal number of clusters (Supplementary Figure 4). Due to their possible biological relevance, other genetic clusters, i.e., K =4, K = 5, and K = 6, were also displayed in Figure 4. Indeed, K = 4 provided the simplest biological explanation for the genetic structure: a first group including *C. aeolica* and *C. pandataria* (aeoST and panVE), a second group encompassing *C. cineraria* (cinCA, cinMI, and cinML) and specimens resembling it (cixVA), a third group with *C. deusta* (deuFI, deuMO, and leuBA), and a last one including populations attributable to *C. tenorei* s.l. (lacCA, monAV, and tenMO) and its intermediates (intSM, intVI, and intVA). The individual labelled as hybVA was confirmed as a hybrid between *C. cineraria* and *C. deusta*. At K = 5, the coastal population intVI was further separated from the others; similarly, at K = 6, the other coastal population, lacCA was segregated from the *C. tenorei* aggregate. At K = 7, the population cixVA was separated from the other *C. cineraria* samples. As a consequence, the specimen hybVA, which is sympatric with population cixVA, is obviously a hybrid between the latter population and *C. deusta*. For all Ks, *C. montaltensis* and *C. tenorei* appeared very similar to each other and remarkably more admixed with *C. deusta* than all the other populations of the complex.
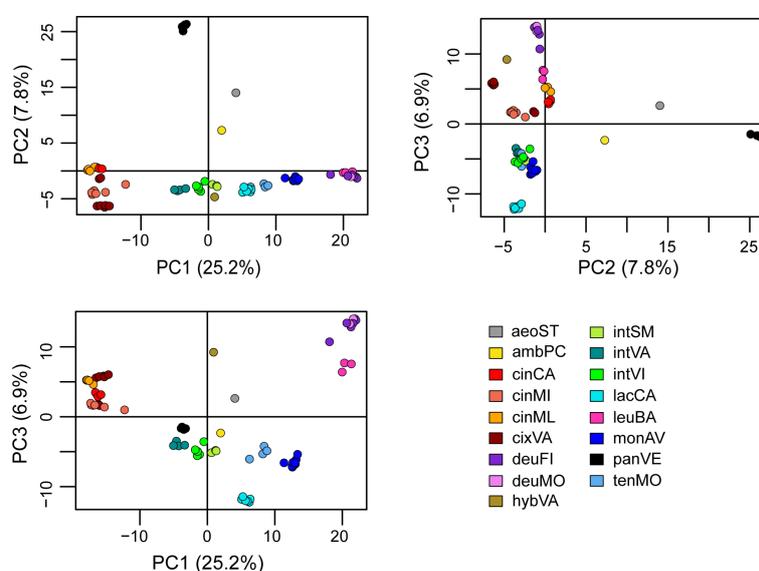

FIGURE 2
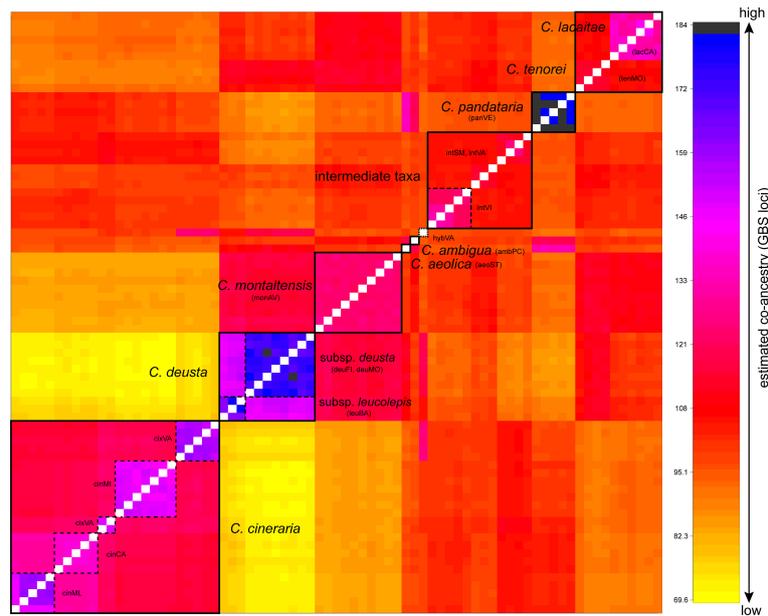PCA analysis. For explanation of legend see Table 1.

FIGURE 3
Co-ancestry plot from fineRADstructure analysis. Boxes with dashed lines highlight population samples or subspecies, while boxes with filled lines indicate well-defined groups. Co-ancestry levels increase from yellow to black.

F_{ST} analysis (Supplementary Table 5) showed that between *C. montaltensis* and *C. tenorei* (0.092) a low genetic differentiation is present, altogether comparable to that between the two populations of *C. deusta* subsp. *deusta* (0.089). A slightly higher differentiation is observed between lacCA and monAV/tenMO (0.178 and 0.191, respectively). The plants with intermediate features (intVA, intSM, intVI) had F_{ST} values < 0.25 with *C. tenorei* s.l., with the only exception of lacCA vs. intVI (0.277); genetic structure was higher when comparing it with that of *C. cineraria* (~0.35) and *C. deusta* (~0.48). Within *C. cineraria*, the lowest differentiation was found between the populations from cinCA and cinML (geographically near), a result also evident in the fineRADstructure analysis.

