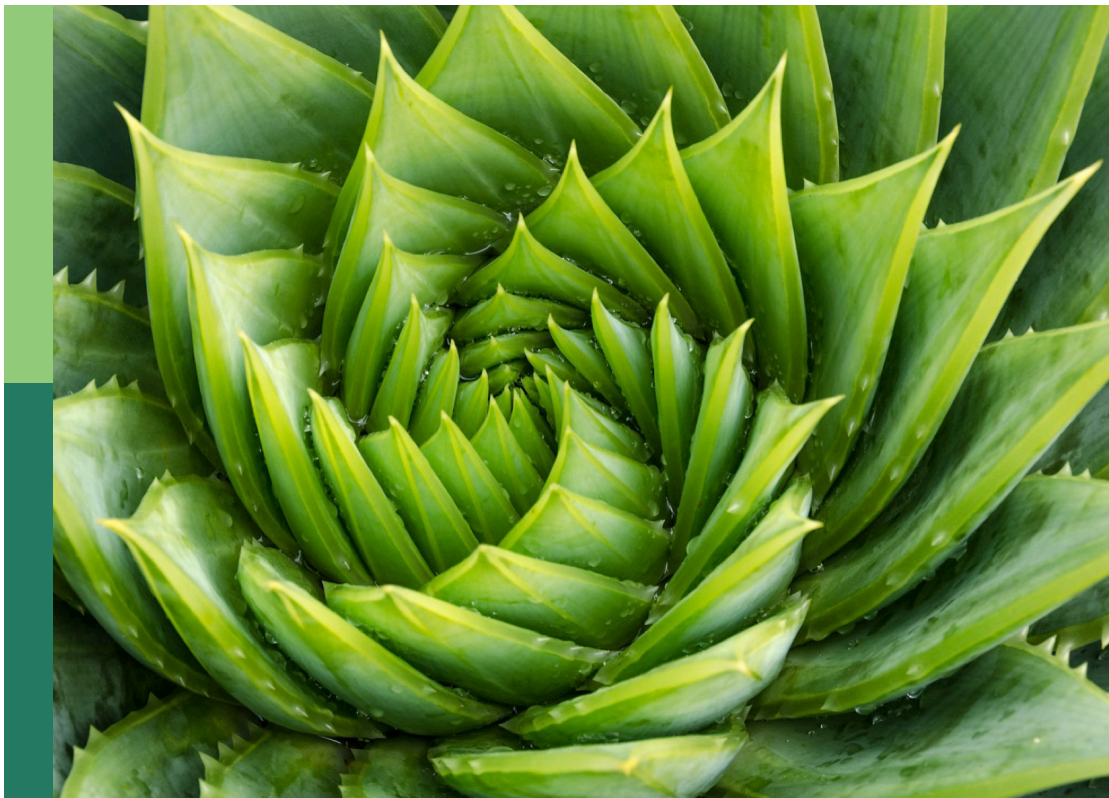
frontiers | Research Topics

Understanding plant diversity and evolution in the mediterranean basin

Edited by
Nico Cellinese, Božo Frajman, Andrew A. Crowl and Gonzalo Nieto Feliner

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-2196-0
DOI 10.3389/978-2-8325-2196-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Understanding plant diversity and evolution in the mediterranean basin

Topic editors

Nico Cellinese — University of Florida, United States
Božo Frajman — University of Innsbruck, Austria
Andrew A. Crowl — Duke University, United States
Gonzalo Nieto Feliner — Real Jardín Botánico (RJB, CSIC), Spain

Citation

Cellinese, N., Frajman, B., Crowl, A. A., Feliner, G. N., eds. (2023). *Understanding plant diversity and evolution in the mediterranean basin*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2196-0

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 14 February 2023
DOI 10.3389/fpls.2023.1152340

OPEN ACCESS

EDITED AND REVIEWED BY
Susann Wicke,
Humboldt University of Berlin, Berlin, Germany

*CORRESPONDENCE
Gonzalo Nieto Feliner
nieto@rjb.csic.es

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

RECEIVED 27 January 2023
ACCEPTED 07 February 2023
PUBLISHED 14 February 2023

CITATION
Nieto Feliner G, Cellinese N, Crowl AA and Frajman B (2023) Editorial: Understanding plant diversity and evolution in the Mediterranean Basin.
Front. Plant Sci. 14:1152340.
doi: 10.3389/fpls.2023.1152340

Editorial: Understanding plant diversity and evolution in the Mediterranean Basin

Gonzalo Nieto Feliner[1]*, Nico Cellinese[2], Andrew A. Crowl[3] and Božo Frajman[4]

[1]Department of Biodiversity and Conservation, Real Jardín Botánico (RJB), CSIC, Madrid, Spain, [2]Florida Museum of Natural History, University of Florida, Gainesville, FL, United States, [3]Department of Biology, Duke University, Durham, NC, United States, [4]Department of Botany, University of Innsbruck, Innsbruck, Austria

KEYWORDS

Mediterranean Basin, systematics, evolution, biogeography, phylogeography, phylogenetics, hybridization, polyploidy

Editorial on the Research Topic
Understanding plant diversity and evolution in the Mediterranean Basin

The biota of the Mediterranean Basin (MB), characterized by a high species richness and endemism (Greuter, 1991; Blondel et al., 2010), has drawn the attention of biologists for centuries. While representing only 1.6% of Earth's surface area, this hotspot of biodiversity houses 7–10% of the world's plant diversity, 60% of which is restricted to the region (Myers et al., 2000; Thompson, 2005). Two main factors which have been shown to be important for high diversity in other Mediterranean climate zones–characterized by mild wet winters and warm dry summers– are low rates of extinction and an association with fire (Rundel et al., 2018). However, many additional features and forces shape biodiversity patterns in the MB (e.g., Nieto Feliner, 2014; Onstein et al., 2015).

Abiotic factors, in particular a complex and active geo-climatic history (Krijgsman, 2002; Rosenbaum et al., 2002), have played a crucial role in shaping the MB flora. Three climatic events are considered as fundamental historical landmarks in plant evolution within the region. These are the Messinian Salinity Crisis (MSC) c. 5.9 Ma (Bocquet et al., 1978), the advent of Mediterranean climate c. 3.2 Ma (Suc, 1984) and the climatic oscillations during the Pleistocene c. 2.5 Ma onwards (Hewitt, 2000). These abiotic events have been tested as drivers of diversification in numerous phylogenetic studies (e.g., Fiz-Palacios and Valcárcel, 2013; Crowl et al., 2015). However, in a biota as rich and diverse as that found in the MB, an accurate picture of diversification drivers across taxonomic groups and geographic areas requires consideration of additional factors. These include evolutionary consequential mechanisms such as polyploidy and hybridization (Thompson, 2005; Marques et al., 2018), the importance of annual habit and associated selfing reproduction (Stebbins, 1970), the heterogeneous orography that provided refugial areas in periods of adverse climates during the Pleistocene (Médail and Diadema, 2009), and the structure of the landscape, composed of a mosaic of habitat patches (Blondel et al., 2010). This structure could partly underlie features such as the importance of chasmophytic species (Davis, 1951), the role of ecological speciation (Thompson et al., 2005), in addition to shaping distribution limits of narrow endemics, which are highly represented in the MB (Thompson, 2005).

Another crucial factor that greatly influenced the MB biota across time is human impact. Flanking the fertile crescent, cradle of western civilizations, the impact of humans on the MB over, at least, the last 10,000 years has probably been the most intense on the planet leading to landscape modification, habitat destruction and species range fragmentation (Blondel et al., 2010; Thompson, 2020). In addition, the MB acted as a corridor for the rapid spread of useful wild species such as chestnut (Fineschi et al., 2000) or stone pine (Vendramin et al., 2008) and domesticated species or varieties (Zohary et al., 2012) although not always along an expected straight east-to-west direction, but rather involving the preservation of ancestral local lineages preceding agriculture (Besnard et al., 2018; Baumel et al., 2022).

Despite a considerable body of knowledge on plant evolution in the MB, we still have an insufficient understanding about the various potential drivers of diversification that, over time, have interacted with other numerous intervening factors to produce the rich plant biodiversity we observe. This is no doubt partly due to the complexity of the evolutionary history in the basin (Nieto Feliner, 2014). The present special issue is a reflection of the broad array of research foci and methodological approaches that the scientific community use to tackle the fascinating questions related to plant evolution in the MB.

Eight of the fifteen papers included in this special issue adopt macroevolutionary perspectives attempting to understand diversification patterns, taxonomic circumscriptions, and biogeographic connections of species lineages occurring in the MB and tested under explicit phylogenetic and phylogenomic frameworks. One of the factors generating uncertainty in taxonomy and phylogenetic relationships is polyploidy, which is documented in five of these studies. Cetlová et al. (2021) shows a combination of representative features for the MB in *Alyssum*, ranging from annual lifespan to multiple events of allopolyploidy, but also highlighting a strong human influence shaping its current distribution, and an impact of range shifts caused by sea-level changes, resulting in complex patterns. The same complexity is found in the *Euphorbia nicaeensis* alliance, which, as several other plant groups (e.g., Manafzadeh et al., 2014), colonized the MB from Western Asia and the Irano-Turanian region, and owes its partly unsettled taxonomy —including cryptic diversity— to diversification driven by vicariance in three main European Pleistocene refugia, ecological adaptation and, to a lesser extent, polyploidy (Stojilkovič et al., 2022). Polyploidy often evolves independently in different groups (Segraves et al., 1999). The following two studies identified different polyploidization patterns across the MB over time. Tomasello and Oberprieler (2022) find that elevational range shifts and secondary contact fostered allopolyploidy in Iberian populations of the genus *Leucanthemopsis*, whereas in other parts of its Mediterreanean-Alpine range, only autopolyploidy occurred. In the genus *Centaurium*, with tetraploids mainly distributed in northern temperate areas and hexaploids in southern arid areas, Maguilla et al. (2021) infer through diversification estimates, biogeographic analysis and chromosome evolution, that diploid lineages remained in the area of origin. By contrast, whole genome duplication events —recent and old— could have facilitated colonization and establishment in other areas. Aiming to tackle the complexity derived from polyploidy, Slenker et al. (2021) developed a new phylogenomic approach to identifying merged subgenomes from different progenitors, which allow them to infer an allopolyploid origin for the Greek endemic *Cardamine barbaraeoides*. Naciri et al. (2022) conclude that the chasmophyte habit, important in the MB, evolved independently several times in *Silene* sect. *Italicae*. The remaining two phylogenetic studies find high diversification rates: first, in the Iberian-centered genus *Antirrhinum*, associated with a geographic mode of speciation and multiple acquisition of key taxonomic characters (Otero et al., 2021); and second, in the Mediterranean lineage of *Limonium*, associated with the MSC, the onset of Mediterranean climate, Plio-Pleistocene sea-level fluctuations, and apomixis (Koutroumpa et al., 2021).

In line with the diversity of possible evolutionary outcomes, hybridization can be addressed under different approaches. The contribution by López de Heredia et al. (2021) exemplifies how in two hybridizing species of the best known syngameon —*Quercus* (Cannon and Petit, 2020)— cytonuclear incompatibilities may have shaped the combination of plastid and nuclear genomes across the species ranges. Abdelaziz et al. provide evidence of a stable unimodal hybrid zone in the genus *Erysimum*, one of the possible evolutionary outcomes of hybridization, which however has been insufficiently studied in plants (Abbott, 2017).

One of the unwritten goals of the present special issue was to enrich our knowledge concerning the Balkan Peninsula, a critical area for the evolution of MB biota, as well as for the postglacial recolonization of central Europe (Hewitt, 2011). In addition to Slenker et al. (2021), four other papers focused on this region. Rešetnik et al. (2022) and Torre et al. (2022) unveil the role of the Balkans as a cradle of diversification and a source for dispersal into other areas in *Aurinia saxatilis* and *Ulmus laevis*, respectively. Lazarević et al. (2022) explore the consequences of interploidy hybridization in the areas of contact between two Tertiary relict species of the long-lived genus *Ramonda* using AFLPs and genome size estimation. Using an integrative approach that includes environmental modelling, Faltner et al. (2023) provide clues to the causes behind very narrowly-distributed endemics in the MB and clarify the unforeseen phylogenetic position of one of them. Focusing on *Euphorbia orphanidis*, an endangered species restricted to a few patches of limestone screes in Mt. Parnassos (Greece), this study underlines the role of microrelief in heterogeneous mountain environments. Finally, a phylogeographic study focuses on a narrow endemic coastal species, *Eokochia saxicola*, which is exceptional in the MB for being wind-pollinated, and concludes that, despite small population sizes, this rare species does not face immediate conservation risks (Strumia et al., 2021).

The Mediterranean Basin is undoubtedly a remarkable biogeographic region with a rich but fragile flora. We hope that this collection of studies provides the inspiration and motivation for much needed additional research in this biodiversity hotspot.

Author contributions

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Funding

This study was supported by grant MCIN/ AEI/ FEDER (PID2021-125432NB-I00), from the Spanish Ministry of Science and Innovation, to GNF.

Acknowledgments

We thank all reviewers, who have helped us edit the articles included in this Research Topic.

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

References

Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *J. Syst. Evol.* 55, 238–258. doi: 10.1111/jse.12267

Baumel, A., Nieto Feliner, G., Médail, F., La Malfa, S., Di Guardo, M., Bou Dagher Kharrat, M., et al. (2022). Genome-wide footprints in the carob tree (*Ceratonia siliqua*) unveil a new domestication pattern of a fruit tree in the Mediterranean. *Mol. Ecol.* 31, 4095–4111. doi: 10.1111/mec.16563

Besnard, G., Terral, J. F., and Cornille, A. (2018). On the origins and domestication of the olive: A review and perspectives. *Ann. Bot.-London* 121, 385–403. doi: 10.1093/aob/mcx145

Blondel, J., Aronson, J., Bodiou, J. Y., and Boeuf, G. (2010). *The Mediterranean region: biological diversity in space and time* (New York: Oxford University Press).

Bocquet, G., Widler, B., and Kiefer, H. (1978). The messinian model, a new outlook for the floristics and systematics of the Mediterranean area. *Candollea* 33, 269–287.

Cannon, C. H., and Petit, R. J. (2020). The oak syngameon: more than the sum of its parts. *New Phytol.* 226, 978–983. doi: 10.1111/nph.16091

Crowl, A. A., Visger, C. J., Mansion, G., Hand, R., Wu, H.-H., Kamari, G., et al. (2015). Evolution and biogeography of the endemic *Roucela* complex (Campanulaceae: *Campanula*) in the Eastern Mediterranean. *Ecol. Evol.* 5, 5329–5343. doi: 10.1002/ece3.1791

Davis, P. H. (1951). Cliff vegetation in the eastern Mediterranean. *J. Ecol.* 39, 63–93. doi: 10.2307/2256628

Fineschi, S., Taurchini, D., Villani, F., and Vendramin, G. G. (2000). Chloroplast DNA polymorphism reveals little geographical structure in *Castanea sativa* mill. (Fagaceae) throughout southern European countries. *Mol. Ecol.* 9, 1495–1503. doi: 10.1046/j.1365-294x.2000.01029.x

Fiz-Palacios, O., and Valcárcel, V. (2013). From messinian crisis to Mediterranean climate: a temporal gap of diversification recovered from multiple plant phylogenies. *Perspect. Plant Ecol.* 15, 130–137. doi: 10.1016/j.ppees.2013.02.002

Greuter, W. (1991). Botanical diversity, endemism, rarity, and extinction in the Mediterranean area: an analysis based on the published volumes of med-checklist. *Bot. Chron.* 10, 63–79.

Hewitt, G. M. (2000). The genetic legacy of the quaternary ice ages. *Nature* 405, 907–913. doi: 10.1038/35016000

Hewitt, G. M. (2011). "Mediterranean Peninsulas: the evolution of hotspots," in *Biodiversity hotspots*. Eds. F. E. Zachos and J. C. Habel (Berlin, Heidelberg: Springer), 123–147.

Krijgsman, W. (2002). The Mediterranean: Mare nostrum of earth sciences. *Earth Planet. Sc. Lett.* 205, 1–12. doi: 10.1016/S0012-821X(02)01008-7

Manafzadeh, S., Salvo, G., and Conti, E. (2014). A tale of migrations from east to west: The irano-turanian floristic region as a source of Mediterranean xerophytes. *J. Biogeogr.* 41, 366–379. doi: 10.1111/jbi.12185

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

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

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

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

Onstein, R. E., Carter, R. J., Xing, Y., Richardson, J. E., and Linder, H. P. (2015). Do Mediterranean-type ecosystems have a common history?–insights from the buckthorn family (Rhamnaceae). *Evolution* 69, 756–771. doi: 10.1111/evo.12605

Rosenbaum, G., Lister, G. S., and Duboz, C. (2002). Reconstruction of the tectonic evolution of the western Mediterranean since the oligocene. *J. Virtual Explor.* 8, 107–130. doi: 10.3809/jvirtex.2002.00053

Rundel, P. W., Arroyo, M. T., Cowling, R. M., Keeley, J. E., Lamont, B. B., Pausas, J. G., et al. (2018). Fire and plant diversification in Mediterranean-climate regions. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00851

Segraves, K. A., Thompson, J. N., Soltis, P. S., and Soltis, D. E. (1999). Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia*. *Mol. Ecol.* 8, 253–262. doi: 10.1046/j.1365-294X.1999.00562.x

Stebbins, G. L. (1970). Adaptive radiation of reproductive characteristics in angiosperms, i. pollination mechanisms. *Annu. Rev. Ecol. Syst.* 1, 307–326. doi: 10.1046/j.1365-294X.1999.00562.x

Suc, J. P. (1984). Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature* 307 (5950), 429–432. doi: 10.1038/307429a0

Thompson, J. D. (2005). *Plant evolution in the Mediterranean* (Oxford: Oxford University Press).

Thompson, J. D. (2020). *Plant evolution in the Mediterranean: insights for conservation* (USA: Oxford University Press).

Thompson, J. D., Lavergne, S., Affre, L., Gaudeul, M., and Debussche, M. (2005). Ecological differentiation of Mediterranean endemic plants. *Taxon* 54, 967–976. doi: 10.2307/25065481

Vendramin, G. G., Fady, B., González-Martínez, S. C., Hu, F. S., Scotti, I., Sebastiani, F., et al. (2008). Genetically depauperate but widespread: the case of an emblematic Mediterranean pine. *Evolution* 62, 680–688. doi: 10.1111/j.1558-5646.2007.00294.x

Zohary, D., Hopf, M., and Weiss, E. (2012). *Domestication of plants in the old world: The origin and spread of domesticated plants in southwest Asia, Europe, and the Mediterranean basin* (Oxford: Oxford University Press).

frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 September 2020
doi: 10.3389/fpls.2020.564414

ddRAD Sequencing-Based Identification of Genomic Boundaries and Permeability in *Quercus ilex* and *Q. suber* Hybrids

Unai López de Heredia[1], Fernando Mora-Márquez[1], Pablo G. Goicoechea[2], Laura Guillardín-Calvo[1†], Marco C. Simeone[3] and Álvaro Soto[1]*

[1] G.I. Genética, Fisiología e Historia Forestal, Dpto. Sistemas y Recursos Naturales, ETSI Montes, Forestal y del Medio Natural, Universidad Politécnica de Madrid, Madrid, Spain, [2] Department of Forestry, NEIKER-BRA, Vitoria-Gasteiz, Spain, [3] Dipartimento di Scienze Agrarie e Forestali (DAFNE), Università degli Studi della Tuscia, Viterbo, Italy

Hybridization and its relevance is a hot topic in ecology and evolutionary biology. Interspecific gene flow may play a key role in species adaptation to environmental change, as well as in the survival of endangered populations. Despite the fact that hybridization is quite common in plants, many hybridizing species, such as *Quercus* spp., maintain their integrity, while precise determination of genomic boundaries between species remains elusive. Novel high throughput sequencing techniques have opened up new perspectives in the comparative analysis of genomes and in the study of historical and current interspecific gene flow. In this work, we applied ddRADseq technique and developed an *ad hoc* bioinformatics pipeline for the study of ongoing hybridization between two relevant Mediterranean oaks, *Q. ilex* and *Q. suber*. We adopted a local scale approach, analyzing adult hybrids (*sensu lato*) identified in a mixed stand and their open-pollinated progenies. We have identified up to 9,251 markers across the genome and have estimated individual introgression levels in adults and seedlings. Estimated contribution of *Q. suber* to the genome is higher, on average, in hybrid progenies than in hybrid adults, suggesting preferential backcrossing with this parental species, maybe followed by selection during juvenile stages against individuals with higher *Q. suber* genomic contribution. Most discriminating markers seem to be scattered throughout the genome, suggesting that a large number of small genomic regions underlie boundaries between these species. In adult hybrids 273 markers (3%) showed allelic frequencies very similar to one of the parental species, and very different from the other; these loci could be relevant for understanding the hybridization process and the occurrence of adaptive introgression. Candidate marker databases developed in this study constitute a valuable resource to design large scale re-sequencing experiments in Mediterranean sclerophyllous oak species and could provide insight into species boundaries and adaptive introgression between *Q. suber* and *Q. ilex*.

Keywords: ddRADseq, *Quercus*, hybridization, introgression, genomic boundaries, SNPs

OPEN ACCESS

Edited by:
Gonzalo Nieto Feliner, Real Jardín Botánico (RJB, CSIC), Spain

Reviewed by:
John McVay, Florida Department of Agriculture and Consumer Services, United States
Rowan Schley, Royal Botanic Gardens, Kew, United Kingdom

****Correspondence:***
Álvaro Soto
alvaro.soto.deviana@upm.es

†Present address:
Laura Guillardín-Calvo, Department of Plant Sciences, University of Oxford, Oxford, United Kingdom

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

Received: *21 May 2020*
Accepted: *13 August 2020*
Published: *04 September 2020*

Citation:
López de Heredia U, Mora-Márquez F, Goicoechea PG, Guillardín-Calvo L, Simeone MC and Soto Á (2020) ddRAD Sequencing-Based Identification of Genomic Boundaries and Permeability in Quercus ilex and Q. suber Hybrids. Front. Plant Sci. 11:564414. doi: 10.3389/fpls.2020.564414

INTRODUCTION

The relevance of hybridization as an adaptive and evolutionary driving force is a controversial topic. Even the different concepts of species are tightly linked to the relevance authors concede to gene exchange among species (reviewed in Harrison and Larson, 2014). The Biological Species Concept (Mayr, 1982) relies mainly on reproductive isolation among species, while other perspectives admit the "porous" nature of genomes and stress ecological differentiation instead (f. i., Mallet, 2007). Actually, hybridization and introgression are common in plants, where interspecific gene flow is supposed to be an ongoing process in approximately 25% of species (Mallet et al., 2016). Hybridization is especially frequent in certain taxa, as is the case of *Quercus* species, which have been proposed as models for the Ecological Species Concept (Burger, 1975; van Valen, 1976), and are commonly regarded as "syngameons" (Cannon and Petit, 2020; Kremer and Hipp, 2020). Temperate oaks from Europe and North America have become references for hybridization studies and they have shown that hybridization might be crucial to understand their successful evolutionary trajectories (f.i., Petit et al., 2003; Lepais et al., 2009; Eaton et al., 2015; Goicoechea et al., 2015; Ortego et al., 2018; Crowl et al., 2019).

Hybridization can allow one species to incorporate genes or alleles from another one. Some of these transferred genes can confer a selective advantage to the first species, improving its performance or increasing its ecological range (Sexton et al., 2009). This process is known as adaptive introgression, recently reviewed for plants by Suárez-González et al. (2018). Human-mediated hybridization has been largely used in agriculture, and analysis of spontaneous introgression to search for useful adaptations for crop breeding has been recently claimed, for instance, by Burgarella et al. (2019). Additionally, hybridization has probably played a key role in plant species demography. Threats to small populations can be intensified by lower individual fitness (Allee effect), including reproductive success, maybe related to inbreeding depression. Recurrent hybridization can increase effective population size, overrunning this effect. Moreover, hybridization can also act as a colonization mechanism over other species range (Potts and Reid, 1988; Petit et al., 2003).

Notwithstanding this ample capacity of interspecific gene flow, *Quercus* species still maintain their integrity (f.i., Valbuena-Carabaña et al., 2005; Moran et al., 2012). Cannon and Scher (2017) presented a well-constructed hypothesis accounting for this apparent paradox. The authors focused on the behavior of male hybrid gametes in the diploid stigma of a pure individual of one of the parental species. In such a situation, it is argued that hybrid gametes most similar to those of that parental species have advantage over the other hybrid gametes. Thus, resulting backcrosses would be quite similar to parental species, bearing, at maximum, a low proportion of introgressed alleles. This would be the reason why species boundaries do not get blurred so often even when there is hybridization. However, such hypothesis does not consider the opposite situation, with hybrids acting as mother trees. In this case, in order to parallel Cannon and Scher hypothesis, it would not be male gametes (assumed to come from parental species), but the female gametes and/or the almost pure zygotes which should show better fitness on a hybrid ovary.

Quercus ilex (holm oak) and *Q. suber* (cork oak) constitute a very suitable experimental system for hybridization analysis. These two Mediterranean species are easily distinguishable due to their phenotype, mainly according to the characteristic corky bark of *Q. suber*. Hybridization between both species is well known since ancient times, but it is considered to be rather infrequent. Therefore, their introgression should be more easily traceable than, for instance, the more frequent hybridization and introgression between white oaks, which often hampers correct identification of individuals (Muir et al., 2000; Valbuena-Carabaña et al., 2005). Hybridization is supposed to have played a key role in the survival of *Q. suber* in Southern France and Eastern Spain, in a general landscape of calcareous soils, not suitable for *Q. suber* (Lumaret et al., 2002; Jiménez et al., 2004; López de Heredia et al., 2007). In 2009, current hybridization was estimated in less than 2%, in a study using eight nSSR and including natural pure and mixed stands, as well as provenance trials covering the whole distribution range of *Q. suber* (Burgarella et al., 2009). Nevertheless, recent studies have shown that classification/identification power of the set of nSSR used is too low, and thus, hybridization and introgression could have been underestimated (López de Heredia et al., 2018a; Soto et al., 2018). Therefore, more powerful marker sets are needed in order to obtain accurate estimations.

In recent years, rapid advancement and affordability of next generation sequencing (NGS) methodologies have largely increased the scope and accuracy of genome-wide analysis and its application in phylogenetics, adaptation, hybridization, and breeding studies. For oak species, significant efforts have been made in the white oak complex, producing valuable genomic resources: ESTs, QTLs, SNP panels, or linkage maps, among others (Saintagne et al., 2004; Ueno et al., 2010; Lepoittevin et al., 2015; Lesur et al., 2015; Bodénès et al., 2016; Lesur et al., 2018; Lang et al., 2020). The ongoing progress in NGS technologies has resulted in a first draft genome assembly of *Q. suber* (Ramos et al., 2018), and in a more complete assembly for *Q. robur* (Plomion et al., 2018). The latter is the first oak genome assembly that contains information at the chromosome level, thus easing comparative genomic analysis in less studied oak species (see for instance Goicoechea et al., 2015).

The unavailability of reference genome assemblies constrains the scope of genomic studies in non-model species, as is the case for the vast majority of woody plant species (Falk et al., 2018). In this context, the use of reduced representation genotyping methodologies, such as double-digested RAD sequencing (ddRADseq; Peterson et al., 2012), largely facilitates genomic analysis (Hirsch et al., 2014), and has been already used in the phylogenetic analysis (Hipp et al., 2020) and genetic mapping of oak species (Konar et al., 2017). The use of these methodologies, together with the aforementioned NGS-based genomic resources and specifically designed bioinformatic pipelines, opens up new perspectives in oak species genomics.

Here we report the application of ddRADseq to the study of introgression and ongoing hybridization between *Q. ilex* and *Q. suber*, using adult hybrids identified in a natural population and

their progenies, since natural hybrids can be quite useful for mapping differences and reproductive barriers in non-model organisms (Baack and Rieseberg, 2007). The specific aims of the study were: (1) to identify candidate markers for the study of ongoing hybridization and adaptive introgression and (2) to provide insight into the reproductive behavior of hybrids as mother trees.

MATERIALS AND METHODS

Plant Material and DNA Extraction

The sampling area is located in Fregenal de la Sierra (Badajoz province, southwestern Spain). Twenty-two adult hybrid individuals were identified according to morphological characters, especially those regarding bark, described by Natividade (1936). Leaves from these hybrids, along with leaves from 98 *Q. suber* and 99 *Q. ilex* adult individuals were collected and stored at −80°C for further DNA extraction.

In addition, in November 2016, we collected acorns from nine of these hybrid trees, eight of them located in the same estate, in a mixed *Q. ilex*-*Q. suber* stand. A total of 1,270 acorns (492 from hybrids, 397 from four cork oaks, and 381 from four holm oaks, presumably unrelated) were collected directly from the canopies, so as to be sure of the identity of mother trees. Acorns were first sown on perlite and seedlings were transplanted to 3 l. containers, in peat:perlite (3:1), and grown in nursery conditions. Hybrid acorns showed lower germination rate, so that the final sampling size for DNA extraction and genotyping consisted of 302 hybrids (22 adults and 280 seedlings), 467 cork oaks (98 adults and 369 seedlings), and 439 holm oaks (99 adults and 340 seedlings) (**Table 1**).

Leaf discs (6 mm diameter) from adults and seedlings (N = 1,208 samples) were used for DNA extraction on a Genespin™ platform, and for further library preparation and sequencing (LGC Genomics GmbH, Germany).

Identification of Hybrids' Plastid Lineages

The trnH-psbA intergenic spacer of chloroplast DNA was amplified and sequenced following Simeone et al. (2016) in the four *Q. suber* and the four *Q. ilex* from which acorns were sampled, the 22 adult hybrids and 18 of the hybrid progeny individuals. Four additional adult hybrid individuals from Central Spain (López de Heredia et al., 2018a) were also analyzed. Electropherograms were edited and eye-checked with Chromas 2.6.2 (http://technelysium.com.au). The obtained dataset was enriched with all GenBank accessions showing 100% identity in a nucleotide BLAST search on NCBI (https://www.ncbi.nlm.nih.gov/), and with the trnH-psbA haplotypes representing all the known evolutionary lineages of the West Eurasian *Quercus* sections *Ilex* and *Cerris* (Simeone et al., 2016; Vitelli et al., 2017; Simeone et al., 2018). A multiple alignment was built with MEGA7 (Kumar et al., 2016) and a haplotype list was computed with DnaSP 5.1 (Librado and Rozas, 2009). A median-joining (MJ) haplotype network of the entire dataset was inferred with Network 4.6.1.1 (http://www.fluxus-engineering.com/). The MJ algorithm was invoked with default parameters (equal weight of transversion/transition), treating gaps as 5th state.

TABLE 1 | Number of seedlings in open-pollinated families, identified by the name of each mother tree.

	No. of seedlings		No. of seedlings
Quercus ilex		*Hybrids*	
E28	82	FS08	12
E31	84	FS14	19
E41	81	FS16	47
E96	93	FS17	26
Quercus suber		FS18	16
A05	71	FS19	57
A07	94	FS20	30
A09	101	FS21	5
A10	103	FS22	68

ddRADseq Library Construction and Sequencing

For ddRADseq analysis of all samples, a pilot study was conducted using the protocol implemented on RADdesigner (Guillardín-Calvo et al., 2019) to assess the optimal enzyme combination, type of library, and number of reads per sample for the aims of the study. Based on this pilot study, PstI/MspI were selected as restriction enzymes. The use of methylation-sensitive restriction enzyme combinations such as PstI/MspI increases the fraction of coding DNA among the sequenced fragments, which is of particular interest in complex plant genomes (Pootakham et al., 2015). ddRADseq libraries were constructed and sequenced by LGC Genomics GmbH (Germany), including a barcode to identify the fragments corresponding to each sample. Libraries were prepared in 14 96-well PCR plates, with 96 inline-barcoded adaptors, and 10 ul of each library were amplified using MyTaq (Bioline GmbH, Germany) and TrueSeq primers (Illumina, USA) in 20 ul PCR reactions. Five ul of each amplified library from the same PCR plate were pooled and cleaned up using Agencourt AMPure XP bead purifications (Beckman Coulter, USA) and MinElute columns (QIAGEN, Netherlands). Normalization was done using Evrogen Trimmer kit (Evrogen, Russia) and reamplified in 100 ul using MyTaq (Bioline GmbH, Germany). Two hundred to 400 bp fragments were selected on Blue Pippin and LMP-agarose gel, in order to produce fewer, yet sufficient, fragments at higher depth. Sequencing was performed on an Illumina NextSeq 500 V2 platform (Illumina Inc., USA) and several runs were performed in order to obtain ~1.5 M 150 bp single-ended reads per sample.

Bioinformatic Analysis

All the scripts to perform the bioinformatic analysis detailed below (**Figure 1**) are stored in the ddRAD-CORK-HYB package (https://github.com/GGFHF/ddRAD-CORK-HYB). A README file provides instructions to install software and dependencies, and includes information about each script parameter and how to run scripts properly.

Read Pre-Processing and Filtering

Read preprocessing consisted in a demultiplexing step of all library groups using the Illumina BclFastQC 2.17.1.14 software (http://support.illumina.com/downloads/bcl2fastqconversion-software-v217.html).

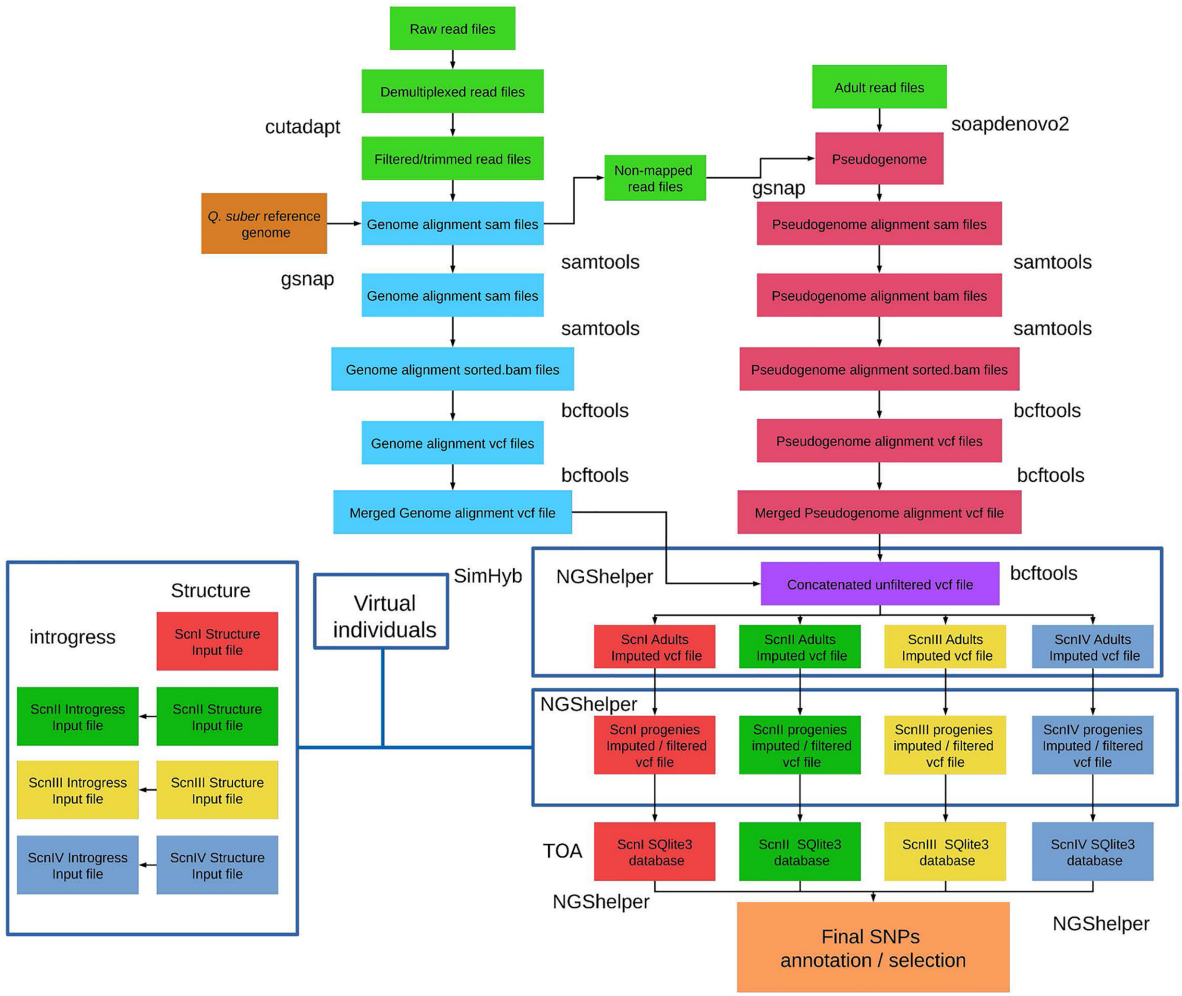


FIGURE 1 | Bioinformatics strategy for ddRADseq data analysis. BASH/R scripts and instructions to install software and dependencies are available by downloading the ddRAD-CORK-HYB package at https://github.com/GGFHF/ddRAD-CORK-HYB.

A maximum of two mismatches or Ns in the barcode read for demultiplexing between libraries was set when the barcode distances between all libraries on the lane allowed for it. Demultiplexing of library groups into samples was performed according to their inline barcodes. No mismatches or Ns were allowed.

Filtering was performed with CUTADAPT (Martin, 2011), and sequencing adapter remnants were clipped from all reads. Reads with final length <20 bp and reads with 5′ ends not matching the restriction enzyme site were discarded. Finally, reads were filtered by quality, removing reads containing Ns and trimming those reads at the 3′-end to get a minimum average Phred quality score of 20 over a ten-base window. Again, reads with a final length <20 bp were discarded. Quality of pre-processed read files was checked with FASTQC (Andrews, 2010).

Read Alignment and Variant Calling

GSNAP (Wu et al., 2016) was used to align pre-processed and filtered reads to the genome assembly of *Q. suber* (Ramos et al., 2018). Reads that did not map to the *Q. suber* genome assembly were aligned back with GSNAP to a pseudogenome built with SOAPDENOVO2 genomic assembler (Luo et al., 2012) using the reads from the adult individuals, and filtered by size to keep only contigs >100 bp.

Genome and pseudogenome alignments in *sam* format were converted into compressed *bam* files and sorted with SAMTOOLS (Li et al., 2009). Variant calling was performed individually with SAMTOOLS *mpileup* and BCFTOOLS (Li, 2011) *call* options. Variant calling files of each sample from genome and pseudogenome alignments were concatenated with BCFTOOLS to obtain individual *vcf* files. Finally, all samples were merged with BCFTOOLS *merge* to obtain a single concatenated-merged unfiltered variant calling file for all the samples (adults, progenies, and replicates).

Determination of Marker Positions in *Q. robur* Linkage Groups

Multi-fasta sequences of the flanking regions 200 bp to each side of the selected loci positions were extracted from the *Q. suber*

reference genome assembly and from the pseudogenome using the *get-flanking-regions.py* utility from NGSHELPER (https://github.com/GGFHF/NGShelper). The resulting multifasta file was subjected to a nucleotide BLAST search against *Q. robur* v. PM1N genome assembly with the local version of BLAST+ (Camacho et al., 2009) using an e-value threshold of 1E-6, to obtain the genomic positions in the *Q. robur* genome of the identified markers.

Further, we built a linear function to correlate the genomic coordinates in bp of the SNPs used to build linkage groups in *Q. robur* (Bodénès et al., 2016) with the positions of those SNPs in 12 *Q. robur* linkage groups in cM. All this information was freely available at the oak genome sequencing web (http://www.oakgenome.fr) and at the *Quercus* portal (https://arachne.pierroton.inra.fr/cgi-bin/cmap/). Thus, we were able to identify the putative location of our marker loci in the linkage groups of *Q. robur*, and to visualize them with the R package LINKAGEMAPVIEW (Ouellette et al., 2017).

Marker Databases

SQLITE3 databases for each imputation scenario (see below) were constructed with NGSHELPER including information parsed from *vcf* files (type of marker, genomic coordinates, etc.) and functional annotation of selected markers. Functional annotations of the markers identified in the genome were collected from *Q. suber* genome *gff* files and descriptions of gene IDs was obtained from NCBI 'All_Plants.gene.info' file. The markers corresponding to the pseudogenome were annotated with TOA (Mora-Márquez et al., 2020), through nucleotide BLAST searches against NCBI NT databases.

Two-Step Imputation/Filtering Procedure: Scenarios

Missing data must be considered for *vcf* files processing steps. There is no standard missing data cutoff for excluding loci, and filtering criteria need to be established depending on the aims of the study (Huang and Knowles, 2016). Excessive data filtration can have unforeseen consequences, such as the truncation of loci with higher mutation rates or missing relevant loci. Besides, filtering protocols must ensure sufficient depth coverage of the remaining loci (Eaton et al., 2017). A common practice is to test for different scenarios encompassing more restrictive and more relaxed filtering criteria.

The procedures to perform filtering and imputation to account for missing data and/or null alleles are implemented in NGSHELPER. Input data consist of the single concatenated-merged unfiltered variant calling file generated in the previous step. The basic procedure included two steps. Firstly, the proportion of missing data in adult individuals of each parental species was considered. For loci with a low proportion of missing data in one of the parental species (md < 0.05) but with a high proportion in the other (md > 0.90), a null allele was imputed to missing data of the adult individuals of the latter species. Successive filtrations of imputed loci were performed, resulting in different imputation scenarios (see *Results* section). In a second step, genotypes of the seedlings were corrected according to their mother tree imputation state.

Estimation of Introgression Levels

Estimation of the introgression level of adult individuals and progenies was performed using the Bayesian approach implemented in STRUCTURE v. 2.3.4 (based on Pritchard et al., 2000), as well as by means of the hybrid index provided by INTROGRESS (Gompert and Buerkle, 2010). Firstly, we checked the performance and accuracy of both classification methods using virtual individuals with known introgression levels. These individuals were simulated with SIMHYB (Soto et al., 2018; https://github.com/GGFHF/SimHyb), based on the allele frequencies of the adult *Q. ilex* and *Q. suber* populations. Later on, we estimated contribution of parental species to the genome of each individual in two scenarios of prevalence of hybrids: 1 and 10%. In both cases, additional virtual *Q. ilex* and *Q. suber* individuals were included as reference. Number of real and virtual individuals used, according to their introgression level, are provided in Supplementary Information (**Supplementary Table 1**). In all the cases, we allowed STRUCTURE to estimate admixture proportions in the unknown samples. For all jobs, we used 10,000 burn-in steps followed by 100,000 repetitions of the MCMC chains, and the number of populations was set to K = 2, following the methodology described by Evanno et al. (2005).

RESULTS

Adult Hybrid Identification and Germination Rates

Twenty-two adult individuals, identified as *Q. ilex*–*Q. suber* hybrids according to their phenotypic characters, were localized in the field, in different estates in the south of Badajoz (SW Spain). Acorns were collected from the canopy of these hybrid trees, as well as from unrelated *Q. ilex* and *Q. suber* adults (four of each species). While parent species showed high germination rates (89.24% for *Q. ilex* and 92.95% for *Q. suber*), only 56.92% of hybrid acorns germinated successfully.

Chloroplast DNA analysis of these trees revealed that the four cork oak adult individuals gathered in a single haplotype (H12), corresponding to a widespread, likely ancestral haplotype, shared by multiple *Cerris* section species extending from Anatolia to the central Mediterranean. The four holm oak adult samples grouped with all hybrid adults and progenies in the 'Euro-Med' haplotype group (sub-lineage V) (Simeone et al., 2018), i.e. an exclusive lineage of the central and western Mediterranean members of section *Ilex* (*Q. ilex* and *Q. coccifera*) (**Figure 2** and **Supplementary List 1**). This result confirms the directionality in *Q. ilex*–*Q. suber* hybridization, in which holm oak acts as mother tree, and therefore hybrids carry holm oak chloroplasts. The trnH-psbA haplotype sequences generated in this study are available on GenBank under accession numbers LR797857-LR797862.

Read Alignment, Variant Filtering, and Imputation

A mean number of 1.74 million single-ended reads per sample was obtained after removal of adapters and read filtering by quality. Sequences are available at NCBI SRA database (BioProject: PRJNA628590). From these, 64.9% of all filtered reads mapped

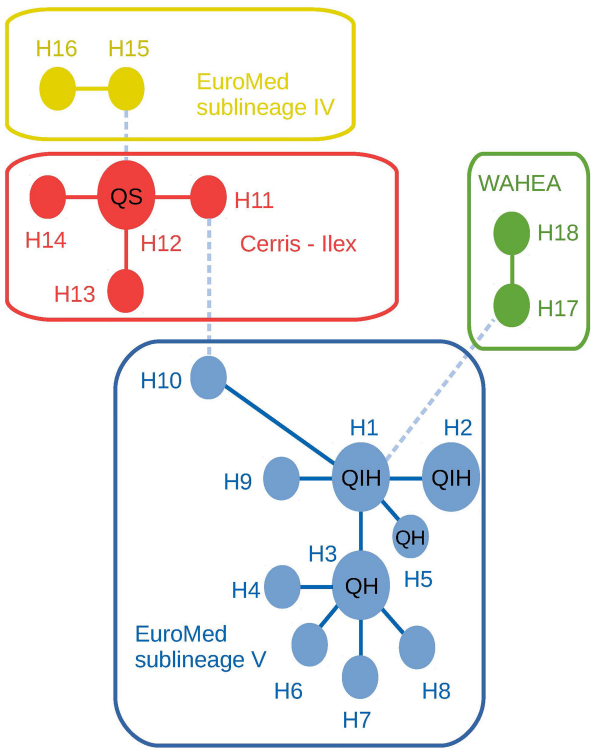


FIGURE 2 | Median-joining haplotype network of trnH-psbA marker combining the investigated samples and published haplotypes of each major *Ilex* and *Cerris* lineages. Lineage and sub-lineage identity according to Vitelli et al., 2017 and Simeone et al., 2018. WAHEA: haplotypes from the West Asian, Himalayan, and East Asian lineages. Line types correspond to the number of mutations separating each haplotype: solid lines = 1–3 mutations; dashed lines 6–10 mutations. Coding within haplotype circles indicate the presence of samples of the present work: QS, Adult *Q. suber* individuals; QIH, Adult *Q. ilex* and hybrid individuals; QH, only hybrid individuals.

uniquely to the *Q. suber* genome assembly (**Figure 3**). However, we found noticeable differences in the total number of aligned reads among progenies and adult individuals of each species. As expected, *Q. suber* libraries aligned better than *Q. ilex*, while the hybrids showed intermediate behavior. Reads showing ambiguous mapping were discarded, and reads that did not map to the *Q. suber* genome assembly were aligned to the pseudogenome, improving total alignment percentages to 67.8% of all filtered reads. The best alignments to the pseudogenome were obtained for those *Q. ilex* libraries with the poorest mapping to the *Q. suber* genome.

After individual variant calling, the number of variants ranged between 14,593 and 524,458 for the genome and between 212 and 66,680 for the pseudogenome alignments. The final concatenated-merged variant calling file had 16,234,798 variants, of which 97.4 % were SNPs and <5% multi-allelic sites. This massive *vcf* file was subjected to filtering and imputation according to four scenarios.

As a basic filtering, we removed loci with low coverage (DP, total number of reads across all samples, <6000). On average, ~70 reads per sample were obtained for each remaining locus. Additionally, for those loci with a low proportion of missing data in one of the parental species (md < 0.05) but with a very high proportion in the other parental species (md > 0.90), a null allele was assigned to such missing data and to one of the alleles in homozygous individuals in the latter species; this was considered the scenario I (ScnI). These loci (hereinafter referred to as imputed loci) can correspond to alterations in one or both restriction sites (including methylation), yielding no scorable fragments in the sequencing phase, or to larger indels in the genome of one of the species; in any case, they can be highly informative. Nevertheless, in order to avoid an excessive impact of imputed loci, we conceived a second scenario (ScnII), in which we kept only one imputed locus per gene or intergenic fragment of the *Q. suber* genome (Ramos et al., 2018). The location of a locus within an intergenic fragment was assessed considering a proximity criterion using a 10 kb sliding window. This way, ScnII kept approximately 80% of the loci imputed under ScnI. In ScnIII we removed all the imputed alleles, and considered them as missing data. Finally, in the most restrictive scenario, we discarded all the loci with a proportion of missing data >90% in either of the adult *Q. suber* or *Q. ilex* populations (ScnIV). Later on, we also corrected the genotypes of the progenies according to each imputation scenario, and considered as missing data those genotypes incompatible with that of the corresponding mother tree.

The number of final recovered loci varied depending on the scenario (**Figure 4A**). For ScnI and ScnIII we obtained up to 9,251 loci, with 2,026 (21.9%) imputed ones in ScnI. Under ScnII we considered 8,901 loci, with 18.8% of them imputed. The more restrictive ScnIV kept 7,225 unimputed loci. The annotation to the *Q. suber* genome allowed precise location of many of these loci in specific genes of known function (Ramos et al., 2018) and intergenic regions (**Figure 4B**). Loci fromScnI, ScnII and ScnIII were located in 3,396 fragments, of which 2,566 were genic and 811 intergenic. For Scn IV the number of fragments dropped to 1,829, of which 1,540 (84.2%) corresponded to genic regions and 279 (15.3%) to intergenic ones. In all the scenarios, loci corresponding to genic regions were mostly exonic c. 72%), although a significant percentage of loci (c. 28%) occurred in introns (**Figure 4B**). The remaining loci (0.3%) were located in 29 fragments that could not be mapped to the Q. suber genome assembly. Databases of markers are available at Zenodo Repository (López de Heredia et al., 2020).

Distribution of Markers Across the Genome

In order to evaluate their distribution across the genome, and provided there is no information regarding chromosomal distribution of *Q. suber* nor *Q. ilex* genomes, we checked the homology of these loci on the *Q. robur* genome, for which 12 linkage groups (corresponding to the 12 chromosomes) have been identified (Bodénès et al., 2012; Bodénès et al., 2016). A total of 8,774 loci were successfully mapped against *Q. robur* genome; of these, 8,004 showed homology with loci included

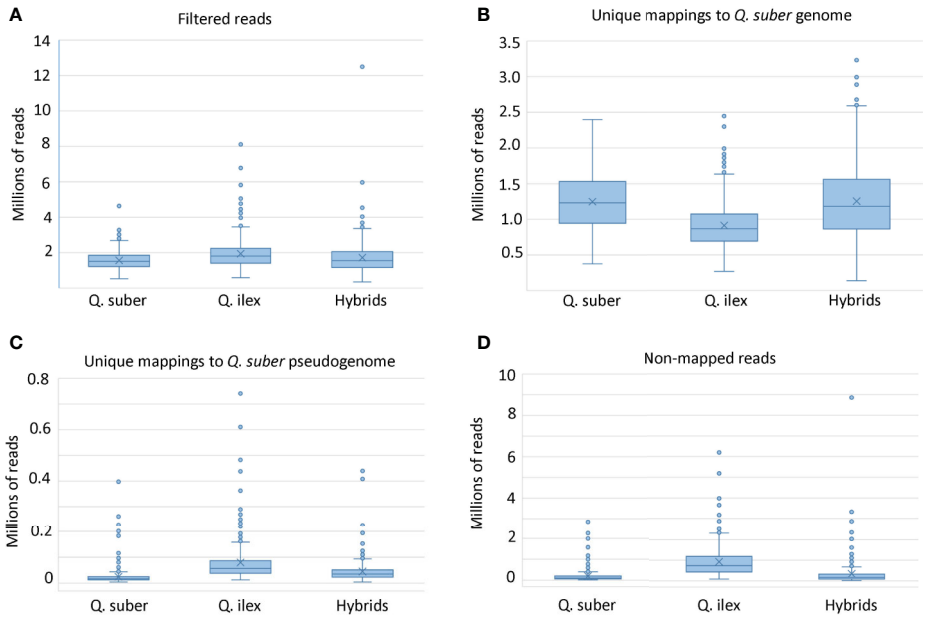


FIGURE 3 | Box and whisker plots of mapping sequences. **(A)** Total number of reads employed in the ddRADseq pipeline after adapter removal and quality filtering. **(B)** Number of reads uniquely mapped to *Q. suber* genome assembly (Ramos et al., 2018). **(C)** Number of reads uniquely mapped to the pseudogenome generated with SoapDeNovo2. **(D)** Final number of non-mapped reads.

in the 12 linkage groups. These loci belong to 2,932 genomic fragments: 2,264 genic and 668 intergenic. We found a rather even distribution of these loci among the 12 linkage groups, with an average distribution of almost 670 loci per linkage group, approximately 10.55 loci/Mb (**Figure 5**). Assuming a high synteny among species of the genus (Bodénès et al., 2012; Konar et al., 2017), these results suggest an unbiased distribution of our loci across *Q. suber* and *Q. ilex* genomes.

Introgression Levels

The Bayesian approach implemented in STRUCTURE v. 2.3.4 (Pritchard et al., 2000), as well as the hybrid index provided by the INTROGRESS R-package (Gompert and Buerkle, 2010), were then used to classify individuals according to their introgression levels. Firstly, and following the procedure established by Burgarella et al. (2009), accuracy of the classification was evaluated using virtual individuals of known pedigree created with SIMHYB (Soto et al., 2018). ScnIII yielded the same results as ScnIV, due to the distribution of missing data among species and the way both programs consider them. Therefore, ScnIII was discarded for further analysis of real individuals (**Figure 6**).

Once accuracy and reliability of classification tools had been checked, introgression levels of real hybrid adults were assessed. Estimation was performed considering 1% and 10% of hybrids in the analyzed population. INTROGRESS and STRUCTURE yielded similar results in each situation, and very small differences were detected between both hybrid prevalence situations. Under the four imputation scenarios, estimations for adult individuals were roughly compatible with F1 hybrids (except for FS-01, which could be classified as a backcross with *Q. suber*) (**Figure 7**). As for the hybrid progenies, a large variability of introgression levels was detected, although we scored, on average, a higher contribution of *Q. suber* to the genome of the seedlings for all the hybrid families, compared to their mother trees (**Table 2**).

Following López de Heredia et al. (2018a), we compared STRUCTURE estimates of the *Q. suber* contribution to the genome of each hybrid seedling to that of its mother tree ($q_{soffspring}/q_{smother}$). Values higher than 1 point to likely fertilizations by *Q. suber*, while values below 1 suggest fertilization by *Q. ilex*. This comparison suggests that hybrid mother trees can be fertilized by both *Q. ilex* and Q. *suber* pollen, although more frequently by the latter species, at least in the studied season (**Table 3**).

DISCUSSION

Recent advances in NGS methodologies along with the release of reference genome assemblies have opened up new perspectives in the study of open pollinated plant hybridizing systems, such as the *Quercus* syngameon (Kremer and Hipp, 2020). In this work, we have used ddRADseq methodology and have developed *ad hoc* bioinformatic pipelines to identify genomic variants that may shed light in the study of ongoing hybridization and genomic boundaries between *Q. suber* and *Q. ilex*. This insight may complement the current view derived from other presumably more introgressed *Quercus* complexes, such as the intensively studied European white oaks.

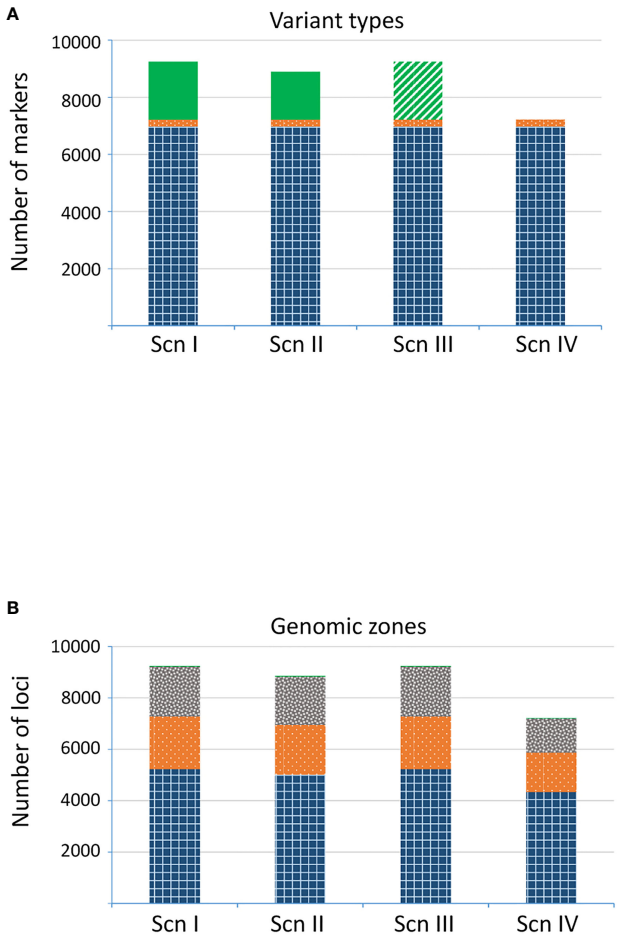


FIGURE 4 | Number of loci detected for each scenario. **(A)** Variant types: SNP (gridded blue), indel (dotted orange), and imputed loci (plain green; treated as missing data in ScnIII). **(B)** Genomic region: known exonic region (gridded blue), known intronic region (dotted orange), intergenic region (grey confetti), not assigned region (plain green).

Most hybridization studies in woody plant species have been conducted to assess ancient introgression from a phylogenetic perspective (McVay et al., 2017) or ongoing introgression by sampling individuals exploring ample hybrid zones (Lepais et al., 2009; Zeng et al., 2011). In this work, we have adopted a local scale approach, aiming to block the impact of environmental or IBD related variation in candidate loci inference. The sampling scheme of adults and progenies of both parental species and hybrids used here has revealed powerful to identify candidate gene markers for the study of ongoing hybridization in Mediterranean sclerophyllous oaks.

Candidate Marker Loci Identification

Obtaining candidate loci in reduced genome representation studies (ddRADseq), particularly if they are aimed to perform individual classification of hybrids, requires careful experimental design to ensure sufficient coverage of the genome of the focal species and rigorous filtering of variants in order to minimize the potential sources of error in marker identification (O'Leary et al., 2018). In the present work, an experimental design derived from a pilot study (Guillardín-Calvo et al., 2019), together with an *ad hoc* bioinformatic pipeline has enabled the identification of a significant number of genomic markers to analyze ongoing hybridization between *Q. ilex* and *Q. suber*.

We employed an enzyme combination (PstI/MspI) that prioritizes the representation of loci within or in the proximity of genic regions that could be relevant for the study of introgression patterns, enriching for hypomethylated gene space and reducing the number of fragments from highly repetitive genomic regions (Emberton et al., 2005; Nelson et al., 2008; Poland and Rife, 2012). Moreover, a combination of methylation-sensitive rare- and common-cutting enzymes, such as PstI/MspI, has proven to show greater uniformity of read depth across loci providing high quality genotype information in plants (Pootakham et al., 2016). We have also optimized ddRADseq protocols by performing a size selection step to ensure the collection of robust polymorphic loci at sufficient depth. Actually, genome mapping and variant calling using *Q. suber* genome assembly as a reference have confirmed that most candidate polymorphic markers (*c.* 80%) correspond to genic regions, more than 55% of loci are located in exons and *c.* 22% in introns. Approximately 20% of loci were located in intergenic regions, and, comparatively few candidate loci (0.3%) were obtained from the pseudogenome mapping. However, we cannot discard that some of these regions could contain genes or regulatory regions in our focal species and further re-sequencing experimental work will be necessary to confirm this point.

To gain robustness in candidate marker inference, we have applied a filtering-imputation procedure to yield four different scenarios. Using restrictive filteringcriteria (ScnIV), we have obtained 7,225markers that correspond to 1,540 genic fragments of known function, 279 intergenic fragments and 10 fragments that could not be assigned to *Q. suber* genome assembly. Nevertheless, strict application of these criteria could have discarded informative loci, for instance, those present in the genome of one of the parental species but absent in the other. To avoid missing this information, we designed an imputation procedure, giving rise to what we have called scenarios I and II. This way we identified up to 2,026 additional loci, with imputed null alleles, under ScnI. These loci, which could be highly informative for introgression studies, belonged to 1,264 genic fragments of known function, 584 intergenic fragments and 9 fragments that could not be assigned to *Q. suber* genome assembly. It is noteworthy that many of these null alleles were imputed to *Q. ilex*.

The markers identified in this work provide a well-distributed coverage of the whole genome, assuming synteny among genomes of the genus (Bodénès et al., 2012; Konar et al., 2017), since their orthologs appear distributed throughout the 12 linkage groups described for *Q. robur* (Plomion et al., 2018) (**Figure 5**). Furthermore, we have investigated the distribution of the most discriminating genomic regions between *Q. suber* and

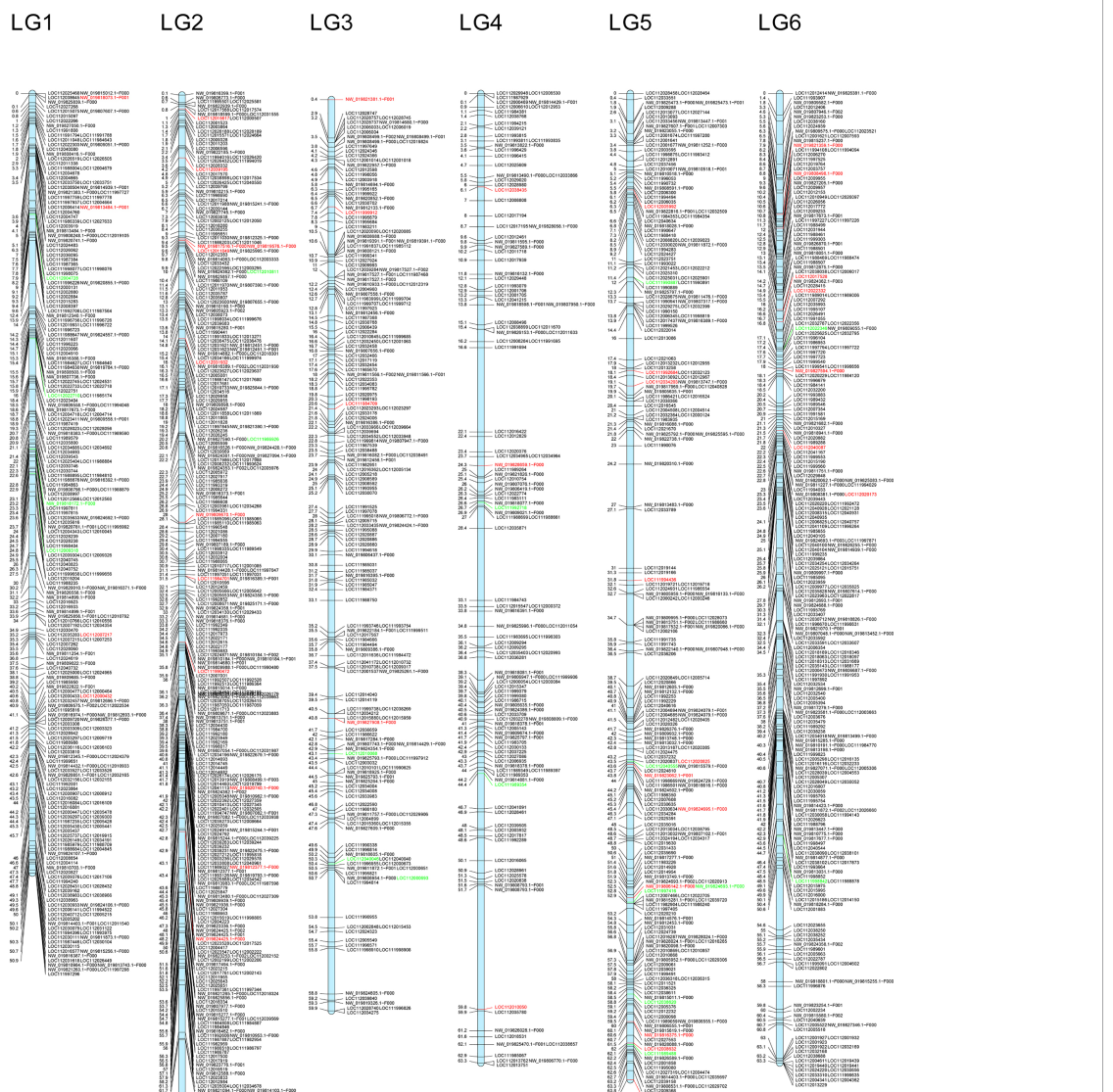


FIGURE 5 | Continued

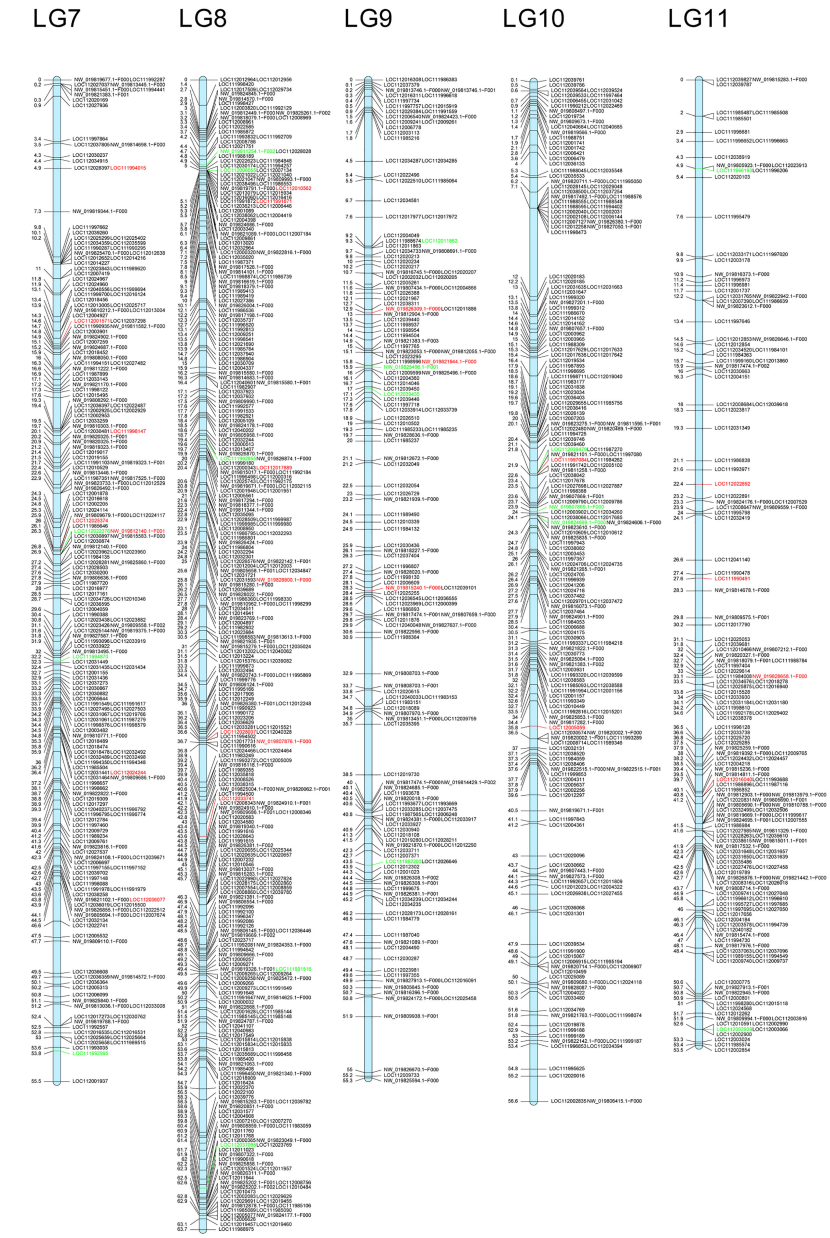


FIGURE 5 | Putative location in the *Q. robur* linkage groups of the genomic fragments including markers from ScnI. Location of markers with allelic frequencies in the adult hybrids very similar to one of the parental species is highlighted in red (*Q. ilex*) or in green (*Q. suber*).

Q. ilex. Thus, in addition to the 2,026 imputed loci, up to 2,830 non-imputed ones show very different patterns in both species, with frequencies of the most common allele 0.9 in one of the species and 0.2 in the other one. Regarding the hybrids, under ScnI 190 loci (167 imputed) show allelic frequencies in the hybrids quite similar to those of *Q. ilex* and very different from *Q. suber*, while 83 loci (17 imputed) show frequencies in the hybrids very similar to *Q. suber* and different from *Q. ilex*. These loci, presumably linked to introgression directionality and genomic barriers between *Q. suber* and *Q. ilex*, are scattered throughout the 12 linkage groups (**Figure 5**). This feature seems to be consistent with previous results for *Q. robur* and *Q. petraea* based on QTL (Saintagne et al., 2004), suggesting that species boundaries are underpinned by a large number of small genomic regions, rather than on few large blocks.

Individual Introgression Levels

Most adult hybrids could be classified as F1 hybrids under all the imputation scenarios considered. Inclusion of imputed loci does not entail a significant difference in the estimation of the contribution of parental species to the genome of hybrid individuals. For the adult hybrids, only slightly lower values of

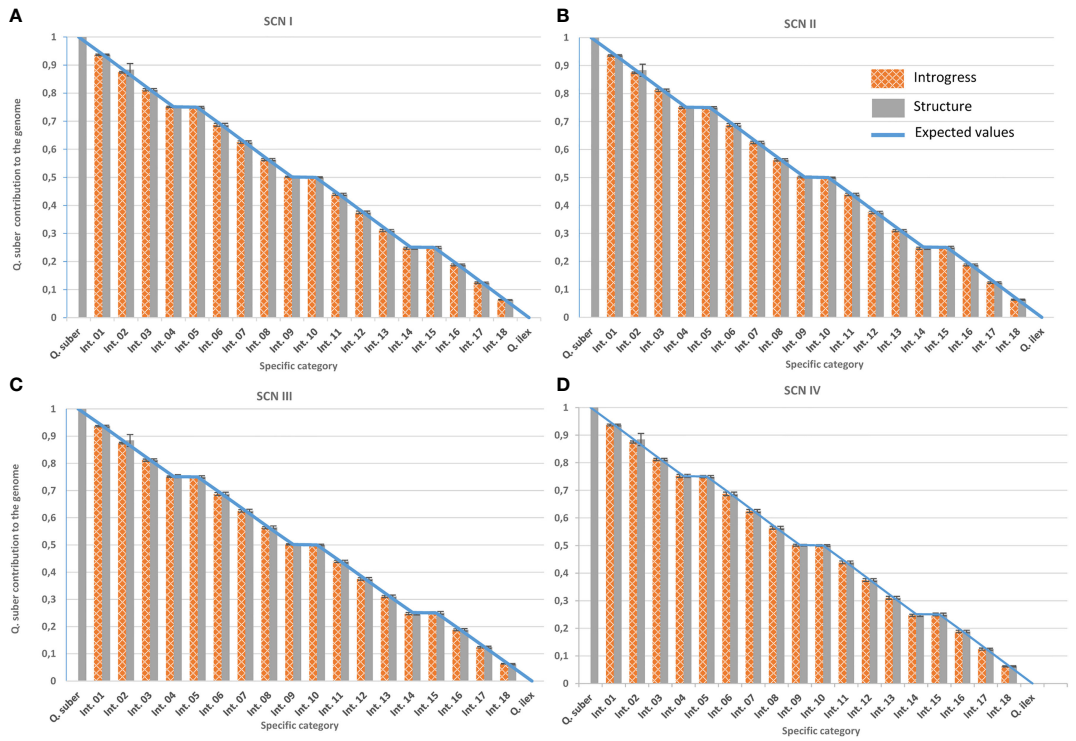


FIGURE 6 | Performance of classification tools under the different imputation scenarios. Genomic contribution of *Q. suber* is estimated using Structure's qs (plain grey) and Introgress hybrid index (gridded orange) on virtual individuals generated with SimHyb (pure species and 18 intermediate categories). Expected values are represented by a blue line. Standard deviations are also indicated. **(A)** Imputation scenario I; **(B)** imputation scenario II; **(C)** imputation scenario III; **(D)** imputation scenario IV.

Q. suber contribution are obtained under ScnI and ScnII. Different results are observed for the hybrid progenies. Individuals with higher estimated *Q. suber* contributions under ScnIV show lower values when imputed loci are considered, while the opposite is observed for individuals with lower estimations. Since most of the null alleles are imputed to *Q. ilex*, this latter result must be due to a higher proportion of non-imputed, "suber" alleles in heterozygosity in these loci in these individuals. Taking into account PstI/MspI sensitivity to methylation, hybridization-mediated alteration of epigenetic characters could also contribute to these results. This way, methylated epialleles in the restriction sites, which would yield no scorable reads and, therefore, would have been imputed with a null allele, could have turned out to be unmethylated and therefore scorable in hybrids, or vice-versa. This could be the case at least of the 184 markers for which very high frequencies of the imputed allele are recorded in adult hybrids (0.75), no matter their global classification as F1 hybrids. Such an alteration of epigenetic characters has been reported, for instance, in *Spartina* (Salmon et al., 2005) or in *Solanum* (Marfil et al., 2006). On the contrary, Radosavljevic et al. (2019) obtained similar estimations of introgression levels using markers sensitive or insensitive to methylation in *Salvia officinalis* and *S. fruticosa* hybrids. Further research will be needed to disentangle the effects of *Q. ilex* and *Q. suber* hybridization in epigenetic patterns.

On the other hand, estimated contribution of *Q. suber* increases in the genome of hybrid offspring compared with their mother trees, for all the families and in all the scenarios considered. These results suggest a preferential fertilization of hybrid trees by *Q. suber* pollen. Different factors, including phenology can account for this feature. In fact, earlier flowering of *Q. ilex* (Gómez-Casero et al., 2007) favors that initial hybridization takes place with *Q. ilex* acting as mother tree; additionally, pre-zygotic incompatibilities for *Q. ilex* pollen in *Q. suber* pistil have also been reported (Boavida et al., 2001). Consistently, all the 22 adult hybrid individuals included in this study carry *ilex* chloroplasts. Actually, *ilex* chloroplasts have been found in all hybrid trees analyzed to date (Lumaret et al., 2002; Burgarella et al., 2009; López de Heredia et al., 2018a). Thus, hybrid mother trees could be preferentially fertilized by *Q. suber*. The only study performed to date, to our knowledge, reports a similar reproductive phenology for hybrid trees and *Q. suber* (Perea Garcia-Calvo, 2006), so that backcrossing with this species would be more likely. Additionally, pre- and post-zygotic incompatibilities or a better performance of *Q. suber* pollen tubes in the hybrid pistil could also account for such a preferential fertilization. In spite of this, López de Heredia et al. (2018a) reported that hybrids can be effectively pollinated both by *Q. suber* and by *Q. ilex* and that backcrossing directionality is largely driven by pollen availability. Consistently, and

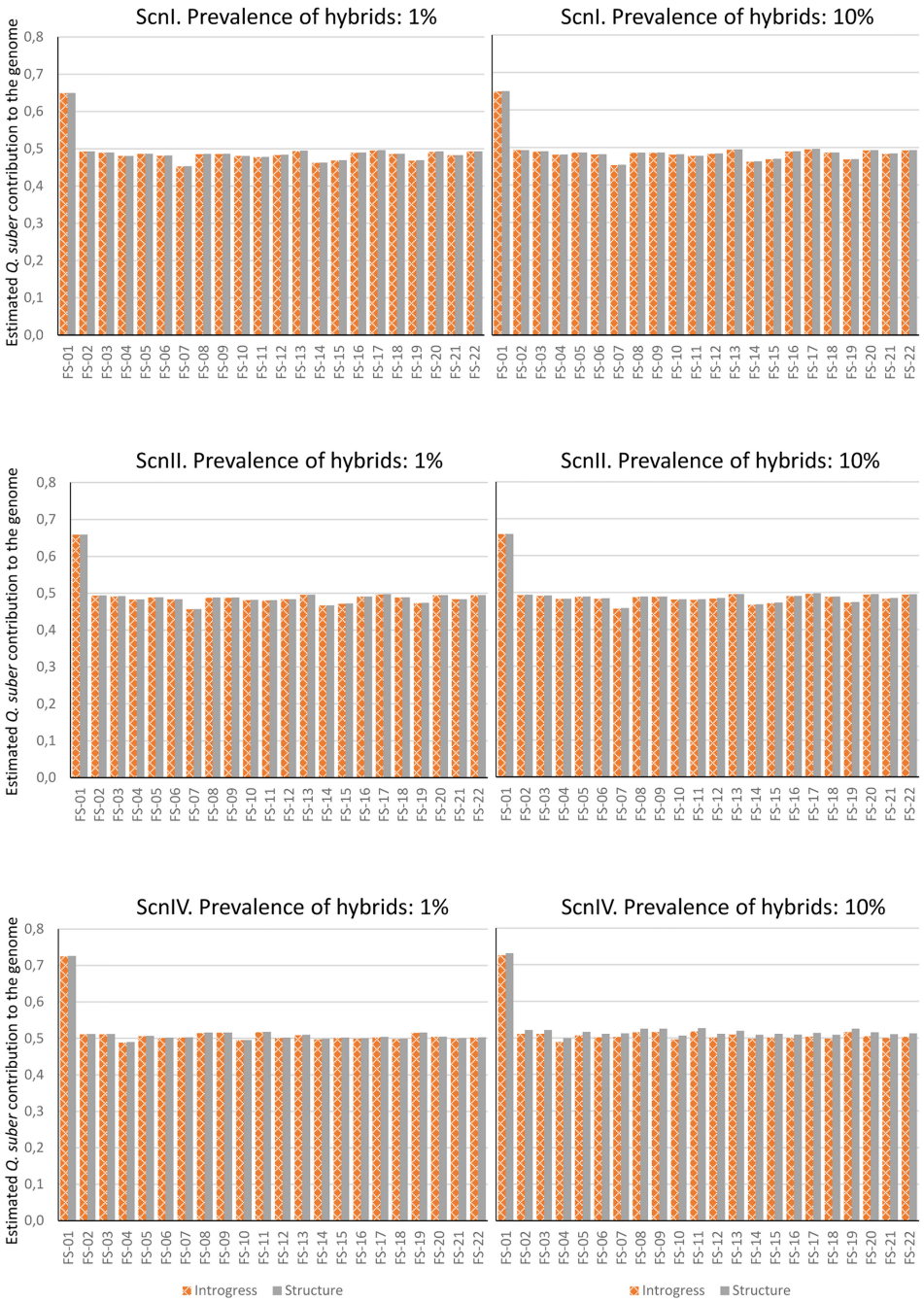


FIGURE 7 | Contribution of Q. suber to the genome of adult hybrids under each scenario, estimated by means of STRUCTURE's qs and Hybrid Index of INTROGRESS, with 1% and 10% of hybrids.

notwithstanding the preferential fertilization by *Q. suber*, in the present work we also report some hybrid seedlings presumably derived from *Q. ilex* pollen in most families. This result is also in accordance with the detection of putatively introgressed holm oaks reported by Burgarella et al. (2009). Moreover, our analyses yield intermediate contribution of parental species to most hybrid seedlings (according to both STRUCTURE and INTROGRESS). This feature in hybrid mother trees contrasts with Cannon and Scher (2017) hypothesis for hybrid pollen and should not be disregarded. Actually, it may be more likely that hybrid trees scattered in stands of parental species reproduce acting as mother trees, pollinated by parental species,

TABLE 2 | STRUCTURE's q_s and INTROGRESS hybrid index estimates for the progenies of the open-pollinated hybrid families under each scenario.

Mother tree	Scn I			Scn II			Scn IV		
	q_s	Max	min	q_s	Max	min	q_s	Max	min
FS08	0.633 (0.049)	0.714	0.543	0.644 (0.051)	0.726	0.548	0.731 (0.067)	0.821	0.606
FS14	0.648 (0.096)	0.776	0.322	0.658 (0.100)	0.791	0.319	0.731 (0.125)	0.904	0.316
FS16	0.626 (0.058)	0.713	0.317	0.634 (0.061)	0.726	0.313	0.700 (0.075)	0.818	0.311
FS17	0.655 (0.073)	0.758	0.361	0.665 (0.077)	0.773	0.358	0.738 (0.096)	0.878	0.365
FS18	0.662 (0.035)	0.718	0.596	0.672 (0.037)	0.731	0.602	0.748 (0.048)	0.824	0.659
FS19	0.587 (0.152)	0.769	0.155	0.595 (0.157)	0.784	0.149	0.657 (0.190)	0.893	0.114
FS20	0.663 (0.058)	0.773	0.521	0.674 (0.061)	0.789	0.524	0.752 (0.079)	0.901	0.561
FS21	0.625 (0.013)	0.640	0.607	0.633 (0.014)	0.648	0.614	0.701 (0.019)	0.722	0.677
FS22	0.633 (0.056)	0.721	0.359	0.643 (0.058)	0.736	0.373	0.718 (0.068)	0.835	0.487

Mean, standard deviation (in brackets), maximum and minimum values per family are provided.

TABLE 3 | Ratio between STRUCTURE's q_s of the offspring and q_s of their mothers under each scenario.

Mother tree	Scn I			Scn II			Scn IV		
	$q_s o/q_s m$	Max	Min	$q_s o/q_s m$	Max	Min	$q_s o/q_s m$	Max	Min
FS08	1.303	1.469	1.117	1.316	1.485	1.121	1.392	1.564	1.154
FS14	1.400	1.676	0.695	1.405	1.690	0.682	1.438	1.780	0.622
FS16	1.277	1.455	0.647	1.290	1.479	0.637	1.375	1.607	0.611
FS17	1.321	1.528	0.728	1.335	1.552	0.719	1.439	1.712	0.712
FS18	1.359	1.474	1.224	1.374	1.495	1.231	1.472	1.622	1.297
FS19	1.252	1.640	0.330	1.254	1.654	0.314	1.252	1.701	0.217
FS20	1.346	1.568	1.057	1.358	1.591	1.056	1.459	1.750	1.089
FS21	1.290	1.322	1.254	1.304	1.336	1.266	1.375	1.416	1.327
FS22	1.284	1.462	0.728	1.298	1.487	0.754	1.403	1.631	0.951

Mean, maximum and minimum values per hybrid family are provided. Mean for the open-pollinated hybrid families by scenario. Values >1 point to likely fertilizations by Q. suber, while values <1 suggest fertilization by Q. ilex (López de Heredia et al., 2018a).

than as pollen donors. In that case, there is no reason to suppose any advantage for pure parental gametes in a diploid, hybrid stigma or for "pure" or "almost pure" zygotes and seed development in the hybrid sporophyte (mother tree).

The larger contribution from *Q. suber* to hybrid progenies' genomes compared to the adults could be due to greater availability of *Q. suber* pollen in the only campaign analyzed in the present study. However, a selective advantage for individuals more similar to *Q. ilex*, from the seedling to the adult stage cannot be discarded either. Such a selection would not be unexpected. Fertilization of hybrids can give rise to a huge amount of genomic combinations. Large part of them can be incompatible since the first developmental stages, hampering the formation of embryos. Later on, other combinations can also be deleterious, as confirmed by the low germination rate of hybrid acorns in this study (56.92% in hybrids, compared to 89.24% in *Q. suber* and up to 92.95% in *Q. ilex*), consistent with previous studies in other areas (López de Heredia et al., 2018a). Although many morphological anomalies already described in *Q. suber x ilex* hybrid seedlings (López de Heredia et al., 2017; López de Heredia et al., 2018b) were present in part of the seedlings, we have scored low mortality in nursery conditions over three years since seedlings emergence. Nevertheless, selection against certain genomic combinations—late-acting genetic incompatibilities—during the juvenile phases of hybrid individuals with major *Q. suber* ancestry could also account for the higher contribution of *Q. ilex* to the genome of surviving hybrids. In addition, extrinsic selection during pre-adult stages, mitigated under nursery conditions, cannot be discarded.

CONCLUSION AND FUTURE PROSPECTS

Our work reports a case study of hybridization and introgression in two non-model forest tree species, *Q. suber* and *Q. ilex*, using genome-wide NGS techniques, and provides a pipeline and scripts for this kind of studies. Out of the 9,251 marker loci identified in this study, 4,856 are highly discriminant between both species, and 2,026 of these are apparently absent in one of the species (*Q. ilex* in most cases). This can be due to alterations in restriction enzyme target sites or to real indels. Interestingly, for 9.1%of them adult hybrids show patterns quite similar to one of the parental species (8.3% to *Q. ilex*, while only 0.8% to *Q. suber*), suggesting selection of those alleles in backcrosses or hybridization-mediated alterations of the methylation patterns. In any case, these loci deserve further attention, since they could be linked to viability of hybrid individuals and/or to selective advantages.

Future Prospects

Most discriminant loci and their adjacent regions, as well as those putatively involved in selection of hybrids, should be

analyzed both in hybrid individuals and in parental species. These genomic regions are good candidates to be involved in genome permeability, interspecific barriers and, eventually, in adaptive introgression. Further characterization of these loci, as well as of those with different allele frequencies in hybrid adults and seedlings could provide insight in species boundaries and on adaptive introgression between *Q. suber* and *Q. ilex*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, LR797857-LR797862 https://www.ncbi.nlm.nih.gov/, PRJNA628590.

AUTHOR CONTRIBUTIONS

AS conceived the idea and drafted the manuscript. AS and ULH designed the experiments and collected the plant material. LG-C handled the plant material. MS performed the cpDNA analysis. FM-M, PGG, ULH, and AS performed the simulations and bioinformatics analysis. All the authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the projects AGL2015-67495-C2-2-R (Spanish Ministry of Economy and Competitiveness) and PID2019-110330GB-C22 (Spanish Ministry of Science and Innovation).

ACKNOWLEDGMENTS

The authors thank Francisco Martínez Moreno for his help in hybrids localization and sampling in the field.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.564414/full#supplementary-material

REFERENCES

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (Accessed March 15, 2019).

Baack, E. J., and Rieseberg, L. H. (2007). A genomic view of introgression and hybrid speciation. *Curr. Opin. Genet. Dev.* 17, 513–518. doi: 10.1016/j.gde.2007.09.001

Boavida, L. C., Silva, J. P., and Feijó, J. A. (2001). Sexual reproduction in the cork oak (*Quercus suber* L). II. Crossing intra- and interspecific barriers. *Sex Plant Reprod.* 14 (3), 143–152. doi: 10.1007/s004970100100

Bodénès, C., Chancerel, E., Gailing, O., Vendramin, G. G., Bagnoli, F., Durand, J., et al. (2012). Comparative mapping in the Fagaceae and beyond with EST-SSRs. *BMC Plant Biol.* 12:153. doi: 10.1186/1471-2229-12-153

Bodénès, C., Chancerel, E., Ehrenmann, F., Kremer, A., and Plomion, C. (2016). High-density linkage mapping and distribution of segregation distortion regions in the oak genome. *DNA Res.* 23-2, 115–124. doi: 10.1093/dnares/dsw001

Burgarella, C., Lorenzo, Z., Jabbour-Zahab, R., Lumaret, R., Guichoux, E., Petit, R. J., et al. (2009). Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). *Heredity* 102, 442–452. doi: 10.1038/hdy.2009.8

Burgarella, C., Barnaud, A., Kane, N. A., Jankowski, F., Scarcelli, N., Billot, C., et al. (2019). Adaptive introgression: an untapped evolutionary mechanism for crop adaptation. *Front. Plant Sci.* 10:4. doi: 10.3389/fpls.2019.00004

Burger, W. C. (1975). The species concept in *Quercus*. *Taxon* 24, 45–50. doi: 10.2307/1218998

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinform.* 10, 1–9. doi: 10.1186/1471-2105-10-421

Cannon, C. H., and Petit, R. J. (2020). The oak syngameon: more than the sum of its parts. *New Phytol.* 226 (4), 978–983. doi: 10.1111/nph.16091

Cannon, C. H., and Scher, C. L. (2017). Exploring the potential of gametic reconstruction of parental genotypes by F1 hybrids as a bridge for rapid introgression. *Genome* 60 (9), 713–719. doi: 10.1139/gen-2016-0181

Crowl, A. A., Manos, P. A., McVay, J. D., Lemmon, A. R., Moriarty Lemmon, E., and Hipp, A. L. (2019). Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). *New Phytol.* 226 (4), 1158–1170. doi: 10.1111/nph.15842

Eaton, D. A. R., Hipp, A. L., González-Rodríguez, A., and Cavender-Bares, J. (2015). Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution* 69 (10), 2587–2601. doi: 10.1111/evo.12758

Eaton, D. A. R., Spriggs, E. L., Park, B., and Donoghue, M. J. (2017). Misconceptions on Missing Data in RAD-seq Phylogenetics with a Deep-scale Example from Flowering Plants. *Syst. Biol.* 66 (3), 399–412. doi: 10.1093/sysbio/syw092

Emberton, J, Ma, J., Yuan, Y., SanMiguel, P., and Bennetzen, J. L. (2005). Gene enrichment in maize with hypomethylated partial restriction (HMPR) libraries. *Genome Res.* 15, 1441–1446. doi: 10.1101/gr.3362105

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.2005.02335.x

Falk, T., Herndon, N., Grau, E., Buehler, S., Richter, P., Zaman, S., et al. (2018). Growing and cultivating the forest genomics database, TreeGenes. *Database (Oxford)* 2018, 1–11. doi: 10.1093/database/bay084

Goicoechea, P., Herrán, A., Durand, J., Bodènés, C., Plomion, C., and Kremer, A. (2015). A linkage disequilibrium perspective on the genetic mosaic of speciation in two hybridizing Mediterranean white oaks. *Heredity* 114, 373–386. doi: 10.1038/hdy.2014.113

Gómez-Casero, M. T., Galán, C., and Domínguez-Vilches, E. (2007). Flowering phenology of Mediterranean Quercus species in different locations (Córdoba, SW Iberian Peninsula). *Acta Bot. Malac.* 32, 127–146. doi: 10.24310/abm.v32i0.7033

Gompert, Z., and Buerkle, C. A. (2010). *Introgress*: a software package for mapping components of isolation in hybrids. *Mol. Ecol. Resour.* 10, 374–384. doi: 10.1111/j.1755-0998.2009.02733.x

Guillardín-Calvo, L., Mora-Márquez, F., Soto, Á., and López de Heredia, U. (2019). RADdesigner: a workflow to select the optimal sequencing methodology in genotyping experiments on woody plant species. *Tree Genet. Genomes* 15, 64. doi: 10.1007/s11295-019-1372-3

Harrison, R. G., and Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *J. Hered.* 105, 795–809. doi: 10.1093/jhered/esu033

Hipp, A. L., Manos, P. S., Hahn, M., Avishai, M., Bodénès, C., Cavender-Bares, J., et al. (2020). Genomic landscape of the global oak phylogeny. *New Phytol.* 226 (4), 1198–1212. doi: 10.1111/nph.16162

Hirsch, C. D., Evans, J., Buell, C. R., and Hirsch, C. N. (2014). Reduced representation approaches to interrogate genome diversity in large repetitive plant genomes. *Brief. Funct. Genomics* 13 (4), 257–267. doi: 10.1093/bfgp/elt051

Huang, H., and Knowles, L. L. (2016). Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation study of RAD Sequences. *Syst. Biol.* 65, 357–365. doi: 10.1093/sysbio/syu046

Jiménez, M. P., López de Heredia, U., Collada, C., Lorenzo, Z., and Gil, L. (2004). High variability of chloroplast DNA in three Mediterranean evergreen oaks indicates complex evolutionary history. *Heredity* 93, 510–515. doi: 10.1038/sj.hdy.6800551

Konar, A., Choudhury, O., Bullis, R., Fiedler, L., Kruser, J. M., Stephens, M. T., et al. (2017). High-quality genetic mapping with ddRADseq in the non-model tree *Quercus rubra*. *BMC Genomics* 18, 417. doi: 10.1186/s12864-017-3765-8

Kremer, A., and Hipp, A. L. (2020). Oaks: an evolutionary success story. *New Phytol.* 226 (4), 987–1011. doi: 10.1111/nph.16274

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi: 10.1093/molbev/msw054

Lang, T., Abadie, P., Leger, V., Decourcelle, T., Frigerio, J. M., Burban, C., et al. (2020). High-quality SNPs from genic regions highlight introgression patterns among European white oaks (*Quercus petraea* and *Q. robur*). *bioRxiv*. doi: 10.1101/388447

Lepais, O., Petit, R., Guichoux, E., Lavabre, J., Alberto, F., Kremer, A., et al. (2009) Species relative abundance and direction of introgression in oaks. *Mol. Ecol.* 18, 2228–2242. doi: 10.1111/j.1365-294X.2009.04137.x

Lepoittevin, C., Bodénès, E., Chancerel, E., Villate, L., Lang, T., Lesur, I., et al. (2015). Single-nucleotide polymorphism discovery and validation in high-density SNP array for genetic analysis in European white oaks. *Mol. Ecol. Resour.* 15 (6), 1446–1459. doi: 10.1111/1755-0998.12407

Lesur, I., Le Provost, G., Bento, P., Da Silva, C., Leplé, J. C., Murat, F., et al. (2015). The oak gene expression atlas: insights into Fagaceae genome evolution and the discovery of genes regulated during bud dormancy release. *BMC Genomics* 16, 112. doi: 10.1186/s12864-015-1331-9

Lesur, I., Alexandre, H., Boury, C., Chancerel, E., Plomion, C., and Kremer, A. (2018). Development of target sequence capture and estimation of genomic relatedness in a mixed oak stand. *Front. Plant Sci.* 9, 996. doi: 10.3389/fpls.2018.00996

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27 (21), 2987–2993. doi: 10.1093/bioinformatics/btr509

Librado, P., and Rozas, J. (2009). DnaSP v5: A Software for Comprehensive Analysis of DNA Polymorphism Data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

López de Heredia, U., Jiménez, P., Collada, C., Simeone, M. C., Bellarosa, R., Schirone, B., et al. (2007). Multi-Marker Phylogeny of Three Evergreen Oaks Reveals Vicariant Patterns in the Western Mediterranean. *Taxon*. 56 (4), 1209–1220. doi: 10.2307/25065912

López de Heredia, U., Vázquez, F. M., and Soto, Á. (2017). "The role of hybridization on the adaptive potential of Mediterranean sclerophyllous oaks: the case of the Quercus ilex x Q . suber complex," in *Oaks Physiological Ecology. Exploring the Functional Diversity of Genus*. Eds. L. Quercus, E. Gil-Pelegrín, J. J. Peguero-Pina and D. Sancho-Knapik (Cham, Switzerland: Springer International Publishing), 239–260.

López de Heredia, U., Sánchez, H., and Soto, Á. (2018a). Molecular evidence of bidirectional introgression between *Quercus suber* and *Quercus ilex*. *iForest* 11, 338–343. doi: 10.3832/ifor2570-011

López de Heredia, U., Duro-García, M. J., and Soto, Á. (2018b). Leaf morphology of progenies in *Q. suber, Q. ilex*, and their hybrids using multivariate and geometric morphometric analysis. *iForest* 11, 90–98. doi: 10.3832/ifor2577-010

López de Heredia, U., Mora-Márquez, F., Goicoechea, P. G., Guillardín-Calvo, L., Simeone, M. C., and Soto, Á. (2020). SQLite3 databases of candidate marker loci from ddRADseq in Quercus suber, Quercus ilex and their hybrids. *Zenodo Repos.* doi: 10.5281/zenodo.3786309

Lumaret, R., Mir, C., Michaud, H., and Raynal, V. (2002). Phylogeographical variation of chloroplast DNA in holm oak (*Quercus ilex* L.). *Mol. Ecol.* 11, 2327–2336. doi: 10.1046/j.1365-294X.2002.01611.x

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1 (1):18. doi: 10.1186/2047-217X-1-18

Mallet, J., Besansky, N., and Hahn, M. W. (2016). How reticulated are species? *Bioessays* 38, 140–149. doi: 10.1002/bies.201500149

Mallet, J. (2007). Hybrid speciation. *Nature* 446, 279–283. doi: 10.1038/nature05706

Marfil, C. F., Masuelli, R. W., Davison, J., and Comai, L. (2006). Genomic instability in *Solanum tuberosum* x *Solanum kurtzianum* interspecific hybrids. *Genome* 49, 104–113. doi: 10.1139/g05-088

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17 (1), 10–12. doi: 10.14806/ej.17.1.200

Mayr, E. (1982). *The growth of biological thought: diversity, evolution, and inheritance* (Cambridge (MA), USA: Belknap Press).

McVay, J. D., Hipp, A. L., and Manos, P. S. (2017). A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proc. Biol. Sci.* 284 (1854), 20170300. doi: 10.1098/rspb.2017.0300

Mora-Márquez, F., Chano, V., Vázquez-Poletti, J., and López de Heredia, U. (2020). TOA: a software package for automated functional annotation in non-model plant species. *Authorea*. doi: 10.22541/au.159611047.70067764

Moran, E. V., Willis, J., and Clark, J. S. (2012). Genetic evidence for hybridization in red oaks (Quercus sect. Lobatae, Fagaceae). *MAm. J. Bot.* 99 (1), 92–100. doi: 10.3732/ajb.1100023

Muir, G., Flemming, C. C., and Schlötterer, C. (2000). Species status of hybridizing oaks. *Nature* 405 (6790), 1016. doi: 10.1038/35016640

Natividade, J. V. (1936). Estudo histológico das peridermes do híbrido *Quercus ilex x suber*, P. *Cout. Publ. Dir. G. Serv. Flor. III* 1, 32 p. Lisbon, Portugal.