Discussion

Ploidy, speciation, and diversification in *Centaurea tenorei* group

Despite current taxonomy distinguishes three microspecies within *Centaurea tenorei* s.l. also on account of their different ploidy (Peruzzi and Scoppola, 2008; Santangelo et al., 2017), our results show that all the taxa of this group are consistently tetraploid. In fact, after five direct counts (Table 2) and 50 NGS estimations in this group (Supplementary Table 1) we detected or postulated only 17 triploids in the group, and no diploids. Tetraploids were first indicated by Peruzzi and Cesca (2002) for
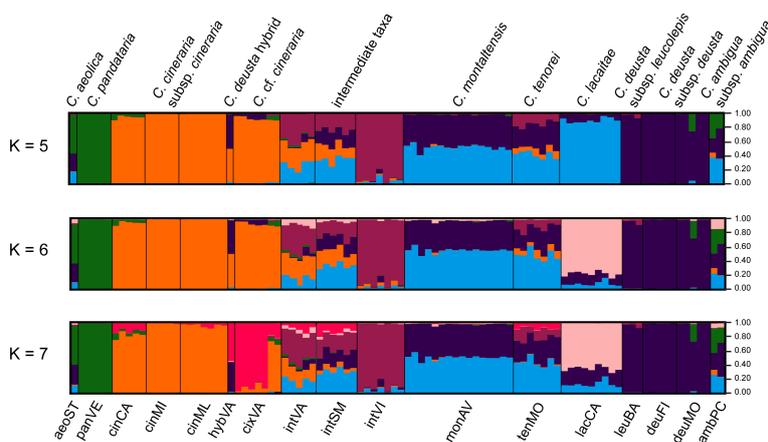

FIGURE 4
STRUCTURE plots. The analysis was carried out using 96 individuals and a selected panel of 267 SNPs.

C. montaltensis and *C. lacaitae*. However, the same authors reported as diploids a population "near Scala and Ravello" and another from "Mt. Sant'Angelo a Tre Pizzi" (*C. tenorei*). It is to be noted that the latter population corresponds to the tetraploid tenMO, whereas we could not find the former after several field trips. On Mt. Molare, we found diploid only the individuals belonging to *C. deusta* subsp. *deusta*, which is sympatric in that locality with *C. tenorei* and is widespread also elsewhere. Therefore, contrary to what previously supposed (Peruzzi and Scoppola, 2008; Pignatti, 2018), the remarkable morphological variability of the group is independent of ploidy. Such variability is especially surprising in a narrow endemic, and it is not shared with any other Campanian endemic (Del Guacchio et al., 2020b). If on one hand, previous scholars (Lacaita, 1922; Guadagno, card-index; Barone Lumaga and Menale, 2000) indicate that this variation has a large environmental component (especially regarding the vegetative features), on the other hand, STRUCTURE analysis (Figure 4) reveals that this variability is mainly due to the genetic influence of *C. cineraria* and *C. deusta*. Hybridization and introgression are obvious responsible of great variability in plants in general and also in *Centaurea* (López-Pujol et al., 2012; Mameli et al., 2014; Requena et al., 2020). Hybridization phenomena in *C. tenorei* s.l. were already hypothesized by several authors founding on morphological observations (Lacaita, 1922; Guadagno, 1932; Del Guacchio et al., 2022), but a definitive proof was still lacking. In this regard, they pointed especially to *C. cineraria* and considered hybrids between *C. tenorei* s.l. and *C. deusta* as rare. In particular, Guadagno (1932) considered *C. tenorei* s.l. the stabilized hybrid between *C. dissecta* Ten. (= *C. ambigua* s.l.) and *C. cineraria*. Unexpectedly, our results revealed a greater involvement of *C. deusta*: according to STRUCTURE and fineRADstructure (Figures 3 and 4), admixture of *C. deusta* occurs throughout the populations ascribed to our study group and its intermediates, with the partial exception of intVI, more influenced by *C. cineraria* (Figure 4, K =4). Despite the clinal distribution of monAV from 600 to 1000 m a.s.l., this population was genetically homogeneous. Half of the allelic pool of monAV is shared with *C. deusta* subsp. *deusta*. Marginally, we note that this parentage was possibly perceived by Lacaita (1922), who described the basal leaves of *C. tenorei* as "somewhat resembling that of *C. deusta*" and essentially invariant throughout the different populations examined by him. Moreover, he regretted that this useful feature was difficult to compare because the basal leaves wither at flowering time in this and in allied species. Even if the two subspecies of *C. deusta* do not differ in their allelic pattern, we exclude that the tetraploid *C. deusta* subsp. *leucolepis* is involved in the current gene flow with *C. tenorei* s.l., because of geographic distance. On the contrary, the only literature tetraploid count known for *C. cineraria* subsp. *cineraria* (Damboldt and Matthäs, 1975) is from a locality ("Rocky slope at a lookout point coastal road about 5.7 km west of Positano"), where intermediate populations, which resulted tetraploid in our analysis, may occur.

These considerations entail an intriguing issue: as *C. tenorei* s.l. is tetraploid, why we observe such a wide gene flow involving both *C. cineraria* (2×) and *C. deusta* subsp. *deusta* (2×)? This is difficult to explain by frequent hybridizations between these two latter taxa

and diploid individuals of *C. tenorei* s.l., (Peruzzi and Cesca, 2002) because these latter must be very rare (and indeed were not found by us). Equally unlikely is the hypothesis of a tetraploid ancestor of the group, later hybridized with diploid taxa, as this would not justify the current tetraploid populations. The only parsimonious hypothesis is that a diploid ancestor of *C. tenorei* s.l. was sufficiently compatible with both *C. cineraria* and *C. deusta* to produce homoploid hybrids. The obvious consequence of this scenario would be progressive introgression of hybrid populations towards the more abundant diploid parentals, and the final dilution of the gene pool of the postulated ancestor. However, if some of these hybrids developed into autotetraploids, may have originated established hybridogenous populations. Polyploidy, in fact, is unanimously regarded as one of the driving forces of plant evolution and polyploids (autotetraploids, in our case) perform better in colonizing new habitats (Soltis et al., 2016; Becher et al., 2020). This can easily explain both the morphological variability of *C. tenorei* s.l., and its wide altitudinal range (from 1400 m to 10 m a.s.l.), a plasticity shared with *C. deusta*.

The origin of the postulated ancestor of *C. tenorei* s.l. remains obscure and might be clarified only by an extensive study involving the allied taxa of the group; this in turn is a rather insidious issue, because the obscure phylogeny of the section does not allow to identify them easily (Hilpold et al., 2014b). However, in our opinion, the hypothesis first suggested by Guadagno (1932) about the involvement of *C. ambigua* s.l. or of a taxon allied to it, remains the most suggestive. In fact, *C. ambigua* s.l. and *C. tenorei* s.l. share unquestionable morphological similarity (Pignatti and Lausi, 1982). Moreover, *C. ambigua* s.l. is geographically nearest to *C. tenorei* s.l. among the similar species (Lacaita, 1922); in this respect, we mention that it has been recently collected by us on Lago Matese (northern Campania), and possibly it occurs even southward (Terracciano, 1874; PFI [Portale della Flora d'Italia], 2022, sub *C. dissecta*). It has to be noted, however, that the most widespread subspecies, i.e., *C. ambigua* subsp. *ambigua*, as far as it is known, is tetraploid (Table 2).

The allelic frequencies are substantially identical in each population of *C. tenorei* s.l. (intermediates included) but different across them (Figures 3 and 4). This fact suggests that most of the gene flow occurred in the past and present-day populations have been founded by parentals with different introgression degrees. This is clearly reflected by the scattered distribution of the populations of the group and by their morphological distinctness, which induced several taxonomists to split the group into several taxa, ranking from species to forms (Guadagno, card-index; Santangelo et al., 2017). A remarkable exception to this general genetic distinctiveness of single populations can be found in the locality called "Valico di Chiunzi", a pass in the hills between the northern and the southern slopes of the Peninsula of Sorrento, and therefore a natural way of passage. Incidentally, we note that it is densely traveled by man, who doubtlessly facilitated the contact of different populations by preparing new paths in the slopes, digging caves and building towers (which are preferred habitats of these plants). It is suggestive that in this locality, we find two different genetic pools of *C. cineraria*: the most common is private to that population,

whereas two individuals share more affinity with the other populations of the same taxon (Figure 4). At Valico di Chiunzi, hybridization has been documented since a century ago (Guadagno, 1932; Del Guacchio et al., 2020a) and is still very active, involving *C. deusta* subsp. *deusta*, *C. tenorei* s.l., and *C. cineraria* subsp. *cineraria* (Figures 3 and 4). We can add that, according to literature (e.g., Pignatti and Lausi, 1982), the natural range of *C. cineraria* is strictly linked to the coast. More precisely, Guadagno (card-index) did not observe it above 350 m a.s.l., even if he gathered on the pass as well. The plant might have been overlooked by him or confused with *C. tenorei* s.l. but this seems unlikely. Indeed, we cannot exclude that *C. cineraria* has reached the pass only recently, possibly also driven by climate change. The altitudinal shift upward has been documented for several mountain plants (Lenoir et al., 2008; Telwala et al., 2013; Zu et al., 2021) but, to the best of our knowledge, not well known in coastal plants.

Taxa delimitation in *C. tenorei* s.l.

The populations ascribed to *C. tenorei* s.l. (i.e., tenMO, monAV, lacCA) and those indicated as 'intermediate' (i.e., intVI, intSM, intVA) can be grouped in the PCA graphs (Figure 2, especially in the PC1\PC3), but with a rather dispersed pattern. In detail, tenMO and monAV, i.e., the typical populations of *C. tenorei* and *C. montaltensis* respectively, are very close (Figure 2). The F_{ST} value between tenMo and monAV (0.092, Supplementary Table 5) is lower than the average value found between other plant species and in the average for different varieties (Huang et al., 2020; Chambers et al., 2021); also for the other taxa of *C. tenorei* s.l. and its intermediate forms, F_{ST} values are in agreement with subspecific ranks in other studies (e.g., Yu et al., 2022). The genetic structure of the two populations, however, is not identical, because that of *C. tenorei* is more complex (Figure 4). Morphologically, *Centaurea tenorei* and *C. montaltensis* are very weakly distinguishable (Lacaita, 1922; Peruzzi and Scoppola, 2008; Santangelo et al., 2017, Table 1), and several scholars do not recognize any taxonomic distinctness (e.g., Fiori, 1927). Moreover, the presumed difference about the ratio between the pappus and the cypsela, as well as the color of the bracteal ciliae vary independently within several population (Guadagno, card-index; pers. obs.), while the head width of *C. montaltensis* is fully included in the range of *C. tenorei*. Therefore, after our result on the tetraploid chromosome number shared by both taxa, distinctness of *C. montaltensis* does not seem furtherly sustainable. The population lacCA appears as more distinct, what would suggest in turn a taxonomic recognition. In this regard, we note that, according to our results, lacCA is one of the least admixed populations, and maybe the one preserving the plesiomorphic features of the unknown ancestor. However, traces of the genetic contribution by *C. deusta* are present as well (Figure 4). More importantly, several other similar populations included in *C. lacaitae* occur along the coasts (Santangelo et al., 2017), and we suspect that their inclusion in our study would have furtherly diluted the genetic isolation of lacCA. For these reasons, any taxonomic distinction of the typical *C. lacaitae* from *C. tenorei*