Nelson, W., Luo, M., Ma, J., Estep, M., Estill, J., He, R., et al. (2008). Methylation-sensitive linking libraries enhance gene-enriched sequencing of complex genomes and map DNA methylation domains. *BMC Genomics* 9, 621. doi: 10.1186/1471-2164-9-621

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., and Portnoy, D. S. (2018). These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Mol. Ecol.* 27 (16), 3193–3206. doi: 10.1111/mec.14792

Ortego, J., Gugger, P. F., and Sork, V. L. (2018). Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytol.* 218, 804–818. doi: 10.1111/nph.14951

Ouellette, L. A., Reid, R. W., Blanchard, J. S. G., and Brouwer, C. R. (2017). LinkageMapView - Rendering High Resolution Linkage and QTL Maps. *Bioinformatics* 34 (2), 306–307. doi: 10.1093/bioinformatics/btx576

Perea Garcia-Calvo, R. (2006). *Estudio de la estructura de masa de una dehesa de encina con alcornoque en "El Deheson del Encinar"* (Toledo: [dissertation]: Universidad Politécnica de Madrid).

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and 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 (5), e37135. doi: 10.1371/journal.pone.0037135

Petit, R. J., Bodénès, C., Ducousso, A., Roussel, G., and Kremer, A. (2003). Hybridization as a mechanism of invasion in oaks. *New Phytol.* 161, 151–164. doi: 10.1046/j.1469-8137.2003.00944.x

Plomion, C., Aury, J. M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., et al. (2018). Oak genome reveals facets of long lifespan. *Nat. Plants* 4, 440–452. doi: 10.1038/s41477-018-0172-3

Poland, J. A., and Rife, T. W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome* 5, 92–102. doi: 10.3835/plantgenome2012.05.0005

Pootakham, W., Ruang-Areerate, P., Jomchai, N., Sonthirod, C., Sangsrakru, D., Yoocha, T., et al. (2015). Construction of a high-density integrated genetic linkage map of rubber tree (Hevea brasiliensis) using genotyping-by-sequencing (GBS). *Front. Plant Sci.* 6, 367. doi: 10.3389/fpls.2015.00367

Pootakham, W., Sonthirod, C., Naktang, C., Jomchai, N., Sangsrakru, D., and Tangphatsornruang, S. (2016). Effects of methylation-sensitive enzymes on the enrichment of genic SNPs and the degree of genome complexity reduction in a two-enzyme genotyping-by-sequencing (GBS) approach: a case study in oil palm (*Elaeis guineensis*). *Mol. Breed.* 36 (11), 154. doi: 10.1007/s11032-016-0572-x

Potts, B., and Reid, J. B. (1988). Hybridization as a dispersal mechanism. *Evolution* 42, 1245–1255. doi: 10.1111/j.1558-5646.1988.tb04184.x

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genomics* 155, 945–959. doi: 10.1111/j.1471-8286.2007.01758.x

Radosavljevic, I., Bogdanovic, S., Celep, F., Filipovic, M., Satovic, Z., Surina, B., et al. (2019). Morphological, genetic and epigenetic aspects of homoploid hybridization between *Salvia officinalis* L. and *Salvia fruticose* Mill. *Sci. Rep.* 9, 3276. doi: 10.1038/s41598-019-40080-0

Ramos, A., Usié, A., Barbosa, P., Barros, P. M., Capote, T., Chaves, I., et al. (2018). The draft genome sequence of cork oak. *Sci. Data* 5, 180069. doi: 10.1038/sdata.2018.69

Saintagne, C., Bodenes, C., Barreneche, T., Pot, D., Plomion, C., and Kremer, A. (2004). Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity* 92 (1), 20–30. doi: 10.1038/sj.hdy.6800358

Salmon, A., Ainouche, M. L., and Wendel, J. F. (2005). Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Mol. Ecol.* 14, 1163–1175. doi: 10.1111/j.1365-294X.2005.02488.x

Sexton, J. P., McIntyre, P. J., Angert, A. L., and Rice, K. J. (2009). Evolution and ecology of species range limits. *Annu. Rev. Ecol. Evol. Syst.* 40, 415–436. doi: 10.1146/annurev.ecolsys.110308.120317

Simeone, M. C., Grimm, G. W., Papini, A., Vessella, F., Cardoni, S., Tordoni, E., et al. (2016). Plastome data reveal multiplegeographic origins of *Quercus* Group Ilex. *PeerJ* 4, e1897. doi: 10.7717/peerj.1897

Simeone, M. C., Cardoni, S., Piredda, R., Imperatori, F., Avishai, M., Grimm, G. W., et al. (2018). Comparative systematics and phylogeography of Quercus Section Cerris in western Eurasia: inferences from plastid and nuclear DNA variation. *PeerJ* 6, e5793. doi: 10.7717/peerj.5793

Soto, Á., Rodríguez-Martínez, D., and López de Heredia, U. (2018). SimHyb: a simulation software for the study of the evolution of hybridizing populations. Application to *Quercus ilex* and *Q. suber* suggests hybridization could be underestimated. *iForest* 11, 99–103. doi: 10.3832/ifor2569-011

Suárez-González, A., Lexer, C., and Cronk, Q. C. B. (2018). Adaptive introgression: a plant perspective. *Biol. Lett.* 14 (3), 20170688. doi: 10.1098/rsbl.2017.0688

Ueno, S., Le Provost, G., Leger, V., Klopp, C., Noirot, C., Frigerio, J. M., et al. (2010). Bioinformatics analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics* 11, 650. doi: 10.1186/1471-2164-11-650

Valbuena-Carabaña, M., González-Martínez, S. C., Sork, V. L., Collada, C., Soto, Á., Goicoechea, P. G., et al. (2005). Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. *Heredity* 95, 457–465. doi: 10.1038/sj.hdy.6800752

van Valen, L. (1976). Ecological species, multispecies and oaks. *Taxon* 25, 233–239. doi: 10.2307/1219444

Vitelli, M., Vessella, F., Cardoni, S., Pollegioni, P., Denk, T., Grimm, G. W., et al. (2017). Phylogeographic structuring of plastome diversity in Mediterranean oaks (Quercus Group Ilex, Fagaceae). *Tree Genet. Genomes* 13, 3. doi: 10.1007/s11295-016-1086-8

Wu, T. D., Reeder, J., Lawrence, M., Becker, G., and Brauer, M. J. (2016). GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. *Methods Mol. Biol.* 1418, 283–334. doi: 10.1007/978-1-4939-3578-9_15

Zeng, Y. F., Liao, W. J., Petit, R. J., and Zhang, D. Y. (2011). Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. *Mol. Ecol.* 20 (23), 4995–5011. doi: 10.1111/j.1365-294X.2011.05354.x

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.

The reviewer JM declared a past co-authorship with one of the authors MS to the handling Editor.

frontiers in Plant Science

CORRECTION
published: 10 January 2022
doi: 10.3389/fpls.2021.786695

Corrigendum: ddRAD Sequencing-Based Identification of Genomic Boundaries and Permeability in *Quercus ilex* and *Q. suber* Hybrids

Unai López de Heredia[1], *Fernando Mora-Márquez*[1], *Pablo G. Goicoechea*[2], *Laura Guillardín-Calvo*[1†], *Marco C. Simeone*[3] *and Álvaro Soto*[1*]

[1] G.I. Genética, Fisiología e Historia Forestal, Dpto. Sistemas y Recursos Naturales, ETSI Montes, Forestal y del Medio Natural, Universidad Politécnica de Madrid, Madrid, Spain, [2] Department of Forestry, NEIKER-BRA, Vitoria-Gasteiz, Spain, [3] Dipartimento di Scienze Agrarie e Forestali (DAFNE), Università degli Studi della Tuscia, Viterbo, Italy

Keywords: ddRADSeq, *Quercus*, hybridization, introgression, genomic boundaries, SNPs

Edited and reviewed by:
Gonzalo Nieto Feliner, Real Jardín Botánico (RJB, CSIC), Spain

***Correspondence:**
Álvaro Soto
alvaro.soto.deviana@upm.es

†Present address:
Laura Guillardín-Calvo, Department of Plant Sciences, University of Oxford, Oxford, United Kingdom

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

Received: *30 September 2021*
Accepted: *18 November 2021*
Published: *10 January 2022*

Citation:
López de Heredia U, Mora-Márquez F, Goicoechea PG, Guillardín-Calvo L, Simeone MC and Soto Á (2022) Corrigendum: ddRAD Sequencing-Based Identification of Genomic Boundaries and Permeability in Quercus ilex and Q. suber Hybrids. Front. Plant Sci. 12:786695. doi: 10.3389/fpls.2021.786695

A Corrigendum on

ddRAD Sequencing-Based Identification of Genomic Boundaries and Permeability in *Quercus ilex* and *Q. suber* Hybrids

by López de Heredia, U., Mora-Márquez, F., Goicoechea, P.G., Guillardín-Calvo, L., Simeone, M. C., and Soto, Á. (2020). Front. Plant Sci. 11:564414. doi: 10.3389/fpls.2020.564414

The article originally published contained errors caused by mistakes made in the automated variant calling process, which led to the misidentification of several polymorphic loci. This affected mainly to the exact number of markers of different classes, reported in several sections of the text. The details of the affected parts of the article are specified here.

1. In the *Abstract*

The sentence "We have identified up to 9,435 markers across the genome and have estimated individual introgression levels in adults and seedlings." should read "We have identified up to 9,251 markers across the genome and have estimated individual introgression levels in adults and seedlings."

Furthermore, the sentences "A noticeable proportion of the markers (26%) showed allelic frequencies in adult hybrids very similar to one of the parental species, and very different from the other; a finding that seems relevant for understanding the hybridization process and the occurrence of adaptive introgression. Candidate marker databases developed in this study constitute a valuable resource to design large scale re-sequencing experiments in Mediterranean sclerophyllous oak species and could provide insight in species boundaries and on adaptive introgression between *Q. suber* and *Q. ilex.*" should read "In adult hybrids 273 markers (3%) showed allelic frequencies very similar to one of the parental species, and very different from the other; these loci could be relevant for understanding the hybridization process and the occurrence of adaptive introgression. Candidate marker databases developed in this study constitute a valuable resource to design large scale re-sequencing experiments in Mediterranean sclerophyllous oak species and could provide insight into species boundaries and adaptive introgression between *Q. suber* and *Q. ilex*."

2. In the *Results*, Subsection *Read Alignment, Variant Filtering and Imputation*

In the second paragraph, the sentences "After individual variant calling, the number of variants ranged between 14,666 and 539,229 for the genome and between 217 and 71,716

for the pseudogenome alignments. The final concatenated-merged variant calling file had 17,289,128 variants, of which >99.5% were SNPs and <8% multi-allelic sites." should read "After individual variant calling, the number of variants ranged between 14,593 and 524,458 for the genome and between 212 and 66,680 for the pseudogenome alignments. The final concatenated-merged variant calling file had 16,234,798 variants, of which 97.4 % were SNPs and <5% multi-allelic sites."

In the third paragraph, the sentence "This way, ScnII kept approximately 2/3 of the loci imputed under ScnI." should read "This way, ScnII kept approximately 80% of the loci imputed under ScnI."

In the fourth paragraph, the sentences "The number of final recovered loci varied depending on the scenario (**Figure 4A**). For ScnI and ScnIII we obtained up to 9,435 loci, with 36.6% of imputed ones in ScnI. Under ScnII we considered 8,175 loci, with 26.6% of them imputed. The more restrictive ScnIV kept 6,001 unimputed loci." should read "The number of final recovered loci varied depending on the scenario (**Figure 4A**). For ScnI and ScnIII we obtained up to 9,251 loci, with 2,026 (21.9%) imputed ones in ScnI. Under ScnII we considered 8,901 loci, with 18.8% of them imputed. The more restrictive ScnIV kept 7,225 unimputed loci."

Furthermore, in the fourth paragraph, the sentences "Loci from ScnI/ScnIII were located in 3,156 fragments, of which 2,406 were genic and 750 intergenic. Under ScnII only two intergenic fragments were completely discarded, resulting in a total of 3,154 identified fragments. For ScnIV the number of fragments dropped to 2,166, of which 1,577 (72.8%) corresponded to genic regions and 589 (27.2%) to intergenic ones. In all the scenarios, loci corresponding to genic regions were mostly exonic (>50%), although a significant percentage of loci (*c.* 20%) occurred in introns (**Figure 4B**). The remaining loci (3–4%) were located in 107 fragments that could not be mapped to the *Q. suber* genome assembly." should read "Loci from ScnI, ScnII and ScnIII were located in 3,396 fragments, of which 2,566 were genic and 811 intergenic. For Scn IV the number of fragments dropped to 1,829, of which 1,540 (84.2%) corresponded to genic regions and 279 (15.3%) to intergenic ones. In all the scenarios, loci corresponding to genic regions were mostly exonic *c.* 72%), although a significant percentage of loci (*c.* 28%) occurred in introns (**Figure 4B**). The remaining loci (0.3%) were located in 29 fragments that could not be mapped to the *Q. suber* genome assembly."

3. In the *Results*, Subsection *Distribution of Markers Across the Genome*

The sentences "A total of 8,210 loci were successfully mapped against *Q. robur* genome; of these, 7,559 showed homology with loci included in the 12 linkage groups. These loci belong to 2,764 genomic fragments: 2,110 genic, 646 intergenic, and 8 fragments not found in the *Q. suber* genome. We found a rather even distribution of these loci among the 12 linkage groups, with an average distribution of more than 600 loci per linkage group, approximately 10.55 loci/Mb (**Figure 5**)." should read "A total of 8,774 loci were successfully mapped against *Q. robur* genome; of these, 8,004 showed homology with loci included in the 12 linkage groups. These loci belong to 2,932 genomic fragments: 2,264 genic and 668 intergenic. We found a rather even distribution of these loci among the 12 linkage groups, with an average distribution of almost 670 loci per linkage group, approximately 10.55 loci/Mb (**Figure 5**)."

4. In the *Results*, Subsection *Introgression Levels*

In the first paragraph, the sentences "ScnIV provided a fairly accurate classification of virtual hybrid individuals. ScnII, and, most of all, ScnI, provided even more precise classifications. On the contrary, ScnIII yielded large deviations for virtual individuals. Therefore, ScnIII was discarded for further analysis of real individuals (**Figure 6**)." should read "ScnIII yielded the same results as ScnIV, due to the distribution of missing data among species and the way both programs consider them. Therefore, ScnIII was discarded for further analysis of real individuals (**Figure 6**)."

In the second paragraph, the sentences, "Estimation was performed considering 1 and 10% of hybrids in the analyzed population. INTROGRESS and STRUCTURE yielded similar results in each situation, and very small differences were detected between both hybrid prevalence situations. On the contrary, noticeably different results were obtained for ScnI and ScnII on one hand, and ScnIV on the other. A much larger contribution of *Q. ilex* was estimated under ScnI and ScnII. Only FS-01 showed a roughly similar contribution of both parental species while the rest of hybrids could be rather classified as backcrosses with *Q. ilex*. Under ScnIV, estimations for adult individuals were roughly compatible with F1 hybrids (except for FS-01, which could be classified as a backcross with *Q. suber*) (**Figure 7**)." Should read "Estimation was performed considering 1% and 10% of hybrids in the analyzed population. INTROGRESS and STRUCTURE yielded similar results in each situation, and very small differences were detected between both hybrid prevalence situations. Under the four imputation scenarios, estimations for adult individuals were roughly compatible with F1 hybrids (except for FS-01, which could be classified as a backcross with *Q. suber*) (**Figure 7**)."

5. In the *Discussion*, Subsection *Candidate Marker Loci Identification*

In the second paragraph, the sentences "Actually, genome mapping and variant calling using *Q. suber* genome assembly as a reference have confirmed that most candidate polymorphic markers (*c.* 73%) correspond to genic regions, more than 50% of loci are located in exons and *c.* 20% in introns. Approximately 25% of loci were located in intergenic regions, and, comparatively few candidate loci (*c.* 3%) were obtained from the pseudogenome mapping." should read "Actually, genome mapping and variant calling using *Q. suber* genome assembly as a reference have confirmed that most candidate polymorphic markers (*c.* 80%) correspond to genic regions, more than 55% of loci are located in exons and *c.* 22% in introns. Approximately 20% of loci were located in intergenic regions, and, comparatively few candidate loci (0.3%) were obtained from the pseudogenome mapping."

In the third paragraph, the sentence "Using restrictive filtering criteria (ScnIV), we have obtained 6,001 markers that correspond to 1,577 genic fragments of known function, 489 intergenic fragments, and 107 fragments that could not be assigned to *Q. suber* genome assembly." should read "Using restrictive filtering

criteria (ScnIV), we have obtained 7,225 markers that correspond to 1,540 genic fragments of known function, 279 intergenic fragments and 10 fragments that could not be assigned to *Q. suber* genome assembly."

Furthermore, in the third paragraph, the sentences "This way we identified up to 3,434 additional loci, with imputed null alleles, under ScnI. These loci, which could be highly informative for introgression studies, belonged to 2,406 genic fragments of known function, 750 intergenic fragments, and 107 fragments that could not be assigned to *Q. suber* genome assembly. It is noteworthy that many of these null alleles were imputed to *Q. suber*. Given the large number of imputed loci and their asymmetric distribution between both species, we prepared an additional filtering of imputed loci (ScnIII), considering as missing data the imputed alleles from ScnI. However, estimations of the introgression levels for simulated individuals showed a poor accuracy under ScnIII; therefore, it was discarded in further analysis." should read "This way we identified up to 2,026 additional loci, with imputed null alleles, under ScnI. These loci, which could be highly informative for introgression studies, belonged to 1,264 genic fragments of known function, 584 intergenic fragments and 9 fragments that could not be assigned to *Q. suber* genome assembly. It is noteworthy that many of these null alleles were imputed to *Q. ilex*."

In the fourth paragraph, the sentence "Thus, 2,457 loci under ScnI show allelic frequencies in the hybrids quite similar to those of *Q. ilex* and very different from *Q. suber*, while just 34 loci show frequencies in the hybrids very similar to *Q. suber* and different from *Q. ilex*." should read "Thus, in addition to the 2,026 imputed loci, up to 2,830 non-imputed ones show very different patterns in both species, with frequencies of the most common allele $\geq$0.9 in one of the species and $\leq$0.2 in the other one. Regarding the hybrids, under ScnI 190 loci (167 imputed) show allelic frequencies in the hybrids quite similar to those of *Q. ilex* and very different from *Q. suber*, while 83 loci (17 imputed) show frequencies in the hybrids very similar to *Q. suber* and different from *Q. ilex*."

6. In the *Discussion*, Subsection *Individual Introgression Levels*

In the first paragraph the sentences "As pointed out above, the estimations under ScnI on one side and under ScnIV on the other constitute the limits between which real introgression levels probably lie. Under ScnIV, which considers up to 6,001 markers, most adult hybrids could be classified as F1 hybrids. On the contrary, it is noteworthy that inclusion of imputable loci in the analysis (ScnI and, to a lesser extent, ScnII), yields a higher contribution of *Q. ilex* to adult hybrid genomes compared to ScnIV. Since most of the null alleles are imputed to *Q. suber*, this result must be due to a higher proportion of non-imputed, "*ilex*" alleles in heterozygosity in these loci in adult hybrids. Taking into account PstI/MspI sensitivity to methylation, hybridization-mediated alteration of epigenetic characters could also contribute to the apparent higher contribution of *Q. ilex* to the genome of hybrid individuals. This way, methylated epialleles in the restriction sites in *Q. suber*, which would yield no scorable reads and, therefore, would have been imputed with a null allele, could have turned out to be unmethylated and therefore scorable in hybrids, yielding an apparent higher contribution of *Q. ilex* even to F1 hybrids." should read "Most adult hybrids could be classified as F1 hybrids under all the imputation scenarios considered. Inclusion of imputed loci does not entail a significant difference in the estimation of the contribution of parental species to the genome of hybrid individuals. For the adult hybrids, only slightly lower values of *Q. suber* contribution are obtained under ScnI and ScnII. Different results are observed for the hybrid progenies. Individuals with higher estimated *Q. suber* contributions under ScnIV show lower values when imputed loci are considered, while the opposite is observed for individuals with lower estimations. Since most of the null alleles are imputed to *Q. ilex*, this latter result must be due to a higher proportion of non-imputed, "*suber*" alleles in heterozygosity in these loci in these individuals. Taking into account PstI/MspI sensitivity to methylation, hybridization-mediated alteration of epigenetic characters could also contribute to these results. This way, methylated epialleles in the restriction sites, which would yield no scorable reads and, therefore, would have been imputed with a null allele, could have turned out to be unmethylated and therefore scorable in hybrids, or vice-versa. This could be the case at least of the 184 markers for which very high frequencies of the imputed allele are recorded in adult hybrids ($\geq$0.75), no matter their global classification as F1 hybrids."

7. In the Section *Conclusion and Future Prospect*

The first and second paragraphs "Our work reports a case study of hybridization and introgression in two non-model forest tree species, *Q. suber* and *Q. ilex*, using genome-wide NGS techniques, and provides a pipeline and scripts for this kind of studies. We have identified up to 9,435 marker loci in *Q. suber* and *Q. ilex*. Among them, allelic frequencies of 2,457 are quite similar in hybrid adult individuals and in *Q. ilex*, while only 34 are quite similar in hybrids and *Q. suber*, consistently with the estimated higher contribution of this latter species to the genome of adult hybrids.

Additionally, we have detected 3,434 highly discriminating loci for which a species-specific null allele has been imputed. In most cases, the fragment was scored in *Q. ilex* samples, and absent in *Q. suber*. This can be due to alterations in restriction enzyme target sites or to real indels. Interestingly, in many cases hybrid individuals show the presence of *Q. ilex* variants, rather than *Q. suber* variants, suggesting a selection of these alleles in backcrosses or hybridization-mediated alterations of the methylation patterns. In any case, these loci deserve further attention, since they could be linked to viability of hybrid individuals or to selective advantages." should read (after being merged into a single paragraph). "Our work reports a case study of hybridization and introgression in two non-model forest tree species, *Q. suber* and *Q. ilex*, using genome-wide NGS techniques, and provides a pipeline and scripts for this kind of studies. Out of the 9,251 marker loci identified in this study, 4,856 are highly discriminant between both species, and 2,026 of these are apparently absent in one of the species (*Q. ilex* in most cases). This can be due to alterations in restriction enzyme target sites or to real indels. Interestingly, for 9.1%

of them adult hybrids show patterns quite similar to one of the parental species (8.3% to *Q. ilex*, while only 0.8% to *Q. suber*), suggesting selection of those alleles in backcrosses or hybridization-mediated alterations of the methylation patterns. In any case, these loci deserve further attention, since they could be linked to viability of hybrid individuals and/or to selective advantages.

8. *Materials and Methods*, Subsection *Estimation of Introgression Levels*

In the original article, the correct website for downloading the SIMHYB software was not cited. Therefore, the sentence "These individuals were simulated with SIMHYB (Soto et al., 2018), based on the allele frequencies of the adult *Q. ilex* and *Q. suber* populations." should read "These individuals were simulated with SIMHYB (Soto et al., 2018; https://github.com/GGFHF/SimHyb), based on the allele frequencies of the adult *Q. ilex* and *Q. suber* populations."

9. Table Errors

In the original article, there were some mistakes in **Tables 2**, **3** as published. The tables corresponded to the results obtained after an incorrect variant calling procedure. The corrected **Tables 2**, **3** appear here.

10. Figure Errors

In the original article, there were some mistakes in **Figures 4**–**7** as published. The figures corresponded to the results obtained after an incorrect variant calling procedure. The corrected **Figures 4**–**7** appear here. Furthermore, in the original article, there were some mistakes in the legends for **Figures 4**–**7** as published. The figures corresponded to the results obtained after an incorrect variant calling procedure, and the legends were phrased accordingly. The correct legends appear here.

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

REFERENCES

López de Heredia, U., Sánchez, H., Soto, Á. (2018a). Molecular evidence of bidirectional introgression between Quercus suber and Quercus ilex. *iForest*. 11, 338–343. doi: 10.3832/ifor2570-011

Soto, Á., Rodríguez-Martínez, D., López de Heredia, U. (2018). SimHyb: a simulation software for the study of the evolution of hybridizing populations. Application to Quercus ilex and Q. suber suggests hybridization could be underestimated. *iForest*. 11, 99–103. doi: 10.3832/ifor2569-011

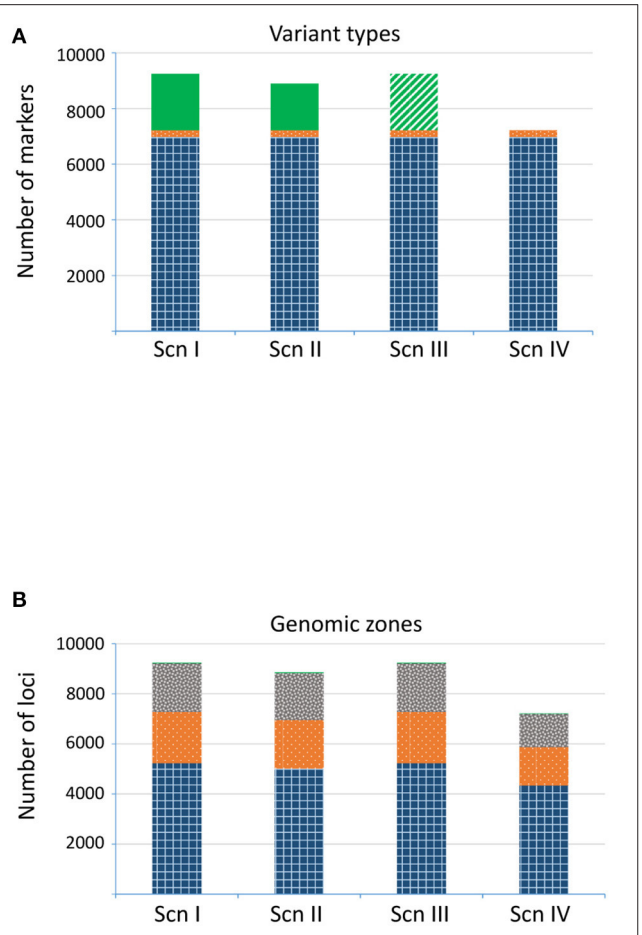


FIGURE 4 | Number of loci detected for each scenario. **(A)** Variant types: SNP (gridded blue), indel (dotted orange), and imputed loci (plain green; treated as missing data in ScnIII). **(B)** Genomic region: known exonic region (gridded blue), known intronic region (dotted orange), intergenic region (gray confetti), not assigned region (plain green).

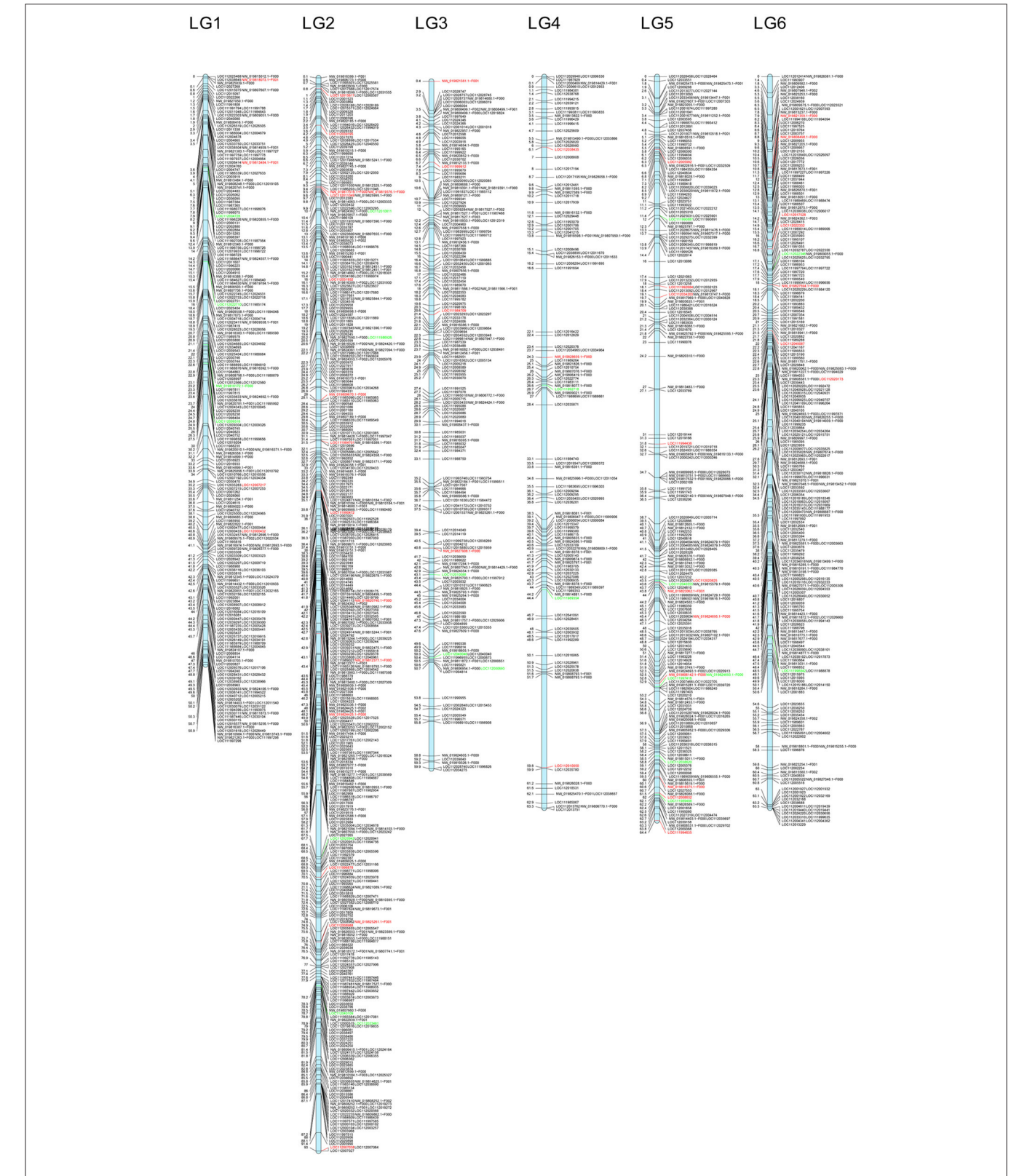


FIGURE 5 | Continued

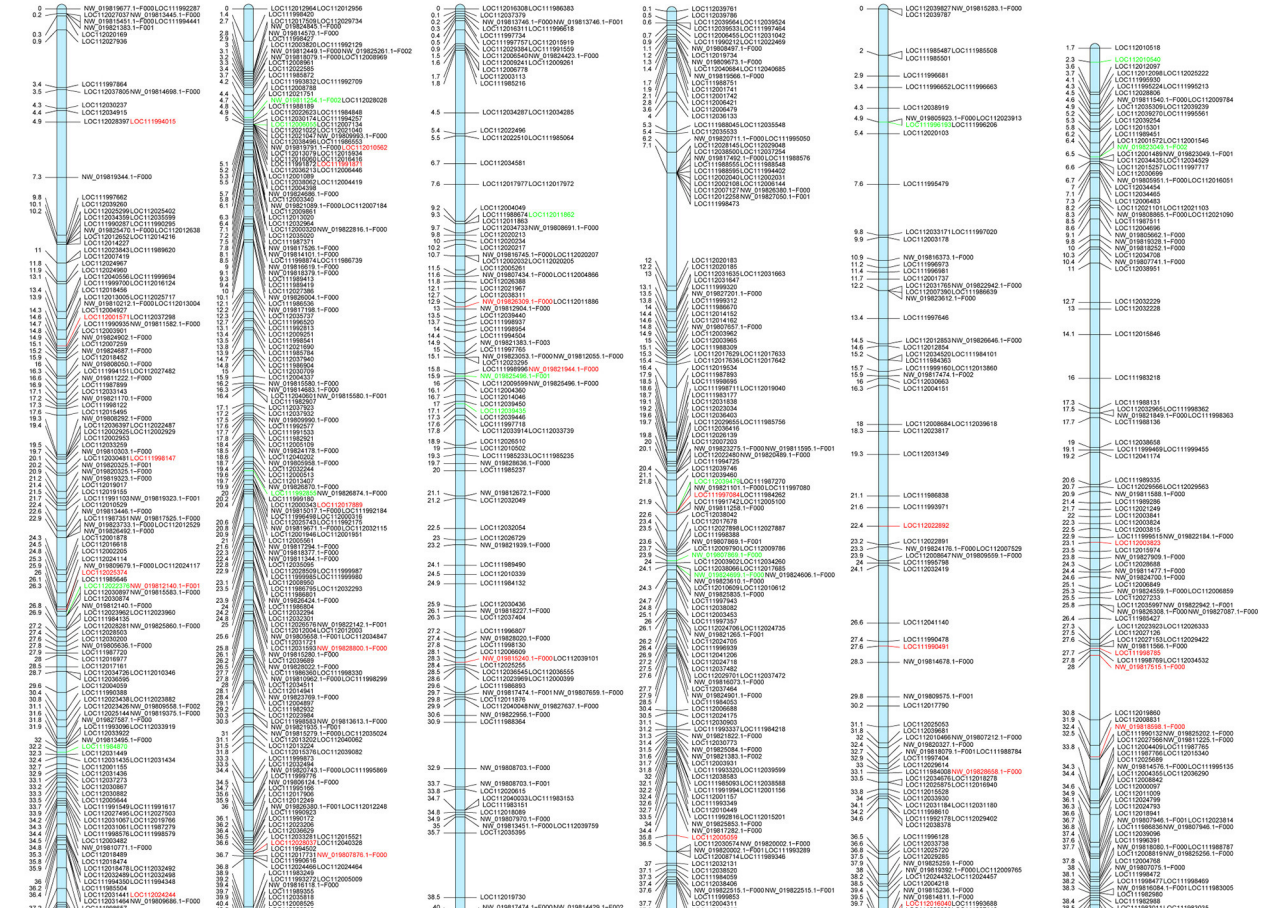


FIGURE 5 | Putative location in the *Q. robur* linkage groups of the genomic fragments including markers from ScnI. Location of markers with allelic frequencies in the adult hybrids very similar to one of the parental species is highlighted in red (*Q. ilex*) or in green (*Q. suber*).

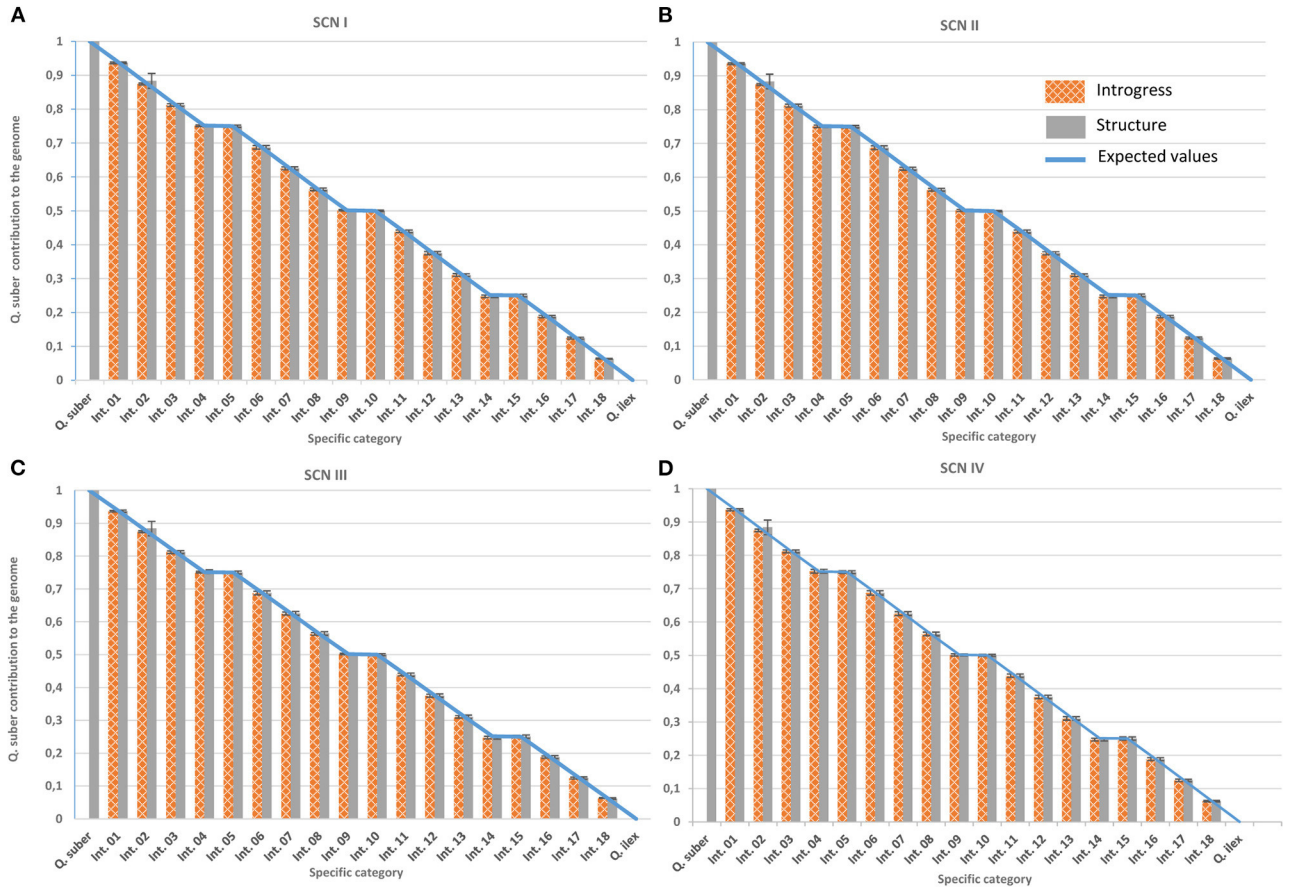


FIGURE 6 | Performance of classification tools under the different imputation scenarios. Genomic contribution of *Q. suber* is estimated using STRUCTURE's q_s (plain grey) and INTROGRESS hybrid index (gridded orange) on virtual individuals generated with SimHyb (pure species and 18 intermediate categories). Expected values are represented by a blue line. Standard deviations are also indicated. **(A)** imputation scenario I; **(B)** imputation scenario II; **(C)** imputation scenario III; **(D)** imputation scenario IV.

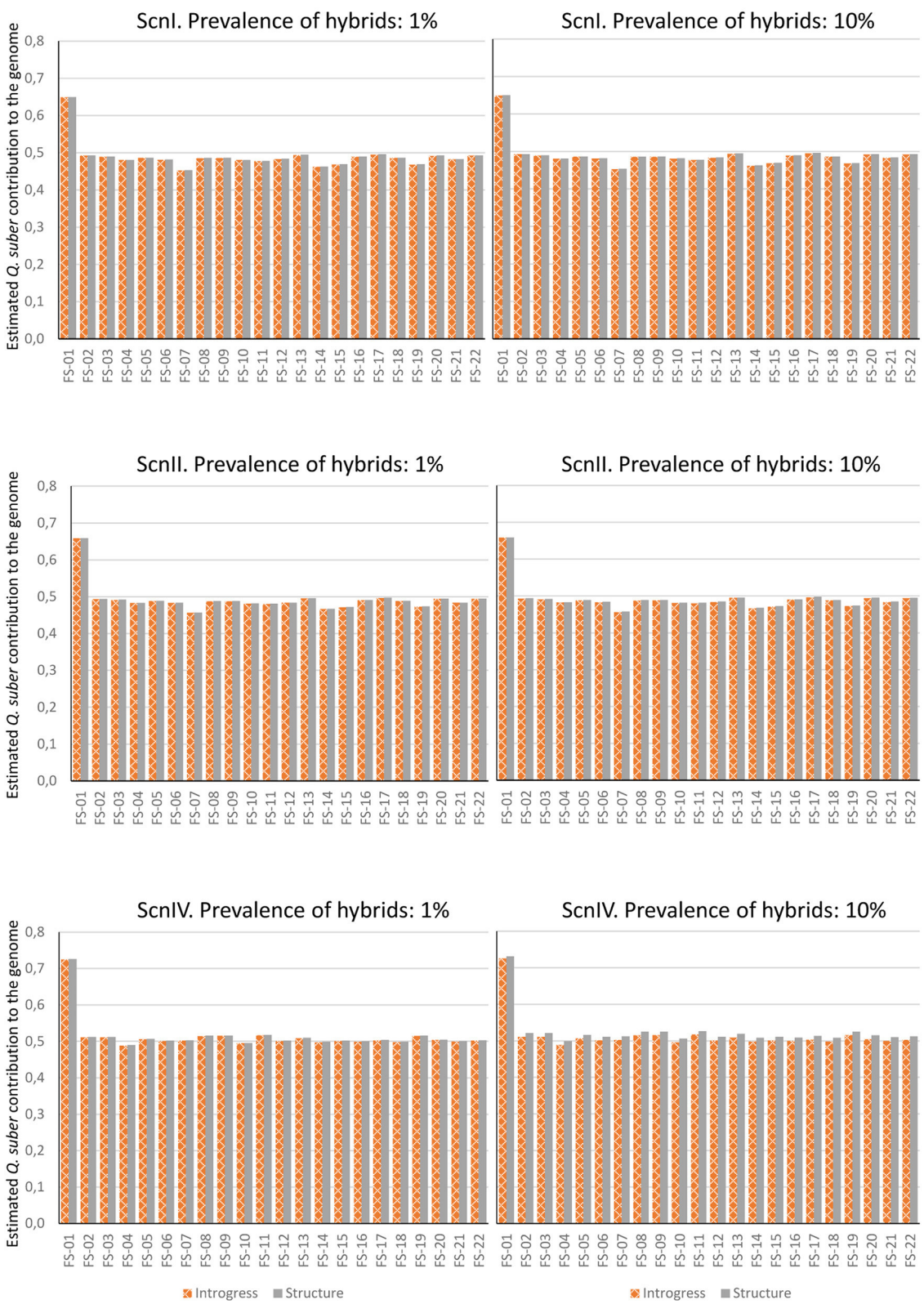


FIGURE 7 | Contribution of *Q. suber* to the genome of adult hybrids under each scenario, estimated by means of STRUCTURE's qs and Hybrid Index of INTROGRESS, with 1% and 10% of hybrids.

TABLE 2 | STRUCTURE's q_s and INTROGRESS hybrid index estimates for the progenies of the open-pollinated hybrid families under each scenario.

Mother tree	Scn I			Scn II			Scn IV		
	q_s	Max	Min	q_s	Max	Min	q_s	Max	Min
FS08	0.633 (0.049)	0.714	0.543	0.644 (0.051)	0.726	0.548	0.731 (0.067)	0.821	0.606
FS14	0.648 (0.096)	0.776	0.322	0.658 (0.100)	0.791	0.319	0.731 (0.125)	0.904	0.316
FS16	0.626 (0.058)	0.713	0.317	0.634 (0.061)	0.726	0.313	0.700 (0.075)	0.818	0.311
FS17	0.655 (0.073)	0.758	0.361	0.665 (0.077)	0.773	0.358	0.738 (0.096)	0.878	0.365
FS18	0.662 (0.035)	0.718	0.596	0.672 (0.037)	0.731	0.602	0.748 (0.048)	0.824	0.659
FS19	0.587 (0.152)	0.769	0.155	0.595 (0.157)	0.784	0.149	0.657 (0.190)	0.893	0.114
FS20	0.663 (0.058)	0.773	0.521	0.674 (0.061)	0.789	0.524	0.752 (0.079)	0.901	0.561
FS21	0.625 (0.013)	0.640	0.607	0.633 (0.014)	0.648	0.614	0.701 (0.019)	0.722	0.677
FS22	0.633 (0.056)	0.721	0.359	0.643 (0.058)	0.736	0.373	0.718 (0.068)	0.835	0.487

Mean, standard deviation (in brackets), maximum and minimum values per family are provided.

TABLE 3 | Ratio between STRUCTURE's q_s of the offspring and q_s of their mothers under each scenario.

Mother tree	Scn I			Scn II			Scn IV		
	$q_s o/q_s m$	Max	Min	$q_s o/q_s m$	Max	Min	$q_s o/q_s m$	Max	Min
FS08	1.303	1.469	1.117	1.316	1.485	1.121	1.392	1.564	1.154
FS14	1.400	1.676	0.695	1.405	1.690	0.682	1.438	1.780	0.622
FS16	1.277	1.455	0.647	1.290	1.479	0.637	1.375	1.607	0.611
FS17	1.321	1.528	0.728	1.335	1.552	0.719	1.439	1.712	0.712
FS18	1.359	1.474	1.224	1.374	1.495	1.231	1.472	1.622	1.297
FS19	1.252	1.640	0.330	1.254	1.654	0.314	1.252	1.701	0.217
FS20	1.346	1.568	1.057	1.358	1.591	1.056	1.459	1.750	1.089
FS21	1.290	1.322	1.254	1.304	1.336	1.266	1.375	1.416	1.327
FS22	1.284	1.462	0.728	1.298	1.487	0.754	1.403	1.631	0.951

Mean, maximum and minimum values per hybrid family are provided. Mean for the open-pollinated hybrid families by scenario. Values > 1 point to likely fertilizations by Q. suber, while valuses < 1 suggest fertilization by Q. ilex (López de Heredia et al., 2018a).

frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fpls.2020.612258

Geo-Climatic Changes and Apomixis as Major Drivers of Diversification in the Mediterranean Sea Lavenders (*Limonium* Mill.)

Konstantina Koutroumpa[1†], Ben H. Warren[1,2], Spyros Theodoridis[1,3], Mario Coiro[1], Maria M. Romeiras[4,5], Ares Jiménez[1,6] and Elena Conti[1*]*

[1]*Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland,* [2]*Institut de Systematique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, UA, Paris, France,* [3]*Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft für Naturforschung, Frankfurt am Main, Germany,* [4]*Linking Landscape, Environment, Agriculture and Food (LEAF), Instituto Superior de Agronomia (ISA), Universidade de Lisboa, Lisboa, Portugal,* [5]*Centre for Ecology, Evolution and Environmental Changes (cE3c), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal,* [6]*IES Pedra da Auga, Ponteareas, Spain*

OPEN ACCESS

Edited by:
Andrew A. Crowl,
Duke University, United States

Reviewed by:
Mario Fernández-Mazuecos,
Real Jardín Botánico (RJB), Spain
Santiago Martín-Bravo,
Universidad Pablo de Olavide, Spain

****Correspondence:***
Konstantina Koutroumpa
konstantina.koutroumpa@systbot.uzh.ch
Elena Conti
elena.conti@systbot.uzh.ch

†Present address:
Konstantina Koutroumpa,
Botanischer Garten und Botanisches Museum Berlin (BGBM), Freie Universität Berlin, Berlin, Germany

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

Received: *30 September 2020*
Accepted: *07 December 2020*
Published: *12 January 2021*

Citation:
Koutroumpa K, Warren BH, Theodoridis S, Coiro M, Romeiras MM, Jiménez A and Conti E (2021) Geo-Climatic Changes and Apomixis as Major Drivers of Diversification in the Mediterranean Sea Lavenders (Limonium Mill.). Front. Plant Sci. 11:612258. doi: 10.3389/fpls.2020.612258

The Mediterranean realm, comprising the Mediterranean and Macaronesian regions, has long been recognized as one of the world's biodiversity hotspots, owing to its remarkable species richness and endemism. Several hypotheses on biotic and abiotic drivers of species diversification in the region have been often proposed but rarely tested in an explicit phylogenetic framework. Here, we investigate the impact of both species-intrinsic and -extrinsic factors on diversification in the species-rich, cosmopolitan *Limonium*, an angiosperm genus with center of diversity in the Mediterranean. First, we infer and time-calibrate the largest *Limonium* phylogeny to date. We then estimate ancestral ranges and diversification dynamics at both global and regional scales. At the global scale, we test whether the identified shifts in diversification rates are linked to specific geological and/or climatic events in the Mediterranean area and/or asexual reproduction (apomixis). Our results support a late Paleogene origin in the proto-Mediterranean area for *Limonium*, followed by extensive *in situ* diversification in the Mediterranean region during the late Miocene, Pliocene, and Pleistocene. We found significant increases of diversification rates in the "Mediterranean lineage" associated with the Messinian Salinity Crisis, onset of Mediterranean climate, Plio-Pleistocene sea-level fluctuations, and apomixis. Additionally, the Euro-Mediterranean area acted as the major source of species dispersals to the surrounding areas. At the regional scale, we infer the biogeographic origins of insular endemics in the oceanic archipelagos of Macaronesia, and test whether woodiness in the Canarian *Nobiles* clade is a derived trait linked to insular life and a biotic driver of diversification. We find that *Limonium* species diversity on the Canary Islands and Cape Verde archipelagos is the product of multiple colonization events followed by *in situ* diversification, and that woodiness of the Canarian endemics is indeed a derived trait but is not associated with a significant shift to higher diversification rates. Our study expands knowledge on how the interaction between abiotic and biotic drivers shape the uneven distribution of species diversity across taxonomic and geographical scales.

Keywords: Messinian salinity crisis, Mediterranean climate, sea-level fluctuations, asexual reproduction, *in situ* diversification, island biogeography, Macaronesia, long-distance dispersal

INTRODUCTION

Biodiversity on Earth is unevenly distributed. Species richness varies across habitats, geographic regions, and taxonomic groups, raising long-standing questions about the ecological and evolutionary mechanisms underpinning this variation. Factors that drive speciation and extinction (i.e., species diversification) have altered biodiversity in space and time. Drivers of diversification include abiotic factors, for example, tectonic processes and climatic events (Barnosky, 2001), and biotic factors, for example, reproductive traits, ploidy levels, hybridization, and habit (Soltis et al., 2019). Numerous studies have focused on identifying a single key trait and linking it to shifts in diversification rates (e.g., Mayhew et al., 2008; de Vos et al., 2014; Howard et al., 2020). However, both biotic and abiotic factors can act synergistically toward changes in diversification (e.g., Bouchenak-Khelladi et al., 2015; Donoghue and Sanderson, 2015; Condamine et al., 2018). Analyses of spatio-temporal evolution and drivers of diversification in species-rich lineages are crucial to clarify the role of past biotic and abiotic changes in the origins of species diversity and predict how lineages will be affected by ongoing environmental changes.

Flowering plant diversity is partitioned taxonomically, geographically, and environmentally. Angiosperms comprise more than 13,000 genera (Christenhusz and Byng, 2016) ranging in size from a single to thousands of species, yet only about 70 genera are characterized as species rich (with ≥500 species; Frodin, 2004; Mabberley, 2017). Furthermore, plant diversity is concentrated in 36 global biodiversity hotspots that cover only 16% of Earth's surface but harbor more than 50% of endemic vascular plants and are undergoing remarkable loss of habitat (Myers et al., 2000; Mittermeier et al., 2011). The Mediterranean has been identified as one of the world's biodiversity hotspots by Mittermeier et al. (2011) and Critical Ecosystem Partnership Fund (2019). These authors defined the Mediterranean hotspot as comprising the Mediterranean region (i.e., mainland areas surrounding the Mediterranean Sea and the islands in it) and Macaronesia (i.e., Canaries, Cape Verde, Azores, Madeira, and Selvagen Islands). The Mediterranean region and Macaronesia combined are also often referred to as the Mediterranean realm (e.g., Comes, 2004; Mansion et al., 2009). This hotspot covers only 1.6% of the Earth's surface, yet accommodates 10% of its total plant species richness, representing the third richest hotspot with approximately 25,000 species, more than half of which are endemic (Medail and Quezel, 1997; Blondel et al., 2010; Critical Ecosystem Partnership Fund, 2019). Its geographic location at the crossroads of three continents (Europe, Africa, and Asia) makes the Mediterranean a large contact zone for taxa of different biogeographic origins (e.g., Eurasian Circumboreal, Irano-Turanian, and Saharo-Arabian), which, together with taxa that originated and diversified *in situ*, form its remarkably diverse flora (Blondel and Aronson, 1999).

The Mediterranean region has undergone multiple geological and climatic upheavals (Thompson, 2005). Geologically, the region originated from two ancient, independent ocean basins: the Alpine Tethys Ocean (opened during the Middle to Late Jurassic and related to the opening of the Central Atlantic) in the Northwest and the Neotethys Ocean (opened from the Triassic to the Jurassic between Laurasia and Gondwana) in the Southeast (van Hinsbergen et al., 2020 and references therein). From the Cretaceous to the early Miocene, a continuing convergence of tectonic plates brought Europe and Africa progressively closer (Rosenbaum et al., 2002). In the late Miocene, uplift at the continental margins of Iberia and Africa triggered extensive basin desiccation (Duggen et al., 2003; Gargani and Rigollet, 2007). This period, known as the Messinian Salinity Crisis (MSC, from ca. 5.96 to 5.33 Ma; Krijgsman et al., 1999), has been described as "one of the most dramatic events on Earth during the Cenozoic era" (Hsü et al., 1973; Duggen et al., 2003).

The MSC and the onset of the Mediterranean climate (3.2–2.8 Ma; Suc, 1984) were landmark events in the evolution of diversity in the Mediterranean region (Fiz-Palacios and Valcárcel, 2013). The creation of saline deserts during the MSC (Hsü et al., 1973) produced land bridges between islands and continental areas that potentially facilitated migrations of plants with the necessary dispersal properties and salt-tolerance (e.g., halophytes). The MSC is considered to have facilitated speciation in arid-adapted lineages and extinction in sub-tropical Tertiary lineages (Rodríguez-Sánchez et al., 2008; Jiménez-Moreno et al., 2010; Crowl et al., 2015). The refilling of the basin at the end of the MSC disrupted previously formed land bridges, thus promoting vicariance, and mitigated aridity (García-Castellanos et al., 2009), thus possibly causing extinction of arid-adapted lineages (Fiz-Palacios and Valcárcel, 2013). While the effects of MSC on the Mediterranean flora are still debated, the positive effects of the emergence of the Mediterranean climatic regime on diversification are corroborated by multiple studies (e.g., Valente et al., 2010; Fiz-Palacios and Valcárcel, 2011). Furthermore, several plant lineages show a temporal period of reduced diversification rate from the Messinian event to the onset of the Mediterranean climate that has been variably attributed to either mass extinction, rate stasis, or a combination of the two (Fiz-Palacios and Valcárcel, 2013).

During the late Pliocene-early Pleistocene, cooler and dryer conditions were implicated in several extinctions (Bessedik et al., 1984), while Pleistocene glacial cycles and eustatic sea-level changes (2.58–0.01 Ma; Lisiecki and Raymo, 2007) further impacted Mediterranean plant diversification and distributions. Pleistocene geoclimatic oscillations caused species range contractions and expansions as plant populations fragmented and merged in response to the appearance and disappearance of dispersal barriers through time, with contrasting effects on diversification of different Mediterranean plant lineages (Nieto Feliner, 2014). Furthermore, different types of islands (oceanic and continental), substrates, and microclimates provided opportunities for adaptation and speciation driven by both ecological and geographical barriers. Overall, geomorphological and climatic processes, coupled with a long history of human activities in the Mediterranean, created a mosaic of heterogeneous habitats, where a diversity of abiotic factors had profound impacts on diversification.

In addition to extrinsic environmental factors, inherent biological features also play a key role in diversification. For example, hybridization, polyploidy, plant habit, and reproductive strategies have all been invoked to explain species divergence and eco-geographical differentiation in plants (Rieseberg et al., 2003; Goldberg et al., 2010; Goldberg and Igić, 2012; Soltis et al., 2013). In particular, asexual reproduction *via* apomixis (i.e., cloning through seeds; Asker and Jerling, 1992) can promote rapid diversification by enabling reproductively isolated genotypes to form rapidly, providing reproductive assurance in the absence of pollinators and/or mates, and offering an escape from sterility in newly formed polyploid hybrid individuals (Baker, 1955; Darlington, 1958; Majeský et al., 2017). Apomixis can enable the establishment of a new population from a single individual, an analogous effect to that of selfing as proposed by Baker (1955). Furthermore, in oceanic island systems, such as Macaronesia, the evolution of woodiness from herbaceous ancestors following island colonization has been proposed as a key driver of insular radiations (Nürk et al., 2019). Thus, elucidating the timing, space, and rates of diversification of Mediterranean and Macaronesian lineages and attempting to correlate them with likely biotic and abiotic drivers is essential to explain the evolution of biodiversity in the Mediterranean hotspot (Comes, 2004).

To achieve a deeper understanding of diversification dynamics in the Mediterranean realm, it is necessary to focus on widely distributed plant groups with high species diversity in this region and biotic traits that could act as triggers of diversification given the unique geo-climatic history of the area. *Limonium* Mill. (sea lavender; Plumbaginaceae) qualifies as such a group. It is a species-rich genus (ca. 600 species; Koutroumpa et al., 2018) in the top 0.005% of angiosperm genera for size (according to Frodin's, 2004 criteria) and has a cosmopolitan distribution with center of diversity in the Mediterranean (ca. 70% of the species, mostly endemics, occur in the region). *Limonium* species, characterized as facultative halophytes (i.e., salt tolerant), grow predominantly on saline and metal-rich soils of mainland and coastal areas (Erben, 1978). They represent an important component of coastal vegetation in Mediterranean ecosystems and contribute the dominant species in several vegetation types (e.g., *Crithmo-Limonietea* class; Brullo et al., 2017). *Limonium* displays marked variation of chromosome numbers, ranging from 12 to 18 in diploids and from 24 to 72 in polyploids (Erben, 1979; Brullo and Erben, 2016). Sea lavenders can reproduce both sexually and asexually *via* apomixis, with most sexual species characterized by pollen-stigma dimorphism and sporophytic self-incompatibility (Baker, 1966). Apomixis occurs exclusively in polyploid taxa, some of presumed hybrid origin (Erben, 1979). The combined effects of polyploidy, hybridization, and apomixis have been suggested as having shaped *Limonium* diversity in the Mediterranean region (e.g., Ingrouille, 1984; Palacios et al., 2000; Lledó et al., 2005). Previous phylogenetic and taxonomic analyses concluded that all described apomictic species of *Limonium*, together with some sexual species, belong to a single large clade formed by the vast majority of Mediterranean endemics, named the "Mediterranean lineage" by Koutroumpa et al. (2018). The same study placed the Macaronesian endemics in four clades of the *Limonium* tree, with the majority of them included in the *Nobiles* and *Jovibarba-Ctenostachys* clades. The *Nobiles* clade consists entirely of Canarian endemics with a woody (suffruticose) habit. The *Jovibarba-Ctenostachys* clade comprises endemics of the Canaries and Cape Verde archipelagos together with endemics in Morocco and Hispaniola (Malekmohammadi et al., 2017; Koutroumpa et al., 2018). A previous phylogenetic study, limited to ca. 8% of *Limonium* species and based on a single biogeographic calibration, inferred a late Miocene (ca. 6–7 Ma) origin for the Mediterranean clade of *Limonium* with most diversification during the Messinian and Pliocene (Lledó et al., 2005). The same study inferred the diversification of Macaronesian clades between the late Miocene and Pliocene. Owing to its large size, worldwide distribution, and uneven diversity among regions, *Limonium* is an ideal group to elucidate the factors shaping the partitioning of biodiversity through space and time, warranting a new study with enhanced sampling and methodology.

Here, we generate and time calibrate the largest phylogeny to date of *Limonium* and Plumbaginaceae. By reconstructing ancestral areas of distribution, inferring the tempo of diversification, and estimating diversification rate dynamics, we address the following questions. At the global scale we ask: (1) When and where did the genus originate and diversify? (2) Can we detect significant shifts of diversification rates and link them to specific abiotic (major geological and/or climatic events) and/or biotic factors (apomixis)? More specifically, we test whether diversification rates are constant or heterogeneous in *Limonium*, and whether potential rate changes are concomitant with changes in extrinsic and intrinsic variables. At the regional scale, we focus on island biogeography, trait evolution, and diversification of Macaronesian *Limonium* by asking: (1) What are the biogeographic origins of island endemics in Macaronesia? (2) Did the transition from herbaceousness to woodiness precede or follow island colonization in the Canarian *Nobiles* clade? Specifically, we hypothesize that woodiness is a derived state linked to insular life, as suggested by several studies (e.g., Carine et al., 2010; Lens et al., 2013), rather than an ancestral state preserved in islands, as proposed by the "islands-as-museums" hypothesis (Cronk, 1997). (3) Did the transition to woodiness trigger an increase in diversification rate in the Canarian *Nobiles* clade, as found in other insular clades (Nürk et al., 2019)? Our findings highlight the importance of both extrinsic and intrinsic factors in generating the remarkable plant diversity of a global biodiversity hotspot, the Mediterranean realm.

MATERIALS AND METHODS

Taxon and Molecular Sampling

We sampled more than one-third of all *Limonium* species (i.e., 216 taxa representing 214 species: 212 species were identified at the species level, one species was identified at the subspecies level, and one species was represented by three varieties) together with 66 species of 20 other Plumbaginaceae genera.

Additionally, 20 species from the sister family of Polygonaceae (both subfamilies sampled) were used as outgroup taxa for phylogenetic inference (**Supplementary Table S1**). Our sampling of *Limonium* includes representatives from all taxonomic subdivisions and phylogenetic clades (Malekmohammadi et al., 2017; Koutroumpa et al., 2018) and spans its cosmopolitan distribution. Sampling of Plumbaginaceae and Polygonaceae provided appropriate nodes for fossil and secondary calibrations. Overall, our molecular dataset comprised 302 taxa and a total of 1,039 sequences from one nuclear (261 ITS sequences) and three chloroplast (291 *trnL-F*, 243 *rbcL*, and 244 *matK* sequences) markers. Here, we increase taxon sampling by 15 taxa compared to the previously published phylogeny of *Limonium* and Plumbaginaceae of Koutroumpa et al. (2018). Molecular methods for 70 newly generated sequences and alignments followed Koutroumpa et al. (2018).

Phylogenetic and Dating Analyses

Phylogenies were estimated using maximum likelihood (ML) and Bayesian inference (BI) as implemented in RAxML 8.2.10 (Stamatakis, 2014) and MrBayes 3.2.6 (Ronquist et al., 2012), respectively, following conditions described in Koutroumpa et al. (2018; except for BI of supermatrices for which chains ran for 30 million generations). Trees were inferred for each locus separately and in concatenation (concatenated datasets: 3-loci cpDNA matrix and 4-loci "Supermatrix"). We inspected the resulting cpDNA and ITS trees to detect incongruences between well-supported clades (as defined in Koutroumpa et al., 2018) and identify "rogue taxa" (i.e., taxa placed in different well-supported clades in the cpDNA vs. the ITS tree). Given the small number of conflicts thus identified, we additionally inferred phylogenies for a Supermatrix from which we excluded ITS sequences of "rogue clades and taxa" (hereafter "Supermatrix-cpDNA-like") and a Supermatrix from which we excluded cpDNA sequences of "rogue clades and taxa" (hereafter "Supermatrix-ITS-like"). Trees from these two trimmed supermatrices have higher resolution than those from the three-loci cpDNA matrix and ITS matrix.

Divergence time estimates were performed in BEAST 2.5.1 (Bouckaert et al., 2019) for ITS, cpDNA, "Supermatrix," "Supermatrix-ITS-like," and "Supermatrix-cpDNA-like" datasets, using five calibration points. Following a comprehensive review of literature for the Plumbaginaceae-Polygonaceae fossil record and dated angiosperm phylogenies, and following guidelines for nodal assignment of fossils (Magallon and Sanderson, 2001; Rutschmann et al., 2007), we calibrated four nodes (two within Plumbaginaceae and two within Polygonaceae), and the stem node of Plumbaginaceae, forming a total of five calibration points (see **Supplementary Figures S1**, **S2**); the details of the calibration process are described below. Although fossils for Plumbaginaceae are sparse, two internal nodes could be calibrated: an *Armeria*-type pollen fossil from upper Miocene (Van Campo, 1976; Muller, 1981) was used as a minimum age constraint (5.333 Ma) for the Limonieae crown-node (the origin of *Armeria*-type pollen according to Costa et al.'s, 2019 ancestral state estimates), and a pollen fossil of *Ceratolimon* cf. *feei* (former *Limoniastrum*; Beucher, 1975; Muller, 1981) from the Pliocene was used as minimum age constraint (2.58 Ma) for the *Ceratolimon* stem-node. We also used two fossils from Polygonaceae: a *Coccoloba* pollen fossil from the upper Miocene (Graham, 1976; Muller, 1981) was used to assign a minimum age (5.333 Ma) to the *Coccoloba* stem-node and *Muehlenbeckia*-type pollen (†*Rhoipites muehlenbeckiaformis*; Macphail, 1999) from the upper Eocene was used as minimum age constraint (33.9 Ma) for the *Muehlenbeckia* stem-node. Finally, we used a secondary calibration for the stem of Plumbaginaceae based on age estimates from the dated angiosperm phylogeny of Magallón et al. (2015) [mean: 67.9 Ma and 95% Highest Posterior Density (HPD): 65.63–78.21 Ma]. We employed uniform priors for fossil calibrations with a hard maximum bound of 78.21 Ma (Magallón et al.'s, 2015 HPD max age for Plumbaginaceae-Polygonaceae split) and minimum bounds mentioned above. Employing the conservative approach of uniform priors, we consider that fossils can provide only minimum ages and that paucity of fossil records hinders the use of more informative priors (similar to other angiosperm studies, e.g., Bouchenak-Khelladi et al., 2014; Boucher et al., 2016). For the secondary calibration, we ran analyses using either a normal (mean age of 67.9 Ma and SD of 5.25 Myr) or a uniform prior (65.63–78.21 Ma), which produced very similar age estimates (**Supplementary Table S2**). Thus, we present results from analyses with uniform priors following Schenk (2016), who demonstrated reduced error in uniform vs. normal priors for secondary calibrations.

Divergence times were inferred using a relaxed uncorrelated lognormal clock and a nucleotide substitution model-averaging method (bModelTest tool; Bouckaert and Drummond, 2017) for each of the two partitions (ITS vs. cpDNA). bModelTest integrates over 203 time-reversible models while simultaneously estimating other parameters (estimates weighted by the support of each model). Independent runs of 200 million generations were combined (after removing up to 22.5% as burn-in) using LogCombiner 2.5.1, and maximum clade credibility (MCC) trees were constructed in TreeAnnotator 2.5.1 (Drummond et al., 2012). Chain convergence was verified in Tracer 1.7.1 [Effective Sample Size (ESS) >200 for all parameters; Rambaut et al., 2018]. The 95% HPD intervals of inferred ages for the five datasets overlapped with each other (**Supplementary Table S2**). Thus, we performed all subsequent analyses on the *Limonium* clade pruned from the "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" MCC trees (see also justification above), unless otherwise specified. BEAST and MrBayes analyses were performed using the CIPRES Science Gateway (Miller et al., 2010).

Biogeographic Analyses

We estimated ancestral ranges for *Limonium* on "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" MCC trees using the R package BioGeoBEARS 1.1.2 (Matzke, 2013; R Core Team, 2018). Considering current species distributions, patterns of endemism, and the floristic regions occupied by the species, we identify nine major biogeographic areas: South Africa, Euro-Mediterranean, North Africa, Irano-Turanian, Macaronesia, East Asia-Australia, Circumboreal, America, and Arabia-NE Africa.

Taxa were coded as present or absent in these areas (**Figure 1** and **Supplementary Table S3**). We tested DEC (Dispersal-Extinction-Cladogenesis, Ree and Smith, 2008) and DIVA-like models (Dispersal-Vicariance Analysis, Ronquist, 1997). We chose not to use the BayArea-like model (Landis et al., 2013) since it does not allow for vicariance or "subset-within-area" speciation (allowed in DIVA-like and/or DEC models), which are both plausible considering species distributions and the broad areas defined in our study.

Time-stratified analyses with dispersal multiplier matrices were also employed to account for differences in dispersal probabilities due to changes in distances and connectivities between areas over time (**Supplementary Table S4**). In all biogeographic analyses, Macaronesia was allowed as an ancestral range only after 27 Ma, corresponding to the age of the oldest extant volcanic islands, i.e., the Selvagen Islands (Borges et al., 2008; Carine et al., 2010; Triantis et al., 2010; Fernández-Palacios et al., 2011). Considering the uncertainty surrounding the reconstruction of past sea levels (Miller et al., 2005; Rohling et al., 2014), we chose the age of 27 Ma as the most widely accepted, hence conservative estimate for the oldest emerged island in Macaronesia. Unlikely disjunct ranges (i.e., non-neighboring, geographically unconnected areas over the inferred existence of *Limonium*, starting ca. 40 Ma) were removed from the list of possible states. All models were also run with *j* parameter, which accounts for founder-event speciation. However, due to criticism by Ree and Sanmartín (2018) concerning "degenerate" inferences of +*j* models and inappropriate statistical comparisons of models with and without *j*, and considering that our inferences with and without *j* parameter closely match each other, we present only results of models without *j* and compare their likelihoods using AIC and AICc. We additionally estimated the type and number of biogeographical events using Biogeographical Stochastic Mapping (BSM; Matzke, 2016; Dupin et al., 2017) in BioGeoBEARS 1.1.2, providing the best-fitting model for each MCC tree and summarizing event frequencies over 200 discrete stochastic maps.

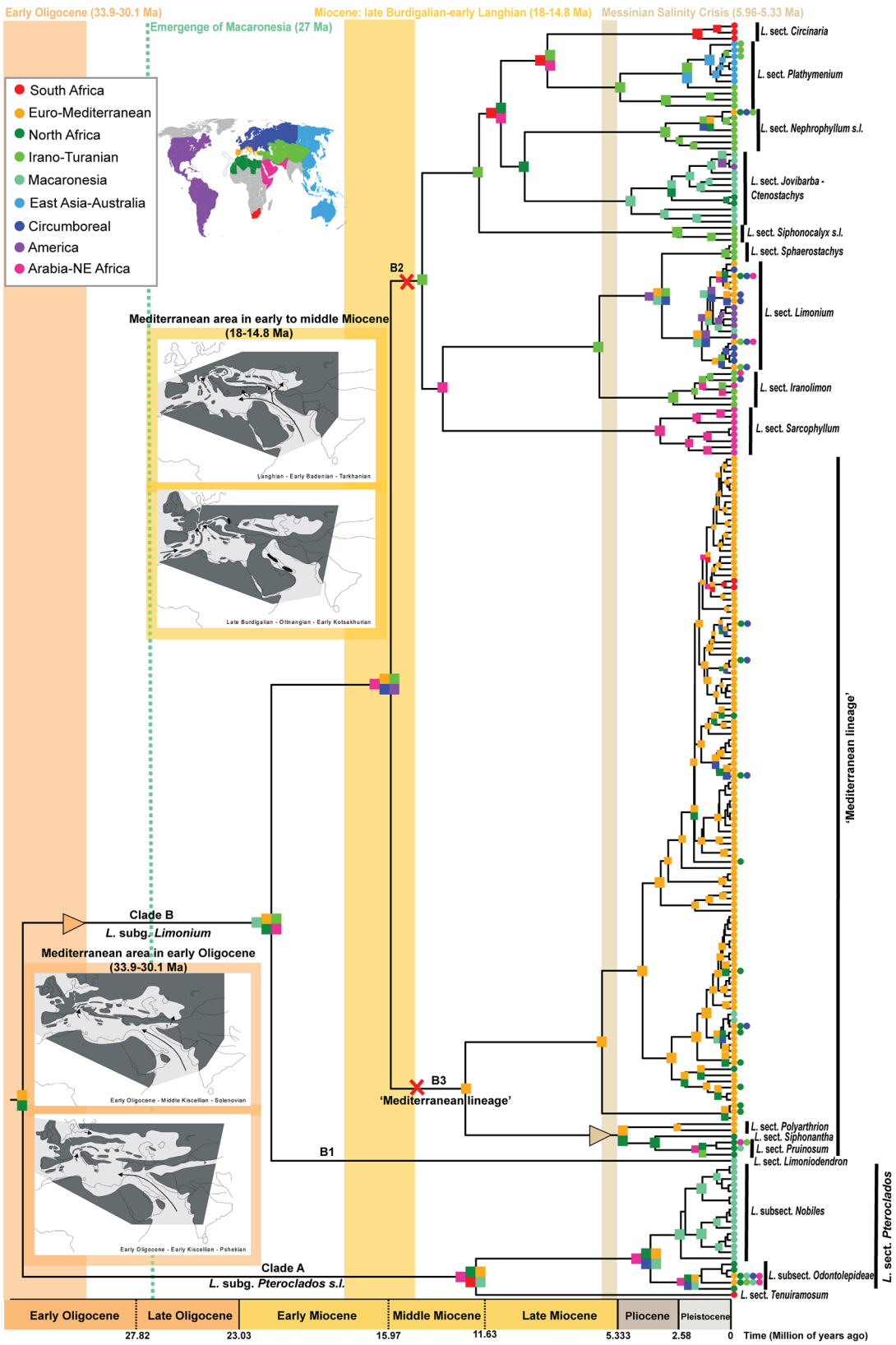
At a finer biogeographic scale, we investigated the origins of island endemics in Macaronesia and analyzed a clade (*Jovibarba-Ctenostachys* clade) comprising endemics in the Canaries, Cape Verde, Morocco, and Hispaniola in the Caribbean (four coded areas). Since the *Jovibarba-Ctenostachys* clade shows topological differences between datasets, we pruned it from cpDNA, ITS and "Supermatrix" MCC trees, ran DEC, DIVA-like, and BayArea-like models and compared results for all three trees. In contrast to the genus-wide biogeographic analyses, we included also the BayArea-like model for the biogeographic analyses of the *Jovibarba-Ctenostachys* clade because the absence of vicariance and "subset-within-area" speciation of this model was not considered a limitation for a lineage consisting exclusively of endemics, most of which are insular. Their biogeographic history could only be explained by dispersal, extinction, and sympatric speciation, all of which are accounted for in the BayArea-like model.

Diversification Analyses

We estimated speciation and extinction rate dynamics of *Limonium* applying recently developed methods to both "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" MCC trees. First, we tested for shifts in diversification rates along tree branches and through time, using Bayesian Analysis of Macroevolutionary Mixture (BAMM 2.5; Rabosky et al., 2014). Sampling fractions per *Limonium* clade/section (following the classification of Koutroumpa et al., 2018: Supplementary Table S3) were provided in order to account for missing species (**Supplementary Table S5**). The R package BAMMtools 2.1.6 (Rabosky et al., 2014) was used to choose appropriate priors for BAMM analyses, analyze, and visualize results. BAMM was run with four independent chains of 100 million generations each and convergence was confirmed with ESS >200 for all parameters. The absence of recent, comprehensive monographic treatments for *Limonium* prevents us from performing analyses that account for potential discrepancies among different taxonomic treatments (e.g., Faurby et al., 2016). Moreover, taxonomic uncertainties could especially affect the apomictic species of the "Mediterranean lineage" (Haveman, 2013; Majeský et al., 2017). Acknowledging that taxonomic uncertainties could affect estimates of diversification rate dynamics (Faurby et al., 2016; Fernández-Mazuecos et al., 2019), we ran preliminary sensitivity analyses in BAMM, where we arbitrarily assumed that the number of species for the "Mediterranean lineage" is only half of the currently accepted number of "Mediterranean lineage" species (i.e., ca. 200 vs. 400 species), and thus assigned a higher sampling fraction to this lineage to account for missing species in analyses of both MCC trees (i.e., ½ instead of ¼; the former being the fraction used in the preliminary analyses). The signal in our data was so strong that even in these preliminary analyses that underestimated the number of species in the "Mediterranean lineage," a shift in rates was recovered for this lineage in both MCC trees, despite the extreme and rather unrealistic taxonomic scenario.

Second, we tested for episodic tree-wide rate shifts caused, for example, by a mass extinction event, such as the one hypothesized for the temporal gap of diversification from the MSC to the onset of the Mediterranean climate in other Mediterranean plant lineages. We used the CoMET algorithm implemented in the R package TESS 2.1 (Höhna et al., 2015) to test for signals of mass extinction events and for shifts in speciation and extinction rates over time. Changes in speciation and extinction are modeled to occur at discrete time points and affect all clades in a tree simultaneously (episodic birth-death). We ran models assuming one or two rate shifts and allowing mass extinction events. A global sampling fraction of *Limonium* was provided to account for sampling probability of extant taxa ("Supermatrix-ITS-like": 0.35 and "Supermatrix-cpDNA-like": 0.36). We also used the automatic stopping functionality and specified a minimum ESS of 1,000 to allow MCMC simulations to reach convergence.

Third, we fitted different diversification models with constant or varying speciation and/or extinction rates through time and in relation to paleoenvironmental variables (Morlon et al., 2010; Condamine et al., 2013) using the R package RPANDA 1.6 (Morlon et al., 2016). Analyses were repeated using different initial parameters and choosing the solution with the highest likelihood for each model. Models were compared using AICc


Early Oligocene (33.9-30.1 Ma)
Miocene: late Burdigalian-early Langhian (18-14.8 Ma)
Messinian Salinity Crisis (5.96-5.33 Ma)
Emergenge of Macaronesia (27 Ma)
South Africa
Euro-Mediterranean
North Africa
Irano-Turanian
Macaronesia
East Asia-Australia
Circumboreal
America
Arabia-NE Africa
Mediterranean area in early to middle Miocene (18-14.8 Ma)
Langhian - Early Badenian - Tarkhanian
Late Burdigalian - Ottnangian - Early Kotsakhurian
B2
B3
'Mediterranean lineage'
B1
Clade B
L. subg. Limonium
Mediterranean area in early Oligocene (33.9-30.1 Ma)
Early Oligocene - Middle Kiscellian - Solenovian
Early Oligocene - Early Kiscellian - Pshekian
Clade A
L. subg. Pteroclados s.l.
L. sect. Circinaria
L. sect. Plathymenium
L. sect. Nephrophyllum s.l.
L. sect. Jovibarba - Ctenostachys
L. sect. Siphonocalyx s.l.
L. sect. Sphaerostachys
L. sect. Limonium
L. sect. Iranolimon
L. sect. Sarcophyllum
'Mediterranean lineage'
L. sect. Polyarthrion
L. sect. Siphonantha
L. sect. Pruinosum
L. sect. Limoniodendron
L. subsect. Nobiles
L. subsect. Odontolepideae
L. sect. Tenuiramosum
L. sect. Pteroclados
Early Oligocene
Late Oligocene
Early Miocene
Middle Miocene
Late Miocene
Pliocene
Pleistocene
27.82
23.03
15.97
11.63
5.333
2.58
0
Time (Million of years ago)

(Continued)

FIGURE 1 | Time calibrated phylogeny and biogeography of *Limonium* based on "Supermatrix-ITS-like" dataset. The nine areas used for biogeographic inference are color coded as in the world map and corresponding list in the upper left inset. Colored squares on nodes indicate the most likely ancestral areas; due to space limitations, some nodes have no colored squares: in these cases, ranges are the same as in their respective ancestral nodes. Triangles and X's along branches indicate range expansions and contractions, respectively, that occurred when geological events significantly impacted *Limonium*'s distribution and diversification (i.e., in the early Oligocene, early to middle Miocene and during the Messinian). These periods are marked by vertical color bands. In two of these periods, namely in the early Oligocene and the early to middle Miocene, paleomaps show the Mediterranean geographic configuration in each time frame (paleomaps are from the article of Rögl, 1999 published in Geologica Carpathica journal under the CC BY-NC-ND 4.0 license). Uncertainties in ancestral range estimates are in **Supplementary Figure S4**.

and a multimodel inference approach was employed for parameter estimates of time-dependent analyses (Burnham and Anderson, 2002). According to Morlon et al. (2011), rate heterogeneity can affect diversification rate estimates if such heterogeneity is not accounted for. Thus, apart from fitting models to the entire tree, RPANDA allowed us to test and account for rate heterogeneity by fitting models to subclades for which shifts were inferred in BAMM analyses (see section Results) and to the remaining phylogeny after removing these subclades. Models with or without rate shifts were then compared using the AICc. Global or clade-specific sampling fractions were assigned in analyses of either the entire phylogeny or the subclades defined above (in heterogeneity analyses), respectively, to account for incomplete taxon sampling. Paleoenvironmental variables used in RPANDA analyses as potential correlates of diversification rates were (i) global temperature throughout the Cenozoic (Zachos et al., 2008), fitted to the entire tree or all parts of the tree in heterogeneity analyses and (ii) past sea-level estimates (Miller et al., 2005; Rohling et al., 2014), fitted to Mediterranean subclade(s) that comprise coastal species (heterogeneity analyses; see environmental curves in **Supplementary Figure S3**). Specifically, the recent sea-level reconstructions of Rohling et al. (2014) in the Mediterranean are limited to the past 5.3 Myr and thus were fitted to the temporally corresponding part of the Mediterranean subclade, requiring that a few early-diverging taxa were removed from the subclade.

State-Dependent Diversification

We tested for the effect of reproductive system on diversification rates using several state-dependent speciation and extinction (SSE) models. We assembled available information on reproductive strategies, coding 175 *Limonium* taxa (out of the 216 taxa included in the trees) as sexual or apomictic based on the literature (see **Supplementary Table S6**). Tips with missing reproductive information were dropped from "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" MCC trees before analyses.

First, we used BiSSE (Maddison et al., 2007) implemented in the R package diversitree 0.9–11 (FitzJohn et al., 2009; FitzJohn, 2012) to evaluate 10 models in which speciation, extinction, and transition rates varied or remained equal between states; two of these models had zero extinction (pure-birth models). Since we lack information on reproductive system for many *Limonium* species, we conducted three analyses with three different sampling fractions for each state (sexual vs. apomictic) to account for missing species in the phylogeny and tested the robustness of the results. In the first analysis, we assumed that the same proportion of sexuals vs. apomicts sampled in the entire *Limonium* phylogeny also applies to unsampled taxa (i.e., global sampling fraction). In the other two analyses, we incorporated prior knowledge that all described apomicts are included in the "Mediterranean lineage" (while species in all other lineages are sexual) and assumed either 50–50 or 40–60% of sexuals vs. apomicts in this lineage; we also accounted for uneven sampling in major lineages of *Limonium* by using the "make.bisse.uneven" function. We consider the latter sampling scenario as more realistic, since the predominance of apomicts over sexuals in the Mediterranean has been documented by Brullo and Erben (2016). Finally, we accounted for confidence intervals of parameter estimates by running a Bayesian MCMC with 50,000 steps using an exponential prior (following FitzJohn, 2012) and the best-fitting model for each analysis. Chain convergence was checked using the R package coda (Plummer et al., 2006) and results were plotted after removing 5,000 steps as burn-in.

Second, due to criticism of BiSSE concerning incorrect assignments of rate differences for neutral observed states in phylogenies, where rate heterogeneity is caused by other factors (Maddison and FitzJohn, 2015; Rabosky and Goldberg, 2015), we applied the Hidden SSE (HiSSE; Beaulieu and O'Meara, 2016) model to specifically account for other unmeasured factors impacting diversification rates. We tested seven models (model details in Beaulieu and O'Meara, 2016): a full BiSSE-like model, two HiSSE models with equal or varying transition rates (one rate for transitions among hidden states and two rates for transitions between sexual reproduction and apomixis), and four character-independent diversification (CID) models assuming that diversification is trait-independent but not constant across the tree (CID-2 and CID-4 models match complexity of BiSSE and HiSSE models, respectively). We also used the same three sampling scenarios for sexuals vs. apomictic states explained above, but without assigning uneven lineage-specific sampling probabilities (not implemented in HiSSE). Furthermore, we inferred the model-averaged rate estimates for all tips in both MCC trees and for all analyses to examine whether diversification rates varied between sexuals and apomicts. In BiSSE and HiSSE analyses, the tested models were compared with AIC.

Third, we used FiSSE (Rabosky and Goldberg, 2017) as a nonparametric test for reproductive system-dependent diversification in *Limonium*. This test is used for additional validation of results obtained from the other two state-dependent models (Rabosky and Goldberg, 2017). FiSSE shows the lowest "false positive" rates compared to BiSSE and HiSSE, but its power in detecting state-dependent diversification is much lower than BiSSE for trees with <300 tips (Rabosky and Goldberg, 2017). Although our trees are relatively small for the power limits of FiSSE, we expect that a significant state-dependent

diversification, if detected, would show that the signal of our data is strong enough to overcome FiSSE's power limitations. Conversely, finding non-significant state-dependent diversification does not allow the possibility that this result is due to FiSSE's power limitations to be excluded, especially when the other two methods support state-dependent diversification. Here, we did not account for incomplete taxon sampling because this is not implemented in FiSSE.

Since most *Limonium* species occur in the Euro-Mediterranean area, we also tested the effect of this range on diversification rates using the fast Hidden Geographical SSE model (fGeoHiSSE; Caetano et al., 2018). *Limonium* species were coded as present only in the Euro-Mediterranean area or only in other areas or widespread in both the Euro-Mediterranean and other areas. Missing species were accommodated by providing sampling fractions for each of these three range categories. We fitted four models: two range-independent and two range-dependent diversification models, each either with or without a hidden area. Models were compared *via* AIC.

At shallower phylogenetic levels, we pruned the *Limonium* subgen. *Pteroclados s.l.* lineage (clade A in **Figure 1**; see also Koutroumpa et al., 2018) from "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" MCC trees, and tested whether woodiness (i.e., suffruticose habit) of Canarian *Limonium* endemics in *Nobiles* clade (Bramwell and Bramwell, 1974, 1990; Mesa et al., 2001; Marrero and Almeida, 2003; Kunkel, 2012; Jiménez et al., 2017) had an impact on diversification rates. In this fully sampled *Pteroclados s.l.* lineage, all other clades apart from *Nobiles* consist of herbaceous species. We used BiSSE and HiSSE analyses, with models described above, to test for habit-dependent diversification in the *Pteroclados s.l.* lineage. Simulation studies have suggested that while it is feasible to use SSE methods for analyses of small-size clades such as *Pteroclados s.l.* (e.g., Gamisch, 2016), their power in detecting state-dependent diversification depends on the strength of speciation rate asymmetry in the tree. If state-dependent diversification were detected, this would mean that speciation rate asymmetry between woody and herbaceous taxa is sufficiently high (i.e., >2.5-fold) to be detected despite small clade size, as demonstrated by simulations (Gamisch, 2016; Kodandaramaiah and Murali, 2018). Conversely, if state-independent diversification were supported, this would not automatically exclude the existence of moderate levels of speciation rate asymmetry (i.e., <2.5-fold), but might simply mean that such levels are insufficient to be recovered by the SSE methods in small clades. We also tested whether woodiness is a derived state for the *Pteroclados s.l.* lineage employing model-averaged ancestral state estimations implemented in HiSSE package.

RESULTS

Molecular Phylogenies, Divergence Time, and Biogeographic Estimates

Phylogenies inferred with ML and BI for each dataset recovered very similar topologies. Only a small number of conflicts were found between cpDNA and ITS trees, all located within *Limonium* (i.e., "rogue clades": subclades within "Mediterranean lineage," *Circinaria* clade and *Jovibarba-Ctenostachys* clade, and six "rogue taxa"; see **Supplementary Figures S1**, **S2** and Malekmohammadi et al., 2017; Koutroumpa et al., 2018). The 50% majority-rule BI trees for all three Supermatrices showed better resolution ("Supermatrix": 181/300, "Supermatrix-ITS-like": 177/293, and "Supermatrix-cpDNA-like": 176/300 nodes resolved, i.e., have posterior probabilities ≥0.5) than ITS and cpDNA trees (154/259 and 142/298 nodes resolved, respectively). Our results strongly support monophyly for Plumbaginaceae subfamilies (Plumbaginoideae and Limonioideae), tribes (Aegialitideae and Limonieae) and genera (except for *Plumbago* and *Acantholimon*; both non-monophyletic), and for major clades of *Limonium* corresponding to infrageneric subdivisions (except for section *Schizhymenium*). Hereafter, results for "Supermatrix-ITS-like" tree and "Supermatrix-cpDNA-like" tree will be reported first and second, respectively, separated by a vertical bar, unless otherwise specified.

Divergence time estimates were almost identical between "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" MCC trees (detailed age estimates and HPDs in **Table 1**; **Supplementary Table S2** and **Supplementary Figures S1**, **S2**). The origin of Plumbaginaceae is placed between the Upper Cretaceous and the early Eocene (95% HPD: 52–77 Ma), with a median age at early Paleocene (ca. 65.7 Ma). The median crown ages for Plumbaginoideae and Limonioideae are ca. 29.5 Ma (95% HPD: ca. 18–43 Ma) and ca. 57 Ma (95% HPD: ca. 43–71 Ma), respectively. Combining divergence time and biogeographic estimates for *Limonium* (DEC without dispersal-multiplier matrices selected as best-fitting model; **Figure 1**; **Supplementary Table S7** and **Supplementary Figures S4**, **S5**), we infer a late Paleogene origin (ca. 33 Ma median crown age; 95% HPD: ca. 24–44 Ma) in the proto-Mediterranean region (i.e., Euro-Mediterranean and North-Africa|North Africa: most likely states, yet with low probability) for the genus. The two subgenera *Pteroclados s.l.* and *Limonium* originated during Miocene, ca. 12 Ma (95% HPD: ca. 6–20 Ma) and ca. 21 Ma (95% HPD: ca. 15–29 Ma), respectively, within widespread regions. A widespread range was also inferred (with low to moderate probabilities) for the most recent common ancestor (MRCA) of clades B2 (comprising mostly non-Mediterranean species) and B3 ("Mediterranean lineage") in the middle Miocene (ca. 15.5 Ma; 95% HPD: ca. 11–21 Ma). This widespread ancestor gave rise to an endemic Euro-Mediterranean MRCA of the "Mediterranean lineage" (B3 crown-node; ca. 12 Ma) and either an endemic Irano-Turanian MRCA for the mostly non-Mediterranean B2 clade according to "Supermatrix-ITS-like" tree (ca. 14.4 Ma; **Figure 1**; **Supplementary Figure S4**) or a widespread MRCA for the B2 clade in the same ancestral widespread range according to "Supermatrix-cpDNA-like" tree (ca. 14 Ma; **Supplementary Figure S5**). In clade B2, dispersals, vicariance events and/or range contractions from middle Miocene through Plio-Pleistocene produced ancestors with endemic ranges in the Irano-Turanian, Arabia-NE Africa, East Asia-Australia, South Africa, America, and Macaronesia. Early divergence events of the "Mediterranean lineage" (clade B3) in late Miocene-early Pliocene were accompanied by dispersal from Euro-Mediterranean to North Africa, while the origin of the larger subclade of B3 (ca. 6 Ma

TABLE 1 | Divergence time estimates for major clades in the "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" maximum clade credibility (MCC) trees.

Crown-nodes	Median ages (95% HPD)	
	"Supermatrix-ITS-like"	"Supermatrix-cpDNA-like"
Plumbaginaceae-Polygonaceae split (Root)	75.69 (69.03–78.21)	75.78 (69.18–78.21)
Plumbaginaceae	65.68 (52.01–77.03)	65.68 (52.07–77.2)
Plumbaginoideae	29.42 (18.71–42.76)	29.6 (18.36–43.23)
Limonioideae	56.92 (42.83–70.6)	57.2 (42.56–70.65)
Limonieae	40.35 (29.88–51.58)	40.1 (29.65–51.22)
Limonium	33.07 (23.68–44.23)	33.18 (23.39–43.76)
Limonium subg. *Pteroclados s.l.*	11.89 (6.25–19.97)	11.82 (6.2–19.45)
Limonium sect. *Pteroclados*	3.74 (2.11–5.86)	3.71 (2.11–5.82)
Limonium sect. *Pteroclados* subsect. *Nobiles*	2.32 (1.31–362)	2.29 (1.29–3.54)
Limonium subg. *Limonium*	21.46 (14.79–29.74)	21.12 (14.69–28.96)
Limonium clades B2-B3 split*	15.87 (11.11–21.52)	15.56 (11.12–21.12)
Limonium clade B2*	14.41 (10.03–19.57)	14.12 (9.89–19.05)
Limonium clade B3: "Mediterranean lineage"*	12.39 (8.09–17.33)	12.23 (8.05–16.93)
Larger subclade of "Mediterranean lineage"**	6 (3.72–8.99)	5.68 (3.41–8.46)

Median ages and 95% Highest Posterior Density (HPD) intervals (in parentheses) for the crown-nodes of clades are in million years (Ma). Dating results are from analyses with uniform priors assigned to Plumbaginaceae stem-node (secondary calibration).
**Coding of clades within Limonium follow Koutroumpa et al. (2018).*
***The subclade has slightly different species composition between the two Supermatrices, due to the incongruent position of some Aegean species (for details see Koutroumpa et al., 2018).*

median crown age) coincides with the MSC. Extensive speciation within the Euro-Mediterranean area during the Pleistocene was inferred for clade B3, with only little dispersal from Euro-Mediterranean to Macaronesia, North-Africa, Circumboreal and South Africa. Some of these biogeographic results should be interpreted with caution given the limited phylogenetic resolution within the "Mediterranean lineage." Detailed nodal biogeographic reconstructions and uncertainties in ancestral range estimates for both trees are in **Supplementary Figures S4**, **S5**.

BSM analyses estimated 191|196 (SD: ±8| ± 7.79) within-area speciation events, 67|63 (±2.71| ± 2.06) range expansion events and 16|18 (±2.18| ± 1.71) vicariance events. The large number of sympatric speciation events (ca. 70% of all biogeographic events) was expected, given the large size of the defined areas. Most dispersal events occurred from Euro-Mediterranean to North Africa (ca. 14) and from Euro-Mediterranean to Circumboreal (ca. 5), while all other combinations of areas had <3.5 events (**Figure 2**; **Supplementary Table S8**). Overall, the major source area for dispersals was the Euro-Mediterranean (ca. 40% of estimated dispersal events), followed by North Africa (ca. 15%) and Irano-Turanian (ca. 14%), while the most common sink was North Africa (ca. 28%), followed by Circumboreal (16%). The longest-distance dispersals occurred from Macaronesia to the Americas, and from Arabia-NE Africa and North Africa (or Euro-Mediterranean region) to South Africa (**Supplementary Table S8**). Movements between areas were largely asymmetrical for most pairs of areas. For example, dispersals from Euro-Mediterranean to North Africa were ca. eight times more common than those in the opposite direction.

Focusing on island biogeography in Macaronesia, origins of island endemics included both neighboring continents and archipelagos (**Supplementary Figures S4**, **S5**; **Figure 3**). Macaronesian endemics were placed in five distinct clades, indicating at least five independent long distance dispersal (LDD) events (**Supplementary Figures S4**, **S5**). The *Nobiles* clade, comprising 16 Canarian endemics, originated ca. 2.3 Ma, while colonization of the Macaronesian area predated the origin of this clade and occurred from a widespread ancestor (Euro-Mediterranean and North Africa|Euro-Mediterranean, North and South Africa, Circumboreal and Arabia-NE Africa) between late Oligocene and early Pliocene. The Canarian endemic *L. dendroides* (clade B1) represents an early diverged lineage (ca. 21 Ma) that originated *via* dispersal from neighboring areas (i.e., North Africa or Euro-Mediterranean). The same dispersal pattern holds true for the closely related *L. bollei* and *L. lowei* (nested within "Mediterranean lineage") endemic to Canaries and Madeira, respectively, which diversified in the late Pleistocene. The Azorean endemic (*L. eduardi-diasii*) originated in the late Pleistocene in a clade comprising American endemics, either *via* a recent colonization from the Americas ("Supermatrix-cpDNA-like") or an earlier colonization (late Miocene-Pliocene) following range expansion from Irano-Turanian toward Euro-Mediterranean, Circumboreal, Macaronesia, and the Americas ("Supermatrix-ITS-like"). Fine-scale biogeographic estimations for the *Jovibarba-Ctenostachys* clade (DIVA-like selected as best-fitting model; **Figure 3**; **Supplementary Table S9**) that diversified during Plio-Pleistocene revealed two independent colonizations of Cape Verde (one probably from North Africa and the other from Canaries) and a LDD event from Cape Verde to Hispaniola. In this clade, the biogeographic origin of Canarian endemics is either from North Africa or Cape Verde and for Moroccan endemics either from Canaries or Cape Verde (inferences varied due to unresolved relationships). BSM reconstructs an overall total of seven dispersal events to Macaronesia (giving rise to native Macaronesian species as well as endemics) including two from North Africa and two from the Euro-Mediterranean. Furthermore, Macaronesia is inferred to be the source pool for four colonizations of continental areas, including one to the Americas and one to North Africa (**Figure 2**; **Supplementary Table S8**).

Diversification Dynamics

BAMM analyses strongly reject constant diversification along clades and through time for *Limonium* (**Supplementary Figure S6**). Instead, they reveal significant shift(s) toward higher net diversification rates within the "Mediterranean lineage" (**Figure 4**; **Supplementary Figure S7**). In both trees, there is a shift close to the origin of the larger subclade of this lineage (crown age of subclade: ca. 6 Ma), while an additional shift within the

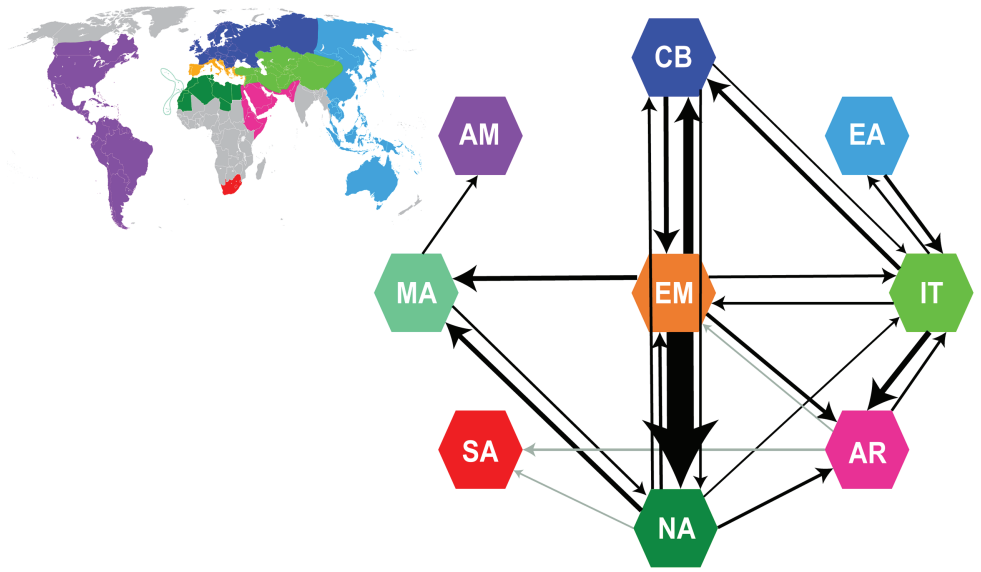


FIGURE 2 | Summary of dispersal events from Biogeographical Stochastic Mapping (BSM) analyses. Arrows and their thickness represent the direction and frequency of dispersals between areas, respectively. Only event counts with a mean of ≥0.9 in both trees ("Supermatrix-ITS-like" and "Supermatrix-cpDNA-like") are depicted as arrows (detailed results of BSM analyses in **Supplementary Table S8**). Gray arrows represent dispersal events with a mean ≥0.9 in only one of the two trees. The nine areas used in BSM analyses are color coded as in the world map in the upper left inset (same color coding of areas as in **Figure 1**). AM, America; AR, Arabia-NE Africa; CB, Circumboreal; EA, East Asia-Australia; EM, Euro-Mediterranean; IT, Irano-Turanian; MA, Macaronesia; NA, North Africa; SA, South Africa.

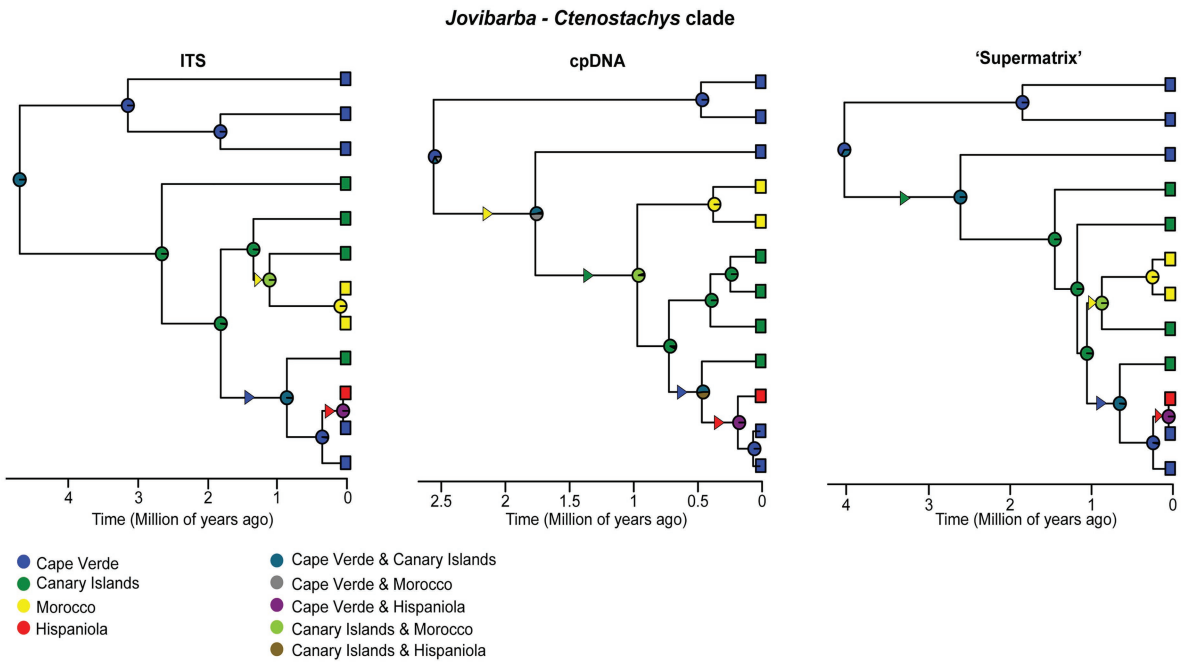


FIGURE 3 | Ancestral range estimates for the *Jovibarba – Ctenostachys* clade of *Limonium* cropped from the ITS, cpDNA and "Supermatrix" MCC trees. The four areas used for biogeographic analyses and their combined ranges are color coded as in the left and right column, respectively, of the inset at the bottom. Pie charts on the nodes of trees represent relative probabilities of the inferred ancestral ranges. Triangles along branches indicate range expansions and their colors indicate the colonized area.

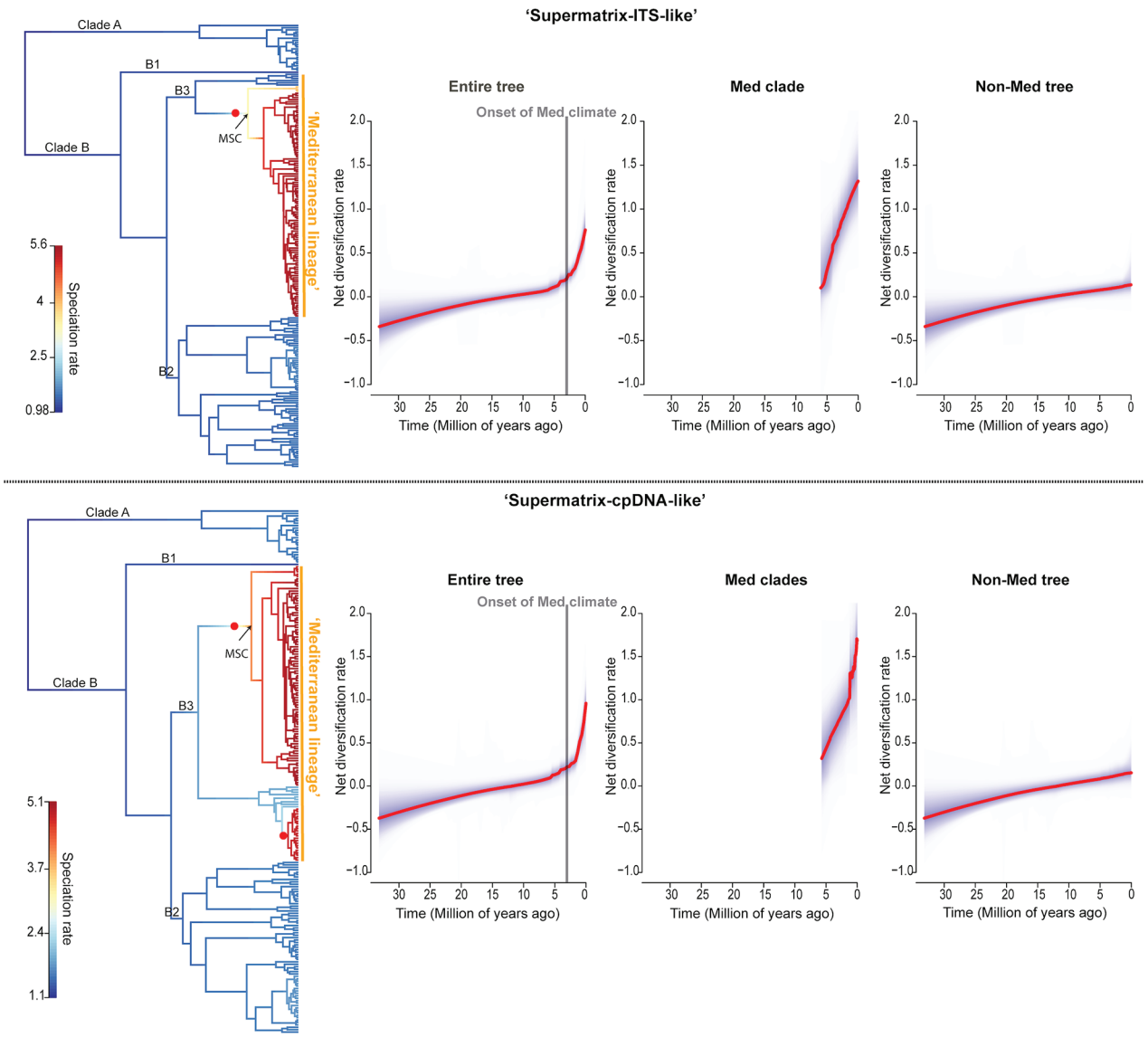


FIGURE 4 | Diversification dynamics of *Limonium* across clades and through time. Results of BAMM analyses for "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" trees are presented in the top and bottom row, respectively. The first figure at the left of each row shows rate heterogeneity across clades for *Limonium* (phylorate plot) and the other three figures show the net diversification rates through time (rate-through-time plots). In the phylorate plots, branches are colored from blue to red to indicate low to high speciation rates, respectively, and the best rate-shift configuration is denoted by a red circle (the 95% credible set of rate shift configurations are in **Supplementary Figure S7**). The rate-through-time plots show the net diversification rates for the entire *Limonium* phylogeny ("Entire tree"), the Mediterranean clade(s) for which a shift was detected by BAMM ("Med clade(s)"), and the remaining *Limonium* phylogeny after removing the Mediterranean clade(s) ("Non-Med tree"). Significant shift(s) in diversification rates were found within the "Mediterranean lineage" of *Limonium*, while the steep increase in net diversification rates over the past 2–3 Myr is concomitant with the Plio-Pleistocene climatic oscillations and the establishment of the Mediterranean climate. MSC, Messinian Salinity Crisis.

smaller subclade for "Supermatrix-cpDNA-like" tree is explained by the conflicting topologies obtained from plastid and nuclear genomes for some Aegean species. Further evidence supports an 18–20-fold increase in net diversification rate for Mediterranean subclades compared to background rate (1.07 vs. 0.06|1.18 vs. 0.06). The rate-through-time plots show an increase in net diversification rates starting ca. 6 Ma that intensifies during the past 2–3 Myr (**Figure 4**).

TESS analyses found no significant mass extinction event or episodic change of speciation or extinction rates that could concurrently affect all clades of *Limonium* (Bayes Factors: 2lnBF<6; **Supplementary Figure S8**). RPANDA analyses of rate heterogeneity further corroborated results from BAMM. Models with rate shift(s), allowing distinct patterns of rate variation for the "Mediterranean lineage" subclade(s), were strongly supported over models assuming no rate shift for *Limonium* (ΔAICc>30; **Table 2**).

TABLE 2 | Diversification rate heterogeneity analyses in *Limonium* for both "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" trees (RPANDA analyses).

A

"Supermatrix-ITS-like" tree	No rate shift	Heterogeneity analysis One rate shift (Med subclade)	ΔAICc
Log-likelihood	−257.734	−241.163	
AICc	523.664	492.679	30.985
"Supermatrix-cpDNA-like" tree	**No rate shift**	**Heterogeneity analysis Two rate shifts (Med subclades)**	**ΔAICc**
Log-likelihood	−247.652	−227.627	
AICc	503.493	468.141	35.352

B

"Supermatrix-ITS-like" tree			
	Entire tree	**Med subclade**	**Non-Med tree**
Speciation rate	4.86	6.91	3.07
Extinction rate	4.75	5.70	3.03
Net diversification rate	0.11	1.21	0.04

"Supermatrix-cpDNA-like" tree				
	Entire tree	**Med1 subclade**	**Med2 subclade**	**Non-Med tree**
Speciation rate	5.17	5.46	9.18	3.27
Extinction rate	5.08	4.35	3.23	3.23
Net diversification rate	0.09	1.11	5.95	0.04

(A) *Likelihoods and AIC comparisons between a model assuming no shifts across the tree ("No rate shift") and a model assuming one and two rate shifts for "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" tree, respectively ("Heterogeneity analysis");* ***(B)*** *Multimodel averaged speciation, extinction and net diversification rate estimates of time-dependent diversification analyses on the entire tree, the Mediterranean subclade(s) ("Med subclade" and "Med1 subclade": the larger Mediterranean subclade in "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" trees, respectively, for which a shift was detected by BAMM; "Med2 subclade": the small Mediterranean subclade for which a shift was detected by BAMM in "Supermatrix-cpDNA-like" tree) and the tree after removing the Mediterranean subclade(s) ("Non-Med tree").*

We inferred a ca. 30-fold higher net diversification rate for the larger "Mediterranean lineage" subclade compared to the background rate (1.21 vs. 0.04|1.11 vs. 0.04; **Table 2**). For this subclade, a model of constant speciation and exponentially decreasing extinction over time ("Supermatrix-ITS-like") or a pure-birth model with exponentially increasing speciation over time ("Supermatrix-cpDNA-like") had the lowest AICc values (**Supplementary Table S10**). In the slightly reduced version of this subclade (with Rohling et al.'s, 2014 sea-level data fitted), the highest support was assigned to a model with constant speciation and extinction positively correlated to past sea-level for "Supermatrix-ITS-like," while for "Supermatrix-cpDNA-like" that the model was the second best-fitting after a pure-birth model (**Supplementary Table S10**). For the backbone tree [i.e., remaining tree after removing "Mediterranean lineage" subclade(s)], a constant birth-death model received the lowest AICc values, while for the entire tree a model with exponentially varying speciation and extinction, both positively correlated to paleo-temperature, was best-fitting. In all analyses, we observed model uncertainty, with several alternative models explaining diversification dynamics (i.e., ΔAICc<2; **Supplementary Table S10**). Paleoenvironment-dependent models always received some support, with past temperature and sea-level having similar impacts on speciation and extinction rates.

BiSSE, HiSSE, and FiSSE analyses consistently supported the reproductive strategy-dependent diversification for *Limonium*, with apomixis leading to higher speciation and net diversification rates. In BiSSE, a model with varying speciation and transition rates and equal extinction rates between sexual reproduction and apomixis received the highest support in all analyses (i.e., for both MCC trees and different sampling scenarios), except for one analysis that supported a full BiSSE model (**Supplementary Table S11**). In most BiSSE analyses, up to two additional models received some statistical support (i.e., ΔAIC<2), yet all these alternative models supported the state-dependent diversification. Bayesian parameter estimations revealed higher speciation and net diversification rates for apomixis, and higher transition rates from apomixis to sexual reproduction than in the opposite direction (**Figure 5**). Additionally, higher extinction rate for apomixis was observed in the analysis supporting a full BiSSE model (**Figure 5**; **Supplementary Table S11**). In HiSSE, either a BiSSE-like model or a HiSSE model with three varying transition rates (see Section Materials and Methods) was supported in all analyses (**Supplementary Table S12**). Model-averaged parameter estimates for species in our trees recovered higher mean speciation, extinction, and net diversification rates for apomicts across all trees and analyses (**Table 3**). Given that a precise estimation of extinction from phylogenies of extant taxa is challenging (Pyron and Burbrink, 2013; but see Morlon et al., 2011 showing inference of realistic extinction rates), we focus our discussion on the general patterns rather than the estimates of extinction rates *per se* that need to be taken with caution. Complementary FiSSE analyses reported a ca. 2-fold higher speciation rate for apomixis (2.2|1.9 vs. 1.2|1.3), with the difference in rates being either significant (p = 0.016; "Supermatrix-ITS-like" tree) or only marginally non-significant (p = 0.0599; "Supermatrix-ITS-like" tree). Conversely, range-dependent diversification analysis (fGeoHiSSE) found no effect of presence in the Euro-Mediterranean area on diversification rates. Instead, analyses supported a model with diversification rate shifts across the tree, yet independent of the ranges (**Supplementary Table S13**).

State-dependent diversification analyses in *Pteroclados s.l.* clade showed no effect of habit (herbaceous habit vs. woodiness)

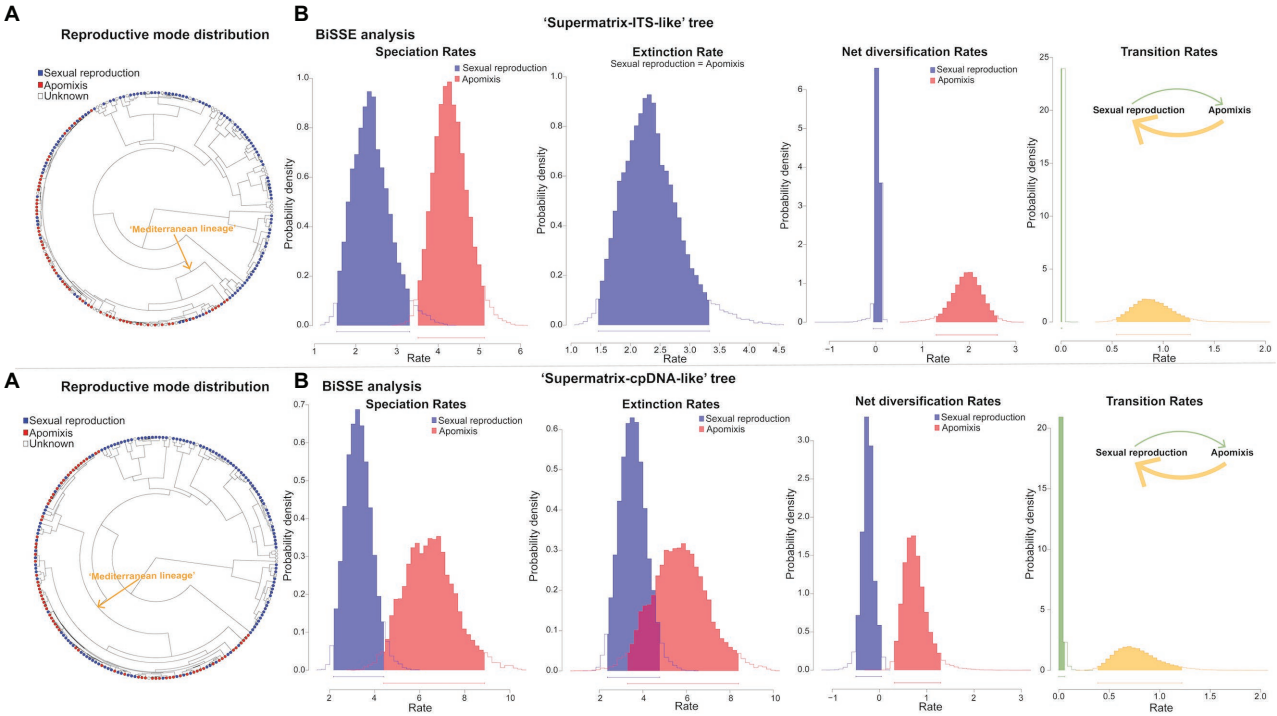


FIGURE 5 | Diversification rate analyses in *Limonium* dependent upon apomictic vs. sexual reproduction. Results for "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" trees are presented in the top and bottom rows, respectively. **(A)** Phylogenetic distribution of sexual reproduction and apomixis. **(B)** Bayesian parameter inferences from BiSSE analyses employing the best fitting model under the sampling scenario that assumes a 40–60% of sexuals vs. apomicts in the "Mediterranean lineage." For "Supermatrix-ITS-like" tree, a model with equal extinction rates was the best (see the single parameter estimate for both reproductive modes in the Extinction Rate graph in the top row) and for "Supermatrix-cpDNA-like" tree, a model with all rates varying was the best (**Supplementary Table S11**). The results show that speciation and net diversification rates are higher for apomictic than sexual taxa in both trees, extinction rates are also higher for apomicts but only in "Supermatrix-cpDNA-like" tree, while transitions from apomixis to sexual reproduction were more common that in the opposite direction.

on diversification rates (**Supplementary Tables S14**, **S15**). BiSSE analyses reported the lowest AIC for the constant rates model, although some state-dependent models with varying rates between woodiness and herbaceousness received partial support (i.e., ΔAIC<2). This result could be attributed to state-independent rate heterogeneity within *Pteroclados s.l.* clade, which BiSSE cannot test/account for, as revealed by HiSSE (i.e., the highest support for the CID-2 model with equal transition rates; **Figure 6**). However, caution should be taken when interpreting these results due to known limitations of SSE models in analyses of small-size clades such as the *Pteroclados s.l.* (see sections Materials and Methods and Discussion). Model-averaged ancestral state estimations combined with biogeographic results inferred woodiness as a derived state linked to insular life (**Figure 6**).

DISCUSSION

Identifying biotic and abiotic factors that drive the compartmentalization of biodiversity is central to biology (Benton, 2009, 2015). With its large size (ca. 600 species), wide geographic range (six continents), and species richness concentrated in the Mediterranean region (ca. 70% of all species), *Limonium* is a paramount example of uneven taxonomic and geographic plant diversity. Our macroevolutionary study on sea lavenders is one of the few to test the impact of both species-extrinsic and -intrinsic factors on diversification by using carefully selected methods of biogeographic and diversification-rate analyses (Lagomarsino et al., 2016; Condamine et al., 2018; Letsch et al., 2018), thus expanding the knowledge of the tempo and mode of plant diversification. We found significant increases of diversification rates associated with both geo-climatic events (namely, the MSC, onset of Mediterranean climate, and Pleistocene sea-level fluctuations) and intrinsic species traits, such as apomixis. The significant role of apomixis in shaping the Mediterranean radiation of *Limonium* had been previously proposed and is here explicitly tested for the first time. Our study provides new insights into the origins of Mediterranean biodiversity and highlights a significant role and interplay of both biotic and abiotic factors in promoting species diversification.

Geological and Climatic Processes Shaping Species Distributions and Diversification

Our biogeographic and dating analyses of the largest *Limonium* phylogeny to date, combined with paleo-geological, paleo-climatic, and paleo-vegetational evidence (e.g., Rögl, 1999; Böhme, 2003;

TABLE 3 | Mean speciation, extinction, and net diversification rate values from model-averaged estimates at the tips of each tree ("Supermatrix-ITS-like" and "Supermatrix-cpDNA-like"), as inferred from ancestral state reconstructions in HiSSE for the two MCC trees and under three different sampling scenarios of apomicts and sexuals.

"Supermatrix-ITS-like" tree		
Global Sampling Fraction		
	Sexuals	**Apomicts**
Speciation	4.18	7.57
Extinction	4.34	7.09
Net diversification	−0.16	0.47
Sampling Fractions with 50–50% sexuals vs. apomicts in "Mediterranean lineage"		
	Sexuals	**Apomicts**
Speciation	3.68	10.40
Extinction	4.22	9.87
Net diversification	−0.53	0.53
Sampling Fractions with 40–60% sexuals vs. apomicts in "Mediterranean lineage"		
	Sexuals	**Apomicts**
Speciation	2.96	9.88
Extinction	3.41	9.45
Net diversification	−0.46	0.43
"Supermatrix-cpDNA-like" tree		
Global Sampling Fraction		
	Sexuals	**Apomicts**
Speciation	4.27	7.99
Extinction	4.65	6.98
Net diversification	−0.38	1.01
Sampling Fractions with 50–50% sexuals vs. apomicts in "Mediterranean lineage"		
	Sexuals	**Apomicts**
Speciation	4.26	7.99
Extinction	4.65	6.97
Net diversification	−0.39	1.02
Sampling Fractions with 40–60% sexuals vs. apomicts in "Mediterranean lineage"		
	Sexuals	**Apomicts**
Speciation	3.40	12.27
Extinction	4.03	11.65
Net diversification	−0.63	0.62

Thompson, 2005; Pound and Salzmann, 2017), allow us to propose likely processes that shaped the spatio-temporal evolution of the genus. Our analyses support the origin of *Limonium* in the late Paleogene (ca. 33 Ma) most probably in the proto-Mediterranean region (**Figure 1**; **Table 1**; **Supplementary Figures S4**, **S5**). During this period, major tectonic activities changed the configuration of Eurasia. Europe, formerly an archipelago during the Eocene, became more continental, and the Mediterranean Sea separated from the newly formed intercontinental Paratethys Sea (Rögl, 1999; **Figure 1**). The newly created land corridors may have facilitated the initial range expansion of *Limonium* from the Euro-Mediterranean into the Irano-Turanian region (**Figure 1**). Furthermore, the origin of *Limonium* in the early Oligocene is marked by a climatic transition from sub-tropical to more seasonal conditions along the Euro-Mediterranean coasts. At that time, vegetation representatives of seasonal climates (i.e., warm temperate conifer forests and sclerophyll woodland) spread from the Iberian Peninsula to the coasts of France, and xerophytic shrubland occurred for the first time in the central Mediterranean coasts partially replacing subtropical paleo-biomes formerly dominating the Mediterranean area (Pound and Salzmann, 2017).

The divergence of the "Mediterranean lineage" (clade B3) of *Limonium* from its mostly non-Mediterranean sister lineage (clade B2) during the Miocene (**Table 1**; **Figure 1**) follows a temporal pattern of divergence found in many other Mediterranean angiosperm lineages (Vargas et al., 2018) and is concomitant with paleo-geological and -climatic changes. These two largest lineages of *Limonium*, clades B2 and B3, diverged during the middle Miocene (ca. 15.5 Ma) from an ancestor that was widespread in the Euro-Mediterranean, Irano-Turanian, and other regions (**Figure 1**; **Supplementary Figure S5**). Prior to this split, in the early Miocene and specifically late Burdigalian, the Arabian plate that was connected to Africa collided with the Anatolian plate for the first time, cutting off the seaway between the Mediterranean Sea and the Indian Ocean (**Figure 1**). In the early Langhian and concomitant with the divergence between B2 and B3 clades, a marine transgression flooded the entire Mediterranean and Paratethys causing the re-opening of the seaway between South Anatolia and Arabia and the formation of other seaways in Eastern Anatolia and probably also in the Aegean or Western Anatolian regions (Rögl, 1999**; Figure 1**). These short-lived seaways may have enabled vicariant speciation followed by range contraction giving origin to a Euro-Mediterranean ancestor for clade B3 and an Irano-Turanian ancestor for clade B2 ("Supermatrix-ITS-like" tree; **Figure 1**), or peripatric speciation (i.e., temporary peripheral isolate in Euro-Mediterranean) giving origin to a Euro-Mediterranean ancestor for clade B3 and a widespread ancestor in the same ancestral range for clade B2 ("Supermatrix-cpDNA-like" tree; **Supplementary Figure S5**). Moreover, the early Langhian split between B2 and B3 clades occurred during the Mid-Miocene Climatic Optimum (i.e., a global warming event at ca. 17–15 Ma; Böhme, 2003) and specifically in a period of increased seasonality in Central Europe and the Mediterranean area (up to 6 dry months; Böhme, 2003; Thompson, 2005). At that time, the Mediterranean flora began to resemble the current vegetation due to the extinction of several tropical and subtropical elements in the region (Thompson, 2005).

The diverse Mediterranean flora is composed of a combination of elements that immigrated to the region from other areas, and those that arose *in situ* (Thompson, 2005). Elucidating the relative roles of repeated colonization vs. *in situ* speciation is necessary to understand the processes that shaped Mediterranean diversity. Our results show that the great majority of Euro-Mediterranean *Limonium* species have an autochthonous origin (ca. 100 sympatric speciation vs. ca. 7 dispersal events; see BSM results above), supporting Thompson's (2005) hypothesis that *in situ* diversification contributed the main component of the heterogeneous Mediterranean flora. Additionally, the Euro-Mediterranean area received a few *Limonium* species, mostly from the neighboring Irano-Turanian, North African, and Circumboreal regions, consistent with previous studies on other taxa (e.g., Mansion et al., 2008, 2009; Salvo et al., 2010; Manafzadeh et al., 2014). The

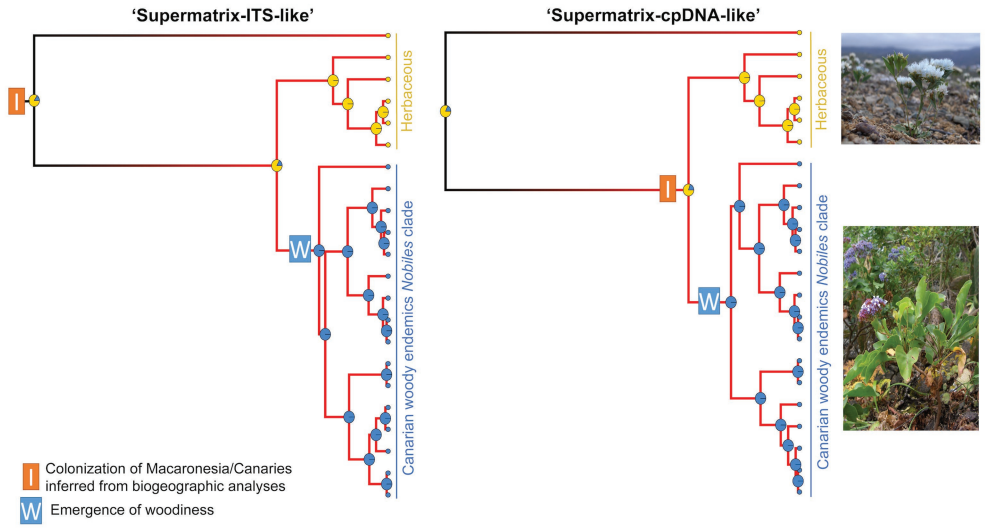


FIGURE 6 | Evolution of habit in relation to island colonization for *Pteroclados s.l.* clade of *Limonium* from "Supermatrix-ITS-like" and "Supermatrix-cpDNA-like" MCC trees (model-averaged HiSSE analyses). Yellow at the tips and nodes denotes herbaceous habit and blue denotes woody habit. The pies at the nodes represent the relative probabilities of the inferred ancestral states. Net-diversification rates increase from black to red along the branches. Shift to woodiness followed island colonization but did not drive a significant increase of diversification rates. Photos of the herbaceous *L. lobatum* (top photo) and the woody-suffruticose *L. arboreum* (bottom photo) from Ares Jiménez.

Euro-Mediterranean area also served as a major source of dispersals to other areas (see also tribe Antirrhineae: Gorospe et al., 2020), especially North Africa (**Figure 2**). Accordingly, diversity in this latter region was shaped mostly by dispersals (ca. 18 events, of which 14 from the Euro-Mediterranean area) and less by *in situ* speciation (ca. 3 events). Strong dispersal directionalities between areas (**Figure 2**) were also documented in Solanaceae by Dupin et al. (2017), who found a close relationship between species richness of an area and the number of dispersals from this area, a pattern congruent with our results. The inferred higher number of species dispersals from the Euro-Mediterranean "source" to the North African "sink" than to any other area could be explained by geoclimatic events that connected these two areas in the past, and by the availability of suitable habitats on their coastlines. Overall, the directional asymmetry of dispersals between the Euro-Mediterranean and other areas, consisting in more emigration than immigration to the region, suggests that extensive *in situ* speciation is the pump that promotes species movement from one main area (here the Euro-Mediterranean) to many others.

Proposed abiotic landmarks in the evolution of Mediterranean biodiversity include the Messinian Salinity Crisis (ca. 6–5.33 Ma), the onset of the Mediterranean climate (3.2–2.8 Ma), and Pleistocene geo-climatic fluctuations (Fiz-Palacios and Valcárcel, 2013). Indeed, we found that these events played a key role in the diversification of the "Mediterranean lineage" of *Limonium*. Although Mediterranean sea lavenders originated already in the middle Miocene (ca. 12 Ma; **Figure 1**; **Table 1**), all major extant lineages started to diversify only about 6 Ma concomitantly with the onset of the MSC, as a product of a significant, clade-specific shift toward higher diversification rates in the largest subclade of the "Mediterranean lineage" (phylorate plots; **Figure 4**). Thus, the salinity crisis that massively increased the availability of saline habitats across the Mediterranean represented an ecological opportunity for a salt tolerant genus like *Limonium* to experience a major episode of increased diversification rates. Conversely, the MSC apparently caused the extinction of other genera in the region, including the subtropical mangrove *Avicennia* (Thompson, 2005). Furthermore, the land corridors available across the Mediterranean at that time (see Mediterranean Paleogeographic Reconstruction During the MSC in Anzidei et al., 2014) might have provided venues for the colonization of North Africa from the Euro-Mediterranean (**Figures 1**, **2**), or vice versa, as documented for *Borago* (Mansion et al., 2009). The refilling of the Mediterranean Sea at the end of the MSC may have promoted further vicariant speciation of *Limonium* in the Euro-Mediterranean and North Africa (**Figure 1**; see also Romeiras et al., 2016).

Diversification continued to increase exponentially also in the past 2–3 Myr, after the establishment of the Mediterranean climate and during the Pleistocene climatic oscillations (rate-through-time plots; **Figure 4**), corroborating results from other Mediterranean plant taxa (e.g., Valente et al., 2010; Fiz-Palacios and Valcárcel, 2011). Here, we explicitly tested the impact of paleo-temperature and past sea-level on diversification rates for the larger Mediterranean subclade. A pattern supported in all analyses was that extinction rates were higher when sea-level and temperature were higher (**Supplementary Table S10**). Additionally, our results supported heterogeneous, yet range-independent diversification of *Limonium* (**Supplementary Table S13**), demonstrating that occurrence in the Euro-Mediterranean region *per se* cannot explain the burst of diversification rates. Rather, we conclude that it was the combination of environmental changes experienced by *Limonium* in this region and intrinsic biotic traits (i.e., apomixis) that triggered

the increase in diversification rates of the Mediterranean lineage, giving rise to its current high diversity of more than 400 species (see below).

Apomixis as a Biotic Driver of Diversification and the Interplay Between Biotic and Abiotic Factors in Bolstering Diversification Rates

In explaining the origins of the Mediterranean flora, questions have been raised about the relative role of geological and climatic history, and intrinsic biological processes such as polyploidization and hybridization (Thompson, 2005). In *Limonium*, as in other angiosperms (Hörandl and Hojsgaard, 2012), polyploidy and hybridization are usually associated with apomixis (Erben, 1978, 1979), and the combined effects of these three biotic factors were proposed to have shaped the Mediterranean radiation species (e.g., Palacios et al., 2000; Lledó et al., 2005, 2011). While several apomictic *Limonium* species are characterized as obligate apomicts (e.g., Palacios and González-Candelas, 1997; Arrigoni and Diana, 1999; Palacios et al., 1999; Khan et al., 2012; Róis et al., 2016), facultative apomixis (i.e., occasional sexual reproduction in apomicts) has also been documented for some species (e.g., D'Amato, 1949; Hjelmqvist and Grazi, 1964; Artelari, 1989; Artelari and Georgiou, 2002; Georgakopoulou et al., 2006). At the macroevolutionary level, our analyses show for the first time that apomixis is associated with an acceleration of diversification rates. Speciation rates are consistently higher in apomictic than sexual taxa, while extinction rates are either equal or higher in apomicts (**Figure 5**; **Table 3**; **Supplementary Table S11**).

The higher extinction rate inferred for apomicts than sexuals is congruent with the labile nature of apomicts proposed by Darlington (1958) and Holsinger (2000), who regarded apomixis as a "blind alley of evolution" due to reduced recombination, hence lower genetic variation in apomicts, eventually driving them to extinction. This view was based on the assumption that apomixis is an irreversible, derived trait. However, our results suggest that transitions from apomixis to sexuality are very common in *Limonium* (**Figure 5**) and corroborate recent studies that support apomixis-to-sex reversals enabled by the usually facultative nature of apomixis in angiosperms (Hörandl et al., 2007; Hörandl and Hojsgaard, 2012; Hand and Koltunow, 2014; Hojsgaard et al., 2014; Hojsgaard and Hörandl, 2015; Carman et al., 2019; Sharma and Bhat, 2020). While the result that transitions from apomixis to sexuality are more common than vice versa is strongly supported for both MCC trees in *Limonium*, it should be interpreted with caution because the limited phylogenetic resolution in the "Mediterranean lineage" could affect the inferred transition rates. The molecular genetic basis of apomixis is unknown in *Limonium*. However, in some plant genera, the molecular pathway of apomixis seems to be superimposed onto the pathway of sexual reproduction, rather than being completely independent, thus apomixis can revert to sexuality relatively easily (e.g., Hand and Koltunow, 2014 and references therein). This explanation for the molecular basis of apomixis is consistent with the occurrence of both apomictic and sexually developed seeds in facultative apomicts, suggesting that, if an ovule fails to initiate the apomictic pathway, sexual reproduction is activated, since the sexual pathway remains available. Importantly, in addition to enabling apomixis-to-sex reversals, the occurrence of "a little bit of sex" in apomicts can prevent the genomic decay caused by the absence of meiotic recombination, which can lead to extinction (Hörandl and Hojsgaard, 2012; Hojsgaard and Hörandl, 2015; Hodač et al., 2019).

Our results point toward a synergistic relationship between apomixis and sea-level fluctuations in driving the diversification of Mediterranean sea lavenders. In the Euro-Mediterranean region, the majority of *Limonium* species occur in coastal areas (which are vast, including many islands). Geo-climatic changes caused intense sea-level oscillations that had the highest frequency during the Plio-Pleistocene (**Supplementary Figure S3**), directly impacting coastal habitats. High sea levels triggered inundation and loss of available coastal habitats for *Limonium*, thus increasing extinction (see positive correlation of extinction rates with past sea levels for the "Mediterranean lineage" in **Supplementary Table S10**). Additionally, when sea levels are high, some populations may split and diversify allopatrically. Conversely, low sea levels create new areas that allow populations to expand, occasionally coming in contact and hybridizing. The newly created polyploid hybrids can establish new populations by single individuals through apomixis, and further diversify (see Negative Correlation of Speciation Rates With Past Sea Levels for the "Mediterranean lineage" in **Supplementary Table S10**). Apomixis provides both the required escape from sterility for hybrid polyploids and reproductive assurance in the absence of pollinators or mates in newly colonized habitats (Baker, 1955; Darlington, 1958). Thus, sea-level fluctuations cause repeated events of species-range contraction and fragmentation when sea level is high, and expansion and reconnection when sea level is low, promoting the origin of incipient species that can become established through apomixis. Long-term survival of apomictic species can be enhanced through occasional sex enabling the filtering of deleterious mutations *via* purging selection (Hojsgaard and Hörandl, 2015; Hodač et al., 2019).

Island Biogeography of Macaronesian Endemics and Insular Woodiness

Oceanic islands emerge from the sea empty of life. Their biodiversity is attributed to a combination of dispersals across the sea of species from neighboring areas and local diversification, producing high levels of endemicity (Cowie and Holland, 2006; Warren et al., 2015). *Limonium* endemics in Macaronesian archipelagos are of recent Plio-Pleistocene origin, except for *L. dendroides*, which diverged much earlier during the early Miocene (**Supplementary Figures S4**, **S5**). *Limonium dendroides* is an endangered Canarian endemic that is morphologically and taxonomically unique. It is characterized by at least two striking autapomorphies, i.e., arborescent habit and salt-glands in spikelet, and is the sole species of *L.* sect. *Limoniodendron* (**Figure 1**). The phylogenetic distinctiveness of *L. dendroides*, indicated by its isolated long branch sister to the large clade

comprising all other species of subgenus *Limonium*, suggests that *L. dendroides* is either an old species or it is the only surviving species of a once larger clade that has undergone extensive extinction (Warren et al., 2018). Many plant taxa in Macaronesia comprise multiple endemic species that stemmed from a single long-distance dispersal followed by *in situ* diversification (e.g., Böhle et al., 1996; Francisco-Ortega et al., 1996; Barber et al., 2007; Kim et al., 2008; Caujapé-Castells, 2011). In *Limonium*, however, Macaronesian diversity was shaped by multiple LDD events (ca. 7), mostly from the neighboring Euro-Mediterranean and North African areas, followed by *in situ* speciation (ca. 22 events; BSM results). The Canarian and Cape Verde archipelagos were colonized repeatedly (at least four times and twice, respectively), as found in several studies of other Macaronesian taxa (Carine et al., 2004; Sanmartín et al., 2008 and references therein; Navarro-Pérez et al., 2015; Jaén-Molina et al., 2020). The availability of different ecological niches in oceanic islands may facilitate independent colonization by different species, possibly *via* reducing competition. For example, the two *Limonium* endemic lineages stemming from independent dispersals to Cape Verde occupy distinct habitats, namely wet mountainous cliffs and coasts (Romeiras et al., 2015). Thus, habitat heterogeneity of oceanic island systems favors colonization by a diversity of species pre-adapted to contrasting ecological niches.

While oceanic islands have been traditionally regarded as major sinks of biodiversity (Carlquist, 1974), their role as sources of biodiversity for neighboring continents or archipelagos has been rarely documented (e.g., Mort et al., 2002; Carine et al., 2004). Our results support four LDD events for *Limonium* from Macaronesia including at least one to North Africa, but also Trans-Atlantic LDD to the Americas (**Figure 2**; **Supplementary Table S8**). Fine-scale analysis of *Jovibarba-Ctenostachys* clade reveals a dispersal event from Macaronesia to Hispaniola and another to Morocco, and one or two inter-archipelago dispersals between Canaries and Cape Verde (**Figure 3**). These results highlight the significance of oceanic archipelagos in Macaronesia as a source flora for LDD to archipelagos within and outside the region, and to neighboring continents (see also Gruenstaeudl et al., 2017; Jaén-Molina et al., 2020). Furthermore, the noteworthy long-distance dispersal from Cape Verde across the Atlantic Ocean to Hispaniola (Caribbean) was probably facilitated by both the light diaspores of these species and sea currents, such as the North Equatorial Current, as suggested for other angiosperms (Renner, 2004). Our results further suggest that different dispersal properties in Macaronesian lineages may explain the geographically broad range of the *Jovibarba-Ctenostachys* clade vs. the within-archipelago range of the Canarian *Nobiles* clade. Indeed, reduced dispersal abilities have been documented for the latter clade (Jiménez et al., 2017), supporting the hypothesis of post-colonization loss of dispersal ability in colonists of oceanic islands followed by extensive *in situ* diversification (Carlquist, 1966; MacArthur and Wilson, 1967).

Woodiness is very common in oceanic islands (Darwin, 1859; Wallace, 1878; Carlquist, 1965, 1974; Lens et al., 2013; Burns, 2019). Two competing hypotheses about the origin of insular woodiness have been proposed, which differ in whether woodiness evolved before or after island colonization. The "islands-as-museums" hypothesis views woodiness as a relictual trait already present in continental taxa that colonized islands and later went extinct from the continent (Cronk, 1992, 1997). Alternatively, woodiness has been interpreted as a derived trait that evolved *in situ* from herbaceous colonists subsequent to island colonization (e.g., Nürk et al., 2019). Our results in *Limonium* support the latter hypothesis by showing that woodiness is a derived trait in *Pteroclados s.l.* that emerged after colonization of the Canaries (woody *Nobiles* clade; **Figure 6**). The derived nature of woodiness is also supported by most studies on woody insular plants (e.g., *Echium*: Böhle et al., 1996; *Lavatera*: Fuertes-Aguilar et al., 2002; *Sideritis*: Barber et al., 2002; *Sanctambrosia*: Kool and Thulin, 2017), while fewer studies support its relictual nature (e.g., *Tolpis*: Moore et al., 2002; *Descurainia*: Goodson et al., 2006).

Woodiness on oceanic islands has been regarded as a key innovation linked to higher diversification rates when it is associated with disparity in growth forms (e.g., arborescent shrubs, subshrubs, trees, cushion forms, woody lianas, and giant rosette plants), because it enables the exploration of a broader niche space (Nürk et al., 2019). However, our results suggest that insular woodiness in the Canarian *Nobiles* clade of *Limonium* is either not linked to accelerated diversification (**Figure 6**; **Supplementary Tables S14**, **S5**) or a shift in rates for woody vs. herbaceous taxa is too moderate for the SSE methods to detect it, considering their limited power in analyses of small-size clades (Gamisch, 2016; Kodandaramaiah and Murali, 2018). A possible explanation for the result that the evolution of woodiness on the Canary Islands did not trigger a significant shift to much higher diversification rates may be that, in Canarian *Limonium*, woodiness is not associated with a diversity of woody growth forms (as found by Nürk et al., 2019), since all species in *Nobiles* clade have similar subshrub suffruticose habit. The lack of diversity in woody forms observed in species of the *Nobiles* clade may have limited their ability to radiate in a way that can be detected as a significant shift of diversification rates.

CONCLUSION

At the global scale, our study shows that the evolution of *Limonium* in space and time was shaped by major geologic and climatic processes that resulted in its cosmopolitan distribution and extensive diversification in the Mediterranean area. The increased diversification of Mediterranean sea lavenders was enabled by the ability of these taxa to reproduce asexually, *via* apomixis, which allowed them to survive and diversify during climatic oscillations. Our results show that the joint effect of biotic and abiotic factors is responsible for the current diversity of *Limonium*. At the regional scale, focusing on the insular endemics of Macaronesia, diversity on oceanic islands was shaped by multiple colonization events from neighboring continents and archipelagos, and *in situ* speciation. The Macaronesian islands have also served as sources for dispersals

to other archipelagos (Caribbean) and continents (North Africa and America). In addition, woodiness in the Canarian sea lavenders (*Nobiles* clade) is a derived trait linked to insularity but not too much higher diversification rates. Our study highlights the importance of analyzing multiple abiotic and biotic factors, and their interactions, to achieve an in-depth understanding of evolution in hotspots of biodiversity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

EC and KK contributed to conception and design of the study. AJ, BW, KK, and MR provided material and data for analyses. KK performed all analyses with help from BW, MC, and ST and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the final version.

FUNDING

Financial support was given by the University of Zurich (Department of Systematic and Evolutionary Botany) and Georges-und-Antoine-Claraz-Schenkung.

ACKNOWLEDGMENTS

We are grateful to Prof. Dr. Peter Linder, Dr. Yanis Bouchenak-Khelladi, Dr. Fabien Condamine, Dr. Hélène Morlon, Dr. Isabel Sanmartín, and Ass. Prof. Mónica Maria Tavares Moura for their valuable comments and useful insights in this study. We also thank the two reviewers of this study for their valuable comments and suggestions to improve the final version of this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.612258/full#supplementary-material

REFERENCES

Anzidei, M., Lambeck, K., Antonioli, F., Furlani, S., Mastronuzzi, G., Serpelloni, E., et al. (2014). Coastal structure, sea-level changes and vertical motion of the land in the Mediterranean. *Geol. Soc. Lond., Spec. Publ.* 388, 453–479. doi: 10.1144/SP388.20

Arrigoni, P. V., and Diana, S. (1999). Karyology, chorology and bioecology of the genus *Limonium* (Plumbaginaceae) in Sardinia. *Plant Biosyst.* 133, 63–71. doi: 10.1080/11263509909381533

Artelari, R. (1989). Biosystematic study of the genus *Limonium* (Plumbaginaceae) in the Aegean area (Greece). I. some Limonium species from the Kikladhes islands. *Willdenowia* 18, 399–408.

Artelari, R., and Georgiou, O. (2002). Biosystematic study of the genus *Limonium* (Plumbaginaceae) in the Aegean area, Greece. III. Limonium on the islands Kithira and Antikithira and the surrounding islets. *Nord. J. Bot.* 22, 483–502. doi: 10.1111/j.1756-1051.2002.tb01402.x

Asker, S., and Jerling, L. (1992). *Apomixis in plants*. Boca Raton, FL: CRC Press.

Baker, H. G. (1955). Self-compatibility and establishment after "long-distance" dispersal. *Evolution* 9, 347–349. doi: 10.2307/2405656

Baker, H. G. (1966). The evolution, functioning and breakdown of heteromorphic incompatibility systems. I. the Plumbaginaceae. *Evolution* 20, 349–368. doi: 10.2307/2406635

Barber, J. C., Finch, C. C., Francisco-Ortega, J., Santos-Guerra, A., and Jansen, R. K. (2007). Hybridization in Macaronesian *Sideritis* (Lamiaceae): evidence from incongruence of multiple independent nuclear and chloroplast sequence datasets. *Taxon* 56, 74–88. doi: 10.2307/25065737

Barber, J. C., Francisco-Ortega, J., Santos-Guerra, A., Turner, K. G., and Jansen, R. K. (2002). Origin of Macaronesian *Sideritis* L.(Lamioideae: Lamiaceae) inferred from nuclear and chloroplast sequence datasets. *Mol. Phylogenet. Evol.* 23, 293–306. doi: 10.1016/S1055-7903(02)00018-0

Barnosky, A. D. (2001). Distinguishing the effects of the red queen and court jester on Miocene mammal evolution in the northern Rocky Mountains. *J. Vertebr. Paleontol.* 21, 172–185. doi: 10.1671/0272-4634(2001)021[0172:DTEOTR]2.0.CO;2

Beaulieu, J. M., and O'Meara, B. C. (2016). Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst. Biol.* 65, 583–601. doi: 10.1093/sysbio/syw022

Benton, M. J. (2009). The red queen and the court jester: species diversity and the role of biotic and abiotic factors through time. *Science* 323, 728–732. doi: 10.1126/science.1157719

Benton, M. J. (2015). Exploring macroevolution using modern and fossil data. *Proc. R. Soc. B* 282:20150569. doi: 10.1098/rspb.2015.0569

Bessedik, M., Guinet, P., and Suc, J. P. (1984). Données paléofloristiques en Méditerranée Nord-occidentale depuis l'Aquitanien. *Rev. Paléobiol.* 25–31.

Beucher, F. (1975). *Étude palynologique de formations néogènes et quaternaires au Sahara Nord-occidental. Vol. 20*. Paris: Centre national de la recherche scientifique.

Blondel, J., and Aronson, J. (1999). *Biology and wildlife of the Mediterranean region*. New York: Oxford University Press.

Blondel, J., Aronson, J., Bodiou, J. Y., and Boeuf, G. (2010). *The Mediterranean region: Biological diversity in space and time*. New York: Oxford University Press.

Böhle, U. R., Hilger, H. H., and Martin, W. F. (1996). Island colonization and evolution of the insular woody habit in *Echium* L. (Boraginaceae). *Proc. Natl. Acad. Sci. U. S. A.* 93, 11740–11745. doi: 10.1073/pnas.93.21.11740

Böhme, M. (2003). The Miocene climatic optimum: evidence from ectothermic vertebrates of Central Europe. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 195, 389–401. doi: 10.1016/S0031-0182(03)00367-5

Borges, P. A., Abreu, C., Aguiar, A. M. F., Carvalho, P., Jardim, R., Melo, I., et al. (eds.) (2008). "A list of the terrestrial fungi, flora and fauna of Madeira and Selvagens archipelagos" in *Direcção regional do Ambiente da Madeira and Universidade dos Açores*. Horta: Angra do Heroísmo and Ponta Delgada.

Bouchenak-Khelladi, Y., Muasya, A. M., and Linder, H. P. (2014). A revised evolutionary history of Poales: origins and diversification. *Bot. J. Linn. Soc.* 175, 4–16. doi: 10.1111/boj.12160

Bouchenak-Khelladi, Y., Onstein, R. E., Xing, Y., Schwery, O., and Linder, H. P. (2015). On the complexity of triggering evolutionary radiations. *New Phytol.* 207, 313–326. doi: 10.1111/nph.13331

Boucher, F. C., Zimmermann, N. E., and Conti, E. (2016). Allopatric speciation with little niche divergence is common among alpine Primulaceae. *J. Biogeogr.* 43, 591–602. doi: 10.1111/jbi.12652

Bouckaert, R. R., and Drummond, A. J. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol. Biol.* 17:42. doi: 10.1186/s12862-017-0890-6

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al. (2019). BEAST 2.5: an advanced software platform

for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650. doi: 10.1371/journal.pcbi.1006650

Bramwell, D., and Bramwell, Z. I. (1974). *Wild flowers of the Canary Islands*. London: Stanley Thornes.

Bramwell, D., and Bramwell, Z. I. (1990). *Flores silvestres de las Islas Canarias*. *Vol. 1*. Madrid: Ed. Rueda.

Brullo, S., Brullo, C., Cambria, S., Del Galdo, G. G., and Minissale, P. (2017). Phytosociological investigation on the class Crithmo maritimi-Limonieteain Greece. *Plant Sociol.* 54, 3–57. doi: 10.7338/pls2017541/01

Brullo, S., and Erben, M. (2016). The genus *Limonium* (Plumbaginaceae) in Greece. *Phytotaxa* 240, 1–212. doi: 10.11646/phytotaxa.240.1.1

Burnham, K. P., and Anderson, D. R. (2002). *Model selection and multimodel inference. A practical information-theoretic approach. 2nd Edn*. New York: Springer.

Burns, K. C. (2019). *Evolution in isolation: The search for an island syndrome in plants*. Cambridge: Cambridge University Press.

Caetano, D. S., O'Meara, B. C., and Beaulieu, J. M. (2018). Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* 72, 2308–2324. doi: 10.1111/evo.13602

Carine, M. A., Russell, S. J., Santos-Guerra, A., and Francisco-Ortega, J. (2004). Relationships of the Macaronesian and Mediterranean floras: molecular evidence for multiple colonizations into Macaronesia and back-colonization of the continent in *Convolvulus* (Convolvulaceae). *Am. J. Bot.* 91, 1070–1085. doi: 10.3732/ajb.91.7.1070

Carine, M. A., Santos-Guerra, A., Guma, I. R., and Reyes-Betancourt, J. A. (2010). "Endemism and evolution of the Macaronesian flora" in *Beyond cladistics: The branching of a paradigm*. eds. D. M. Williams and S. Knapp (Oakland, CA: University of California Press), 101–124.

Carlquist, S. (1965). *Island life; a natural history of the islands of the world*. New York: The Natural History Press.

Carlquist, S. (1966). The biota of long-distance dispersal. III. Loss of dispersibility in the Hawaiian flora. *Brittonia* 18, 310–335. doi: 10.2307/2805148

Carlquist, S. (1974). *Island biology*. New York: Columbia University Press.

Carman, J. G., Mateo de Arias, M., Gao, L., Zhao, X., Kowallis, B. M., Sherwood, D. A., et al. (2019). Apospory and diplospory in diploid *Boechera* (Brassicaceae) may facilitate speciation by recombination-driven apomixis-to-sex reversals. *Front. Plant Sci.* 10:724. doi: 10.3389/fpls.2019.00724

Caujapé-Castells, J. (2011). "Jesters, red queens, boomerangs and surfers: a molecular outlook on the diversity of the Canarian endemic flora" in *The biology of island floras*. eds. D. Bramwell and J. Caujapé-Castells (Cambridge: Cambridge University Press), 284–324.

Christenhusz, M. J., and Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa* 261, 201–217. doi: 10.11646/phytotaxa.261.3.1

Comes, H. P. (2004). The Mediterranean region-a hotspot for plant biogeographic research. *New Phytol.* 164, 11–14. doi: 10.1111/j.1469-8137.2004.01194.x

Condamine, F. L., Rolland, J., Höhna, S., Sperling, F. A., and Sanmartín, I. (2018). Testing the role of the red queen and court jester as drivers of the macroevolution of Apollo butterflies. *Syst. Biol.* 67, 940–964. doi: 10.1093/sysbio/syy009

Condamine, F. L., Rolland, J., and Morlon, H. (2013). Macroevolutionary perspectives to environmental change. *Ecol. Lett.* 16, 72–85. doi: 10.1111/ele.12062

Costa, J., Torices, R., and Barrett, S. C. (2019). Evolutionary history of the buildup and breakdown of the heterostylous syndrome in Plumbaginaceae. *New Phytol.* 224, 1278–1289. doi: 10.1111/nph.15768

Cowie, R. H., and Holland, B. S. (2006). Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. *J. Biogeogr.* 33, 193–198. doi: 10.1111/j.1365-2699.2005.01383.x

Critical Ecosystem Partnership Fund (2019). Biodiversity hotspots. Available at: https://www.cepf.net/our-work/biodiversity-hotspots (Accessed November 15, 2019).

Cronk, Q. C. B. (1992). Relict floras of Atlantic islands: patterns assessed. *Biol. J. Linn. Soc. Lond.* 46, 91–103. doi: 10.1111/j.1095-8312.1992.tb00852.x

Cronk, Q. C. B. (1997). Islands: stability, diversity, conservation. *Biodivers. Conserv.* 6, 477–493. doi: 10.1023/A:1018372910025

Crowl, A. A., Visger, C. J., Mansion, G., Hand, R., Wu, H. H., Kamari, G., et al. (2015). Evolution and biogeography of the endemic *Roucela* complex (Campanulaceae: campanula) in the eastern Mediterranean. *Ecol. Evol.* 5, 5329–5343. doi: 10.1002/ece3.1791

D'Amato, F. (1949). Triploidia e apomissia in *Statice oleaefolia* Scop. Var. *confusa* Godr. *Caryologia* 2, 71–84. doi: 10.1080/00087114.1949.10797527

Darlington, C. D. (1958). *The evolution of genetic systems. 2nd Edn*. Edinburgh: Oliver & Boyd.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London: J Murray.

de Vos, J. M., Hughes, C. E., Schneeweiss, G. M., Moore, B. R., and Conti, E. (2014). Heterostyly accelerates diversification via reduced extinction in primroses. *Proc. R. Soc. B* 281:20140075. doi: 10.1098/rspb.2014.0075

Donoghue, M. J., and Sanderson, M. J. (2015). Confluence, synnovation, and depauperons in plant diversification. *New Phytol.* 207, 260–274. doi: 10.1111/nph.13367

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Duggen, S., Hoernle, K., Van Den Bogaard, P., Rüpke, L., and Morgan, J. P. (2003). Deep roots of the Messinian salinity crisis. *Nature* 422, 602–606. doi: 10.1038/nature01553

Dupin, J., Matzke, N. J., Särkinen, T., Knapp, S., Olmstead, R. G., Bohs, L., et al. (2017). Bayesian estimation of the global biogeographical history of the Solanaceae. *J. Biogeogr.* 44, 887–899. doi: 10.1111/jbi.12898

Erben, M. (1978). Die Gattung Limonium im südwestmediterranen Raum. *Mitt. Bot. Staatssamml. München.* 14, 361–631.

Erben, M. (1979). Karyotype differentiation and its consequences in Mediterranean «*Limonium*». *Webbia* 34, 409–417. doi: 10.1080/00837792.1979.10670178

Faurby, S., Eiserhardt, W. L., and Svenning, J. C. (2016). Strong effects of variation in taxonomic opinion on diversification analyses. *Methods Ecol. Evol.* 7, 4–13. doi: 10.1111/2041-210X.12449

Fernández-Mazuecos, M., Blanco-Pastor, J. L., Juan, A., Carnicero, P., Forrest, A., Alarcón, M., et al. (2019). Macroevolutionary dynamics of nectar spurs, a key evolutionary innovation. *New Phytol.* 222, 1123–1138. doi: 10.1111/nph.15654

Fernández-Palacios, J. M., De Nascimento, L., Otto, R., Delgado, J. D., García-del-Rey, E., Arévalo, J. R., et al. (2011). A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *J. Biogeogr.* 38, 226–246. doi: 10.1111/j.1365-2699.2010.02427.x

FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* 3, 1084–1092. doi: 10.1111/j.2041-210X.2012.00234.x

FitzJohn, R. G., Maddison, W. P., and Otto, S. P. (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* 58, 595–611. doi: 10.1093/sysbio/syp067

Fiz-Palacios, O., and Valcárcel, V. (2011). Imbalanced diversification of two Mediterranean sister genera (*Bellis* and *Bellium*, Asteraceae) within the same time frame. *Plant Syst. Evol.* 295, 109–118. doi: 10.1007/s00606-011-0468-5

Fiz-Palacios, O., and Valcárcel, V. (2013). From Messinian crisis to Mediterranean climate: a temporal gap of diversification recovered from multiple plant phylogenies. *Perspect. Plant Ecol. Evol. Syst.* 15, 130–137. doi: 10.1016/j.ppees.2013.02.002

Francisco-Ortega, J., Jansen, R. K., and Santos-Guerra, A. (1996). Chloroplast DNA evidence of colonization, adaptive radiation, and hybridization in the evolution of the Macaronesian flora. *Proc. Natl. Acad. Sci. U. S. A.* 93, 4085–4090. doi: 10.1073/pnas.93.9.4085

Frodin, D. G. (2004). History and concepts of big plant genera. *Taxon* 53, 753–776. doi: 10.2307/4135449

Fuertes-Aguilar, J., Ray, M. F., Francisco-Ortega, J., Santos-Guerra, A., and Jansen, R. K. (2002). Molecular evidence from chloroplast and nuclear markers for multiple colonizations of *Lavatera* (Malvaceae) in the Canary Islands. *Syst. Bot.* 27, 74–83. doi: 10.1043/0363-6445-27.1.74

Gamisch, A. (2016). Notes on the statistical power of the binary state speciation and extinction (BiSSE) model. *Evol. Bioinforma.* 12, 165–174. doi: 10.4137/EBO.S39732

García-Castellanos, D., Estrada, F., Jiménez-Munt, I., Gorini, C., Fernández, M., Vergés, J., et al. (2009). Catastrophic flood of the Mediterranean after the Messinian salinity crisis. *Nature* 462, 778–781. doi: 10.1038/nature08555

Gargani, J., and Rigollet, C. (2007). Mediterranean Sea level variations during the Messinian salinity crisis. *Geophys. Res. Lett.* 34:L10405. doi: 10.1029/2007GL029885

Georgakopoulou, A., Manousou, S., Artelari, R., and Georgiou, O. (2006). Breeding systems and cytology in Greek populations of five *Limonium* species (Plumbaginaceae). *Willdenowia* 36, 741–750. doi: 10.3372/wi.36.36209

Goldberg, E. E., and Igić, B. (2012). Tempo and mode in plant breeding system evolution. *Evolution* 66, 3701–3709. doi: 10.1111/j.1558-5646.2012.01730.x

Goldberg, E. E., Kohn, J. R., Lande, R., Robertson, K. A., Smith, S. A., and Igić, B. (2010). Species selection maintains self-incompatibility. *Science* 330, 493–495. doi: 10.1126/science.1194513

Goodson, B. E., Santos-Guerra, A., and Jansen, R. K. (2006). Molecular systematics of *Descurainia* (Brassicaceae) in the Canary Islands: biogeographic and taxonomic implications. *Taxon* 55, 671–682. doi: 10.2307/25065643

Gorospe, J. M., Monjas, D., and Fernández-Mazuecos, M. (2020). Out of the Mediterranean region: worldwide biogeography of snapdragons and relatives (tribe Antirrhineae, Plantaginaceae). *J. Biogeogr.* 47, 2442–2456. doi: 10.1111/jbi.13939

Graham, A. (1976). Studies in neotropical paleobotany. II. The Miocene communities of Veracruz, Mexico. *Ann. Mo. Bot. Gard.* 63, 787–842. doi: 10.2307/2395250

Gruenstaeudl, M., Carstens, B. C., Santos-Guerra, A., and Jansen, R. K. (2017). Statistical hybrid detection and the inference of ancestral distribution areas in *Tolpis* (Asteraceae). *Biol. J. Linn. Soc. Lond.* 121, 133–149. doi: 10.1093/biolinnean/blw014

Hand, M. L., and Koltunow, A. M. (2014). The genetic control of apomixis: asexual seed formation. *Genetics* 197, 441–450. doi: 10.1534/genetics.114.163105

Haveman, R. (2013). Freakish patterns-species and species concepts in apomicts. *Nord. J. Bot.* 31, 257–269. doi: 10.1111/j.1756-1051.2013.00158.x

Hjelmqvist, H., and Grazi, F. (1964). Studies on variation in embryo sac development. *Bot. Notiser* 117, 141–166.

Hodač, L., Klatt, S., Hojsgaard, D., Sharbel, T. F., and Hörandl, E. (2019). A little bit of sex prevents mutation accumulation even in apomictic polyploid plants. *BMC Evol. Biol.* 19:170. doi: 10.1186/s12862-019-1495-z

Höhna, S., May, M. R., and Moore, B. R. (2015). TESS: an R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. *Bioinformatics* 32, 789–791. doi: 10.1093/bioinformatics/btv651

Hojsgaard, D., and Hörandl, E. (2015). A little bit of sex matters for genome evolution in asexual plants. *Front. Plant Sci.* 6:82. doi: 10.3389/fpls.2015.00082

Holsinger, K. E. (2000). Reproductive systems and evolution in vascular plants. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7037–7042. doi: 10.1073/pnas.97.13.7037

Hojsgaard, D., Klatt, S., Baier, R., Carman, J. G., and Hörandl, E. (2014). Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics. *Crit. Rev. Plant Sci.* 33, 414–427. doi: 10.1080/07352689.2014.898488

Hörandl, E., Grossniklaus, U., van Dijk, P. J., and Sharbel, T. (2007). *Apomixis: Evolution, mechanisms and perspectives*, Regnum Vegetabile. *Vol. 147*. Rugell, Liechtenstein: Gantner Verlag.

Hörandl, E., and Hojsgaard, D. (2012). The evolution of apomixis in angiosperms: a reappraisal. *Plant Biosyst.* 146, 681–693. doi: 10.1080/11263504.2012.716795

Howard, C. C., Landis, J. B., Beaulieu, J. M., and Cellinese, N. (2020). Geophytism in monocots leads to higher rates of diversification. *New Phytol.* 225, 1023–1032. doi: 10.1111/nph.16155

Hsü, K. J., Ryan, W. B., and Cita, M. B. (1973). Late Miocene desiccation of the Mediterranean. *Nature* 242, 240–244. doi: 10.1038/242240a0

Ingrouille, M. J. (1984). A taxometric analysis of *Limonium* (Plumbaginaceae) in Western Europe. *Plant Syst. Evol.* 147, 103–118. doi: 10.1007/BF00984583

Jaén-Molina, R., Marrero-Rodríguez, Á., Caujapé-Castells, J., and Ojeda, D. I. (2020). Molecular phylogenetics of *Lotus* (Leguminosae) with emphasis in the tempo and patterns of colonization in the Macaronesian region. *Mol. Phylogenet. Evol.* 154:106970. doi: 10.1016/j.ympev.2020.106970

Jiménez, A., Weigelt, B., Santos-Guerra, A., Caujapé-Castells, J., Fernández-Palacios, J. M., and Conti, E. (2017). Surviving in isolation: genetic variation, bottlenecks and reproductive strategies in the Canarian endemic *Limonium macrophyllum* (Plumbaginaceae). *Genetica* 145, 91–104. doi: 10.1007/s10709-017-9948-z

Jiménez-Moreno, G., Fauquette, S., and Suc, J. P. (2010). Miocene to Pliocene vegetation reconstruction and climate estimates in the Iberian Peninsula from pollen data. *Rev. Palaeobot. Palynol.* 162, 403–415. doi: 10.1016/j.revpalbo.2009.08.001

Khan, Z., Santpere, G., and Traveset, A. (2012). Breeding system and ecological traits of the critically endangered endemic plant *Limonium barceloi* (Gil and Llorens) (Plumbaginaceae). *Plant Syst. Evol.* 298, 1101–1110. doi: 10.1007/s00606-012-0619-3

Kim, S. C., McGowen, M. R., Lubinsky, P., Barber, J. C., Mort, M. E., and Santos-Guerra, A. (2008). Timing and tempo of early and successive adaptive radiations in Macaronesia. *PLoS One* 3:e2139. doi: 10.1371/journal.pone.0002139

Kodandaramaiah, U., and Murali, G. (2018). What affects power to estimate speciation rate shifts? *PeerJ* 6:e5495. doi: 10.7717/peerj.5495

Kool, A., and Thulin, M. (2017). A giant spurrey on a tiny island: on the phylogenetic position of *Sanctambrosia manicata* (Caryophyllaceae) and the generic circumscriptions of *Spergula*, *Spergularia* and *Rhodalsine*. *Taxon* 66, 615–622. doi: 10.12705/663.6

Koutroumpa, K., Theodoridis, S., Warren, B. H., Jiménez, A., Celep, F., Doğan, M., et al. (2018). An expanded molecular phylogeny of Plumbaginaceae, with emphasis on *Limonium* (sea lavenders): taxonomic implications and biogeographic considerations. *Ecol. Evol.* 8, 12397–12424. doi: 10.1002/ece3.4553

Krijgsman, W., Hilgen, F. J., Raffi, I., Sierro, F. J., and Wilson, D. S. (1999). Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400, 652–655. doi: 10.1038/23231

Kunkel, G. (2012). *Biogeography and ecology in the Canary Islands*. The Hague: Springer Science & Business Media.

Lagomarsino, L. P., Condamine, F. L., Antonelli, A., Mulch, A., and Davis, C. C. (2016). The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytol.* 210, 1430–1442. doi: 10.1111/nph.13920

Landis, M. J., Matzke, N. J., Moore, B. R., and Huelsenbeck, J. P. (2013). Bayesian analysis of biogeography when the number of areas is large. *Syst. Biol.* 62, 789–804. doi: 10.1093/sysbio/syt040

Lens, F., Davin, N., Smets, E., and del Arco, M. (2013). Insular woodiness on the Canary Islands: a remarkable case of convergent evolution. *Int. J. Plant Sci.* 174, 992–1013. doi: 10.1086/670259

Letsch, H., Gottsberger, B., Metzl, C., Astrin, J., Friedman, A. L., McKenna, D. D., et al. (2018). Climate and host-plant associations shaped the evolution of ceutorhynch weevils throughout the Cenozoic. *Evolution* 72, 1815–1828. doi: 10.1111/evo.13520

Lisiecki, L. E., and Raymo, M. E. (2007). Plio-Pleistocene climate evolution: trends and transitions in glacial cycle dynamics. *Quat. Sci. Rev.* 26, 56–69. doi: 10.1016/j.quascirev.2006.09.005

Lledó, M. D., Crespo, M. B., Fay, M. F., and Chase, M. W. (2005). Molecular phylogenetics of *Limonium* and related genera (Plumbaginaceae): biogeographical and systematic implications. *Am. J. Bot.* 92, 1189–1198. doi: 10.3732/ajb.92.7.1189

Lledó, M. D., Karis, P. O., Crespo, M. B., Fay, M. F., and Chase, M. W. (2011). "Endemism and evolution in macaronesian and mediterranean Limonium taxa" in *The biology of island floras*. eds. D. Bramwell and J. Caujapé-Castells (Cambridge: Cambridge University Press), 325–337.

Mabberley, D. J. (2017). *Mabberley's plant-book: A portable dictionary of plants, their classification and uses. 4th Edn*. Cambridge: Cambridge University Press.

MacArthur, R. H., and Wilson, E. O. (1967). *Island biogeography*. Princeton: Princeton University Press.

Macphail, M. K. (1999). Palynostratigraphy of the Murray Basin, inland southeastern Australia. *Palynology* 23, 197–240. doi: 10.1080/01916122.1999.9989528

Maddison, W. P., and FitzJohn, R. G. (2015). The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst. Biol.* 64, 127–136. doi: 10.1093/sysbio/syu070

Maddison, W. P., Midford, P. E., and Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56, 701–710. doi: 10.1080/10635150701607033

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., and Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* 207, 437–453. doi: 10.1111/nph.13264

Magallon, S., and Sanderson, M. J. (2001). Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762–1780. doi: 10.1111/j.0014-3820.2001.tb00826.x

Majeský, Ľ., Krahulec, F., and Vašut, R. J. (2017). How apomictic taxa are treated in current taxonomy: a review. *Taxon* 66, 1017–1040. doi: 10.12705/665.3

Malekmohammadi, M., Akhani, H., and Borsch, T. (2017). Phylogenetic relationships of *Limonium* (Plumbaginaceae) inferred from multiple chloroplast and nuclear loci. *Taxon* 66, 1128–1146. doi: 10.12705/665.8

Manafzadeh, S., Salvo, G., and Conti, E. (2014). A tale of migrations from east to west: the Irano-Turanian floristic region as a source of Mediterranean xerophytes. *J. Biogeogr.* 41, 366–379. doi: 10.1111/jbi.12185

Mansion, G., Rosenbaum, G., Schoenenberger, N., Bacchetta, G., Rosselló, J. A., and Conti, E. (2008). Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean Basin by the angiosperm family Araceae. *Syst. Biol.* 57, 269–285. doi: 10.1080/10635150802044029

Mansion, G., Selvi, F., Guggisberg, A., and Conti, E. (2009). Origin of Mediterranean insular endemics in the Boraginales: integrative evidence from molecular dating and ancestral area reconstruction. *J. Biogeogr.* 36, 1282–1296. doi: 10.1111/j.1365-2699.2009.02082.x

Marrero, Á., and Almeida, R. (2003). Novedades taxonómicas del género *Limonium* mill. Subsecc. Nobiles en gran Canaria (islas Canarias) (Plumbaginaceae-Staticoideae). *Vieraea* 31, 391–406.

Matzke, N. J. (2013). Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* 5, 242–248. doi: 10.21425/F5FBG19694

Matzke, N. J. (2016). "Stochastic mapping under biogeographical models." PhyloWiki BioGeoBEARS website, 2016. Available at: http://phylo.wikidot.com/biogeobears#stochastic_mapping (Accessed September 15, 2020).

Mayhew, P. J., Jenkins, G. B., and Benton, T. G. (2008). A long-term association between global temperature and biodiversity, origination and extinction in the fossil record. *Proc. R. Soc. B* 275, 47–53. doi: 10.1098/rspb.2007.1302

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

Mesa, R., Santos, A., Oval, J. P., and Voggenreiter, V. (2001). *Limonium relicticum*, una nueva especie Para La Gomera, islas Canarias (Plumbaginaceae). *Vieraea* 29, 111–118.

Miller, K. G., Kominz, M. A., Browning, J. V., Wright, J. D., Mountain, G. S., Katz, M. E., et al. (2005). The Phanerozoic record of global sea-level change. *Science* 310, 1293–1298. doi: 10.1126/science.1116412

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), Nov 14, 2010, New Orleans, LA, 1–8.

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

Moore, M. J., Francisco-Ortega, J., Santos-Guerra, A., and Jansen, R. K. (2002). Chloroplast DNA evidence for the roles of island colonization and extinction in *Tolpis* (Asteraceae: Lactuceae). *Am. J. Bot.* 89, 518–526. doi: 10.3732/ajb.89.3.518

Morlon, H., Lewitus, E., Condamine, F. L., Manceau, M., Clavel, J., and Drury, J. (2016). RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods Ecol. Evol.* 7, 589–597. doi: 10.1111/2041-210X.12526

Morlon, H., Parsons, T. L., and Plotkin, J. B. (2011). Reconciling molecular phylogenies with the fossil record. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16327–16332. doi: 10.1073/pnas.1102543108

Morlon, H., Potts, M. D., and Plotkin, J. B. (2010). Inferring the dynamics of diversification: a coalescent approach. *PLoS Biol.* 8:e1000493. doi: 10.1371/journal.pbio.1000493

Mort, M. E., Soltis, D. E., Soltis, P. S., Francisco-Ortega, J., and Santos-Guerra, A. (2002). Phylogenetics and evolution of the Macaronesian clade of Crassulaceae inferred from nuclear and chloroplast sequence data. *Syst. Bot.* 27, 271–288. doi: 10.1043/0363-6445-27.2.271

Muller, J. (1981). Fossil pollen records of extant angiosperms. *Bot. Rev.* 47, 1–140.

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

Navarro-Pérez, M. L., Vargas, P., Fernández-Mazuecos, M., López, J., Valtueña, F. J., and Ortega-Olivencia, A. (2015). Multiple windows of colonization to Macaronesia by the dispersal-unspecialized *Scrophularia* since the late Miocene. *Perspect. Plant Ecol. Evol. Syst.* 17, 263–273. doi: 10.1016/j.ppees.2015.05.002

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

Nürk, N. M., Atchison, G. W., and Hughes, C. E. (2019). Island woodiness underpins accelerated disparification in plant radiations. *New Phytol.* 224, 518–531. doi: 10.1111/nph.15797

Palacios, C., and González-Candelas, F. (1997). Lack of genetic variability in the rare and endangered *Limonium cavanillesii* (Plumbaginaceae) using RAPD markers. *Mol. Ecol.* 6, 671–675. doi: 10.1046/j.1365-294X.1997.00232.x

Palacios, C., Kresovich, S., and González-Candelas, F. (1999). A population genetic study of the endangered plant species *Limonium dufourii* (Plumbaginaceae) based on amplified fragment length polymorphism (AFLP). *Mol. Ecol.* 8, 645–657. doi: 10.1046/j.1365-294X.1999.t01-1-00597.x

Palacios, C., Rosselló, J. A., and González-Candelas, F. (2000). Study of the evolutionary relationships among *Limonium* species (Plumbaginaceae) using nuclear and cytoplasmic molecular markers. *Mol. Phylogenet. Evol.* 14, 232–249. doi: 10.1006/mpev.1999.0690

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news* 6, 7–11.

Pound, M. J., and Salzmann, U. (2017). Heterogeneity in global vegetation and terrestrial climate change during the late Eocene to early Oligocene transition. *Sci. Rep.* 7:43386. doi: 10.1038/srep43386

Pyron, R. A., and Burbrink, F. T. (2013). Phylogenetic estimates of speciation and extinction rates for testing ecological and evolutionary hypotheses. *Trends Ecol. Evol.* 28, 729–736. doi: 10.1016/j.tree.2013.09.007

Rabosky, D. L., and Goldberg, E. E. (2015). Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.* 64, 340–355. doi: 10.1093/sysbio/syu131

Rabosky, D. L., and Goldberg, E. E. (2017). FiSSE: a simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution* 71, 1432–1442. doi: 10.1111/evo.13227

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., et al. (2014). BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* 5, 701–707. doi: 10.1111/2041-210X.12199

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

R Core Team (2018). R: a language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Available at: https://www.R-project.org/ (Accessed September 15, 2020).

Ree, R. H., and Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *J. Biogeogr.* 45, 741–749. doi: 10.1111/jbi.13173

Ree, R. H., and Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57, 4–14. doi: 10.1080/10635150701883881

Renner, S. (2004). Plant dispersal across the tropical Atlantic by wind and sea currents. *Int. J. Plant Sci.* 165, S23–S33. doi: 10.1086/383334

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., et al. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301, 1211–1216. doi: 10.1126/science.1086949

Rodríguez-Sánchez, F., Pérez-Barrales, R., Ojeda, F., Vargas, P., and Arroyo, J. (2008). The strait of Gibraltar as a melting pot for plant biodiversity. *Quat. Sci. Rev.* 27, 2100–2117. doi: 10.1016/j.quascirev.2008.08.006

Rögl, F. (1999). Mediterranean and paratethys. Facts and hypotheses of an Oligocene to Miocene paleogeography (short overview). *Geol. Carpath.* 50, 339–349.

Rohling, E. J., Foster, G. L., Grant, K. M., Marino, G., Roberts, A. P., Tamisiea, M. E., et al. (2014). Sea-level and deep-sea-temperature variability over the past 5.3 million years. *Nature* 508, 477–482. doi: 10.1038/nature13230

Róis, A. S., Sádio, F., Paulo, O. S., Teixeira, G., Paes, A. P., Espírito-Santo, D., et al. (2016). Phylogeography and modes of reproduction in diploid and tetraploid halophytes of *Limonium* species (Plumbaginaceae): evidence for a pattern of geographical parthenogenesis. *Ann. Bot.* 117, 37–50. doi: 10.1093/aob/mcv138

Romeiras, M. M., Monteiro, F., Duarte, M. C., Schaefer, H., and Carine, M. (2015). Patterns of genetic diversity in three plant lineages endemic to the Cape Verde Islands. *AoB Plants* 7, 1–11. doi: 10.1093/aobpla/plv051

Romeiras, M. M., Vieira, A., Silva, D. N., Moura, M., Santos-Guerra, A., Batista, D., et al. (2016). Evolutionary and biogeographic insights on the Macaronesian *Beta-Patellifolia* species (Amaranthaceae) from a time-scaled molecular phylogeny. *PLoS One* 11:e0152456. doi: 10.1371/journal.pone.0152456

Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* 46, 195–203. doi: 10.1093/sysbio/46.1.195

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, L., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Rosenbaum, G., Lister, G. S., and Duboz, C. (2002). Relative motions of Africa, Iberia and Europe during alpine orogeny. *Tectonophysics* 359, 117–129. doi: 10.1016/S0040-1951(02)00442-0

Rutschmann, F., Eriksson, T., Salim, K. A., and Conti, E. (2007). Assessing calibration uncertainty in molecular dating: the assignment of fossils to alternative calibration points. *Syst. Biol.* 56, 591–608. doi: 10.1080/10635150701491156

Salvo, G., Ho, S. Y., Rosenbaum, G., Ree, R., and Conti, E. (2010). Tracing the temporal and spatial origins of island endemics in the Mediterranean region: a case study from the citrus family (*Ruta* L., Rutaceae). *Syst. Biol.* 59, 705–722. doi: 10.1093/sysbio/syq046

Sanmartín, I., Van Der Mark, P., and Ronquist, F. (2008). Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *J. Biogeogr.* 35, 428–449. doi: 10.1111/j.1365-2699.2008.01885.x

Schenk, J. J. (2016). Consequences of secondary calibrations on divergence time estimates. *PLoS One* 11:e0148228. doi: 10.1371/journal.pone.0148228

Sharma, R., and Bhat, V. (2020). "Role of Apomixis in Perpetuation of Flowering Plants: Ecological Perspective" in *Reproductive Ecology of Flowering Plants: Patterns and Processes*. eds. R. Tandon, K. Shivanna and M. Koul (Singapore: Springer), 275–297.

Soltis, P. S., Folk, R. A., and Soltis, D. E. (2019). Darwin review: angiosperm phylogeny and evolutionary radiations. *Proc. R. Soc. B* 286:20190099. doi: 10.1098/rspb.2019.0099

Soltis, D. E., Mort, M. E., Latvis, M., Mavrodiev, E. V., O'Meara, B. C., Soltis, P. S., et al. (2013). Phylogenetic relationships and character evolution analysis of Saxifragales using a supermatrix approach. *Am. J. Bot.* 100, 916–929. doi: 10.3732/ajb.1300044

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Suc, J. P. (1984). Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature* 307, 429–432. doi: 10.1038/307429a0

Thompson, J. D. (2005). *Plant evolution in the Mediterranean*. New York: Oxford University Press.

Triantis, K. A., Borges, P. A., Hortal, J., and Whittaker, R. J. (2010). "The Macaronesian province: patterns of species richness and endemism of arthropods" in *Terrestrial arthropods of Macaronesia-biodiversity, ecology and evolution*. eds. A. R. M. Serrano, P. A. V. Borges, M. Boieiro and P. Oromí (Lisboa: Sociedade Portuguesa de Entomologia), 49–71.

Valente, L. M., Savolainen, V., and Vargas, P. (2010). Unparalleled rates of species diversification in Europe. *Proc. R. Soc. B* 277, 1489–1496. doi: 10.1098/rspb.2009.2163

Van Campo, É. (1976). La flore sporopollinique du gisement Miocène terminal de Venta del Moro. dissertation. Montpellier: Université Montpellier II Sciences et Techniques du Languedoc.

Van Hinsbergen, D. J., Torsvik, T. H., Schmid, S. M., Maţenco, L. C., Maffione, M., Vissers, R. L., et al. (2020). Orogenic architecture of the Mediterranean region and kinematic reconstruction of its tectonic evolution since the Triassic. *Gondwana Res.* 81, 79–229. doi: 10.1016/j.gr.2019.07.009

Vargas, P., Fernández-Mazuecos, M., and Heleno, R. (2018). Phylogenetic evidence for a Miocene origin of Mediterranean lineages: species diversity, reproductive traits and geographical isolation. *Plant Biol.* 20, 157–165. doi: 10.1111/plb.12626

Wallace, A. R. (1878). *Tropical nature and other essays*. London: MacMillan & Co.

Warren, B. H., Hagen, O., Gerber, F., Thébaud, C., Paradis, E., and Conti, E. (2018). Evaluating alternative explanations for an association of extinction risk and evolutionary uniqueness in multiple insular lineages. *Evolution* 72, 2005–2024. doi: 10.1111/evo.13582

Warren, B. H., Simberloff, D., Ricklefs, R. E., Aguilée, R., Condamine, F. L., Gravel, D., et al. (2015). Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecol. Lett.* 18, 200–217. doi: 10.1111/ele.12398

Zachos, J. C., Dickens, G. R., and Zeebe, R. E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451, 279–283. doi: 10.1038/nature065

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 February 2021
doi: 10.3389/fpls.2021.631178

Evolution in the Model Genus *Antirrhinum* Based on Phylogenomics of Topotypic Material

Ana Otero, Mario Fernández-Mazuecos and Pablo Vargas*

Real Jardín Botánico (RJB-CSIC), Madrid, Spain

Researchers in phylogenetic systematics typically choose a few individual representatives of every species for sequencing based on convenience (neighboring populations, herbarium specimens, samples provided by experts, garden plants). However, few studies are based on original material, type material or topotypic material (living specimens from the locality where the type material was collected). The use of type or topotypic material in phylogenetic studies is paramount particularly when taxonomy is complex, such as that of *Antirrhinum* (Plantaginaceae). In this paper, we used topotypic materials of *Antirrhinum* at the species level (34 species proposed by previous authors), 87 specimens representing the species distributions and >50,000 informative nucleotide characters (from ∼4,000 loci) generated by the genotyping-by-sequencing (GBS) technique: (i) to test two explicit taxonomic hypotheses widely followed by local taxonomic treatments; (ii) to robustly estimate phylogenetic relationships; (iii) to investigate the evolution of key morphological characters and biogeographic centers of differentiation. Two GBS phylogenies based on two datasets (87 localities and 34 topotypic specimens) revealed that: (1) Sutton's (1988) taxonomic account is the most congruent with phylogenetic results, whereas division of *Antirrhinum* into three major clades disagrees with Rothmaler's (1956) infrageneric classification; (2) monophyly of populations currently included in the same species is primarily supported; (3) the historically recognized *Antirrhinum majus* group is not monophyletic; (4) sister-group relationships are robust for eight species pairs; (5) the evolutionary radiation of 26 species since the Pliocene is underpinned given a high rate of diversification (0.54 spp. Myr^{-1}); (6) a geographic pattern of speciation is reconstructed, with northern Iberia as the center of early diversification followed by more recent speciation in southeastern Iberia; and (7) multiple acquisitions of key taxonomic characters in the course of *Antirrhinum* diversification are strongly supported, with no evidence of hybridization between major clades. Our results also suggest incipient speciation in some geographic areas and point to future avenues of research in evolution and systematics of *Antirrhinum*.

Keywords: biogeography, genotyping-by-sequencing, high-throughput sequencing, *locus classicus*, molecular dating, phylogeny, Plantaginaceae, systematics

OPEN ACCESS

Edited by:
Andrew A. Crowl,
Duke University, United States

Reviewed by:
Ezgi Ogutcen,
Université de Genève, Switzerland
Sonia Herrando,
Botanical Institute of Barcelona, Spain

****Correspondence:***
Ana Otero
aotero@rjb.csic.es;
anaoterogomez@gmail.com

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

Received: *19 November 2020*
Accepted: *06 January 2021*
Published: *12 February 2021*

Citation:
Otero A, Fernández-Mazuecos M and Vargas P (2021) Evolution in the Model Genus Antirrhinum Based on Phylogenomics of Topotypic Material. Front. Plant Sci. 12:631178. doi: 10.3389/fpls.2021.631178

INTRODUCTION

In botanical systematics and taxonomy, validly published species names are based on type materials, which are deposited in reference collections (herbaria). This way, any species is anchored to a single name and type material, i.e., a single specimen designated by taxonomists following the rules of the *International Code of Nomenclature for Algae, Fungi, and Plants*[1]. As a result, researchers can name any individuals or populations based on key characters contained in the type specimen, with the assistance of any other original materials.

Plant taxonomists have historically found two significant patterns: (1) not all plant groups have clear-cut characters circumscribing populations into species; and (2) unrelated species share similar morphological characters that have been independently acquired in the course of evolution (convergence and parallel evolution). The advent of molecular phylogenetics helped to tackle these two taxonomic obstacles. Taxonomy takes advantage of DNA sequence data not only for improving species circumscription, but also for species-level identification with the development of DNA barcoding (Li et al., 2015). In contemporary taxonomic studies, DNA sequences are frequently included to validate new species, whereas many historical specimens do not have sequences associated. These historical specimens include original material (the very specimens used for the first description of a species) and type material (physical specimen that serves as exemplar for any formal species name, preferentially selected from among the original material). However, sequencing of original and type materials usually faces two problems: (i) tissue destruction during DNA extraction, which may cause permanent damage to unique specimens (herbarium policies); and (ii) high DNA fragmentation because of specimen age, which hinders sequencing (Staats et al., 2011). Consequently, the utility of old herbarium specimens (including original and type materials) in phylogenetic studies is limited. To circumvent these two problems, an alternative approach relies on topotypic specimens, i.e., those collected at the type locality of a species, which usually corresponds to the *locus classicus* (the locality where plants were originally collected and studied for description of a new species). Topotypic specimens are expected to retain the genetic identity of the type material better than plants collected in any other locality. Unfortunately, researchers do not commonly use topotypic specimens, but rather identify species of a natural group (e.g., a genus) using the most recent taxonomic studies and then choose a few individual representatives of each species for sequencing based on convenience (neighboring populations, herbarium specimens, samples provided by experts, and garden individuals). Most phylogenetic reconstructions unfortunately ignore that any species name is ultimately linked to the type material, which is particularly critical when taxonomy is complex. In other words, to which degree can we rely on species-level phylogenetic relationships based on sequenced specimens whose genetic makeup may differ significantly from that of the type materials?

[1]https://www.iapt-taxon.org/nomen/main.php

Taxonomy is particularly complex for plant groups actively evolving in biodiversity hotspots such as the Mediterranean region (Vargas et al., 2018). One of these groups is the model genus *Antirrhinum* L. (snapdragons), of which around 40 species have been historically proposed to circumscribe morphological differentiation of populations primarily distributed throughout the western Mediterranean. Published studies have repeatedly failed to obtain a well-supported phylogenetic structure using Sanger sequencing of nuclear and plastid DNA regions (Vargas et al., 2004, 2009; Carrió et al., 2010). A more resolved phylogeny was obtained by analyzing AFLP data, but species delimitation was still largely unresolved (Wilson and Hudson, 2011). A phylogenomic analysis based on genome-wide data is, however, promising to disclose evolutionary relationships among populations and species.

In this study, we investigated the phylogenetic structure of snapdragons (*Antirrhinum* species) by analyzing genome-wide genotyping-by-sequencing (GBS) data obtained from topotypic material of most species (36 specimens). In addition, materials of widespread species were collected from distant localities covering species distributions to test the monophyly of currently recognized taxa of *Antirrhinum*. Therefore, the main aim was to obtain a solid species-level phylogeny using specimens that are strongly linked to valid names. To this end, we tested two explicit taxonomic hypotheses proposed by the two worldwide taxonomic accounts of *Antirrhinum* published to date (Rothmaler, 1956; Sutton, 1988). We expected that species relationships and clade composition would help to propose a more consistent taxonomic and evolutionary framework for *Antirrhinum* at the species and supra-specific levels. Specific objectives were: (1) to analyze phylogenetic relationships of the populations and species described by previous authors; (2) to reconstruct primary and secondary centers of species diversification in a spatio-temporal framework; (3) to investigate the evolution of key morphological characters; and (4) to test the radiation hypothesis (abundant speciation from a common ancestor in a short period of time) proposed by previous studies.

MATERIALS AND METHODS

Study System

Snapdragons (species of *Antirrhinum*) have been considered a model system for the study of plant genetics, development, and evolution since the 20th century because of their easy cultivation and morphological diversity (Schwarz-Sommer et al., 2003; Fernández-Mazuecos and Glover, 2017). The complexity of *Antirrhinum* systematics has historically been ascribed to the recent diversification of the genus, putatively accompanied by hybridization and introgression. In fact, dissimilar taxonomic treatments exist, of which those of Rothmaler (1956) and Sutton (1988) are the most complete to date. Rothmaler (1956) divided *Antirrhinum* into two sections: sect. *Saerorrhinum* and sect. *Antirrhinum*. Sect. *Saerorrhinum* comprised North American plants that are currently included in the genus *Sairocarpus* and other related genera, while sect. *Antirrhinum* comprised all the species currently included in the genus

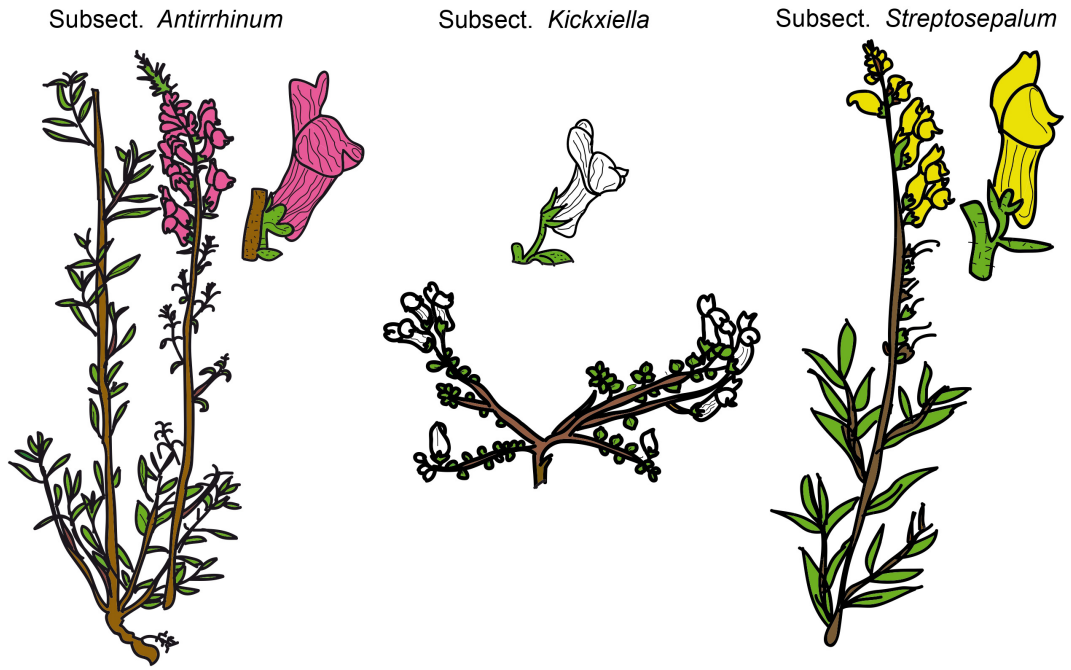


FIGURE 1 | Schematic drawings of the three morphotypes of *Antirrhinum* based on the Rothmaler (1956) subsections. Drawings have been adapted from Güemes (2009).

Antirrhinum. Rothmaler (1956) also split sect. *Antirrhinum* into three subsections: subsect. *Antirrhinum*, subsect. *Kickxiella* and subsect. *Streptosepalum*. These three subsections represent three major morphotypes/ecotypes: (1) species in subsect. *Antirrhinum* grow in sandy soils and have an upright tall habit, long thin leaves with no hairs and magenta or yellow flowers; (2) species in subsect. *Kickxiella* inhabit rocks and are small, prostrate, xerophytic woody plants; and (3) species in subsect. *Streptosepalum* occur on rocky substrates and have long thin leaves, glandular hairs only on the inflorescence and large yellow flowers (**Figures 1**, **2**; Wilson and Hudson, 2011). The Iberian Peninsula appears to be the center of diversification of snapdragons (Vargas et al., 2014), although the distribution range of *Antirrhinum* encompasses most of Mediterranean Europe, northern Africa and the Middle East (Sutton, 1988).

Specimen Sampling

The sample was designed to obtain the phylogenetic structure of *Antirrhinum* based on the majority of the taxa validly described at the species level. A total of 32 of the 34 species included in the taxonomic treatments of Rothmaler (1956) and Sutton (1988) were sampled. Three more species recognized by specialists after Sutton's (1988) account, were sampled: *A. onubensis* (Fernández-Casas, 1987), *A. subbaeticum* (Güemes et al., 1993), and *A. rothmaleri* (García-Barriuso et al., 2011) (see Güemes, 2009; **Table 1** and **Supplementary Table 1**, in Supporting Information). The only two species in Rothmaler (1956) and Sutton (1988) of which no material was found were *A. ambiguum* Lange and *A. martenii* (Font Quer) Rothm. In the *locus classicus* of the former species (El Escorial, Madrid) we collected some snapdragons, but they had morphological characters more similar to those of *A. graniticum* than to the description (protologue) of *A. ambiguum*. Unfortunately, *A. martenii* from Morocco has not been found since description (1956), including unsuccessful collecting campaigns in recent times (E. Carrió pers. com.).

The type species of the genus (*A. majus*) was described from cultivated specimens (Linnaeus, 1753), and thus it is not possible to localize its original source with certainty. Sutton (1988) indicates that the most characteristic morphological features of *A. majus* are found in wild populations of southern France and north-eastern Spain, where we sampled three specimens (see **Supplementary Table 1**). For the remaining species, a minimum of two topotypic specimens per species were sampled. When we failed to sample at the *locus classicus* itself, plants from nearby locations were collected. To test monophyly of each species, additional populations (from two to five depending on species range size) were sampled from distant locations. Based on distribution ranges reviewed by Vargas et al. (2009), three range size categories were considered: (I) widely distributed species (four to six samples); (II) moderately distributed species (three to five samples); and (III) narrowly distributed species (two to three samples) (see **Table 1**). As a result, 108 samples of *Antirrhinum* were initially included as the ingroup. Ten samples of genera belonging to the sister clade to *Antirrhinum* (Fernández-Mazuecos et al., 2019) were sampled as the outgroup (**Supplementary Table 1**). Most materials were collected in multiple field campaigns (2005–2019) or obtained from herbarium specimens (JACA, LEB, MA, MGC, PI, PO, and SALA).