results problematic and should be considered with caution. It is to be noted that *C. lacaitae*, corresponding to *C. dissecta* Ten. var. *maritima*, as circumscribed by Lacaita (1922) and recent authors (Peruzzi and Scoppola, 2008; Santangelo et al., 2017) includes not only the typical population lacCA, but also the 'intermediate' ones (intVI, intSM, intVA), formerly distinct as *C. dissecta* var. *maritima* f. *cinerarioides* Lacaita ex Fiori. Guadagno (card-index), for example, did not accept var. *maritima* as a separate taxon, but was aware that f. *cinerarioides* would deserve recognition.

Centaurea deusta

According to all results, the populations ascribed to *C. deusta* group together, also including the subsp. *leucolepis*. This latter remarkable subspecies differs from the typical *C. deusta* on account of a combination of morphological features independently and sporadically occurring also in the subsp. *deusta* (i.e., pale, muticous involucral bracts, whitish flowers, often fleshy leaves), of the maritime ecology and tetraploid chromosome number (Del Guacchio et al., 2022). Hilpold et al. (2015, sub *C. corensis*) suggested an allopolyploid origin of this taxon. Nevertheless, our results better support a direct differentiation from *C. deusta* by local autotetraploidy as an adaptation to coastal environments (see also Del Guacchio et al., 2022). In fact, even at highest K values in the Structure analysis (Supplementary Figure 5), the genetic structure of the investigated population (leuBA), is very homogeneous and apparently not different from that of two separate mountain populations of *C. deusta* subsp. *deusta* (deuMO, deuAV). Taxonomically, the results of the co-ancestry analysis with fineRADstructure (Figure 3) and of PCA (Figure 2) concur with the subspecific rank already proposed by Del Guacchio et al. (2022) on morphological grounds.

New hybrids

Hybrids between the homoploid taxa *C. cineraria* s.l. and *C. deusta* subsp. *deusta* seem very rare, despite the natural range of *C. cineraria* (south-western coasts of Italy) is entirely included in that of the other species (Italy, Switzerland, and Balkan peninsula). Indeed, a single hybrid between them is documented here for the first time (the individual labelled as hybVA). We may presume that *C. deusta* and *C. cineraria*, belonging to two different subsections (Hilpold et al., 2014a), are poorly compatible and not closely related. This, however, might be the consequence of the slightly different habitat of the two species, or of overlooked hybrids. The results concur with the hypothesis by Del Guacchio et al. (2022) about the hybrid between *C. tenorei* (mainly tetraploid) and *C. deusta* subsp. *deusta*, i.e., *C. ×cavarae* Guadagno ex Del Guacchio, Cennamo & Caputo, only known for a gathering in the same place (Valico di Chiunzi). A form morphologically similar to *C. ×cavarae* still occurs in its *locus classicus*. Further and finer analyses are required to assess the precise identity of these forms.

Conclusion

This study demonstrates that the taxa belonging to the *C. tenorei* aggregate, i.e., *C. tenorei*, *C. lacaitae*, and *C. montalensis*, including the intermediate forms, are chiefly tetraploid and that the current recognition of three microspecies is not supported by genetic data. These taxa appear of hybridogenous origin, with an important contribution of *C. deusta* (or its ancestor) and, to a lesser extent, *C. cineraria*. Some populations of this aggregate, however, have unique genetic features and may be regarded as evolutionary significant units (ESUs). We also find that the locality "Valico di Chiunzi" is particularly interesting from the evolutionary and ecological point of views because it harbors populations of *C. cineraria* out of their ecological range that hybridize with *C. deusta* and *C. cineraria*, creating an intricate pattern of relationships across species. Furthermore, this study confirms that SNPs markers are a powerful tool to analyze the genetic structure of difficult plant systems where mixed-ploidy, hybridization and introgression occur and to determine ploidy at individual level in the absence of chromosome counts.

Data availability statement

The raw Illumina sequences used in the study are available in the Sequence Read Archive (SRA) repository under NCBI BioProject PRJNA928762. Voucher specimens are deposited at the Herbarium of the Botanical Garden of Naples (NAP) and available upon request. Supporting information to the paper is available as Supplementary Material.