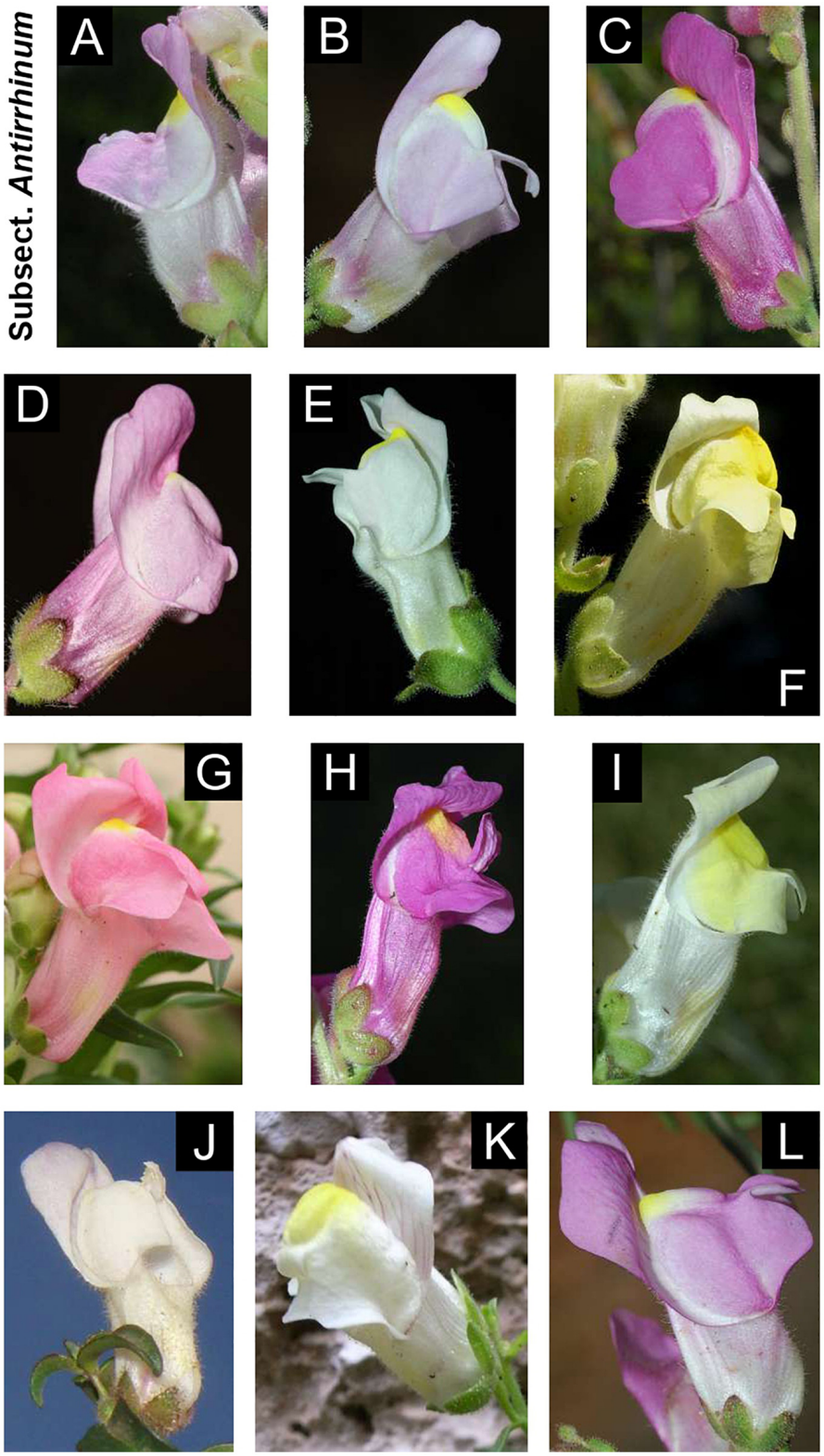


FIGURE 2 | Continued

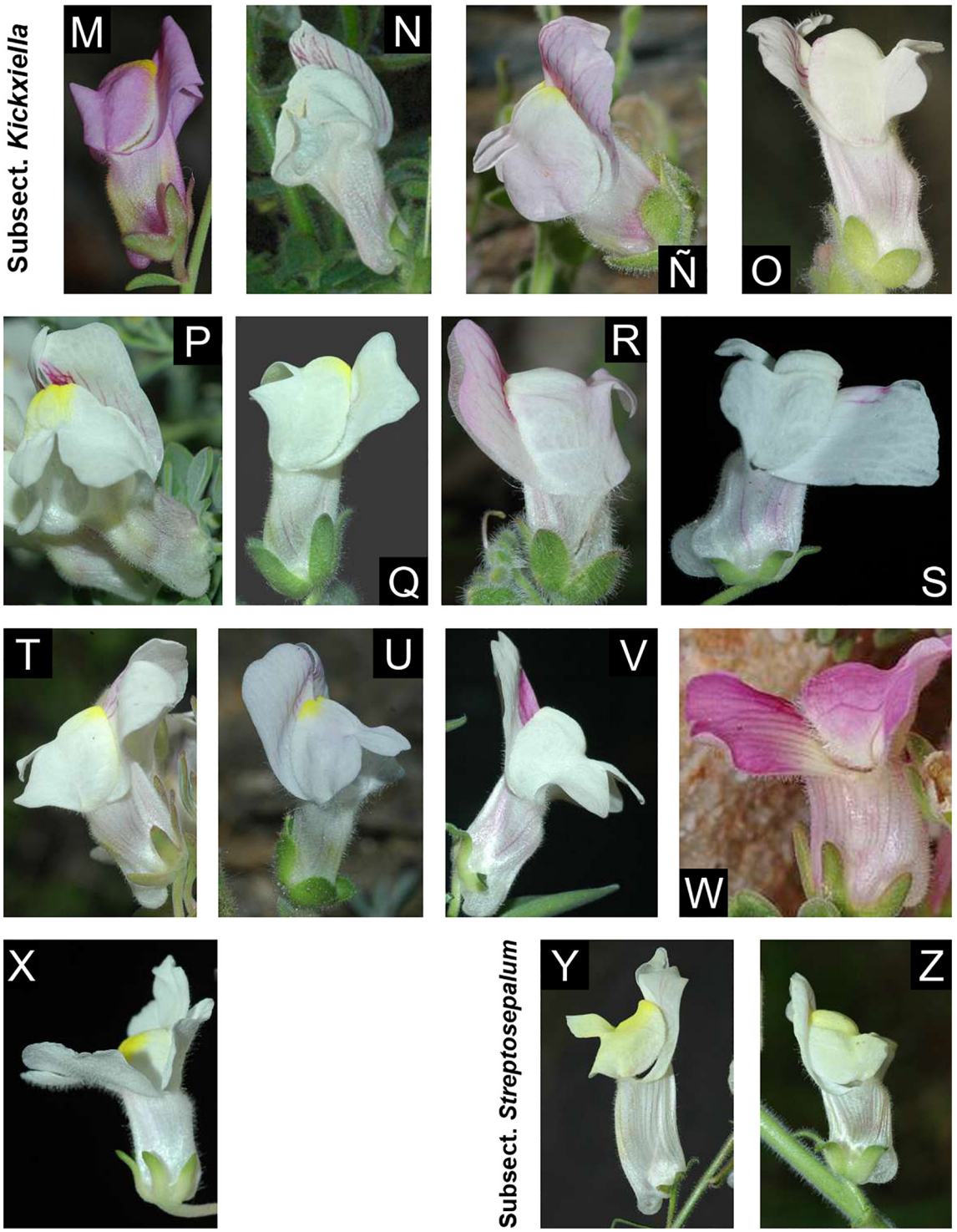


FIGURE 2 | Diversity of flower color and shape in the 26 recognized species in this study. The photographs are arranged by subsections: subsect. *Antirrhinum* **(A–L)**; subsect *Kickxella* **(M–X)**; and subsect. *Streptosepalum* (Y-Z). Photographs: **(A)** *A. australe*; **(B)** *A. barrelieri*; **(C)** *A. cirrhigerum*; **(D)** *A. controversum*; **(E)** *A. graniticum* subsp. *graniticum*; **(F)** *A. latifolium*; **(G)** *A. linkianum*; **(H)** *A. majus* subsp. *majus*; **(I)** *A. majus* subsp. *striatum*; **(J)** *A. onubensis*; **(K)** *A. siculum*; **(L)** *A. tortuosum*; **(M)** *A. charidemi*; **(N)** *A. grosii*; **(Ñ)** *A. hispanicum*; **(O)** *A. lopesianum*; **(P)** *A. microphyllum*; **(Q)** *A. molle*; **(R)** *A. mollissimum*; **(S)** *A. pertegasii*; **(T)** *A. pulverulentum*; **(U)** *A. rupestre*; **(V)** *A. sempervirens*; **(W)** *A. subbaeticum*; **(X)** *A. valentinum*; **(Y)** *A. braun-blanquetii*; and **(Z)** *A. meonanthum*. All photographs were taken in Spain by Pablo Vargas, except for **(F)** van der Strate, Saxifraga Foundation; **(G)** (Luis Nunes; Wikipedia); **(K)** (Denis Barthel; Wikipedia), and **(W)** (José Quiles; www.florasilvestre.es). Locations for each photograph is detailed on **Supplementary Material** (see **Supplementary Data Sheet 1**).

TABLE 1 | Species considered for this study, including species names from the two worldwide accounts of *Antirrhinum* (Rothmaler, 1956; Sutton, 1988) and new species described after Sutton (1988).

Species considered in this study	Species considered in Rothmaler (1956)	Species considered in Sutton (1988)	Valid names after Sutton (1988)	*Locus classicus*	Sampled Material from *locus classicus*: Voucher number	Distance in km from sampled material to *Locus classicus*)	Geographic distribution (geographic range size: I, II or III)	Total number of specimens sampled	Main phylogenetic clade (yes/no species monophyly)	Ecology
A. ambiguum Lange	*A. ambiguum* Lange (2)	–	–	Castella Nova supra Escorial, Cerro Las Machotas, Madrid (Spain)	Not available	Not available	Sierra de Guadarrama, Serra da Estrela (I)	Not available	Not available	Rocky substrates
A. australe Rothm.	*A. australe* Rothm. (3)	*A. australe* Rothm.	–	Benacoaz, Cádiz (Spain)	3IML12(1) (2)	*	SE Spain (II)	4	III (yes)	Sandy soils
A. barrelieri Boreau	*A. majus* subsp. *litigiosum* (Pau) Rothm. (3)	*A. majus* subsp. *litigiosum* (Pau) Rothm.	–	Calatayud, Zaragoza (Spain)	141PV08(4)	Nuévalos (20.37 km)	NE Spain (II)	5	III (yes)	Sandy soils
A. boissieri Rothm.	*A. boissieri* Rothm. (3)	*A. hispanicum* Chav.	–	Granada, Silla del Moro (Spain)	6IML13(1) (2)	*	SE Spain (I)	3	III (yes)	Sandy soils
A. braun-blanquetii Rothm.	*A. braun-blanquetii* Rothm. (2)	*A. braun-blanquetii* Rothm.	–	La Guiana, Dolomitklippen de Apóstoles, León (Spain)	MA777373/ MA345884	*	N Iberian Peninsula (II)	6	II (yes)	Rocky substrates
A. caroli-paui Rothm.	*A. caroli-paui* Rothm. (1)	*A. molle* var. *marianum* Pau	–	Doña María de Ocaña, Almería (Spain)	53PV07	Abrucena (10 Km)	SE Spain (I)	2	III (–)	Rocky substrates
A. charidemi Lange	*A. charidemi* Lange (1)	*A. charidemi* Lange	–	Promontorio Charidemi, Cabo de Gata, Almería (Spain)	PV05	*	SE Spain (I)	2	III (–)	Rocky substrates
A. cirrhigerum (Welv. ex Ficalho) Rothm.	*A. linkianum* var. *ramosissimum* Willk	*A. majus* subsp. *cirrhigerum* (Welv. ex Ficalho) Franco	–	Int. Sines et Milfontes (Portugal)	3AO19(7)	*	SW Iberian Peninsula, NW Norocco (II)	3	III (–)	Sandy soils
A. controversum Pau	*A. barrelieri* Boreau (3)	–	–	Sierra de Cártama, Málaga (Spain)	MA855349	Pizarra (9 Km)	S Spain (I)	6	III (yes)	Sandy soils
A. dielsanum Rothm.	*A. dielsanum* Rothm. (3)	–	–	Siracusa, Sicilly (Italy)	MA646409	*	Sicilly (Italy) (I)	1	III (–)	Sandy soils

(Continued)

TABLE 1 | Continued

Species considered in this study	Species considered in Rothmaler (1956)	Species considered in Sutton (1988)	Valid names after Sutton (1988)	*Locus classicus*	Sampled Material from *locus classicus*: Voucher number	Distance in km from sampled material to *Locus classicus*)	Geographic distribution (geographic range size: I, II or III)	Total number of specimens sampled	Main phylogenetic clade (yes/no species monophyly)	Ecology
A. graniticum Rothm. subsp. *graniticum*	*A. graniticum* Rothm. (3)	*A. graniticum* Rothm.	–	Castelo Branco, Lardosa prope. Soalheira (Portugal)	1AO19(2)/ 2AO19(2)	*	E Portugal to C Spain (III)	6	III (yes)	Sandy soils
A. graniticum Rothm. subsp. *brachycalyx* D. A. Sutton	–	*A. graniticum* Rothm. subsp. *brachycalyx* D. A. Sutton	–	Madrid, near Valdemoro (Spain)	104PV19(14)	*	Madrid, Toledo (Spain) (I)	2	III (–)	Sandy soils
A. grosii Font Quer	*A. grosii* Font Quer (1)	*A. grosii* Font Quer	–	Sierra de Gredos, Riscos del Morezón, Ávila (Spain)	133PV10(1)/ 138PV10(7)	*	Sierra de Gredos, Ávila (Spain) (I)	2	II (yes)	Rocky substrates
A. hispanicum Chav.	*A. hispanicum* Chav. (1)	*A. hispanicum* Chav.	–	Hispania Tournefort s.n (holo. P-Tournefort, fragment JE!	19IML12(2)/ 7IML13(3)	*	SE Spain (II)	2	III (no)	Rocky substrates
A. latifolium Mill.	*A. majus subsp. latifolium* (Mill.) Rouy (3)	*A. latifolium* Mill.	–	In ins Archipielagi Toscana (Italy)	PI009621	*	Slovenia and C Italy (III)	4	III (yes)	Sandy soils
A. linkianum Boiss. & Reut.	*A. majus* subsp. *linkianum* (Boiss. & Reut.) Rothm. (3)	*A. majus* subsp. *linkianum* (Boiss. & Reut.) Rothm.	–	Ad sepes Olyssoponi (Sintra, Lisbon, Portugal)	3MF13(2) (5)	*	NW and CW Portugal (II)	3	III (yes)	Sandy soils
A. lopesianum Rothm.	*A. lopesianum* Rothm. (1)	*A. lopesianum* Rothm.	–	Vimioso prope Argoselo (Portugal)	MA824747	*	NE Portugal, NW Spain (I)	4	II (yes)	Rocky substrates
A. majus L.	*A. majus* subsp. *majus* (3)	*A. majus* L.	–	Eur. Austr. (Linné).	46PV12(1)	*	SW Europe and Mediterranean region (widely planted as ornamental) (II)	5	III (no)	Sandy soils
A. martenii (Font Quer) Rothm.	*A. martenii* (Font Quer) Rothm.	*A. martenii* (Font Quer) Rothm.	–	Rupibus Bar-er-Ruida, (Morocco)	Not available	Not available	N Morocco (I)	Not available	Not available	Rocky substrates

(Continued)

TABLE 1 | Continued

Species considered in this study	Species considered in Rothmaler (1956)	Species considered in Sutton (1988)	Valid names after Sutton (1988)	*Locus classicus*	Sampled Material from *locus classicus*: Voucher number	Distance in km from sampled material to *Locus classicus*)	Geographic distribution (geographic range size: I, II or III)	Total number of specimens sampled	Main phylogenetic clade (yes/no species monophyly)	Ecology
A. meonanthum Hoffmanns. & Link	*A. meonanthum* Hoffmanns. & Link (2)	*A. meonanthum* Hoffmanns. & Link	–	Lusitania, ad ripas Durii (Douro River near Oporto)	PO60051	Cinfães (35 Km)	N Portugal, W and C Spain (II)	5	II (no)	Rocky substrates
A. microphyllum Rothm.	*A. microphyllum* Rothm. (1)	*A. microphyllum* Rothm.	–	Sacedón, Guadalajara (Spain)	39PV08(2)/ 40PV08(2)	*	NE Spain (I)	2	I (yes)	Rocky substrates
A. molle L.	*A. molle* L. (1)	*A. molle* L.	–	Pyrenees (Spain, France)	MA895768/ MA756423	*	NE Spain, Pyrenees and adjoining mountains (II)	3	III (yes)	Rocky substrates
A. mollissimum (Pau) Rothm.	*A. mollissimum* (Pau) Rothm. (1)	*A. mollissimum* (Pau) Rothm.	–	Barranco del Caballar, Almería (Spain)	73PV06	*	SE Spain (I)	3	III (yes)	Rocky substrates
A. onubensis (Fern. Casas) Fern. Casas	–	–	*A. onubensis* (Fern. Casas) Fern. Casas (Fernández-Casas, 1987)	Sierra Aracena, Huelva (Spain)	104PV09/ 146PJM13(3)	*	Sierra Aracena, Huelva (Spain) (I)	2	III (yes)	Rocky substrates
A. pertegasii Pau ex Rothm.	*A. pertegasii* Pau ex Rothm. (1)	*A. pertegasii* Pau ex Rothm.	–	Mont Caro, Montes de Tortosa, Tarragona (Spain)	36PV11/ 38PV11(12)	La Senia (18.81 Km)/ La Pobla de Benifassa (22.86 km)	E Spain (I)	2	I (yes)	Rocky substrates
A. pulverulentum Lázaro Ibiza	*A. pulverulentum* Lázaro Ibiza (1)	*A. pulverulentum* Lázaro Ibiza	–	Monasterio de Piedra, Zaragoza (Spain)	15IML12(4)/ (5)	*	E Spain (II)	3	I (no)	Rocky substrates
A. rothmalieri (Pinto da Silva) Amich, Bernardos & García-Barriuso	–	–	*A. rothmalieri* (Pinto da Silva) Amich, Bernardos & García-Barriuso (García-Barriuso et al., 2011)	Macedos de Cavaleiros (Portugal)	12IML12(4) (2)	*	NW Iberian Peninsula (I)	2	II (–)	Sandy soils

(Continued)

TABLE 1 | Continued

Species considered in this study	Species considered in Rothmaler (1956)	Species considered in Sutton (1988)	Valid names after Sutton (1988)	*Locus classicus*	Sampled Material from *locus classicus*: Voucher number	Distance in km from sampled material to *Locus classicus*)	Geographic distribution (geographic range size: I, II or III)	Total number of specimens sampled	Main phylogenetic clade (yes/no species monophyly)	Ecology
A. rupestre Boiss. & Reut	*A. rupestre* Boiss. & Reut (1)	*A. hispanicum* Chav.	–	Sierra Nevada, Granada (Spain)	9IML13(3) (5)	*	SE Iberian Peninsula (I)	2	III (–)	Rocky substrates
A. sempervirens Lapeyr.	*A. sempervirens* Lapeyr. (1)	*A. sempervirens* Lapeyr.	–	Hautes Pyreneés, Gèdre (France)	104PV10(7)	Panticosa (Spain) (25.77 Km)	Pyrenees (Spain, France) (II)	4	I (–)	Rocky substrates
A. siculum Mill.	*A. siculum* Mill. (3)	*A. siculum* Mill.	–	Palermo, Sicily (Italy)	MA705546/ 33bisPV2015(1)	*	S Italy and Sicily (III)	4	III (yes)	Sandy soils
A. striatum Lam.	*A. majus* subsp. *striatum* (DC) Rothm. (3)	*A. latifolium* subsp. *intermedium* (Debeaux) Nyman	*A. latifolium* var. striatum DC (Güemes, 2009)	Perpignan (France)	55PV07(1)	Limoes (France) (66.54 km)	SE France and NE Spain (I)	2	III (–)	Sandy soils
A. subbaeticum Güemes, Mateu & Sánchez Gómez	–	–	*A. subbaeticum* Güemes, Mateu & Sánchez Gómez (Güemes et al., 1993)	Bogarra, Albacete (Spain)	MA705104/ MA593205	*	SE Spain (I)	2	I (yes)	Rocky substrates
A. tortuosum Bosc ex Lam.	*A. majus* subsp. *tortuosum* (Bosc ex Vent.) Rouy (3)	*A. majus* subsp. *tortuosum* (Bosc ex Vent.) Rouy	–	Described from material cultivated at Paris of Italian origin	MA589750/ PI010588	*	Mediterranean Region (III)	7	III (yes)	Sandy soils
A. valentinum Font Quer	*A. valentinum* Font Quer (1)	*A. valentinum* Font Quer	–	Mt Mondúber, supra Gandía, Valencia (Spain)	27PV11(6)/ 32PV11(3)	*	SE Spain (I)	2	I (yes)	Rocky substrates

The locus classicus for each species is transcribed from each original protologue. Numbers in brackets in the column for Rothmaler (1956) names indicate species subsections: (1) subsect. Kickxiella; (2) subsect. Streptosepalum; and (3) subsect. Antirrhinum. Voucher numbers and indication of locus classicus (asterisk) for each species or distance (in km) where samples were taken from the locus classicus are indicated in two more columns. Species distribution ranges categorized in three types are shown (I: restricted; II: moderate; III: wide), which help to define sampling intensity for each species (see Species sampling). Information about phylogenetic clade to which each species belongs and whether monophyly is retrieved [or dash line (–) when monophyly cannot be assessed] according to our phylogenetic results is also indicated. Plant collectors (AO, Ana Otero; IML, Isabel María Liberal; MF, Mario Fernández-Mazuecos; PV, Pablo Vargas; PJM, Pedro Jiménez Mejías) and herbaria (MA, Royal Botanical Garden of Madrid; PI, Pisa; PO, Porto Herbarium) that provided plant material. For details about Antirrhinum see www.rjb.csic.es/snapdragons/.

DNA Extraction and GBS Library Preparation

DNA was extracted from leaf tissue (c. 20 mg) using a modified CTAB protocol (Doyle and Doyle, 1987; Cullings, 1992) and extractions were quantified at the Next Generation Sequencing Lab (Real Jardín Botánico, CSIC, Madrid, Spain) using a Qubit Fluorometer (Thermo Fisher, Waltham, MA, United States). The 118 genomic DNA samples (ingroup plus outgroup samples, 500 ng of DNA per sample when available) were sorted into two 96-well plates, and four samples per plate were replicated in order to evaluate possible biases derived from the lab technique, thus raising the total number of samples to 126. These were used to prepare two separate GBS libraries using the *Pst*I-HF restriction enzyme and following the previously published procedures of Fernández-Mazuecos et al. (2018), which were adapted from Elshire et al. (2011), Escudero et al. (2014), and Grabowski et al. (2014) with some modifications. The 500 ng of genomic DNA per sample were combined with 0.6 pmol of a sample specific barcode adapter and 0.6 pmol of common adapter. Four units of *Pst*I-HF (NEB, MA, United States) were used for the digestion step at 37°C overnight. The ligation step was done using 400 units of T4 DNA Ligase (NEB, MA, United States) at room temperature for 4 h. Then, 50 ng of each sample were pooled and purified with Agencourt AMPure XP beads (Beckman Coulter, CA, United States). DNA pools of each library were amplified for 19 PCR cycles using NEB 2x Taq MasterMix (NEB, MA, United States). The PCR products were purified using different ratios of AMPure beads. Concentration and fragment sizes were assessed in a 2100 Bioanalyzer (Agilent Technologies, CA, United States) at the genomic facilities of Real Jardín Botánico (CSIC, Madrid, Spain). The optimal purification ratio of AMPure beads (1:1) yielded a concentration of 2.13 and 2.63 ng/μl, with an averaged fragment size of 498.5 and 561 bp for each library, respectively. Both libraries were submitted to Macrogen Inc. (Seoul, South Korea) for 150 bp HiSeqX Illumina paired-end sequencing. Quality control of sequencing results was conducted in FastQC 0.11.7 (Andrews, 2010).

Data Assembly

The FASTQ sequence files from the two libraries were processed in ipyrad v.0.9.4 (Eaton and Overcast, 2020). Ipyrad is an assembly pipeline implementing seven sequential steps that have been specifically developed for the processing of high-throughput sequencing results derived from restriction digest-based methods such as GBS. We first demultiplexed and filtered reads from both libraries separately (steps one and two) using the default threshold of 33 quality score for filtering of reads, and then combined all samples from both libraries for subsequent assembly steps. Data were assembled using the reference mapping method, incorporating BWA and Bedtools algorithms (Li and Durbin, 2009; Quinlan and Hall, 2010), in which GBS reads are mapped to a reference genome to determine homology and all sequences that do not match the reference genome are discarded. As reference, we used the genome of the cultivar line JI7 of *A. majus* published by Li et al. (2019). Samples with low-quality results (<100 sequenced loci in preliminary assemblies) were discarded, leading to a total of 87 samples in the final datasets (trimmed dataset). Additional datasets were produced including topotypic specimens only (topotypic dataset, 38 samples; see **Table 2**). To assess the sensitivity of the results to data matrix completeness and number of loci, we tested four values of minimum taxon coverage, i.e., the minimum number of sequenced samples for a locus to be considered in the final matrix (m4, m6, m18, and m36). Thus, a total of eight assembled datasets were produced (see **Table 2**).

Phylogenomic Analyses

Maximum likelihood (hereafter ML) analyses of the eight matrices of concatenated loci were performed through RAxML BlackBox (Stamatakis et al., 2008) by using the GTRCAT model of nucleotide substitution and automatic bootstrap stopping as recommended in CIPRES Portal (Miller et al., 2010). Among the four topologies obtained for each dataset (trimmed and topotypic dataset), we chose the one with the highest average bootstrap support (BS) for downstream phylogeny-based analyses. The matrices producing the best RAxML trees were also analyzed using Bayesian inference in ExaBayes 1.5 (Aberer et al., 2014), implemented in the CIPRES. We specified two runs, four coupled chains, one million generations with a sampling frequency of 500 and a burn-in proportion of 0.1. Convergence and effective sample size (ESS) values for all parameters were assessed using Tracer v.1.6 (Rambaut et al., 2014).

We additionally implemented the coalescent-based method SVDquartets (Chifman and Kubatko, 2014) in PAUP

TABLE 2 | Parameter combination of nucleotides assembled GBS datasets generated in ipyrad and used in the phylogenetic analyses.

Data set	Minimum taxon coverage	Number of loci	Concatenated length (bp)	Missing data (%)	Number of parsimony-informative sites
Trimmed dataset: 87 taxa					
refm4	4	6524	1696692	78.8	52991
refm6	6	4663	1276470	73	51140
refm18	18	2707	807146	61.7	42203
refm36	36	1989	626043	56.8	35381
Topotypic dataset: 38 taxa					
ref_clas_m4	4	5240	1262819	64	39093
ref_clas_m6	6	4016	1008597	57.3	37713
ref_clas_m18	18	2344	632938	43.8	30014
ref_clas_m36	36	237	78284	32.6	3969

(Swofford, 2001). Individuals were grouped according to current species circumscriptions. All possible quartets were evaluated with 100 bootstrap replicates.

Estimates of Divergence Times

We used penalized likelihood as implemented in TreePL (Smith and O'Meara, 2012) to estimate a time-calibrated tree for the topotypic dataset using the best RAxML tree. TreePL is suitable for divergence time estimation when dealing with large amounts of data, such as those yielded by GBS (Zheng and Wiens, 2015). The tree was pruned to include a single topotypic individual per species of *Antirrhinum* by choosing the individual with the highest number of loci retrieved from the assembly. Two calibration points were used: (i) root age (minimum age = 11.0583; maximum age = 21.58182) and (ii) crown node of *Antirrhinum* (ingroup) (minimum age = 2.1168; maximum age = 5.6736) based on the 95% Highest Posterior Density (HPD) intervals inferred in Fernández-Mazuecos et al. (2019) and Gorospe et al. (2020). We first conducted an analysis under the "prime" option to select the optimal set of parameter values. Then we set the gradient-based (opt) optimizer to one, and autodifferentiation-based (optad) and autodifferentiation crossvalidation-based optimizers (optcvad) to two. Then, we ran a second analysis using random subsample and replicate crossvalidation (RSRCV) to identify the best value for the smoothing parameter. We ran the final analysis by setting the best chi-square value for the smoothing parameter (smoothing = 1e-129). For all runs, we used the thorough analysis option and set 200,000 iterations for penalized likelihood and 5,000 iterations for cross-validation. Based on estimated crown age, net diversification rate was calculated following the whole-clade method of Magallón and Sanderson (2001), implemented in the R package Geiger (Harmon et al., 2007).

Biogeographic Reconstruction

For biogeographic analyses we used the ultrametric tree resulting from TreePL after pruning the outgroup to avoid anomalous inferences of ancestral areas that may have resulted from the difference in sampling depth between outgroup and ingroup lineages. This approach also circumvents the potential effect of extinction between the outgroup and the ingroup expected after millions of years since the Oligocene (Gorospe et al., 2020). We additionally excluded four taxa originally described as different species but with limited phylogenetic and morphological distinctiveness that may have produced biogeographic bias: *A. rothmalieri*, *A. dielsianum, A. caroli-paui*, and *A. boissieri*. Areas for biogeographic reconstruction were based on the set of areas of endemicity for *Antirrhinum* proposed by Vargas et al. (2009). As a result, we considered six areas: (1) northwest of the Iberian Peninsula; (2) northeast of the Iberian Peninsula; (3) southwest of the Iberian Peninsula; (4) southeast of the Iberian Peninsula; (5) northern Africa; and (6) remaining circun-Mediterranean areas (non-Iberian Europe and SW Asia). We tested dispersal-extinction-cladogenesis (DEC, Ree and Smith, 2008) and dispersal-vicariance (DIVA, Ronquist, 1997) models using the "BioGeoBEARS" package (Matzke, 2013) in R (R Core Team, 2013) with no constraints in dispersal rates and adjacency between areas. In particular, the "BioGeoBEARS" package implements a likelihood interpretation of the parsimony-based DIVA method (DIVALIKE), i.e., processes are modeled as in DEC modeling, but including only the parameters considered in DIVA (widespread vicariance but not subset sympatry; Ronquist and Sanmartín, 2011). AIC was calculated for the two models although, given the inadequacy of likelihood comparison between different biogeographic models, DEC and DIVA results were discussed based on biological criteria (see Ree and Sanmartín, 2018).

Ancestral State Reconstruction

To evaluate the evolution of taxonomic characters traditionally used for *Antirrhinum* species delimitation, we selected four traits that best represent vegetative and reproductive variation used in dichotomous keys of the two main *Antirrhinum* monographs (Rothmaler, 1956; Sutton, 1988). Based on the consistent use of three reproductive characters in the two keys to *Antirrhinum* species, discrete traits were selected and codified as: (1) color of the corolla (yellow or pink/purple/white); (2) corolla size (large: >30 mm; or small: <30 mm); (3) capsule size (using the midpoint of the range provided by the keys, large: >10 mm; or small: <10 mm). In addition, we reconstructed the evolution of three major morphotypes as a discrete character with three states (*Antirrhinum*, *Kickxiella*, and *Streptosepalum*; see **Figure 1**) based on the subsectional classification of Rothmaler (1956). Ancestral state reconstructions were conducted using a stochastic character mapping approach (SIMMAP) implemented in the R package "PHYTOOLS" (Revell, 2012; R Core Team, 2013) and using the ultrametric tree resulting from TreePL as the input (after pruning the outgroup and the four conflicting taxa as previously done for biogeographic analyses, see above). For these analyses we tested up to three models of trait evolution differing in the transition rates among states: (i) equal rates (ER; rates are equal for all transitions among states); (ii) symmetric rates (SYM; transition rates vary, but backward and forward rates for each transition are equal); and (iii) different rates (ARD; all rates are different, including different backward and forward rates for each transition). We fit Mk models using the function "fitMk" of PHYTOOLS. The three models of trait evolution were tested for the three major morphotypes. For those traits with only two states, i.e., corolla color, and corolla and capsule size, model SYM was not applicable, and therefore only ER and ARD were tested. The best-fit model for each trait was chosen according to the Akaike Information Criterion (AIC). Two hundred stochastic reconstructions were simulated using the function "make.simmap." As a result, a summary tree of each of the 200 simulations was obtained from each character reconstruction.

Introgression Tests

We evaluated the potential introgression between early-diverging lineages in clade III and species in clades I + II (see **Figure 3** for major clades) by conducting four-taxon *D*-statistic tests (Durand et al., 2011) in PyRAD v. 3.0.6. D-statistic tests compare the

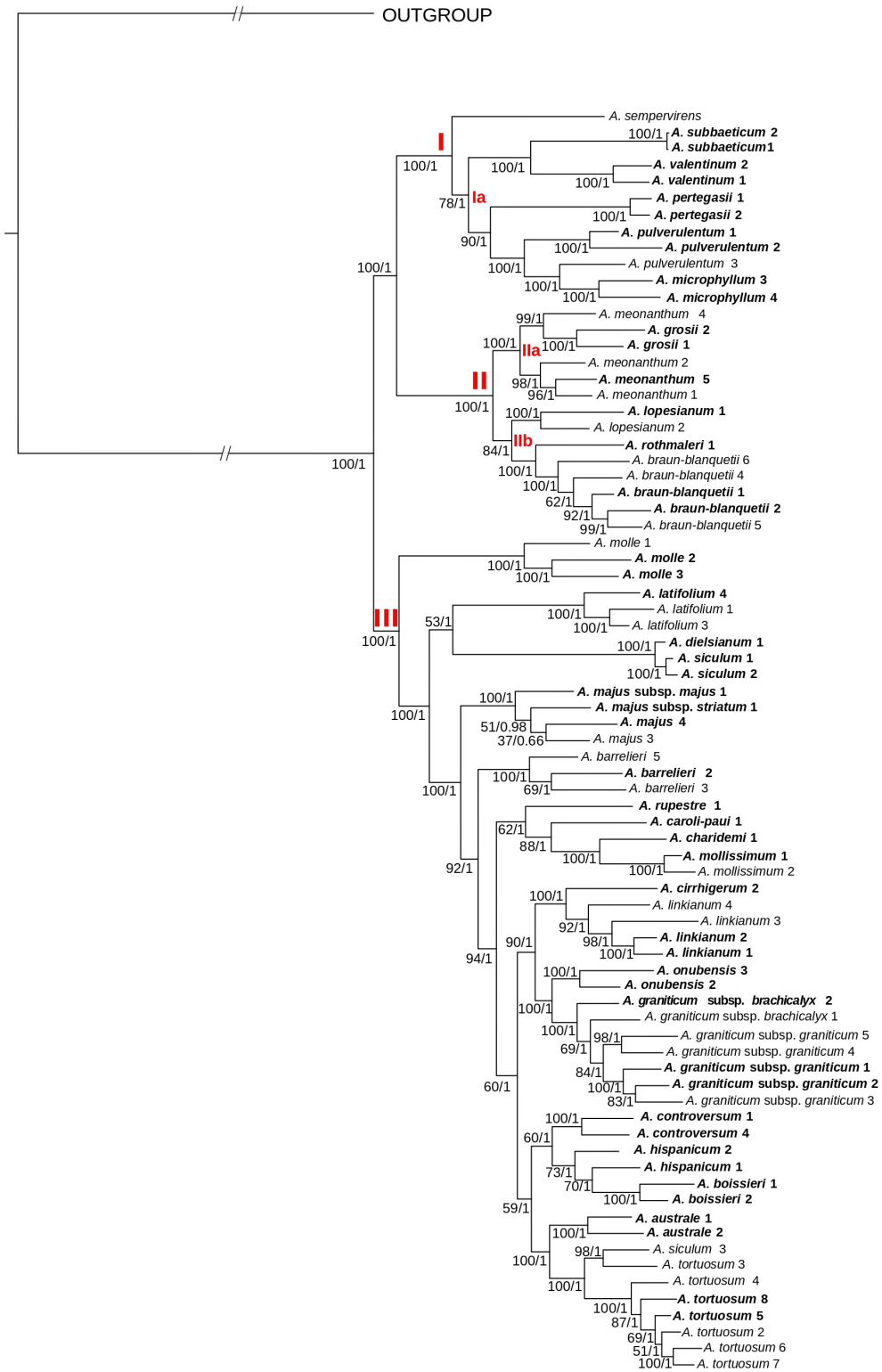


FIGURE 3 | Phylogenetic reconstruction of *Antirrhinum* based on concatenated DNA sequences of 4,663 genotyping-by-sequencing loci (c90m6 dataset) obtained from a wide sample of specimens representing the geographic distribution of species. The best scoring of the maximum likelihood tree from RAxML analysis is shown. Three major clades (I–III) and subclades (Ia, IIa, and IIb) are indicated in red letters. Values at branches are bootstrap support values from the maximum likelihood (RaxML) analysis followed by posterior probabilities from the Bayesian (ExaBayes) analysis. In bold, populations collected from topotypic localities, i.e., those where the type material was collected. Further voucher information for each specimen (indicated with name and numbers) is detailed in **Supplementary Table 1** of the Supplementary Files.

occurrence of "ABBA" versus "BABA" patterns based on a four taxon pectinate topology: {[(P1, P2), P3], O} (Martin et al., 2015). Introgression either between P1 and P3 (BABA) or between P2 and P3 (ABBA) is suggested when one of those patterns is significantly more frequent than the other (Eaton et al., 2015). In particular, we tested introgression for *A. molle, A. latifolium* and *A. siculum* as suggested by previous results (Wilson and Hudson, 2011). This three species are involved in the early divergence in clade III (see **Figures 3**, **4**). Therefore, we focused on three specific introgression hypotheses: (i) introgression between *A. molle* and clade I + II; (ii) introgression between *A. latifolium* and clade I + II; and (iii) introgression between *A. siculum* and clade I + II. To test these hypotheses, our pectinated four-taxon tree comprised: (1) the outgroup (O), which included the nine outgroup taxa of phylogenomic analyses; (2) P3, corresponding to a selection of two individuals per species from clade I + II, including at least one topotypic individual (see **Supplementary Table 2**); (3) P2, including all individuals from the three early-diverging lineages of clade III (*A. molle, A. latifolium*, and *A. siculum*); and (4) P1, comprising a selection of individuals from the remaining taxa of clade III following the same criteria as for P3 (see **Supplementary Table 2**). Combination of different individuals yielded a total of 46,980 ABBA-BABA tests. As input, we used the dataset that had produced the tree with the highest bootstrap average. Two hundred bootstrap replicates per test were run. We calculated percentages of tests with significant results (based on adjusted *p*-values calculated with the Holm–Bonferroni method; $\alpha = 0.05$) for the three set of tests corresponding to the three introgression hypotheses.

RESULTS

Data Assembly

Illumina sequencing provided c. 700 millions of reads for each of the two libraries, with a GC content between 42–44% and around 96% of bases with quality >Q20. FastQC analysis showed high quality of reads with low signal of adapters or other contaminants. Reference mapping assembly yielded between 237 and 6,524 loci, concatenated lengths between 78,284 bp and 1.69 Mbp, and percentages of missing data varying from 32.6 to 78.8% depending on the set of localities included (all localities or topotypic localities only) and minimum taxon coverage (see **Table 2**).

Phylogenetic Analysis and Divergence Time Estimates

Topologies resulting from ML analyses were mostly congruent among the eight matrices, with differences in bootstrap support (BS) values at some nodes (**Supplementary Figures 1**, **2**). The highest BS values were obtained for the m6 matrices (minimum coverage of six samples per locus). Consequently, these matrices were used for further phylogenetic analyses (**Figures 3**, **4**). In general, bootstrap support values were high (the majority at 100) except for some medium-terminal nodes (BS < 70). Monophyly of multiple populations in species recognized by Sutton (1988) was reconstructed in many cases. In addition, phylogenetic distinctiveness and morphological characters led us to recognize one more species (*A. rupestre*). For the sake of brevity, we show results of 26 species herein recognized for the genus *Antirrhinum*, which includes all the species of Sutton (1988) and those recognized after this publication (see Güemes, 2009), plus *A. rupestre* (**Figure 3**). Additional taxonomic reassessment includes the transference of yellow-flowered plants from Morocco previously identified as *A. siculum* (Fiz et al., 2000) to the polymorphic *A. tortuosum* based on morphological characters other than flower color.

Regarding the topology of the tree including all samples, three well-supported main clades were obtained: (1) clade I, including *A. sempervirens* as the earliest-diverging lineage, sister to a subclade (subclade Ia) formed by *A. subbaeticum*, *A. valentinum*, *A. pertegasii*, *A. pulverulentum,* and *A. microphyllum* (the latter nested within *A. pulverulentum*); (2) clade II, sister to clade I, and in turn formed by the two sister subclades IIa (*A. meonanthum–A. grosii*) and IIb (*A. lopesianum–A. rothmalerii*, *A. braun-blanquetii*); and (3) clade III, containing all remaining species. The latter clade showed a repeated pattern of nested lineage differentiation in which *A. molle* was inferred to be the earliest-diverging lineage (see **Figure 3**).

The same three main clades were reconstructed in the tree of topotypic specimens (**Figure 4**). Topology was congruent except for the position of *A. barrelieri* and phylogenetic relationships among some subclades, although with moderate bootstrap support (*A. meonanthum–A. grosii; A. linkianum-cirrhigerum, A. graniticum-onubensis*, and *A. hispanicum–A. controversum–A. australe–A. tortuosum*) (**Figures 3**, **4**). For the tree including all localities, *A. barrelieri* was inferred as an isolated early-diverging lineage, sister to *Antirrhinum* species from S and SW Iberian Peninsula plus *A. tortuosum*, whereas for tree of topotypic specimens, *A. barrelieri* was inferred as nested within the S-SW clade, although with marginal bootstrap support value (BS = 77) (**Figure 4**). Bayesian inference in ExaBayes yielded the same topologies as ML, and most nodes showed maximum Bayesian posterior probabilities (BPP = 1) (**Figure 4**). Topologies from coalescent-based analyses using the SVDquartets method were congruent with concatenation-based phylogenies (**Figure 4**). The three major clades were also inferred with high BS, although lower BS values were obtained for internal subclades within clade III (**Supplementary Figure 3**).

Estimates of divergence times obtained from TreePL are shown in **Supplementary Figure 4**. A stem age of 16.85 myr and a crown age of 4.77 myr were inferred for *Antirrhinum*, leading to a diversification rate estimate of 0.54 spp. Myr^{-1} assuming an extant diversity of 26 species. A stem age of 4.32 myr was inferred for diversification of clades I and II, while estimated crown ages were 2.43 and 3.49, respectively. The estimated crown age for clade III was 4.36 myr. Divergence times from the Pleistocene onward (<2 myr) were inferred for most *Antirrhinum* species, with some exceptions in the Pliocene such as *A. sempervirens*, *A. molle*, *A. siculum*, *A. latifolium*, and *A. majus* (**Supplementary Figure 4**).

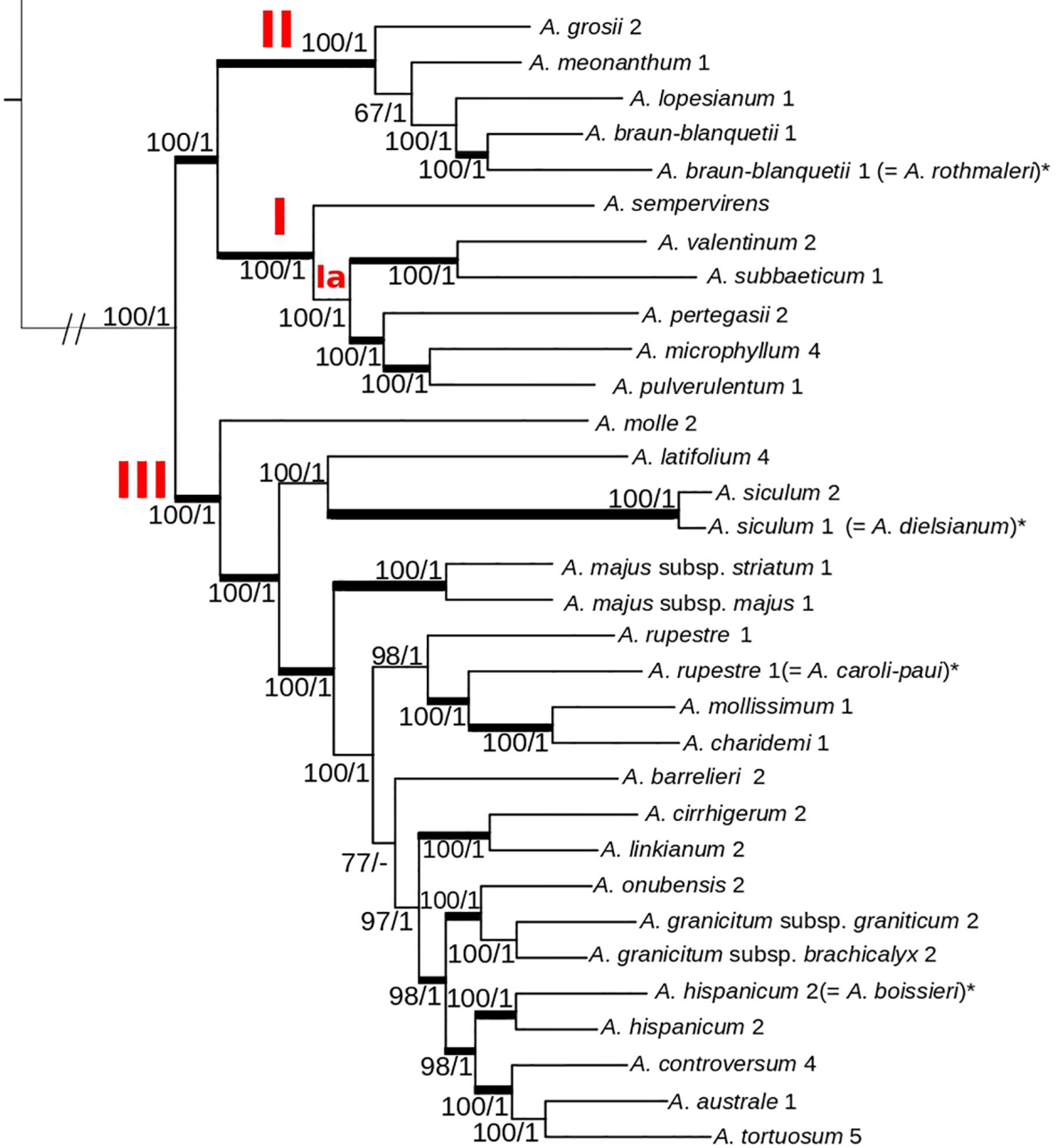


FIGURE 4 | Phylogenetic relationships of the 26 *Antirrhinum* species herein recognized based on concatenated DNA sequences of 4,016 genotyping-by-sequencing loci (c90m6 dataset) obtained from topotypic specimens. The best scoring maximum likelihood tree from RAxML analysis is shown. The three major clades (I–III) and a subclade (Ia) are indicated in red letters. Values at branches are bootstrap support values from the maximum likelihood (RAxML) analysis followed by posterior probabilities from the Bayesian (ExaBayes) analysis. Thicker branches are also supported by coalescent-based analyses in SVDquartets. An asterisk (*) at some taxon names (in brackets) indicates current synonyms for particular species. Further voucher information for each specimen is detailed in **Supplementary Table 1** of the Supplementary Files.

Biogeographic Analyses

The DEC model showed lower AIC values than the DIVA-like model (132.6 and 138.6, respectively), with $\Delta AIC > 2$ (Anderson and Burnham, 2002). Overall, DEC and DIVA models resulted in biologically congruent reconstructions, although DEC resulted in lower uncertainty at some deep nodes (**Figure 5**). The DEC model inferred a widespread ancestral range for the genus *Antirrhinum*, most likely including the NW, NE, and SE of the Iberian Peninsula (pACD = 0.41; **Figure 5A**), whereas the DIVA model showed a widespread ancestral range including northern Iberia and non-Iberian Europe but with higher ambiguity (pADF = 0.33; **Figure 5B**). Ancestral range for the common ancestor of clades I and II was inferred to be most likely in northern and SE Iberia (pACD = 0.52) for the DEC model, and only in northern Iberia for the DIVA model (pAD = 0.84; **Figure 5**). Meanwhile, a NW Iberian ancestral range was congruently inferred for clade II by DEC and DIVA models (DEC: $p = 0.99$; DIVA: $p = 0.99$), while an eastern (DEC: $p = 0.69$) or NE (DIVA: pD = 0.82) Iberian ancestral range was obtained for clade I. A geographic split was inferred for the eastern widespread ancestor of the subclade Ia, which led to lineage differentiation in NE (*A. pertegasii–A. pulverulentum–A. microphyllum*) and SE (*A. subbaeticum–A. valentinum*) Iberia, with maximum probabilities for both DEC and DIVA models. Ancestral range for clade III was inferred as most likely in NE Iberia with ($p = 0.44$ in DIVA) or without ($p = 0.50$ in DEC) other areas. For both DEC and DIVA, lineage colonization within clade III was inferred from NE Iberia to SE Iberia, accompanied by speciation in the latter area (*A. rupestre*, *A. charidemi*, and *A. mollissimum*). Later on, the subclade formed by *A. barrelieri* and the remaining species had a widespread eastern ancestor that also colonized SW Iberia. Within this subclade, lineage colonization led to differentiation of SW lineages on one side (*A. linkianum*, *A. cirrhigerum,* and *A. onubensis*) and SE lineages on the other side (*A. hispanicum*, *A. boissieri*, *A. controversum*, and *A. australe*), while *A. graniticum* expanded to a larger area. Three independent events of colonization of northern Africa were inferred in clade III: (1) from non-Iberian Europe by *A. siculum*; (2) from SW Iberia by *A. cirrhigerum*; and (3) from SE Iberia or non-Iberian Europe leading to differentiation of the widely distributed *A. tortuosum* (**Figure 5**). Lineage expansion was inferred for *A. linkianum* because it has an ancestral origin in SW Iberia, with subsequent colonization of NW Iberia.

Ancestral State Reconstruction

Ancestral state reconstruction estimated pink/purple/white corollas as the most likely state for all ancestors, with yellow corollas evolving six times independently (in *A. meonanthum*, *A. braun-blanquetii*, *A. siculum–A. latifolium,* and yellow-flowered populations of *A. tortuosum* and *A. majus*; **Figure 6A**). Large corollas were inferred as the most likely ancestral state for *Antirrhinum*, with multiple changes to smaller ones during the evolutionary history of the genus (**Figure 6B**). Acquisition of small corollas took place in the ancestors of clades I and II, but not clade III. This state was maintained to the present in all extant species of clades I and II except for *A. braun-blanquetii*, which showed a recent reversion to large corollas (**Figure 6B**). Within clade III at least six changes to small corollas were inferred (**Figure 6B**). Shifts in capsule size were inferred from the ancestral state of large capsules to small sizes for Clades I and II including a reversion to large size for *A. braun-blanquetii* (**Figure 6C**). Clade III maintained the ancestral state of large capsule with at least six shifts to small capsules toward the tips (*A. molle*, *A. rupestre–A. charidemi–A. mollissimum*; *A. barrelieri*; *A. onubense*; *A. hispanicum*; and *A. controversum*; **Figure 6C**). The *Kickxiella* morphotype (small prostate and xerophytic woody plants) was inferred as the most likely ancestral habit not only for the genus but also for the three main clades. The *Antirrhinum* morphotype (upright tall habit, long thin leaves with no hairs and magenta or yellow flowers growing in sandy soils) was acquired soon after divergence of clade III, and reverted twice to the *Kickxiella* morphotype in the evolution of four species. The *Streptosepalum* morphotype appeared to have evolved twice, giving rise to two species of clade II (**Figure 6D**).

Introgression Tests

D-statistic analyses showed no significant ancestral introgression between any of the three early-diverging lineages of clade III (*A. molle, A. latifolium,* and *A. siculum*) and the clades I + II (see **Table 3**). However, ABBA-BABA tests suggested a certain level of introgression between remaining species of clade III and certain lineages of clade II. In particular, *A. lopesianum* was involved in 276 of 406 tests that yielded significant results for introgression of clade II and lineages from clade III (see **Supplementary Table 2**). None of the individuals of clade I was suggested to have experienced significant introgression with clade III.

DISCUSSION

A phylogenetic structure of *Antirrhinum* has been elusive for a long time. Possible causes for lack of phylogenetic resolution included rapid speciation and hybridization (Vargas et al., 2004, 2009). In particular, *Antirrhinum* phenotypes seemed to be constrained by selection, which might reflect contrasting adaptations to life on bare rock faces or sandy soils (Wilson and Hudson, 2011). Indeed, recent speciation and hybridization appear to be the main causes of plant diversity in Mediterranean-type ecosystems (Rundel et al., 2016), and they may also be responsible for complex taxonomy of angiosperms (Vargas et al., 2018). Nevertheless, none of the previous phylogenetic studies of *Antirrhinum* included a considerable number of type or topotypic materials to contrast hypotheses (but see Vargas et al., 2009). This has also prevented the accurate naming of some *Antirrhinum* populations. In contrast, our study offers a new phylogenetic hypothesis that helps to elucidate systematics and evolution thanks to a high number of phylogenetically informative nucleotide characters (>50,000 bp) taken from genome-wide (GBS) sequences, together with DNA sequences that are anchored to original *Antirrhinum* names (Bell et al., 2020).

Our approach leads to strongly encourage researchers in plant systematics to not only use a high number of loci (as provided

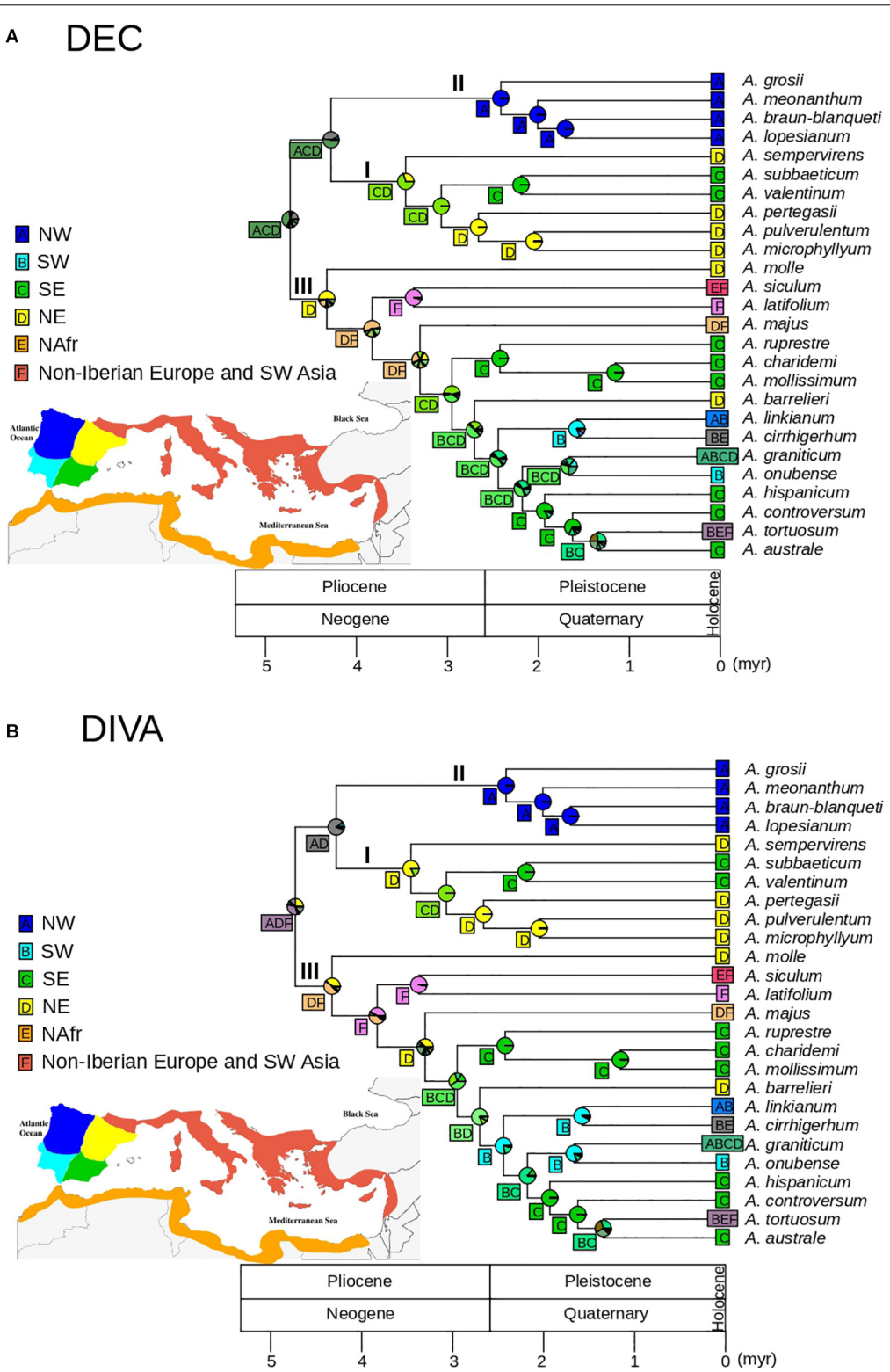


FIGURE 5 | Biogeographic patterns for *Antirrhinum* estimated from the GBS phylogeny of topotypic specimens (see **Figure 4**) and inferred using a set of six areas previously defined: NW Iberia (A), SW Iberia (B), SE Iberia (C), NE Iberia (D), N Africa (E), and non-Iberian Europe and SW Asia (F) (see Vargas et al., 2009). Pie charts represent probabilities of alternative ancestral ranges at nodes based on the DEC **(A)** and DIVA **(B)** models. The most likely range areas at each node are indicated in a square.

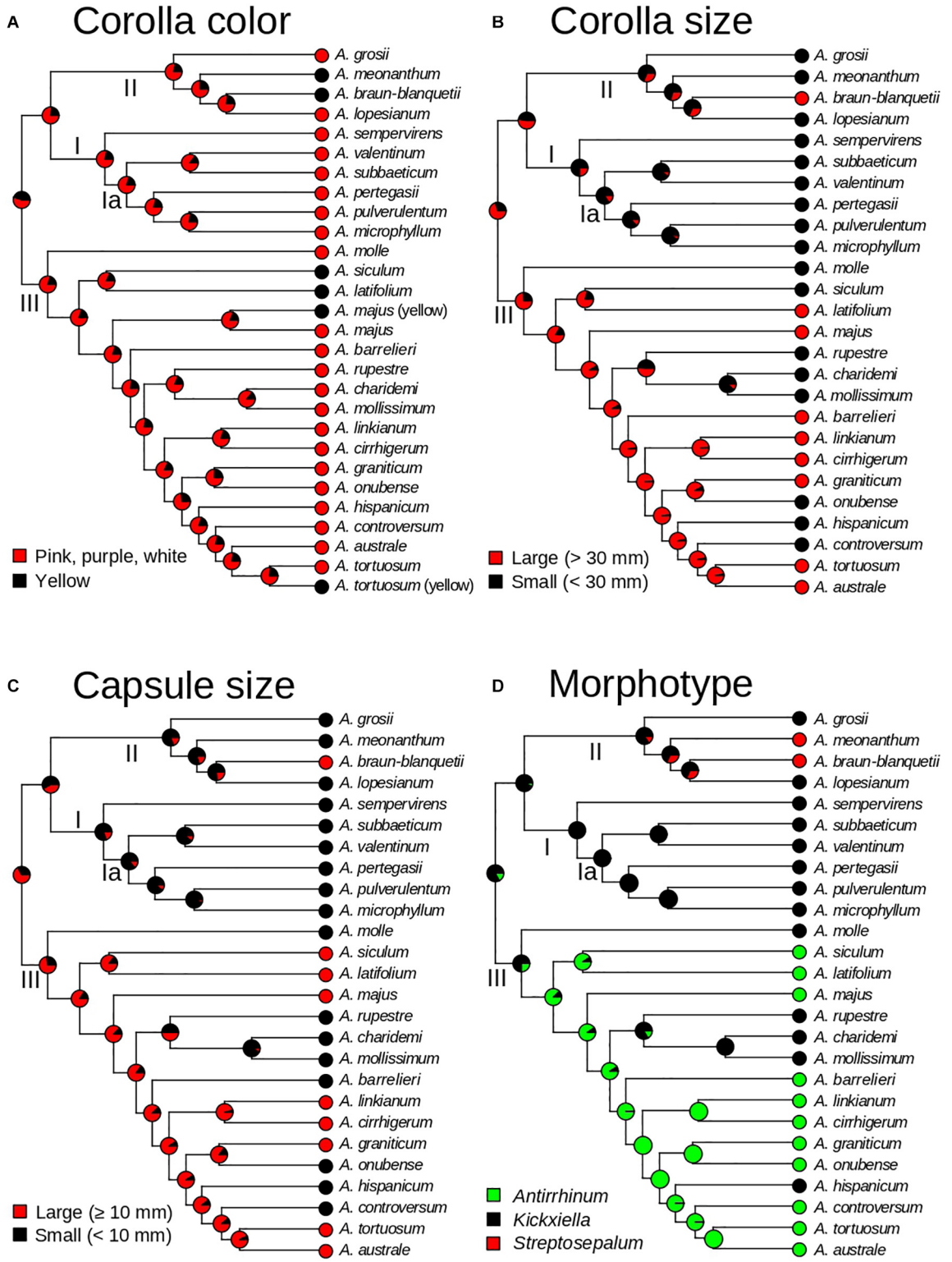


FIGURE 6 | Evolution of morphological traits of *Antirrhinum* based on a stochastic character mapping approach (SIMMAP) and the GBS phylogeny of topotypic specimens (see **Figure 4**). Pie charts represent probabilities of alternative ancestral states at nodes after analyses of four traits: corolla color **(A)**, corolla size **(B)**, capsule size **(C)**, and morphotype **(D)**. Morphotypes are defined according to the three taxonomic subsections of Rothmaler (1956).

TABLE 3 | *D*-statistic summary for the introgression test given a four-taxon tree {[(P1,P2),P3],O}.

		Proportion of significative tests for introgression with clade I + II (P3)
Early-diverging lineages of clade III (P2): BABA pattern	Hypothesis I: *A. molle*	0
	Hypothesis II: *A. latifolium*	0
	Hypothesis III: *A. siculum*	0
Remaining lineages of clade III (P1): ABBA pattern	Hypothesis I: *A. molle*	105/ 15660
	Hypothesis II: *A. latifolium*	227/ 15660
	Hypothesis III: *A. siculum*	74/ 15660

Proportion of significant tests between individuals of clade I and II (P3) and either early-diverging individuals from clade III (P2; BABA pattern) or remaining individuals from clade III (P1; ABBA pattern).

by the GBS technique) but also include topotypic material in phylogenetic analyses to ensure a correct naming of populations and stable evolutionary inferences.

Contribution to the Systematics of *Antirrhinum*

The family of restriction digest-based techniques (including GBS, RAD-Seq and similar techniques) is highly recommended for studies in systematics of diverse plant groups inasmuch as the combination of cost-effectiveness, high resolution and strong statistical support enables a robust overview of phylogenetic relationships among lineages. However, the nature of reduced representation sequencing based on restriction sites can result in high rates of missing data in concatenated matrices and lower repeatability of targeted genomic regions (Harvey et al., 2016). In our case, the percentages of missing data are similar to those analyzed in other studies that also used GBS data (e.g., Fernández-Mazuecos et al., 2018; Martín-Hernanz et al., 2019) and found robust phylogenetic support after discarding some low quality samples (see section "Results").

The two worldwide taxonomic treatments of *Antirrhinum* (Rothmaler, 1956; Sutton, 1988) included species circumscriptions that mostly agree with our phylogenetic results. However, our results suggest a number of taxonomic re-arrangements, particularly at the species level. On the one hand, some species recognized by Rothmaler (1956) were not considered by Webb (1971) and Sutton (1988) or further authors, but should be re-visited (**Table 1**). On the other hand, new narrow endemic species found in the field were described after Sutton (1988), recognized in more recent floras (e.g., Güemes, 2009) and supported by our phylogenetic analyses. For the sake of brevity, we next discuss major phylogenetic results that disagree with those four treatments.

Monophyly and Species Recognition

Numerous (17) monophyletic groups of populations were congruently associated with current species (Vargas et al., 2004; Carrió et al., 2010; Wilson and Hudson, 2011). Some other species (4) fall in a pattern of paraphyly (*A. microphyllum* embedded in *A. pulverulentum* lineages; *A. grosii* embedded in *A. meonanthum* lineages) following Sutton's (1988) circumscription of species (**Figure 3**). This suggests a pattern of budding speciation typically found in Iberia (Otero et al., 2019). The hypothesis of a general pattern of budding speciation in *Antirrhinum* needs to be tested with a higher number of populations, particularly from widespread species.

Infrageneric Taxa

The hypothesis of three main groups (subsections) of *Antirrhinum* proposed by Rothmaler (1956), which was not adopted by either Webb (1971) or Sutton (1988), is not supported by our phylogenetic analyses. The phylogenetic structure of *Antirrhinum* based on GBS data revealed two main clades (I + II and III, see **Figures 3**, **4**), both of them with species of subsect. *Kickxiella* (**Figure 3**). In particular, the two species of subsect. *Streptosepalum* (*A. braun-blanquetii* and *A. meonanthum*) grouped together with most species of subsect. *Kickxiella*, whereas all the species of subsect. *Antirrhinum* grouped together with a few species of subsect. *Kickxiella*. This is partially congruent with Wilson and Hudson (2011) and prevents us from accepting any subsections based solely on classical morphological characters.

Antirrhinum majus Group

The type species of the genus is *A. majus* (Sutton, 1988). This species is an important model for plant genetics, development and evolution (see Schwarz-Sommer et al., 2003; Li et al., 2019). Despite the importance of this model plant, taxonomy of the "*A. majus* group" has neither met consensus yet nor been revisited using molecular phylogenetics. During most of the 20th century, researchers primarily considered six subspecies (*barrelieri*, *cirrhigerum, linkianum*, *majus*, *striatum*, and *tortuosum*) circumscribed in the *A. majus* group (Rothmaler, 1956; Webb, 1971; Sutton, 1988). The use of topotypic samples reveals that all these taxa do not form a natural group because they are placed in different subclades of clade III (**Figure 4**). This result agrees with a taxonomic treatment at the species level following the most recent treatment of *Antirrhinum* for the Iberian Peninsula (Güemes, 2009). The question remains as to whether laboratories using *Antirrhinum* as a model plant are employing plants correctly assigned to the true *A. majus* or any other species of the former *A. majus* group.

Further Taxonomic Research

Our phylogenetic results help taxonomic decision making at the supraspecific level, but also suggest further investigation in certain subclades that contain poorly studied species: (i) *A. rupestre* and *A. caroli-paui* are basal-most lineages of the subclade of *A. mollissimum–A. charidemi*; (ii) *A. striatum* is certainly unrelated to *A. latifolium* (see Liberal et al., 2014) and

now considered within *A. majus* at the subspecies level (Khimoun et al., 2013); and (iii) four taxa described as independent species (*A. bolosii*, *A. fernandezcasasii*, *A. saccharatum*, and *A. ternatum*) need to be phylogenetically analyzed for systematic purposes given that there is no consensus among authors. High uncertainty about these species is illustrated by the fact that the same author who first described *A. bolosii* (Fernández-Casas, 1972) did not consider it shortly after (Fernández-Casas, 1974). In addition, as these species have never been included in a key to species of *Antirrhinum* or in a comparative table, a taxonomic study using their type specimens and topotypic populations is needed.

Radiation and Geographical Speciation of *Antirrhinum*

The hypothesis of an evolutionary radiation of *Antirrhinum* based on ITS and plastid sequences (Vargas et al., 2009) is supported by our phylogenetic analysis based on GBS data, which found a high rate of diversification (0.54 spp. Myr^{-1}) into at least 26 species (Valente et al., 2010; Bell et al., 2012) since the Pliocene (Vargas et al., 2009). The combination of terrain complexity and eco-climatic novelty seems to explain why the Mediterranean basin contains numerous examples of explosive radiations in the angiosperms (Valente et al., 2010).

The hypothesis of a primarily geographic pattern of snapdragon speciation in Iberia has been put forward based on: (i) endemicity for the majority of *Antirrhinum* species (Vargas et al., 2009); (ii) numerous narrow, endangered endemics in small mountain ranges (Carrió et al., 2010); (iii) a high number of unique haplotypes and haplotype clades restricted to small geographic areas (Vargas et al., 2009). In particular, previous phylogeographic results are supported by our biogeographic analysis (**Figure 5**), in which a likely primary center of diversification in northern Iberia (or even out of the Iberian Peninsula) was followed by a secondary center of diversification in SE Iberia (Vargas et al., 2009). Six speciation events occurred unequivocally in SE Iberia during the Quaternary based on our biogeographic reconstruction and time-calibrated phylogeny (**Figure 5**). The mountains of eastern Andalusia (SE Iberia) contain one of the richest areas of Europe in terms of number of species and endemics (Cueto et al., 2018) and form indeed a main center of recent angiosperm diversification (Buira et al., 2020). One more argument for the strong pattern of recent geographic speciation in SE Iberia is shown by the considerable number of sister species pairs (*A. subbaeticum–A. valentinum*; *A. mollissimum–A. charidemi*; *A. australe–A. tortuosum*) in nearby areas (**Figure 3**). In addition, two more clades of NW and NE Iberian species support a strong pattern of geographic speciation for *Antirrhinum* (Vargas et al., 2009), which fits into a general pattern for Mediterranean plants. A pattern of geographic differentiation in the Mediterranean appears to be the rule rather than the exception because of unique opportunities for spatial isolation given the numerous islands, peninsulas, and high mountains of southern Europe. In particular, the complex geography and orography of Iberia provides a suitable spatial framework for speciation (Rundel et al., 2016; Vargas et al., 2018). Spatial differentiation has also been documented for early stages of speciation between *A. majus* subsp. *majus* and subsp. *striatum* separated by mountains (Pujol et al., 2017), albeit ecological conditions may have secondarily contributed (Khimoun et al., 2013). One more line of evidence that supports predominant speciation by geographic isolation reinforced by ecological factors is given by a strong geographic pattern of pollinator types associated with the bee-specialized flowers of *Antirrhinum* (Vargas et al., 2010). In particular, two pollinator systems (Mediterranean vs. temperate areas of Europe) were proposed based on two consistent pollinator niches that include 11 bee species and most *Antirrhinum* species (Vargas et al., 2017).

Hybridization vs. Convergent Evolution of Key Morphological Characters

None of the states of the three morphological characters (corolla color, corolla size, and fruit size) used in all taxonomic keys to *Antirrhinum* species turned out to be synapomorphic (**Figure 6**). This result was already observed in previous analyses, but poor resolution prevented from a detailed description of character homoplasy (Vargas et al., 2004; Wilson and Hudson, 2011). Our ancestral state reconstruction analyses revealed a strong pattern of convergent evolution for corolla color, corolla size, and capsule size (**Figure 6**). One of the fastest and most frequent mechanisms of phenotypic convergence is hybridization (Stern, 2013). Hybridization in *Antirrhinum* has been extensively documented based on the observation of plants with intermediate morphological characters between co-occurring species (see Güemes, 2009), production of viable offspring from crosses between multiple species (Günther and Rudolph, 1970) and molecular results (Vargas et al., 2009). In particular, plastid haplotype sharing among species (Vargas et al., 2009; Wilson and Hudson, 2011) and the occurrence of numerous nucleotide additivities in nuclear ribosomal ITS sequences (Vargas et al., 2009; Carrió et al., 2010) have been interpreted as evidence of recent hybridization in *Antirrhinum*. Indeed, both recent and ancient hybridization events have been proposed based on shallow and profound incongruence between morphology, plastid, and nuclear markers, which produce large phylogenetic polytomies (Vargas et al., 2004, 2009; Carrió et al., 2010; Wilson and Hudson, 2011). Our analyses to test a hybridization signal between main clades potentially resulting in the early-diverging position of three species of clade III (*A. molle, A. latifolium*, and *A. siculum*) failed to find a significant result. More recent hybridization may be easier to reconstruct based on the genetic makeup of current populations. At this level, no unequivocal support for hybridization was, however, found for some species of *Antirrhinum* based on nuclear fingerprints (RAPD) and morphology (Jiménez et al., 2005). Along the same lines, unique allozyme profiles were interpreted as long-term isolation between three species with similar morphotypes (Mateu-Andrés, 1999), while *A. charidemi* also shows unique plastid sequences and SSR profiles (Forrest et al., 2017). Sporadic hybrids in the field, together with rare cases of hybrid swarms indicate that hybridization between

main clades is not currently playing an important role in *Antirrhinum* (Sutton, 1988; Mateu-Andrés, 1999). Indeed, reproductive experiments revealed that post-zygotic barriers are at play at least for two co-occurring species of *Antirrhinum* (*A. controversum*, *A. valentinum*) that are distant in our phylogenetic reconstruction (Carrió and Güemes, 2014).

In sum, our study provides a clear cladogenetic pattern for *Antirrhinum* thanks to a high number of loci obtained with the GBS technique. Besides, stability of the phylogeny is guaranteed by the use of topotypic material anchored to original *Antirrhinum* names. Lack of resolution using less variable DNA sequences appear to have primarily been the result of evolutionary radiation in the Pleistocene rather than hybridization between plants of different major lineages. A strong signal of monophyly and phylogenetic distinctiveness for currently recognized species was obtained, but the analysis of a higher number of populations is still needed for a more solid proposal of the systematics of snapdragons. Although some plants may have a hybrid origin, intermediate key morphological characters (corolla color and size, fruit size) in numerous species are better explained by parallel evolution due to gene reutilization, a hypothesis that has been put forward for some model genera including *Antirrhinum* (Preston et al., 2011). All sources of data strongly suggest a general pattern of divergence promoted by isolation in mountains, which accounts for a high number of endemics to the Iberian Peninsula. Geographical isolation is, therefore, the most plausible driver of speciation in *Antirrhinum*, followed by ecological factors such as pollination systems, soil preference and climate conditions.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Short Read Archive (SRA), under the reference number PRJNA690200. SRA accessions for each sample are indicated in **Supplementary Material** at **Supplementary Table 1**.

AUTHOR CONTRIBUTIONS

PV conceived the idea and drafted the manuscript. AO led lab work. AO conducted the data analysis and summarized results with contributions of MF-M. PV, AO, and MF-M interpreted and discussed the results, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

We gratefully acknowledge financial support by two Spanish projects: *En busca de áreas de biodiversidad en Sierra Nevada: las técnicas de barcoding aplicadas a angiospermas (tribu Antirrhineae), polinizadores (abejas), y sus interacciones* (Organismo Autónomo de Parques Nacionales 005/2008); and *Evolución de la flor personada* (Ministerio de Ciencia e Innovación CGL2009-10031). AO and MF-M were supported by two Special Intramural Projects of the Spanish National Research Council (CSIC, references 201730E029 and 201930E078 respectively).

ACKNOWLEDGMENTS

The authors thank Isabel Liberal, Jose Luis Blanco-Pastor, Emilio Cano, Yolanda Ruiz, Mónica García-Gallo, Concha Baranda, Leopoldo Medina, Alberto Coello, Pedro Jiménez-Mejías, Alberto Herrero, Jaime Güemes, Marcos Egea, Ferderico Selvi, Lorenzo Peruzzi, Chiara Nepi, Daniel Gómez, Luis Calderón, Francisco Javier Salgueiro, Víctor Manuel Pizarro, Manuel Joao Pinto, Ana Isabel de Vasconcelos, Ana Isabel Correia, Jose García, Cristiana Costa, Paulo Ventura, Miguel Porto, Jesus Riera, Francisco Javier Hernández, Estrella Alfaro, Ester Manjavacas, and all the authors who collected *Antirrhinum* material deposited in public herbaria (FI, JACA, LEB, MA, MGC, PI, PO, SALA, and VAL).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.631178/full#supplementary-material

Supplementary Figure 1 | The best scoring maximum likelihood tree from RAxML analysis of all *Antirrhinum* taxa set under three different numbers of minimum taxa for a locus: m4, m18, and m36. Bootstrap support values are indicated at the nodes.

Supplementary Figure 2 | The best scoring maximum likelihood tree from RAxML analysis of *Antirrhinum* topotypic specimens under three different numbers of minimum taxa for a locus: m4, m18, and m36. Bootstrap support values are indicated at the nodes.

Supplementary Figure 3 | Consensus coalescent-based tree obtained from SVDquartets. Individuals were grouped according to current species circumscriptions. Bootstrap values are indicated at the nodes.

Supplementary Figure 4 | Time-calibrated tree of topotypic specimens of *Antirrhinum* obtained from TreePL analysis. Inferred ages are indicated at nodes.

Supplementary Table 1 | Data information and NCBI SRA accessions of all individuals sampled for GBS.

Supplementary Table 2 | Significant *D*-statistic tests given a four-taxon tree {[(P1,P2),P3],O}. Abbreviation for each individual is described at **Supplementary Table 1**. Tests are differentiated according to the three introgression hypotheses tested (MOLLE: hybrid origin for *A. molle*; LAT: hybrid origin for *A. latifolium*; and *SIC*: hybrid origin for *A. siculum*). *D*-statistic value (D), standard deviation [(std. (D)], *z*-score (Z), proportion of BABA and ABBA, number of loci involved for each test (nloci), number of bootstrap replicates (nboot), significant pattern (pattern), *p*-value and adjusted *p*-value through Bonferroni–Holm method are shown.

Supplementary Data Sheet 1 | Locations of *Antirrhinum* photographs included in **Figure 2** of the main text.

REFERENCES

Aberer, A. J., Kobert, K., and Stamatakis, A. (2014). ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Mol. Biol. Evol.* 31, 2553–2556. doi: 10.1093/molbev/msu236

Anderson, D. R., and Burnham, K. P. (2002). Avoiding pitfalls when using information-theoretic methods. *J. Wildlife Manag.* 66, 912–918. doi: 10.2307/3803155

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data". Babraham Bioinformatics*. Cambridge: Babraham Institute.

Bell, C. D., Mavrodiev, E. V., Soltis, P. S., Calaminus, A. K., Albach, D. C., Cellinese, N., et al. (2012). Rapid diversification of *Tragopogon* and ecological associates in Eurasia. *J. Evol. Biol.* 25, 2470–2480. doi: 10.1111/j.1420-9101.2012.02616.x

Bell, R. C., Mulcahy, D. G., Gotte, S. W., Maley, A. J., Mendoza, C., Steffensen, G., et al. (2020). The type locality project: collecting genomic-quality, topotypic vouchers and training the next generation of specimen-based researchers. *Syst. Biodivers.* 18, 1–16.

Buira, A., Fernández-Mazuecos, M., Aedo, C., and Molina-Venegas, R. (2020). The contribution of the edaphic factor as a driver of recent plant diversification in a Mediterranean biodiversity hotspot. *J. Ecol.* 1–13. doi: 10.1111/1365-2745.13527

Carrió, E., Forrest, A. D., Güemes, J., and Vargas, P. (2010). Evaluating species nonmonophyly as a trait affecting genetic diversity: a case study of three endangered species of *Antirrhinum* L.(Scrophulariaceae). *Plant Syst. Evol.* 288, 43–58. doi: 10.1007/s00606-010-0311-4

Carrió, E., and Güemes, J. (2014). The effectiveness of pre-and post-zygotic barriers in avoiding hybridization between two snapdragons (*Antirrhinum* L.: plantaginaceae). *Bot. J. Linn. Soc.* 176, 159–172.

Chifman, J., and Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30, 3317–3324. doi: 10.1093/bioinformatics/btu530

Cueto, M., Melendo, M., Gimenez, E., Fuentes, J., Carrique, E. L., and Blanca, G. (2018). First updated checklist of the vascular flora of Andalusia (S of Spain), one of the main biodiversity centres in the Mediterranean Basin. *Phytotaxa* 339, 1–95. doi: 10.11646/phytotaxa.339.1.1

Cullings, K. (1992). Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. *Mol. Ecol.* 1, 233–240. doi: 10.1111/j.1365-294x.1992.tb00182.x

Doyle, J., and Doyle, J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28, 2239–2252.

Eaton, D. A., Hipp, A. L., González-Rodríguez, A., and Cavender-Bares, J. (2015). Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution* 69, 2587–2601. doi: 10.1111/evo.12758

Eaton, D. A., and Overcast, I. (2020). ipyrad: interactive assembly and analysis of RADseq datasets. *Bioinformatics* 36, 2592–2594. doi: 10.1093/bioinformatics/btz966

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

Escudero, M., Eaton, D. A. R., Hahn, M., and Hipp, A. L. (2014). Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: a case study in *Carex* (Cyperaceae). *Mol. Phylogenet. Evol.* 79, 359–367. doi: 10.1016/j.ympev.2014.06.026

Fernández-Casas, J. (1972). Dos especies nuevas del género *Antirrhinum* L. *Cuad. Cienc. Biol.* 2, 43–45.

Fernández-Casas, J. (1974). De flora hispanica II. *Candollea* 29, 327–335.

Fernández-Casas, J. (1987). Asientos para una flora occidental, 7. *Fontqueria* 15, 39.

Fernández-Mazuecos, M., and Glover, B. J. (2017). The evo-devo of plant speciation. *Nat. Ecol. Evol.* 1:0110.

Fernández-Mazuecos, M., Mellers, G., Vigalondo, B., Sáez, L., Vargas, P., and Glover, B. J. (2018). Resolving recent plant radiations: power and robustness of genotyping-by-sequencing. *Syst. Biol.* 67, 250–268. doi: 10.1093/sysbio/syx062

Fernández-Mazuecos, M., Blanco-Pastor, J. L., Juan, A., Carnicero, P., Forrest, A., Alarcón, M., et al. (2019). Macroevolutionary dynamics of nectar spurs, a key evolutionary innovation. *New Phytol.* 222, 1123–1138. doi: 10.1111/nph.15654

Fiz, O., Valcarcel, V., Martinez, J., Vargas, P., and Güemes, J. (2000). *Antirrhinum siculum* Mill.(Scrophulariaceae) in Morocco: a new record for Africa. *An. Jard. Bot. Madr.* 58, 362–363.

Forrest, A., Escudero, M., Heuertz, M., Wilson, Y., Cano, E., and Vargas, P. (2017). Testing the hypothesis of low genetic diversity and population structure in narrow endemic species: the endangered *Antirrhinum charidemi* (Plantaginaceae). *Bot. J. Linn. Soc.* 183, 260–270. doi: 10.1093/botlinnean/bow002

García-Barriuso, M., Nabais, C., Crespí, A. L., Fernández-Castellano, C., Bernardos, S., and Amich, F. (2011). Morphology and karyology of *Antirrhinum rothmaleri* comb. & stat. nov.(Plantaginaceae), a plant endemic to the NW Iberian Peninsula. *Ann. Bot. Fenn.* 48, 409–421.

Gorospe, J. M., Monjas, D., and Fernández-Mazuecos, M. (2020). Out of the Mediterranean Region: worldwide biogeography of snapdragons and relatives (tribe Antirrhineae, Plantaginaceae). *J. Biogeogr.* 47, 2442–2456. doi: 10.1111/jbi.13939

Grabowski, P. P., Morris, G. P., Casler, M. D., and Borevitz, J. O. (2014). Population genomic variation reveals roles of history, adaptation and ploidy in switchgrass. *Mol. Ecol.* 23, 4059–4073. doi: 10.1111/mec.12845

Güemes, J. (2009). "*Antirrhinum* L," in *Flora Ibérica. Plantas vasculares de la Península Ibérica e Islas Baleares. Plantaginaceae-Scrophulariaceae*, Vol. 13, eds C. C. Benedí, E. Rico, J. Güemes, and A. Herrero (Madrid: Real Jardín Botánico, CSIC), 134–166.

Güemes, J., Andrés, I. M., and Gómez, P. S. (1993). *Antirrhinum subbaeticum* güemes, mateu & sánchez-gómez (Scrophulariaceae), especie nueva de la península Ibérica. *An. Jard. Bot. Madr.* 51, 237–247.

Günther, E., and Rudolph, L. (1970). Kreuzungen zur Ermittlung genetischer Beziehungen innerhalb der Gattung *Antirrhinum*. *Biol. Zentralblatt* 89, 735–750.

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., and Challenger, W. (2007). GEIGER: investigating evolutionary radiations. *Bioinformatics* 24, 129–131. doi: 10.1093/bioinformatics/btm538

Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., and Brumfield, R. T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Syst. Biol.* 65, 910–924. doi: 10.1093/sysbio/syw036

Jiménez, J., Sánchez-Gómez, P., Güemes, J., and Rosselló, J. (2005). Isolated populations or isolated taxa? A case study in narrowly-distributed snapdragons (*Antirrhinum* sect. *Sempervirentia*) using RAPD markers. *Plant Syst. Evol.* 252, 139–152. doi: 10.1007/s00606-004-0250-z

Khimoun, A., Cornuault, J., Burrus, M., Pujol, B., Thebaud, C., and Andalo, C. (2013). Ecology predicts parapatric distributions in two closely related *Antirrhinum majus* subspecies. *Evol. Ecol.* 27, 51–64. doi: 10.1007/s10682-012-9574-2

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, M., Zhang, D., Gao, Q., Luo, Y., Zhang, H., Ma, B., et al. (2019). Genome structure and evolution of *Antirrhinum majus* L. *Nat. Plants* 5, 174–183.

Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., and Chen, S. (2015). Plant DNA barcoding: from gene to genome. *Biol. Rev.* 90, 157–166. doi: 10.1111/brv.12104

Liberal, I. M., Burrus, M., Suchet, C., Thébaud, C., and Vargas, P. (2014). The evolutionary history of *Antirrhinum* in the Pyrenees inferred from phylogeographic analyses. *BMC Evol. Biol.* 14:146. doi: 10.1186/1471-2148-14-146

Linnaeus, C. (1753). *Species Plantarum*. Stockholmiae: Laurentii Salvii.

Magallón, S., and Sanderson, M. J. (2001). Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762–1780. doi: 10.1554/0014-3820(2001)055[1762:adriac]2.0.co;2

Martin, S. H., Davey, J. W., and Jiggins, C. D. (2015). Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32, 244–257. doi: 10.1093/molbev/msu269

Martín-Hernanz, S., Aparicio, A., Fernández-Mazuecos, M., Rubio, E., Reyes-Betancort, J. A., Santos-Guerra, A., et al. (2019). Maximize resolution or minimize error? Using genotyping-by-sequencing to investigate the recent

diversification of *Helianthemum* (Cistaceae). *Front. Plant Sci.* 10:1416. doi: 10.3389/fpls.2019.01416

Mateu-Andrés, I. (1999). Allozymic variation and divergence in three species of *Antirrhinum* L.(Scrophulariaceae-Antirrhineae). *Bot. J. Linn. Soc.* 131, 187–199. doi: 10.1111/j.1095-8339.1999.tb01849.x

Matzke, N. J. (2013). *BioGeoBEARS: Biogeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts*. R Package, Version 0.2 1. Berkeley, CA: University of California.

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES science gateway for inference of large phylogenetic trees," in *Proceedings of the Gateway Computing Environments Workshop*, New Orleans, LA, 1–8.

Otero, A., Vargas, P., Valcárcel, V., Fernández-Mazuecos, M., Jiménez-Mejías, P., and Hipp, A. L. (2019). A snapshot of progenitor-derivative speciation in action in *Iberodes* (Boraginaceae). *bioRxiv* [preprint]. doi: 10.1101/823641

Preston, J. C., Hileman, L. C., and Cubas, P. (2011). Reduce, reuse, and recycle: developmental evolution of trait diversification. *Am. J. Bot.* 98, 397–403. doi: 10.3732/ajb.1000279

Pujol, B., Archambeau, J., Bontemps, A., Lascoste, M., Marin, S., and Meunier, A. (2017). Mountain landscape connectivity and subspecies appurtenance shape genetic differentiation in natural plant populations of the snapdragon (*Antirrhinum majus* L.). *Bot. Lett.* 164, 111–119. doi: 10.1080/23818107.2017.1310056

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Rambaut, A., Suchard, M., Xie, D., and Drummond, A. (2014). *Tracer v1. 6*. Available online at: http://beast.bio.ed.ac.uk/Tracer (accessed May 13, 2018).

R Core Team (2013). *R: a Language and Environment for Statistical Computing*. Version 3.2.3. Vienna, Austria: R Foundation for Statistical Computing.

Ree, R. H., and Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *J. Biogeogr.* 45, 741–749. doi: 10.1111/jbi.13173

Ree, R. H., and Smith, S. A. (2008). Lagrange: software for likelihood analysis of geographic range evolution. *Syst. Biol.* 57, 4–14. doi: 10.1080/10635150701883881

Revell, L. J. (2012). Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210x.2011.00169.x

Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* 46, 195–203. doi: 10.1093/sysbio/46.1.195

Ronquist, F., and Sanmartín, I. (2011). Phylogenetic methods in biogeography. *Annu. Rev. Ecol. Evol. Syst.* 42, 441–464. doi: 10.1146/annurev-ecolsys-102209-144710

Rothmaler, W. (1956). *Taxonomische Monographie der Gattung Antirrhinum. Feddes Repertorium Specierum Novarum Regni Vegetabilis*, Vol. 136. Berlin: Akademie Verlag.

Rundel, P. W., Arroyo, M. T., Cowling, R. M., Keeley, J. E., Lamont, B. B., and Vargas, P. (2016). Mediterranean biomes: evolution of their vegetation, floras, and climate. *Annu. Rev. Ecol. Evol. Syst.* 47, 383–407. doi: 10.1146/annurev-ecolsys-121415-032330

Schwarz-Sommer, Z., Davies, B., and Hudson, A. (2003). An everlasting pioneer: the story of *Antirrhinum* research. *Nat. Rev. Genet.* 4, 655–664. doi: 10.1038/nrg1127

Smith, S. A., and O'Meara, B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28, 2689–2690. doi: 10.1093/bioinformatics/bts492

Staats, M., Cuenca, A., Richardson, J. E., Vrielink-van Ginkel, R., Petersen, G., Seberg, O., et al. (2011). DNA damage in plant herbarium tissue. *PLoS One* 6:e28448. doi: 10.1371/journal.pone.0028448

Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* 57, 758–771. doi: 10.1080/10635150802429642

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nat. Rev. Genet.* 14, 751–764. doi: 10.1038/nrg3483

Sutton, D. (1988). *A Revision of the Tribe Antirrhineae*. Oxford, UK: Oxford University Press.

Swofford, D. L. (2001). *Paup*: Phylogenetic Analysis Using Parsimony (and other Methods)* 4.0. B5. Duke: Department of Biological Sciences, Duke University.

Valente, L. M., Savolainen, V., and Vargas, P. (2010). Unparalleled rates of species diversification in Europe. *Proc. R. Soc. B Biol. Sci.* 277, 1489–1496. doi: 10.1098/rspb.2009.2163

Vargas, P., Carrió, E., Guzmán, B., Amat, E., and Güemes, J. (2009). A geographical pattern of *Antirrhinum* (Scrophulariaceae) speciation since the Pliocene based on plastid and nuclear DNA polymorphisms. *J. Biogeogr.* 36, 1297–1312. doi: 10.1111/j.1365-2699.2008.02059.x

Vargas, P., Fernández-Mazuecos, M., and Heleno, R. (2018). Phylogenetic evidence for a Miocene origin of Mediterranean lineages: species diversity, reproductive traits and geographical isolation. *Plant Biol.* 20, 157–165. doi: 10.1111/plb.12626

Vargas, P., Liberal, I., Ornosa, C., and Gómez, J. M. (2017). Flower specialisation: the occluded corolla of snapdragons (*Antirrhinum*) exhibits two pollinator niches of large long tongued bees. *Plant. Biol.* 19, 787–797. doi: 10.1111/plb.12588

Vargas, P., Ornosa, C., Ortiz-Sánchez, F. J., and Arroyo, J. (2010). Is the occluded corolla of *Antirrhinum* bee-specialized?. *J. Nat. Hist.* 44, 1427–1443. doi: 10.1080/00222930903383552

Vargas, P., Rosselló, J., Oyama, R., and Güemes, J. (2004). Molecular evidence for naturalness of genera in the tribe Antirrhineae (Scrophulariaceae) and three independent evolutionary lineages from the new world and the old. *Plant Syst. Evol.* 249, 151–172. doi: 10.1007/s00606-004-0216-1

Vargas, P., Valente, L. M., Blanco-Pastor, J. L., Liberal, I., Guzmán, B., Cano, E., et al. (2014). Testing the biogeographical congruence of palaeofloras using molecular phylogenetics: snapdragons and the Madrean-Tethyan flora. *J. Biogeogr.* 41, 932–943. doi: 10.1111/jbi.12253

Webb, D. (1971). Taxonomic notes on *Antirrhinum* L. *Bot. J. Linn. Soc.* 64, 271–275.

Wilson, Y., and Hudson, A. (2011). The evolutionary history of *Antirrhinum* suggests that ancestral phenotype combinations survived repeated hybridizations. *Plant J.* 66, 1032–1043. doi: 10.1111/j.1365-313x.2011.04563.x

Zheng, Y., and Wiens, J. J. (2015). Do missing data influence the accuracy of divergence-time estimation with BEAST? *Mol. Phylogenet. Evol.* 85, 41–49. doi: 10.1016/j.ympev.2015.02.002

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 March 2021
doi: 10.3389/fpls.2021.650551

Polyploidy Expands the Range of *Centaurium* (Gentianaceae)

Enrique Maguilla[1], Marcial Escudero[1], Vania Jiménez-Lobato[2], Zoila Díaz-Lifante[1], Cristina Andrés-Camacho[1] and Juan Arroyo[1]*

[1]*Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, Seville, Spain,* [2]*Laboratorio Nacional de Análisis y Síntesis Ecológica, Escuela Superior de Desarrollo Sustentable, Universidad Autónoma de Guerrero – CONACYT, Chilpancingo de los Bravo, Mexico*

The Mediterranean region is one of the most important worldwide hotspots in terms of number of species and endemism, and multiple hypotheses have been proposed to explain how diversification occurred in this area. The contribution of different traits to the diversification process has been evaluated in different groups of plants. In the case of *Centaurium* (Gentianaceae), a genus with a center of diversity placed in the Mediterranean region, polyploidy seems to have been an important driver of diversification as more than half of species are polyploids. Moreover, ploidy levels are strongly geographically structured across the range of the genus, with tetraploids distributed towards more temperate areas in the north and hexaploids in more arid areas towards the south. We hypothesize that the diversification processes and biodiversity patterns in *Centaurium* are explained by the coupled formation of polyploid lineages and the colonization of different areas. A MCC tree from BEAST2 based on three nuclear DNA regions of a total of 26 taxa (full sampling, of 18 species and 8 subspecies) was used to perform ancestral area reconstruction analysis in "BioGeoBEARS." Chromosome evolution was analyzed in chromEvol and diversification in BAMM to estimate diversification rates. Our results suggest that two major clades diverged early from the common ancestor, one most likely in the western Mediterranean and the other in a widespread area including west and central Asia (but with high uncertainty in the exact composition of this widespread area). Most ancestral lineages in the western clade remained in or around the western Mediterranean, and dispersal to other areas (mainly northward and eastward), occurred at the tips. Contrarily, most ancestral lineages in the widespread clade had larger ancestral areas. Polyploidization events in the western clade occurred at the tips of the phylogeny (with one exception of a polyploidization event in a very shallow node) and were mainly tetraploid, while polyploidization events occurred in the widespread clade were at the tips and in an ancestral node of the phylogeny, and were mainly hexaploid. We show how ancestral diploid lineages remained in the area of origin, whereas recent and ancestral polyploidization could have facilitated colonization and establishment in other areas.

Keywords: biogeography, *Centaurium*, chromosome, diversification, Mediterranean, niche, polyploidy

OPEN ACCESS

Edited by:
Andrew A. Crowl,
Duke University, United States

Reviewed by:
Blanca M. Rojas-Andrés,
Leipzig University, Germany
Virginia Valcárcel,
Autonomous University of Madrid, Spain

***Correspondence:**
Marcial Escudero
amesclir@gmail.com

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

Received: *07 January 2021*
Accepted: *17 February 2021*
Published: *10 March 2021*

Citation:
Maguilla E, Escudero M, Jiménez-Lobato V, Díaz-Lifante Z, Andrés-Camacho C and Arroyo J (2021) Polyploidy Expands the Range of Centaurium (Gentianaceae). Front. Plant Sci. 12:650551. doi: 10.3389/fpls.2021.650551

INTRODUCTION

Some areas on Earth have experienced high radiations and become especially rich in the number of species compared to others. These areas of high species richness and elevated percentage of endemism are known as biodiversity hotspots (Myers, 1988). Many factors have been claimed as drivers of increased biodiversity in hotspots, such as long-standing climatic and environmental stability (Fordham et al., 2019), strong spatial variability leading to habitat heterogeneity (e.g., Noroozi et al., 2018), specialized biotic interactions (e.g., Duffy and Johnson, 2017), and prevalence of reproductive systems promoting adaptive novel genetic combinations (Barrett, 2002). One of the drivers which have been proposed is polyploidization, which facilitates both reproductive isolation and adaptation to novel environments (Tate et al., 2005).