Author contributions

EDG and PCE contributed to conception and design of the study. Mainly EDG and PCE field-collected plant material, but all authors collected at least in one locality. DL carried out molecular work and data analysis. LP was responsible for chromosome counts. DDL, EDG, and PCA wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version. PCA supervised the whole work.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1130889/full#supplementary-material

References

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Allendorf, F. W., Luikart, G. H., and Aitken, S. N. (2013). *Conservation and the genetics of populations. 2nd Edition* (Chichester: Wiley-Blackwell).

Arnelas, I., Pérez-Collazos, E., Devesa, J. A., López, E., and Catalan, P. (2018). Phylogeny of highly hybridogenous Iberian *Centaurea* l. (Asteraceae) taxa and its taxonomic implications. *Plant Biosyst.* 152, 1182–1190. doi: 10.1080/11263504.2018.1435569

Arnelas, I., Pérez-Collazos, E., Devesa, J. A., Manzaneda, A. J., and Catalan, P. (2020). Taxonomic differentiation of Iberian knapweeds (*Centaurea* sects. *Jacea* and *Lepteranthus*, asteraceae) and genetic isolation of infraspecific floral morphotypes. *Ann. Mo. Bot. Gard.* 105, 481–501. doi: 10.3417/2020389

Barone Lumaga, M. R., and Menale, B. (2000). Morpho-functional features of leaves of *Centaurea tenorei* Guss. ex Lacaita (Asteraceae). *Delpinoa n.s.* 39-40 (1997-1998), 85–98. Available at: https://www.ortobotanico.unina.it/OBN4/mv/biblioteca/Delpinoa/volumi/39-40%20ocr/Delpinoa_39-40_1997-98_085-098.pdf.

Becher, H., Brown, M. R., Powell, G., Metherell, C., Riddiford, N. J., and Twyford, A. D. (2020). Maintenance of species differences in closely related tetraploid parasitic *Euphrasia* (Orobanchaceae) on an isolated island. *Plant Commun.* 6, 100105. doi: 10.1016/j.xplc.2020.100105

Bedini, G., and Peruzzi, L. (2021, onwards) *Chrobase.it – chromosome numbers for the Italian flora v. 2.0*. Available at: http://bot.biologia.unipi.it/chrobase/ (Accessed 6 December 6, 2022]).

Blair, A. C., and Hufbauer, R. A. (2010). Hybridization and invasion: one of north america's most devastating invasive plants shows evidence for a history of interspecific hybridization. *Evol. Appl.* 3, 40–51. doi: 10.1111/j.1752-4571.2009.00097.x

Bona, M. (2015). Systematic implications of achene characteristics in genera *Centaurea* l., *Cyanus* mill., *Psephellus* cass. and *Rhaponticoides* vaill. (Asteraceae). *Bangladesh J. Plant Taxon.* 22, 125–136. doi: 10.3329/bjpt.v22i2.26074

Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge–accurate paired shotgun read merging via overlap. *PloS One* 12, e0185056. doi: 10.1371/journal.pone.0185056

Chambers, A., Cibrián-Jaramillo, A., Karremans, A. P., Martinez, D. M., Hernandez-Hernandez, J., Brym, M., et al. (2021). Genotyping-By-Sequencing diversity analysis of international vanilla collections uncovers hidden diversity and enables plant improvement. *Plant Sci.* 311, 111019. doi: 10.1016/j.plantsci.2021.111019

Cheng, J., Kao, H., and Dong, S. (2020). Population genetic structure and gene flow of rare and endangered *Tetraena mongolica* maxim. revealed by reduced representation sequencing. *BMC Plant Biol.* 20, 391. doi: 10.1186/s12870-020-02594-y

Damboldt, J., and Matthäs, U. (1975). Chromosomenzahlen einiger mediterraner und mitteleuropaeischer *Centaurea*-arten (Asteraceae). *Plant Syst. Evol.* 123, 107–115. doi: 10.1007/BF00981769

De Castro, O., Del Guacchio, E., Di Iorio, E., Di Maio, A., Di Martino, L., Bartolucci, F., et al. (2020). Barcoding helps threatened species: the case of *Iris marsica* (Iridaceae) from the protected areas of the Abruzzo (Central Italy). *Plant Biosyst.* 154, 961–972. doi: 10.1080/11263504.2020.1762786

Del Guacchio, E., Cennamo, P., Paino, L., and Caputo, P. (2022). Further remarks on the narrow endemic *Centaurea pandataria* (Asteraceae, Cardueae). *Fl. Medit.* 31 (Special Issue), 469–475. doi: 10.7320/FlMedit31SI.469

Del Guacchio, E., Iamonico, D., Cennamo, P., and Caputo, P. (2020a). Nomenclatural and taxonomic notes on some *Centaurea* taxa (Asteraceae) from southern Italy. *Turk. J. Bot.* 44, 441–454. doi: 10.3906/bot-1912-9

Del Guacchio, E., Innangi, M., Giacò, A., Peruzzi, L., and Caputo, P. (2020b). Taxa endemic to Campania (southern italy): nomenclatural and taxonomic notes. *Phytotaxa* 449, 217–231. doi: 10.11646/phytotaxa.449.3.2

Dufresne, F., Stift, M., Vergilino, R., and Mable, B. K. (2014). Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Mol. Ecol.* 23, 40–69. doi: 10.1111/mec.12581

Earl, D. A., and von Holdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

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

Favre, F., Jourda, C., Besse, P., and Charron, C. (2021). Genotyping-by-Sequencing technology in plant taxonomy and phylogeny. *Methods Mol. Biol.* 2222, 167–178. doi: 10.1007/978-1-0716-0997-2_10

Fiori, A. (1927). *Nuova flora analitica d'Italia 2(5)* (Firenze: Tipografia di M. Ricci).

Foll, M., and Gaggiotti, O. E. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180, 977–993. doi: 10.1534/genetics.108.092221

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

Font, M., Garnatje, T., Garcia-Jacas, N., and Susanna, A. (2002). Delineation and phylogeny of *Centaurea* sect. *Acrocentron* based on DNA sequences: a restoration of the genus *Crocodylium* and indirect evidence of introgression. *Plant Syst. Evol.* 234, 15–26. doi: 10.1007/s00606-002-0203-3

Garcia-Jacas, N. (1998). *Centaurea kunkelii* (Asteraceae, Cardueae), a new hybridogenic endecaploid species of sect. *Acrocentron* from Spain. *Ann. Bot. Fenn.* 35, 159–167. Available at: https://www.jstor.org/stable/23726556

Garcia-Jacas, N., Soltis, P. S., Font, M., Soltis, D. E., Vilatersana, R., and Susanna, A. (2009). The polyploid series of *Centaurea toletana*: glacial migrations and introgression revealed by nrDNA and cpDNA sequence analyzes. *Mol. Phylogenet. Evol.* 52, 377–394. doi: 10.1016/j.ympev.2009.03.010

Garcia-Jacas, N., Uysal, T., Romashchenko, K., Suárez-Santiago, V. N., Ertuğrul, K., and Susanna, A. (2006). *Centaurea* revisited: a molecular survey of the *Jacea* group. *Ann. Bot.* 98, 741–753. doi: 10.1093/aob/mcl157

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv* 1207.3907v2 [q-bio.GN]. doi: 10.48550/arXiv.1207.3907

Gilbert, K. J., Andrew, R. L., Bock, D. G., Franklin, M. T., Kane, N. C., Moore, J. S., et al. (2012). Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. *Mol. Ecol.* 21, 4925–4930. doi: 10.1111/j.1365-294X.2012.05754.x

Gompert, Z., and Mock, K. E. (2017). Detection of individual ploidy levels with genotyping-by-sequencing (GBS) analysis. *Mol. Ecol. Resour.* 17, 1156–1167. doi: 10.1111/1755-0998.12657

Guadagno, M. (1932). Flora caprearum Nova [II]. *Arch. Bot. (Forlì)* 8, 65–80, 143–158, 275–295.

Guarino, C., and Rampone, S. (2006). A morphometric analysis of *Centaurea* sect. *Dissectae* (Compositae). *Bocconea* 16, 1–16. Available at: https://www.herbmedit.org/bocconea/19-077.pdf.

Hatmaker, E. A., Staton, M. E., Dattilo, A. J., Hadziabdic, Đ., Rinehart, T. A., Schilling, E. E., et al. (2018). Population structure and genetic diversity within the endangered species *Pityopsis ruthii* (Asteraceae). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00943

Herrando-Moraira, S., Calleja, J. A., Galbany-Casals, M., Garcia-Jacas, N., Liu, J. Q., López-Alvarado, J., et al. (2019). Nuclear and plastid DNA phylogeny of tribe Cardueae (Compositae) with hyb-seq data: a new subtribal classification and a temporal diversification framework. *Mol. Phylogenet. Evol.* 137, 313–332. doi: 10.1016/j.ympev.2019.05.001

Hilpold, A., Garcia-Jacas, N., Vilatersana, R., and Susanna, A. (2014a). Taxonomical and nomenclatural notes on *Centaurea*: a proposal of classification, a description of new sections and subsections, and a species list of the redefined section centaurea. *Collect. Bot.* 33, e001. doi: 10.3989/collectbot.2013.v33.001

Hilpold, A., López-Alvarado, J., Garcia-Jacas, N., and Farris, E. (2015). On the identity of a *Centaurea* population on procida island, Italy: *Centaurea corensis* rediscovered. *Plant Biosyst.* 149, 1025–1035. doi: 10.1080/11263504.2014.983578

Hilpold, A., Schönswetter, P., Susanna, A., Garcia-Jacas, N., and Vilatersana, R. (2011). Evolution of the central Mediterranean *Centaurea cineraria* group (Asteraceae): evidence for relatively recent, allopatric diversification following transoceanic seed dispersal. *Taxon* 60, 528–538. doi: 10.1002/tax.602019

Hilpold, A., Vilatersana, R., Susanna, A., Meseguer, A. S., Boršić, I, Constantinidis, T., et al. (2014b). Phylogeny of the *Centaurea* group (*Centaurea*, compositae) – geography is a better predictor than morphology. *Mol. Phylogenet. Evol.* 77, 195–215. doi: 10.1016/j.ympev.2014.04.022

Huang, L., Wang, X., Dong, Y., Long, Y., Hao, C., Yan, L., et al. (2020). Resequencing 93 accessions of coffee unveils independent and parallel selection during *Coffea* species divergence. *Plant Mol. Biol.* 103, 51–61. doi: 10.1007/s11103-020-00974-4

Hyun, D. Y., Sebastin, R., Lee, K. J., Lee, G. A., Shin, M. J., Kim, S. H., et al. (2020). Genotyping-by-sequencing derived single nucleotide polymorphisms provide the first well-resolved phylogeny for the genus *Triticum* (Poaceae). *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00688