Polyploidy has been described to be an important process leading to speciation in plants (Stebbins, 1947; Tate et al., 2005). This mechanism consists of acquiring new sets of chromosomes by hybridization and duplication of the genome (allopolyploidy) or by self-duplication of the whole genome (autopolyploidy). This is a common process; in angiosperms, for example, most species have experienced one or more rounds of polyploidization and subsequent post-polyploid diploidization (Wendel, 2015; Escudero and Wendel, 2020), and at least 47% of species have undergone a recent polyploidy event (Wood et al., 2009). This process can drive speciation promoting reproductive isolation or increasing the probability for new genetic combinations as the species' genetic content increases. Additionally, polyploids have been demonstrated to tolerate wider climatic and ecological conditions and colonize new areas, facilitated by their increased phenotypic plasticity and higher genetic diversity (see Thompson, 2020).

With about 25,000 plant species, of which about 50% are endemic, the Mediterranean basin region has been considered among the most important biodiversity hotspots (Myers et al., 2000). The Mediterranean climate, rugged topography, heterogeneous geological and geomorphological setting, the isolation provided by many mountains and islands, and historical variations have been considered important causal factors for the huge number of species and fast diversification in the region (Buerki et al., 2012; Rundel et al., 2016). For example, with respect to geography, the formation of the current insular system, especially after the Messinian salinity crisis (5.96–5.33 Ma; Duggen et al., 2003) played a key role on the formation of new species by isolation (Krijgsman et al., 1999; Buerki et al., 2012). Moreover, Anatolia, the Balkans, and the Iberian and Italian peninsulas, have served as refugia for many species during the Quaternary glaciations (Hewitt, 2004; Míguez et al., 2017; Thompson, 2020). Thus, the region became not only an important refuge for many animals and plants but also a source area for the colonization of other regions during the Holocene, after glaciations (Hewitt, 2004; Ali et al., 2019; Thompson, 2020). Climatic oscillations during the Quaternary also led species' ranges to contract and expand, producing diversification (Avise, 2000; Maguilla et al., 2017; Schneeweiss et al., 2017). Many studies have attempted to understand how diversification takes place in the Mediterranean region using different study cases (e.g., Barres et al., 2013; Maguilla et al., 2017; Affenzeller et al., 2018, etc.). The importance of different traits in the process of diversification depends on the organism under study, and many traits have been proposed to have an impact on speciation or extinction rates (Fitzjohn, 2010).

The genus *Centaurium* Hill (Gentianaceae Juss.), with ca. 26 taxa (18 species and eight subspecies, Mansion and Struwe, 2004; Díaz-Lifante, 2012) has its center of diversity in the Mediterranean basin, although the range of the genus includes Asia, Europe, North-central Africa, and North America (Mansion and Struwe, 2004). In addition to their natural distribution, members of the genus have been introduced into areas of South America, Australia, and New Zealand with similar climates (Mansion and Struwe, 2004). In this genus, 60% of taxa are polyploids (Zeltner, 1970; Mansion et al., 2005), which have been demonstrated to be important in the evolution of *Centaurium* (Mansion et al., 2005). More interestingly, ploidy levels in the genus seem to conform to a geographical pattern, with diploid species (2*n*) close to the Mediterranean basin, tetraploids (4*n*) mainly in northern Europe and eastern Asia, and hexaploids (6*n*) generally distributed in India, the Arabic Peninsula, and southwestern Mediterranean basin reaching the Canary Islands (Mansion et al., 2005). The hexaploids (50% of polyploid taxa) have been proposed to be allopolyploids, while the tetraploids have been suggested to be autopolyploids (Zeltner, 1970; Mansion et al., 2005). However, in many cases, it is very difficult to distinguish allopolyploids from autopolyploids because it is difficult to determine whether gene flow during the course of evolution has occurred between two species or two lineages within the same species. So far, only one clear case of allopolyploidy has been reported for the genus *Centaurium*, in *Centaurium discolor*, in which the two parental species from two different main clades, *Centaurium maritium* and *Centaurium tenuiflorum* formed a new allopolyploid species (Guggisberg et al., 2006).

In this study, we aim to reconstruct the biogeographic history of *Centaurium* as well as its diversification and chromosome evolution in order to understand the geographic pattern of polyploidy and how polyploidization contributed to the evolution, dispersal, and diversification of the genus from its origin in concert with the range expansion of the genus. We hypothesize that the diversification processes and biodiversity patterns in *Centaurium* are explained by the coupled formation of polyploid lineages and the colonization of new areas. To our knowledge, this is a very innovative study with only one recently published study addressing the same macroevolutionary hypothesis but with a different approach (Han et al., 2020).

MATERIALS AND METHODS

Sampling

Samples were collected in the field and from the following herbaria: MA, MGC, NEU, SANT, SEV, and VAL (Thiers, 2015) by Jiménez-Lobato et al. (2019). A Maximum Clade Credibility (MCC) tree constructed by Jiménez-Lobato et al. (2019) using BEAST 2.4.0 (Bouckaert et al., 2014) was used as input for

our analyses after removing outgroup species. This MCC tree was based on nuclear DNA regions CPR1, ETS, and ITS (see details in Jiménez-Lobato et al., 2019) and includes all known taxa within genus *Centaurium* following taxonomic treatment by Mansion and Struwe (2004), and Díaz-Lifante (2012) for species occurring in the Iberian Peninsula. In total, 26 taxa (18 species and 8 subspecies) were considered for the genus.

Ancestral Area Reconstruction

The package BioGeoBEARS 1.1 (Matzke, 2013) implemented in R 3.2.2 (R Core Team, 2015) was used for the ancestral area reconstruction analyses after coding species' occurrences within six biogeographic areas (see species scoring in **Figure 1**): (1) Western Mediterranean, (2) Eastern Mediterranean, (3) northern Europe, (4) northern Africa excluding the Mediterranean region, (5) western and central Asia, and (6) North America. These areas were delimited based on climatic and paleogeographic data following Buerki et al. (2012) with modifications according to species' distribution patterns in *Centaurium*. The main modification was the split of the Mediterranean into two areas; the western and the eastern Mediterranean, as proposed by other authors (e.g., Takhtajan, 1986) and because many species in the genus are distributed only in the western Mediterranean. Introduced areas were excluded from the scoring of the species' occurrence matrix.

Three biogeographic models were tested in BioGeoBEARS: Dispersal-extinction-cladogenesis (DEC; Richard and Stephen, 2008), dispersal-vicariance-like (DIVA; Ronquist, 1997), and BayArea-like (Bayesian analysis; Landis et al., 2013). Range expansion and contraction (parameters "*d*" and "*e*") were allowed to vary during the analyses. Additionally, biogeographic reconstructions were done including founder events or jump dispersal as an additional free parameter ("*j*"). This reached a total of six combinations of models tested. Moreover, all these analyses were repeated including a connectivity matrix based on Hilpold et al. (2014), where dispersal probability between areas was set to one for adjacent areas, 0.5 for areas with an intermediate area between them, and 0.1 for areas separated by more than

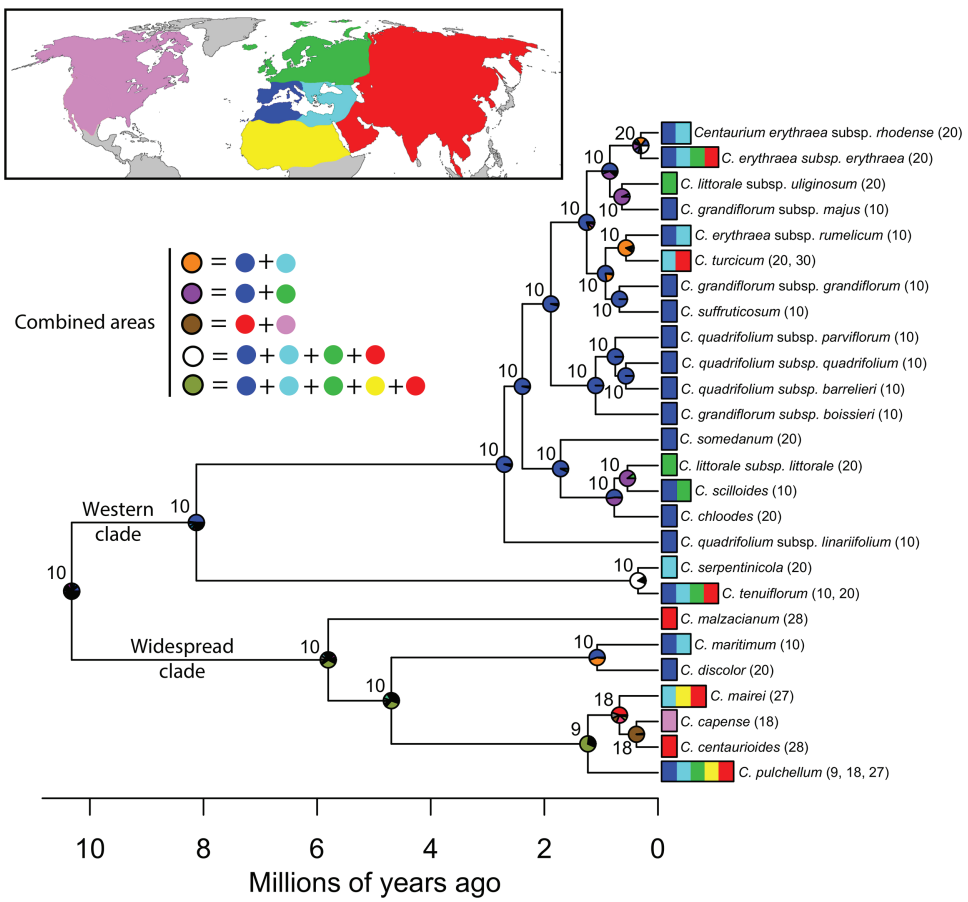


FIGURE 1 | Ancestral area reconstruction under constrained DEC model in *Centaurium* resulting from BioGeoBEARS analysis (Matzke, 2013). Node charts represent the probability of the occurrence of the most recent common ancestor (MRCA) of that node in the area represented by the same color on the map, and combined areas are represented by colors in the legend. Node numbers indicate the reconstructed haploid chromosome numbers retrieved by ChromEvol (Glick and Mayrose, 2014). Tip labels indicate species names and the area where each species occurs following colors in the map. Haploid chromosome number (*n*, in parenthesis) and ploidy levels for the species are also indicated. Lower time-scale shows Millions of years ago, from the origin of the genus to the present.

one area (**Supplementary Table 1**). In these cases, the parameter "*w*" was considered as a free parameter to reduce subjectivity induced by the assignment of a value for dispersal between geographic areas. The Akaike Information Criterion weight (AICw; Akaike, 1974) was used to compare among the 12 model combinations, selecting the best fit model. As broadly discussed by Ree and Sanmartín (2018), models that include jump dispersal ("*j*") tend to overestimate cladogenetic dispersal (or jump dispersal) as the main mechanism explaining species' distribution, considering in many cases null anagenetic rates for the model. For this reason, models including the parameter "*j*" were interpreted with caution and were included mainly to test whether this parameter significantly affected the results. Additionally, to avoid phylogenetic uncertainty bias, we ran in BioGeoBEARS the best fitting model in 100 independent post-burn-in trees from BEAST, and later estimated average probabilities from all trees.

Finally, biogeographic stochastic mapping (BSM; Matzke, 2014) was done in BioGeoBEARS after 50 runs to characterize the number and nature of biogeographical events in the best fitting model. We tested the confidence of the BSM by comparing states probabilities with those obtained under a maximum likelihood (ML) model.

Chromosome Evolution

The evolution of the chromosome number across the phylogeny was predicted using ChromEvol 2.0 (Mayrose et al., 2010; Glick and Mayrose, 2014). Haploid chromosome numbers (*n*) for all species were obtained from Mansion et al. (2005) and Díaz-Lifante (2012). In the case of species with different ploidy levels, the lowest chromosome number was assumed to be ancestral for the analyses (although only three out of 26 taxa had more than one ploidy level: *Centaurium pulchellum*, *Centaurium turcicum*, and *C. tenuiflorum*). A total of 10 models of chromosome evolution were tested. The analyses were performed following Escudero et al. (2014; but see also Glick and Mayrose, 2014; Aparicio et al., 2019). Akaike's information criterion (AIC; Akaike, 1974) was used for comparing and choosing the best fitting model of chromosome evolution, which was later used for the reconstruction and plotting of chromosome numbers in the phylogeny (Glick and Mayrose, 2014).

Diversification Analyses

Diversification rates across the phylogeny were estimated using Bayesian analyses of macroevolutionary mixtures (BAMM; Rabosky et al., 2013, 2014a; Shi and Rabosky, 2015). These analyses were done using BAMM 2.5 in the R package BAMMtools (Rabosky et al., 2014b), allowing changes over time and shifts in diversification rates for five million generations. Markov-Chain Monte Carlo (MCMC) convergence was checked using the R package coda (Plummer et al., 2006), and results of the model with the highest posterior probability (PP) were summarized using BAMMtools.

Pagel's Model for Correlated Evolution

Dependent or independent evolution modes of polyploidy and geographic distribution were also tested using Pagel's method (Pagel, 1994) using the function "fitPagel" from the R package phytools (Revell, 2012). This function searches for correlated evolution of two binary traits. Species and subspecies were coded as diploids against polyploids (including with more than one level of ploidy). Regarding geographic distribution, first, species and subspecies were coded as narrow distributed if they only occur in one area or widespread if they occur in two or more areas. Second, the narrow distributed were coded as exclusively within the area of origin vs. dispersed outside the area of origin, following the biogeographic reconstruction obtained in BioGeoBEARS. For the western clade, the interpretation of the western Mediterranean basin as ancestral range was of high confidence. However, the uncertainty of the ancestral area in the widespread clade was very high. The species with narrow distributions *Centaurium malzacianum*, *C. discolor*, and *Centaurium centauroides* were interpreted as been in their original range as inferred ancestral ranges were widespread ranges including the range of these narrow distributed species. The species *Centaurium capense* from North America was coded as out of the ancestral range based on the reconstruction (as North America was excluded as the possible ancestral distribution in the deeper nodes of the widespread clade). We grouped the widespread species and narrow species out of the ancestral range in one category and the narrow species in the ancestral range in another category. Using these scorings, we tested for the correlation between polyploidy and widespread distribution and/or the expansion to new areas outside of the area of origin of the species or subspecies. We ran models with different rates of transitions between character states ("ARD"), assuming null rates of transition from a polyploid state to a diploid state. Additionally, we performed all analyses using three different models: (i) assuming that the variable polyploidy depends on distribution, (ii) that distribution depends on polyploidy, or (iii) that the two variables depend on each other.

RESULTS

Ancestral Areas

The best biogeographic model based on AICc weights (Akaike, 1974) was DEC + Jc (AICc weight = 0.98; **Table 1**; **Supplementary Figure 1**); this was a constrained DEC model using the dispersal matrix and setting "*j*" as a free parameter. Nevertheless, given the overestimation of founder events explained above, we interpreted this model with caution. The constrained DEC model (DECc; including dispersal matrix), was the best fitting model with an AICc weight of 0.82 when comparing models without "*j*" as a free parameter (**Table 1**). In this case, the model showed an extinction rate ("*e*") of 0.019 events/Myr, and an anagenetic dispersal rate ("*d*") of 0.085 events/Myr (**Table 1**). These results were almost identical to the model obtained under the DECc model in BioGeoBEARS in 100 independent post-burn-in trees (see summarize result in **Supplementary Figure 2**; accordingly, we used the MCC tree results from here).

After the most recent common ancestor (MRCA) of the whole genus *Centaurium* around 10 million years ago (Ma), this phylogeny showed an early split into two major clades

TABLE 1 | Biogeographical models tested in BioGeoBEARS (Matzke, 2013) for *Centaurium*.

Model	LnL	Free params	*d*	*e*	*w*	*j*	AICc	AICcwt	AICcwt without "*j*"
No constraints									
DEC	−72.61	2	0.057	2.0e-09	1	0	149.2	0.0022	0.14
DEC+J	−70.39	3	0.051	1.0e-12	1	0.019	164.8	0.0074	–
DIVALIKE	−78.9	2	0.081	0.019	1	0	161.8	4.1e-06	0.0003
DIVALIKE+J	−78.14	3	0.07	0.0046	1	0.016	162.3	3.2e-06	–
BAYAREALIKE	−81.61	2	0.086	0.3	1	0	167.2	2.7e-07	1.8e-05
BAYAREALIKE+J	−77.18	3	0.054	0.15	1	0.034	160.4	8.3e-06	–
Dispersal multipliers									
DECc	−69.87	3	0.085	0.019	0.43	0	145.7	0.012	0.82
DEC+Jc	−64.5	4	0.084	1.0e-12	2.07	0.032	137	0.98	–
DIVALIKEc	−72.92	3	0.13	0.021	1.58	0	151.8	0.0006	0.039
DIVALIKE+Jc	−72.08	4	0.11	0.0085	1.54	0.027	152.2	0.0005	–
BAYAREALIKEc	−80.92	3	0.1	0.28	0.12	0	167.8	2.0e-07	1.3e-05
BAYAREALIKE+Jc	−73.25	4	0.09	0.17	1.05	0.062	154.5	0.0002	–

Dispersal-extinction-cladogenesis (DEC), dispersal-vicariance like (DIVA), and Bayesian binary-like (BAYAREALIKE) models were tested with and without the "j" parameter which allows founder speciation events (cladogenetic dispersal rate). Additionally, all analyses were performed adding a dispersal multiplier matrix where dispersal probability was set to 1 for adjacent areas, 0.5 for areas with an intermediate area between them, and 0.1 for areas separated by more than one area. In these cases, "w" was considered as a free parameter to reduce subjectivity induced by the assignment of a given value for dispersal between geographic areas. LnL indicates Log-likelihood for each model tested, numbers of free parameters included as "Free Params," "d" for anagenetic dispersal rate, "e" for extinction, Akaike's information criterion corrected for small sample size (AICc), Akaike's information criterion weights for all eight models (AICcwt) and considering only models without "j" parameter. The best fitted model without "j" is highlighted in bold type.

(**Figure 1**). For the first major clade (hereafter the western clade) comprising most species and subspecies from the western Mediterranean and some species and subspecies from surrounding areas (the rest of the Palaearctic), the MRCA was most likely predicted to be from the western Mediterranean (**Figure 1**; other possible ancestral ranges were unlikely and included western Mediterranean as part of the ancestral distribution). The origin of the MRCA for the second major clade (hereafter the widespread clade) was unresolved, as occurred with other internal nodes within this clade, suggesting wider ancestral areas. This second major clade was composed of *C. capense*, *C. centaurioides*, *C. discolor*, *Centaurium mairei*, *C. malzacianum*, *Centaurium maritimum* and the broadest distributed species in the genus, *C. pulchellum*, which is present in all codified areas (but North America). In the evolution of *Centaurium*, anagenetic dispersal was apparently the most important mechanism explaining the current distribution and biogeographic history of the genus (a mean of 21.28 ± 2.02 events after 50 BSMs; **Figure 2**). The area of origin of most *Centaurium* species and subspecies, the western Mediterranean, acted as a source for many dispersals to other areas (41.64% of all dispersal events; **Figures 1, 3**). Most of these dispersals took place from the western Mediterranean to the eastern Mediterranean and northern Europe (a mean of 3.64 and 3.46 events, respectively; **Figure 3**), and most of them took place towards the tips of the phylogeny (**Figure 1**). This mechanism was followed in importance by within-area speciation, which showed a mean of 12.62 ± 1.48 events after 50 BSM runs. Less important were subset within-area speciation (a mean of 8.18 ± 1.72 events) and vicariance (4.2 ± 1.55 events; **Figure 2**). Finally, a comparison of state probabilities obtained using Bayesian stochastic mapping and ML approaches were compared to test the suitability of the model, obtaining an R^2 of 0.985 (**Supplementary Figure 3**), which confirmed that states predicted after the 50 BSMs were strongly congruent with a ML approach.

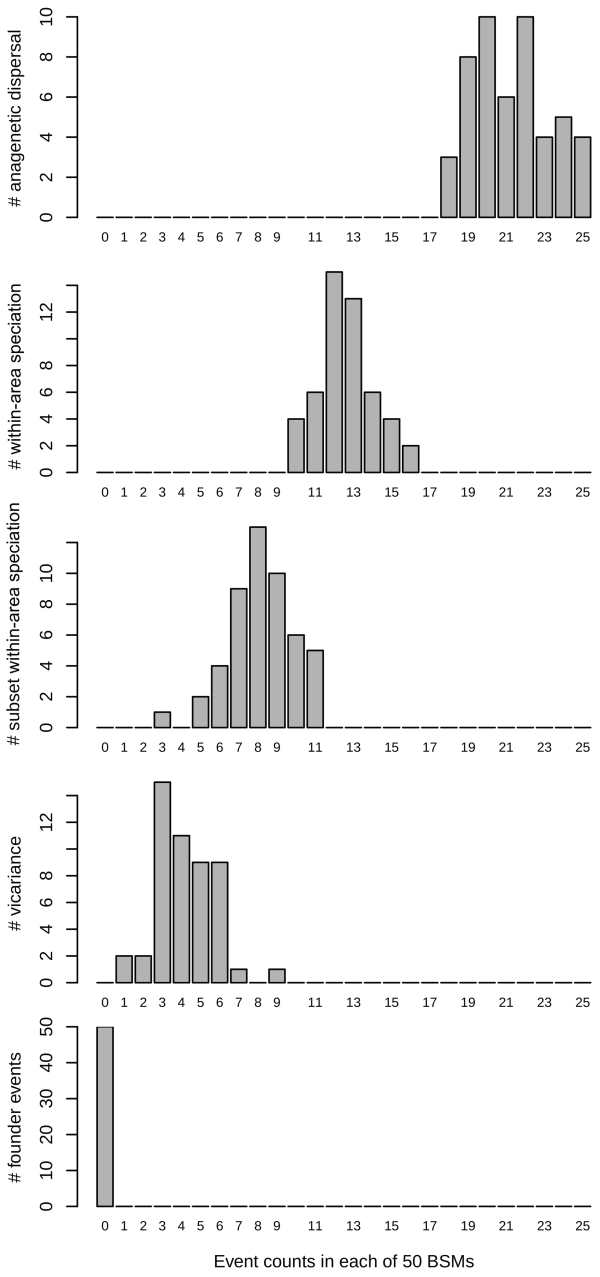


FIGURE 2 | Biogeographical events count after 50 biogeographic stochastic mappings (BSMs) in the best fitting model obtained in BioGeoBEARS (constrained DEC model) in *Centaurium*. The horizontal (*x*) axis represents the number of events, and the vertical (*y*) axis the number of BSM analyses.

Diversification Rates

The BAMM analyses showed that diversification rates were slightly increasing towards the present in *Centaurium* (**Figure 4**). A scenario without shifts in diversification rates was the most plausible (probability = 0.811), followed by a scenario with one shift in diversification rates, though this scenario had a probability of only 0.137. The speciation rate for the most supported model (0-shifts) was 0.654 lineage per million years (lin./Myr; lam1), with an extinction rate of 0.546 lin./Myr (mu1) and speciation growth parameter of 0.018 lin./Myr (lam2; **Figure 4**).

Chromosomal Evolution and Biogeography

The best-fitting model for chromosome evolution based on AIC values was "BASE_NUM_DUPL" (AIC = 116.989; see **Supplementary Table 2** for AIC values of all models tested), which considered that chromosome changes were driven by the gain or loss of a chromosome pair, whole-genome duplication (polyploidization) or changes in the basic number (Glick and Mayrose, 2014). This scenario of chromosome evolution showed a rate of 0.52 loss events/Myr, and a gain of 0.00 events/Myr. The rate for duplication of the chromosome number was 0.81 duplications/Myr. For the MRCA of *Centaurium*, the estimated chromosome number was n = 10 (**Figure 1**). This is expected to be the ancestral chromosome number for the genus. Our results showed how polyploidization occurred mainly in the tips of the phylogeny in the case of the large western clade (only one polyploidization event inferred in the very shallow

	WMED	EMED	NEUR	NAFR	WCA	NAM	
WMED	-	3.64 (1.14)	3.46 (0.93)	0.4 (0.7)	1.28 (0.95)	0.08 (0.27)	8.86 (41.64%)
EMED	1.16 (0.93)	-	0.54 (0.71)	0.66 (0.66)	1.84 (1)	0	4.2 (19.74%)
NEUR	0.58 (0.7)	0.96 (0.78)	-	0.12 (0.33)	0.86 (0.76)	0.06 (0.24)	2.58 (12.12%)
NAFR	0.22 (0.51)	0.44 (0.64)	0.14 (0.35)	-	0.56 (0.84)	0.02 (0.14)	1.38 (6.48%)
WCA	0.3 (0.58)	1.3 (1.13)	0.52 (0.62)	0.88 (0.72)	-	0.98 (0.38)	3.98 (18.7%)
NAM	0.04 (0.2)	0.02 (0.14)	0	0.02 (0.14)	0.2 (0.45)	-	0.28 (1.32%)
	2.3 (10.81%)	6.36 (29.89%)	4.66 (21.9%)	2.08 (9.77%)	4.74 (22.27%)	1.14 (5.36%)	21.28 (100%)

FIGURE 3 | Average number of dispersals and standard deviation (in parenthesis) after 50 BSMs obtained from BioGeoBEARS (Matzke, 2013) under the constrained DEC model in *Centaurium*. Source and sink areas are indicated in the *y* and *x* axes respectively, codified as follows: WMED for western Mediterranean; EMED, eastern Mediterranean; NEUR, northern Europe; NAFR, northern Africa; WCA, west and central Asia; and NAM for North America. The frequency of each event is also indicated by the cell color (green colors for more frequent events, and warmer colors for infrequent or null events). The last row and column summarize the dispersal events or destination of each area, respectively, indicating also the percentage of all events.

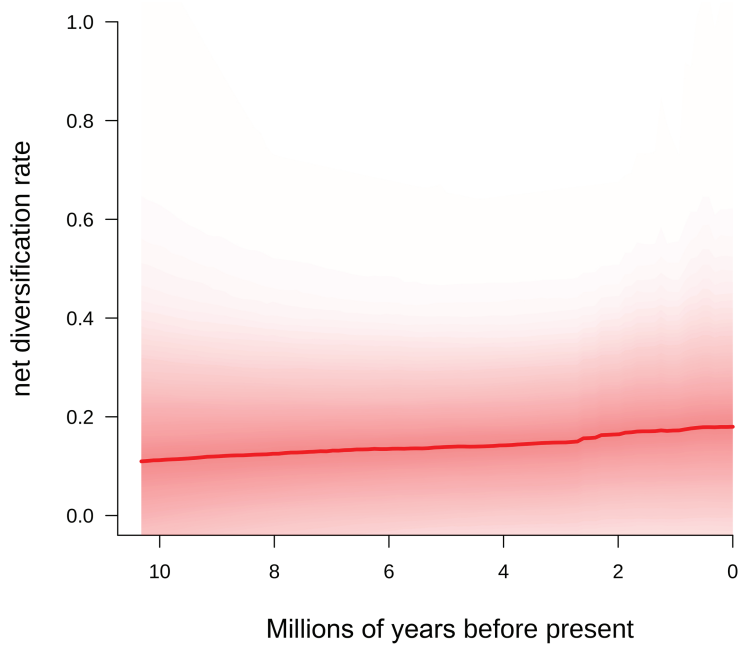


FIGURE 4 | Diversification rate from the origin of the genus *Centaurium* to the present obtained using BAMM (Rabosky et al., 2013, 2014a; Shi and Rabosky, 2015).

node of the subspecies *Centaurium erythraea* ssp. *erythraea* and ssp. *rhodense*), whereas in the widespread clade, there is one ancestral polyploidization event reconstructed at the MRCA node of the species *C. mairei*, *C. capense*, and *C. centauroides*. Specifically, a chromosome loss event was predicted for the MRCA of the clade composed of *C. capense*, *C. centaurioides*, *C. mairei*, and *C. pulchellum* (base number shifts from 10 to 9), and a later polyploidization event in the case of the MRCA of *C. capense*, *C. centaurioides*, and *C. mairei*.

The analyses using Pagel's method retrieved dependent correlated evolution between polyploidy and geography, where polyploidy depended on the area where the species and subspecies occurs (**Figure 5**). Initially, the dispersal rate was independent of polyploidy (from the area of origin to new

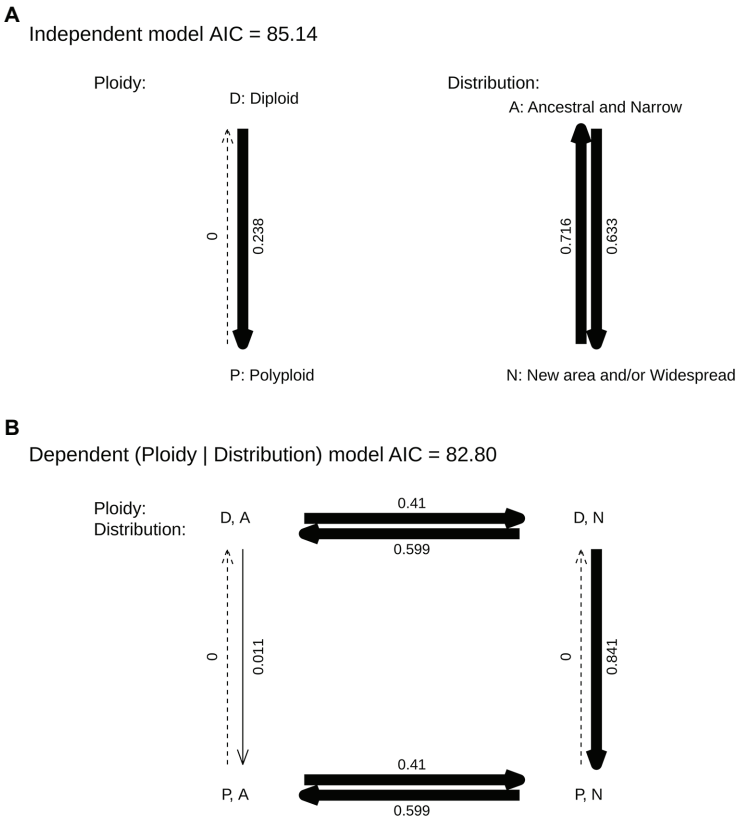


FIGURE 5 | Pagel's binary correlation test of polyploidy (D for diploid, P for polyploid) and the presence in the area of origin of the species (A if the species is in the ancestral area, N if the species has colonized a new area even if it is still present in its ancestral area and/or have widespread distribution) in *Centaurium*. Independent **(A)** and dependent **(B)** models are shown. Arrow thickness indicates rate proportion.

areas 0.410 events/Myr and *vice-versa*, 0.599 events/Myr). Second, whereas transitions from diploidy to polyploidy were almost null in the area of origin (0.011 events/Myr), the rate of transition from diploidy to polyploidy in species and subspecies inhabiting new areas or with widespread distribution was 0.841 events/Myr.

DISCUSSION

Origin and Evolution of *Centaurium*

Based on the best fitting model (DECc, AICw = 0.82; **Table 1**), two main clades – the western and widespread clade – originated from their MRCA. Nevertheless, the best fitting model (regardless of whether founder event "*j*" was included as a free parameter) could not confidently reconstruct the origin of the MRCA of the whole genus *Centaurium* (**Table 1**; **Figure 1**; **Supplementary Figure 1**), originated during the Miocene (Jiménez-Lobato et al., 2019). The western clade that originated most likely in the western Mediterranean, which includes almost 70% of extant taxa in the genus (**Figure 1**), started diversifying about 8.1 Ma (Jiménez-Lobato et al., 2019). Most taxa in this clade occur in the western Mediterranean, only one species with two subspecies (*Centaurium littorale*) is not present in the Mediterranean but only in northern Europe. The other clade (with ca. 30% of extant taxa) has up to 40% of the species and subspecies with widespread ranges and combining different distribution patterns including the six coding areas in our analyses. For example, *C. pulchellum* is present in all of the codified areas (but North America), *C. mairei* in the eastern Mediterranean, Africa, and west and central Asia, and *C. maritimum* in the whole Mediterranean region, whereas other species have more restricted distribution; for example, *C. capense* is found in North America and *C. centaurioides* and *C. malzacianum* in west and central Asia, and *C. discolor* in the western Mediterranean (**Figure 1**). This complex and wide range of species occurrences made BioGeoBEARS unable to reconstruct the geographic origin of deeper nodes in this clade, leaving the biogeographic history of these species partially unknown. The ancestral distribution of the widespread clade was inferred as a broad distribution including west and central Asia but the specific composition of this inferred broad ancestral distribution is very uncertain.

In *Centaurium*, as in many other groups of plants, the Mediterranean basin acted as a source for later colonization of adjacent areas after glaciations (Thompson, 2020). The origin of most dispersal events in the genus was the western Mediterranean (**Figure 3**). The most common destinations were

adjacent areas, such as the eastern Mediterranean and northern Europe when the origin was the western Mediterranean (with a mean of 3.64 and 3.46 dispersal events, respectively; **Figure 3**), and the western Mediterranean and west and central Asia in the case of dispersal from the eastern Mediterranean (a mean of 1.16 and 1.84 events, respectively; **Figure 3**). Interestingly, most dispersal events occurred in recent times in the phylogeny, and all species from the Mediterranean dispersed northward and eastward. Thus, the colonization of southern latitudes, i.e., northern Africa, occurred mainly from western and central Asia and not from the western Mediterranean (**Figure 3**). Regarding the colonization of North America, the MRCA of *C. capense* and *C. centaurioides* presumably dispersed from Asia to North America over the Bering Land Bridge (BLB), when this was open during glacial periods and later diversified into two species: *C. capense* from North America, and *C. centaurioides* in west and central Asia, probably during the past million years (**Figures 1, 3**). The BLB connected Asia and North America during periods where the sea level descended as a consequence of ice formation during glacial periods (Hopkins, 1972; Brikiatis, 2014). The formation of the BLB occurred alternatively during glacial and interglacial periods since the Pliocene (Harris, 2005) and was used by many plant species as a route of dispersal from one continent to the other (e.g., Harris, 2007; Waltari et al., 2007; Mao et al., 2010; Maguilla et al., 2018; among others). The crown node of the separation between *C. capense* and *C. centaurioides* was dated by Jiménez-Lobato et al. (2019) sometime during the Pleistocene. This coincides with a period where the BLB was successively closed and then Asia and North America separated (Harris, 2007; **Figure 2**), supporting the role of the Bering Strait and, more particularly, the alternate opening and closure of the BLB as a dispersal and speciation driver in this clade (**Figure 1**).

Thus, the importance of the Mediterranean basin for *Centaurium* lies not only in the fact that it is the place of origin of the MRCA of 70% of the genus (**Figure 1**), but also as a region where lineages diversified, showing a constant increase in diversification rates from the origin to the present (**Figure 4**). Presumably, the role of some parts of the Mediterranean as glacial refugia (Petit et al., 2003) allowed diversification of the genus in the region, and were the origins for later dispersals, mainly to adjacent areas during interglacial periods, where more favorable niches were available in upper latitudes.

Polyploidy as a Key to Expand the Range

Diversification in *Centaurium* was dominated by multiple events of anagenetic dispersals, followed by within-area speciation, then less frequent subset within-area speciation (this is, speciation within part of the distribution of the ancestor but not over the whole area) and rarely, vicariance, as commented above (**Figure 2**). According to biogeographic reconstruction (**Figure 1**), most within-area speciation occurred in the western Mediterranean, where most of the species and subspecies in the genus originated. Later, lineages diversified at a constantly increasing rate from the origin to the tips of the phylogeny (**Figure 4**), while at the same time dispersing, mainly into adjacent areas (**Figure 3**). When a plant species disperses or colonizes a new area or niche, the success of the event depends on multiple traits. High phenotypic plasticity, capacity for sexual reproduction, high competitiveness with other species or wide range of climatic tolerance are only a few examples of factors affecting the success of establishment in a new area (Baker, 1965; Levin, 2000; Maguilla et al., under review). Polyploidization has been described to have important evolutionary consequences and correlate strongly with range expansion in plants (Levin, 1983; Hijmans et al., 2007; Treier et al., 2009; Te Beest et al., 2012; Soltis et al., 2015). In the Mediterranean region, polyploidy is common among plants, and polyploidization has been demonstrated to be an important trait for diversification in species' lineages within the region (Vilatersana et al., 2000; Escudero et al., 2018; Thompson, 2020), although it can either increase or decrease rates of diversification (Escudero et al., 2018).

Cytogenetic studies reported two basic chromosome numbers for *Centaurium* (x = 9 and x = 10; Zeltner, 1970) and during the evolutionary history of the genus, chromosome numbers remained invariable at deeper nodes from an ancestral chromosome number of n = 10 (**Figure 1**). In the western clade, with most species and subspecies occurring in the area of origin, polyploidization events took place at the tips (with only one exception in a very shallow clade that contains two subspecies), whereas in the widespread clade, with up to ca. 40% of species broadly distributed and inferred broad ancestral areas, a polyploidization event has been reconstructed to be older, as reflected in **Figure 1**. It is interesting that the widespread clade where up to 40% of species are broadly distributed is also the clade where the percentage of polyploidization events is highest and events deeper in our chromosome number reconstruction (**Figure 1**). Whereas in the western clade the percentage of polyploids is ca. 50% (mostly tetraploids), the percentage of polyploids in the widespread clade rises to 70% (mostly hexaploids that curiously mostly avoid the Mediterranean region with a clear preference for more arid regions). This widespread clade includes for example *C. pulchellum*, with a widespread distribution across the northern hemisphere, occurring in all five areas codified for our analyses but North America (**Figure 1**). Except for *C. maritimum* from the western and eastern Mediterranean, all other species in this clade (i.e., *C. capense*, *C. centaurioides*, *C. discolor*, *C. mairei*, *C. malzacianum*, and *C. pulchellum*) show some level of polyploidization (**Figure 1**).

When analyzing the possible correlation between polyploidy and distribution, we found a strong dependence of polyploidy on distribution (**Figure 5**). Thus, our results confirmed that polyploidy occurred in species and subspecies that are widespread or dispersed out of the area of origin, and suggest that it may have facilitated their establishment and success in the newly colonized area. Thus, polyploidization does not seem to facilitate the dispersal event *per se*, but rather success in the new or expanded area (**Figure 5**). Additionally, Mansion et al. (2005) represented geographically the ploidy level of *Centaurium* species in the Mediterranean basin showing that tetraploids dominated upper latitudes, whereas hexaploids are more common in southern latitudes and diploids in the core of the Mediterranean

basin. This is congruent with polyploidy favoring adaptation to different niches or new areas.

CONCLUSION

We demonstrated how *Centaurium* species in the western clade remaining in their ancestral area, the western Mediterranean, are diploids. Both recent and ancestral polyploidization events have facilitated the expansion of the distribution range and/or the colonization of new areas outside the areas of origin. Interestingly, *Centaurium* tetraploid species were more successful colonizing north of the Mediterranean in a more temperate climatic regime. Hexaploid species seem to be more successful growing in southern places in a more arid climatic regime (although our results are not conclusive on whether or not the current range of the hexaploid species is ancestral or recently colonized).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JA and ME conceived the research topic. CA-C, VJ-L, and ZD-L gathered chromosome data and helped assessing plant taxonomy and distributions. EM and ME performed the analyses. EM wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

Project (PGC2018-099608-B-100) funded by: FEDER/Ministerio de Ciencia e Innovación – Agencia Estatal de Investigación. The present work was funded by the Spanish Government and FEDER funds (European Commission) through granted projects CGL2013-45037-P and PGC2018-099608-B-100 to JA, ME, CA-C, EM, and ZD-L, and fellowship FJCI-2017-32314 ("Juan de la Cierva – Formación") to EM granted by the Spanish Government.

ACKNOWLEDGMENTS

We are grateful to and dedicated this manuscript to Dr. L. Zeltner for sharing cytogenetic information about the genus *Centaurium*, which was useful for clarifying the ploidy level of the species. We thank B. M. Rojas-Andrés and V. Valcarcel as reviewers, and associate editor A. A. Crowl for their valuable comments improving the first versions of this manuscript. General Research Services at Universidad de Sevilla (CITIUS2, herbarium and DNA lab), for allowing lab work, and CICA (Scientific Andalusian Informatics Center) for providing computational resources for the analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.650551/full#supplementary-material

REFERENCES

Affenzeller, M., Kadereit, J. W., and Comes, H. P. (2018). Parallel bursts of recent and rapid radiation in the Mediterranean and Eritreo-Arabian biodiversity hotspots as revealed by *Globularia* and *Campylanthus* (Plantaginaceae). *J. Biogeogr.* 45, 552–566. doi: 10.1111/jbi.13155

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–7230.

Ali, T., Muñoz-Fuentes, V., Buch, A. K., Çelik, A., Dutbayev, A., Gabrielyan, I., et al. (2019). Out of Transcaucasia: origin of Western and central Palearctic populations of *Microthlaspi perfoliatum*. *Flora Morphol. Distrib. Funct. Ecol. Plants* 253, 127–141. doi: 10.1016/j.flora.2019.02.012

Aparicio, A., Escudero, M., Valdés-Florido, A., Pachón, M., Rubio, E., Albaladejo, R. G., et al. (2019). Karyotype evolution in *Helianthemum* (Cistaceae): Dysploidy, achiasmate meiosis and ecological specialization in *H. squamatum*, a true gypsophile. *Bot. J. Linn. Soc.* 191, 484–501. doi: 10.1093/botlinnean/boz066

Avise, J. C. (2000). *Phylogeography: The history and formation of species*. Cambridge, Massachussets: Harvard Un.

Baker, H. G. (1965). "Characteristics and modes of origin of weeds" in *The genetics of colonizing species*. eds. H. G. Baker and G. L. Stebbins (NY: Academic Press Inc.), 147–172.

Barres, L., Sanmartín, I., Anderson, C. L., Susanna, A., Buerki, S., Galbany-Casals, M., et al. (2013). Reconstructing the evolution and biogeographic history of tribe Cardueae (compositae). *Am. J. Bot.* 100, 867–882. doi: 10.3732/ajb.1200058

Barrett, S. C. H. (2002). The evolution of plant sexual diversity. *Nat. Rev. Genet.* 3, 274–284. doi: 10.1038/nrg776

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* 10, 1–6. doi: 10.1371/journal.pcbi.1003537

Brikiatis, L. (2014). The de geer, thulean and beringia routes: key concepts for understanding early cenozoic biogeography. *J. Biogeogr.* 41, 1036–1054. doi: 10.1111/jbi.12310

Buerki, S., Jose, S., Yadav, S. R., Goldblatt, P., Manning, J. C., and Forest, F. (2012). Contrasting biogeographic and diversification patterns in two mediterranean-type ecosystems. *PLoS One* 7:e39377. doi: 10.1371/journal.pone.0039377

Díaz-Lifante, Z. (2012). "Centaurium" in *Flora Iberica 11*. eds. C. Romero and A. Quintanar (Madrid: Real Jardín Botánico, CSIC), 49–81.

Duffy, K. J., and Johnson, S. D. (2017). Specialized mutualisms may constrain the geographical distribution of flowering plants. *Proc. R. Soc. B Biol. Sci.* 284:20171841. doi: 10.1098/rspb.2017.1841

Duggen, S., Hoernie, K., Van den Bogaard, P., Rüpke, L., and Morgan, J. P. (2003). Deep roots of the Messinian salinity crisis. *Nature* 422, 602–606. doi: 10.1038/nature01553

Escudero, M., Balao, F., Martín-Bravo, S., Valente, L., and Valcárcel, V. (2018). Is the diversification of Mediterranean Basin plant lineages coupled to karyotypic changes? *Plant Biol.* 20, 166–175. doi: 10.1111/plb.12563

Escudero, M., Martín-Bravo, S., Mayrose, I., Fernández-Mazuecos, M., Fiz-Palacios, O., Hipp, A. L., et al. (2014). Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. *PLoS One* 9:e85266. doi: 10.1371/journal.pone.0085266

Escudero, M., and Wendel, J. F. (2020). The grand sweep of chromosomal evolution in angiosperms. *New Phytol.* 228, 805–808. doi: 10.1111/nph.16802

Fitzjohn, R. G. (2010). Quantitative traits and diversification. *Syst. Biol.* 59, 619–633. doi: 10.1093/sysbio/syq053

Fordham, D. A., Brown, S. C., Wigley, T. M. L., and Rahbek, C. (2019). Cradles of diversity are unlikely relics of regional climate stability. *Curr. Biol.* 29, R356–R357. doi: 10.1016/j.cub.2019.04.001

Glick, L., and Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* 31, 1914–1922. doi: 10.1093/molbev/msu122

Guggisberg, A., Bretagnolle, F., and Mansion, G. (2006). Allopolyploid origin of the Mediterranean endemic, *Centaurium bianoris* (Gentianaceae), inferred by molecular markers. *Syst. Bot.* 31, 368–379. doi: 10.1600/036364406777585937

Han, T. S., Zheng, Q. J., Onstein, R. E., Rojas-Andrés, B. M., Hauenschild, F., Muellner-Riehl, A. N., et al. (2020). Polyploidy promotes species diversification of *Allium* through ecological shifts. *New Phytol.* 225, 571–583. doi: 10.1111/nph.16098

Harris, S. A. (2005). Thermal history of the Arctic Ocean environs adjacent to North America during the last 3.5 Ma and a possible mechanism for the cause of the cold events (major glaciations and permafrost events). *Prog. Phys. Geogr.* 29, 218–238. doi: 10.1191/0309133305pp444ra

Harris, S. A. (2007). Biodiversity of the alpine vascular flora of the N.W. north American cordillera: the evidence from phyto-geography. *Erdkunde* 61, 344–356. doi: 10.3112/erdkunde.2007.04.05

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the quaternary. *Philos. Trans. R. Soc. B Biol. Sci.* 359, 183–195. doi: 10.1098/rstb.2003.1388

Hijmans, R. J., Gavrilenko, T., Stephenson, S., Bamberg, J., Salas, A., and Spooner, D. M. (2007). Geographical and environmental range expansion through polyploidy in wild potatoes (*Solanum* section *Petota*). *Glob. Ecol. Biogeogr.* 16, 485–495. doi: 10.1111/j.1466-8238.2007.00308.x

Hilpold, A., Vilatersana, R., Susanna, A., Meseguer, A. S., Boršić, I., Constantinidis, T., et al. (2014). 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

Hopkins, D. M. (1972). The paleogeography and climatic history of Beringia during late Cenozoic time. *Inter-Nord* 12, 121–150.

Jiménez-Lobato, V., Escudero, M., Díaz Lifante, Z., Andrés Camacho, C., De Castro, A., Mansion, G., et al. (2019). Evolution of reproductive traits and selfing syndrome in the sub-endemic Mediterranean genus *Centaurium* Hill (Gentianaceae). *Bot. J. Linn. Soc.* 191, 216–235. doi: 10.1093/botlinnean/boz036

Krijgsman, W., Hilgen, F. J., Raffi, I., Sierro, F. J., and Wilson, D. S. (1999). Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400, 652–655. doi: 10.1038/23231

Landis, M., Matzke, N., Mooer, B., and Huelsenbeck, J. (2013). Bayesian analysis of biogeography when the number of areas is large. *Syst. Biol.* 62, 789–804. doi: 10.1093/sysbio/syt040

Levin, D. A. (1983). Polyploidy and novelty in flowering plants. *Am. Nat.* 122, 1–25. doi: 10.1086/284115

Levin, D. A. (2000). *The origin, expansion, and demise of plant species*. Oxford: Oxford Univ. Press.

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

Maguilla, E., Escudero, M., and Luceño, M. (2018). Vicariance versus dispersal across Beringian land bridges to explain circumpolar distribution: a case study in plants with high dispersal potential. *J. Biogeogr.* 45, 771–783. doi: 10.1111/jbi.13157

Mansion, G., and Struwe, L. (2004). Generic delimitation and phylogenetic relationships within the subtribe *Chironiinae* (Chironieae: Gentianaceae), with special reference to *Centaurium*: evidence from nrDNA and cpDNA sequences. *Mol. Phylogenet. Evol.* 32, 951–977. doi: 10.1016/j.ympev.2004.03.016

Mansion, G., Zeltner, L., and Bretagnolle, F. (2005). Phylogenetic patterns and polyploid evolution within the Mediterranean genus *Centaurium* (Gentianaceae - Chironieae). *Taxon* 54, 931–950. doi: 10.2307/25065479

Mao, K., Hao, G., Liu, J., Adams, R. P., and Milne, R. I. (2010). Diversification and biogeography of *Juniperus* (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. *New Phytol.* 188, 254–272. doi: 10.1111/j.1469-8137.2010.03351.x

Matzke, N. J. (2013). Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* 5, 242–248. doi: 10.21425/F55419694

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Syst. Biol.* 63, 951–970. doi: 10.1093/sysbio/syu056

Mayrose, I., Barker, M. S., and Otto, S. P. (2010). Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst. Biol.* 59, 132–144. doi: 10.1093/sysbio/syp083

Míguez, M., Gehrke, B., Maguilla, E., Jiménez-Mejías, P., and Martín-Bravo, S. (2017). *Carex* sect. *Rhynchocystis* (Cyperaceae): a Miocene subtropical relict in the western Palaearctic showing a dispersal derived Rand Flora pattern. *J. Biogeogr.* 44, 2211–2224. doi: 10.1111/jbi.13027

Myers, N. (1988). Threatened biotas: "hot spots" in tropical forests. *Environmentalist* 8, 187–208.

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

Noroozi, J., Talebi, A., Doostmohammadi, M., Rumpf, S. B., Linder, H. P., and Schneeweiss, G. M. (2018). Hotspots within a global biodiversity hotspot-areas of endemism are associated with high mountain ranges. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-018-28504-9

Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. B Biol. Sci.* 255, 37–45. doi: 10.1098/rspb.1994.0006

Petit, R. J., Aguinagalde, I., de Beaulieu, J. -L., Bittkau, C., Brewer, S., Cheddadi, R., et al. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300, 1563–1565. doi: 10.1126/science.1083264

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News* 6, 7–11.

Rabosky, D. L., Donnellan, S. C., Grundler, M., and Lovette, I. J. (2014a). Analysis and visualization of complex macroevolutionary dynamics: an example from Australian Scincid lizards. *Syst. Biol.* 63, 610–627. doi: 10.1093/sysbio/syu025

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., et al. (2014b). BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* 5, 701–707. doi: 10.1111/2041-210X.12199

Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., et al. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* 4:1958. doi: 10.1038/ncomms2958

R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Found. Stat. Comput.

Ree, R. H., and Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC *via* model selection. *J. Biogeogr.* 45, 741–749. doi: 10.1111/jbi.13173

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Richard, H. R., and Stephen, A. S. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57, 4–14. doi: 10.1080/10635150701883881

Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* 46, 195–203. doi: 10.1093/sysbio/46.1.195

Rundel, P. W., Arroyo, M. T., Cowling, R. M., Keeley, J. E., Lamont, B. B., and Vargas, P. (2016). Mediterranean biomes: evolution of their vegetation, floras, and climate. *Annu. Rev. Ecol. Evol. Syst.* 47, 383–407. doi: 10.1146/annurev-ecolsys-121415-032330

Schneeweiss, G. M., Winkler, M., and Schönswetter, P. (2017). Secondary contact after divergence in allopatry explains current lack of ecogeographical isolation in two hybridizing alpine plant species. *J. Biogeogr.* 44, 2575–2584. doi: 10.1111/jbi.13071

Shi, J. J., and Rabosky, D. L. (2015). Speciation dynamics during the global radiation of extant bats. *Evolution* 69, 1528–1545. doi: 10.1111/evo.12681

Soltis, P. S., Marchant, D. B., Van de Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi: 10.1016/j.gde.2015.11.003

Stebbins, G. L. (1947). Types of polyploids: their classification and significance. *Adv. Genet.* 1, 403–429. doi: 10.1016/S0065-2660(08)60490-3

Takhtajan, A. (1986). *Floristic regions of the world*. Berkeley: University of California Press.

Tate, J. A., Soltis, P. S., and Soltis, D. E. (2005). "Polyploidy in plants" in *The evolution of the genome*. eds. T. R. Gregory (Elsevier Science & Technology, Academic Press), 371–426.

Te Beest, M., Le Roux, J. J., Richardson, D. M., Brysting, A. K., Suda, J., Kubešová, M., et al. (2012). The more the better? The role of polyploidy in facilitating plant invasions. *Ann. Bot.* 109, 19–45. doi: 10.1093/aob/mcr277

Thiers, B. (2015). *Index Herbariorum: A global directory of public herbaria and associated staff*. New York: Botanical Garden's Virtual Herbarium.

Thompson, J. D. (2020). *Plant evolution in the Mediterranean*. Oxford: Oxford University Press.

Treier, U. A., Broennimann, O., Normand, S., Guisan, A., Schaffner, U., Steinger, T., et al. (2009). Shift in cytotype frequency and niche space in the invasive plant *Centaurea maculosa*. *Ecology* 90, 1366–1377. doi: 10.1890/08-0420.1

Vilatersana, R., Susanna, A., Garcia-Jacas, N., and Garnatje, T. (2000). Karyology, generic delineation and dysploidy in the genera *Carduncellus*, *Carthamus* and *Phonus* (Asteraceae). *Bot. J. Linn. Soc.* 134, 425–438. doi: 10.1006/bojl.2000.0349

Waltari, E., Hoberg, E. P., Lessa, E. P., and Cook, J. A. (2007). Eastward ho: Phylogeographical perspectives on colonization of hosts and parasites across the Beringian nexus. *J. Biogeogr.* 34, 561–574. doi: 10.1111/j.1365-2699.2007.01705.x

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102, 1753–1756. doi: 10.3732/ajb.1500320

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., and Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci.* 106, 13875–13879. doi: 10.1073/pnas.0811575106

Zeltner, L. (1970). Recherches de biosystématique Sur les genres *Blackstonia* Huds. Et *Centaurium* Hill (Gentianaceae). *Bull. la Société Neuchâteloise des Sci. Nat.* 93, 1–164.

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 27 April 2021
doi: 10.3389/fpls.2021.627909

Multiple Drivers of High Species Diversity and Endemism Among *Alyssum* Annuals in the Mediterranean: The Evolutionary Significance of the Aegean Hotspot

*Veronika Cetlová[1], Judita Zozomová-Lihová[1], Andrea Melichárková[1], Lenka Mártonfiová[2] and Stanislav Španiel[1,3]**

[1] Institute of Botany, Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Bratislava, Slovakia, [2] Botanical Garden of P. J. Šafárik University in Košice, Košice, Slovakia, [3] Department of Botany, Faculty of Science, Charles University, Prague, Czechia

OPEN ACCESS

Edited by:
Božo Frajman, University of Innsbruck, Austria

Reviewed by:
Pau Carnicero, University of Innsbruck, Austria
Virginia Valcárcel, Autonomous University of Madrid, Spain
Andreas Tribsch, University of Salzburg, Austria

***Correspondence:**
Stanislav Španiel
stanislav.spaniel@savba.sk

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

Received: *10 November 2020*
Accepted: *22 March 2021*
Published: *27 April 2021*

Citation:
Cetlová V, Zozomová-Lihová J, Melichárková A, Mártonfiová L and Španiel S (2021) Multiple Drivers of High Species Diversity and Endemism Among Alyssum Annuals in the Mediterranean: The Evolutionary Significance of the Aegean Hotspot. Front. Plant Sci. 12:627909. doi: 10.3389/fpls.2021.627909

The Mediterranean Basin is a significant hotspot of species diversity and endemism, with various distribution patterns and speciation mechanisms observed in its flora. High species diversity in the Mediterranean is also manifested in the monophyletic lineage of *Alyssum* annuals (Brassicaceae), but little is known about its origin. These species include both diploids and polyploids that grow mainly in open and disturbed sites across a wide elevational span and show contrasting distribution patterns, ranging from broadly distributed Eurasian species to narrow island endemics. Here, we investigated the evolution of European representatives of this lineage, and aimed to reconstruct their phylogeny, polyploid and genome size evolution using flow cytometric analyses, chloroplast and nuclear high- and low-copy DNA markers. The origin and early diversification of the studied *Alyssum* lineage could be dated back to the Late Miocene/Pliocene and were likely promoted by the onset of the Mediterranean climate, whereas most of the extant species originated during the Pleistocene. The Aegean region represents a significant diversity center, as it hosts 12 out of 16 recognized European species and comprises several (sub)endemics placed in distinct phylogenetic clades. Because several species, including the closest relatives, occur here sympatrically without apparent niche differences, we can reject simple allopatric speciation via vicariance as well as ecological speciation for most cases. Instead, we suggest scenarios of more complex speciation processes that involved repeated range shifts in response to sea-level changes and recurrent land connections and disconnections since the Pliocene. In addition, multiple polyploidization events significantly contributed to species diversity across the entire distribution range. All seven polyploids, representing both widespread species and endemics to the western or eastern Mediterranean, were inferred to be allopolyploids. Finally, the current distribution patterns have likely been affected also by the human factor (farming and grazing). This study illustrates the complexity of evolutionary and speciation processes in the Mediterranean flora.

Keywords: Aegean area, allopolyploidy, *Alyssum*, annual species, endemics, Mediterranean, phylogeny, sympatry

INTRODUCTION

The Mediterranean Basin represents one of the world's major biodiversity hotspots with an exceptionally high species diversity and endemism rate (Myers et al., 2000; Thompson, 2020). This species richness can be attributed to a variety of different factors. Complex geological history, high topographical and ecological heterogeneity of this region have promoted allopatric and ecological diversification and speciation (Médail and Diadema, 2009; Hewitt, 2011; Nieto Feliner, 2014; Thompson, 2020). In addition, the impact of Quaternary climatic oscillations was less severe in the Mediterranean than in central and northern Europe, allowing preservation and accumulation of diversity through small-scale range shifts with minimized population and species extinctions (Nieto Feliner, 2011). Several major paleoclimatic and paleogeologic events have exerted crucial influences on Mediterranean flora, especially the Mediterranean Sea desiccation during the Messinian salinity crisis (ca. 5.96–5.33 mya; Krijgsman et al., 1999), the onset of the Mediterranean climate (ca. 3.2 mya, Suc, 1984), and sea-level oscillations that occurred during the glacial and interglacial periods of the Pleistocene (Lambeck et al., 2014). Sea-level drops resulted in the formation of temporary land connections between major peninsulas (e.g., between the Balkan and Apennine Peninsulas, Correggiari et al., 1996), between islands, and between the mainland and islands (e.g., Anatolia and East Aegean islands, Simaiakis et al., 2017). At the latest Pleistocene, the Aegean and Ionian sea level was ca. 120 m lower than today, and Aegean islands formed a kind of land bridges connecting mainland Greece and Turkey, separated locally by sea channels (Perissoratis and Conispoliatis, 2003), which enhanced plant migrations. Sea-level oscillations also fostered distribution range shifts, which led to secondary contacts and gene flow between otherwise isolated species and lineages (Nieto Feliner, 2011, 2014).

The Mediterranean Basin and the adjacent mountain ranges represent the center of species diversity for *Alyssum* L. (Dudley, 1965; Ball and Dudley, 1993; Jalas et al., 1996), which is the largest genus in the tribe Alysseae and comprises approximately 114 species, including 30 annuals and 84 perennials (Španiel et al., 2015, but still subject of taxonomic revisions). The genus *Alyssum*, in its traditional taxonomic treatment, was nearly twice as large (Al-Shehbaz, 2012); however, many species have recently been transferred into the genera *Cuprella* Salmerón-Sánchez, Mota & Fuertes, *Meniocus* Desv., and *Odontarrhena* C.A.Mey. ex Ledeb. (Španiel et al., 2015). In the current delimitation, the genus *Alyssum* includes three of its six original sections (Dudley, 1964b): *Alyssum* sect. *Alyssum*, *A.* sect. *Gamosepalum* (Hausskn.) T.R.Dudley, and *A.* sect. *Psilonema* (C.A.Mey.) Hook.f. Recent phylogenetic analyses, however, have revealed two major clades in *Alyssum* s.str., which do not agree with the traditional sectional classification (Rešetnik et al., 2013; Li et al., 2015; Salmerón-Sánchez et al., 2018). One clade includes annual species of *A.* sect. *Psilonema* and both annual and perennial species of *A.* sect. *Alyssum*, whereas the other clade includes perennial species from *A.* sect. *Alyssum* and *A.* sect. *Gamosepalum* and an annual species *Alyssum dasycarpum* Stephan ex. Willd. from *A.* sect. *Psilonema.* The present study focuses on the former clade, which comprises all annuals from sections *Alyssum* and *Psilonema* (except for the phylogenetically distant *A. dasycarpum*) and a nested *Alyssum montanum–Alyssum repens* perennial species complex. Whereas the taxonomy and evolutionary history of the *A. montanum–A. repens* species complex have been thoroughly explored in a series of recent studies (e.g., Španiel et al., 2011, 2017a,b, 2019; Magauer et al., 2014; Zozomová-Lihová et al., 2014, 2020; Arrigo et al., 2016; Melichárková et al., 2017), little is known about the phylogenetic relationships and speciation processes of the annual taxa. Here, we attempted to fill this gap by exploring the evolutionary history of all 16 annual taxa of *Alyssum* reported in Europe, some of which extend also to northern Africa or Asia (see **Table 1**). Altogether, the studied clade comprises tentatively 29 annual species, thus about 13 extra-European annuals distributed mainly in the Irano-Anatolian area, Caucasus and Central Asia (Dudley, 1964a,b, 1965; Rechinger, 1968) were not included in our study. Their circumscription and taxonomic treatment is namely uncertain in several cases, and detailed field and herbarium research is needed before including them in phylogenetic studies, which was beyond the scope of the present study.

The here studied *Alyssum* annuals display contrasting distribution patterns in Europe, ranging from narrow endemics to widespread species (**Table 1**). Much species diversity is concentrated in the Aegean islands and adjacent mainland areas (Dudley, 1965; Hartvig, 2002; Strid, 2016). All these species grow preferentially on open, disturbed ground and, therefore, can be found on pastures, gravelly screes, rocky or stony slopes, phrygana, macchia, and typically (but not exclusively) on limestone at various elevations. Some species are more ubiquitous (e.g., *Alyssum alyssoides* and *Alyssum minutum*), whereas others are typically present in particular types of habitats (see **Table 1** for details). Quite commonly three or even more annual *Alyssum* species can grow in the same locality within a close distance of few meters (field observations and the present sampling, see **Table 1** and **Supplementary Table 1**). Small yellow flowers arranged in racemes are visited by a variety of insects, but nothing is known about the mating system of these *Alyssum* species. Phenological shifts observed between some co-occurring species suggest the existence of premating isolating barriers. Seed dispersal occurs primarily through gravity and zoochory, especially sheep and goats, as the seeds become mucilaginous under wet conditions and easily attach to the animal body (field observations). Three different ploidy levels were previously reported for these annuals, ranging from diploids to hexaploids (**Table 2**, Španiel et al., 2015).

Some hypotheses regarding species relationships can be drawn from morphological resemblance and the previously published phylogeny of Alysseae (Rešetnik et al., 2013). For example, available data suggest that the hexaploid species *Alyssum siculum* is either an allopolyploid derived from the tetraploid *A. alyssoides* and the diploid *Alyssum simplex* or an autopolyploid of *A. alyssoides* (Persson, 1971; Rešetnik et al., 2013). Another hexaploid taxon, *A. granatense*, morphologically resembles the tetraploid *A. alyssoides* (Küpfer and Nieto Feliner, 1993). Morphological similarity indicates close evolutionary relationships between the hexaploid *Alyssum hirsutum* and the diploid *Alyssum pogonocarpum* (Carlström, 1984), between

TABLE 1 | Overview of the studied *Alyssum* annual species (listed alphabetically), including their distribution ranges, habitat characteristics and site coexistence with other species.

Species	Distribution	Area codes	Habitat	Site coexistence
A. alyssoides (L.) L.	Europe; N Africa; SW, C Asia	ABCEFGHIJ	Steppes, rocky slopes, quarries, disturbed ground, pastures, roadsides; limestone, but also other substrates (schist, flysch, serpentine), 0–2,300 m a.s.l.	*hirsutum, minutum, simplex, turkestanicum, umbellatum*
A. collinum Brot.	SW Europe; N Africa	FJ	Pastures, steppes, open dry sites, cultivated areas, roadsides, gravel, screes, rocks, sands; calcareous and other substrates, 50–1,850 m a.s.l.	*granatense*
A. foliosum Bory & Chaub.	GR-m, Ae; MK; CY; wTR	ABCE	Less exposed microsites between rocks or spiny shrubs in dry rocky scrubland, screes, pastures; limestone, 400–1,800 m a.s.l.	*fulvescens, minutum, simplex, simulans, smyrnaeum, xiphocarpum*
A. fulvescens Sm.	GR-Ae; wTR; ?CY	AB	Rocky and gravelly ground in open scrubland, pastures and rocky slopes; limestone (in Turkey reportedly also on volcanic substrates and sand dunes), 20–1,300 m a.s.l.	*foliosum, simplex, smyrnaeum, umbellatum*
A. granatense Boiss. & Reut.	SW Europe; N Africa	FJ	Gravelly pastures, open disturbed ground, roadsides; calcareous substrates, schist, granite, 40–2,700 m a.s.l.	*collinum, minutum*
A. hirsutum M.Bieb.	E, SE Europe; TR	BEH	Seashores, sandy, rocky or gravelly places, riverbeds, ruderal sites, cultivated areas; mainly calcareous substrates, 0–1,300 m a.s.l.	*alyssoides, minutum, turkestanicum*
A. minutum Schltdl. ex DC.	S, E Europe (incl. Ae); N Africa; SW Asia	ABCDEF*G*H*IJ	Steppes, open ground, dry ruderal sites, sandy seashores, gravelly pastures; mainly limestone, 0–2,700 m a.s.l.	*alyssoides, foliosum, granatense, hirsutum, siculum, simplex, simulans, smyrnaeum, umbellatum, xiphocarpum*
A. pogonocarpum Carlström	GR-Ae (Rhodos); sTR, ?cTR	AB	Ultramafic (serpentine) gravelly slopes, ca. 250 m a.s.l.	*simplex, strigosum*
A. siculum Jord.	GR-m, Ae (Crete); IT (Sicily)	ACD	Rocky slopes, gravel, open disturbed ground, pastures; mainly limestone, 600–2,200 m a.s.l.	*minutum, simplex*
A. simplex Rudolphi	S, E Europe (incl. Ae); SW, C Asia; CY; ?N Africa	ABCDEFGHI(?J)	Pastures, open ground, rocky and gravelly slopes; calcareous and other substrates (including serpentine), 0–2,000 m a.s.l.	*alyssoides, foliosum, fulvescens, minutum, pogonocarpum, siculum, simulans, smyrnaeum, strigosum*
A. simulans Runemark ex Hartvig	GR-m, Ae (Crete)	AC	Rocky slopes, screes, open disturbed ground, pastures; limestone, 600–1,900 m a.s.l.	*alyssoides, foliosum, minutum, siculum, simplex, smyrnaeum*
A. smyrnaeum C.A.Mey.	GR-m, Ae; TR; Crimea	ABC	Rocky slopes, screes, dry grasslands; calcareous substrates, 150–1,400 m a.s.l.	*foliosum, fulvescens, minutum, simplex, simulans*
A. strigosum Banks & Sol.	S Europe (incl. Ae); SW Asia; CY; ?N Africa	ABCEFGHI(?J)	Steppes, gravel, fallow fields; calcareous and other substrates (including serpentine), 0–1,200 m a.s.l.	*alyssoides, pogonocarpum, simplex, turkestanicum*
A. szovitsianum Fisch. & C.A.Mey.	SW Asia	I	Dry rocky sites, open ground, cultivated areas, 200–2,800 m a.s.l.	
A. turkestanicum Regel & Schmalh. (including *A. desertorum* Stapf)	C, E, SE Europe (incl. Ae); SW, C, E Asia;	ABCEHI	Steppes, rocky and sandy places, dry ruderal sites, pastures; mainly limestone and serpentine, 0–2,000 m a.s.l.	*alyssoides, hirsutum, simplex, strigosum*
A. umbellatum Desv.	S, E Balkan (incl. Ae); CY; Crimea; wTR; wSY	ACEI	Steppes, dry ruderal sites, gravel roads, riverbeds, seashores; a variety of substrates (including serpentine), 0–900 m a.s.l.	*alyssoides, foliosum, fulvescens, minutum*
A. xiphocarpum P. Candargy	GR-Ae (Lesvos)	A	Rocky slopes, dry rocky scrubland; limestone, ca. 730 m a.s.l.	*foliosum, minutum*

Geographic distributions are reported according to Dudley (1965), Küpfer and Nieto Feliner (1993), Rybinskaya (1994), Jalas et al. (1996), Zhou et al. (2001), Hartvig (2002), Al-Shehbaz (2010), Marhold (2011), Strid (2016), Rešetnik and Španiel (2018), and Cetlová et al. (2019). In the Distribution column, country abbreviations follow standardized two-letter codes (GR, Greece; MK, North Macedonia; CY, Cyprus; TR, Turkey; IT, Italy; SY, Syria); abbreviations 'w' and 's' stand for western and southern parts, respectively, and distribution in Greece is given separately for the mainland ('m') and Aegean islands ('Ae'). Species distribution is summarized also in the column Area codes: A, Aegean islands; B, Turkey (mainland); C, Greece (mainland); D, Sicily; E, Balkan; F, southwestern Europe (Iberian Peninsula, France); G, Apennine Peninsula; H, central and eastern Europe; I, Asia (except Turkey); J, northern Africa. Asterisks given for A. minutum indicate some exceptions: absent from central Europe, France and very rare in the Apennine Peninsula. Habitat data are based on Küpfer and Nieto Feliner (1993) and Hartvig (2002) and our own field observations. Data on site coexistence are based on our field observations, exceptionally on literature sources (site coexistence of A. pogonocarpum with A. strigosum in Rhodos according to Hassler, 2004-2021). For the list and details of the analyzed populations, see ***Supplementary Table 1****.*

TABLE 2 | Chromosome numbers, ploidy level data and relative genome size values of the studied *Alyssum* species.

Species	Chrom. number	Ploidy level	2C ± *SD*	Cx ± *SD*
A. alyssoides	2*n* = 32*[a]	4*x*	0.699 ± 0.011	0.175 ± 0.003
A. collinum	2*n* = 32*	4*x*	1.094 ± 0.012	0.273 ± 0.003
A. foliosum	2*n* = 16*	2*x*	0.507 ± 0.010	0.253 ± 0.005
A. fulvescens	2*n* = 16*[b]	2*x*	0.583 ± 0.020	0.292 ± 0.010
A. granatense	2*n* = 48	6*x*	0.966 ± 0.027	0.161 ± 0.004
A. hirsutum	2*n* = 46, 48*[c]	6*x*	1.991 ± 0.045	0.332 ± 0.008
A. minutum	2*n* = 16*	2*x*	0.496 ± 0.006	0.248 ± 0.003
A. pogonocarpum	2*n* = 16	2*x*	0.948 ± 0.011	0.474 ± 0.005
A. siculum	2*n* = 48*	6*x*	1.229 ± 0.018	0.205 ± 0.003
A. simplex	2*n* = 16*[d]	2*x*	0.602 ± 0.008	0.301 ± 0.005
A. simulans	2*n* = 32*	4*x*	1.015 ± 0.032	0.254 ± 0.008
A. smyrnaeum	2*n* = 16*	2*x*	0.508 ± 0.010	0.253 ± 0.005
A. strigosum	2*n* = 16*[e]	2*x*	0.617 ± 0.007	0.308 ± 0.003
A. szovitsianum	2*n* = 16[f]	2*x*	0.505 ± 0.002	0.253 ± 0.001
A. turkestanicum	2*n* = 32*[g]	4*x*	0.713 ± 0.009	0.178 ± 0.002
A. umbellatum	2*n* = 14*, 16	2*x*	0.701 ± 0.015	0.351 ± 0.007
A. xiphocarpum	–	2*x*	0.770 ± 0.010	0.385 ± 0.005

Data on chromosome numbers include both previously published records (the original literature sources are listed in AlyBase, Španiel et al., 2015) and those determined in the present study (marked by asterisks, see ***Supplementary Table 1****). Superscripts a–g indicate some previously published deviating, rare and/or probably erroneous records (see below). Data on ploidy level and relative genome size values are based on the present flow-cytometric (DAPI) measurements, listed as species-level averages of 2C (total) and Cx (monoploid) values given in arbitrary units ± standard deviation (SD). For the details of the analyzed populations and the population-level genome size values, see* ***Supplementary Table 1****.*

[a]Rare and probably erroneous records of 2n = 16 (Crete) and of 2n = 24 (Italy).

[b]A single record of 2n = 32 reported from Turkey, needs verification.

[c]A single record of 2n = 16 from Georgia, needs verification.

[d]A doubtful record of 2n = 24 from Iran; rare records of 2n = 32 from Iran and Tajikistan.

[e]A doubtful record of 2n = 64 from Turkey.

[f]Rare and probably erroneous records of 2n = 14, 32 from Iran.

[g]Rare and doubtful records of 2n = 16, 24, 48 from Iran and Afghanistan.

the diploids *A. simplex* and *Alyssum strigosum,* and between the diploids *Alyssum fulvescens* and *Alyssum smyrnaeum* (Hartvig, 2002). The tetraploid *Alyssum simulans* is a suspected allopolyploid of the diploids *Alyssum foliosum* and *A. minutum,* due to its intermediate morphology (Hartvig, 2002). *Alyssum xiphocarpum* (the ploidy level of which has not yet been published) is morphologically very similar to the sympatric diploid *Alyssum umbellatum* and, therefore, has been treated by many authors as a synonym of the latter species (e.g., Dudley, 1965; Jalas et al., 1996); however, *A. xiphocarpum* exhibits much larger petals, anthers and styles than *A. umbellatum* (Hartvig, 2002). In contrast, the tetraploid *Alyssum turkestanicum* does not show obvious morphological affinity to any particular European species, and its phylogenetic relationships are unknown.

In the present study, we aimed to resolve the evolutionary history of all 16 European *Alyssum* annuals described above. To accomplish this goal, we performed the following: (1) assessed the ploidy levels and genome size variation within and among the species; (2) inferred phylogenetic relationships among these species based on multilocus DNA data, including divergence time estimates; (3) proposed scenarios of speciation events based on distribution, phylogenetic, and polyploidy patterns; and (4) formulated consistent hypotheses for the origin of polyploids.

MATERIALS AND METHODS

Plant Material

Material sampling comprised all 16 annual *Alyssum* species reported in Europe (**Table 1** and **Supplementary Table 1**). In addition, two species from the well-explored perennial *A. montanum-A. repens* species complex, which is nested as a monophyletic clade within the lineage of the studied annuals, and one southwestern Asian annual species, *Alyssum szovitsianum*, were included. On the other hand, *A. dasycarpum*, which is an annual species that also occurs in Europe but is phylogenetically unrelated to the lineage being explored here (see section "Introduction" and Rešetnik et al., 2013) was not investigated. The sampling focused on the European Mediterranean, and particularly Greece and the Aegean area, which represent the regions with the highest species diversity of *Alyssum* annuals within Europe (**Table 1**). Two species of the related genus *Odontarrhena* were used as outgroups in the molecular analyses. In total, 191 populations of annual species were sampled and used for flow-cytometric (FCM) analyses. Of these, 76 populations, representing all sampled 17 annual species, were subjected to molecular analyses. A full list of the population samples, including detailed descriptions of their localities, can be found in **Supplementary Table 1** and is shown on the map in **Figure 1**. Tissue samples were collected from five to ten individuals in each population. Leaves and sterile shoots were dried and preserved in silica gel for further use in FCM and DNA extraction procedures. All sampled individuals were analyzed by FCM, and typically, two individuals per population were subjected to molecular analyses. In several cases, when individuals at the sampled localities were in the fruiting stage, or the leaves were withered or damaged, we collected seeds. These seeds were later sown, and the plantlets that sprouted from the seeds were used to obtain vegetative tissues for FCM and molecular analyses. In addition, seeds were collected from most species (whenever available at localities) for chromosome counting. Voucher specimens of the plants analyzed during the present study have been deposited in the herbarium of the Institute of Botany, SAS, Bratislava, Slovakia (SAV).

Chromosome Counting and Flow Cytometry

Chromosome numbers were determined in the mitotic metaphases of cells from root tips, following the protocol described in Španiel et al. (2018). FCM was used to determine relative genome sizes and to infer the ploidy levels of the investigated annual taxa and populations. Each sampled individual was analyzed separately, measured together with an internal standard (see below). Ploidy levels were inferred by comparing the relative genome size of each sample with the values measured in conspecific plants with known, counted

FIGURE 1 | Map of the sample sites of the analyzed annual *Alyssum* species. Population assignments to the species are indicated by different symbol colors and shapes. Ploidy levels are indicated next to the species names. *Alyssum szovitsianum,* which was represented by one population sample (289TRJ-sz) from Iran, is not shown on the map. Population codes follow **Supplementary Table 1** and are written in either black script (population samples used only for the FCM analyses) or red script (population samples used for both FCM and molecular analyses).

chromosome numbers. In the case of four taxa for which we were unable to obtain chromosome numbers (*A. granatense*, *A. pogonocarpum*, *A. szovitsianum,* and *A. xiphocarpum*), their ploidy levels were deduced from the previously published chromosome counts for these species and/or through comparisons with the relative genome sizes of their closest

relatives, for which the chromosome numbers were determined in this study. The relative genome size was measured based on the fluorescence intensity of nuclei stained with the AT-selective fluorochrome 4′,6-diamidino-2-phenylindole (DAPI). The relative genome size (relative 2C value) of a sample was computed as the ratio between the mean G0/G1 peak value of the sample and the mean G0/G1 peak value of the standard with a known genome size. A monoploid relative genome size (relative C*x* value) was also inferred, representing the relative DNA content of a monoploid, unreplicated genome that comprises one set of chromosomes and corresponds to the base chromosome number (*x*) of the taxon (Greilhuber et al., 2005). One of three internal standards was selected for each individual measurement, depending on the genome size of the measured species. *Lycopersicon esculentum* 'Stupické polní rané' (2C = 1.96 pg; Doležel et al., 1992) was used as the primary standard for most taxa. In cases when *Lycopersicon* was not suitable (due to overlapping G0/G1 or G2 peaks between the sample and the standard), we used secondary standards, either *Bellis perennis* L. (2C = 3.38 pg; Schönswetter et al., 2007) or *Solanum pseudocapsicum* L. (2C = 2.59 pg; Temsch et al., 2010). The relative genome sizes of all samples that were measured using secondary standards were eventually recalculated relative to the primary standard to obtain a final value for the relative genome size (in arbitrary units), as reported in Section "Results." The recalculation was performed based on the ratio between the relative genome sizes of the primary and secondary standards. This ratio was calculated based on repeated simultaneous measurements of the primary standard with each secondary standard. In addition to individual sample measurements, selected samples from two or three different species that showed divergent genome size values at the same ploidy level (see section "Results") were analyzed simultaneously. These measurements were performed to prove and illustrate the differences in genome size between species. All analyses were performed using a Partec Cyflow ML instrument equipped with an HBO-100 mercury arc lamp (Partec, Münster, Germany) following the protocol described in Španiel et al. (2011). The relative genome sizes of the perennials *A. repens* and *A. montanum* were known and taken from previous studies (Španiel et al., 2018; Melichárková et al., 2019), used here in the analysis exploring phylogenetic patterns of relative genome size variation (see below).

PCR Amplification, Cloning, and Sequencing

Genomic DNA was isolated using the DNeasy 96 Plant Kit, according to the manufacturer's protocol (Qiagen, Hilden, Germany). One chloroplast and two nuclear DNA regions were amplified and sequenced: the *rpoB-trnC*^{GCA} intergenic spacer of chloroplast DNA (cpDNA), the multicopy internal transcribed spacer (ITS) region of rDNA (ITS1-5.8S-ITS2), and the single-copy *DET1* (de-etiolated 1) nuclear gene. Universal primers were used to amplify the *rpoB-trnC* spacer (primers rpoB and trnC^{GCA}R; Shaw et al., 2005) and the ITS region (primers ITS4 and ITS5; White et al., 1990). The partial sequence of the *DET1* gene was amplified with the primer set published by Kuittinen et al. (2002), which binds to exons 5 and 8 in *Arabidopsis thaliana* and has previously been shown to perform well also in *Alyssum* (Melichárková et al., 2017). Polymerase chain reaction (PCR) amplifications were performed in a total volume of 20 μl containing 1 μl of gDNA, 0.58 U *Pfu* DNA polymerase (Thermo Fisher Scientific Inc., Waltham, MA, United States), 1× *Pfu* buffer with $MgSO_4$, 0.2 mM dNTPs, and 0.2 μM each forward and reverse primers. The *DET1* reaction mix also included 5% dimethyl sulfoxide (DMSO) to reduce possible recombination events during PCR. The following PCR conditions were used: 80°C for 5 min, 30 × (95°C for 1 min, 50°C for 1 min, followed by a 5% ramp to 65°C, 65°C for 4 min), 65°C for 5 min for the *rpoB-trnC* amplification; 94°C for 3 min, 35 × (94°C for 30 s, 50°C for 30 s, 72°C for 1 min), 72°C for 10 min for the ITS amplification; 94°C for 5 min, 30 × (94°C for 30 s, 51°C for 30 s, 72°C for 2 min), 72°C for 7 min for the *DET1* amplification.

PCR products were purified either enzymatically (prior to direct sequencing) using a mixture of Exonuclease I and FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific Inc.) or with NucleoSpin Gel and PCR Clean-up column kit (Macherey-Nagel, Düren, Germany), when subjected to cloning before sequencing. ITS products from all individuals and *DET1* products from all diploids were first sequenced directly, and if sequence polymorphisms (double peaks or unreadable parts) were detected in the electropherograms, the PCR products were cloned. *DET1* products from the polyploid individuals were all cloned prior to sequencing. Purified PCR products were cloned with the CloneJET PCR Cloning kit using the pJET1.2/blunt vector, following the manufacturer's protocol (Thermo Fisher Scientific), with transformation into JM109 competent cells (Promega, Madison, United States). Direct PCR from bacterial colonies was performed in a total volume of 13 μl, using 0.2 μM pJET1.2 sequencing primers, 0.65 U of DreamTaq DNA polymerase (Thermo Fisher Scientific), 1× DreamTaq Buffer with $MgCl_2$, and 0.2 mM dNTPs. The cells were lysed by 95°C for 3 min, followed by 35 cycles (95°C for 30 s, 50°C for 30 s and 72°C for 2 min), and a 10-min final incubation step at 72°C. The PCR products were purified by enzymatic clean-up (as specified above) prior to sequencing. Multiple clones per sample were sequenced in an effort to recover all alleles and identify any potential PCR-mediated recombinations: six to eight clones per diploid, 16 clones per tetraploid, and 20 clones per hexaploid sample. Sequencing was performed using the ABI 3730xl DNA analyzer at Eurofins Genomics Company (Konstanz, Germany). The sequences were edited and aligned using Geneious software R7.1.9 (Biomatters Ltd., Auckland, New Zealand) and submitted to the GenBank nucleotide database (**Supplementary Table 1**). Indels present in the sequence alignments were coded as binary characters (except for those occurring in highly variable homopolymeric stretches that were ignored) according to the simple indel coding approach (Simmons and Ochoterena, 2000) using FastGap 1.2 software (Borchsenius, 2009). The indel coding datasets were appended to the nucleotide datasets and were included in the phylogenetic analyses. The final ITS and *rpoB-trnC*

alignments contained sequence data from 152 individuals (76 populations); the *DET1* alignment comprised data from 73 individuals (from a selection of 38 populations; **Table 3** and **Supplementary Table 1**).

Phylogenetic Analyses of the *rpoB-trnC*, ITS, and *DET1* Sequences

The statistical parsimony-based TCS method (Clement et al., 2000) in PopART (Leigh and Bryant, 2015) was used for the identification of different alleles and haplotypes within the datasets, and to determine their frequencies and sharing patterns. A NeighborNet (NN) network (Huson and Bryant, 2006) was generated in SplitsTree4 v. 4.14.4 based on uncorrected P-distances. Phylogenetic trees were inferred using maximum-likelihood (ML, GARLI v.2.01; Zwickl, 2006) and Bayesian analyses (MrBayes v.3.2.6, Huelsenbeck and Ronquist, 2001), both of which were run at the CIPRES Science Gateway (Miller et al., 2010). The three studied DNA markers were analyzed separately, and three data partitions were defined within each dataset as follows: intergenic spacer of *rpoB-trnC*, a fragment of the *rpoB* gene, and indels within the *rpoB-trnC* dataset; non-coding ITS1 and ITS2, 5.8S of rDNA, and indels within the ITS dataset; introns, exons, and indels within the *DET1* dataset. Best-fit models of nucleotide substitutions were assessed in jModelTest v.2.1.10 (Darriba et al., 2012) separately for each nucleotide data partition, using the Akaike information criterion (AIC; Akaike, 1974). Bayesian analyses were conducted with four MCMC (Markov chain Monte Carlo algorithm) chains for 10–15 million generations, with a sampling frequency of every 100th generation. The first 10% of trees were discarded as burn-in and a consensus tree was generated from the remaining trees, computing also the Bayesian posterior probabilities (BPP) for each node. For ML analyses, multiple searches were performed with random starting trees and by setting the program to stop after 20,000 generations if no improvement of the log-likelihood was detected ($\leq$0.01), with a maximum of 215 million generations. Branch support was assessed by 2,000 bootstrap replicates (bootstrap support, BS).

Inference of Coalescent-Based Species Trees

A Bayesian coalescent-based approach was employed to estimate a species tree using STACEY package v. 1.2.2 (Jones, 2017) for BEAST v. 2.5.1 (Drummond and Rambaut, 2007). Since the multispecies coalescent accounts for incomplete lineage sorting as the primary source of gene tree discordance but not for hybridization events, we assembled two datasets avoiding reticulate evolutionary patterns in the markers and samples (especially in the case of *DET1* data for polyploids, see section "Results"). The first dataset included the *rpoB-trnC* and ITS sequence data, obtained from both diploids and polyploids, whereas the second dataset included all three regions, *rpoB-trnC*, ITS, and *DET1*, from diploids only. The purpose of performing these analyses was to infer species relationships at the diploid level, as well as, to a limited extent, with polyploids involved and confront them with the multilabelled *DET1* gene tree reconstruction. Two populations of diploid *A. fulvescens* sampled from different islands (Samos and Chios islands) were *a priori* defined as two 'species', because they were resolved in two divergent clades in each of the gene trees. The tetraploid *A. simulans* was excluded from the analyses because it displayed two divergent ITS copy variants, indicating an allopolyploid origin. All other species (both diploids and polyploids) appeared genetically coherent, without any indication of reticulation or polyphyletic origins when considering only the ITS and *rpoB-trnC* gene trees.

The tool BEAUti 2 (Bouckaert et al., 2014) was used to create a STACEY input file using the following settings and parameters: multiple unlinked data partitions (as defined above), the best-fit nucleotide substitution models, as determined with jModelTest (see above), lognormal relaxed clock, Yule species tree model, and lognormal priors. We ran three independent Markov chain Monte Carlo (MCMC) analyses, with 75 million generations each

TABLE 3 | Characteristics of the nucleotide alignments used in the present study.

Locus – partition	Alignment length	Number of					Evolutionary model
		Individuals	Sequences	Alleles[b]	PI sites[c]	Indels	
rpoB-trnC	1,013 bp	152	152	61	94 (9.28%)		
- *rpoB* gene	−12 bp				2 (16.67%)		JC
- spacer	−1,001 bp				92 (9.19%)	10	TIM1 + G
ITS rDNA	676 bp	152	265	131	136 (20.12%)		
- 5.8S[a]	−190 bp				3 (1.58%)		TrNef + I
- ITS1 + ITS2	−486 bp				133 (27.37%)	3	TIM2 + I+G
DET1	603 bp	73	154	75	163 (27.03%)		
- exons	−293 bp				50 (17.06%)		TIM2 + G
- introns	−309 bp				113 (36.57%)	3	GTR + G

[a] *Including short fragments of the 18S and 28S genes flanking the ITS1 and ITS2 regions.*
[b] *The total number of different haplotypes (ribotypes, alleles) identified in the alignment.*
[c] *The number of parsimony informative (PI) sites is given in absolute numbers and as percentages of the aligned nucleotide positions.*
Outgroup sequences are not included in this summary data.

and a sampling frequency of 10,000 trees, using BEAST v. 2.5.1 through the CIPRES Science Gateway. The BEAST output was analyzed in Tracer v.1.7.1 (Rambaut et al., 2018) and verified for MCMC convergence and the ESS (effective sample sizes) values of parameters. Multiple runs were combined using LogCombiner v.2.5.2, after discarding the first 20% of the sampled trees as burn-in. A maximum clade credibility (MCC) tree was obtained in TreeAnnotator v.2.5.2 and visualized in FigTree v.1.4.4.

To explore the presence of a phylogenetic signal in the measured genome size values of the studied species (described above in section "Chromosome Counting and Flow Cytometry"), we calculated the tree transformation statistics, λ (Pagel, 1999), which estimates the statistical (in)dependence of trait (genome size) evolution and the reconstructed phylogeny. It ranges from 0 to 1, with values close to 0 indicating phylogenetic independence and values close to 1 indicating a phylogenetic signal in the examined trait. It was computed in BayesTraits v.3.0.2 (Pagel et al., 2004) on the combined post-burn-in set of 18,000 trees generated from the three BEAST runs used for MCC tree as described above (the dataset including both diploids and polyploids), with two independent MCMC runs each of 1 million generations and 100,000 burn-in. Since the two runs gave virtually the same results, the posterior distribution of λ values resulting from one MCMC run was depicted in a histogram.

Phylogenetic inference based on the concatenation of the three sequence alignments was discarded in the present study, as this approach can lead to biased topologies with false support, especially in recently diverged lineages (Kubatko and Degnan, 2007). The concatenation was also technically unfeasible, as it is generally not possible to know *a priori* which allele (*DET1*) should be concatenated with which ribotype (ITS) when multiple sequence variants are revealed even in diploid accessions (see Results).

Estimation of Divergence Times Based on ITS Sequence Data

To estimate the divergence times of the studied species, we assembled an additional ITS alignment with a broader, tribe-wide taxonomic coverage. This alignment followed the dataset compiled by Salmerón-Sánchez et al. (2018), which comprised representatives of almost all genera of the tribe Alysseae (Rešetnik et al., 2013; Španiel et al., 2015), complemented by the ITS sequences of annual *Alyssum* species generated in the present study and the ITS dataset of the perennial *A. montanum–A. repens* species complex (generated in Melichárková et al., 2019). For the tree calibration, we applied an approach described by Huang et al. (2020), who reconstructed a dated phylogeny of the Rosidae clade based on complete plastid sequences and multiple reliable non-Brassicales fossil records and, subsequently, inferred three secondary calibration points within Brassicaceae. These secondary calibrations were used as temporal anchor points for the molecular dating analyses of Brassicaceae tribes based on ITS sequences, representing major lineage splits within the family. Following this approach, the ITS sequences of two outgroup species (*A. thaliana*, tribe Camelineae, lineage I; and *Clausia aprica*, tribe Dontostemoneae, lineage III) were included in the assembled Alysseae dataset to define the ages of the respective nodes (node B representing the split between lineage I and Alysseae; node C representing the split between lineage III and Alysseae, following Huang et al., 2020). In total, the ITS alignment used in the present study comprised 702 sequences from 123 species and 25 genera (**Supplementary Table 2**). The tree topology and divergence times were estimated using the Bayesian MCMC approach, as implemented in BEAST v. 2.6.1 and run through the CIPRES Science Gateway (Miller et al., 2010). The input file was generated using the tool BEAUti 2. For the tree priors, we selected the general time-reversible (GTR + G) substitution model, the uncorrelated relaxed lognormal clock model, the birth-dead speciation model, and we set the ages of the two nodes with a normal distribution prior, as described by Huang et al. (2020). MCMC was run with 100 million generations and a sampling frequency of 10,000 trees. The BEAST output was analyzed in Tracer v.1.7.1 and checked for MCMC convergence and the ESS values of parameters. An MCC tree was obtained in TreeAnnotator v.2.6.0, discarding the first 25% of the sampled trees as burn-in, and visualized in FigTree v.1.4.4.

RESULTS

Chromosome Numbers

Chromosome numbers were determined in 15 populations of 13 annual species (**Supplementary Table 1**, **Table 2**, and **Figure 2**). For the remaining four species (*A. granatense*, *A. pogonocarpum*, *A. szovitsianum,* and *A. xiphocarpum*), the seeds obtained from the studied populations did not germinate (were unripe or too old) and, therefore, were unusable for chromosome counting. Except for *A. xiphocarpum*, the chromosome numbers of these species were already known from the previously published records (see **Table 2**). Four different chromosome numbers were detected, $2n$ = 14, 16, 32, and 48, with the base chromosome number x = 8 or, exceptionally, x = 7 in *A. umbellatum*. They represented the diploid, tetraploid and hexaploid levels.

Ploidy Level and Genome Size Variation

The relative genome sizes were measured for all sampled 17 annual species (**Table 2** and **Supplementary Table 1**). Both the chromosome number data and genome size values demonstrated uniform ploidy levels within individual species. The chromosome number and ploidy level of *A. xiphocarpum* were previously unknown; therefore, the diploid level inferred here represents the first ploidy record for this species. We also found that the studied annual species showed significant differences in their monoploid genome sizes, exhibiting approximately 2.9-fold variation (**Table 2**). Differences in genome sizes between several species (and groups within *A. fulvescens*) within the same ploidy level were large enough to yield separate peaks in the FCM histograms during simultaneous analyses, as shown in **Figure 3**. This variation among species also demonstrated some phylogenetic patterns (see below section "Coalescent-Based Species Trees" for details). The largest relative monoploid genome sizes were detected in

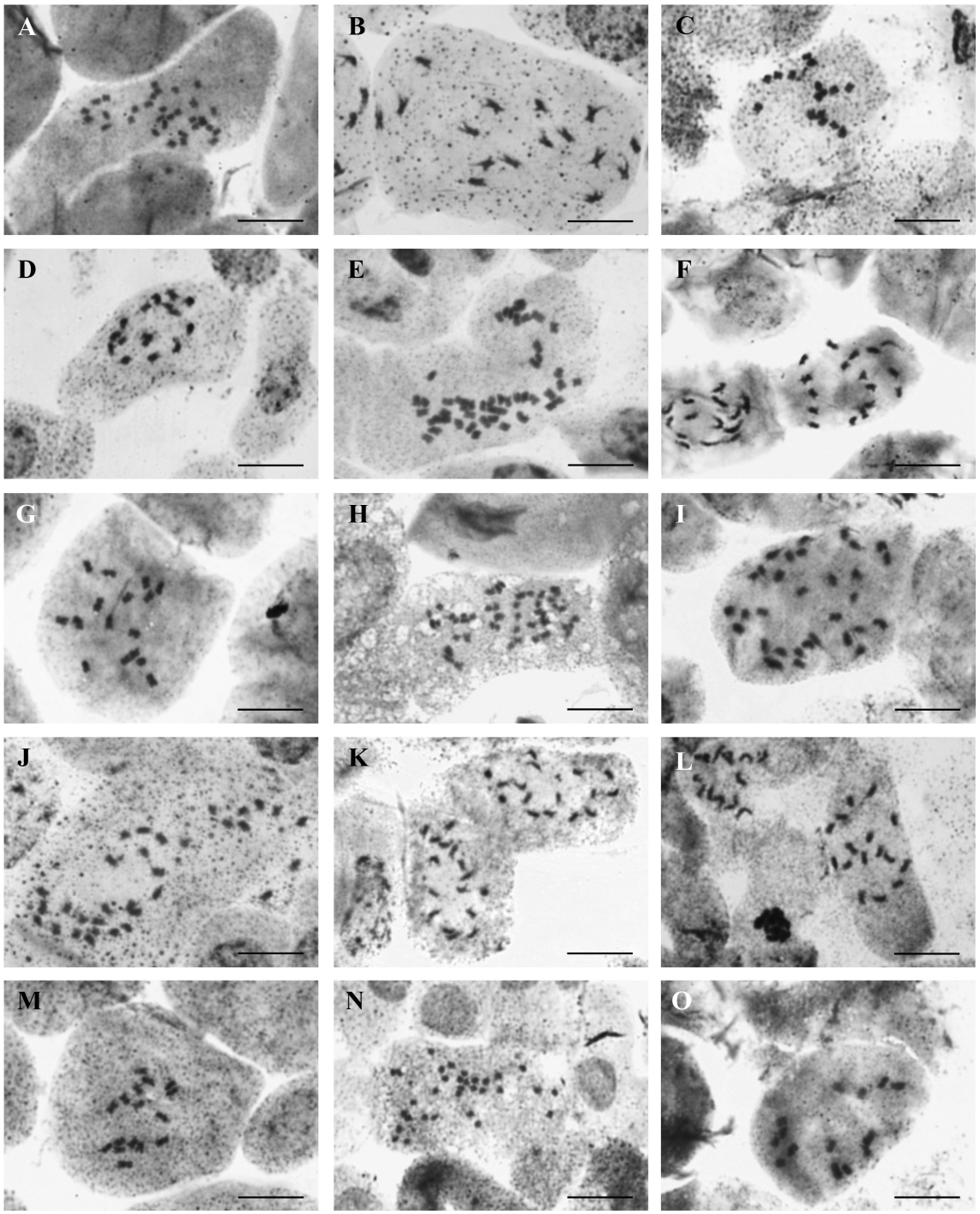


FIGURE 2 | Photographs of chromosome metaphase plates from **(A)** *Alyssum alyssoides* (427AMB-*al*), 2*n* = 32; **(B)** *Alyssum foliosum* (151KAT-*fo*), 2*n* = 16; **(C)** *Alyssum fulvescens* (594PEO-*fu*), 2*n* = 16; **(D)** *A. fulvescens* (595KEK-*fu*), 2*n* = 16; **(E)** *Alyssum hirsutum* (506DRK-*hi*), 2*n* = 48; **(F)** *Alyssum minutum* (319LOM-*mi*), 2*n* = 16; **(G)** *Alyssum simplex* (551RBS-*sx*), 2*n* = 16; **(H)** *Alyssum siculum* (549QUA-*sc*), 2*n* = 48; **(I)** *Alyssum collinum* (261UXA-*cl*), 2*n* = 32; **(J)** *Alyssum simulans* (481PTK-*ss*), 2*n* = 32; **(K)** *Alyssum smyrnaeum* (544DTI-*sy*), 2*n* = 16; **(L)** *Alyssum strigosum* (457SVU-*st*), 2*n* = 16; **(M)** *A. strigosum* (477VAD-*st*), 2*n* = 16; **(N)** *Alyssum turkestanicum* (429ORT-*t*), 2*n* = 32; and **(O)** *Alyssum umbellatum* (458PSS-*um*), 2*n* = 14. Scale bar = 10 μm.