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

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., et al. (2017). The k= 2 conundrum. *Mol. Ecol.* 26, 3594–3602. doi: 10.1111/mec.14187

Jombart, T. (2008). *Adegenet*: a r package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129

Jombart, T., and Ahmed, I. (2011). *Adegenet 1.3-1*: new tools for the analysis of genome-wide SNP data. *Bioinformatics.* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521

Kienberg, O., and Becker, T. (2017).). differences in population structure require habitat-specific conservation strategies in the threatened steppe grassland plant *Astragalus exscapus*. *Biol. Conserv.* 211 (A), 56–66. doi: 10.1016/j.biocon.2017.05.002

Kolář, F. (2021). Tracing evolutionary history and admixture in mixed-ploidy systems. *Mol. Ecol. Resour.* 21, 1413–1415. doi: 10.1111/1755-0998.13390

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191. doi: 10.1111/1755-0998.12387

Koutecký, P. (2007). Morphological and ploidy level variation of *Centaurea phrygia* agg. (Asteraceae) in the Czech republic, Slovakia and Ukraine. *Folia Geobot* 42, 77–102. doi: 10.1007/BF02835103

Lacaita, C. (1922). Piante italiane critiche o rare. LXXIVLXXXIII. *Nuovo Giorn. Bot. Ital. n.s.* 29, 174–194.

Lenoir, J., Gégout, J. C., Marquet, P. A., Ruffray, P., and de Brisse, H. (2008). A significant upward shift in plant species optimum elevation during the 20th century. *Science* 320, 1768–1771. doi: 10.1126/science.1156831

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

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

López-Alvarado, J., Sáez, L., Filigheddu, R., Garcia-Jacas, N., and Susanna, A. (2014). The limitations of molecular markers in phylogenetic reconstruction: the case of *Centaurea* sect. *Phrygia* (Compositae). *Taxon* 63, 1079–1091. doi: 10.12705/635.6

López-Pujol, J., Garcia-Jacas, N., Susanna, A., and Vilatersana, R. (2012). Should we conserve pure species or hybrid species? delimiting hybridization and introgression in the Iberian endemic *Centaurea podospermifolia*. *Biol. Conserv.* 152, 271–279. doi: 10.1016/j.biocon.2012.03.032

Malinsky, M., Trucchi, E., Lawson, D. J., and Falush, D. (2018). RADpainter and fineRADstructure: population inference from RADseq data. *Mol. Biol. Evol.* 35, 1284–1290. doi: 10.1093/molbev/msy023

Mameli, G., López-Alvarado, J., Farris, E., Susanna, A., Filigheddu, R., and Garcia-Jacas, N. (2014). The role of parental and hybrid species in multiple introgression events: evidence of homoploid hybrid speciation in *Centaurea* (Cardueae, asteraceae). *Bot. J. Linn. Soc* 175, 453–467. doi: 10.1111/boj.12177

Merle, H., Garmendia, A., and Ferriol, M. (2022). Genotyping-by-Sequencing and morphology revealed the role of polyploidization and hybridization in the diversification of the *Centaurea aspera* l. complex of section *Seridia* (Juss.) DC. (Asteraceae). *Plants* 11, 1919. doi: 10.3390/plants11151919

Mráz, P., Garcia-Jacas, N., Gex-Fabry, E., Susanna, A., Barres, L., and Müller-Schärer, H. (2012). Allopolyploid origin of highly invasive *Centaurea stoebe* s.l. (Asteraceae). *Mol. Phylogenet. Evol.* 62, 612–623. doi: 10.1016/j.ympev.2011.11.006

Novaković, J., Janaćković, P., Susanna, A., Lazarević, M., Boršić, I., Milanovici, S., et al. (2022). Molecular insights into the *Centaurea calocephala* complex (Compositae) from the balkans – does phylogeny match systematics? *Diversity* 14, 394. doi: 10.3390/d14050394

Paetzold, C., Wood, K. R., Eaton, D. A., Wagner, W. L., and Appelhans, M. S. (2019). Phylogeny of Hawaiian *Melicope* (Rutaceae): RAD-seq resolves species relationships and reveals ancient introgression. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01074

Pembleton, L. W., Cogan, N. O. I., and 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. doi: 10.1111/1755-0998.12129

Peruzzi, L., and Cesca, G. (2002). "Report: 1302," in *Mediterranean Chromosome number reports*. Eds. G. Kamari, C. Blanché and F. Garbari Fl. Medit. 12, 465. https://www.herbmedit.org/flora/12-443.pdf.

Peruzzi, L. (2008). "Notulae: 1472-1473," in *Notulae alla checklist della flora italiana*, 5. Eds. C. Nepi, L. Peruzzi and A. Scoppola Inform. Bot. Ital., 114–115. https://www.societabotanicaitaliana.it/SBI/IBI%2040%20(1)%202008/097-115%20Notulae%20alla%20checklist%20della%20flora%20vascolare%20italiana%205%20(1420-1474).pdf.

PFI [Portale della Flora d'Italia] (2022). *Centaurea lacaitae \ Centaurea montaltensis \ Centaurea tenorei* (accessed December 5, 2022). Available at: http://dryades.units.it/floritaly.

Pignatti, S. (2018). *Flora d'Italia*. ed. 2 vol. 3 (Bologna: Edagricole).

Pignatti, S., and Lausi, D. (1982). "Centaurea L.," in *Flora d'Italia*, vol. volume 3 . Ed. S. Pignatti (Bologna: Edagricole), 173–209.

Pisanu, S., Mameli, G., Farris, E., Binelli, G., and Filigheddu, R. (2011). A natural homoploid hybrid between *Centaurea horrida* and *Centaurea filiformis* (Asteraceae) as revealed by morphological and genetic traits. *Folia Geobot.* 46, 69–86. doi: 10.1007/s12224-010-9085-2

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

Pritchard, J. K., Wen, X., and Falush, D. (2010) *Documentation for structure software: version 2.3*. Available at: https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/structure_doc.pdf.

Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350

Ranjbar, M., and Negaresh, K. (2014). A revision of *Centaurea* sect. *Centaurea* (Asteraceae) from Iran. *Turk. J. Bot.* 38, 969–987. doi: 10.3906/bot-1304-26

Requena, J., López-Pujol, J., Carnicero, P., Susanna, A., and Garcia-Jacas, N. (2020). The *Centaurea alba* complex in the Iberian peninsula: gene flow, introgression, and blurred genetic boundaries. *Plant Syst. Evol.* 306, 1–21. doi: 10.1007/s00606-020-01669-5

Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., et al. (2015). The chromosome counts database (CCDB) – a community resource of plant chromosome numbers. *New Phytol.* 206, 19–26. doi: 10.1111/nph.13191

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

Rossi, G., Orsenigo, S., Gargano, D., Montagnani, C., Peruzzi, L., Fenu, G., et al. (2020). *Lista rossa della flora italiana. 2 endemiti e altre specie minacciate* (Roma: Ministero dell'Ambiente e della Tutela del Territorio e del Mare). Available at: http://www.iucn.it/pdf/LISTAROSSAvol-2-FLORAITALIANA.pdf.

Santangelo, A., Del Guacchio, E., Cennamo, P., and Caputo, P. (2017). Reassembling the *Centaurea tenorei* group (Asteraceae) puzzle: typification of the names. *Phytotaxa* 298, 119–133. doi: 10.11646/phytotaxa.298.2.2

Soltis, D. E., Visger, C. J., Blaine Marchant, D., and Soltis, P. S. (2016). Polyploidy: pitfalls and paths to a paradigm. *Am. J. Bot.* 103, 1146–1166. doi: 10.3732/ajb.1500501

Soraggi, S., Rhodes, J., Altinkaya, I., Tarrant, O., Balloux, F., Fisher, M. C., et al. (2022). HMMploidy: inference of ploidy levels from short-read sequencing data. *Peer Community J.* 2, e60. doi: 10.24072/pcjournal.178

Stift, M., Kolář, F., and Meirmans, P. G. (2019). Structure is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity* 123, 429–441. doi: 10.1038/s41437-019-0247-6

Suarez-Santiago, V. N., Salinas, M. J., Garcia-Jacas, N., Soltis, P. S., Soltis, D. E., and Blanca, G. (2007). Reticulate evolution in the *Acrolophus* subgroup (*Centaurea* l., compositae) from the Western Mediterranean: origin and diversification of section *Willkommia* Blanca. *Mol. Phylogenet. Evol.* 43, 156–172. doi: 10.1016/j.ympev.2006.08.006

Telwala, Y., Brook, B. W., Manish, K., and Pandit, M. K. (2013). Climate-induced elevational range shifts and increase in plant species richness in a Himalayan biodiversity epicentre. *PloS One* 8, e57103. doi: 10.1371/journal.pone.0057103

Terracciano, N. (1874). *Terza relazione intorno alle peregrinazioni botaniche fatte nella provincia di Terra di Lavoro per disposizione della deputazione provinciale dal dott. N. Terracciano* (Caserta: Nobile & C.o).

Turland, N. J., Wiersema, J. H., Barrie, F. R., Greuter, W., Hawksworth, D. L., Herendeen, P. S., et al. (2018). International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code). *Regnum Vegetabile* (Glashütten: Koeltz Botanical Books) 159. doi: 10.12705/Code.2018

Weiß, C. L., Pais, M., Cano, L. M., Kamoun, S., and Burbano, H. A. (2018). nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinform.* 19, 122. doi: 10.1186/s12859-018-2128-z

Weir, B. S., and Cockerham, C. C. (1984). Estimating f-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x

Wuyun, T., Amo, H., Xu, J., Ma, T., Uematsu, C., and Katayama, H. (2015). Population structure of and conservation strategies for wild *Pyrus ussuriensis* maxim. in China. *PloS One* 10, e0133686. doi: 10.1371/journal.pone.0133686

Yu, Q., Ling, Y., Xiong, Y., Zhao, W., Xiong, Y., Dong, Z., et al. (2022). RAD-seq as an effective strategy for heterogenous variety identification in plants–a case study in Italian ryegrass (*Lolium multiflorum*). *BMC Plant Biol.* 22, 231. doi: 10.1186/s12870-022-03617-6

Zu, K., Wang, Z., Zhu, X., Lenoir, J., Shrestha, N., Lyu, T., et al. (2021). Upward shift and elevational range contractions of subtropical mountain plants in response to climate change. *Sci. Total Environ.* 783, 146896. doi: 10.1016/j.scitotenv.2021.146896