A. hirsutum, *A. umbellatum*, *A. xiphocarpum*, and, especially, *A. pogonocarpum* (the chromosomes, unfortunately, could not be counted in this species; however, previous records reported the diploid level for this species). In contrast, the smallest relative C*x* values were revealed in *A. alyssoides*, *A. siculum*, *A. turkestanicum*, and *A. granatense* (**Table 2**).

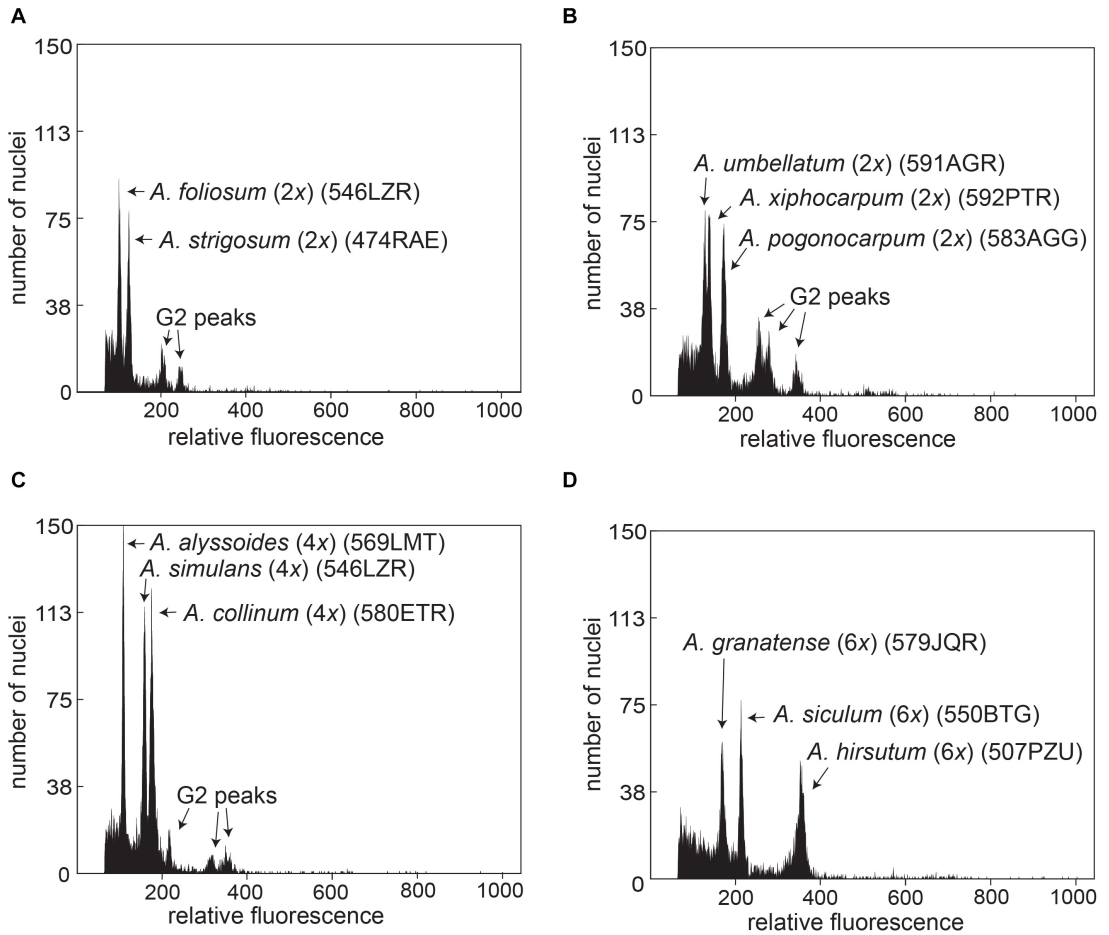


FIGURE 3 | Simultaneous FCM analyses of *Alyssum* species, illustrating the differences in relative genome size values (relative fluorescence) between selected diploid **(A,B)**, tetraploid **(C)**, and hexaploid **(D)** species. For population codes, see **Supplementary Table 1**.

DNA Sequence Data Variation

The detailed characteristics of the *rpoB-trnC*, ITS, and *DET1* alignments, such as the alignment length, number of sequences, alleles, polymorphic sites, and indels, are summarized in **Table 3**.

The *rpoB-trnC* dataset comprised 152 sequences (individuals) with 61 different haplotypes. Haplotype sharing most commonly occurred within populations, as well as among different populations of the same species. In two cases, haplotype sharing was observed between different species: between the tetraploid *A. alyssoides* and the hexaploid *A. siculum*, and between the diploid *A. foliosum* and the tetraploid *A. simulans*. The analyzed species displayed one to six different haplotypes, and the most diverse species were the diploid *A. simplex* and the hexaploids *A. granatense* and *A. siculum*.

Molecular cloning revealed multiple ITS copy variants (ribotypes) within several accessions, in both diploids and polyploids (**Supplementary Table 1**). Up to eight different ribotypes were, for instance, found in one individual of the hexaploid *A. granatense*; and up to five ribotypes in one individual of the diploid *A. fulvescens*. Altogether, the final ITS alignment comprised 265 sequences retrieved from 152 individuals, which represented 131 different ribotypes. The ribotype sharing patterns were similar to those observed in the *rpoB-trnC* dataset. The same ribotypes were frequently revealed within populations and among different populations of the same species. Ribotype sharing between different species was observed only in two cases: between the tetraploid *A. alyssoides* and the hexaploid *A. siculum*, and between the diploid *A. minutum* and the tetraploid *A. simulans*. The analyzed species displayed one to 15 different ribotypes; the most diverse species were the diploid *A. fulvescens* (12 ribotypes) and the polyploids *A. granatense* (15 ribotypes), and *A. alyssoides* (11 ribotypes).

The *DET1* alignment included four exons and introns among them (although the first intron was missing in several accessions, which was also observed for some perennials by Melichárková et al., 2017). All diploid accessions, except for *A. fulvescens*, appeared homozygous. Multiple alleles were revealed by cloning in the polyploids, with up to six alleles in the hexaploids (see **Supplementary Table 1**). Altogether, the alignment comprised 154 sequences from 73 individuals, which represented 75 different alleles. No alleles were shared between different diploid species, but allele sharing occurred

in several cases between diploids and polyploids (reflecting progenitor-polyploid derivative relationships, see below), and between different polyploid species [those with presumably shared progenitor(s)].

Phylogenetic Analyses of the *rpoB-trnC*, ITS, and *DET1* Sequences

The NN networks, the ML, and Bayesian phylogenetic trees exhibited largely congruent topologies for each of the three markers used. Here we present the ML trees (**Figures 4–6**) with both the bootstrap support (BS) and the Bayesian posterior probability (BPP) values plotted on them. Only the clades that were supported by BS (≥50%) and BPP (≥0.85) are described and interpreted. The NN networks are shown in the **Supplementary Figures 1–3**.

The ML tree based on the chloroplast *rpoB-trnC* data (**Figure 4**) showed a relatively well-resolved structure, in which most of the species harbored monophyletic sequences (but see *A. fulvescens*), and several clades could be highlighted. One clade (denoted as clade 1) was composed of four Aegean (sub-)endemics, the diploids *A. foliosum*, *A. fulvescens* from Chios island (594PEO population) and *A. smyrnaeum*, and the tetraploid *A. simulans*. In contrast, *A. fulvescens* from Samos island (595KEK) was placed in another clade together with the more widespread *A. minutum* (clade 2). *A. xiphocarpum* endemic to Lesvos formed a clade with the more widespread Balkan to western Asian *A. umbellatum* (clade 3a). As for the polyploids, the widespread hexaploid *A. hirsutum* was sister to the Aegean diploid subendemic *A. pogonocarpum* (clade 3b). The other four polyploids, *A. turkestanicum, A. granatense, A. alyssoides*, and *A. siculum* (clades 4–6) were clearly differentiated from all here analyzed diploids. While the first two polyploids formed two distinct clades (clades 4 and 5), the last two polyploids appeared completely intermingled (clade 6).

The ML tree based on the ITS data displayed poorly supported backbone relationships (**Figure 5**), therefore, here we focus only on some well supported patterns. Multiple ITS copy variants obtained from several individuals by cloning were phylogenetically close even in polyploids, placed within the same clades (with a single exception of tetraploid *A. simulans*, see below). Two Aegean diploid subendemics, *A. fulvescens* from Chios island and *A. smyrnaeum*, formed a well-supported clade in a sister position to another Aegean subendemic, *A. pogonocarpum* (clade 1). *A. fulvescens* from Samos island, in congruence with cpDNA data, was placed in a different clade together with *A. minutum* (clade 2). *A. xiphocarpum* endemic to Lesvos together with *A. umbellatum* were resolved in a clade with the more widespread *A. simplex* (clade 3). As for the polyploids, clearly polyphyletic ITS sequences were observed only in tetraploid *A. simulans*, placed intermingled either in the clade with *A. minutum* (clade 2) or in the clade with *A. foliosum* (clade 4), the latter position being in congruence with the *rpoB-trnC* data. The tetraploid *Alyssum collinum*, so far reported from southwestern Europe and northern Africa, was placed with much support within the clade of Asian *A. szovitsianum* (clade 5). Four polyploids, *A. turkestanicum, A. granatense,*

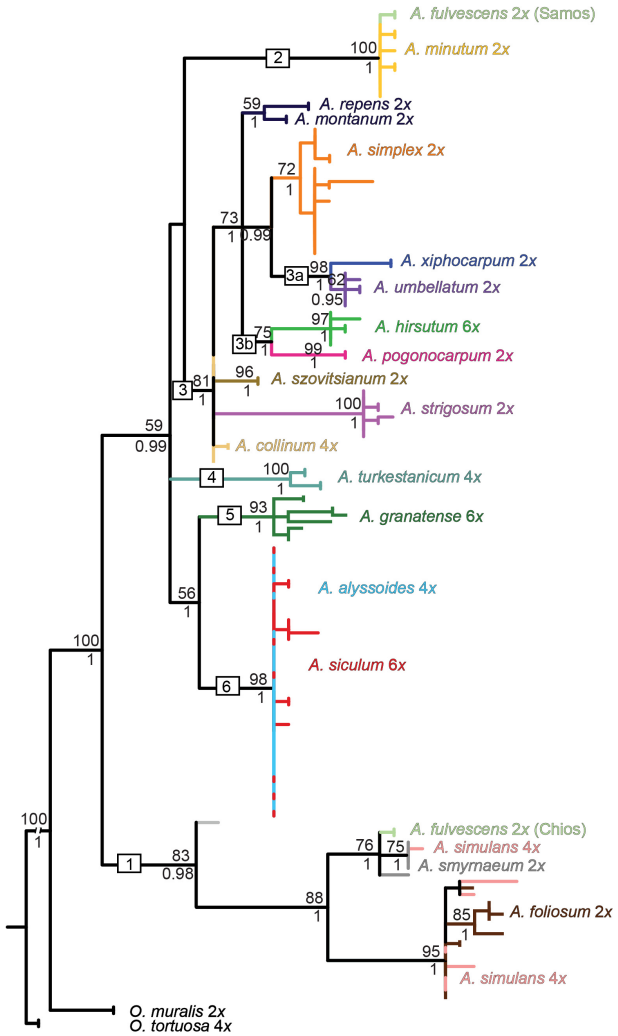


FIGURE 4 | Maximum-likelihood tree based on the *rpoB-trnC* sequences of cpDNA for the studied *Alyssum* species. The values above the branches are bootstrap support (BS) ≥ 50%; those below the branches are Bayesian posterior probability (BPP) values ≥ 0.75 taken from the Bayesian majority-rule consensus tree (not shown). BS and BPP are indicated only for major and species-specific clades, whereas they are skipped for the most terminal clades. Branches and clades are colored according to the species assignment, omitting the terminal labels of individual sequences for the sake of readability (see **Supplementary Figure 1A** for the fully labeled version of the tree). Each species name is followed by its ploidy level. The geographic origins of the two populations of diploid *A. fulvescens* that exhibited genetic divergence in each of the three DNA regions are indicated (Samos and Chios islands). The numbers 1-6 in boxes denote the clades, as described in the text.

A. alyssoides, and *A. siculum*, similarly as in the *rpoB-trnC* data, appeared differentiated from all here analyzed diploids, but only *A. turkestanicum* formed a separate well-supported clade. The NN network (**Supplementary Figure 2B**) displayed a structure that supported the topology of the ML tree, but, in addition, indicated conflicting, reticulated patterns for two polyploids: the tetraploid *A. simulans* against the clades 2 and 4; and the hexaploid *A. hirsutum* against the clades 1 and 3.

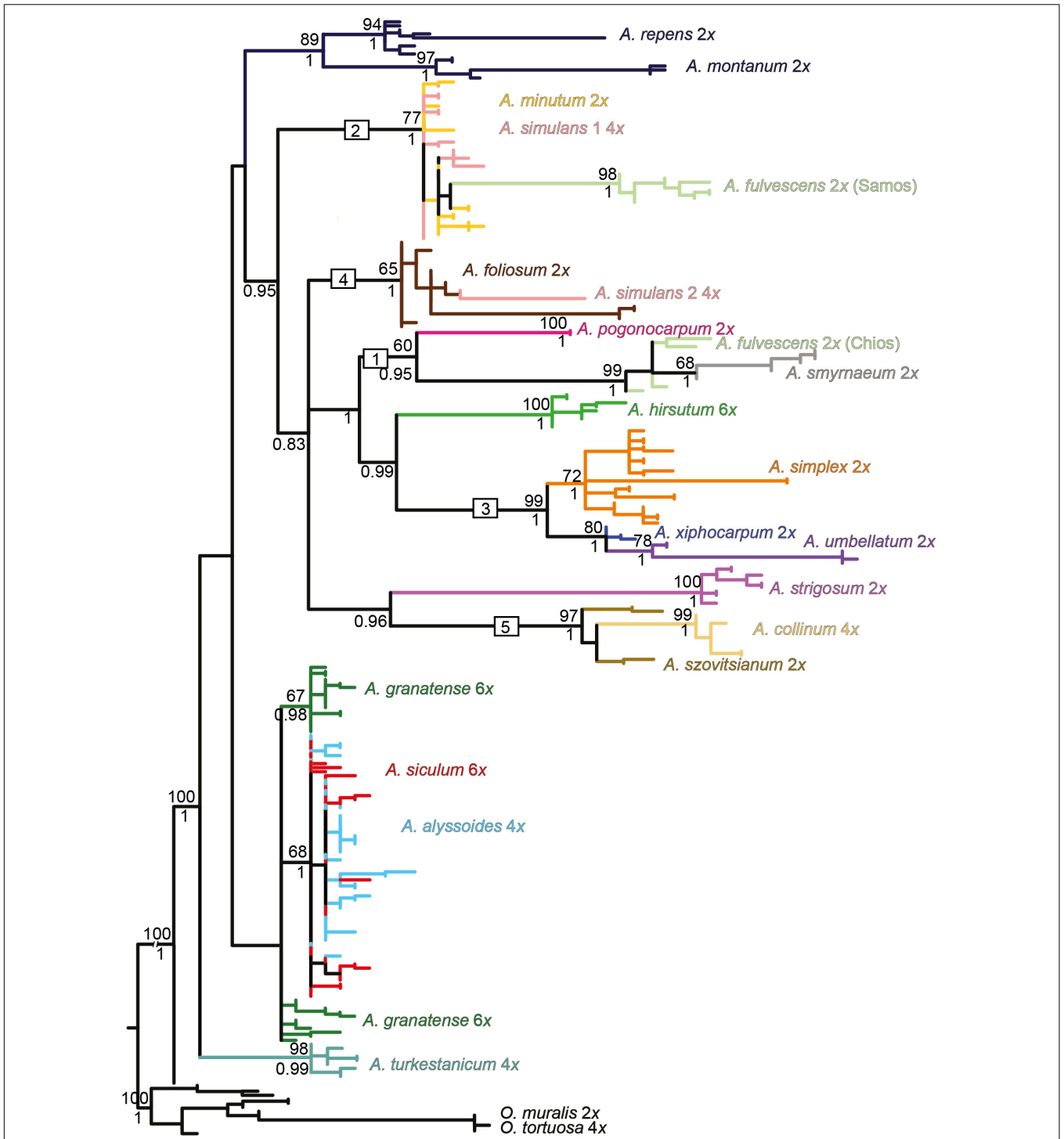


FIGURE 5 | Maximum-likelihood tree based on the ITS sequences of nrDNA for the studied *Alyssum* species. For the tree description, see the legend of **Figure 4**. Divergent ITS sequences were observed in the tetraploid *A. simulans*, which are placed in distinct clades (nr. 2 and 4), marked here as *A. simulans* 1 and *A. simulans* 2. The fully labeled version of the tree is shown in **Supplementary Figure 2A**.

The ML tree based on the *DET1* data (**Figure 6**) showed low resolution at the backbone (in congruence with the NN network structure, see **Supplementary Figure 3B**), but more terminal clades and the relationships among them generally received high support. The diploid species were placed in five distinct clades (clades 1–5), in agreement with the splits resolved in the NN network. In congruence with the ITS data, two Aegean subendemics, *A. fulvescens* from Chios island and *A. smyrnaeum*,

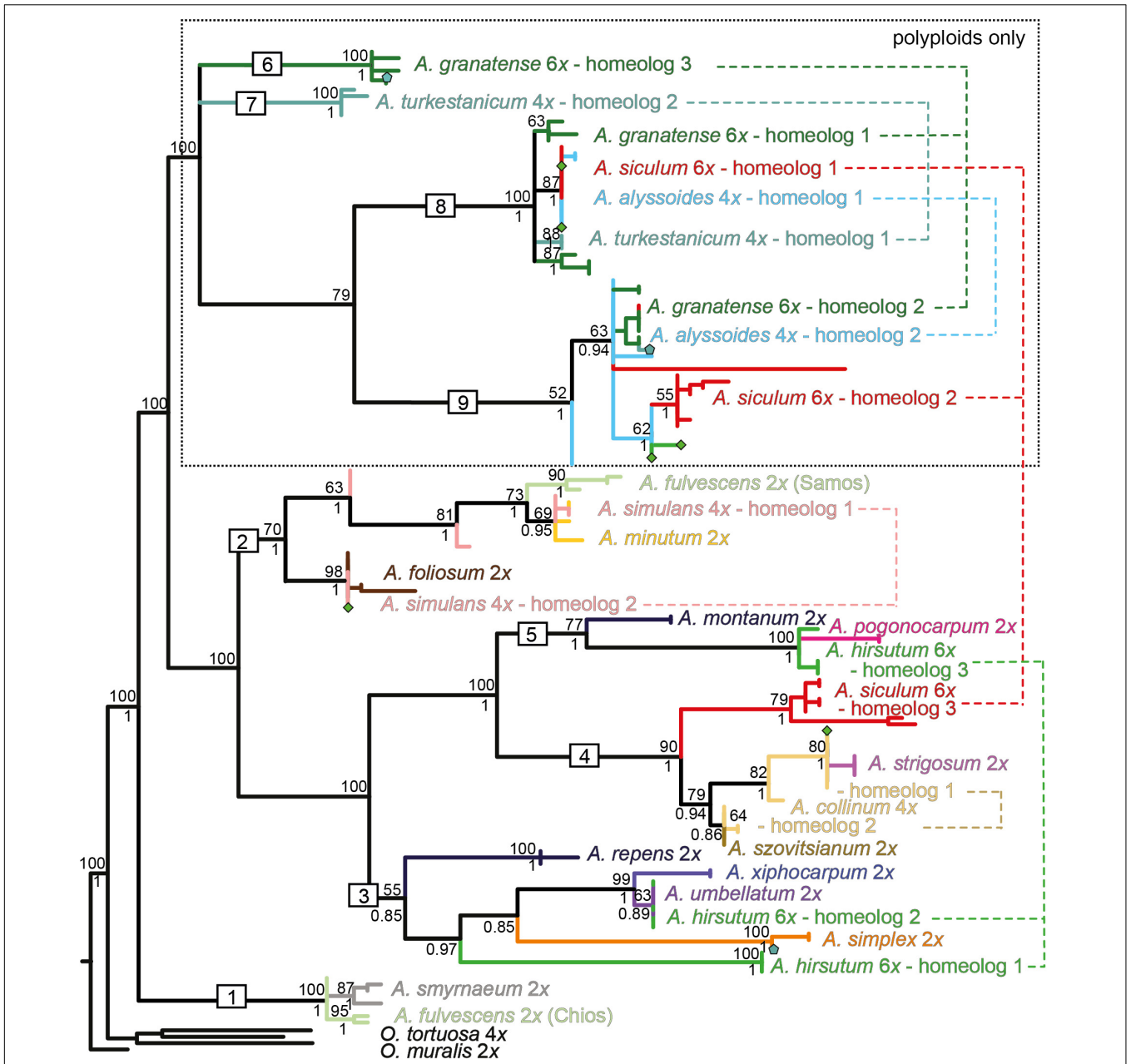


FIGURE 6 | Maximum-likelihood tree based on the *DET1* sequence data for the studied *Alyssum* species. For the tree description, see the legend of **Figure 4**. The different homeologs identified in the polyploid species are numbered and connected by dashed lines. Symbols (pentagons and diamonds) show the alleles that deviated from the observed homeolog variation in *A. turkestanicum* and *A. hirsutum*, respectively. The fully labeled version of the tree is shown in **Supplementary Figure 3A**.

formed one clade (clade 1), whereas *A. fulvescens* from Samos island was sister to *A. minutum* (within clade 2). *A. xiphocarpum* endemic to Lesvos together with *A. umbellatum* were resolved in a clade with the more widespread *A. simplex* (clade 3). Finally, the remaining three annual diploids, *A. pogonocarpum, A. strigosum*, and *A. szovitsianum*, were placed within the clades 4 and 5. The sequences obtained from the polyploids were scattered across the tree, placed within four of the five above-mentioned clades of the diploids, plus in four additional clades (clades 6–9) comprising only polyploids. Within each polyploid, we revealed divergent alleles, which were placed in separate clades of the trees, indicating the presence of two or three different homeologs, i.e., gene copies derived from the parental species (**Figure 6**). In four polyploids, *A. alyssoides, A. granatense, A. siculum,* and *A. turkestanicum*, we observed two to three divergent homeologs, which were placed in the clades comprising exclusively polyploids, i.e., not matching any of the analyzed diploids.

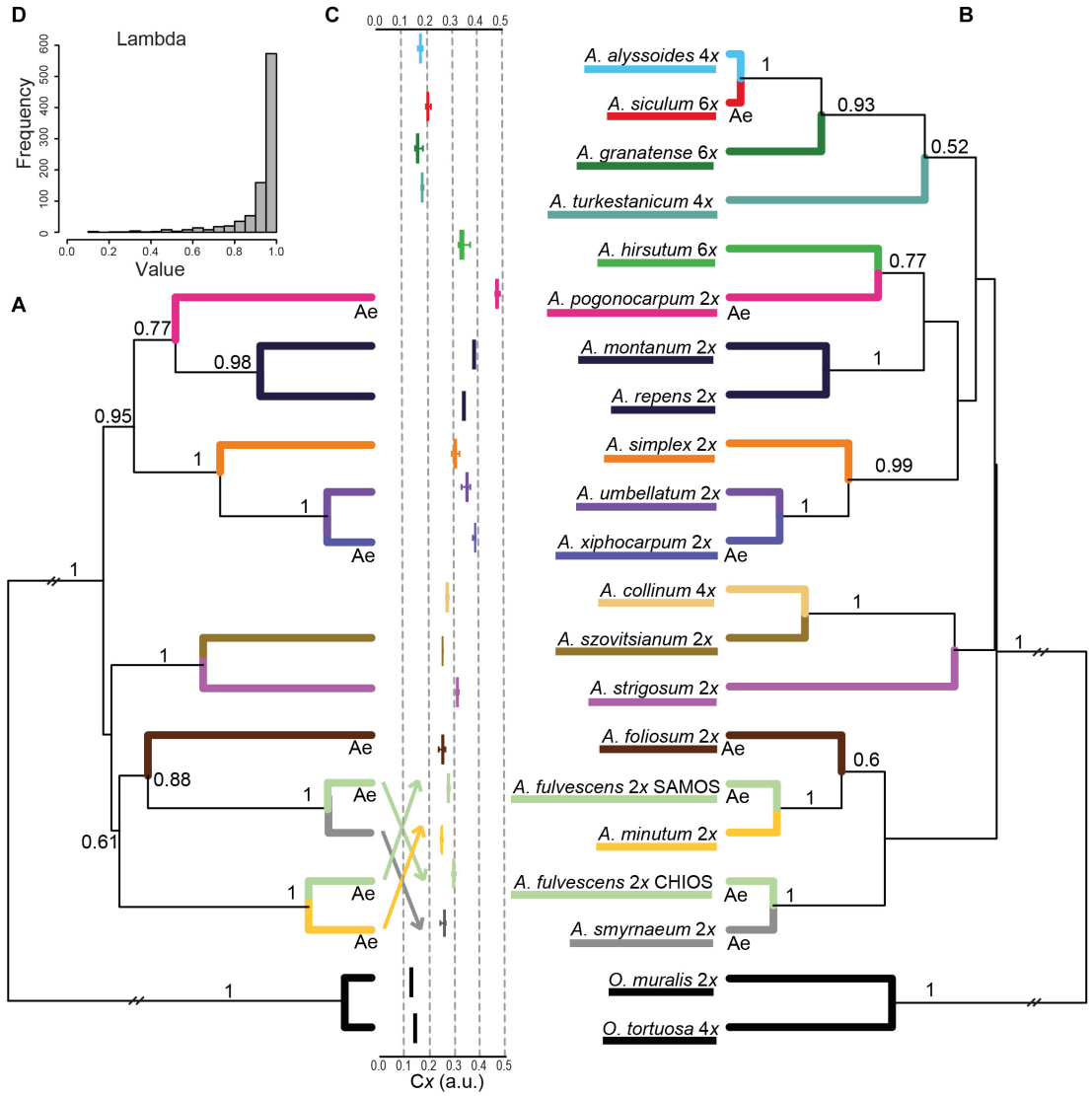


FIGURE 7 | Results of the species tree inference in the studied *Alyssum* species. The maximum clade credibility species trees, which were obtained from the coalescent analysis in BEAST are shown, based on **(A)** cpDNA, ITS, and *DET1* datasets for diploids, and **(B)** cpDNA and ITS datasets for both diploids and polyploids (excluding the tetraploid *A. simulans*, see section "Materials and Methods"). Posterior probability values ≥ 0.5 are indicated above the branches. Ploidy levels are shown next to the species names. 'Ae' below branches indicates the placements of Aegean subendemics referred to in the text. **(C)** Boxplots depicting the relative monoploid genome size (C*x*) values measured for the studied species. See **Table 2** and **Supplementary Table 1** for the precise population- and species-level C*x* values. **(D)** Posterior distribution of the tree transformation statistic λ (Pagel, 1999), computed by MCMC analyses in BayesTraits, quantifying the phylogenetic signal in genome size variation.

Coalescent-Based Species Trees

The species tree inferred from all three DNA markers of diploids is depicted in **Figure 7A**. It includes one larger clade (BPP = 0.95) comprising two Aegean (sub)endemics (*A. pogonocarpum* and *A. xiphocarpum*) and two species with wider ranges (*A. simplex* and *A. umbellatum*), together with the nested *A. montanum* and *A. repens* perennials. The rest of the tree displayed low resolution at higher-level relationships, and only smaller clades can be highlighted as follows. The widespread southern Europe-southwestern Asian *A. strigosum* formed a clade with southwestern Asian *A. szovitsianum* (BPP = 1). In agreement with the individual gene trees, *A. fulvescens* from Samos was resolved in the sister position to widespread *A. minutum* (BPP = 1), whereas *A. fulvescens* from Chios was sister to the Aegean subendemic *A. smyrnaeum* (BPP = 1). Another Aegean subendemic, *A. foliosum*, clustered with the latter two species, albeit with somewhat lower support (BPP = 0.88). The species tree based on cpDNA and ITS alignments and including both diploids and polyploids (**Figure 7B**) showed virtually the same topology as the former species tree (in terms of the relationships among the diploids), but with lower BPP for some clades. Two polyploids, *A. hirsutum* and *A. collinum*, were placed within the clades of

diploids. In contrast, the polyploids *A. granatense, A. siculum,* and *A. alyssoides* formed a distinct clade (BPP = 0.93), and together with another polyploid *A. turkestanicum*, were separated from the diploids.

The boxplots showing the relative monoploid genome size values (C*x*) measured in the studied species (**Figure 7C**) and plotted next to the species tree indicate that the observed genome size variation is phylogenetically structured. The presence of a strong phylogenetic signal in the genome size data was, indeed, supported by the posterior distribution of λ values, which were skewed toward 1 (**Figure 7D**).

Estimation of Divergence Times Based on ITS Sequence Data

The dated phylogeny based on the ITS sequence data (**Supplementary Figure 4**) suggested that the most recent common ancestor (MRCA) of the genus *Alyssum* evolved during the Tortonian Age of the Late Miocene epoch [median age of 9.58 mya, 95% high posterior density (HPD): 7.11–14.09]. The origin and early diversification of the annual lineage examined in this study, including the nested subclade of the perennial *A. montanum-A. repens* species complex, could be dated back to the Late Miocene (Messinian Age) up to the Early Pliocene (median age of the MRCA: 5.38 mya, 95% HPD: 4.15–7.85). The annuals were split into three clades. The first clade, comprising most of the species, originated and began to diversify as early as in the Early Pliocene (median age of 4.66 mya, 95% HPD: 3.36–6.44), followed by further speciation during the Pliocene and Pleistocene. The MRCA of the two other clades, comprising the remaining four and three species, were inferred to appear later (median ages of 3.51 mya, 95% HPD: 2.59–5.39; and 1.71 mya, 95% HPD: 0.91–3.47, respectively). In summary, when confronted with the clades resolved in the species tree (**Figure 7A**, see also **Table 4**), we can derive that the annual species examined in this study have likely originated in the Late Pliocene and the Pleistocene epochs.

DISCUSSION

Distribution and Diversity Patterns of the Studied *Alyssum* Annuals: An Evolutionary Hotspot in the Aegean Region

The phylogenetic lineage of the *Alyssum* species examined in the present study includes 16 species growing in Europe. Their distribution is centered in southern Europe, especially in the eastern Mediterranean and the Aegean archipelago (Jalas et al., 1996; Hartvig, 2002; Strid, 2016). As many as 12 species occur in the Aegean region, seven of which are subendemics that are either restricted to this region or extend only to the adjacent mainland regions of Greece, North Macedonia, western Turkey or Sicily (Dudley, 1965; Hartvig, 2002; Strid, 2016). Our phylogenetic reconstructions revealed that these Aegean subendemics do not form a single monophyletic group but are placed in several distinct clades across the inferred species trees (**Figure 7**), and include both diploids and polyploids. Along with the divergence time estimates, these results suggest that the Aegean subendemics likely originated at different times and through different modes. Thus, it seems that the Aegean region has repeatedly acted as a significant evolutionary center favoring multiple speciation events in this phylogenetic lineage. Indeed, this area is recognized to be one of the major biodiversity hotspots of the Mediterranean Basin, featuring high species richness and a significant proportion of narrow endemics (Médail and Quézel, 1997; Panitsa et al., 2018). Several factors have been highlighted that are likely to drive plant diversification and distribution patterns in the Aegean area, especially its complex climatic and geological history, high topographical and geological heterogeneity, which contributes to the formation of large habitat diversity, as well as the long-term presence and influence of humans (Sfenthourakis and Triantis, 2017; Panitsa et al., 2018).

In this study, we have inferred that the origin and early diversification of the studied lineage of annual *Alyssum* species could be dated back to the Late Miocene and Pliocene, between the Messinian salinity crisis (Krijgsman et al., 1999) and the establishment of the Mediterranean climate (Suc, 1984). The onset of the Mediterranean climate was a major climatic change, which likely promoted the further diversification of this lineage, in a similar manner as identified for some other Mediterranean-adapted plant groups (Fiz-Palacios and Valcárcel, 2013; see also Carnicero et al., 2017). Most extant *Alyssum* species, including those that are distributed throughout the Aegean, originated

TABLE 4 | Age estimates (median and 95% of the highest posterior density, HPD) in millions of years for the nodes of the given clades of the studied *Alyssum* species, as inferred from the relaxed molecular-clock analysis, performed in BEAST and based on ITS sequence data (see **Supplementary Figure 4** and **Supplementary Table 2**).

Clade	Median	95% HPD lower – upper interval	BPP
A. smyrneum + *A. fulvescens* (Chios)	0.85	0.42–1.70	1.00
A. minutum + *A. fulvescens* (Samos)	1.71	0.91–3.47	1.00
A. szovitsianum + *A. collinum* + *A. strigosum*	3.27	2.15–5.33	0.97
A. szovitsianum + *A. collinum*	1.22	0.71–2.33	1.00
A. umbellatum + *A. xiphocarpum* + *A. simplex*	1.40	0.89–2.39	1.00
A. umbellatum + *A. xiphocarpum*	0.53	0.26–1.14	1.00
A. alyssoides + *A. siculum* + *A. granatense* + *A. turkestanicum**	3.51	2.59–5.39	1.00
A. alyssoides + *A. siculum* + *A. granatense*	1.20	0.77–2.20	1.00
A. alyssoides + *A. siculum*	0.68	0.44–1.21	1.00
All studied annual species incl. the perennial *A. montanum–A. repens* species complex	5.38	4.15–7.85	1.00

**The clade on the calibrated ITS tree includes also nested A. montanum-A. repens species group.*

*Only the clades resolved on the coalescence species trees (see **Figure 7**) are listed here. The BPP column shows posterior probability values for the given clades on the calibrated ITS tree (**Supplementary Figure 4**).*

at the end of the Pliocene or during the Pleistocene. This was the era when the climatic oscillations caused sea-level changes and recurrent land connections and disconnections (Perissoratis and Conispoliatis, 2003). They affected the speciation processes by facilitating species dispersal during sea-level drops in glacial periods, and increasing the degree of range fragmentation and population isolation during interglacial and postglacial periods (Nieto Feliner, 2014; Simaiakis et al., 2017; Panitsa et al., 2018). The *Alyssum* species do not bear specific adaptations to facilitate the long-distance dispersal of seeds; instead, field observations suggest that seed dispersal occurs primarily through gravity and zoochory (sheep and goats), and thus also through human activities, mainly cattle farming and grazing. Therefore, geographic distance and barriers may efficiently reduce gene flow, which is manifested by the existence of a few narrow endemics. Interestingly, most species that occur in the Aegean region are largely sympatric (**Table 1**), with several species recorded within one island (e.g., four species in Lesvos and Samos, and six in Crete), with as many as five species recorded from a single locality in Crete (**Supplementary Table 1**). Thus, a simple vicariance scenario associated with allopatric differentiation that has generally been favored in the fragmented Mediterranean landscape (Thompson, 2020) and has been observed for a number of Mediterranean plant groups (e.g., Bittkau and Comes, 2009; Crowl et al., 2015), cannot explain the *Alyssum* species diversity observed in the Aegean region. Considering the common co-occurrence of multiple species within a single site, ecological speciation (Coyne and Orr, 2004) does not appear to play a determinant role either, despite the wide range of occupied habitats, the substrate-specificity observed in some cases (Hartvig, 2002; Strid, 2016, see also **Table 1**), and the large elevational gradient. However, detailed ecological niche characterization and modeling to compare the potential and realized niche spaces of the species (see, e.g., López-Jurado et al., 2019; Castro et al., 2020) will be needed in future to address this issue in detail. We assume that the restricted distribution of some narrow endemics is more likely due to low-dispersal ability rather than environmental constraints [as inferred e.g., by Crowl et al. (2015)].

The distribution ranges can change over time, either naturally or due to human factor, which both can be expected in the biogeographically and geologically complex Aegean area with the long history of human presence (Panitsa et al., 2018; Thompson, 2020). When inspecting the phylogenetic relationships in comparison with the distributional patterns, a specific pattern emerged in some cases. Several diploid species and their closest relatives (also diploids) occur parapatrically or sympatrically, with one species characterized by a restricted distribution (being a narrow endemic), whereas the other occupying a much broader range. This finding was observed for the species pair *A. umbellatum* and *A. xiphocarpum*, as well as for species from the clade comprising *A. fulvescens*, *A. smyrnaeum*, *A. minutum,* and *A. foliosum* (**Figure 7**). This pattern resembles a specific type of allopatric speciation, which is referred to as founder-event speciation (Templeton, 2008), and was recently inferred in the Mediterranean genus *Cymbalaria* (Carnicero et al., 2017). This mode of speciation involves the establishment of a new (peripheral) population from a larger ancestral population, and its genetic divergence due to genetic drift, as well as divergent selective response (Naciri and Linder, 2020). We propose that such events may have been facilitated by recurrent range shifts, expansion or fragmentation, in response to sea-level changes in the Pleistocene (Nieto Feliner, 2014). In addition, since the studied *Alyssum* annuals commonly grow on disturbed sites and in pastures, human-driven colonization due to cattle transport may also have shaped and changed the distribution patterns. Thus, even though these sister species now grow sympatrically, they may have evolved through allopatric pathways at a finer spatial scale. The development of some premating isolating barriers (Coyne and Orr, 2004) have also likely played an important role in the speciation process, as phenological shifts have been observed among some of these co-occurring, closely related species (personal observations, see also below).

A few recent biogeographic studies that have focused on the Aegean area have investigated the diversification, speciation, and colonization processes of some plant groups in detail (Crowl et al., 2015; Jaros et al., 2018; Carnicero et al., 2021). These studies have suggested primarily allopatric, non-adaptive lineage diversification and speciation, driven by geographic isolation and random drift. These processes have been associated with the formation of the mid-Aegean Trench in the Middle Miocene (12–9 mya), when the ancient Aegean landmass became fragmented (Sfenthourakis and Triantis, 2017), and the Pleistocene-era sea-level oscillations, depending on the age of the studied species (Bittkau and Comes, 2005, 2009; Crowl et al., 2015; Jaros et al., 2018). In other cases, sympatric ecological speciation and founder-event speciation have also been inferred (Carnicero et al., 2017, 2021). Thus, the few detailed studies that have been performed so far, along with the here presented scenarios, highlight the complexity of speciation processes in the Aegean region.

Multiple Allopolyploid Speciation Events Increased the Species Diversity Across the Whole Distribution Range

We revealed the presence of diploid, tetraploid, and hexaploid levels among the studied *Alyssum* annuals, in accordance with earlier reports (summarized in AlyBase, see Španiel et al., 2015). We did not observe ploidy level variation within populations or within species. The diploids and tetraploids that were traditionally classified within *A. simplex* were recently split into two separate species, *A. simplex* (diploids) and *A. collinum* (tetraploids), which was substantiated by preliminary molecular data, by differences in the monoploid genome size, and some morphological traits (Cetlová et al., 2019). Indeed, the here presented phylogenetic analyses placed these two species in two distinct clades. In addition, the parentage of the tetraploid *A. collinum* did not appear to involve the diploid *A. simplex* (see also below). Some ploidy level variation was previously reported within a few species (see **Table 2**); however, these observations remain to be confirmed by future studies. In this study, all examined annual polyploids (seven species in total) appeared to evolve through allopolyploidy. Interestingly, only three of

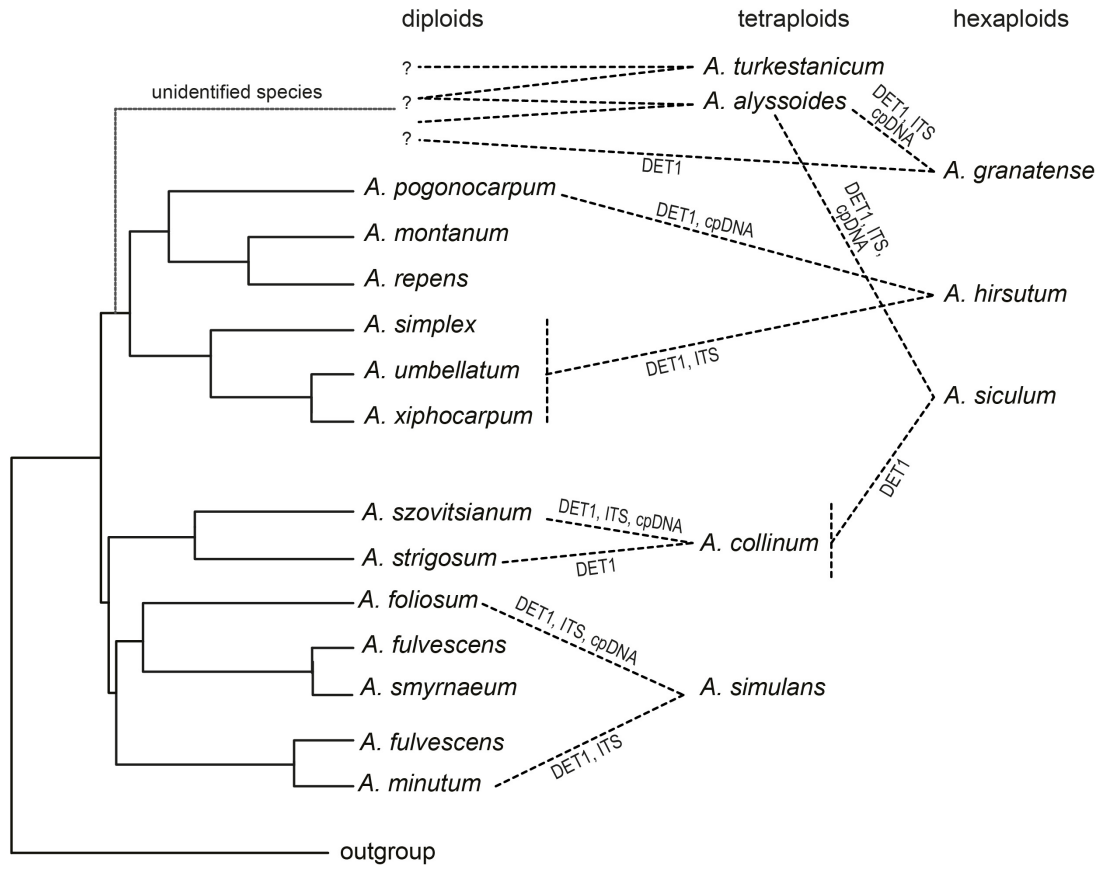


FIGURE 8 | Scheme summarizing the allopolyploidization events in the studied *Alyssum* species, as inferred from DNA sequence data. Diploid species are depicted in the coalescent species tree, as shown in **Figure 7A**, based on all three markers used, whereas polyploids are connected to their inferred diploid or lower-ploidy progenitors. Question marks indicate unidentified diploid progenitors (species currently not occurring in Europe, possibly of Asian origin or extinct). See the text for further details.

these polyploids occur in the Aegean region, thus, it seems that although the polyploidization played a certain evolutionary role in this area as well, it was not the prevalent speciation mechanism there.

The allopolyploid origins of the studied annuals were derived from the cpDNA–ITS incongruence, from ITS intragenomic polymorphisms, and from single-copy (*DET1*) homeolog variation. The combination of multiple markers demonstrating different patterns of inheritance and molecular evolution has been shown to be the most efficient approach for reconstructing reticulate evolution (Naciri and Linder, 2015; e.g., Záveská et al., 2016; Díaz-Pérez et al., 2018; Mandák et al., 2018). In this study, we combined the nrDNA region that is prone to sequence homogenization by concerted evolution and may show complex patterns in polyploids (Álvarez and Wendel, 2003; Nieto Feliner and Rosselló, 2012), cpDNA with uniparental inheritance, and a biparentally inherited single-copy nuclear gene with the potential to retain parental subgenomes (Small et al., 2004; Nieto Feliner and Rosselló, 2012). In some cases, we were able to identify the most likely parental species or at least the phylogenetic lineages involved in the polyploid origins, which are discussed below and summarized in **Figure 8**.

The tetraploid *A. simulans,* which is distributed in southeastern mainland Greece, Peloponnese, Evvia, and Crete, was the only species in which divergent, non-homogenized ITS variants were revealed, clustering with two distinct diploids, Aegean subendemic *A. foliosum* and widespread Mediterranean *A. minutum*, in concert with the identified *DET1* homeologs. The former diploid was revealed as the probable maternal parent, based on the cpDNA patterns. All three species occur in sympatry, and these distribution patterns, non-homogenized nrDNA, and almost identical genome size (monoploid C*x* values) suggested a recent, likely Late Pleistocene or even postglacial allopolyploid origin. *Alyssum simulans* was described only recently, and its hybrid origin was already hypothesized based on morphological intermediacy (Hartvig, 2002). It remains uncertain, however, whether *A. simulans* originated multiple times, i.e., independently in Crete and in mainland Greece, or just at a single site and expanded to its current range after its origin.

The hexaploid *A. hirsutum* is widely distributed across eastern Europe, especially in the Black Sea region, extending to the Caucasus region (Grossheim, 1949; Jalas et al., 1996), but does not grow in the Aegean area. One of its parental species was suggested to originate from the clade comprising *A. xiphocarpum* (which

is endemic to Lesvos), *A. umbellatum* (Aegean subendemic), and *A. simplex* (throughout the entire Mediterranean, reaching the Caucasus). Thus, either *A. simplex* (being more similar to *A. hirsutum* morphologically than the other two species, which have short umbellate fruiting racemes) or a common ancestor of that clade appear to be the most likely parental species. All three species have similar monoploid genome size values, which are also close to *A. hirsutum*. The identity of the other parental species remains puzzling because both cpDNA and *DET1* point to *A. pogonocarpum*, a species restricted to Rhodos and southwestern Turkey (Kleinsteuber et al., 2016), which might be supported by the morphological resemblance between these species (very long hairs on fruits, Carlström, 1984), but occurs allopatrically and also has a markedly larger genome size (**Figure 7C** and **Table 2**). Nevertheless, also an unsampled, related species from the Irano-Anatolian or Caucasus areas could have been involved in the origin of hexaploid *A. hirsutum* (see also below).

The tetraploid *A. collinum* contained ITS and cpDNA sequences that both clustered together with the diploid *A. szovitsianum*. In the *DET1* locus, two slightly divergent groups of alleles were identified, clustering with the sister species *A. szovitsianum* (southwestern Asia) and *A. strigosum* (southern Europe-southwestern Asia). This pattern favors an allopolyploid origin involving progenitors close to these two relatives. Also the monoploid genome size of *A. collinum* is clearly intermediate between these two species (**Figure 7C** and **Table 2**). The distribution of *A. collinum* remains poorly understood because it has only recently been taxonomically segregated from the widespread Mediterranean diploid *A. simplex* (Cetlová et al., 2019). The occurrence of *A. collinum* is so far confirmed only from the western Mediterranean (Cetlová et al., 2019, and the present samples), and no records exist from eastern Europe. Some tetraploid individuals (under the name *A. simplex*) were reported from Iran and Tajikistan (see **Table 2**), but it is unclear if they could be conspecific with *A. collinum* or represent autotetraploids of *A. simplex*, or even another species. Since the present sequence data suggest that the parental species of *A. collinum* came from the eastern regions, it is evident that the taxonomic identity of the Asian tetraploids should be resolved to delimit the distribution range of *A. collinum* completely, and to understand its origin and evolution.

The remaining four polyploids (the tetraploids *A. alyssoides* and *A. turkestanicum*; the hexaploids *A. granatense* and *A. siculum*) did not show closer affinities to any of the European diploids; however, their divergent *DET1* sequences (**Figure 6**), which were placed in clearly distinct clades, strongly favored independent allopolyploid origins. The most widespread tetraploid *A. alyssoides* displayed two divergent *DET1* copies, which were both present in the hexaploids *A. siculum* and *A. granatense,* and one in the tetraploid *A. turkestanicum*. Thus, *A. alyssoides* was most likely involved in the origin of both hexaploids, i.e., the western Mediterranean *A. granatense* (which is in congruence with their similar morphology, Küpfer and Nieto Feliner, 1993) and the eastern Mediterranean *A. siculum*. The third *DET1* homeolog of *A. granatense*, in agreement with the distinct clade formed by its cpDNA sequences, pointed to another progenitor, which, however, remained unidentified. The hexaploid *A. siculum*, based on previous morphological and molecular data, was suggested to be either an allopolyploid, derived from *A. alyssoides* and *A. simplex,* or an autopolyploid of *A. alyssoides* (Persson, 1971; Rešetnik et al., 2013). In this study, we found that *A. siculum*, which is distributed in Crete, Peloponnese, mainland Greece, and Sicily, indeed, displayed lot of shared variation in all three markers with *A. alyssoides*, which may suggest a very recent polyploid origin. In contrast, we did not find any evidence that *A. simplex* could be the second parent. Instead, one of the *DET1* homeologs revealed in *A. siculum* was sister to the clade containing *A. collinum*, *A. szovitsianum,* and *A. strigosum*, which suggests that one of these species may have been the paternal parent of the hexaploid *A. siculum*. Based on the present records, only *A. strigosum* occurs sympatrically (in mainland Greece as well as Crete; Hartvig, 2002; Strid, 2016) with the hexaploid; however, the highest morphological resemblance points to the southwestern Asian *A. szovitsianum* (personal observation). We did not observe any genetic differentiation between the accessions of *A. siculum* from mainland Greece, Crete, and Sicily that could suggest the independent, polytopic origins of this hexaploid. However, the present analyses were based on only a few loci and may not be sufficient to detect multiple polyploid origins. Where this polyploid originated and how it reached its current disjunct distribution also remains an open question. The sea barriers between mainland Greece, Crete, and Sicily have persisted since the late Miocene; therefore, either a natural long-distance dispersal event occurred or introduction by man was possible, as it commonly grows on disturbed sites and pastures. Finally, the tetraploid *A. turkestanicum*, which is distributed from the Balkan Peninsula to eastern Europe, Black Sea Region, Caucasus, Near East, central Asia, and Siberia (Rechinger, 1968 as *Alyssum desertorum*; Rybinskaya, 1994; Jalas et al., 1996; Zhou et al., 2001; German, 2003; Marhold, 2011), formed a distinct clade according to both the ITS and cpDNA-based trees, with no obvious affinities to any European diploids. One *DET1* homeolog was in a clade with *A. alyssoides*, which suggests several scenarios: *A. alyssoides* may have been involved in its origin, both species may share a common diploid ancestor of extra-European origin, or homeolog sharing may also reflect local hybridization events. Interestingly, these four polyploids (*A. alyssoides, A. turkestanicum, A. granatense*, and *A. siculum*) have the smallest monoploid genome sizes from all here analyzed species (**Figure 7C**), which may suggest that their unidentified diploid progenitors also have small genomes, and/or that genome-downsizing occurred as a result of diploidization, a process commonly observed in polyploids (Dodsworth et al., 2016). Genome-downsizing, however, was not observed in the other three polyploids (*A. collinum, A. simulans*, and *A. hirsutum*); their genome sizes were rather correlated with the phylogenetic positions and, thus, reflected the genome size variation among their diploid ancestors (**Table 2** and **Figure 7C**).

The significance of recent polyploidization events (neopolyploidy) for the diversification and speciation processes has been inferred for a number of Mediterranean genera, which have been especially linked to climate-induced range shifts and secondary contacts during the Pleistocene

(e.g., Cecchi et al., 2013; Frajman et al., 2016; Wagner et al., 2019). Except for two polyploids, *A. siculum* and *A. simulans*, which are both narrowly distributed and restricted to certain parts of the Mediterranean regions (SE Greece, Crete, plus Sicily, in the case of *A. siculum*) and supposedly of very recent origins, the other studied polyploids have broader distributions, some also occurring in extra-Mediterranean areas (e.g., the Caucasus, central Asia). This finding may support the hypothesis that allopolyploids, harboring greater genetic variation, complex and dynamic genomes, could become successful colonizers that exhibit larger distribution areas and different, often wider, ecological niches than diploids (Soltis et al., 2014; see e.g., Arrigo et al., 2016; Cornille et al., 2016). Still, it may be premature to favor or reject this hypothesis with the present data, as in the case of these widespread *Alyssum* polyploids, either at least one of its diploid progenitors also shows widespread occurrence (cases of *A. hirsutum* and *A. granatense*), or the diploid progenitors (and their ecological and chorological characteristics) remain unknown (cases of *A. turkestanicum* and *A. alyssoides*).

Phylogenetic Patterns, Cryptic Diversity, and Further Perspectives

The phylogenetic trees that were generated based on the three genetic markers employed in this study differed in their resolution. The best resolved hierarchical structure was obtained from the cpDNA tree, in which most species were monophyletic, and the higher-level relationships received at least moderate support. Few species appeared paraphyletic, which most likely reflects lack of resolution or shared ancestral variation, as this pattern was only observed in some closely related diploids (e.g., *A. minutum* and *A. fulvescens* from Samos; *A. smyrnaeum* and *A. fulvescens* from Chios) or indicated the maternal origins of some polyploids (e.g., *A. siculum* and *A. alyssoides*; *A. simulans* and *A. foliosum*). In contrast, both the ITS- and *DET1*-based trees displayed a low resolution in the tree backbone, whereas the more terminal clades of pairs or groups of closely related species were well-supported and largely congruent among all three markers. The low resolution of the nuclear trees, despite the sufficient number of variable sites, most likely mirrors the more complicated evolution of biparentally inherited markers, in which past interspecific gene flow, during and after speciation, polyploidization, recombination, and, in the case of ITS also processes of sequence homogenization, may have obscured the divergence history, and resulted in conflicting, reticulate patterns (Nieto Feliner and Rosselló, 2012; Naciri and Linder, 2015). In agreement with the individual gene trees, the initial diversification patterns were poorly resolved in the species tree, although the terminal clades of the closest relatives received high support. The close relatedness of some species, as was revealed in this study, was already suggested based on their morphological similarity (e.g., *A. fulvescens* and *A. smyrnaeum*; *A. xiphocarpum* and *A. umbellatum*; Hartvig, 2002). An intriguing case remains the eastern Aegean endemic *A. fulvescens* (distributed in Samos, Chios, and two smaller islands of Patmos and Kalymnos, Strid, 2016). The populations from the Chios and Samos islands that were analyzed in this study were genetically differentiated, despite no obvious morphological distinction. Consistently for each genetic marker, the samples from Chios were genetically closest to *A. smyrnaeum* (the more widespread Aegean subendemic, which grows in both Samos and Chios islands, Strid, 2016 and present records), whereas those from Samos clustered with *A. minutum* (a widespread Mediterranean species, including the Aegean region, but not recorded from Samos and Chios). This pattern cannot be explained by a recent hybridization or introgression event with either of these species, considering also the fact that *A. fulvescens* from Samos grows at the same site, together with *A. smyrnaeum* (pop. 595KEK), but they apparently do not hybridize nowadays and they display some phenological shifts (our field observations). Thus, we can only speculate whether the present patterns are due to past hybridization events that caused genetic divergence within *A. fulvescens* or whether they indicate the existence of two independent and separately evolving entities with low levels of morphological differentiation, also known as cryptic species (Fišer et al., 2018; see, e.g., Crowl et al., 2015, 2017; Padilla-García et al., 2018). Considerable among-population variation was observed in genome size values in this species, although it does not correlate precisely with genetic patterns. Still, these patterns may indicate some hidden evolutionary processes in this species. In the future, more detailed investigations of morphological and genetic variation in *A. fulvescens* should be undertaken to solve this issue.

Genome size variation observed among the analyzed species was congruent with the phylogenetic patterns (**Figure 7**). The monoploid relative genome size of the annual species studied here ranged from 0.161 in *A. granatense* to 0.474 in *A. pogonocarpum* (**Table 2**). Perennial species of the *A. montanum–A. repens* complex, according to previously published studies, showed a much narrower range of values, between 0.296 ± 0.004 (in some populations of *A. repens*, Melichárková et al., 2019) and 0.389 ± 0.006 (in *Alyssum cacuminum*, Španiel et al., 2018, 2019). These values are within the range of those of the here studied annuals, in agreement with the phylogenetic position of the perennial *A. montanum–A. repens* group, nested within the lineage of annuals, suggesting that these perennials could have evolved from the annuals. It has been proposed and also documented in several cases that annuals tend to have smaller genomes compared to perennials (e.g., Frajman et al., 2015; Zahradníček et al., 2018), albeit this correlation may be also due to the common association of annual life history with higher rates of selfing (Albach and Greilhuber, 2004). Here, we did not support this tendency, in congruence with the findings by Cacho et al. (2021), as genome size variation displayed significant phylogenetic correlation, irrespective of the annual or perennial life history. It has been shown that most of genome size variation in plants is attributable to the differential evolution of repetitive DNA components, which may represent a highly dynamic process that is correlated with species relatedness and phylogeny (McCann et al., 2020).

One limitation of the present study, however, is that we did not include species and populations from the extra-European

range (except of *A. szovitsianum*). About 12 species from northern Africa and southwestern to central Asia not sampled here (Dudley, 1964a, 1965), will be needed in future to resolve the evolutionary and biogeographic history of these *Alyssum* annuals completely. So far, up to five extra-European species were included in previous tribus- or genus-wide phylogenetic studies (Li et al., 2015; Huang et al., 2020), which supported their placement within the here studied clade of annuals, and appeared close to some more widespread Eurasian species (*A. minutum, A. simplex*, and *A. strigosum*), similarly as resolved here for *A. szovitsianum* from Iran. Still, the existence of an independent Asian lineage cannot be ruled out either. Taxonomic identification of some Asian species, however, may be tentative, and thorough taxonomic revision of Asian species is necessarily needed as first. The complete Eurasian sampling in future studies may help to identify the missing parental species for some polyploids (see above), reinforce or weaken the phylogenetic signal in genome size, and will allow to gain deeper insights into the colonization and speciation processes in this group.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MW070213–MW070364; https://www.ncbi.nlm.nih.gov/genbank/, MW022541–MW022804; and https://www.ncbi.nlm.nih.gov/genbank/, MW070365–MW070517.

AUTHOR CONTRIBUTIONS

SŠ and JZ-L conceived and designed the study and wrote the manuscript. SŠ and VC sampled plant material. All authors generated data and performed data analyses. All authors have read, revised, and approved the final manuscript.

FUNDING

This study was financially supported by the Slovak Research and Development Agency (APVV; Grant No. APVV-17-0616), Czech Science Foundation (GAČR; Grant No. 19-06632S), and Grant Agency VEGA, Bratislava, Slovakia (Grant No. 2/0133/17).

ACKNOWLEDGMENTS

We are grateful to Marek Šlenker, Pavol Mereďa Jr., Dominik Roman Letz (all three from Bratislava), and Gianniantonio Domina (Palermo) for their assistance during field trips. We also thank to Arne Strid (Ørbæk), Ioannis Bazos (Athens), and Michael Hassler (Bruchsal) for advice on localities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.627909/full#supplementary-material

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Albach, D. C., and Greilhuber, J. (2004). Genome size variation and evolution in Veronica. *Ann. Bot.* 94, 897–911. doi: 10.1093/aob/mch219

Al-Shehbaz, I. A. (2010). "*Alyssum* linnaeus," in *Flora of North America, North of Mexico, Magnoliophyta: Salicaceae to Brassicaceae*, Vol. 7, ed. Flora of North America Editorial Committee (New York, NY: Oxford University Press, Inc.), 247–251.

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

Álvarez, I., and Wendel, J. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogen. Evol.* 29, 417–434. doi: 10.1016/S1055-7903(03)00208-2

Arrigo, N., de La Harpe, M., Litsios, G., Zozomová-Lihová, J., Španiel, S., Marhold, K., et al. (2016). Is hybridization driving the evolution of climatic niche in *Alyssum montanum*? *Am. J. Bot.* 103, 1348–1357. doi: 10.3732/ajb.1500368

Ball, P. W., and Dudley, T. R. (1993). "*Alyssum* L," in *Flora Europaea*, 2nd Edn. Vol. 1, eds T. G. Tutin, N. A. Burges, A. O. Chater, J. R. Edmondson, V. H. Heywood, D. M. Moore, et al. (Cambridge: Cambridge University Press), 359–369.

Bittkau, C., and Comes, H. P. (2005). Evolutionary processes in a continental island system: molecular phylogeography of the Aegean *Nigella arvensis* alliance (Ranunculaceae) inferred from chloroplast DNA. *Mol. Ecol.* 14, 4065–4083. doi: 10.1111/j.1365-294X.2005.02725.x

Bittkau, C., and Comes, H. P. (2009). Molecular inference of a late Pleistocene diversification shift in *Nigella* s. lat. (Ranunculaceae) resulting from increased speciation in the Aegean archipelago. *J. Biogeogr.* 36, 1346–1360. doi: 10.1111/j.1365-2699.2008.02003.x

Borchsenius, F. (2009). *FastGap 1.2*. Denmark: Department of Biosciences, Aarhus University.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537

Cacho, N. I, McIntyre, P. J., Kliebenstein, D. J., and Strauss, S. Y. (2021). Genome size evolution is associated with climate seasonality and glucosinolates, but not life history, soil nutrients or range size, across a clade of mustards. *Ann. Bot.* (in press). doi: 10.1093/aob/mcab028

Carlström, A. (1984). New species of *Alyssum*, *Consolida*, *Origanum* and *Umbilicus* from the SE Aegean sea. *Willdenowia* 14, 15–26.

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

Carnicero, P., Sáez, L., Garcia-Jacas, N., and Galbany-Casals, M. (2017). Different speciation types meet in a Mediterranean genus: the biogeographic history of *Cymbalaria* (Plantaginaceae). *Taxon* 66, 393–407. doi: 10.12705/662.7

Castro, M., Loureiro, J., Figueiredo, A., Serrano, M., Husband, B. C., and Castro, S. (2020). Different patterns of ecological divergence between two tetraploids and their diploid counterpart in a parapatric linear coastal distribution polyploid complex. *Front. Plant Sci.* 11:315. doi: 10.3389/fpls.2020.00315

Cecchi, L., Colzi, I., Coppi, A., Gonnelli, C., and Selvi, F. (2013). Diversity and biogeography of Ni-hyperaccumulators of *Alyssum* section *Odontarrhena* (Brassicaceae) in the central western Mediterranean: evidence from karyology, morphology and DNA sequence data. *Bot. J. Linn. Soc.* 173, 269–289. doi: 10.1111/boj.12084

Cetlová, V., Fuertes-Aguilar, J., Iudova, D., and Španiel, S. (2019). Overlooked morphological variation and a proposal for a new taxonomic circumscription

of *Alyssum simplex* (Brassicaceae). *Phytotaxa* 416, 149–166. doi: 10.11646/phytotaxa.416.2.3

Clement, M., Posada, D., and Crandall, K. (2000). TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1659. doi: 10.1046/j.1365-294x.2000.01020.x

Cornille, A., Salcedo, A., Kryvokhyzha, D., Glémin, S., Holm, K., Wright, S. I., et al. (2016). Genomic signature of successful colonization of Eurasia by the allopolyploid shepherd's purse (*Capsella bursa-pastoris*). *Mol. Ecol.* 25, 616–629. doi: 10.1111/mec.13491

Correggiari, A., Field, M. E., and Trincardi, F. (1996). Late Quaternary transgressive large dunes on the sediment-starved Adriatic shelf. *Geol. Soc. Spec. Publ.* 117, 155–169. doi: 10.1144/GSL.SP.1996.117.01.09

Coyne, J. A., and Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates, Inc.

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

Crowl, A. A., Visger, C. J., Mansion, G., Hand, R., Wu, H. H., Kamari, G., et al. (2015). Evolution and biogeography of the endemic *Roucela* complex (Campanulaceae: Campanula) in the Eastern Mediterranean. *Ecol. Evol.* 5, 5329–5343. doi: 10.1002/ece3.1791

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109

Díaz-Pérez, A., Lopez-Alvarez, D., Sancho, R., and Catalan, P. (2018). Reconstructing the origins and the biogeography of species' genomes in the highly reticulate allopolyploid-rich model grass genus *Brachypodium* using minimum evolution, coalescence and maximum likelihood approaches. *Mol. Phylogen. Evol.* 127, 256–271. doi: 10.1016/j.ympev.2018.06.003

Dodsworth, S., Chase, M. W., and Leitch, A. R. (2016). Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot. J. Linn. Soc.* 180, 1–5. doi: 10.1111/boj.12357

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

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Dudley, T. R. (1964a). Studies in *Alyssum*: Near Eastern representatives and their allies, I. *J. Arnold Arbor.* 45, 57–100.

Dudley, T. R. (1964b). Synopsis of the genus *Alyssum*. *J. Arnold Arbor.* 45, 358–373.

Dudley, T. R. (1965). "*Alyssum* L," in *Flora of Turkey*, Vol. I, ed. P. H. Davis (Edinburgh: Edinburgh University press), 362–409.

Fišer, C., Robinson, C. T., and Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Mol. Ecol.* 27, 613–635. doi: 10.1111/mec.14486

Fiz-Palacios, O., and Valcárcel, V. (2013). From Messinian crisis to Mediterranean climate: a temporal gap of diversification recovered from multiple plant phylogenies. *Perspect. Plant Ecol.* 15, 130–137. doi: 10.1016/j.ppees.2013.02.002

Frajman, B., Rešetnik, I., Niketić, M., Ehrendorfer, F., and 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. doi: 10.1186/s12862-016-0773-2

Frajman, B., Rešetnik, I., Weiss-Schneeweiss, H., Ehrendorfer, F., and Schönswetter, P. (2015). Cytotype diversity and genome size variation in *Knautia* (Caprifoliaceae, Dipsacoideae). *BMC Evol. Biol.* 15:140. doi: 10.1186/s12862-015-0425-y

German, D. A. (2003). Zametki po rodu *Alyssum* L. (Cruciferae) Kazakhstana / Notes on the genus *Alyssum* L. (Cruciferae) in Kazakhstan. *Turczaninowia* 6, 45–57. [in Russian],

Greilhuber, J., Doležel, J., Lysak, M. A., and Bennett, M. D. (2005). The origin, evolution and proposed stabilization of the terms 'genome size'and 'C-value'to describe nuclear DNA contents. *Ann. Bot.* 95, 255–260. doi: 10.1093/aob/mci019

Grossheim, A. A. (1949). *Opredelitel' Rastenii Kavkaza / Key to the Plants of the Caucasus*. Moscow: Sovietskaia nauka. [in Russian]

Hartvig, P. (2002). "*Alyssum* L," in *Flora Hellenica*, Vol. 2, eds A. Strid and K. Tan (Ruggell: A.R.G. Gantner Verlag K. G.), 199–227.

Hassler, M. (2004-2021). *Flora of Rhodos and Chalki. Picture Atlas and Database. Version 3.05; Last Update 9.1.2021*. Available online at: www.flora-germanica.de/rhodos/ (accessed January 19, 2021).

Hewitt, G. M. (2011). Quaternary phylogeography: the roots of hybrid zones. *Genetica* 139, 617–638. doi: 10.1007/s10709-011-9547-3

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

Huelsenbeck, J. P., and Ronquist, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

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

Jalas, J., Suominen, J., and Lampinen, R. (1996). *Atlas Florae Europaeae*, Vol. 11. Helsinki: The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo.

Jaros, U., Tribsch, A., and Comes, H. P. (2018). Diversification in continental island archipelagos: new evidence on the roles of fragmentation, colonization and gene flow on the genetic divergence of Aegean *Nigella* (Ranunculaceae). *Ann. Bot.* 121, 241–254. doi: 10.1093/aob/mcx150

Jones, G. (2017). Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *J. Math. Biol.* 74, 447–467. doi: 10.1007/s00285-016-1034-0

Kleinsteuber, A., Ristow, M., and Hassler, M. (2016). *Flora von Rhodos und Chalki. Band 1*. Karlsruhe: Verlag Kleinsteuber Books.

Krijgsman, W., Hilgen, F. J., Raffi, I., Sierro, F. J., and Wilson, D. S. (1999). Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400, 652–655. doi: 10.1038/23231

Kubatko, L., and Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* 56, 17–24. doi: 10.1080/10635150601146041

Kuittinen, H., Aguadé, M., Charlesworth, D., Haan, A. D. E., Lauga, B., Mitchell-Olds, T., et al. (2002). Primers for 22 candidate genes for ecological adaptations in Brassicaceae. *Mol. Ecol. Notes* 2, 258–262. doi: 10.1046/j.1471-8286.2002.00210.x

Küpfer, P., and Nieto Feliner, G. (1993). "*Alyssum* L," in *Flora Iberica*, Vol. 4, eds S. Castroviejo, C. Aedo, C. Gómez Campo, M. Laínz, P. Montserrat, R. Morales, et al. (Madrid: Real Jardín Botánico, C.S.I.C.), 167–184.

Lambeck, K., Rouby, H., Purcell, A., Sun, Y., and Sambridge, M. (2014). Sea level and global ice volumes from the last glacial maximum to the Holocene. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15296–15303. doi: 10.1073/pnas.1411762111

Leigh, J. W., and Bryant, D. (2015). PopART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210X.12410

Li, Y., Feng, Y., Lv, G., Liu, B., and Qi, A. (2015). The phylogeny of *Alyssum* (Brassicaceae) inferred from molecular data. *Nord. J. Bot.* 33, 715–721. doi: 10.1111/njb.00588

López-Jurado, J., Mateos-Naranjo, E., and Balao, F. (2019). Niche divergence and limits to expansion in the high polyploid *Dianthus broteri* complex. *New Phytol.* 222, 1076–1087. doi: 10.1111/nph.15663

Magauer, M., Schönswetter, P., Jang, T.-S., and Frajman, B. (2014). Disentangling relationships within disjunctly distributed *Alyssum ovirense*/*A. wulfenianum* group (Brassicaceae), including description of a novel species from the north-eastern Alps. *Bot. J. Linn. Soc.* 176, 486–505. doi: 10.1111/boj.12214

Mandák, B., Krak, K., Vít, P., Lomonosova, M. N., Belyayev, A., Habibi, F., et al. (2018). Hybridization and polyploidization within the *Chenopodium album* aggregate analysed by means of cytological and molecular markers. *Mol. Phylogen. Evol.* 129, 189–201. doi: 10.1016/j.ympev.2018.08.016

Marhold, K. (2011). "Brassicaceae," in *Euro+Med Plantbase – The Information Resource for Euro-Mediterranean Plant Diversity*. Available online at: http://www.emplantbase.org/home.html (accessed June 1, 2020).

McCann, J., Macas, J., Novák, P., Stuessy, T. F., Villaseñor, J. L., and Weiss-Schneeweiss, H. (2020). Differential genome size and repetitive DNA evolution in diploid species of *Melampodium* sect. *Melampodium* (Asteraceae). *Front. Plant Sci.* 11:362. doi: 10.3389/fpls.2020.00362

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

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

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

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

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES science gateway for inference of large phylogenetic trees," in *Proceedings of the Gateway Computing Environments Workshop (GCE)*, (New Orleans, LA: Institute of Electrical and Electronics Engineers (IEEE)), 1–8.

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

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

Naciri, Y., and Linder, H. P. (2020). The genetics of evolutionary radiations. *Biol. Rev.* 4, 1055–1072. doi: 10.1111/brv.12598

Nieto Feliner, G. (2011). Southern European glacial refugia: a tale of tales. *Taxon* 60, 365–372. doi: 10.1002/tax.602007

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

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

Padilla-García, N., Rojas-Andrés, B. M., López-González, N., Castro, M., Castro, S., Loureiro, J., et al. (2018). The challenge of species delimitation in the diploid-polyploid complex *Veronica* subsection *Pentasepalae*. *Mol. Phylogen. Evol.* 119, 196–209. doi: 10.1016/j.ympev.2017.11.007

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature* 401, 877–884.

Pagel, M., Meade, A., and Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* 53, 673–684.

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

Perissoratis, C., and Conispoliatis, N. (2003). The impacts of sea-level changes during latest Pleistocene and Holocene times on the morphology of the Ionian and Aegean seas (SE Alpine Europe). *Mar. Geol.* 196, 145–156. doi: 10.1016/S0025-3227(03)00047-1

Persson, J. (1971). Studies in the Aegean flora XIX. Notes on *Alyssum* and some other genera of Cruciferae. *Bot. Not.* 124, 399–418.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Rechinger, K. H. (1968). *Flora Iranica*, Vol. 57, Graz: Akademische Druck- u. Verlagsanstalt.

Rešetnik, I., Satovic, Z., Schneeweiss, G. M., and Liber, Z. (2013). Phylogenetic relationships in Brassicaceae tribe Alysseae inferred from nuclear ribosomal and chloroplast DNA sequence data. *Mol. Phylogenet. Evol.* 69, 772–786. doi: 10.1016/j.ympev.2013.06.026

Rešetnik, I., and Španiel, S. (2018). The new circumscription of the genus *Alyssum* L. (Brassicaceae) in the flora of Croatia. *Glas. Hrvatskog Bot. Društva* 6, 4–16.

Rybinskaya, E. V. (1994). "*Alyssum* L," in *Flora Sibiri: Berberidaceae–Grossulariaceae*, Vol. 7, eds L. I. Malyshev and G. A. Peshkova (Novosibirsk: VO Nauka), 103–106.

Salmerón-Sánchez, E., Fuertes-Aguilar, J., Španiel, S., Pérez-García, F. J., Merlo, E., Garrido-Becerra, J. A., et al. (2018). Plant evolution in alkaline magnesium-rich soils: a phylogenetic study of the Mediterranean genus *Hormathophylla* (Cruciferae: Alysseae) based on nuclear and plastid sequences. *PLoS One* 13:e0208307. doi: 10.1371/journal.pone.0208307

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

Sfenthourakis, S., and Triantis, K. A. (2017). The Aegean archipelago: a natural laboratory of evolution, ecology and civilisations. *J. Biol. Res. (Thessalon)* 24:4. doi: 10.1186/s40709-017-0061-3

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., et al. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142–166. doi: 10.3732/ajb.92.1.142

Simaiakis, S. M., Rijsdijk, K. F., Koene, E. F., Norder, S. J., Van Boxel, J. H., Stocchi, P., et al. (2017). Geographic changes in the Aegean Sea since the Last Glacial Maximum: postulating biogeographic effects of sea-level rise on islands. *Palaeogeogr. Palaeocl.* 471, 108–119. doi: 10.1016/j.palaeo.2017.02.002

Simmons, M. P., and Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49, 369–381. doi: 10.1093/sysbio/49.2.369

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

Soltis, P. S., Liu, X., Marchant, D. B., Visger, C. J., and Soltis, D. E. (2014). Polyploidy and novelty: Gottlieb's legacy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130351. doi: 10.1098/rstb.2013.0351

Španiel, S., Kaplan, K., Bovio, M., Mártonfiová, L., and Cetlová, V. (2018). *Alyssum rossetii* (Brassicaceae), a new species from the Aosta Valley in Italy based on morphological and genome size data. *Phytotaxa* 360, 269–281. doi: 10.11646/phytotaxa.360.3.7

Španiel, S., Kempa, M., Salmerón-Sánchez, E., Fuertes-Aguilar, J., Mota, J. F., Al-Shehbaz, I. A., et al. (2015). AlyBase: database of names, chromosome numbers, and ploidy levels of Alysseae (Brassicaceae), with a new generic concept of the tribe. *Plant Syst. Evol.* 301, 2463–2491. doi: 10.1007/s00606-015-1257-3

Španiel, S., Marhold, K., Passalacqua, N. G., and Zozomová-Lihová, J. (2011). Intricate variation patterns in the diploid-polyploid complex of *Alyssum montanum-A. repens* (Brassicaceae) in the Apennine Peninsula: evidence for long-term persistence and diversification. *Am. J. Bot.* 98, 1887–1904. doi: 10.3732/ajb.1100147

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

Španiel, S., Marhold, K., and Zozomová-Lihová, J. (2019). Polyphyletic *Alyssum cuneifolium* (Brassicaceae) revisited: morphological and genome size differentiation of recently recognized allopatric taxa. *J. Syst. Evol.* 57, 287–301. doi: 10.1111/jse.12464

Španiel, S., Zozomová-Lihová, J., and Marhold, K. (2017b). Revised taxonomic treatment of the *Alyssum montanum-A. repens* complex in the Balkans: a multivariate morphometric analysis. *Plant Syst. Evol.* 303, 1413–1442. doi: 10.1007/s00606-017-1468-x

Strid, A. (2016). *Atlas of the Aegean Flora*. Berlin: Botanic Garden and Botanical Museum Berlin.

Suc, J. P. (1984). Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature* 307, 429–432. doi: 10.1038/307429a0

Templeton, A. R. (2008). The reality and importance of founder speciation in evolution. *Bioessays* 30, 470–479. doi: 10.1002/bies.20745

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

Thompson, J. D. (2020). *Plant Evolution in the Mediterranean: Insights for Conservation*, 2nd Edn. Oxford: Oxford University Press.

Wagner, F., Ott, T., Zimmer, C., Reichhart, V., Vogt, R., and Oberprieler, C. (2019). 'At the crossroads towards polyploidy': genomic divergence and extent of homoploid hybridization are drivers for the formation of the ox-eye daisy

polyploid complex (*Leucanthemum*, Compositae-Anthemideae). *New Phytol.* 223, 2039–2053. doi: 10.1111/nph.15784

White, T., Bruns, T., Lee, S., and Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protoc.* 18, 315–322. doi: 10.1016/B978-0-12-372180-8.50042-1

Zahradníček, J., Chrtek, J., Ferreira, M. Z., Krahulcová, A., and Fehrer, J. (2018). Genome size variation in the genus *Andryala* (Hieraciinae, Asteraceae). *Folia Geobot.* 53, 429–447. doi: 10.1007/s12224-018-9330-7

Záveská, E., Fér, T., Šída, O., Marhold, K., and Leong-Škorničková, J. (2016). Hybridization among distantly related species: examples from the polyploid genus *Curcuma* (Zingiberaceae). *Mol. Phylogen. Evol.* 100, 303–321. doi: 10.1016/j.ympev.2016.04.017

Zhou, T., Lu, L., Yang, G., Al-Shehbaz, I. A., and Dorofeyev, V. I. (2001). "Brassicaceae (Cruciferae)," in *Flora of China (Brassicaceae Through Saxifragaceae)*, Vol. 8, eds Z. Y. Wu and P. H. Raven (Beijing: Science Press), 1–193.

Zozomová-Lihová, J., Marhold, K., and Španiel, S. (2014). Taxonomy and evolutionary history of *Alyssum montanum* (Brassicaceae) and related taxa in southwestern Europe and Morocco: diversification driven by polyploidy, geographic and ecological isolation. *Taxon* 63, 562–591. doi: 10.12705/633.18

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

Zwickl, D. J. (2006). *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion*. Ph.D. thesis, University of Texas, Austin, TX.

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.

frontiers in Plant Science

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

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

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

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

OPEN ACCESS

Edited by:
Božo Frajman, University of Innsbruck, Austria

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

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

† These authors have contributed equally to this work

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

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

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

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

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

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

INTRODUCTION

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

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

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

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

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

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

MATERIALS AND METHODS

Study Species and Sampling

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

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

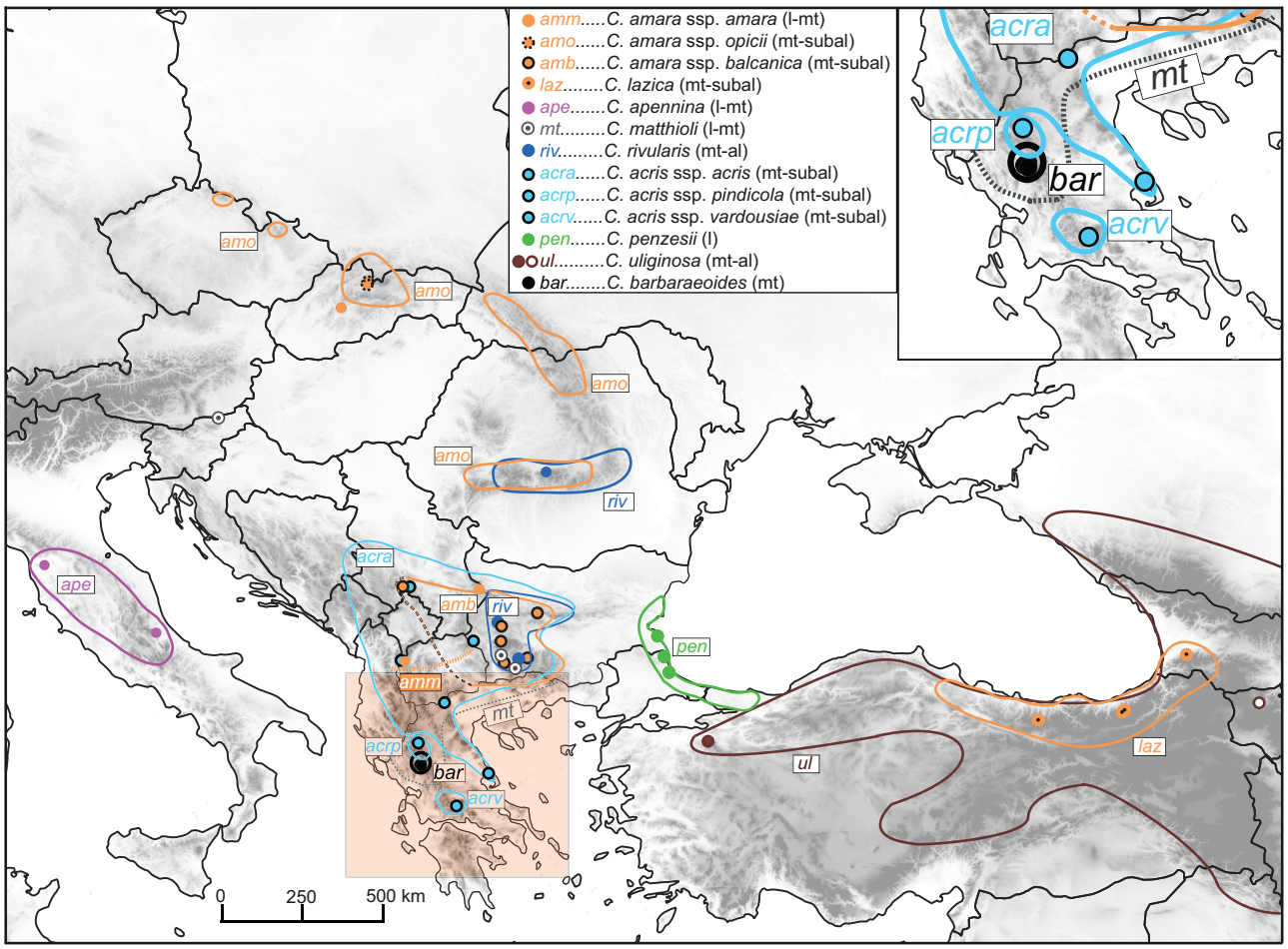


FIGURE 1 | Distribution of the *Cardamine* taxa under study, based on data compiled from floras, herbarium specimens, previous studies, and our own records. The western borders of the area of *Cardamine amara* subsp. *balcanica* remain unknown (marked here by a black–orange dashed line; see Tomović et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

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

Chromosome Counting and Flow Cytometry

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

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

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

PCR Amplification, Molecular Cloning, and Sanger Sequencing

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

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

Hyb-Seq Library Preparation

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

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

Hyb-Seq Data Processing and Phylogenomic Analyses

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

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

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

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

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

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

Extracting Allele Sequences and Identifying Homeologs Inherited From Different Parents

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

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

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

Gene Genealogy Interrogation Analyses

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

Analyses of nrDNA Sequence Data

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

Analyses of Chloroplast Genome Data

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

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

Genomic *in situ* Hybridization

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

RESULTS

Chromosome Numbers and Genome Size Variation

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

Hyb-Seq Data

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

Phylogenomic Analyses of Diploids Based on Target Nuclear Loci

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

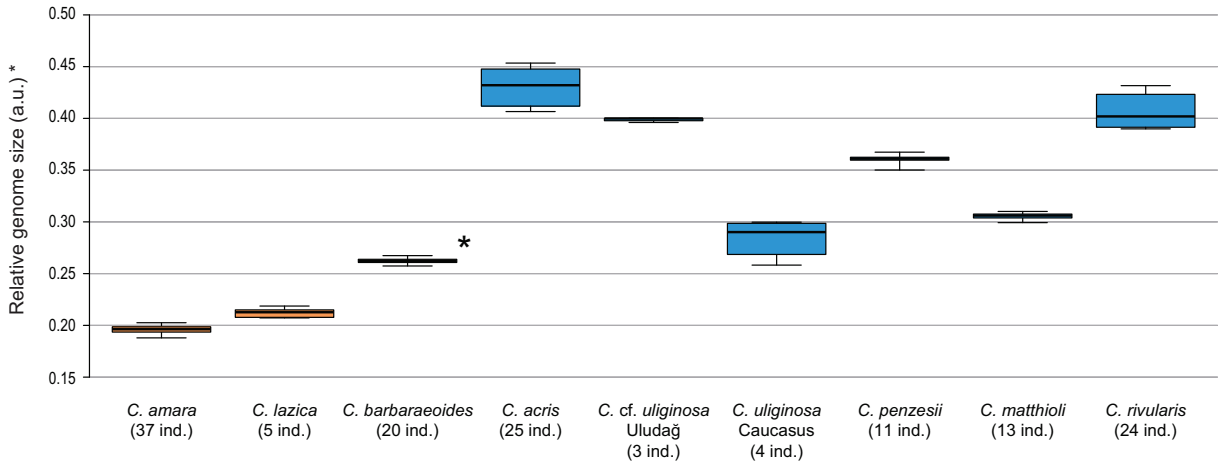


FIGURE 2 | Genome size variation of the *Cardamine* species under study, based on flow cytometric analyses. Relative nuclear DNA content inferred from DAPI measurements is presented, given as the ratio of the sample and standard G1 peaks (2C values in arbitrary units, a.u.). In the tetraploid *Cardamine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

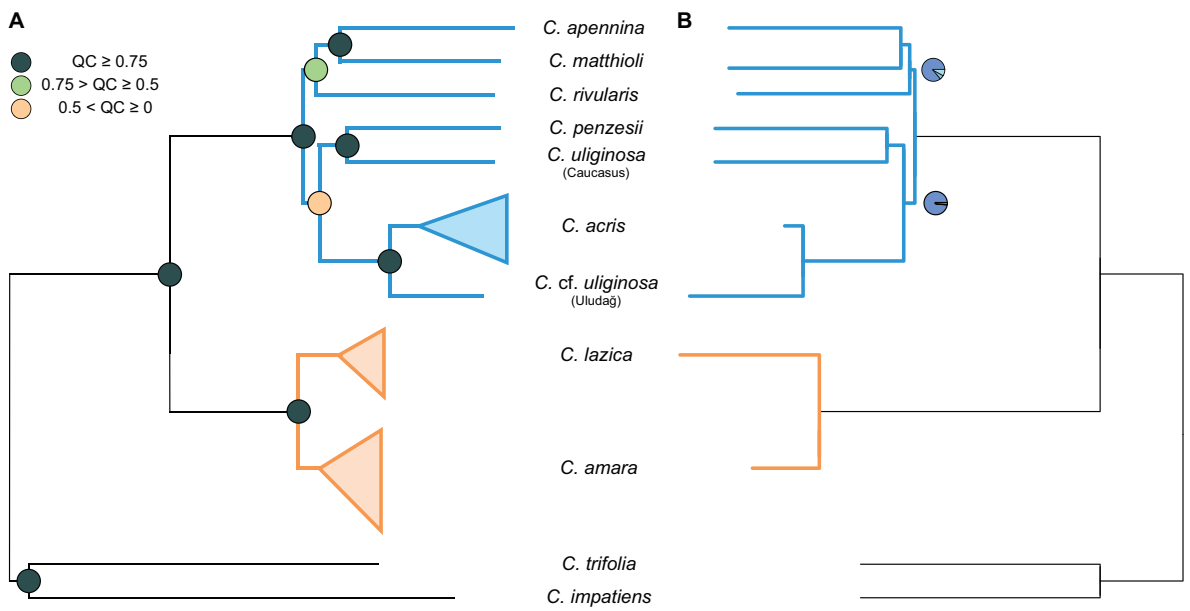


FIGURE 3 | Phylogenetic trees inferred from the complete dataset of 1,168 nuclear genes, based on consensus supercontig sequences of diploid *Cardamine* accessions. Orange color is used for the clades and branches of the *Cardamine amara* group, whereas blue is used for the remaining diploids resolved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

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

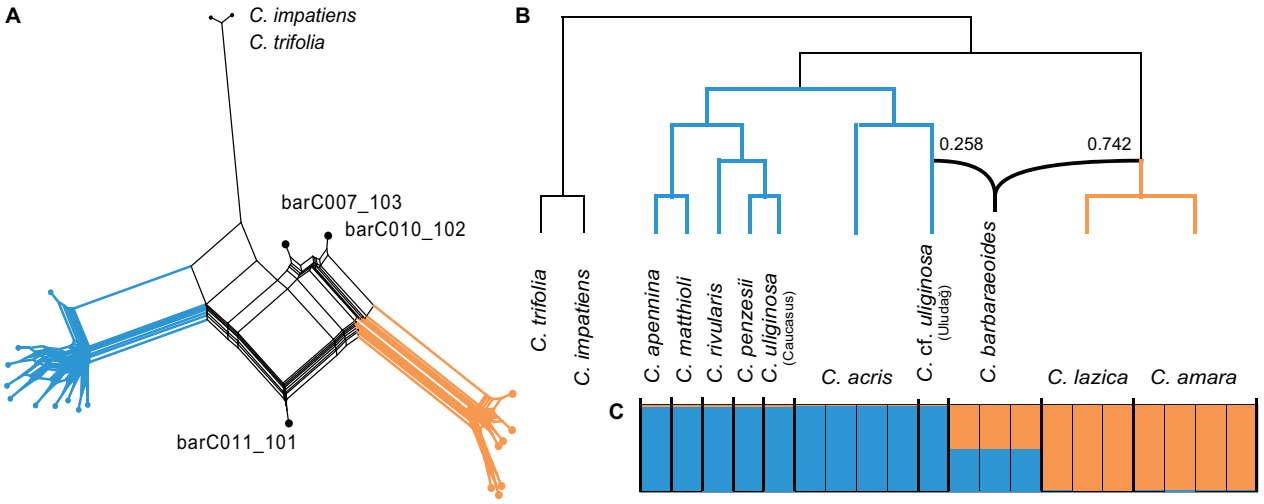


FIGURE 4 | Phylogenetic analyses indicating the hybrid (allopolyploid) origin of the tetraploid *Cardamine barbaraeoides*. Orange color is used for the accessions from the *Cardamine amara* group (*C. amara* and *Cardamine lazica*), blue for the remaining diploids, and black for the tetraploid *C. barbaraeoides*.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

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

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

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

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

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

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

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

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

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

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

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

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

Alternative Topology Testing: GGI Analyses

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

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

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

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

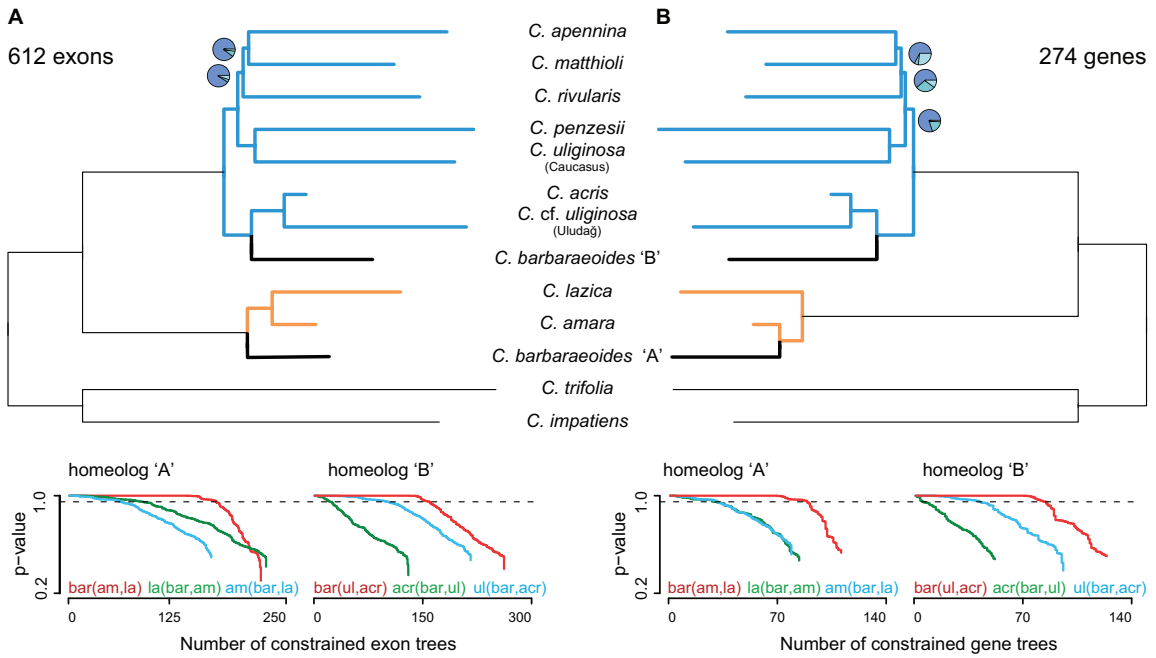


FIGURE 5 | Phylogenetic analyses based on phased allele sequences of 612 exons **(A)** or 274 genes **(B)**, for which alleles of all three accessions of *Cardamine barbaraeoides* were successfully phased and sorted into A and B homeologs. Species trees were inferred in ASTRAL-III. Branch support is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

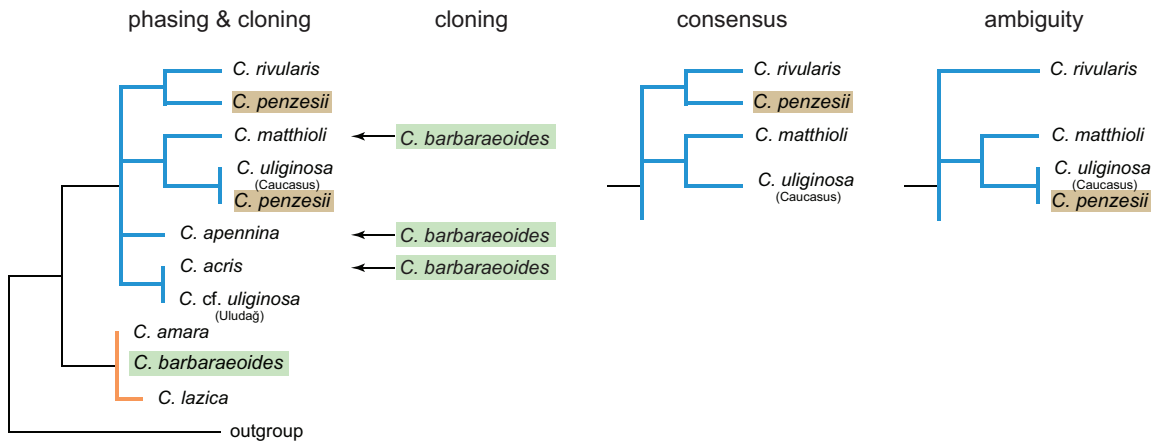


FIGURE 6 | Schematic representation of maximum likelihood (ML) trees inferred from four datasets of nrDNA (ITS1-5.8S-ITS2 region) and their topological differences. Cloning: dataset obtained from molecular cloning and Sanger sequencing; phasing: dataset obtained from read-backed phasing of Hyb-Seq reads,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

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

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

Analyses of Chloroplast Genome Data

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

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

Genomic *in situ* Hybridization

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

DISCUSSION

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

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

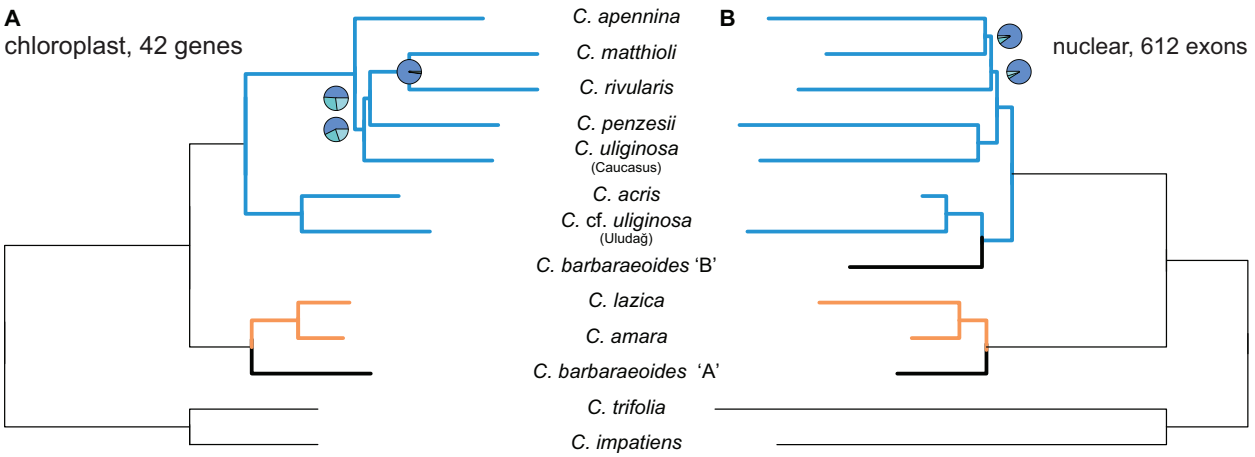


FIGURE 7 | Species trees inferred in ASTRAL-III from the dataset of 42 protein-coding most informative chloroplast genes **(A)** and 612 nuclear exons [**(B)**, as in **Figure 5A**]. Branch support is shown by pie charts, depicting three local posterior probabilities for the given branch (not shown for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

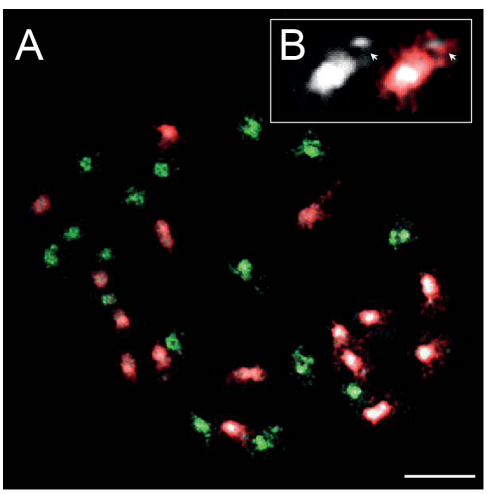


FIGURE 8 | Genomic *in situ* hybridization (GISH) on mitotic chromosomes in the allotetraploid *Cardamine barbaraeoides* (2*n* = 32). **(A)** GISH with total genomic DNA of *Cardamine acris* subsp. *acris* (red fluorescence) and *Cardamine amara* subsp. *balcanica* (green fluorescence) revealed two subgenomes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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

CONCLUSION

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

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

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

FUNDING

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

ACKNOWLEDGMENTS

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

SUPPLEMENTARY MATERIAL

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chen, H., German, D. A., Al-Shehbaz, I. A., Yue, J., and Sun, H. (2020). Phylogeny of Euclidieae (Brassicaceae) based on plastome and nuclear ribosomal DNA data. *Mol. Phylogenet. Evol.* 153:106940. doi: 10.1016/j.ympev.2020.106940

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Lihová, J., Shimizu, K. K., and Marhold, K. (2006). Allopolyploid origin of *Cardamine asarifolia* (Brassicaceae): incongruence between plastid and nuclear ribosomal DNA sequences solved by a single-copy nuclear gene. *Mol. Phylogenet. Evol.* 39, 759–786. doi: 10.1016/j.ympev.2006.01.027

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

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

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

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

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

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

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

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

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

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

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

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

Mandáková, T., Zozomová-Lihová, J., Kudoh, H., Zhao, Y., Lysak, M. A., and Marhold, K. (2019). The story of promiscuous crucifers: origin and genome evolution of an invasive species, *Cardamine occulta* (Brassicaceae), and its relatives. *Ann. Bot. (Oxford)* 124, 209–220. doi: 10.1093/aob/mcz019

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 25 August 2021
doi: 10.3389/fpls.2021.687094

Asymmetric Reproductive Barriers and Gene Flow Promote the Rise of a Stable Hybrid Zone in the Mediterranean High Mountain

Mohamed Abdelaziz[1*†], *A. Jesús Muñoz-Pajares*[1,2,3†], *Modesto Berbel*[1], *Ana García-Muñoz*[1], *José M. Gómez*[3,4] *and Francisco Perfectti*[1,3]

[1] Departamento de Genética, Facultad de Ciencias, Campus Fuentenueva, Universidad de Granada, Granada, Spain, [2] Laboratório Associado, Plant Biology, Research Centre in Biodiversity and Genetic Resources, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade Do Porto, Campus Agrário de Vairão, Fornelo e Vairão, Portugal, [3] Research Unit Modeling Nature, Universidad de Granada, Granada, Spain, [4] Departamento de Ecología Funcional y Evolutiva, Estación Experimental de Zonas Áridas, Consejo Superior de Investigaciones Científicas, Almería, Spain

OPEN ACCESS

Edited by:
Andrew A. Crowl,
Duke University, United States

Reviewed by:
Adrian Christopher Brennan,
Durham University, United Kingdom
Santiago Martín-Bravo,
Universidad Pablo de Olavide, Spain

***Correspondence:**
Mohamed Abdelaziz
mabdelazizm@ugr.es

†These authors have contributed equally to this work and share first authorship

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

Received: *28 March 2021*
Accepted: *15 July 2021*
Published: *25 August 2021*

Citation:
Abdelaziz M, Muñoz-Pajares AJ, Berbel M, García-Muñoz A, Gómez JM and Perfectti F (2021) Asymmetric Reproductive Barriers and Gene Flow Promote the Rise of a Stable Hybrid Zone in the Mediterranean High Mountain. Front. Plant Sci. 12:687094. doi: 10.3389/fpls.2021.687094

Hybrid zones have the potential to shed light on evolutionary processes driving adaptation and speciation. Secondary contact hybrid zones are particularly powerful natural systems for studying the interaction between divergent genomes to understand the mode and rate at which reproductive isolation accumulates during speciation. We have studied a total of 720 plants belonging to five populations from two *Erysimum* (Brassicaceae) species presenting a contact zone in the Sierra Nevada mountains (SE Spain). The plants were phenotyped in 2007 and 2017, and most of them were genotyped the first year using 10 microsatellite markers. Plants coming from natural populations were grown in a common garden to evaluate the reproductive barriers between both species by means of controlled crosses. All the plants used for the field and greenhouse study were characterized by measuring traits related to plant size and flower size. We estimated the genetic molecular variances, the genetic differentiation, and the genetic structure by means of the F-statistic and Bayesian inference. We also estimated the amount of recent gene flow between populations. We found a narrow unimodal hybrid zone where the hybrid genotypes appear to have been maintained by significant levels of a unidirectional gene flow coming from parental populations and from weak reproductive isolation between them. Hybrid plants exhibited intermediate or vigorous phenotypes depending on the analyzed trait. The phenotypic differences between the hybrid and the parental plants were highly coherent between the field and controlled cross experiments and through time. The highly coherent results obtained by combining field, experimental, and genetic data demonstrate the existence of a stable and narrow unimodal hybrid zone between *Erysimum mediohispanicum* and *Erysimum nevadense* at the high elevation of the Sierra Nevada mountains.

Keywords: hybridization, *Erysimum mediohispanicum*, *Erysimum nevadense*, Sierra Nevada, phenotype, reproductive isolation

INTRODUCTION

Evaluating the mechanisms promoting the rise of hybrid zones helps us to understand the nature and dynamics of these interesting evolutionary scenarios. Hybrid zones have been pointed to as one of the most insightful places to study the evolutionary interactions between divergent but related taxa (Barton and Hewitt, 1985; Harrison, 1993). Important effects of hybrid zones were well documented on, for example, morphological traits (Tastard et al., 2012, Keller et al., 2021), behavior (Good et al., 2000), selective regimes (Cruz et al., 2001; Knief et al., 2019), and biodiversity (Whitham et al., 1994, 1999). The hybrids can show a modified phenotype promoting their evolvability and, by altering the reproductive barriers between the parental species, modify the evolvability of these lasts (Parsons et al., 2011). However, an integrative study of the mechanisms, patterns, and consequences of hybridization is not straightforward.

Even before the first conceptual models to explain hybrid zones were proposed, gene transfer between the neighboring species had been recognized as an important evolutionary process (Dobzhansky, 1940; Mayr, 1942; Anderson, 1949). The main classifications of hybrid zones proposed relate to the distribution of the hybrid and the parental phenotypes/genotypes (Harrison and Bogdanowicz, 1997; Jiggins and Mallet, 2000). Therefore, hybrid zones can be classified as unimodal, bimodal, trimodal, and asymmetric (Pickup et al., 2019). Unimodal hybrid zones are characterized by a higher frequency of individuals with intermediate genotypes between both parental species (Jiggins and Mallet, 2000). Hybridization and admixture are predominant in this kind of hybrid zones, so weak reproductive barriers are expected in them (**Figure 1**). A special case of the unimodal hybrid zones appears when the hybrid zone forms a single panmictic population exhibiting individuals with variable degrees of genetic similarity to the parental forms, termed as "hybrid swarms" (Harrison and Bogdanowicz, 1997; Jiggins and Mallet, 2000). Hybrid swarms could present individuals coming from first or second hybrid generations or from different levels of back-crossing. In a bimodal hybrid zone, individuals belonging to the parental species co-occur with a very low frequency of their hybrids (Cruzan and Arnold, 1993; McMillan et al., 1997; Jiggins and Mallet, 2000; Vedenina and Helversen, 2003). This kind of hybrid zone appears when strong reproductive isolation barriers exist between the species previously (Harrison and Bogdanowicz, 1997; Jiggins and Mallet, 2000; Coyne and Orr, 2004; **Figure 1**). However, when reproductive isolation between taxa is only partial, a stable hybrid zone would arise, generating a trimodal hybrid zone that depends, on the one hand, on the differences in fitness between the parentals and the hybrids and, on the other hand, on the migration rates from the parental zones (Key, 1968; Barton and Hewitt, 1985; Jiggins and Mallet, 2000; **Figure 1**). Finally, asymmetric hybrid zones occur when strong asymmetries in gene flow and barrier strength promote hybridization and introgression in one direction, such that gene flow occurs only from one species to another (Pickup et al., 2019; **Figure 1**).

Hybridization is much more frequent and is evolutionarily relevant in plants than in animals, but surprisingly, in the past, it has been more explored in the latter (Ellstrand et al., 1996; Dowling and Secor, 1997). The number of plant hybrid zones studied using an integrative approach, including genetic analyses and the nature of reproductive isolation, is limited (Abbott, 2017). However, interesting examples of contact zones and hybrid populations have been described in plants, for example, those described in *Phlox* (Levin, 1967), *Iris* (Cruzan and Arnold, 1993; Young, 1996; Emms and Arnold, 1997), *Helianthus* (Rieseberg et al., 1998), or *Armeria* (Fuertes Aguilar et al., 1999) and, more recently, those described in *Ipomopsis* (Campbell and Aldridge, 2006), *Silene* (Minder et al., 2007), *Narcissus* (Marques et al., 2010), or *Primula* (Keller et al., 2021), among others. These studies have demonstrated how the rise of hybrids can not only generate new phenotypic values in the contact zone but also modify the parental phenotype through introgression (Owens et al., 2016; Nieto Feliner et al., 2019) or by affecting the selective trajectories in the contact zone and the surrounding areas (Wielstra et al., 2017; Wielstra, 2019). Therefore, it is interesting to explore the phenotypic differences between the parental plants and hybrids not only in natural populations but also in controlled conditions where we could establish the origin and the level of admixture of the hybrid plants. In this sense, the hybrids can show different patterns of phenotypic differentiation. The hybrid phenotype could exhibit higher, lower, or intermediate phenotypic values compared with the parental plants (**Figure 1C**). The hybrid can also show an asymmetric pattern when the hybrids on one parental direction (that is, the parental direction acting as the mother) exhibit significant differences with the hybrid in the other sense (**Figure 1C**). Exploring this question through time in natural populations would allow us to know more about the stability of hybrids and the hybrid zones. However, the longer the life cycle of the organism being experimented on, the more difficult exploring becomes.

Incomplete reproductive barriers (or absence of them) are necessary to the occurrence of hybrids in a contact zone between different species. So, the evaluation of the strength of reproductive barrier components between the hybridizing species sheds light on the mechanisms underlying the rise of hybrid zones. Again, this objective is not always easily reachable, especially when the study systems demand several to many years for blooming. Reproductive barriers are usually estimated by the use of one or a few components of fitness. However, quantifying the different fitness components through the entire life cycle of plants will allow for a more realistic estimation of the intensity of these barriers and their cumulative effects on gene flow between the parental species and hybrids (Baack et al., 2015).

In this study, we examined two plant species belonging to the genus *Erysimum* L., *Erysimum mediohispanicum*, and *Erysimum nevadense*, and the hybrids produced between them in the Sierra Nevada (Southeast Spain). Genus *Erysimum* L. (Brassicaceae) is mainly distributed in the northern hemisphere (Polatschek, 1986), presenting an important diversification center in the western Mediterranean region (Greuter et al., 1986). The genus presents a complex evolutionary history due to events of interspecific hybridization and polyploidization, producing a highly diversified genus (more than 200 species) enriched in species complexes and cryptic species (Clot, 1992;

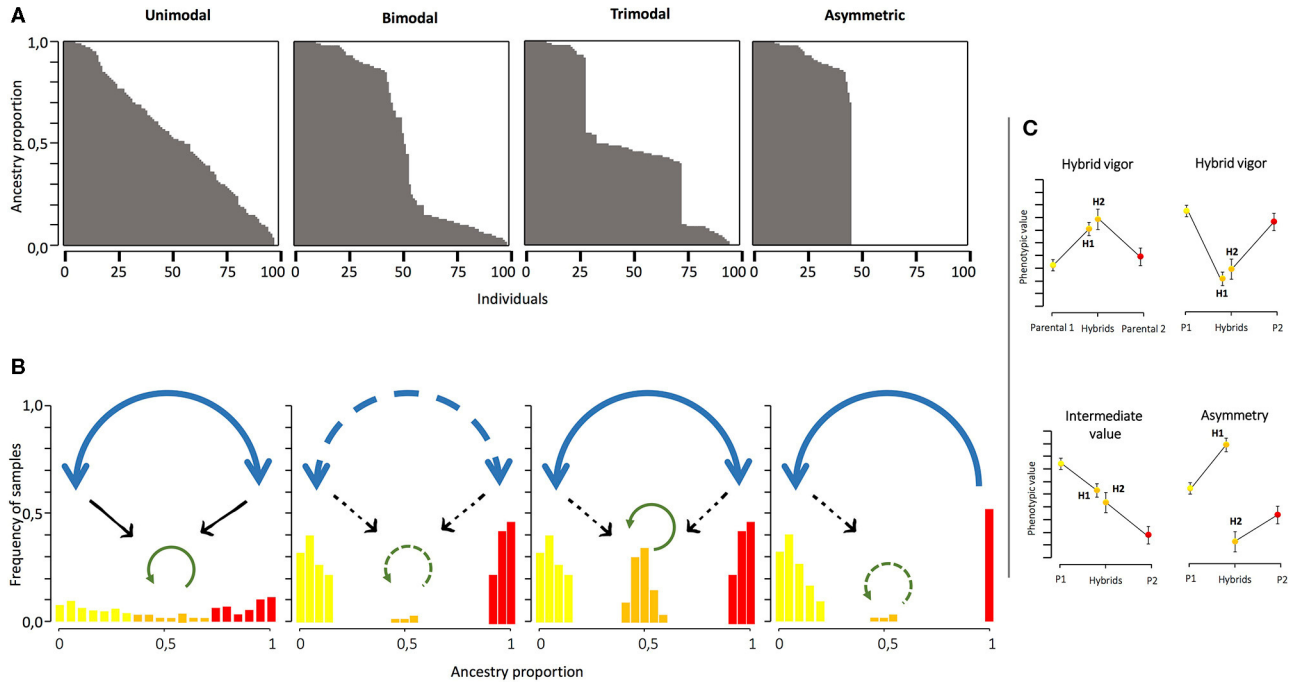


FIGURE 1 | (A) Different models of hybrid zones based on the different patterns of the genetic mixture between divergent taxa. The abscissa axis shows individuals in the hybrid zone and the ordinate axis shows the ancestry proportion exhibited by individuals. **(B)** Frequency of each range of ancestry proportion in the population and the gene flow expected in each situation. Blue, green, and black arrows represent the gene flow promoted by interspecific crosses, backcrosses, and crosses between hybrids, respectively. The persistence of gene flow is indicated by solid lines; the absence of gene flow is indicated by broken lines. **(C)** Patterns of the phenotypic differentiation expected when differences between hybrids and parental plants are found. H1 and H2 are the hybrids obtained by adding pollen on parental plants 1 and 2, respectively. Yellow and red dots and bars correspond to the parental species, while orange represents hybrid values.

Ancev, 2006; Marhold and Lihová, 2006; Turner, 2006; Couvreur et al., 2010, Abdelaziz et al., 2011; Al-Shebaz, 2012). In the Sierra Nevada, *E. mediohispanicum* and *E. nevadense* show contrasting distributions, inhabiting the lowland and the top of the mountains, respectively (Blanca et al., 2009). However, these species make what could be a secondary contact (Abdelaziz et al., 2014) at mid-elevation.

In the present study, we used an integrative approach by combining fieldwork, lab experiments, and molecular data to evaluate the nature and the strength of barriers to hybridization, to characterize the hybrid zone between *E. mediohispanicum* and *E. nevadense* in the Sierra Nevada National Park, and to explore the consequences of this geographic contact on plant phenotypes. The main goals of this study are (1) to explore the genetic and phenotypic differences between both species and their hybrids in the contact zone; (2) to explore the gene flow patterns promoting hybridization in this area; (3) to analyze the stability of the contact zone through time; and (4) to analyze the reproductive barriers promoting or constraining the rise of hybrids between *E. mediohispanicum* and *E. nevadense*.

MATERIALS AND METHODS

Study System

Erysimum mediohispanicum Polatschek is an endemic species of the Iberian Peninsula, where it is distributed in two widespread and almost disconnected regions, located in the North and Southeast of the peninsula. Its life cycle varies among individuals and populations, usually being monocarpic. In the Sierra Nevada (Southeast of Spain), *E. mediohispanicum* is composed of diploid populations (Muñoz-Pajares et al., 2018) of facultative biennial, spending 2–3 years growing at an elevation of up to 2,200 m like a vegetative rosette on calcareous soils (Muñoz-Pajares et al., 2020). After that period, plants display flowers on one to three long stalks, ranging in number from only a few to several hundred (Gómez, 2003). These flowers are visited by a highly diverse assemblage of insects (Gómez et al., 2007; Gomez et al., 2014).

Erysimum nevadense Reut. is also a diploid species (author's unpublished data) and mostly polycarpic herb endemic to the peaks of the Sierra Nevada mountains, where it grows on siliceous soils at an elevation from 2,300 to 2,700-m (Blanca et al., 2009). *E. nevadense* spends 2–3 years like a rosette before displaying anything from a few to several hundred flowers on various short stalks. It is also a pollination-generalist plant, but it does not present a pollinator assemblage as diverse as that of *E. mediohispanicum*, probably due to the harsh conditions of its habitat (Gómez et al., 2007; Mohamed, 2013).

Both species come into contact along a narrow area at an elevation of approximately 2,200 m. In this contact zone, populations from each species are located at a mere hundred meters apart, generating an area where the phenotypic traits defining each species are difficult to identify. In 2007, we

established a line transect of 3.1 km between these species in the north face of the Sierra Nevada encompassing five populations (**Table 1**), two *E. mediohispanicum* populations (Em25 and Em17), two *E. nevadense* populations (En11 and En10), and a putative hybrid population (H01), where both species contact. After 10 years, in 2017, we phenotypically resampled the hybrid population (H01) and the two parental populations more genetically related to the hybrid one (Em25 and En10).

Plant Phenotyping

For about a month from the end of May in 2007, at each of these populations, we phenotyped 90 plants by measuring the following phenotypic traits: (1) number of flowers and (2) number of stalks: counting the total number of flowers and stalks produced by each plant, respectively; (3) stalk height: the height of the tallest stalk of the plant from the ground to the top of the stalk at the end of the flowering period, when no more flowers are expected on the stalk, using a measuring tape (±0.5 mm error); (4) stalk diameter: the basal diameter of the tallest stalk; (5) corolla diameter: the distance between the edge of the two opposite petals; (6) corolla tube width: the diameter of the corolla tube aperture as the distance between the bases of two opposite petals; and (7) corolla tube length: the distance between the corolla tube aperture and the base of the sepals. The stalk diameter and the traits related to the flower size (4–7, above) were quantified in millimeters using a digital caliper (±0.01 mm. error). On approximately the same dates in 2017, all these phenotypic traits were remeasured in the same number of plants in the populations Em25, H01, and En10 (**Table 1**).

For each population and sampled year, we calculated the product–moment correlations, the covariances between each couple of the phenotyped traits, and the variance per trait (**Supplementary Tables 1, 2**). Kruskall–Wallis tests and one-way ANOVAs were used to compare the trait distributions and means of the populations for each phenotypic character, respectively. These analyses were performed using the package stats in R (R Development Core Team, 2014).

DNA Isolation and Genotyping

In 2007, we collected fresh tissues from all the phenotyped plants at the putative hybrid area and also from randomly selected 30 plants out of the 90 phenotyped plants from each of the four parental populations in the transect (in total, 210 plants). The tissue was stored in silica gel for the subsequent DNA isolation, using the GenElute Plant Genomic DNA Miniprep kit (Sigma-Aldrich, St. Louis, MO, USA). The isolated DNA was used for the individual genotyping, carried out with 10 microsatellite markers previously described by Muñoz-Pajares et al. (2011). We performed PCR in a 15-μL reaction mixture containing 0.17 ng/μL of the template genomic DNA, 1X buffer (ref. M0273S, New England BioLabs), 0.16 mM each of dNTP (Sigma-Aldrich), 0.33 μM each of forward (fluorescently tagged) and reverse primers, and 0.02 U/μL *Taq* polymerase (ref. M0273S, New England Biolabs). PCR was conducted in a Gradient Master Cycler Pro S (Eppendorf, Hamburg, Germany) with an initial step of 30 s of denaturation at 94°C followed by 35 cycles at 94°C for 15 s, annealing temperatures for each microsatellite marker described in Muñoz-Pajares et al. (2011) for 30 s, extension at 72°C for 30 s, and a final extension step at 72°C for 3 min. PCR products were diluted at a ratio of 1:15 and sent to MACROGEN (Geumchun-gu, Seoul, South Korea; http://www.macrogen.com) for capillary electrophoresis using 400HD ROX as standard. Alleles were called using the Peak Scanner Software version 1.0 (Applied Biosystems).

Genetic Data Analyses

Once we established the multilocus genotypes, we excluded the individuals presenting levels of missing data higher than 30% from the genetic analyses. This included 3 individuals from the En11 population and 10 individuals from H01 (**Table 2**). To genetically characterize the populations, we estimated the following parameters: (a) number of non-redundant multilocus genotypes (N_G), as the number of genotypes showing at least one different allele, excluding missing data; (b) mean number of alleles per locus (n_a); (c) observed heterozygosity (Ho), as the actual frequency of heterozygous individuals in the sample. We estimated the mean of the individual heterozygosities per population using the ratio between the number of heterozygote loci and the number of successfully genotyped loci; (d) gene diversity (Hs), as the expected proportion of heterozygous individuals assuming the Hardy–Weinberg equilibrium. Gene diversity was calculated using the Nei (1987) estimator; (e) mean allelic richness per locus (Rs), estimated as the probability of sampling the allele i at least once among the 2n genes of a sample, being independent of sample size; (f) the mean private allelic richness (R_P), estimated as the mean number of singular alleles per locus presented at each population. All the previous parameters were calculated using the package hierfstat v. 0.04–6 (Goudet, 2005) or using scripts developed by ourselves, both in R (R Development Core Team, 2014). (g) Inbreeding coefficient (F_{IS}), which provides information about the Hardy–Weinberg equilibrium departures due to either excess or defect of heterozygotes. We estimated F_{IS} by Bayesian inference using BayesAss v3.0 (Wilson and Rannala, 2003) for each population and overall for the two studied species. Analysis lasted for 10 million MCMC iterations, sampling every 1,000 generations and optimizing the mixing parameter for allele frequencies and for inbreeding coefficients. After that, we removed the first 10% of total iterations and checked the trace files with the program Tracer v1.4 (Rambaut and Drummond, 2007) to determine the convergence of the independent Bayesian MCMC runs.

Microsatellite-based genetic differentiation among groups of populations belonging to the same species, among populations within groups, and among individuals within populations was estimated using a hierarchical analysis of molecular variance (AMOVA), as implemented in Arlequin (Excoffier and Lischer, 2010), using 1,000 permutations to test the significance. We performed this analysis two times: the first one excluding the hybrid population and the second one including the hybrid population in its own group. In addition, pairwise comparisons for genetic differentiation between populations were computed using the package hierfstat in R (R Development Core Team, 2014) by the computation of FST statistics (Weir and Cockerham, 1984) and D_{ST} (Jost, 2008). F_{ST} significance was calculated

TABLE 1 | Geographic information and sampling effort in the five populations included in the study (ordered by elevation).

Population characteristics				Sampling effort				
				2007		2009–2010		2017
Population	Latitude	Longitude	Elevation	Phenotyped plants	Genotyped plants	Transplanted plants	Crossed plants	Phenotyped plants
Em25	37°7.230' N	3° 26.082' W	2,064	90	30	30	7	90
Em17	37° 6.698' N	3° 25.450' W	2,182	90	30	30	23	–
H01	37°6.908' N	3° 25.250' W	2,200	90	80	–	–	90
En11	37°6.750' N	3° 25.048' W	2,222	90	27	30	10	–
En10	37°6.658' N	3° 24.301' W	2,322	90	30	30	4	90

from 1,000 permutations, while the DST was estimated as the harmonic mean across loci.

The genetic relationship among genotypes was inferred by Bayesian means using the model-based clustering algorithm, as implemented in Structure v.2.2 (Pritchard et al., 2000; Falush et al., 2003). The number of multilocus genotype clusters (K) was analyzed using diploid setting and using admixture and *prior information* as ancestry models and correlations as the allele frequency model using prior information. We performed simulations with 10 replicates for each K value, ranging from K = 1 to K = 6. Each run consisted of 50,000 Markov Chain Monte Carlo (MCMC) steps after 20,000 burn-in steps. To detect the optimum value of K, we applied the Evanno method (Evanno et al., 2005), as implemented in the Structure Harvester website (Earl and von Holdt, 2012).

Finally, we estimated the gene flow rates among the studied populations and their significance by means of Bayesian inference using BayesAss v3.0 (Wilson and Rannala, 2003). Analysis lasted for 10 million MCMC iterations, with a sampling frequency of every 1,000 generations, optimizing the mixing parameter for allele frequencies and for inbreeding coefficients. After that, we removed the first 10% of the total iterations and checked the trace files with the program Tracer v1.4 (Rambaut and Drummond, 2007) to determine the convergence of the independent Bayesian MCMC runs.

Reproductive Barrier Evaluation

In September 2009, we collected 120 individual plants from the four parental populations (30 plants per population of *E. mediohispanicum* and *E. nevadense*; **Table 1**). These individuals were transplanted to individual pots (11 cm x 11 cm x 15 cm) using the same soil in which they were growing and moved to a common garden in the University of Granada (an elevation of ∼700-m). By May 2010, 44 individuals had survived (30 *E. mediohispanicum* plants and 14 *E. nevadense* plants; **Table 1**). Before the plants started blooming, they were moved to a greenhouse to exclude them from pollinators. During flowering, they were phenotyped following the abovementioned methodology used for the field plants, and several additional flowers per individual were subjected to two different treatments: (a) outcrossing (OC), in which some flowers were emasculated before opening and were pollinated with pollens from a different conspecific individual; and (b) interspecific hybridization (HC), in which some flowers were emasculated before opening and were pollinated with pollens from an individual from the other species. In total, 232 flowers were used in the experiment, with a mean of 5.5 ± 3 flowers per plant (3.2 ± 1.9 for OC and 2.8 ± 1.7 for HC per plant).

Once the blooming period was over, we recorded the number of flowers per plant and the treatment setting as ripe fruits or aborting without producing any fruits. The total number of ovules, unfertilized ovules, aborted seeds, and ripe seeds produced per ripe fruit were recorded in the lab using magnifying glasses. A total of 3,139 seeds were harvested at the end of the experiment. Subsequently, when possible, 15 seeds per plant and treatment were taken at random and sown randomly in a greenhouse. Their germination was recorded two times a week for the first month, and their survival was recorded every month during the next 10 months.

The following predispersal and postdispersal components of the plant reproductive output were quantified for each treatment and plant: (a) fruit set, the proportion of labeled flowers setting fruit; (b) seed production, the number of seeds produced per ovule in a given fruit; (c) seedling emergence, the proportion of sown seeds germinating and emerging as seedlings; and (d) seedling survival, the proportion of seedlings surviving until the end of the experiment. Afterward, we calculated the cumulative pre-dispersal fitness (W_{PRE}), as fruit set x seed production, and the cumulative total fitness (W_{TOT}), as fruit set x seed production x seedling emergence x seedling survival. The F1 generation resulting from OC and HC was grown until blooming (in the spring of 2011) when it was also phenotyped as we did with the natural populations.

Hybrid inviability (HI) was calculated per population by comparing the fitness between the intraspecific and interspecific crosses using the Ågren and Schemske (1993) approach as

$$HI = 1 - w_h/w_o;\ w_h < w_o \quad (1)$$

$$HI = w_h/w_o - 1;\ w_h > w_o \quad (2)$$

where w_o is the fitness of the intraspecific outcrossing treatments and w_h is the fitness of the hybrid crosses. In all cases, we used both predispersal and total fitness. Using this approach, the values range between −1 and +1. Positive values indicate that the hybrid crosses have lower fitness than intraspecific crosses (occurrence of hybrid inviability and hence the rise of reproductive barriers). The significant values of HI were calculated by computing 95% CIs by means of bootstrapping with

1,000 permutations, using package boot in R (Canty and Ripley, 2017).

RESULTS

Phenotypic Differentiation

Using one-way ANOVAs, we found significant differences at the among-population level in the 2007 sample for all the measured phenotypic traits, except for the corolla tube width (**Figure 2A**). The hybrid population showed the highest values for the number of flowering stalks and the number of flowers, both traits related to plant size. However, it showed intermediate values for the rest of the traits, including characters related to plant size (stalk diameter and plant height) and to flower size (**Figure 2A**). The same pattern was found again when we resampled three out of the original populations in 2017 (Em25, H01, and En10; **Table 1** and **Figure 2C**).

The analysis of phenotypic traits performed on the offspring (F1 generation) of the controlled crosses confirmed these trends but also showed different patterns depending on the direction of gene flow producing the hybrid plant. The plants resulting from the hybrid crosses performed on *E. mediohispanicum*, i.e., *E. mediohispanicum* acting as the female, showed significantly higher values than the parental species for the number of flowering stalks and the number of flowers, while the hybrid plants resulting from crosses on *E. nevadense* showed lower values for these same traits (**Figure 2B**). For the rest of the phenotypic traits, the hybrid plants presented intermediate values, not significantly different from the values shown by the parentals for stalk diameter, flower diameter, and flower tube width (**Figure 2B**).

We found significant correlations in every population between the phenotypic traits defining the plant size (number of stalks, stalk diameter, plant height, and the number of flowers) and between the phenotypic traits defining the flower size (corolla diameter, corolla tube length, and corolla tube width) (**Supplementary Table 1**). In some cases, we also found significant positive correlations between the plant-size traits and the flower-size traits. This means that larger plants can also produce larger flowers, but larger plants do not present smaller flowers in any case. This pattern was found for the plants measured in the natural populations in 2007 (**Supplementary Table 1**) and 2017 (**Supplementary Table 2**), as well as in the plants obtained by the controlled crosses. The plants obtained by the intraspecific crosses presented similar number and levels of correlations to the plants phenotyped in natural populations. The experimental within-species cross plants showed similar patterns to those found in plants produced from hybrid crosses, but *E. nevadense* hybrids, where the number of traits significantly correlated, decreased (**Supplementary Table 3**).

Genetic Diversity and Differentiation

A total of 197 plants were successfully genotyped from the five sampled populations in 2007 (**Table 2**). A total of 196 multilocus genotypes were identified from them, finding no identical multilocus genotypes among populations, but only one repeated multilocus genotype at population En11 (**Table 2**).

The mean number of alleles per locus (n_a) ranged between 6.0 and 9.9, corresponding to En11 and H01, respectively, and *E. mediohispanicum* presented slightly higher values than *E. nevadense* (**Table 2**). The observed H_O presented values from 0.50 to 0.63, corresponding to these extremes to En11 and Em25, respectively, and again *E. mediohispanicum* values being higher. Moreover, the mean H_S presented the same pattern shown by H_O but the values of H_S were higher than *Ho* in all the considered populations and species (**Table 2**). The values of the mean allelic richness (R_S) ranged from the minimum value (5.87) at En11 to the maximum (7.54) at H01, presenting higher values for *E. mediohispanicum* populations (**Table 2**). The pattern exhibited by the mean private allelic richness (R_P) was different, being higher for *E. nevadense* and presenting their minimum value at Em17 (0.10) and the maximum value at H01 (1.20). Finally, the inbreeding coefficient (F_{IS}) exhibited its minimum value at the hybrid population (0.139) and the maximum values at both extremes of the transect (Em25 = 0.363 and En10 = 0.360), although the SEs associated with these maximum values were also higher (**Table 2**).

Hierarchical analysis of molecular variance (AMOVA) indicated that population groups (species) were not genetically differentiated for these markers (**Table 3**). Moreover, this also occurred when H01 was included as a separate group. In contrast, the among-population-within–groups level (both including and excluding H01 in the analysis) was significantly differentiated, although it accounted for only low amounts of genetic variance (**Table 3**). The genetic variance at this level decreased when the hybrid population was included (**Table 3**). The within-population level accounted for almost all genetic variance in both cases (**Table 3**).

Pairwise F_{ST} comparisons between populations showed that values ranging from a maximum differentiation of 0.0368 between Em17 and En10 to a minimum differentiation value of 0.0107 exhibited between Em25 and H01 (**Table 4**). Despite the low pairwise F_{ST} values found in our transect, all of them were significant after Bonferroni correction (**Table 4**). In addition, pairwise D_{ST} comparisons between populations indicated a similar pattern of genetic differentiation, with the higher D_{ST} value between Em25 and Em17 and between Em17 and En10 and the lower one between Em25 and H01 (**Table 4**).

Genetic Structure and Gene Flow

The Bayesian inference of the genetic structure assigned the studied populations to two genetic clusters (K = 2), the second-most probable model being the one considering four genetic clusters (K = 4) (**Supplementary Table 4**). Considering K = 2, the plants belonging to the populations Em25, En10, and H01 exhibited very high membership proportions to a given ancestral genetic cluster, while the plants belonging to Em17 and En11 showed higher or medium values of membership proportions to the second cluster (**Figure 3**). However, when we consider the model assuming K = 4, the plants from *E. mediohispanicum* and *E. nevadense* populations showed medium to high membership proportions to four ancestral genetic clusters (**Figure 3**). In

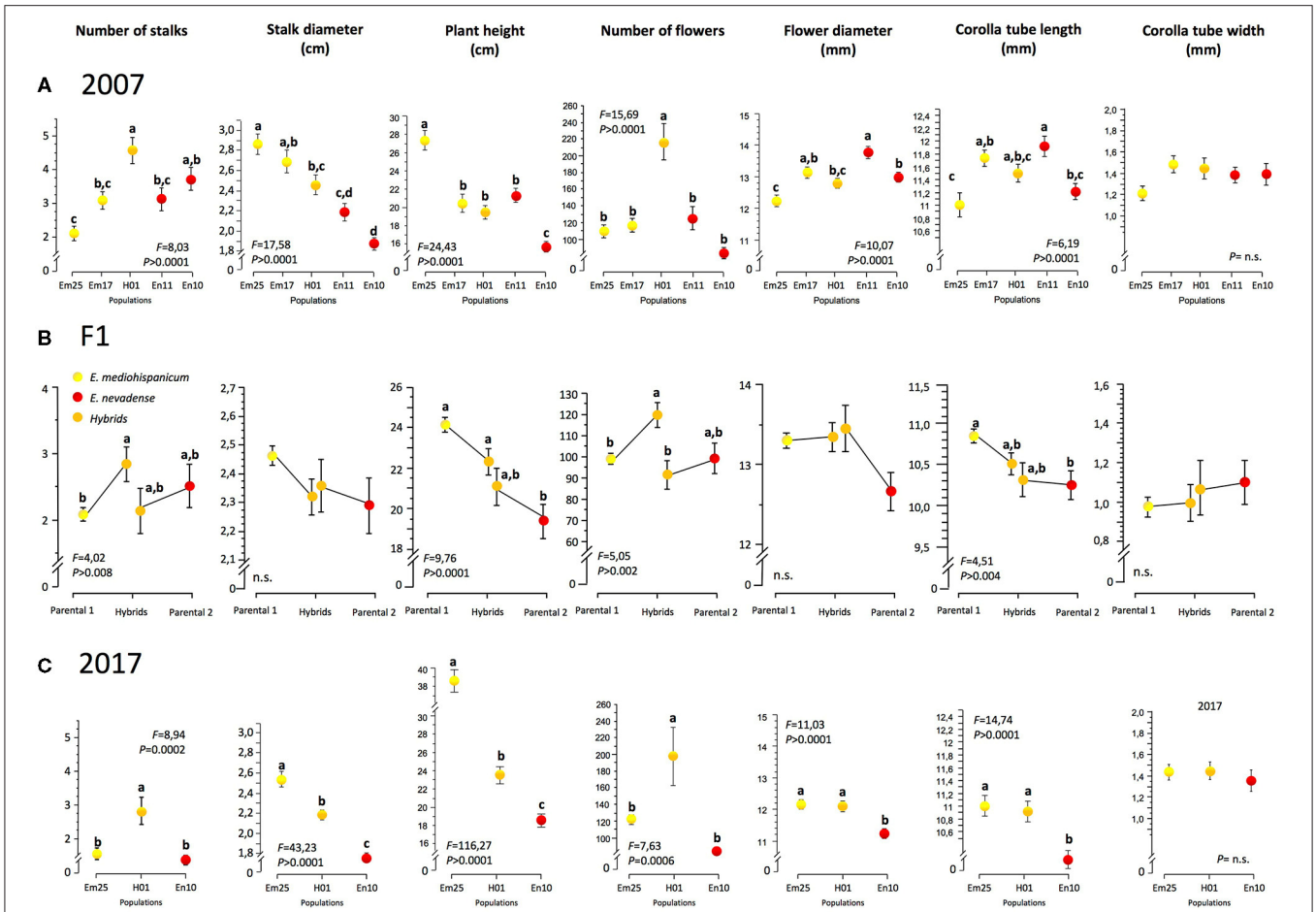


FIGURE 2 | Values of phenotypic traits at each population in 2007 **(A)**, 2017 **(C)**, and the hybrid and parental plants obtained by controlled cross experiments **(B)**. Yellow dots denote plants or populations of *Erysimum mediohispanicum*, orange dots denote the hybrid plants or population, and red dots denote plants or populations of *Erysimum nevadense*. F-ratios refer to one-way ANOVAs. Letters indicate the groups of plants where the differences are significant, according to Tukey's HSD comparison.

TABLE 2 | Genetic diversity parameters and effective population sizes for *Erysimum mediohispanicum* (Em), E*rysimum nevadense* (En), and the hybrid populations (total $N = 197$).

Population	*N*	N_G	n_a	H_O	H_S	R_S	R_P	F_{IS}
Em25	30	30	7.8 ± 1.31	0.63 ± 0.04	0.70 ± 0.06	7.39 ± 1.30	0.50 ± 0.22	0.363 ± 0.248
Em17	30	30	7 ± 1.12	0.54 ± 0.05	0.67 ± 0.07	6.58 ± 0.97	0.10 ± 0.10	0.213 ± 0.045
H01	80	80	9.9 ± 1.57	0.61 ± 0.05	0.70 ± 0.06	7.54 ± 1.12	1.20 ± 0.36	0.139 ± 0.019
En11	27	26	6 ± 1.16	0.50 ± 0.07	0.64 ± 0.06	5.87 ± 1.09	0.20 ± 0.13	0.298 ± 0.068
En10	30	30	7.8 ± 1.57	0.57 ± 0.07	0.68 ± 0.06	7.06 ± 1.34	0.70 ± 0.33	0.360 ± 0.249
Total Em	60	60	7.4 ± 0.85	0.59 ± 0.03	0.68 ± 0.04	6.98 ± 0.79	0.30 ± 0.13	–
Total En	57	56	6.9 ± 0.97	0.53 ± 0.05	0.66 ± 0.04	6.47 ± 0.85	0.45 ± 0.18	–

For each population, the parameters include the number of genotyped plants (N) with 10 nuclear microsatellite loci, the number of multilocus genotypes (N_G), the mean number of alleles per locus (n_a), the mean observed heterozygosity (H_O), the mean gene diversity (H_S), the mean allelic richness (R_S), the mean private allelic richness (R_P), and inbreeding coefficient (F_{IS}). ±SD values are indicated. Mean (±SD) values for each population and species are also given.

contrast, the individuals from the hybrid population exhibited medium to high values for their assignment probabilities to the most frequent genetic clusters in Em25 and En10 (**Figure 3**).

The Bayesian inference indicated that the highest gene flow levels were taking place at the intrapopulational level for all the populations included in the study (**Table 5**). The gene flow among populations was not significant, except for the gene flow from Em25, En11, and En10 to the hybrid population (**Table 5**), the gene flow being higher from Em25 (29%) and En10 (28%) to H01 than that from En11 (14%).

TABLE 3 | Hierarchical analysis of molecular variance (AMOVA).

	A: [Em25–Em17][En11–En10]				B: [Em25–Em17]–H01-[En11–En10]			
Source of variation	***d.f.***	**Variation (%)**	**Fixation index**	***P*-Value**	***d.f.***	**Variation (%)**	**Fixation index**	***P*-Value**
Among groups	1	0.000	0.000	0.69306	2	0.000	0.000	0.93939
Among populations within groups	2	2.31	0.023	0.00489	2	1.48	0.015	0.00782
Within populations	230	97.69	0.034	0.00391	389	98.52	0.033	0.00293
Total	233				393			

(A) Two groups (E. mediohispanicum; E. nevadense) were established, excluding the H01 population. (B) Three groups were established (E. mediohispanicum populations; E. nevadense populations; hybrid population). The degrees of freedom (d.f.), percentage of variation in each hierarchical level, and their respective fixation indexes with their p-values are given.

TABLE 4 | Pairwise F_{ST} (above diagonal) and D_{ST} (below diagonal) values along the transect, estimated by means of Weir and Cockerham methods 1984 and as the harmonic mean across loci, respectively (Weir and Cockerham, 1984).

	Em25	**Em17**	**H01**	**En11**	**En10**
Em25		0.0321**	0.0107*	0.0250*	0.0325**
Em17	0.055		0.0228**	0.0265**	0.0368**
H01	0.013	0.038		0.0179**	0.0128**
En11	0.043	0.042	0.032		0.0267**
En10	0.032	0.054	0.014	0.037	

The significance for F_{ST} values was calculated from 1,000 permutations and the p-values resulted after the Bonferroni correction: $^{*}P < 0.05$, $^{**}P < 0.01$.

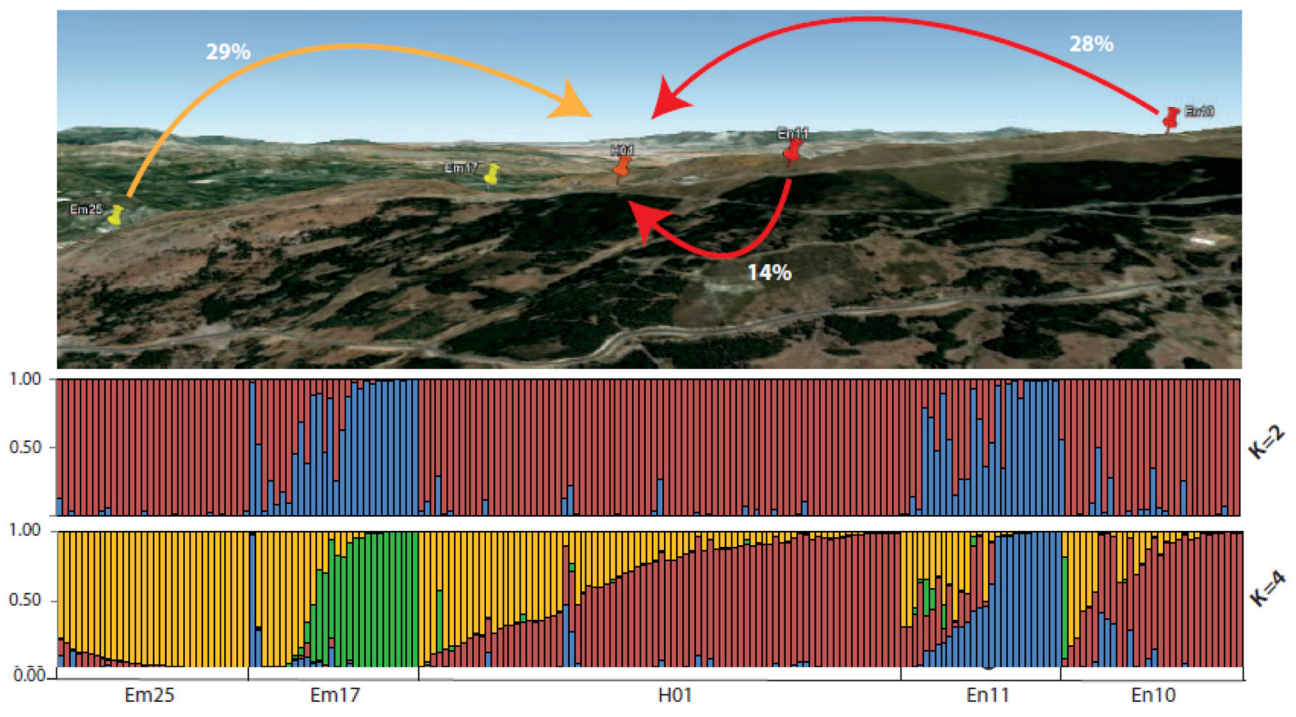


FIGURE 3 | Significant gene flow values among *E. mediohispanicum* and *E. nevadense* populations surrounding the hybrid zone. Bottom: diagram with individual barplots from STRUCTURE (K = 2 and K = 4) based on the variation of 10 unlinked microsatellite loci.

Reproductive Barriers

Significant barriers to reproduction were found between *E. mediohispanicum* and *E. nevadense*, considering the population and species level. However, these barriers were asymmetric. The hybrid crosses performed on *E. mediohispanicum* plants presented a similar fitness to the intraspecific crosses performed in this species, considering both the cumulative predispersal component of the fitness and the

TABLE 5 | Recent migration rates (m_{ij}) and SEs from *i* populations to *j* populations in the transect between *E. mediohispanicum* and *E. nevadense,* inferred from the variation at 10 microsatellite DNA loci using BayesAss.

		Population *j*				
		Em25	**Em17**	**H01**	**En11**	**En10**
Pop *i*	Em25	**0.677 ± 0.019**	0.010 ± 0.019	**0.287 ± 0.042**	0.018 ± 0.028	0.008 ± 0.016
	Em17	0.011 ± 0.022	**0.879 ± 0.076**	0.055 ± 0.065	0.045 ± 0.053	0.010 ± 0.018
	H01	0.004 ± 0.009	0.017 ± 0.021	**0.947 ± 0.046**	0.027 ± 0.041	0.005 ± 0.009
	En11	0.011 ± 0.023	0.019 ± 0.034	**0.141 ± 0.099**	**0.818 ± 0.098**	0.011 ± 0.020
	En10	0.008 ± 0.017	0.017 ± 0.026	**0.279 ± 0.047**	0.017 ± 0.031	**0.677 ± 0.021**

Significant gene flow rates are shown in bold.

total cumulative fitness (**Table 6**). When *E. nevadense* plants received the pollen of *E. mediohispanicum,* we found significant positive HI values at both the population and species levels, independent of the predispersal or cumulative fitness. Hence, hybrid crosses showed a mean decrease in fitness between 36% (considering *Wpre* in En11) and 56% (calculated for *Wtot* for En10). Considering the species level, the *E. nevadense* hybrids showed a mean decrease of 40% compared with the intraspecific crosses (**Table 6**).

DISCUSSION

The consequences of new contacts between the closely related species have been widely highlighted for their interest in understanding the evolutionary processes (Barton and Hewitt, 1985; Harrison, 1993). More recently, they have been pointed out for their importance in assisting the adoption of better conservation policies in the face of rising hybridization (Gómez et al., 2015). In this sense, *E. mediohispanicum* and *E. nevadense* were suggested to be the two non-sister species (Abdelaziz et al., 2014) with contrasting distribution and a karyotypic history in the Iberian Peninsula (Muñoz-Pajares et al., 2017, 2018). However, as in the Sierra Nevada National Park, *E. mediohispanicum* inhabits the low lands, while *E. nevadense* has been described as an endemic species of the Sierra Nevada peaks (Blanca et al., 2009); both species could be the result of a recent or ongoing process of evolutionary divergence by the adaptation to high mountains. This elevational parapatric distribution could promote divergent ecological selection patterns, as those observed in other plant species (Campbell, 2004; Muir et al., 2013), preventing gene flow into the parental areas but creating a stable hybrid zone at the contact. This is supported in our case by the phenotypic stability of the contact zone after 10 years, and it is especially relevant because it makes the scenarios of non-sister species and non-completed speciation compatible as the origin of this hybrid zone. These hybridizing contact zones have important effects on the evolutionary dynamics of the hybridizing species and the biodiversity in the area where the contact happens (Wielstra, 2019; Alves de Moura et al., 2020). The hybridization is not always evident for human, which could have even more pervasive consequences for the biology of the species and their conservation (Gómez et al., 2015; Waldron et al., 2019).

With almost 200 genotyped plants along a transect from *E. mediohispanicum* to *E. nevadense*, we found significant differences between the populations, except for the putative hybrid population. This population seemed to be a mixture between genotypes coming from Em25 and En10, the populations in the extremes of the transect. Most of the genotypes of the hybrid population showed variable admixture proportions with the main genetic clusters of populations Em25 and En10. However, there were also plant genotypes belonging to one of the two inferred genetic sources in Em25 and En10 (**Figure 3**). Therefore, this hybrid zone shows the typical pattern found in the unimodal hybrid zones, with the probable presence of parental plants and/or backcrosses among hybrids, the more frequent plants in the population. These unimodal hybrid zones are frequent when the hybridizing species have not developed strong reproductive barriers (Harrison and Bogdanowicz, 1997). However, Jiggins and Mallet (2000) proposed that the unimodal and bimodal hybrid zones represent different stages of the speciation process: the former would correspond to the early stages in the speciation continuum, while the latter would signify that speciation is nearly completed (Mallet and Dasmahapatra, 2012).

The absence of an interspecific genetic structure and differentiation found in the altitudinal transect from *E. mediohispanicum* to *E. nevadense* seems to be related to the significant and unidirectional levels of gene flow detected for the hybrid population. These patterns are probably related to the highly diverse pollinator assemblage interacting with these species in the Sierra Nevada (Gómez et al., 2009a,b; Abdelaziz et al., unpublished data) and to their ability to promote pollen movement between populations (Tochigi et al., 2021; Abdelaziz et al., unpublished data), as supported by the high number of multilocus genotypes identified in this study. We could identify the hybrid zone despite the nearby distribution of our populations in this area. Usually, the hybrid zones have been studied in extended areas where the researchers identify parental populations and hybrids in a latitudinal gradient (e.g., Lepais et al., 2009; Liu et al., 2014). However, when searching for hybrid contact zones on altitudinal gradients, the approach must be more local: first, because the area at the mountain peaks is reduced with the elevation, and second, because the changes of

TABLE 6 | Reproductive barrier estimation between *E. mediohispanicum* and *E. nevadense*.

	Hybrid inviability						
Species	***N***	W_{PRE}	W_{TOT}	**Pops**	***N***	W_{PRE}	W_{TOT}
E. mediohispanicum	30	0.00 [−0.22, 0.24]	0.12 [−0.16,0.45]				
				Em 25	7	0.10 [−0.5657, 0.9557]	0.38 [−0.2355, 1.4564]
				Em 17	23	−0.01 [−0.2025, 0.2196]	0.02 [−0.2210, 0.3582]
E. nevadense	14	**0.40 [0.18, 0.78]**	**0.41 [0.07, 1.31]**				
				En 11	10	**0.36 [0.1709, 0.8533]**	**0.45 [0.1408, 1.1338]**
				En 10	4	**0.53 [0.2189, 0.7835]**	**0.56 [0.3287, 0.7865]**

The number of plants used in the control-crossing experiment (N) and hybrid inviability (comparing intraspecific and interspecific crosses) at predispersal (W_{PRE}) and total (W_{TOT}) fitness calculated for the hybrid crosses made on E. mediohispanicum and E. nevadense plants. We show the hybrid inviability values per population and per species and their 95% CI. Significant values are stressed in bold.

kilometers in latitude correspond to the changes of meters in elevation for the environmental conditions affecting the plants (Jump et al., 2009). The interest for studying hybrid zones in elevation gradients is increasing in recent years (Aizawa and Iwaizumi, 2020; Mimura and Suga, 2020; Tamaki and Yamada, 2020), although some of these hybrid zones are already in the central body knowledge of hybridization in plants (Aguilar et al., 1999; James and Abbott, 2005).

The gene flow associated with the seed or pollen movements is not enough to generate hybrids in a contact zone. Weak reproductive barriers between the contacting species are also needed, whatever the mechanisms or the way to estimate those (Baack et al., 2015). We calculated pre- and post-dispersal components of reproductive barriers between *E. mediohispanicum* and *E. nevadense*. All of them were coherent and found evidence of the existence of asymmetric reproductive barriers between both species. In this sense, the hybrids obtained by pollinating the plants of *E. nevadense* experienced significant reductions of fitness when compared with the intraspecific crosses at both the predispersal (seed production) and postdispersal (germination and survival) components. In contrast, the hybrid crosses performed on *E. mediohispanicum* showed no fitness differences with the intraspecific crosses. Moreover, the *E. mediohispanicum* hybrids displayed a hybrid vigor for such traits as the number of stalks and the number of flowers, while significantly reduced values for these traits were found for *E. nevadense* hybrids. This means that the *E. nevadense* hybrids suffer from an important reduction in fitness at the prezygotic and postzygotic levels. The traits found to be modified in hybrids are related to floral display and are far from irrelevant. Indeed, they were recurrently demonstrated to play an important role in the attraction of pollinators and being under significant natural selection regimes in *Erysimum* (Gómez et al., 2009a) and other plant species (e.g., Johnston, 1991; Parachnowitsch and Kessler, 2010; Trunschke et al., 2017). Hence, the hybrids of *E. mediohispanicum* probably play an important role in maintaining the hybrid population.

In nature, the plants in the hybrid zone showed vigor only for traits related to the plant size (number of stalks and flowers), but intermediate values were shown for the rest (including all the flower size traits). Reductions in fitness have been described for hybrids mainly associated with sterility and developmental problems (Bomblies and Weigel, 2007; Ouyang et al., 2010; Levin, 2012), promoting the rise of efficient reproductive barriers (Ispolatov and Doebeli, 2009). However, hybrid organisms are also more vigorous than either of the parentals due to heterosis (Huang et al., 2015) or transgressive segregation in the further hybrid generations (Rieseberg et al., 1996, 1999). However, we have found that, while some traits can show heterotic patterns, some other traits can show just intermediate values to both parents. Depending on the trait modified by the hybridization, adaptive trajectories can significantly change and the evolutionary path of the hybrid population can move closer to or further away from the parental species. This fact, together with the absence of the important differences in flower size, may help in the occurrence of hybrids and backcrosses in the natural populations, favoring the continuity or even the spread of the hybrid population. Our phenotypic analysis, 10 years after the first sampling, confirms the same phenotypic pattern found in the transect in the first year and also found in the artificial hybrids developed in the greenhouse. Therefore, we can say that the hybrids are still produced in this contact zone. Studies on the stability of hybrids beyond F1 and on the natural selection acting on the parental and hybrid plants would clarify the stability of this hybrid zone and its effects on the biodiversity of the Sierra Nevada high mountains.

Despite the amount of data included in the current study, we could neither identify the reason why most of the gene flow to the hybrid population is coming from the most distant parental populations nor identify the causes of the asymmetry in the hybrid fitness. On the one hand, the non-homogeneous pattern of gene flow found in our altitudinal gradient could be related to the erratic movements or long-distance flights of the pollinators, as those described for *Syzygium tierneyanum* and *Delphinium nuttallianum* (Hopper, 1980; Schulke and Waser, 2001). The composition of the pollinator assemblage has an important impact on the gene flow and the genetic structure pattern found on the plant species interacting with them (Brunet et al., 2019). On the other hand, the mechanism underlying the asymmetry in hybrid fitness could be related to asymmetries in the interaction between the pollen and the stigma. A similar interaction seems to play a central role in the asymmetric patterns of hybridization

found in different plant species (Rahmé et al., 2009; Pickup et al., 2019). Moreover, the outbreeding depression (Barmentlo et al., 2018) promoted by the intense patterns of local adaptation to the high elevation of the mountain could significantly contribute to the asymmetrical hybridization pattern found in this study.

CONCLUSIONS

The highly coherent results obtained by combining the field, the experimental, and the genetic data demonstrate that, in the altitudinal gradient between *E. mediohispanicum* and *E. nevadense* studied in the present work, there exists a narrow unimodal hybrid zone. The hybrid population appears to be maintained by high levels of the incoming gene flow from mainly two parental populations and by the weak and asymmetrical reproductive barriers between the parental species. *E. mediohispanicum* tolerates the interspecific pollen, but this does not occur with *E. nevadense*. This places *E. mediohispanicum* in a better situation for introgression events than E. nevadense. This pattern was replicated in the greenhouse, where *E. mediohispanicum* hybrids presented some heterotic traits, while *E. nevadense* showed reduced values for these same traits. Finally, a similar pattern was also found in the field when the same traits related to plant sizes showed heterotic values in the hybrid population compared with those in the parental populations. The heterotic phenotypes are probably more frequent in the hybrid zone because of their advantage in attracting pollinators and improving fitness since these traits were demonstrated to be under selection in *E. mediohispanicum* (Gómez et al., 2009a). Finding the same pattern after 10 years is evidence of the stability of this hybrid zone between *E. mediohispanicum* and *E. nevadense* at the top of the Sierra Nevada mountains.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

MA, AJM-P, FP, and JMG conceived and designed the experiments. MA, AJM-P, MB, and AG-M performed the experiments. MA, AJM-P, FP, JMG, MB, and AG-M analyzed the data. MA, AJM-P, FP, and JMG drafted and modified the manuscript. All authors read and approved the manuscript.

FUNDING

This research has been supported by a grant from the Spanish Ministry of Economy and Competitiveness (CGL2014-59886-JIN), the *Organismo Autónomo de Parques Nacionales* (Ref: 2415/2017), and the Ministry of Science and Innovation (PID2019-111294GB-I00/SRA/10.13039/501100011033). FP acknowledges the projects A.RNM.505.UGR18 (FEDER/Junta de Andalucía-Consejería de Economía y Conocimiento) and the Spanish Ministry of Science and Competitiveness (CGL2016-79950-R; CGL2017-86626-C2-2-P), including FEDER funds. MA was supported by the *TransSpeciation* project (CGL2014-59886-JIN). AJM-P was funded by the European Commission under the Marie Sklodowska-Curie Action Cofund 2016 EU agreement 754446 and the UGR Research and knowledge Transfer—Athenea3i. MB was supported by *globalHybrids* project (Ref: 2415/2017) and AG-M was supported by *OUTevolution* project (PID2019-111294GB-I00/SRA/10.13039/501100011033).

ACKNOWLEDGMENTS

The authors thank the Sierra Nevada National Park headquarters for providing the permits to work into the National Park. The authors are grateful to Belén Herrador for her assistance during the lab work. The authors also thank the comments on their work of the associated editor and two reviewers, which certainly improved the previous version of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.687094/full#supplementary-material

REFERENCES

Ågren, J., and Schemske, D. W. (1993). Outcrossing rate and inbreeding depression in two annual monoecious herbs, *Begonia hirsuta* and *B. semiovata*. *Evolution* 47, 125–135. doi: 10.1111/j.1558-5646.1993.tb01204.x

Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *J. Syst. Evol.* 55, 238–258. doi: 10.1111/jse.12267

Abdelaziz, M., Lorite, J., Muñoz-Pajares, A. J., Herrador, M. B., Perfectti, F., and Gómez, J. M. (2011). Using complementary techniques to distinguish cryptic species: a new erysimum (*Brassicaceae*) species from North Africa. *Am. J. Bot.* 98, 1049–1060. doi: 10.3732/ajb.1000438

Abdelaziz, M., Muñoz-Pajares, A. J., Lorite, J., Herrador, M. B., Perfectti, F., and Gómez, J. M. (2014). Phylogenetic relationships of Erysimum (*Brassicaceae*) from the baetic mountains (SE Iberian Peninsula. *Anal. Jardín Botán. Madrid* 1, 1–10. doi: 10.3989/ajbm.2377

Aguilar, J. F., Rossell,ó J. A., and Feliner, G. N. (1999). Molecular evidence for the compilospecies model of reticulate evolution in Armeria (*Plumbaginaceae*). *Syst. Biol.* 48, 735–754. doi: 10.1080/106351599259997

Aizawa, M., and Iwaizumi, M. G. (2020). Natural hybridization and introgression of *Abies firma* and *Abies homolepis* along the altitudinal gradient and genetic insights into the origin of *Abies umbellata*. *Plant Species Biol.* 35, 147–157. doi: 10.1111/1442-1984.12269

Al-Shebaz, I. A. (2012). A generic and tribal synopsis of the *Brassicaceae* (*Cruciferae*). *Taxon* 6, 931–954. doi: 10.1002/tax.615002

Alves de Moura, Y., Alves-Pereira, A., Silva, C. C., de Souza, L. M., de Souza, A. P., and Koehler, S. (2020). Secondary origin, hybridization and sexual reproduction in a diploid-tetraploid contact zone of the facultatively apomictic orchid *Zygopetalum mackayi*. *Plant Biol.* 22, 939–948. doi: 10.1111/plb.13148

Ancev, M. (2006). Polyploidy and hybridization in bulgarian *Brassicaceae*: distribution and evolutionary role. *Phytol. Balcan.* 12, 357–366.

Anderson, E. (1949). *Introgressive Hybridization*. New York, NY: John Wiley. doi: 10.5962/bhl.title.4553

Baack, E., Melo, M. C., Rieseberg, L. H., and Ortiz-Barrientos, D. (2015). The origins of reproductive isolation in plants. *New Phytol.* 207, 968–984. doi: 10.1111/nph.13424

Barmentlo, S. H., Meirmans, P. G., Luijten, S. H., Triest, L., and Oostermeijer, J. G. B. (2018). Outbreeding depression and breeding system evolution in small, remnant populations of Primula vulgaris: consequences for genetic rescue. *Conserv. Genet.* 19, 545–554. doi: 10.1007/s10592-017-1031-x

Barton, N. H., and Hewitt, G. M. (1985). Analysis of hybrid zones. *Annu. Rev. Ecol. Evol. Syst.* 16, 113–148. doi: 10.1146/annurev.es.16.110185.000553

Blanca, G., Cabezudo, B., Cueto, M., López, C. F., and Torres, C. M. (2009). *Flora Vascular de Andalucía Oriental*. Consejería de Medio Ambiente, Junta de Andalucía, Seville, Spain.

Bomblies, K., and Weigel, D. (2007). Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat. Rev. Genet.* 8, 382–393. doi: 10.1038/nrg2082

Brunet, J., Zhao, Y., and Clayton, M. K. (2019). Linking the foraging behavior of three bee species to pollen dispersal and gene flow. *PLoS ONE* 14:e0212561. doi: 10.1371/journal.pone.0212561

Campbell, D., and Aldridge, G. (2006). "Floral biology of hybrid zones," in *Ecology and Evolution of Flowers,* eds L. D. Harder, and S. C. H. Barrett (Oxford: Oxford University Press).

Campbell, D. R. (2004). Natural selection in Ipomopsis hybrid zones: implications for ecological speciation. *New Phytol.* 161, 83–90. doi: 10.1046/j.1469-8137.2003.00919.x

Canty, A., and Ripley, B. (2017). *Package 'Boot'. Bootstrap Functions. Version. 1.3*.

Clot, B. (1992). Caryosystématique de quelques *Erysimum* L. dans le nord de la Péninsule Ibérique. *Anal. Jardín Botán. Madrid* 49, 215–229.

Couvreur, T. L. P., Franzke, A., Al-Shehbaz, I. A., Bakker, F. T., Koch, M. A., and Mummenhoff, K. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (*Brassicaceae*). *Mol. Biol. Evol.* 27, 55–71. doi: 10.1093/molbev/msp202

Coyne, J. A., and Orr, H. A. (2004). *Speciation.* Sunderland, MA: Sinauer Ass., Inc.

Cruz, R., Rolán-Álvarez, E., and García, C. (2001). Sexual selection on phenotypic traits in a hybrid zone of *Littorina saxatilis* (olivi). *J. Evol. Biol.* 14, 773–785. doi: 10.1046/j.1420-9101.2001.00324.x

Cruzan, M. B., and Arnold, M. L. (1993). Ecological and genetic associations in an *Irish* hybrid zone. *Evolution* 47, 1432–1445. doi: 10.1111/j.1558-5646.1993.tb02165.x

Dobzhansky, T. (1940). Speciation as a stage in evolutionary divergence. *Am. Nat.* 74, 312–321. doi: 10.1086/280899

Dowling, T. E., and Secor, C. L. (1997). The role of hybridization and introgression in the diversification of animals. *Annu. Rev. Ecol. Syst.* 28, 593–619. doi: 10.1146/annurev.ecolsys.28.1.593

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

Ellstrand, N. C., Whitkus, R., and Rieseberg, L. H. (1996). Distribution of spontaneous plant hybrids. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5090–5093. doi: 10.1073/pnas.93.10.5090

Emms, S. K., and Arnold, M. L. (1997). The effect of habitat on parental and hybrid fitness: transplant experiments with *Louisiana irises*. *Evolution* 51, 1112–1119. doi: 10.1111/j.1558-5646.1997.tb03958.x

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

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under linux and windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587. doi: 10.1093/genetics/164.4.1567

Fuertes Aguilar, J., Rossell,ó J. A., and Nieto Feliner, G. (1999). Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of Armeria (*Plumbaginaceae*). *Mol. Ecol.* 8, 1341–1346. doi: 10.1046/j.1365-294X.1999.00690.x

Gómez, J. M. (2003). Herbivory reduces the strength of pollinator-mediated selection in the Mediterranean herb Erysimum mediohispanicum: consequences for plant specialization. *Am. Nat.* 162, 242–256. doi: 10.1086/376574

Gómez, J. M., Abdelaziz, M., Camacho, J. P. M., Muñoz-Pajares, A. J., and Perfectti, F. (2009b). Local adaptation and maladaptation to pollinators in a generalist geographic mosaic. *Ecol. Lett.* 12, 672–682. doi: 10.1111/j.1461-0248.2009.01324.x

Gómez, J. M., Bosch, J., Perfectti, F., Fernández, J., and Abdelaziz, M. (2007). Pollinator diversity affects plant reproduction and recruitment: the tradeoffs of generalization. *Oecologia* 153, 597–605. doi: 10.1007/s00442-007-0758-3

Gómez, J. M., González-Megías, A., Lorite, J., Abdelaziz, M., and Perfectti, F. (2015). The silent extinction: climate change and the potential hybridization-mediated extinction of endemic high-mountain plants. *Biodivers. Conserv.* 24, 1843–1857. doi: 10.1007/s10531-015-0909-5

Gomez, J. M., Munoz-Pajares, A. J., Abdelaziz, M., Lorite, J., and Perfectti, F. (2014). Evolution of pollination niches and floral divergence in the generalist plant Erysimum mediohispanicum. *Ann. Bot.* 113, 237–249. doi: 10.1093/aob/mct186

Gómez, J. M., Perfectti, F., Bosch, J., and Camacho, J. P. M. (2009a). A geographic selection mosaic in a generalized plant–pollinator–herbivore system. *Ecol. Monogr.* 79, 245–264. doi: 10.1890/08-0511.1

Good, T. P., Ellis, J. C., Annett, C. A., and Pierotti, R. (2000). Bounded hybrid superiority in an avian hybrid zone: effects of mate, diet, habitat choice. *Evolution* 54, 1774–1783. doi: 10.1111/j.0014-3820.2000.tb00721.x

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Mol. Ecol. Notes* 5, 184–186. doi: 10.1111/j.1471-8286.2004.00828.x

Greuter, W., Burdet, H. M., and Long, G. (1986). *Med-checklist 3, Dicotyledones (Convolvulaceae-Labiatae*. Gen è ve: Conservatoire et Jardin botaniques de la Ville de Gen è ve

Harrison, R. G. (1993). *Hybrid Zones and the Evolutionary Process.* New York, NY: Oxford University Press.

Harrison, R. G., and Bogdanowicz, S. M. (1997). Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution* 51, 493–505. doi: 10.1111/j.1558-5646.1997.tb02437.x

Hopper, S. D. (1980). Pollination of the rain-forest tree *Syzygium tierneyanum* (*Myrtaceae*) at Kuranda, Northern Queensland. *Aust. J. Bot.* 28, 223–237.. doi: 10.1071/BT9800223

Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., et al. (2015). Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.* 6:6258. doi: 10.1038/ncomms7258

Ispolatov, I., and Doebeli, M. (2009). Speciation due to hybrid necrosis in plant–pathogen models. *Evolut. Int. J. Organ. Evolut.* 63, 3076–3084. doi: 10.1111/j.1558-5646.2009.00800.x

James, J. K., and Abbott, R. J. (2005). Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (*Asteraceae*) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution* 59, 2533–2547. doi: 10.1111/j.0014-3820.2005.tb00967.x

Jiggins, C. D., and Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends Ecol. Evol.* 15, 250–255. doi: 10.1016/S0169-5347(00)01873-5

Johnston, M. O. (1991). Natural selection on floral traits in two species of Lobelia with different pollinators. *Evolution* 45, 1468–1479. doi: 10.1111/j.1558-5646.1991.tb02649.x

Jost, L. (2008). Gst and its relatives do not measure differentiation. *Mol. Ecol.* 17, 4015–4026. doi: 10.1111/j.1365-294X.2008.03887.x

Jump, A. S., Mátyás, C., and Peñuelas, J. (2009). The altitude-for-latitude disparity in the range retractions of woody species. *Trends Ecol. Evolut.* 24, 694–701. doi: 10.1016/j.tree.2009.06.007

Keller, B., Ganz, R., Mora-Carrera, E., Nowak, M. D., Theodoridis, S., Koutroumpa, K., et al. (2021). Asymmetries of reproductive isolation are reflected in directionalities of hybridization: integrative evidence on the complexity of species boundaries. *New Phytol.* 229, 1795–1809. doi: 10.1111/nph.16849

Key, K. H. L. (1968). The concept of stasipatric speciation. *Syst. Zool.* 17, 14–22. doi: 10.1093/sysbio/17.1.14

Knief, U., Bossu, C. M., Saino, N., Hansson, B., Poelstra, J., Vijay, N., et al. (2019). Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nat. Ecol. Evolut.* 3, 570–576. doi: 10.1038/s41559-019-0847-9

Lepais, O., Petit, R. J., Guichoux, E., Lavabre, J. E., Alberto, F., Kremer, A., et al. (2009). Species relative abundance and direction of introgression in oaks. *Mol. Ecol.* 18, 2228–2242. doi: 10.1111/j.1365-294X.2009.04137.x

Levin, D. A. (1967). Hybridization between anual species of phlox: population structure. *Am. J. Bot.* 54, 1122–1130. doi: 10.1002/j.1537-2197.1967.tb10742.x

Levin, D. A. (2012). The long wait for hybrid sterility in flowering plants. *New Phytol.* 196, 666–670. doi: 10.1111/j.1469-8137.2012.04309.x

Liu, B., Abbott, R. J., Lu, Z., Tian, B., and Liu, J. (2014). Diploid hybrid origin of ostryopsis intermedia (*Betulaceae*) in the Qinghai-Tibet Plateau triggered by quaternary climate change. *Mol. Ecol.* 23, 3013–3027. doi: 10.1111/mec.12783

Mallet, J., and Dasmahapatra, K. K. (2012). Hybrid zones and the speciation continuum in *Heliconius butterflies*. *Mol. Ecol.* 21, 5643–5645. doi: 10.1111/mec.12058

Marhold, K., and Lihová, J. (2006). Polyploidy, hybridization and reticulate evolution: lessons from the *Brassicaceae*. *Plant Syst. Evolut.* 259, 143–174. doi: 10.1007/s00606-006-0417-x

Marques, I., Feliner, G. N., Draper Munt, D., Martins-Loução, M. A., and Aguilar, J. F. (2010). Unraveling cryptic reticulate relationships and the origin of orphan hybrid disjunct populations in *Narcissus*. *Evolut. Int. J. Organ. Evolut.* 64, 2353–2368. doi: 10.1111/j.1558-5646.2010.00983.x

Mayr, E. (1942). *Systematics and the Origin of Species*. New York, NY: Columbia University Press.

McMillan, W. O., Jiggins, C. D., and Mallet, J. (1997). What initiates speciation in passion-vine butterflies?. *Proc. Natl. Acad. Sci. U.S.A.* 94, 8628–8633. doi: 10.1073/pnas.94.16.8628

Mimura, M., and Suga, M. (2020). Ambiguous species boundaries: Hybridization and morphological variation in two closely related rubus species along altitudinal gradients. *Ecol. Evol.* 10, 7476–7486. doi: 10.1002/ece3.6473

Minder, A. M., Rothenbuehler, C., and Widmer, A. (2007). Genetic structure of hybrid zones between *Silene latifolia* and *Silene dioica* (*Caryophyllaceae*): evidence for introgressive hybridization. *Mol. Ecol.* 16, 2504–2516. doi: 10.1111/j.1365-294X.2007.03292.x

Mohamed, M. A. (2013). *How species are evolutionarily maintained: Pollinator-mediated divergence and hybridization in Erysimum mediohispanicum and Erysimum nevadense* (Doctoral dissertation). Universidad de Granada.

Muir, G., Osborne, O. G., Sarasa, J., Hiscock, S. J., and Filatov, D. A. (2013). Recent ecological selection on regulatory divergence is shaping clinal variation in Senecio on Mount Etna. *Evolution* 67, 3032–3042. doi: 10.1111/evo.12157

Muñoz-Pajares, A. J., Abdelaziz, M., and Picó, F. X. (2020). Temporal migration rates affect the genetic structure of populations in the biennial *Erysimum mediohispanicum* with reproductive asynchrony. *AoB Plants* 12:plaa037. doi: 10.1093/aobpla/plaa037

Muñoz-Pajares, A. J., García, C., Abdelaziz, M., Bosch, J., Perfectti, F., and Gómez, J. M. (2017). Drivers of genetic differentiation in a generalist insect-pollinated herb across spatial scales. *Mol. Ecol.* 26, 1576–1585. doi: 10.1111/mec.13971

Muñoz-Pajares, A. J., Herrador, M. B., Abdelaziz, M., Picó, F. X., Sharbel, T. F., Gómez, J. M., et al. (2011). Characterization of microsatellite loci in Erysimum mediohispanicum (Brassicaceae) and cross-amplification in related species. *Am. J. Bot.* 98, e287–e289. doi: 10.3732/ajb.1100181

Muñoz-Pajares, A. J., Perfectti, F., Loureiro, J., Abdelaziz, M., Biella, P., Castro, M., et al. (2018). Niche differences may explain the geographic distribution of cytotypes in *Erysimum mediohispanicum*. *Plant Biol.* 20, 139–147. doi: 10.1111/plb.12605

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press. doi: 10.7312/nei-92038

Nieto Feliner, G., Rosato, M., Alegre, G., San Segundo, P., Rossell,ó J. A., Garnatje, T., et al. (2019). Dissimilar molecular and morphological patterns in an introgressed peripheral population of a sand dune species (*Armeria pungens*, *Plumbaginaceae*. *Plant Biol.* 21, 1072–1082. doi: 10.1111/plb.13035

Ouyang, Y., Liu, Y. G., and Zhang, Q. (2010). Hybrid sterility in plant: stories from rice. *Curr. Opin. Plant Biol.* 13, 186–192. doi: 10.1016/j.pbi.2010.01.002

Owens, G. L., Baute, G. J., and Rieseberg, L. H. (2016). Revisiting a classic case of introgression: hybridization and gene flow in Californian sunflowers. *Mol. Ecol.* 25, 2630–2643. doi: 10.1111/mec.13569

Parachnowitsch, A. L., and Kessler, A. (2010). Pollinators exert natural selection on flower size and floral display in penstemon digitalis. *New Phytol.* 188, 393–402. doi: 10.1111/j.1469-8137.2010.03410.x

Parsons, K. J., Son, Y. H., and Albertson, R. C. (2011). Hybridization promotes evolvability in African cichlids: connections between transgressive segregation and phenotypic integration. *Evol. Biol.* 38, 306–315. doi: 10.1007/s11692-011-9126-7

Pickup, M., Brandvain, Y., Fraïsse, C., Yakimowski, S., Barton, N. H., Dixit, T., et al. (2019). Mating system variation in hybrid zones: facilitation, barriers and asymmetries to gene flow. *New Phytol.* 224, 1035–1047. doi: 10.1111/nph.16180

Polatschek, A. (1986). "Erysimum," in *Mountain flora of Greece, Vol. l,* ed A. Strid (Cambridge: Cambridge University Press), 239–247.

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

R Development Core Team. (2014). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org

Rahmé, J., Widmer, A., and Karrenberg, S. (2009). Pollen competition as an asymmetric reproductive barrier between two closely related Silene species. *J. Evol. Biol.* 22, 1937–1943. doi: 10.1111/j.1420-9101.2009.01803.x

Rambaut, A., and Drummond, A. J. (2007). *Tracer v1.4 [Computer Program]*. Available online at: http://tree.bio.ed.ac.uk/software/tracer/

Rieseberg, L. H., Archer, M. A., and Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity* 83, 363–372. doi: 10.1038/sj.hdy.6886170

Rieseberg, L. H., Baird, S. J. E., and Desroche, A. (1998). Patterns of mating in wild sunflower hybrid zones. *Evolution* 52, 713–726. doi: 10.1111/j.1558-5646.1998.tb03696.x

Rieseberg, L. H., Sinervo, B., Linder, C. R., Ungerer, M. C., and Arias, D. M. (1996). Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science* 272, 741–745. doi: 10.1126/science.272.5262.741

Schulke, B., and Waser, N. M. (2001). Long-distance pollinator flights and pollen dispersal between populations of *Delphinium nuttallianum*. *Oecologia* 127, 239–245. doi: 10.1007/s004420000586

Tamaki, I., and Yamada, Y. (2020). Environmental pressure rather than ongoing hybridization is responsible for an altitudinal cline in the morphologies of two oaks. *J. Plant Ecol.* 13, 413–422. doi: 10.1093/jpe/rtaa028

Tastard, E., Ferdy, J. B., Burrus, M., Thébaud, C., and Andalo, C. (2012). Patterns of floral colour neighbourhood and their effects on female reproductive success in an antirrhinum hybrid zone. *J. Evol. Biol.* 25, 388–399. doi: 10.1111/j.1420-9101.2011.02433.x

Tochigi, K., Shuri, K., Kikuchi, S., Naoe, S., Koike, S., and Nagamitsu, T. (2021). Phenological shift along an elevational gradient and dispersal of pollen and seeds maintain a hybrid zone between two cherry tree species. *Plant Species Biol.* 36, 230–245. doi: 10.1111/1442-1984.12311

Trunschke, J., Sletvold, N., and Ågren, J. (2017). Interaction intensity and pollinator-mediated selection. *New Phytol.* 214, 1381–1389. doi: 10.1111/nph.14479

Turner, B. L. (2006). Taxonomy and nomenclature of the *Erysimum asperum - E. capitatum* complex (*Brassicaceae*). *Phytologia* 88, 279–287. doi: 10.5962/bhl.part.10454

Vedenina, V. Y., and Helversen, O. V. (2003). Complex courtship in a bimodal grasshopper hybrid zone. *Behav. Ecol. Sociobiol.* 54, 44–54. doi: 10.1007/s00265-003-0595-2

Waldron, B. P., Kuchta, S. R., Hantak, M. M., Hickerson, C. A. M., and Anthony, C. D. (2019). Genetic analysis of a cryptic contact zone between mitochondrial clades of the eastern red-backed salamander, Plethodon cinereus. *J. Herpetol.* 53, 144–153. doi: 10.1670/18-088

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

Whitham, T. G., Martinsen, G. D., Keim, P., Floate, K. D., Dungey, H. S., and Potts, B. M. (1999). Plant hybrid zones affect biodiversity: tools for a genetic-based understanding of community structure. *Ecology* 80, 416–428. doi: 10.1890/0012-9658(1999)0800416:PHZABT2.0.CO;2

Whitham, T. G., Morrow, P. A., and Potts, B. M. (1994). Plant hybrid zones as centers of biodiversity: the herbivore community of two endemic tasmanian eucalypts. *Oecologia* 97, 481–490. doi: 10.1007/BF00325886

Wielstra, B. (2019). Historical hybrid zone movement: More pervasive than appreciated. *J. Biogeogr.* 46, 1300–1305. doi: 10.1111/jbi.13600

Wielstra, B., Burke, T., Butlin, R. K., Avci, A., Üzüm, N., Bozkurt E, et al. (2017). A genomic footprint of hybrid zone movement in crested newts. *Evolut. Lett.* 1, 93–101. doi: 10.1002/evl3.9

Wilson, G. A., and Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163, 1177–1191. doi: 10.1093/genetics/163.3.1177

Young, N. D. (1996). An analysis of the causes of genetic isolation in two Pacific coast iris hybrid zones. *Can. J. Bot.* 74, 2006–2013. doi: 10.1139/b96-241

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 November 2021
doi: 10.3389/fpls.2021.737111

Hopping or Jumping on the Cliffs: The Unusual Phylogeographical and Demographic Structure of an Extremely Narrow Endemic Mediterranean Plant

Sandro Strumia[1], Annalisa Santangelo[2], Teresa Rosa Galise[2], Salvatore Cozzolino[2] and Donata Cafasso[2]*

[1] Department of Environmental, Biological and Pharmaceutical Sciences and Technologies, University of Campania "Luigi Vanvitelli", Caserta, Italy, [2] Department of Biology, University of Naples Federico II, Naples, Italy

OPEN ACCESS

Edited by:
Gonzalo Nieto Feliner, Real Jardín Botánico (RJB), Spanish National Research Council (CSIC), Spain

Reviewed by:
Giovanni Zecca, Independent Researcher, Milan, Italy
Waldir M. Berbel-Filho, Swansea University, United Kingdom

***Correspondence:**
Salvatore Cozzolino
cozzolin@unina.it

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

Received: *06 July 2021*
Accepted: *07 October 2021*
Published: *10 November 2021*

Citation:
Strumia S, Santangelo A, Galise TR, Cozzolino S and Cafasso D (2021) Hopping or Jumping on the Cliffs: The Unusual Phylogeographical and Demographic Structure of an Extremely Narrow Endemic Mediterranean Plant. Front. Plant Sci. 12:737111. doi: 10.3389/fpls.2021.737111

Several past and recent climatic and geological events have greatly influenced the current distribution of coastal species around the Mediterranean Basin. As a consequence, the reconstruction of the distributional history of these species is challenging. In this study, we used both chloroplast and nuclear SNPs to assess the levels of genetic differentiation, contemporary/historical levels of gene flow, and demographic history for the three only known (one mainland and two insular) populations of *Eokochia saxicola*, a rare Mediterranean coastal rocky halophyte. Plastid genome analysis revealed very low intraspecific haplotype variation and partial admixture among Capri and Palinuro populations with at least two independent colonization events for the Strombolicchio islet. Nuclear SNPs variation consistently identified three distinct genetic clusters corresponding to our sampling localities. Furthermore, strong genetic isolation was confirmed by both historical and contemporary levels of migration among the three populations. The DIYABC analysis identified two introductions temporally separated from Palinuro to Capri (ca.25 Mya) and subsequently to Strombolicchio (ca.09 Mya) as the most likely hypothesis for the current distribution of *E. saxicola*. Regardless of their small population sizes, all study sites supported high-genetic diversity maintained by outcrossing and random mating between individuals owing largely to wind pollination, an exclusive trait among Mediterranean narrow endemics. In conclusion, the patterns observed confirm that some Mediterranean endemics are not necessarily "evolutionary dead-ends" but rather represent species that have extensive demographic stability and a strong evolutionary legacy.

Keywords: contemporary and historical gene flow, ddRAD, DIYABC, halophytes, IUCN, long distance dispersal, plant conservation, quaternary sea level oscillations

INTRODUCTION

The Mediterranean Basin is one of the Hotspots of Biodiversity on the Earth (Médail and Quézel, 1999; Myers et al., 2000), largely due to the high number of endemic species (Quezel, 1985; Greuter, 1991). This biodiversity is the result of the current high heterogeneity of environmental matrix, climate, land uses, and geological and climatic events that occurred in the last 15 M years (Thompson, 2005; Blondel et al., 2010).

In particular, the most recent quaternary glacial-interglacial events have greatly influenced the current distribution of many plant species around the Mediterranean Basin (Médail and Diadema, 2009). The role of Glacial Refugia Areas and postglacial migration routes in shaping the current distribution of mainland European plants is well-known (Taberlet et al., 1998; Hewitt, 2001; Heuertz et al., 2004). However, the impact of glaciations on plant species mainly distributed on islands by climatic oscillations remains less explored. The Last Glacial Maximum (LGM) induced the lowering of sea levels, which subsequently led to the creation of several land bridges, facilitating the migration of plant species between the mainland and continental islands (Whittaker and Fernández-Palacios, 2007 quoting Wallace, 1880). Ultimately, the subsequent warming phases led to these temporary land bridges being gradually submerged by rising sea levels (Feliner, 2014). This main isolating barrier favored speciation processes, which resulted in the increase of insular endemics in the Mediterranean Basin (Médail, 2017). In contrast to insular endemics, water level oscillations played a different role in coastal halophytic species. Halophytes are salt-tolerant or salt-resistance species adapted to the very harsh habitat conditions due to high salinity (Ungar, 1991). These species are, therefore, typical of the coastal azonal vegetation [i.e., vegetation occurring where local ecological conditions, such as high salinity, overrule the effect of climate, according to Walter (1985)]. Indeed, for several coastal halophytes, the sea represents the main pathway for gene flow through the dispersal, even at long distances, of ramets or seeds (referred to as hydrochory) (Kadereit and Westberg, 2007; Westberg and Kadereit, 2009). Therefore, the current distribution of halophytic coastal species is the result of both past (main climatic oscillations and related sea-level variations) and recent abiotic factors (i.e., weather conditions, sea storms, and current sea circulations) acting on dispersion processes. Finally, the ecology of both species and plant communities can contribute to the complexity in the interpretation of coastal species distribution (Weising and Freitag, 2007; Feliner, 2014). The reconstruction of the distributional history of coastal halophytes is partly facilitated by their exclusive adaptations to hypersaline habitats that limit their distribution to coastlines, with a linear distribution range and patterns of long-distance dispersal being found to be more influenced by the direction of sea marine currents than by habitat filtering (Kadereit et al., 2005). However, the general pattern largely associated with coastal halophytes inhabiting dune and marshes habitats has not been found in coastal rocky species (Kadereit and Westberg, 2007). Both inland and coastal rocky areas provide suitable habitats for numerous endemic species often with extremely restricted distribution (Davis, 1951; Lavergne et al., 2004; Thompson, 2005) with many of them being considered "narrow endemics" (Kruckeberg and Rabinowitz, 1985; Médail and Baumel, 2018) and threatened by extinction (Montmollin and Strahm, 2005; Orsenigo et al., 2018). Therefore, they play a key role in the biodiversity conservation priority setting (Médail and Baumel, 2018).

Eokochia saxicola (Guss.) Freitag and G. Kadereith (Amaranthaceae) is a perennial evergreen shrub restricted entirely to coastal rocky environments exposed to the salt spray in very few locations around the South Tyrrhenian Sea (Strumia et al., 2020a). Historically, distribution records suggest *E. saxicola* occurred only on three small islands: Strombolicchio (a recent volcanic islet near Sicily), Capri (a calcareous island block), and Ischia (an old volcanic island, where it is now thought to be extinct). However, a new population has recently been discovered on the mainland coastal area along Campania and represents the only known mainland population of *E. saxicola* (Strumia et al., 2015). According to the criteria (less than 500 individuals today distributed in five or less populations) proposed by López-Pujol et al. (2013), *E. saxicola* can be considered an "extremely narrow endemic." This species is a remnant of old lineages of Camphorosmae and is part of the *Chenolea* clade, originating in the early Miocene period during the evolution of the Mediterranean Basin (Kadereit and Freitag, 2011). To date, this clade includes only a few highly disjunct halophytic species, which are typical of warmer temperate climates. *E. saxicola* diverged from two other species [*Chenolea diffusa* Thunb. and *Spirobassia hirsuta* (L) Freitag and G. Kadereith] about 10.2 Mya (Kadereit and Freitag, 2011), i.e., largely before the onset of current Mediterranean climate (about 3.2 Mya, Suc, 1984) and is, therefore, considered a palaeoendemism (Favarger and Contandriopoulos, 1961; Stebbins and Major, 1965). Interestingly, *E. saxicola* distribution sharply contrasts with the general view that coastlines represent linear biogeographic systems connected by sea dispersion as reported for other coastal halophytes (Clausing et al., 2000; Kadereit et al., 2005). Genetic markers are ideal to study seed dispersal and show patterns of populations structuring determined by past and present gene flows (Bermingham and Moritz, 1998). This is particularly true in the case of species without pollen or fossil records. At the same time, the phylogeographical approach may also represent an important tool in conservation (Diniz-Filho et al., 2008) by identifying the populations that possess higher genetic diversity and, therefore, constituting a possible "genetic source" to be considered for future reintroductions (Médail and Baumel, 2018). The main aims of this study were, therefore, to answer the following questions: (1) Do populations of *E. saxicola* show any phylogeographic structure? (2) What are the primary factors that have shaped the phylogeographic structure? (3) What determines the genetic diversity and genetic structure of extant populations? (4) Can the combination of phylogeography and genetic diversity help define *E. saxicola* conservation priorities?

MATERIALS AND METHODS

Study Species

Eokochia saxicola is a multi-branched perennial evergreen shrub found exclusively on maritime rocks (both calcareous and volcanic) close to the sea level in the area occupied by halophilic vegetation (Strumia et al., 2020b). Characteristic features of this species include protogyny, papillose stigma, and recalcitrant pollen, typical of wind pollinated plants (Barone Lumaga et al., 2016). Like other halophytes, the main seed dispersal mechanism is hypothesized to be hydrochory, owing mainly to the presence of diaspores able to float for several days. Moreover, a decrease in the germination rate of aging seeds and the inability for seedlings to anchor at suitable establishment sites have been assessed (Strumia et al., 2020a). According to IUCN, *E. saxicola* is threatened with extinction as it falls in the category *Endangered* (EN) (Santangelo et al., 2012; Orsenigo et al., 2018). The main threats are landslides in the few sites of occurrence and supposed reduction of fertility and reproduction success due to the small number of living individuals. We sampled *E. saxicola* in three locations representing the only known populations of its current distribution: Capri Island near Naples (Campania), Strombolicchio islet (Aeolian Archipelago, Sicily), and Palinuro, along the Cilento and Vallo di Diano and Alburni Nature Park coastline (Campania). Field collections were conducted from 2015 to 2017. Due to the inaccessibility of the cliffs (Strumia et al., 2020a), only the most accessible individuals were sampled. In each sampling site, the maximum estimate number of individuals was recorded through visual census (i.e., by binoculars from a boat) (**Table 1**).

DNA Isolation, Library Preparation, and Sequencing

Genomic DNAs of 51 *Eokochia saxicola* individuals were extracted from preserved leaf material using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), following the protocol of the manufacturer. The quantity and the purity of DNA extractions were checked using the NanoDrop ND 1000 Spectrophotometer (Thermo Fisher Scientific, Delaware), whereas the concentration was assessed with the Qubit fluorometer system (Invitrogen, United States) and the Quant-IT ds-DNA BR Assay kit (Invitrogen). Samples were sequenced using a slight modification to the standard ddRAD protocol (Peterson et al., 2012), with the restriction enzymes *Eco*RI and *Taq*I. A single library with barcoded individuals was sequenced in a single Illumina HiSeq 2500 lane for 150-base single-end reads. From a subset of samples (16 *E. saxicola* + *S. hirsuta*), a "genome skim" was performed (Straub et al., 2012) to gather sufficient data for complete plastid assembly without prior enrichment or isolation of plastid DNA (Coissac et al., 2016; **Supplementary Table SM1**). Libraries were constructed according to the Nextera DNA Library Prep Kit (Illumina) protocol and Nextera XT indexes (Kit v2 Set A), which were used to multiplex the individual samples. Libraries were sequenced in a single Illumina HiSeq 2500 lane for 150-base single-end reads.

TABLE 1 | Sampling sites of *Eokochia saxicola* (abbreviation in brackets).

Site (Abbreviation)	Number of individuals	Sampled individuals
Capri – Grotta dell'Acqua (C)	120	18
Strombolicchio (K)	60	12
Palinuro Porto (P)	40	16
Palinuro Punta Iacco (PIK)	40	1
Palinuro Cala Fetente (CFK)	50	2
Palinuro-Camerota (CAM)	15	2
Total	325	51

A maximum number of individuals (estimated at visual census) and a number of sampled individuals per site.

Preprocessing and Analysis of the Complete Plastome Sequences

The *Eokochia saxicola* and *Spirobassia hirsuta* raw sequence reads were quality checked with the FastQC v0.11.9[1], and low-quality reads and adapters were subsequently trimmed with Trimmomatic v0.39 (Bolger et al., 2014). We implemented a *de novo* assembly of *S. hirsuta* plastome by setting kmer = 121 in Velvet v1.2.10 software (Zerbino and Birney, 2008), and trimmed reads were then further assembled using Pilon v1.23 (Walker et al., 2014). We implemented FASTQ Screen v0.14 (Wingett and Andrews, 2018) for filtering the plastid reads from each *E. saxicola* individual by using a *de novo*-assembled *S. hirsuta* plastome (unpublished), and orphan reads were removed through the python script FASTQ Combine Paired End.py[2]. The *E. saxicola* filtered reads were mapped to *S. hirsuta* plastome by using BOWTIE2 v2.3.5.1 (Li and Durbin, 2009). The alignments (BAM files) produced a single-variant call format (vcf) file for all samples, where the SNPs were identified using bcftools v.1.9 with the setting "mpileup -Ou" and called *via* bcftools using the −mv and –ploidy 1 functions. The vcf file was converted to a Nexus format, and a haplotype network (also including available *Amarantaceae* plastome sequences) was built by using the TCS method implemented in PopArt v1.7 (Leigh and Bryant, 2015).

Preprocessing ddRAD Data

Raw reads were demultiplexed in individual paired-end libraries and filtered for low-quality reads with the process_radtags pipeline in STACKS v2.4 (Rochette et al., 2019). Quality control of demultiplexed reads was performed with FastQC v0.11.9. The catalog RAD-loci was built with *de novo* assembly using the *de novo*_map.pl wrapper in STACKS v2.4 by using parameters $m = 1$, $M = 8$, and $n = 1$, following the optimization procedure described in Paris et al. (2017). The population program in a Denovo_map.pl wrapper was used for collecting the polymorphic SNPs in a vcf file. The vcf files (with 48 individuals) were analyzed with the –missing function of PLINK v1.07 (Purcell et al., 2007) for calculating the percentage of missing data (approximately ∼81% of missing data). Lastly, to avoid the risk of calling false

[1] www.bioinformatics.babraham.ac.uk/projects/fastqc

[2] https://github.com/enormandeau/Scripts/blob/master/fastqCombinePairedEnd.py

SNPs, we reduced the dataset to include all sites with a minor allele frequency (MAF) greater than or equal to 0.05 and with a (*–max-obs-het)* setting of 0.70 (as also suggested in Gargiulo et al., 2021). Finally, to account for the effects of missing data per individual, VCFTOOLS v0.1.14 (Danecek et al., 2011) was implemented to filter the minimum number of ddRAD loci shared by at least 70% of individuals (i.e., the full dataset). To minimize the potential effect of missing data in altering the results, some population demographic analyses were also or only performed on a subsample of individuals by maximizing the number of reads per sample and minimizing the amount of missing data. Therefore, a reduced dataset was constructed with 25 individuals (i.e., only those ranging between 2,000,000 and 500,000 reads, namely 10 individuals from Palinuro, eight from Capri, seven from Strombolicchio, see **Supplementary Text 1.1**), which were filtered so that missing data were present in two or less individuals for each locus (i.e., ddRAD loci present in 95% of the individuals).

Analysis of ddRAD Data

Analyses on the Full (48 Individuals) and Reduced Dataset (25 Individuals)

Maximum Likelihood Tree (ML) trees were generated from both datasets. GTRCAT was used as the substitution model for nucleotide sequences, as implemented in RAxML v8.2.12 (Stamatakis, 2014). The resulting trees were drawn to scale with Figtree v1.4.4[3].

The R package AWclust (Gao and Starmer, 2008) was implemented as a non-parametric distance-based clustering model (allele-shared distance) for the analysis of populations structure with a HapMap format file as input. We estimated the best K using 100 simulations with an interval of K from 1 to 6. To visualize the genetic relationships among individuals, we performed a multidimensional scaling (MDS) based on the optimal number of clusters (best K).

The vcf file (48 individuals) converted into a genind object with the function vcfR2genlight in the R package adegenet v2.02. (Jombart, 2008) was subsequently used for the Discriminant Analysis of Principal Components (DAPC) (**Supplementary Text 1.2** and **Supplementary Figures 5–7**). Lastly, the optimal number of clusters (*K*) was then determined by the lowest Bayesian Information Criterion (BIC), and the DAPC plot was constructed with the function scatter.dapc.

Pairwise Fst among individuals and populations was calculated with the function stamppFst (in the R package adegenet), and the result was visualized in a heatmap plot with function heatmap 0.2. The fineRAD structure was implemented to infer population structure *via* shared ancestry in the ddRAD dataset, focusing on the most recent coalescence (common ancestry) among the sampled individuals (Malinsky et al., 2018). Samples were assigned to populations using 100,000 iterations as burn-in before sampling 100,000 iterations. The trees were built using 10,000 iterations and the output was visualized using fineStructure GUI.

[3]http://tree.bio.ed.ac.uk/software/figtree

Analyses on the Reduced Dataset (25 Individuals)

The following analyses, more sensitive to the excess of missing data, were conducted with the reduced dataset containing 25 individuals and filtered for a minimum number of ddRAD loci shared by 95% of individuals. The function basic.stats in R package hierfstat (Goudet, 2005) was implemented for the estimation of observed heterozygosity (Ho), expected heterozygosity (He), and the within-population inbreeding coefficient (Fis).

We focused on the timing of the connectivity change based on the estimates of historical and contemporary migration by using a coalescent-based method (MIGRATE) and a disequilibrium-based method (BayesAss), respectively. The historical migration rate (**Supplementary Text 1.4** and **Supplementary Table SM3**) was estimated using Bayesian inference in MIGRATE (Beerli, 2006), while the contemporary migration rates (**Supplementary Text 1.3**; **Supplementary Table SM2**; **Supplementary Figures 8, 9**) were estimated from the current generation and two past generations using Bayesian inference in BayesAss v3 (Rannala, 2007).

Alternative scenarios of origin, the direction of migration, and relationships among *E. saxicola* populations were tested on the ddRAD dataset under the coalescent-based approximate Bayesian computation in DIYABC Random Forest v1.0 (Collin et al., 2021). This enables the comparison of different population demographic models and determines the estimation of their parameters without calculating complex likelihood functions (Beaumont, 2010). Different historical demographic scenarios were tested with the following Ne-prior parameters set for each population: Npalinuro (100–500); Ncapri (10–500) and NStrombolicchio (10–50). This low value of Ne priors is based on the rarity of *E. saxicola* among the different locations (**Table 1**).

For *Eokochia saxicola,* a generation time of 10 years was considered (*ex situ* cultivated 5-year-old plants have not bloomed yet). Then, in accordance with the geological age of the sites-hosting current populations (Vezzoli, 1988; Gillot and Keller, 1993; Santangelo et al., 2011), we set the split time (t) as a number of generations from 10 to 50,000. Conditions were set as t1 > t2, according to the timing of events. Bottleneck priors (db time) were set from 10 to 100 generations and the number of founders (Nb) from 10 to 50 individuals. Since previous analyses based on genetic structure did not show admixture events, we did not include admixture events into scenarios.

Four scenarios (**Supplementary Figure 10**) were designed to explore alternative hypotheses regarding the origin and mainland-island migrations (Scenario 1). The colonization of Capri occurred through a long-distance dispersal event of a small fraction of the Palinuro population at time t1. Then, the bottleneck event in the newly founded population (i.e., Capri) was modeled through a limited number of founders (Nb) during a short period (t1-db). At time t2, the Strombolicchio population was born through a long-distance dispersal event of a small fraction of the Capri population. Then, the bottleneck event in the newly founded population (i.e., Strombolicchio) was modeled through a limited number of founders (Nb) during a short period (t2-db) (Scenario 2). The colonization of Capri occurred through a long-distance dispersal event of a small fraction of the

Palinuro population at time t1. Then, the bottleneck event in the newly founded population (i.e., Capri) was modeled through a limited number of founders (Nb) during a short period (t1-db). At time t2, the Strombolicchio population was born through a long-distance dispersal event of a small fraction of the Palinuro population. Then, the bottleneck event in the newly founded population (i.e., Strombolicchio) was modeled through a limited number of founders (Nb) during a short period (t2-db) (Scenario 3). An ancestral population split between Capri and Palinuro populations occurred at time t1. At time t2, Strombolicchio colonization occurred through a long-distance dispersal event of a small fraction of the Capri population. Then, the bottleneck event in the newly founded population (i.e., Strombolicchio) was modeled through a limited number of founders (Nb) during a short period (t2-db) (Scenario 4). An ancestral population split between Capri and Palinuro populations occurred at time t1. At time t2, Strombolicchio colonization occurred through a long-distance dispersal event of a small fraction of the Palinuro population. Then, the bottleneck event in the newly founded population (i.e., Strombolicchio) was modeled through a limited number of founders (Nb) during a short period (t2-db) (**Supplementary Figure 10** and figure caption).

We processed 1,000,000 simulations datasets, ∼250,000 simulations datasets per scenario. For scenario choice, a linear discriminant analysis (LDA) processed on the summary statistics was used before processing Random Forests (RF) predictions (**Supplementary Figure 11**). The 250,000 simulation datasets (i.e., higher than the suggested 20,000 simulation datasets) of the best supported scenario were used to evaluate the model choice and parameter estimation in RF analysis. The number of trees in the constructed random forests were fixed to $n = 1{,}500$, the minimum number to ensure a stable accuracy measure of the best scenario and to estimate the mean value and the lower and upper 95% quartiles of the posterior distributions (Chapuis et al., 2020; **Supplementary Figure 12**).

RESULTS

For the Illumina sequencing for the analysis of SNPs variation in the complete plastome sequences of 16 *E. saxicola* + *S. hirsuta*, 41,187,562 raw reads (150 bp for average read length) were generated; of which, 34,528,403 reads remained after performing the preprocessing procedure (**Supplementary Text 1.1** and **Supplementary Table SM1**). Variant calling revealed 333 SNPs among both *S. hirsuta* and *E. saxicola* individuals as opposed to only three SNPs detected among *E. saxicola* individuals. The location, nucleotide change, annotation, and mutation type of these three SNPs in *E. saxicola* are provided in **Supplementary Table 1**. These SNPs identified three haplotypes in the *E. saxicola* populations, with haplotype H2 present in all three locations and haplotype H1 shared between the mainland and Capri Island, whereas the haplotype H3 was exclusive to Strombolicchio islet (**Figure 1**; **Supplementary Table 1**). All other Amarantaceae plastomes, including the newly assembled *S. hirsuta*, complete plastome (data not shown) differed largely from *E. saxicola* plastome both in size and number of SNPs.

From ddRAD sequencing, a total of 30,022,310 raw reads were obtained. After which, 29,276,067 reads were identified through de-multiplexing and filtering using STACKS pipeline to remove any low-quality read, ambiguous barcodes, and sequences without cut sites. We obtained, on average, 609,918 raw reads per sample and 224,637 mapped reads per sample (**Supplementary Table SM1**).

The pipeline *de novo*_map.pl generated 6,476,170 loci and assembled ∼ 99.9% of loci into contigs. The total length of pair-end contigs (the generated *de novo* assembly) was approximately 175 Mbp, as calculated with the software Bandage (Wick et al., 2015). The average mapping efficiency of our samples to this *de novo* assembly was two reads per locus, as calculated with module gstack of the Stack pipeline. In total, 10,782,592 reads were aligned with the generated genome, i.e., 70% of raw reads did not align (i.e., they were unmapped unique reads). Due to this low coverage, SNPs were called by adopting a variant calling strategy of 3×-minimum coverage. With our filtering approach, in the 48 individuals/3,962 SNPs datasets, we never observed any case of multiple variants in the same locus (suggestive of a diploid status for *E. saxicola*) or of variants called in less than five individuals. With this conservative approach, we may have lost some rare variants, but we have reduced the risk of calling false SNPs due to low coverage.

An ML tree was built with 3,962 SNPs present in at least 70% of samples. The 48 individuals clustered distinctly in three separate clades. The separation of three distinct clades was further emphasized when using the reduced dataset of 25 individuals with 120 SNPs shared by 95% of individuals (**Supplementary Figure 1**).

Genetic Clustering with Awclust software identified $K = 3$ as the optimal number, with each cluster corresponding to a population (**Figure 2A**). The 48 individuals were split into three distinct clusters or geographical regions in the MDS analysis too (**Figure 2B**). Moreover, the AWclust dendrogram plot displays a clade with two clusters (the island populations) and a clade with the mainland Palinuro population (**Figure 2C**).

Visualization of population structure using a DAPC (Discriminant Analysis of Principal Components) concordantly revealed three distinct genetic clusters, which included the two island populations (Strombolicchio and Capri) and the mainland Palinuro population (**Figure 3**; **Supplementary Figure 7**).

The heatmap plot confirmed the existence of the three populations (**Figure 4**). Overall, all populations revealed a moderate level of intrapopulation co-ancestry (**Figure 5**). Comparable results were gathered with the reduced dataset (by only using 120 SNPs present in at least 95% individuals across 25 individuals) but with a higher intrapopulation co-ancestry value as it was found inversely correlated with a number of SNPs and missing data (**Supplementary Text 1.1**; **Supplementary Figure 4**).

The mean genetic diversity parameters among populations were 0.0941 for expected Heterozygosity (He), 0.0841 for observed Heterozygosity (Ho), and 0.0781 for the Fixation index (Fis), respectively. Interestingly, the Palinuro population had the highest inbreeding value (Fis = 0.2343) (**Table 2**).

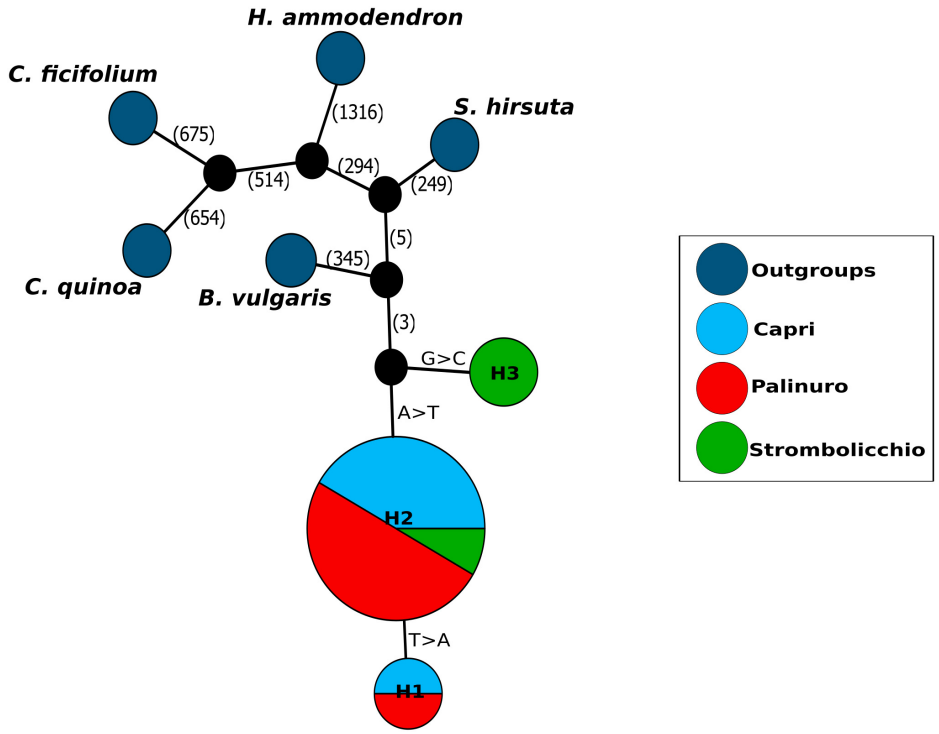


FIGURE 1 | Plastid haplotype network. Each circle represents a different haplotype, where circle size is proportional to the number of individuals that carry the same haplotype. Black circles indicate missing intermediate haplotypes. Mutational steps are reported on the lines of the haplotype networks. Colors show the genotypes belonging to different geographic origins.

Multiple runs of BAYESASS yielded low levels of contemporary gene flow (mc, fraction of individuals that are immigrants) among the three populations. Concordantly, also, the historical rates of migration as calculated in MIGRATE were generally very low (**Figure 6**; **Supplementary Table 2**). Low mc and mh values (mh lower 95% CI includes zero) suggest that populations have become demographically independent in the past. Despite the reduced geographic distance, historical rates of migration between Palinuro and Strombolicchio have the lowest values. Some hidden environmental barriers (likely North heading sea currents) may have impeded any past migration between these two populations.

The *E. saxicola* dispersion was inferred to have occurred by temporally separated introductions from the Palinuro to Capri and by the latter, to Strombolicchio as revealed by DIYABC model selection, scenario 1 (**Figure 7**). This scenario was supported by the model (votes = 781; posterior probability $p = 0.501$). The alternative scenario (2) where Strombolicchio originated from Palinuro received 231 votes. Instead, the scenarios (3 and 4) suggesting the split of an ancestral population between Capri and Palinuro and the subsequent colonization of Strombolicchio from Capri or Palinuro received 308 and 180 of the votes, respectively.

The colonization of Capri from the mainland population was predicted to have occurred around 24,615 generations ago while the following colonization of Strombolicchio only occurred 9,003 generations ago. Interestingly, the relatively narrow posterior parameter distribution of the population introduction from the mainland Palinuro to the Capri Island suggests this single event was followed by a bottleneck period (db) of at least 43 generations with a relatively small founding population of 10 individuals. The bottleneck size of the Strombolicchio population (Nb-Strombolicchio) was around nine individuals. However, at time 0, the effective population size of Strombolicchio is 32 individuals, while the effective population size of Capri is 369 (249–484), i.e., comparable with 379 (150–483) of Palinuro (**Table 3**).

DISCUSSION

The integration of different nuclear and plastid genetic markers provided novel insight into the temporal and spatial distribution, a colonization pathway, and dispersal patterns of *E. saxicola*. In particular, both markers contributed to disclosing the roles of contemporary versus historical processes in shaping the current genetic variation of *E. saxicola* populations. Despite its paleoendemic status, the present distribution and partition of genetic diversity in *E. saxicola* revealed a rather recent dispersion/fragmentation, most likely occurring during the last quaternary climatic oscillations, which resulted in the almost complete interruption of ongoing gene flow among living populations.

It is well-known that different filtering strategies influence both the estimation of genetic diversity and differentiation (Cozzolino et al., 2020; Gargiulo et al., 2021). Therefore,

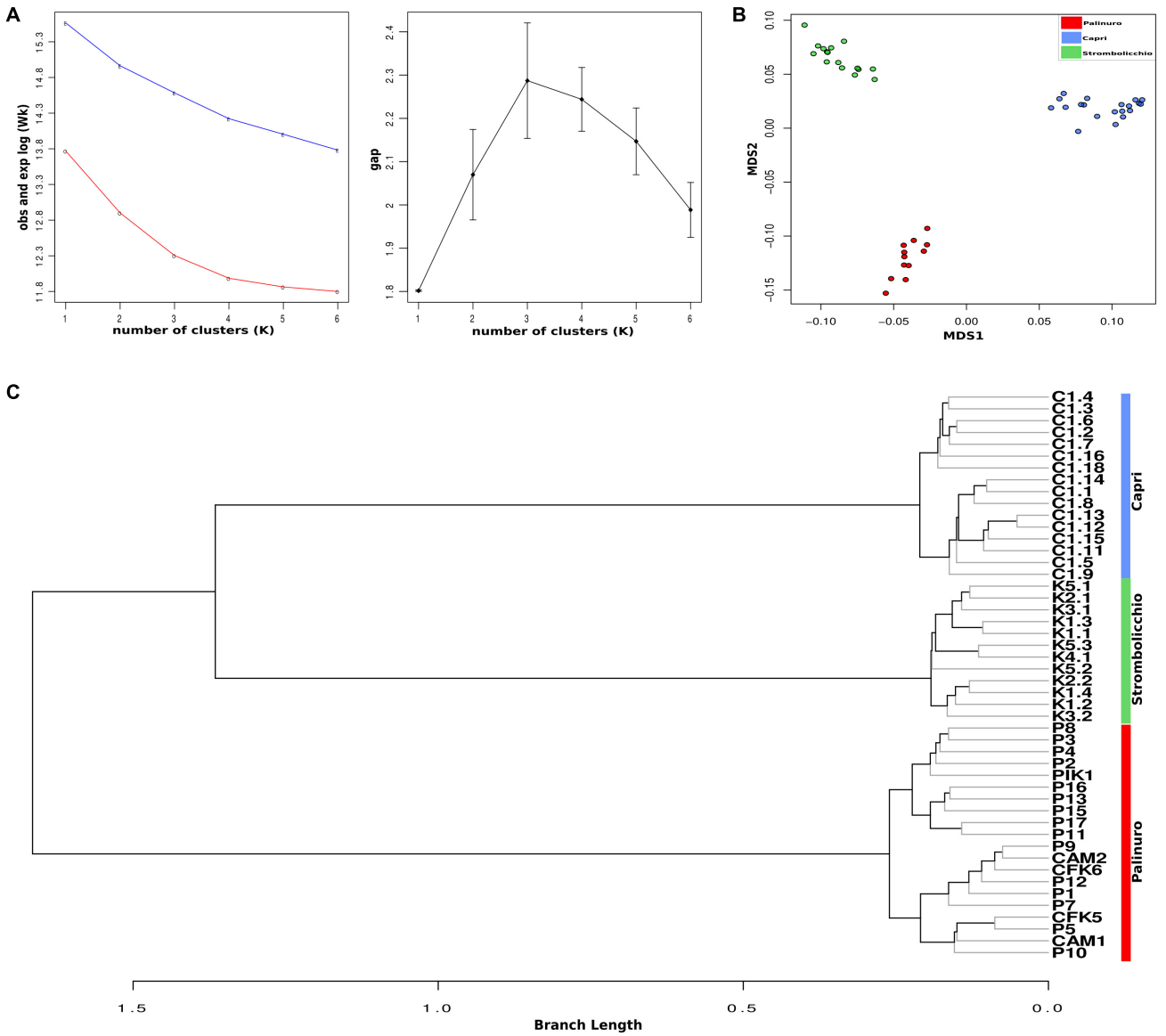


FIGURE 2 | Genetic clustering of sampled individuals. **(A)** Left: The red line represents the log of the pooled within-cluster sum of squares (Wk) from the observed data. The blue line represents the expected log of Wk from a uniform distribution. Right: the actual gap statistic. **(B)** A clustering by size reduction (an MDS plot) is represented. Individuals are distributed as points in a two-dimensional vector space. **(C)** The ancestry-inferred tree with AWclust.

for testing the robustness and reliability of our results using the ML tree, pairwise Fst and co-ancestry among individuals and populations were calculated both with the full dataset (3,962 SNPs across 48 individuals, SNPs present in at least 70% individuals) and with the reduced dataset (25 individuals by selecting 120 SNPs shared by at least 95% of individuals). We found largely overlapping results between the two datasets (**Supplementary Figures 1–3**). Still, as expected, when using different stringency criteria in loci selection, when comparing datasets with SNPs present in 70% vs. 95% of the samples, we found a decrease of between-lineage differentiation of the employed loci as measured by global F_{ST} and an increase in the average overall estimated co-ancestry (**Supplementary Text 1.1**. and **Supplementary Figure 4**). Thus, our conservative approach may have not precisely estimated the values of some population genetic parameters (e.g., heterozygosity, F_{ST}, Ne, and migration rates).

Do Populations Show Any Phylogeographic Structure?

Analysis of full plastid genome variation in *E. saxicola* revealed a surprisingly low level of intraspecific variation (only three haplotypes differing for a maximum of two mutation steps) among the three isolated populations with a considerable number of differences (in total plastome length too) with the

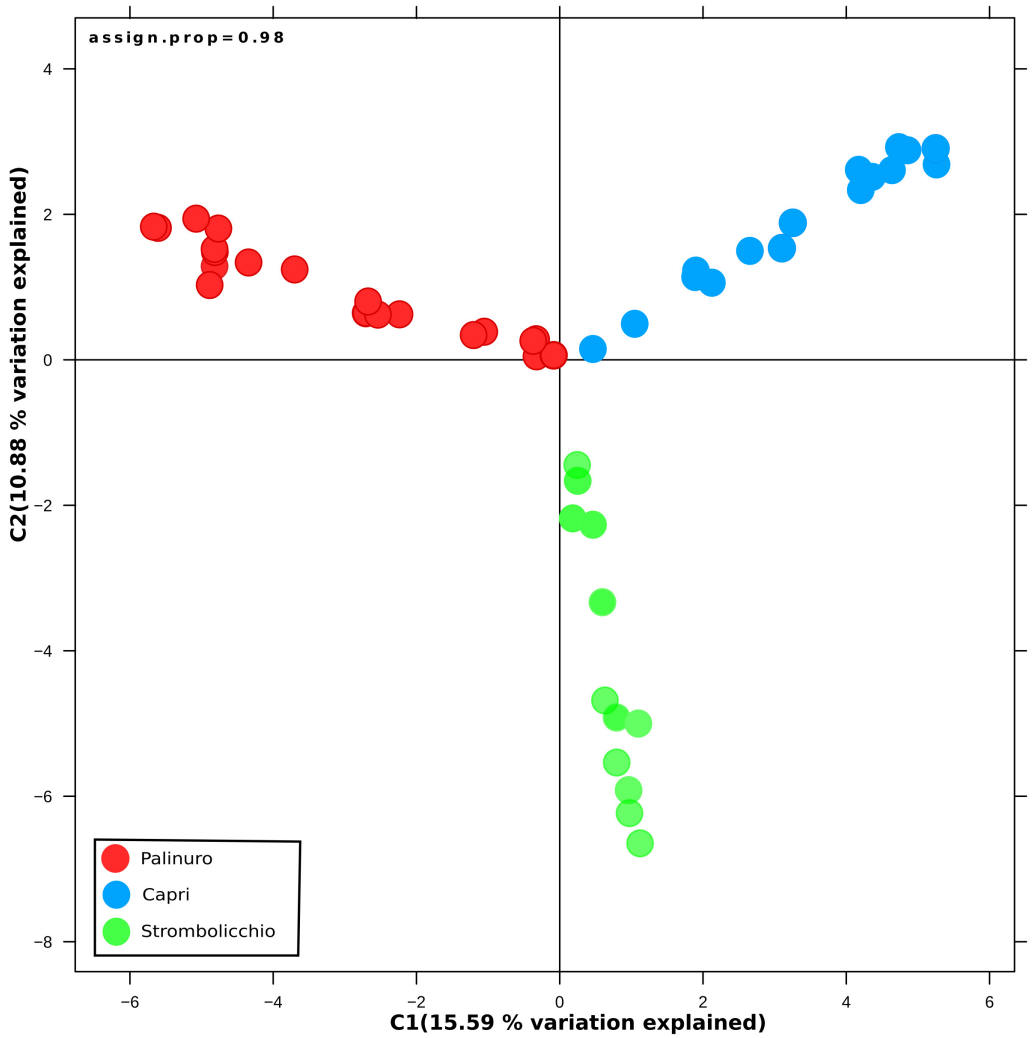


FIGURE 3 | Discriminate analysis of principal components (DAPC). The analysis was drawn using 3,962 SNPs across 48 individuals (SNPs present in at least 70% individuals) and was constructed using 20 principal components (PCs) and two discriminate functions. Dots represent individuals, with colors denoting sampling origin.

related *S. hirsuta*. Since large interspecific differences can be expected between paleoendemic species separated approximately 10 Mya (cfr. Kadereit and Freitag, 2011), both haplotype admixture in populations and very low haplotype variation at the intraspecific level support a recent separation of current *E. saxicola* populations. The Palinuro and Capri populations share two haplotypes (H1 and H2), differing by a single-base mutation (**Figure 1**). The most common haplotype (H2) was found in the Strombolicchio population, but, here, the "distant" haplotype H3 (i.e., differing by two mutational steps) was also detected (**Figure 1**). Even if the scenario of dispersal from Palinuro is the best supported, actual co-occurrence of two related haplotypes in both insular Capri and mainland Palinuro does not rule out an alternative scenario where these populations were likely part of a larger, ancestral population that later split into today isolated populations without a bottleneck phase. Interestingly, this scenario is the second-best supported scenario (i.e., Scenario 3) in the DYABC analysis based on nuclear markers (discussed below).

What Are the Primary Factors That Have Shaped the Phylogeographic Structure?

According to the morpho-bathymetric reconstruction of the Southern Tyrrhenian Sea (Aiello et al., 2015) and considering that, during the last glacial phases, the sea level decreased up to 100 meters, Capri was once connected to the mainland cliffs of Southern Italy. Moreover, the coastline shape between Capri and Palinuro drastically changed during the Quaternary due to the combination of tectonics, sedimentary inputs, volcanism, and sea-level oscillations (Santangelo et al., 2017). In this scenario, which occurred in the last 1.8 Mya, both frequency and surface extension of unsuitable (defined as sandy coasts related to alluvial plains) and suitable (defined as rocky coasts related

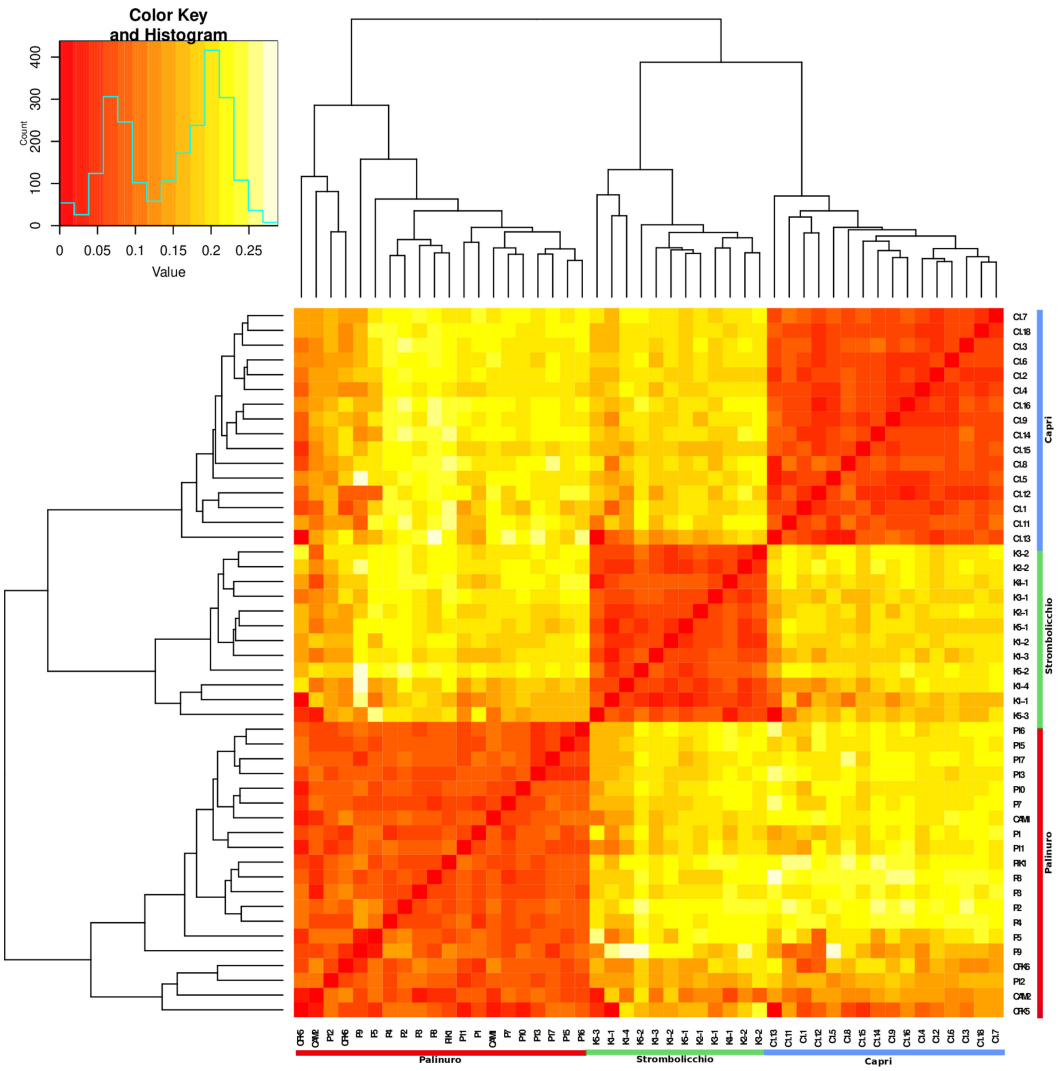


FIGURE 4 | A heatmap graph of Fst was built by using 3,962 SNPs across 48 individuals (SNPs present in at least 70% of individuals). The heatmap illustrates the Fst pairwise matrix based on a color code: orange/red color (low Fst) to white (high Fst). Along the axes of the matrix, phylogenetic trees are constructed using the Neighbor Joining method. The legend (top left) depicts the color-coded Fst values.

to pre-quaternary geological substrates) habitats for *E. saxicola* were transformed from the sea-level oscillations. Moreover, these oscillations most likely reduced the "jumping" distance from the Palinuro cliffs to Capri, favoring a linear "hopping" migration along the coastline, as also reported for several other Mediterranean halophytes (Clausing et al., 2000; Kadereit et al., 2005; Weising and Freitag, 2007). The current distribution of *E. saxicola* along the Palinuro cliffs (ca. < 15 km) further supports the evidence of a "hopping" dispersal ability along with these coastal rocky habitats.

In contrast, the relatively young volcanic islet of Strombolicchio (0.2 Mya) has been completed isolated from any mainland, even during the Quaternary seal-level oscillations. Accordingly, long distance dispersal or "jumping" is the only plausible scenario for the colonization of *E. saxicola* on Strombolicchio. Surprisingly, despite the rarity of *E. saxicola* populations and the small size of the islet, at least two occasional events (two different plastid haplotypes) of colonization likely occurred as founder events for the Strombolicchio population (**Figure 1**). Hydrochory, as for many other halophytes of the Mediterranean Sea (Kadereit et al., 2005; Arafeh and Kadereit, 2006; Weising and Freitag, 2007; De Castro et al., 2020), may have contributed to the long-distance dispersal of *E. saxicola* seeds. Indeed, sea-drifted seeds depend entirely on currents for their dispersal, and there is some evidence from the surface circulation observed in the southern Tyrrhenian Sea (Vetrano et al., 2004; Rinaldi et al., 2010) that is consistent with the known distribution of *E. saxicola*. Changes in sea currents due to the presence of a vortex north of Capri renders the north and south Tyrrhenian Sea "isolated" from each other (Procaccini et al., 2001) and impedes seeds dispersal toward the north.

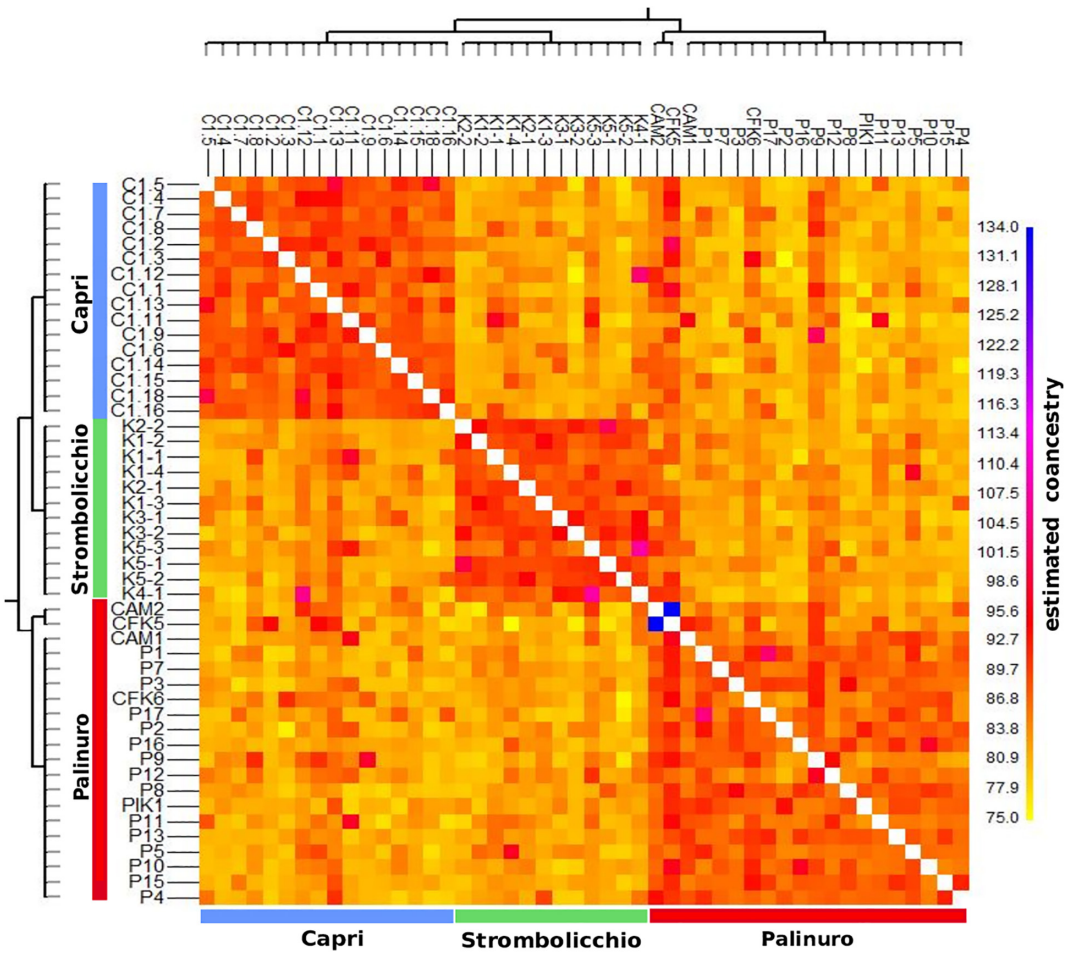


FIGURE 5 | The co-ancestry matrix was shared among 48 individuals (SNPs present in at least 70% individuals). The co-ancestry is calculated as the number of different sequences (i.e., SNPs) between pairs of individuals. The color of each cell in the matrix indicates the expected number of chunks imported from a donor genome (horizontal axis) into a recipient genome (vertical axis). Therefore, highly different cells are indicated in yellow, while indistinguishable cells are represented by blue/purple. Lastly, two phylogenetic trees with the Neighbor-Joining method are represented along the *x* and *y* axes.

According to the average reported speed of south Tyrrhenian surface circulation (Rinaldi et al., 2010), the floating time ability of the *E. saxicola* (Barone Lumaga et al., 2016) would have been sufficient to reach Strombolicchio islet from the Palinuro/Capri coastlines. Haplotype admixture between Capri and Palinuro does not allow detecting the source population, but nuclear data point to the Capri population as more proximate to the Strombolicchio one. Nevertheless, the presence of a second haplotype in the young islet, assuming that this haplotype may not have locally evolved or that it was present in Palinuro/Capri and gone undetected/extinct, points to a second colonization event from an unknown source population. This latter hypothesis is supported by the distinctiveness of exclusive Strombolicchio haplotype compared with those found in Palinuro/Capri (i.e., two mutation steps). The presumed source population, likely distributed along the south Tyrrhenian coastline, could be extinct today or even still existing. Despite the long history of floristic investigations in the Mediterranean Basin, the existence of unknown populations of *E. saxicola* could not be excluded due to the difficulty in recognizing the species (Strumia et al., 2015). Implementation of botanical surveys along the South Tyrrhenian coastlines, considering the Sea surface circulation and suitable rocky habitats, could fill this gap. Furthermore, a previous study on the genus *Limonium* showed that rising sea levels during the warm phases of the glaciation had led to the steady decline of coastal populations (Koutroumpa et al., 2021). We, therefore, cannot exclude that this flooding may have played a similar

TABLE 2 | Estimates of genetic diversity parameters over 25 individuals (SNPs present in at least 95% individuals) implemented with function basic.stats in the hierfstat R package.

Population	Ho	He	Fis
Capri	0.0982	0.1071	0.0833
Strombolicchio	0.0612	0.0704	0.1307
Palinuro	0.0929	0.1048	0.2343
Mean	0.0841	0.0941	0.0781

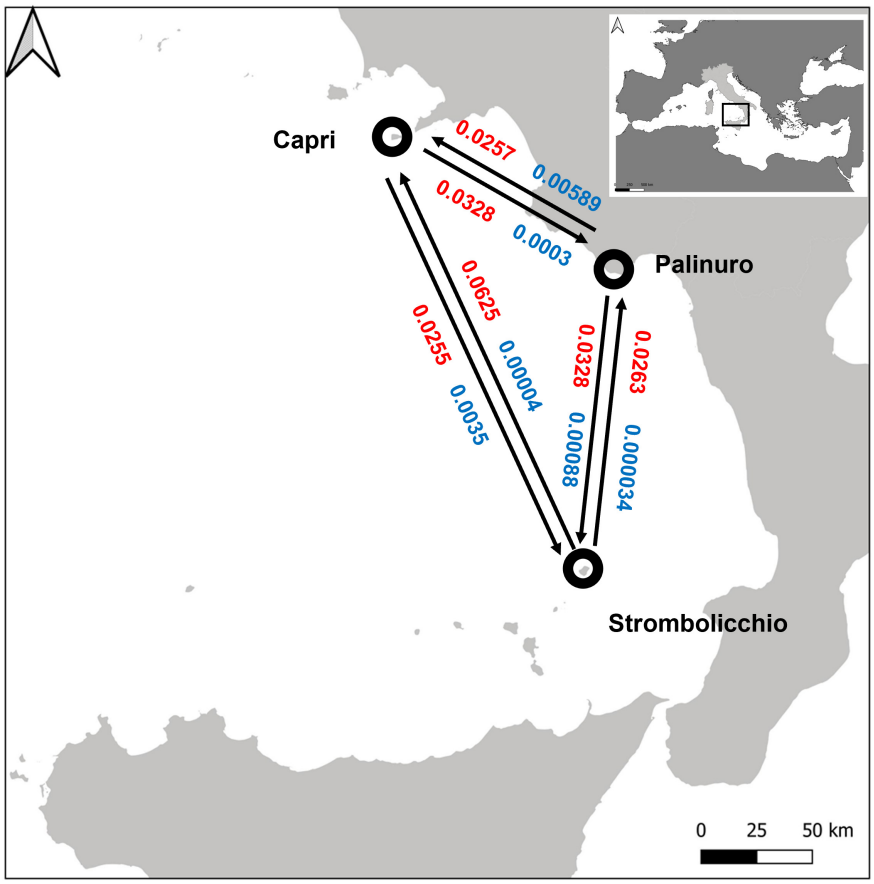


FIGURE 6 | Estimates of the mean posterior recent (blue) and historical (red) migration rate among 25 individuals (SNPs present in at least 95% individuals). The circles represent populations and arrows indicate the direction of migrant individuals between different populations. The inset plot at the top right of the figure reports the position of the map in the Mediterranean Basin.

role in the extinction of some *E. saxicola* populations along the coastline.

What Determines the Genetic Diversity and Genetic Structure of Extant Populations?

The complete analysis of plastid genome variation revealed partial admixture and a very low level of between population differentiation (only 1–2 mutation steps). The different analyses of nuclear markers, performed with different settings, concordantly shows three main genetic clusters matching the three sampling localities of Palinuro, Capri, and Strombolicchio (**Figure 2C**). In contrast to plastidial admixture, the nuclear markers revealed a present story of strong genetic isolation for the three living populations (**Figures 2–4**; **Table 3**). Indeed, all sampled individuals have been correctly assigned to their source population in all analyses, thus almost indicating a complete absence of genetic admixture and ongoing gene flow (with corresponding high Fst values).

MIGRATE and BayesAss estimates revealed that historical (mh) and contemporary (mc) levels of migration between the three genetically distinct populations are very small and similar in magnitude (**Supplementary Table 2**), as also supported by the large generation times estimated by the simulation performed in DIYABC. The low levels of contemporary migration are not surprising because geographic distance and cliff habitat make dispersal between populations highly unlikely. However, as historical migration rates are also very low, concordant patterns strongly imply that the high levels of genetic structure currently observed among extant populations stem from a low colonization ability, which thus seems to be a long-standing life-history trait of *E. saxicola*. One evolutionary implication of similar low contemporary and historical migration rates over time is that *E. saxicola* has a long history of living in relatively small, isolated populations (still preserving enough genetic variation). In contrast to other island narrow endemics (Molins et al., 2009; Mayol et al., 2012; López-Pujol et al., 2013), for which the sea level oscillations represented a switch-off-switch-on the barrier to gene flow, sea levels theoretically do not represent a major barrier for seed dispersal of *E. saxicola*, as well as for other halophytes of both sandy and rocky coasts. Nonetheless, phases of sea lowering could have led to both the increase in available habitat through land emersion and the promotion of

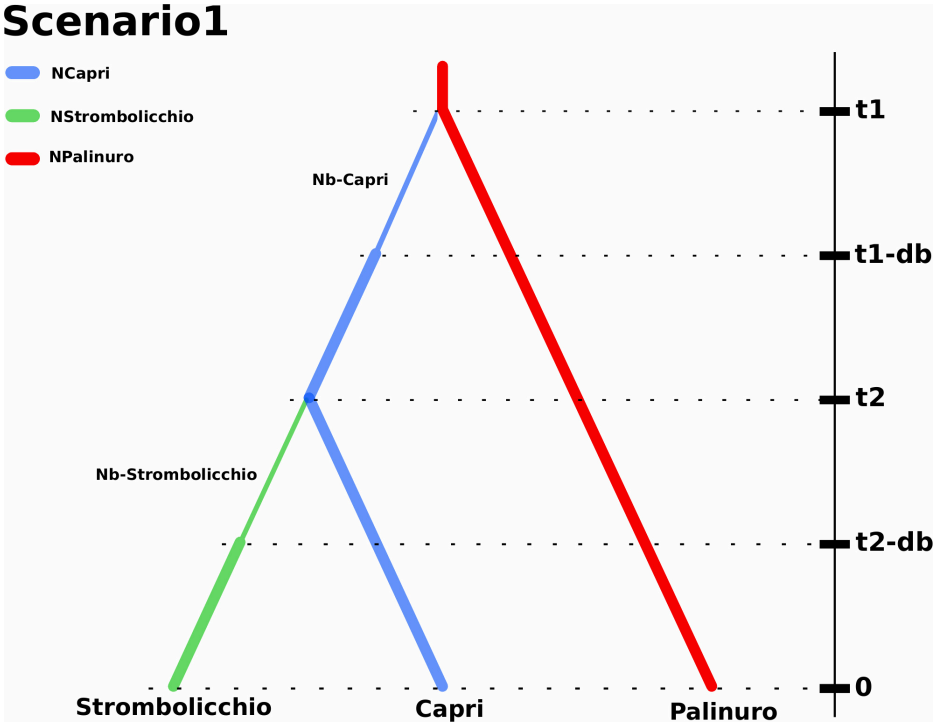


FIGURE 7 | Demographic history of *Eokochia saxicola* populations implemented by DIY ABC. Scenario 1 represents the most likely hypotheses of the demographic history: the Strombolicchio population was derived from the Capri population, which originated from the Palinuro population. The branch colors indicate discrete population size parameters in the model. t1 represents the split time between the Palinuro population and Capri population, while t2 represents the split time between the Capri population and Strombolicchio population. The thin branch width indicates bottlenecks of duration db (t1-db and t2-db) with effective population sizes of Nb (Nb-Capri and Nb-Strombolicchio). Note: Time is not to scale.

TABLE 3 | Prior values (minimum and maximum, with uniform distribution) for the parameters used for the demographic scenarios and posterior values (mean; median; quantile 5%, and 95% and variance) were estimated from Scenario 1 in the DIYABC approach.

	Parameter	Prior parameters		Posterior parameters				
		minimum	Maximum	mean	median	quantile 5%	quantile 95%	variance
Effective population size	Ncapri	10	500	369	371.25	249	484	9242.41
	NStrombolicchio	10	50	32	30.9	15	46	1053.76
	NPalinuro	100	500	379	381.03	150	483	4228.4
	N1b	5	50	10	8	3	18	421.353
	N2b	5	50	9	924.378	2	18	317.939
Time scale in generations	t1	10	50.000	24615	13035	17217.3	47268	8.22E + 07
	t2	10	50.000	9003	10096.1	4261	30256	3.98E + 07
	db	10	100	43	37	12	96	1446.28

hopping colonization by the decrease in distance between isolated populations. However, phases of the current sea rising do not totally halt the potential dispersion of *E. saxicola* but decrease the probability of its occurrence due to the increased distance between a few suitable sites.

The DIYABC analysis identified two introductions temporally separated from the Palinuro to Capri, and from there to Strombolicchio (DIYABC model selection, Scenario 1) as the most likely hypothesis for *E. saxicola* presents distribution. By assuming a generation time of 10 years, we can consider that the introduction to Capri, approximately 0.25 Mya, and the colonization of Strombolicchio (0.09 Mya) took place around the late Pleistocene. The presence of inaccessible known plant spots (on Capri) and the possible occurrence of other unknown plant spots along the Palinuro cliffs suggest we probably have underestimated the census population sizes (N) for the three populations. Still, Ne estimations are near and even larger than N (**Table 3**), i.e., most of the individuals are breeding individuals in the population. Overall, the effective population size of *E. saxicola* is historically low because climatic oscillations during the quaternary glacial-interglacial cycles periodically forced populations to track their optimum into smaller cliff

areas of suitable habitat. The almost exclusive preference of *E. saxicola* for north-facing cliffs suggests an adaptation to cooler conditions, which further highlights the possibility of different phases of population expansion/contraction during these oscillation cycles.

Can the Combination of Phylogeography and Genetic Diversity Help Define *E. saxicola* Conservation Priorities?

While geographic isolation seems the obvious explanation for the current population genetic structure, it partly contrasts with the levels of genetic diversity still detected within each population. Such genetic diversity within populations is unlikely to have resulted from recent or contemporary gene flow but rather points at relative demographic stability (as suggested by Ne estimations) both in recent and ancient timescales. All sampled individuals, despite a clumped distribution due to the accessibility of the cliff faces, were genetically different, indicating that clonality in this species is rare. Moreover, at least at the intrapopulation level, the long lifespan of adult plants, together with their high resprouting ability and wind pollination, may contribute to maintaining a high level of outcrossing and multiple random mating among the few individuals (i.e., observed heterozygosity was comparable to the expected one). Interestingly, wind pollination is an almost unique trait among Mediterranean narrow endemics. Indeed, anemophily is a plesiomorphic trait in *E. saxicola* and Amarantaceae, whereas, in the few other known wind-pollinated endemics, this mechanism is rather a secondary adaptation to the low availability of biotic pollination (Traveset and Navarro, 2017).

Regardless of the extremely small number of individuals in each population, the moderate degree of intrapopulation co-ancestry (**Figure 5**) suggests that a consistent level of crosses among unrelated individuals occurs. Even if the reproductive success of individual plants is unknown, the pattern of genetic relatedness among individuals of these small populations suggests that most of the adult plants contribute to the next generation, i.e., that random mating occurs. This reproductive behavior is consistent with the observation of Ho not different from He and a moderate excess of homozygotes in all populations compared with Hardy–Weinberg equilibrium expectations (Fis average, 0.10). The relative excess of homozygotes (higher Fis) observed in Palinuro denotes some heterogeneity in this mainland population compared to the islets, consisting of populations composed of several subgroups (the so-called Wahlund effect). Indeed, in contrast to the other two insular populations sampled on single cliffs, the Palinuro population was sampled along different cliffs. This subpopulation structure is also partially evident in the heatmap graph of Fst (**Figure 4**).

As reported for other extremely narrow endemics (see Médail and Baumel, 2018), *E. saxicola* still maintains some degree of genetic diversity even at the intrapopulation level (despite its small population sizes); this genetic trait may have represented life insurance for the long-term survivorship of this rare species. This survivorship was further aided by the low level of anthropogenic disturbance and high stability typical of cliff plant communities (Davis, 1951). Indeed, maintaining genetic diversity can reduce the probability of extinction of small populations by providing the standing genetic variation for local adaptation (Leimu and Fischer, 2008). Therefore, the present Palinuro population still holds enough genetic diversity to represent a valuable "donor population" for future reintroduction programs of this species. The high heterozygosity can be relevant in terms of conservation guidelines as genetic diversity of *E. saxicola* can be preserved in the long term even in its current isolated populations, confirming that narrow endemics are not necessarily "evolutionary dead-ends," but rather may represent species that have a strong evolutionary legacy (Médail and Baumel, 2018). From a conservation perspective, we suggest that, at least in the short term, genetic factors may have little impact on the persistence of small populations of this rare paleoendemic species, but, rather, ecological factors (as persistence of suitable habitat) may play a larger role in determining whether populations survive in the long term. In this respect, both the IUCN category and the potential threats of extinction of this species (Santangelo et al., 2012; Orsenigo et al., 2018) should be reconsidered.

CONCLUSION

Even if short and long seed dispersion is still possible, why *E. saxicola* is so rare compared with other littoral halophytes species (Davis, 1951)? Indeed, the strong genetic differentiation detected among living populations suggests that the long-distance colonization typical of halophytes is merely occasional. We recognize that, independently of its colonization potential (through hydrochory), the rarity of this species must be considered also in terms of its specific requirements for habitat type as well as its biological features. Currently, *E. saxicola* has distributed exclusively on north facing (± 45° deviation from North) rocky shores (Strumia et al., 2015), but we do not have experimental evidence to support this restricted distribution. Since long-distance dispersal and colonization are different processes [see Feliner (2014) and references therein], we hypothesize that other factors determine the successful long-distance colonization in *E. saxicola*: namely, landing in a suitable microhabitat, seed germination, and, finally, seedling anchorage (Strumia et al., 2020b). Thus, the rarity of *E. saxicola* must be attributed not only in terms of the longevity and the ability of seeds to float but also, if not predominantly, of the habitat suitability of reached cliffs. These findings confirm the distinctiveness of *E. saxicola* from other halophytes and narrow endemics of the Mediterranean Basin. In conclusion, *E. saxicola* shows habitat specificity with individuals found predominately in small and sporadic populations. However, anemophilous pollination maintains a high-genetic diversity in this species, thus preventing the rapid decline, which is usually expected in small populations, thus ruling out the possibility of genetic impoverishment and inbreeding depression as immediate causes of threat.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: BioProject PRJNA756897.

AUTHOR CONTRIBUTIONS

SS, AS, and DC: conceptualization. SS and AS: plant material collection. SC and DC: data curation. SC, DC, and TG: formal analysis. DC and SS: funding acquisition. DC and TG: methodology. SS, SC, DC, and AS: writing – original draft preparation. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research and the APC were partially funded by a grant provided by Regione Campania – Direzione Generale per l'Ambiente e l'Ecosistema – Dip. 50 DG 06 U.O.D. 07.

ACKNOWLEDGMENTS

The authors thank the Cilento and Vallo di Diano and Alburni National Park and the Sicily Region – Regional Department of Rural and Land Development – Messina Office, Service 15 that authorized the activities in the field. The authors are extremely grateful to Mariacristina Villani for collecting *S. hirsuta* and to Antonio Federico for his kind support on the Capri Island and Giovanni Cammarano, Biagio Fedullo, and Diego Errico for their help in collecting plant materials in Palinuro. The authors are grateful to Marco Balducci for language revision, Simone Fior, and Nunzio D'Agostino for their comments on the manuscript. The authors are extremely grateful to two reviewers for their constructive criticism and helpful comments that significantly improved the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.737111/full#supplementary-material

REFERENCES

Aiello, G., Marsella, E., and D'Insanto, C. (2015). New morpho-bathymetric and tectono-stratigraphic data on Naples and Salerno Gulfs (Southern Tyrrhenian Sea, Italy) derived from bathymetric and sismic data analysis and integrated geological interpretation. *Curr. Dev. Oceanogr.* 8, 1–42. doi: 10.3390/geosciences10080319

Arafeh, R., and Kadereit, J. W. (2006). Long-distance seed dispersal, clone longevity and lack of phylogeographical structure in the European distributional range of the coastal *Calystegia soldanella* (L.) R. Br. (Convolvulaceae). *J. Biogeogr.* 33, 1461–1469. doi: 10.1111/j.1365-2699.2006.01512.x

Barone Lumaga, M. R., Santangelo, A., and Strumia, S. (2016). Morpho-functional traits influencing the fitness of highly endangered *Eokochia saxicola* (Guss.) Freitag & G.Kadereit (Amaranthaceae). *Flora* 218, 11–17. doi: 10.1016/j.flora.2015.11.005

Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol.* 41, 379–406. doi: 10.1146/annurev-ecolsys-102209-144621

Beerli, P. (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22, 341–345. doi: 10.1093/bioinformatics/bti803

Bermingham, E., and Moritz, C. (1998). Comparative phylogeography: concepts and applications. *Mol. Ecol.* 7, 367–369. doi: 10.1046/j.1365-294x.1998.00424.x

Blondel, J., Aronson, J., Bodiou, J., and Boeuf, G. (2010). *The Mediterranean Region: Biological Diversity in Space and Time*, 2nd Edn. New York, NY: Oxford University Press.

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

Chapuis, M. P., Raynal, L., Plantamp, C., Meynard, C. N., Blondin, L., Marin, J. M., et al. (2020). A young age of subspecific divergence in the desert locust inferred by ABC random forest. *Mol. Ecol.* 29, 4542–4558. doi: 10.1111/mec.15663

Clausing, G., Vickers, K., and Kadereit, J. W. (2000). Historical biogeography in a linear system: genetic variation of Sea Rocket (*Cakile maritima*) and Sea Holly (*Eryngium maritimum*) along European coasts. *Mol. Ecol.* 9, 1823–1833. doi: 10.1046/j.1365-294x.2000.01083.x

Coissac, E., Hollingsworth, P. M., Lavergne, S., and Taberlet, P. (2016). From barcodes to genomes: extending the concept of DNA barcoding. *Mol. Ecol.* 25, 1423–1428. doi: 10.1111/mec.13549

Collin, F. D., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., et al. (2021). Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. *Mol. Ecol. Resour.* doi: 10.1111/1755-0998.13413 [Epub ahead of print].

Cozzolino, S., Scopece, G., Roma, L., and Schlüter, P. M. (2020). Different filtering strategies of genotyping-by-sequencing data provide complementary resolutions of species boundaries and relationships in a clade of sexually deceptive orchids. *J. Syst. Evol.* 58, 133–144. doi: 10.1111/jse.12493

Danecek, P., Auton, A., Abecasis, G., Cornelis, G., Albers, C. A., Banks, E., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Davis, P. H. (1951). Cliff vegetation in the Eastern mediterranean. *J Ecol* 39, 63–93. doi: 10.2307/2256628

De Castro, O., Innangi, M., and Menale, B. (2020). Message in a bottle: the Mediterranean Sea currents acted as protagonists in shaping the distribution of the sea daffodil (*Pancratium maritimum*, Amaryllidaceae). *Bot. J. Linn. Soc.* 194, 207–220. doi: 10.1093/botlinnean/boaa037

Diniz-Filho, J. A. F., Campos Telles, M., Bonatto, S., Eizirik, E., De Freitas, T. R., De Marco, P. Jr., et al. (2008). Mapping the evolutionary twilight zone:molecular markers, populations and geography. *J. Biogeogr.* 35, 753–763. doi: 10.1111/j.1365-2699.2008.01912.x

Favarger, C., and Contandriopoulos, J. (1961). Essai sur l'endémisme. *Biol. Soc. Bot. Suisse* 71, 384–408.

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

Gao, X., and Starmer, J. D. (2008). AWclust: point-and-click software for non-parametric population structure analysis. *BMC Bioinformatics* 9:77. doi: 10.1186/1471-2105-9-77

Gargiulo, R., Kull, T., and Fay, M. F. (2021). Effective double-digest RAD sequencing and genotyping despite large genome size. *Mol. Ecol. Resour.* 21, 1037–1055. doi: 10.1111/1755-0998.13314

Gillot, P. Y., and Keller, J. (1993). Age dating of Stromboli. *Acta Vulcanol.* 3, 69–77.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* 5, 184–186.

Greuter, W. (1991). Botanical diversity, endemism, rarity, and extinction in the Mediterranean area: an analysis based on the published volumes of Med-Checklist. *Bot. Chron.* 10, 63–79.

Heuertz, M., Fineschi, S., Anzidei, M., Pastorelli, R., Salvini, D., Paule, L., et al. (2004). Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in Europe. *Mol. Ecol.* 13, 3437–3452. doi: 10.1111/j.1365-294X.2004.02333.x

Hewitt, G. M. (2001). Speciation, hybrid zones and phylogeography: or seeing genes in space and time. *Mol. Ecol.* 10, 537–549. doi: 10.1046/j.1365-294x.2001.01202.x

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

Kadereit, G., and Freitag, H. (2011). Molecular phylogeny of Camphorosmeae (Camphorosmoideae, Chenopodiaceae): implications for biogeography, evolution of C4 photosynthesis and taxonomy. *Taxon* 60, 51–78. doi: 10.1002/tax.601006

Kadereit, J. W., Arafeh, R., Somogyi, G., and Westberg, E. (2005). Terrestrial growth and marine dispersal? Comparative phylogeography of five coastal plant species at a European scale. *Taxon* 54, 861–876. doi: 10.2307/25065567

Kadereit, J. W., and Westberg, E. (2007). Determinants of phylogeographic structure: a comparative study of seven coastal flowering plant species across their European range. *Watsonia* 26, 229–238.

Koutroumpa, K., Warren, B. H., Theodoridis, S., Coiro, M., Romeiras, M. M., Jiménez, A., et al. (2021). Geo-Climatic changes and apomixis as major drivers of diversification in the Mediterranean Sea Lavenders (*Limonium* Mill.). *Front. Plant Sci.* 11:612258. doi: 10.3389/fpls.2020.612258

Kruckeberg, A. R., and Rabinowitz, D. (1985). Biological aspects of endemism in higher plants. *Annu. Rev. Ecol. Syst.* 16, 447–479. doi: 10.1146/annurev.es.16.110185.002311

Lavergne, S., Thompson, J. D., Garnier, E., and Debussche, M. (2004). The biology and ecology of narrow endemic and widespread plants: a comparative study of trait variation in 20 congeneric pairs. *Oikos* 107, 505–518. doi: 10.1111/j.0030-1299.2004.13423.x

Leigh, J. W., and Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. doi: 10.1111/2041-210X.12410

Leimu, R., and Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS One* 3:e4010. doi: 10.1371/journal.pone.0004010

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

López-Pujol, J., Martinell, M. C., Massò, S., Blanché, C., and Sáez, L. (2013). The 'paradigm of extremes': extremely low genetic diversity in an extremely narrow endemic species, *Coristospermum huteri* (Umbelliferae). *Plant Syst. Evol.* 299, 439–446. doi: 10.1007/s00606-012-0732-3

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

Mayol, M., Palau, C., Rosselló, J. A., González-Martínez, S. C., Molins, A., and Riba, M. (2012). Patterns of genetic variability and habitat occupancy in *Crepis triasii* (Asteraceae) at different spatial scales: insights on evolutionary processes leading to diversification in continental islands. *Ann. Bot. London* 109, 429–441. doi: 10.1093/aob/mcr298

Médail, F. (2017). The specific vulnerability of plant biodiversity and vegetation on Mediterranean islands in the face of global change. *Reg. Environ. Change* 17, 1775–1790. doi: 10.1007/s10113-017-1123-7

Médail, F., and Baumel, A. (2018). Using phylogeography to define conservation priorities: the case of narrow endemic plants in the Mediterranean Basin hotspot. *Biol. Conserv.* 224, 258–266. doi: 10.1016/j.biocon.2018.05.028

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

Médail, F., and Quézel, P. (1999). Hot-spots analysis for conservation of plant biodiversity in the Mediterranean Basin. *Cons. Biol.* 13, 1510–1513. doi: 10.1016/j.biocon.2018.05.028

Molins, A., Mayol, M., and Rosselló, J. A. (2009). Phylogeographical structure in the coastal species *Senecio rodriguezii* (Asteraceae), a narrowly distributed endemic Mediterranean plant. *J. Biogeogr.* 36, 1372–1383. doi: 10.1111/j.1365-2699.2009.02108.x

Montmollin, B. D., and Strahm, W. (eds) (2005). "The Top 50 Mediterranean Island plants: wild plants at the brink of extinction, and what is needed to save them," in *IUCN/SSC Mediterranean Islands Plant Specialist Group*, (Gland: IUCN).

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

Orsenigo, S., Montagnani, C., Fenu, G., Gargano, D., Peruzzi, L., Abeli, T., et al. (2018). Red Listing plants under full national responsibility: extinction risk and threats in the vascular flora endemic to Italy. *Biol. Cons.* 224, 213–222. doi: 10.1016/j.biocon.2018.05.030

Paris, J. R., Stevens, J. R., and Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods Ecol. Evol.* 8, 1360–1373. doi: 10.1111/2041-210X.12775

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and 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. doi: 10.1371/journal.pone.0037135

Procaccini, G., Orsini, L., Ruggiero, M. V., and Scardi, M. (2001). Spatialpatterns of genetic diversity in *Posidonia oceanica*, anendemic Mediterranean seagrass. *Mol. Ecol.* 10, 1413–1421. doi: 10.1046/j.1365-294x.2001.01290.x

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Quezel, P. (1985). "Definition of the Mediterranean region and the origin of its flora," in *Plant Conservation in the Mediterranean Area*, ed. C. Gomez Campo (Dordrecht: W. Junk), 9–24.

Rannala, B. (2007). *BayesAss, Version 3.0. User's Manual*. Davis: University of California.

Rinaldi, E., Buongiorno Nardelli, B., Zambianchi, E., Santoleri, R., and Poulain, P.-M. (2010). Lagrangian and Eulerian observations of the surface circulation in the Tyrrhenian Sea. *J. Geophys. Res.* 115:C04024. doi: 10.1029/2009JC005535

Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M. (2019). Stacks 2: analytical methods for paired-end sequencing improve radseq-based population genomics. *Mol. Ecol.* 28, 4737–4754. doi: 10.1111/mec.15253

Santangelo, A., Croce, A., Lo Cascio, P., Pasta, S., Strumia, S., and Troìa, A. (2012). *Eokochia saxicola* (Guss.) Freitag and G. Kadereit. *Inf. Bot. Ital.* 44, 428–431.

Santangelo, N., Di Donato, V., Lebreton, V., Romano, P., and Russo Ermolli, E. (2011). Palaeolandscapes of Southern Apennines during the late early and the Middle Pleistocene. *Quat. Int.* 267, 20–29. doi: 10.1016/j.quaint.2011.02.036

Santangelo, N., Romano, P., Ascione, A., and Russo Ermolli, E. (2017). Quaternary evolution of the Southern Apennines coastal plains. A review. *Geol. Carpath.* 68, 43–56. doi: 10.1515/geoca-2017-0004

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stebbins, G. L., and Major, J. (1965). Endemism and speciation in the *California flora*. *Ecol. Monogr.* 35, 1–35. doi: 10.2307/1942216

Straub, S. C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., and Liston, A. (2012). Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am. J. Bot.* 99, 349–364.

Strumia, S., Buonanno, M., Aronne, G., Santo, A., and Santangelo, A. (2020a). Monitoring of plant species and communities on coastal cliffs: is the use of unmanned aerial vehicles suitable? *Diversity* 12:149. doi: 10.3390/d12040149

Strumia, S., Santangelo, A., and Barone Lumaga, M. R. (2020b). Seed germination and seedling roots traits of four species living on Mediterranean coastal cliffs. *Plant Biosyst.* 154, 990–999. doi: 10.1080/11263504.2020.1837284

Strumia, S., Croce, A., and Santangelo, A. (2015). New distributional data of the rare endemic species *Eokochia saxicola* (Guss.) Freitag and G.Kadereit (Chenopodiaceae): effects on biogeography and conservation. *Plant Biosyst.* 149, 559–564. doi: 10.1080/11263504.2013.870246

Suc, J. P. (1984). Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature* 307, 429–432. doi: 10.1038/307429a0

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., and Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* 7, 453–464. doi: 10.1046/j.1365-294x.1998.00289.x

Thompson, J. D. (2005). *Plant Evolution in the Mediterranean*. Oxford: Oxford University Press.

Traveset, A., and Navarro, L. (2017). State of the art of the plant reproductive ecology and evolution in the Mediterranean Islands. *Plant Biol.* 20(Suppl. 1), 63–77. doi: 10.1111/plb.12636

Ungar, I. A. (1991). *Ecophysiology of Vascular Halophytes*. Boca Raton, FL: CRC Press.

Vetrano, A., Gasparini, G. P., Molcard, R., and Astraldi, M. (2004). Water flux estimates in the central Mediterranean Sea from an inverse box model. *J. Geophys. Res.* 109:C0101903. doi: 10.1029/2003JC001903

Vezzoli, L. (1988). *Island of Ischia*. Roma: Consiglio Nazionale delle Ricerche (CNR).

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wallace, A. R. (1880). *Island Life: or The Phenomena and Causes of Insular Faunas and Floras, Including a Revision and Attempted Solution of the Problem of Ecological Climates*. London: MacMillan and Co.

Walter, H. (1985). *Vegetation of the Earth and Ecological Systems of the Geobiosphere*, 3rd Edn. Berlin: Springer-Verlag.

Weising, K., and Freitag, H. (2007). Phylogeography of halophytes from European coastal and inland habitats. *Zool. Anz.* 246, 279–292. doi: 10.1016/j.jcz.2007.07.005

Westberg, E., and Kadereit, J. W. (2009). The influence of sea currents, past disruption of gene flow and species biology on the phylogeographical structure of coastal flowering plants. *J. Biogogr.* 36, 1398–1410. doi: 10.1111/j.1365-2699.2008.01973.x

Whittaker, R. J., and Fernández-Palacios, J. M. (2007). *Island Biogeography, Ecology, Evolutionand Conservation*, 2nd Edn. Oxford: Oxford University Press.

Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31, 3350–3352.

Wingett, S. W., and Andrews, S. (2018). FastQ screen: a tool for multi-genome mapping and quality control. *F1000Research* 7, 1338. doi: 10.12688/f1000research.15931.2

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 February 2022
doi: 10.3389/fpls.2022.826158

Novel Insights Into Refugia at the Southern Margin of the Distribution Range of the Endangered Species *Ulmus laevis*

Sara Torre[1†], Federico Sebastiani[1†], Guia Burbui[1], Francesco Pecori[1], Alessia L. Pepori[1], Iacopo Passeri[1], Luisa Ghelardini[2], Alberto Selvaggi[3] and Alberto Santini[1]*

[1] Istituto per la Protezione Sostenibile delle Piante, IPSP-CNR, Florence, Italy, [2] Dipartimento di Scienze e Tecnologie Agrarie, Alimentari Ambientali e Forestali (DAGRI), Università di Firenze, Florence, Italy, [3] Istituto per le Piante da Legno e l'Ambiente - I.P.L.A. S.p.A., Turin, Italy

OPEN ACCESS

Edited by:
Andrew A. Crowl, Duke University, United States

Reviewed by:
Aziz Ebrahimi, Purdue University, United States
An Vanden Broeck, Research Institute for Nature and Forest (INBO), Belgium

***Correspondence:**
Federico Sebastiani
federico.sebastiani@ipsp.cnr.it

† These authors have contributed equally to this work and share first authorship

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

Received: *30 November 2021*
Accepted: *14 January 2022*
Published: *15 February 2022*

Citation:
Torre S, Sebastiani F, Burbui G, Pecori F, Pepori AL, Passeri I, Ghelardini L, Selvaggi A and Santini A (2022) Novel Insights Into Refugia at the Southern Margin of the Distribution Range of the Endangered Species Ulmus laevis. Front. Plant Sci. 13:826158. doi: 10.3389/fpls.2022.826158

Riparian ecosystems, in long-time developed regions, are among the most heavily impacted by human activities; therefore, the distribution of tree riparian species, such as *Ulmus laevis*, is highly affected. This phenomenon is particularly relevant at the margins of the natural habitat of the species, where populations are small and rare. In these cases, it is difficult to distinguish between relics or introductions, but it is relevant for the restoration of natural habitats and conservation strategies. The aim of this study was to study the phylogeography of the southern distribution of the species. We sequenced the entire chloroplast (cp) genomes of 54 individuals from five sampled populations across different European regions to highlight polymorphisms and analyze their distribution. Thirty-two haplotypes were identified. All the sampled populations showed private haplotypes that can be considered an indicator of long-term residency, given the low mutation rate of organellar DNA. The network of all haplotypes showed a star-like topology, and Serbian haplotypes were present in all branches. The Balkan population showed the highest level of nucleotide and genetic diversity. Low genetic differentiation between populations was observed but we found a significant differentiation among Serbia vs. other provenances. Our estimates of divergent time of *U. laevis* samples highlight the early split of above all Serbian individuals from other populations, emphasizing the reservoir role of white elm genetic diversity of Serbian population.

Keywords: conservation genetics, phylogeography, *Ulmus laevis*, plastome sequencing, forest trees, refugia, genetic polymorphisms

INTRODUCTION

Modern biodiversity conservation strategies rely, or at least should, on the knowledge of genetic resources (Gepts, 2006; Hoban et al., 2021). Big efforts were made in the exploration of genetic resources in the last decades (Halewood et al., 2018). In Europe, several foundation works were conducted on plant and animal organisms (Toro and Caballero, 2005; Koskela et al., 2013;

Porth and El-Kassaby, 2014), exploring the geographical distribution and structure of genetic diversity and describing the patterns of reduction and expansion of populations mainly related to glaciation and recolonization, respectively (Petit et al., 2003; Heuertz et al., 2004; Mona et al., 2010; Lanier et al., 2015). In forest tree species, three main glacial refugia were identified in southern Europe, corresponding to the Iberian, Italian, and Balkan Peninsulas (Brewer et al., 2002; Heuertz et al., 2004; Magri et al., 2006), and post-glacial colonization routes of temperate species were hypothesized toward central and northern Europe (Petit et al., 2002; Magri, 2008). As a result of the expansion, a reduction in genetic diversity was expected from south to north (Gugerli et al., 2009), and this was partially confirmed, since most divergent populations were found in the Mediterranean areas, but the highest level of genetic diversity was observed in central Europe where the genetic lineages from different refugia got admixture (Petit et al., 2003). The picture that came out progressively gained complexity by considering the effect of preglacial variation distribution (Magri, 2008) and the existence of extra-Mediterranean refugia (Pedreschi et al., 2019). Despite these huge efforts, there are still species that are highly impacted by anthropogenic activity and by pests whose distribution is poorly investigated and that could represent a priority for habitat restoration.

European white elm (EWE) (*Ulmus laevis* Pall. = *Ulmus effusa* Willd. = *Ulmus pedunculata* Foug.) is a characteristic deciduous temperate forest tree species growing in river margins and lowland moist forests, and it can also tolerate dry soils of wooded steppe habitat in the central parts of the distribution. Seeds are generally anemochorous, but in the riparian habitats, hydrochory is an important means of transport, enabling long-distance dispersal (Vakkari et al., 2009). EWE is the only representative of the section *Blepharocarpus* in the Old World. *U. laevis* does not hybridize with other European species, namely, *Ulmus glabra* and *Ulmus minor* (Mittempergher and La Porta, 1991). Natural distribution range of EWE is mainly central and east European. It extends from Ural Mountains in the east to France in the west and from southern Finland in the north to Bosnia in the south (Svejgaard Jensen, 2003). There are even a few relict populations in south-western France (Timbal and Collin, 1999). Quite recently, some native relic scattered populations have been identified in the Iberian Peninsula (Fuentes-Utrilla et al., 2014). In Italy, it is generally considered allochthonous, naturalized in most regions (Pignatti et al., 2017) except in Piedmont (north–west) where it is considered native cryptogenic (Bartolucci et al., 2018) or native (Pignatti et al., 2017), in accordance with the hypothesis supported by regional botanic knowledge.

Elms in Europe are severely damaged by successive epidemic waves of Dutch elm disease (DED), a lethal disease caused by some non-native species of the genus *Ophiostoma*, namely, *Ophiostoma ulmi*, and *Ophiostoma novo-ulmi* ssp. *novo-ulmi* and *O. novo-ulmi* ssp. *americana* (Brasier, 2000). *U. laevis* is susceptible to the fungal agent of DED (*O. novo-ulmi*) (Santini et al., 2005), but it is able to avoid infections since it is far less attractive for the elm bark beetles, the main vector of the disease, than *U. minor* (Sacchetti et al., 1990). It is seldom attacked once other species are not present in an infested area anymore. Then, if the insect vectors do not change their feeding preferences, the genetic resources of *U. laevis* are not really endangered by DED (Collin, 2002). EWE populations are much more threatened by habitat destruction. Riparian forests, where EWE can be commonly found, are the sites of several human activities, namely, continuous cut for preventing floods, drained for reclaiming land for agriculture, and occasionally invaded by alien-human-introduced invasive plants. Therefore, the long-term survival of *U. laevis* throughout most of its distribution has been compromised (Franke et al., 2004). This has induced the generation of many small, isolated populations that result in vulnerability to genetic drift (Fuentes-Utrilla et al., 2014). In western Europe, the appropriate habitat for the species has been greatly reduced, although some small, isolated populations survive.

Typically, only populations that are recognized as native are included in the conservation process, while those that are considered introduced or have an uncertain status are not protected, which are excluded from regional Red Lists of the International Union for Conservation of Nature, even if they are rare and endangered (Decocq et al., 2004). The genetic tools used in most population genetic studies were based on different nuclear and molecular markers, mainly microsatellites (Morgante et al., 1996; Porth and El-Kassaby, 2014). The advent of high throughput sequencing offers from a few years the opportunity to explore genetic diversity at an unprecedented scale by sequencing portions to entire genomes and revealing unexplored genetic variation (Wang et al., 2001; Unamba et al., 2015; Ebrahimi et al., 2021).

Previous phylogeographic analyses in the genus *Ulmus* were based predominantly on neutral markers (Venturas et al., 2013; Zuo et al., 2020; Tamošaitis et al., 2021). In particular, the analysis of chloroplast (cp) markers in EWE populations (Whiteley, 2004) showed only three haplotypes, emphasizing the importance of exploring the entire genome. In recent years, the plastome analysis has facilitated molecular evolutionary studies because of its predominantly uniparental inheritance, small size, and slow nucleotide mutation rates (Wolfe et al., 1987), supplying the appropriate resolution frame to study plant phylogeography at deeper levels (Gitzendanner et al., 2018; Niu et al., 2018). Previous data on EWE genotyping highlighted the need to fill the gap on the southeast distribution of the species (Whiteley, 2004; Fuentes-Utrilla et al., 2014), even more so after the recent finding of an EWE stand in Italy, which was previously excluded from the natural distribution range. This work attempts to use whole plastid genome sequencing to better understand the phylogeography of *U. laevis* populations. The analysis of Serbian samples is critical to confirm the Balkans as the refugial origin of European populations.

Here, we aim to (a) explore the genetic structure of *U. laevis* at the south-western margin of its European distribution, (b) investigate whether Italian stands of white elms are natural or the result of anthropogenic introductions, and (c) discuss, in the light of new outcomes, the consequences for conservation and management strategies for this species.

MATERIALS AND METHODS

Plant Material and DNA Extraction

The samples of *U. laevis* were collected in four areas and included in the natural range of the species: two in France (southwest and northeast), one in Spain, and one in Serbia.[1] In addition, *U. laevis* samples were collected in northwestern Italy (**Supplementary Table 1**). Total genomic DNA was extracted from the leaves of each plant using the Invisorb Spin Plant Mini Kit (Invitek) following the instructions of the manufacturer.

Plastome Sequencing and Assembly

Using a genome skimming approach, all the samples were subjected to DNA shotgun sequencing. Genomic libraries were prepared and sequenced in paired-end mode (2 × 150 bp reads) on an Illumina HiSeq 4000 platform (Novogene Co., Ltd., Beijing, China). Illumina reads were trimmed using Trimmomatic (Bolger et al., 2014) with the following options: trailing: 10, leading: 10, sliding window: 4:20, seed mismatches: 2; palindrome clip threshold: 30, simple clip threshold: 10, and minlen: 40. Clean reads of sample U1 (**Supplementary Table 1**) were used to assemble the cp genome of *U. laevis* using Novoplasty version 3.7 (Dierckxsens et al., 2017) with that of the congeneric species *Ulmus pumila* as reference (Zuo et al., 2017) and the following options: k-mer: 39 and seed: rbcL *U. laevis*. The software Snippy version 3.2-dev[2] was used for reference-based mapping, consensus generation, and variant detection of all other samples. Gene annotation of plastomes was performed with GeSeq (Tillich et al., 2017) and BLAST (Altschul et al., 1990) analysis. The physical map of the *U. laevis* plastid genome was drawn with OGDRAW (Greiner et al., 2019).

Characterization of cpDNA Polymorphisms, Haplotypic Network, and Genomic Diversity Analyses

Plastid sequences were aligned using MAFFT version 7 (Katoh and Standley, 2013), and the resulting alignment was revised manually using CLC Genomics Workbench version 8.5.1 (Qiagen Aarhus, Denmark). Haplotypes were defined as combinations of single nucleotide polymorphism (SNP) and indel (insertion-deletion) variants across the cp genomes. The estimates of haplotype and nucleotide diversity were performed with DnaSP version 6 (Rozas et al., 2017) and PEGAS (Paradis et al., 2015) in R 4.0.4 (R Core Team, 2021). GenoDive version 3.05 (Meirmans and Van Tienderen, 2004) was used to calculate genetic diversity indices. To estimate the relationships between haplotypes, a minimum spanning network of haplotype was constructed using PEGAS by considering all mutation events, both SNPs and indels. Population differentiation based on cpDNA haplotypes was evaluated using two parameters, G_{ST} and N_{ST} (PERMUT 2.0; Pons and Petit, 1996); the presence of phylogeographic structure is highlighted when significantly greater values of N_{ST} (that consider relatedness of haplotypes) in comparison with G_{ST} (that measure the only frequency of haplotypes) are obtained.

A hierarchical analysis of molecular variance (AMOVA; Excoffier et al., 1992) was evaluated using GenAlEx version 6.5 software (Peakall and Smouse, 2012): the molecular variance (Φ_{PT}, an analog of F_{ST}) subdivided into variation among populations and among individuals within populations was evaluated by a permutation test ($n = 999$). The software GenAlEx was used to perform principal coordinates analysis (PCoA), and then plotted in R with the package ggplot2 (Wickham, 2016).

The geographical structure of haplotype diversity was tested with a Bayesian approach implemented in BAPS 6.0 (Corander et al., 2008) using the "clustering with linked loci" analysis. We applied a spatial genetic mixture analysis to the predefined groups of individuals. The software uses Markov chain Monte Carlo (MCMC) simulation to group the sampled populations into K clusters predefined by the user, and the best partitioning is attained based on the highest marginal log-likelihood. After the testing stage, the final analysis was conducted for 10 replicates for K ranging from 2 to 10. Another clustering analysis was conducted by using the "Bayesian approach to phylogeographic clustering" (BPEC) package (Manolopoulou et al., 2016) implemented in R to verify the geographical structure of the populations and the most likely ancestral geographical locations. The analyses were conducted on a two-dimensional dataset (i.e., longitude and latitude), with three levels of parsimony relaxation (ds = 0, ds = 1, and ds = 2), allowing 5 migration events. Two MCMC chains were run for 100 million iterations.

Phylogenetic Analysis of *Ulmus laevis* Populations and Estimation of Divergence Time

The cpDNA sequences of 54 *U. laevis* individuals and 13 plant species within the family Ulmaceae as outgroups (e.g., ten *Ulmus* species, two *Zelkova* species, and one *Hemiptelea* species) were aligned using MAFFT version 7.308 (Katoh and Standley, 2013). We created a time-calibrated phylogenetic tree, reconstructing the phylogeographical relationships and calculating the divergence time of *U. laevis* populations. The plastome phylogeny was inferred by applying the Bayesian inference (BI) method using the program BEAST2 version 2.6.3.0 (Bouckaert et al., 2014). The transversion model (TMV) substitution model was determined to be the best substitution model using a hierarchical likelihood ratio test framework as implemented in jModeltest version 2.1.10 (Darriba et al., 2012). After preliminary runs, we selected the best prior settings for various parameters to achieve an improved effective sample size (ESS) value. We used the lognormal relaxed molecular clock approach and assumed a Yule process of speciation, setting the parameters in BEAUti. The following five priors with normal distribution were selected for calibrations from published data (Zhang et al., 2021): (1) we constrained the crown age of the lineage Ulmaceae to be 85.39 million years ago (Mya) (sigma 1.0, offset: 0.0); (2) *Ulmus* clade was set to 52 Mya (sigma 1.0, offset: 0.0); (3) crown age of *Zelkova* was constrained to be 54.53

[1] http://www.euforgen.org/species/ulmus-laevis/

[2] https://github.com/tseemann/snippy

Mya (sigma 1.0, offset: 0.0); (4) crown age of *Hemiptelea* was set to 72.59 Mya (sigma 1.0, offset: 0.0); and (5) split between *Ulmus americana* and *U. laevis* was constrained to 17.5 Mya. The Bayesian posterior probabilities were sampled using the MCMC algorithm, which was run for 3.0×10^8 generations and sampled every 1,000. The stationarity of the chains and convergence of MCMC simulations were monitored using Tracer version 1.7.1 (Rambaut et al., 2018), analyzing the log output of the BEAST2 analysis. ESSs of all parameters were more than 200, indicating that the estimations were confident. After stationarity was obtained, a maximum clade credibility (MCC) tree was generated using TreeAnnotator version 2.6.3.0 (Helfrich et al., 2019), setting the 20% of trees discarded as burn-in. The MCC tree was visualized using FigTree version 1.4.4 and then was plotted with a geological timescale using the strap package in R (Bell and Lloyd, 2015).

Effective Population Size (N_e)

We estimated the effective population size (N_e) through time for the five populations, applying a non-parametric analysis based on the coalescence theory to cpDNA genome alignments for each population separately. The Bayesian skyline plot (BSP) approach was employed with the TMV substitution site model using a relaxed lognormal molecular clock with rate 1.0 and running 3.0×10^8 MCMC simulations using the software BEAST2.

RESULTS

Ulmus laevis Chloroplast Genome

The complete nucleotide sequence of the cp genome of *U. laevis* was determined through HTS "skim sequencing" (Kumar et al., 2021). Using 864,534 high-quality plastid reads, the cpDNA of one individual (Fsw1) was *de novo* assembled with a depth coverage of 815×. The cpDNA had a size of 159,086 bp, had a guanine-cytosine (GC) content of 35.6%, and its map had the typical quadripartite structure of angiosperms, including two inverted repeats (IRs) of 26,424 bp separated by a large single copy (LSC) of 87,594 bp and a small single copy (SSC) of 18,644 bp (**Figure 1**). The cp genome of *U. laevis* encodes 120 unique genes, of which 19 are duplicated in the IR, giving a total of 139 genes (**Table 1**). The Fsw1 cpDNA was used to map the sequencing reads (ranging from 88,686 to 322,795) of 55 other *U. laevis* samples with an average depth coverage of 304×.

The entire cp genomes of the 56 samples were aligned and haplotypes were identified. The dataset was then reduced to 54 cp genomes: samples U29 and U35 (**Supplementary Table 1**) were the same sample sequenced twice as well as sample U42 was sequenced twice as control for the sequencing process, as expected we obtained the same haplotypes from replicated samples. Only unique samples were kept for further analysis. The 54 plastomes were deposited in GenBank under the accession number PRJNA777116.

Genetic Diversity and Population Differentiation

The alignment of the 54 cp genomes highlighted 190 nucleotide polymorphisms (14 SNPs and 176 indels). Most of the polymorphisms were neutral, and only 14 of these variations (6 SNPs and 8 indels) were in annotated regions. We found 3 mutations (SNPs) into coding sequences, 2 causing synonymous mutations (in *psaB* and *psaI* genes, respectively, both involved in Photosystem I), and 1 non-synonymous mutation in the *rpoC1* gene (that catalyzes a DNA-directed RNA polymerase).

All the polymorphisms allowed identifying 32 haplotypes (**Supplementary Table 2**). Twenty-four haplotypes were private, only 8 haplotypes were represented by more than one sample, and the most frequent haplotype included 11 samples (**Table 2**). The effective number of haplotypes varied across the four populations from 3.6 (France_SW) to 12.0 (Serbia). Haplotypic richness was close to the average for all four populations, ranging from 3.0 (France_SW) to 3.848 (Serbia). Private haplotypes were present in all five populations. The sequencing of the entire plastid genomes unveiled a high number of haplotypes that was reflected in high values of haplotypic diversity: total haplotype diversity was 0.945 and varied from 1.0 (Serbia, France_NE) to 0.801 (Italy). Interestingly, among other polymorphisms, we recovered a 7 bp insertion (TAATAAA) already described in the study by Fuentes-Utrilla et al. (2014) in 5 out of 10 Spain samples and 1 sample from France_SW. Overall nucleotide diversity ($\pi = 1.028 \times 10^{-5}$) was low and ranged from the higher value 2.439×10^{-5} in Serbia to 0 in Spain, where all variants were indels. A comparison of diversity indices obtained with SNPs and indels is reported in **Supplementary Table 3**.

The median-joining network of all haplotypes (**Figure 2**) showed a star-like topology consisting of a central haplotype (I) from which most of the other haplotypes radiated and separated usually by one to two mutations. Serbian haplotypes were present all over the network, close to the center but also with the most diverging haplotypes (XIX, XX).

Total genetic diversity was high ($H_T = 0.960$), but the largest part of cpDNA variation was found within the populations ($H_S = 0.922$). The analyzed *U. laevis* populations showed the low levels of population differentiation ($G_{ST} = 0.050$). Low but significant differentiation was observed among the pairs Italy–Serbia, Spain–Serbia, and Spain–Italy (**Table 3**). In addition, the AMOVA showed that the differences among populations explained 31% of the total variation, while most of the variance (71%) was located within populations (data not shown).

The PCoA scatterplot of the 54 cp genomes is presented in **Figure 3**. The first two PCoA axes account for about 76% of the total variation in the cpDNA, revealing a clear clustering in three groups with multiple haplotypes. The group at the extreme right of the first axis included only Serbian haplotypes. In addition to the three groups, another sample from Serbia (Se14) was resolved.

Using the Bayesian methods, optimum genetic clusters ($k = 4$) were obtained, which did not correspond to the populations. The analysis with BAPS did not display a strong genetic structure along the western part of the sampled species distribution, and two major clusters contained individuals from all four populations (i.e., Spain, France_SW, France_NE, and Italy); nevertheless, two clusters were represented only in Serbia: one with few individuals and the other one included only one individual (**Figure 4A**).

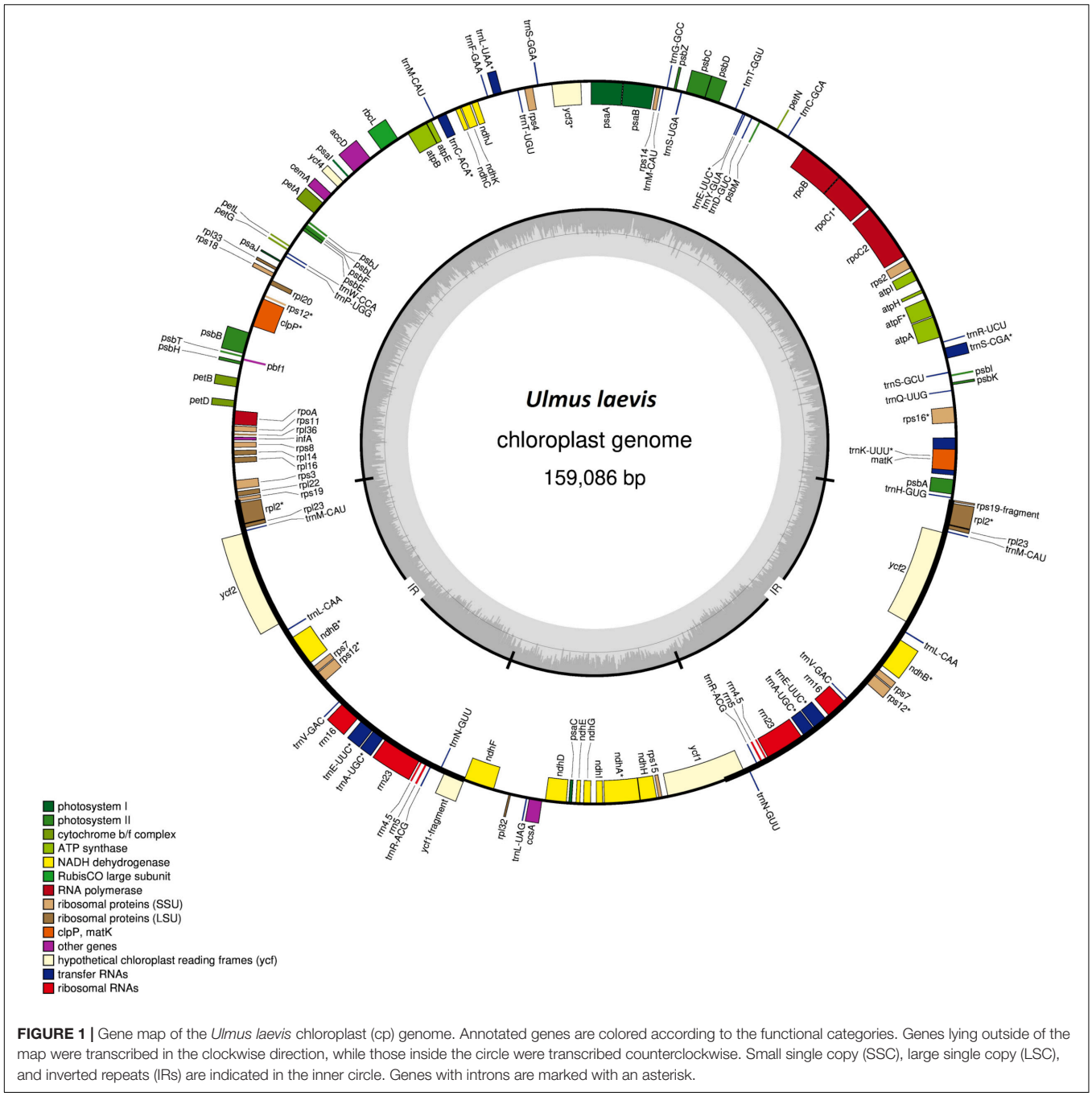


FIGURE 1 | Gene map of the *Ulmus laevis* chloroplast (cp) genome. Annotated genes are colored according to the functional categories. Genes lying outside of the map were transcribed in the clockwise direction, while those inside the circle were transcribed counterclockwise. Small single copy (SSC), large single copy (LSC), and inverted repeats (IRs) are indicated in the inner circle. Genes with introns are marked with an asterisk.

The BPEC clustering showed high uncertainty about the location and number of clusters in the assignment of plastid haplotypes, except for Serbian individuals (**Figure 4B**): haplotypes were assigned to "Serbian" phylogeographic cluster with generally high posterior probabilities (0.64–0.98). The most likely ancestral locations corresponded to sites located in the Ticino valley.

Molecular Dating

The Bayesian phylogenetic inference of cpDNA haplotypes subdivided populations into two main clades, one associated with almost all samples from Serbian populations (Se14, Se15, Se23, Se16, Se22, Se18, and Se24; clade I) and another one composed of other samples (clade II) (**Figure 5**). The molecular dating analysis suggested that the two clades diverged during the Upper Miocene, 17.3 Mya ago [95% highest posterior density (HPD): 12.9–18.7]. Lineages from the Serbian region constituting clade I show more recent allelic divergence, in the range of 5.1–10.1 Mya (95% HPD: 0–6.2 and 3.8–16.7, respectively). In clade I, the Se14 sample started diverging before the others, exhibiting greater genetic differentiation, as confirmed by the other analyses. Regarding clade II, it is noteworthy a subclade including five

TABLE 1 | List of genes present in the *Ulmus laevis* chloroplast genome.

Category	Group of genes	Name of genes
Self-replication	Large subunit of ribosomal protein	*rpl2*, *14*, *16*, *20*, *22*, *23*, *32*, *33*, *36*
	Small subunit of ribosomal proteins	*rps2*, *3*, *4*, *7*, *8*, *11*, *12*, *14*, *15*, *16*, *18*, *19*
	DNA dependent RNA polymerase	*rpoA*, *B*, *C1*, *C2*
	rRNA genes	*rRNA 16S*, *23S*, *4.5S*, *5S*
	tRNA genes	*trnI-TAT*, *trnF-GAA*, *trnR-TCT*, *trnC-GCA*, *trnT-GGT*, *trnN-GTT*, *trnG-GCC*, *trnI-TAT*, *trnS-GGA*, *trnF-GAA*, *trnM-CAT*, *trnV-GAC*, *trnR-ACG*, *trnL-TAG*, *trnK-TT*, *trnN-GTT*, *trnL-CAA*, *trnI-CAT*, *trnV-GAC*, *trnR-ACG*, *trnL-CAA*, *trnN-GTT*, *trnL-CAA*, *trnI-CAT*, *trnP-TGG*, *trnW-CCA*, *trnT-TGT*, *trnfM-CAT*, *trnS-TGA*, *trnE-TTC*, *trnY-GTA*, *trnD-GTC*, *trnI-TAT*, *trnS-GCT*, *trnQ-TTG*, *trnK-CTT*, *trnH-GTG*
Photosynthesis	Photosystem I	*psaA*, *B*, *C*, *I*, *J*
	Photosystem II	*psbA*, *B*, *C*, *D*, *E*, *F*, *H*, *I*, *J*, *K*, *L*, *M*, *T*, *Z*
	Photosystem biogenesis	*pbf1*
	NadH oxidoreductase	*ndhA*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*
	Cytochrome b6/f complex	*petA*, *B*, *D*, *G*, *L*, *N*
	ATP synthase	*atpA*, *B*, *E*, *F*, *H*, *I*
	Rubisco	*rbcL*
Other genes	Maturase	*matK*
	Protease	*clpP*
	Envelop membrane protein	*cemA*
	Subunit Acetyl-CoA-Carboxylate	*accD*
	c-type cytochrome synthesis gene	*ccsA*
	Translation initiation factor IF-1	*infA*
Unknown	Conserved open reading frames	*ycf1*, *2*, *3*, *4*

samples (i.e., It42, Fsw3, It38, Fsw1, and It37) from Italian and French_SW populations that also show a relatively recent divergence time, from 4.9 to 5.96 Mya (95% HPD: 0.01–6.1 and 0.1–8.0, respectively). Among all 13 outgroups, *Hemiptelea* was confirmed to be the most genetically divergent, and this is consistent with earlier studies (Ueda et al., 1997; Sytsma et al., 2002; Zhang et al., 2021), being a sister of the *Ulmus* and *Zelkova* genera.

Profiles for the effective population size (N_e) are shown in the four Skyline plots in **Figure 6** (it was not possible to calculate a Skyline plot for the Spanish population due to the lack of informative variations). The BSPs suggest that the median estimate of population size of all four European populations remained constant throughout the time. Only the Serbian population revealed a quite small increase in the effective population size (N_e) just after the coalescence, while no significant changes in N_e were observed in the other populations. French (NE and SW) and Italian populations share a stable ancestral size.

TABLE 2 | Sample size in each population (N), number of haplotypes detected in each population (A), number of private haplotypes (P), effective number of haplotypes (N_e), and genetic diversity suggested by Nei (1987) corrected for sample size (hapDiv) and nucleotide diversity (nucDiv) in chloroplast DNA sequences of *Ulmus laevis* populations.

Population	N	A	P	N_e	hapDiv	nucDiv
France_SW	5	4	2	3.6	0.900	6.286e-06
France_NE	8	8	6	8.0	1.000	3.143e-06
Serbia	12	12	10	12.0	1.000	2.439e-05
Italy	19	9	5	4.1	0.801	1.764e-06
Spain	10	6	5	5.5	0.911	NA

DISCUSSION

This study sought to explore the genetic structure of the species *U. laevis* in the southern part of its distribution range. Two main focuses were identified, namely, (i) the unexplored glacial refugium in the Balkan Peninsula and (ii) whether small Italian stands in the north-west (Piedmont and Ticino valley) are relic native ones or allochthonous.

This study is among the first to use large-scale intraspecific plastome sequencing of a European forest tree species to explore genetic diversity at population level. The plastid genome of *U. laevis* was recently sequenced (Zhang et al., 2021), but this study provides the first complete annotated plastid genome of EWE and the assessment of intraspecific diversity together with the elucidation of the phylogeography of this species in its southern European range. The cp markers have long been used in forest tree population genetics but were often limited to few polymorphic loci (Scotti et al., 2008; Bagnoli et al., 2009; Dias et al., 2020). The cpDNA sequences of *U. laevis* were highly conserved in terms of size, gene number, and arrangement; however, the full cpDNA sequencing of *U. laevis* unveiled unexplored genetic variation. In fact, 190 polymorphic sites were

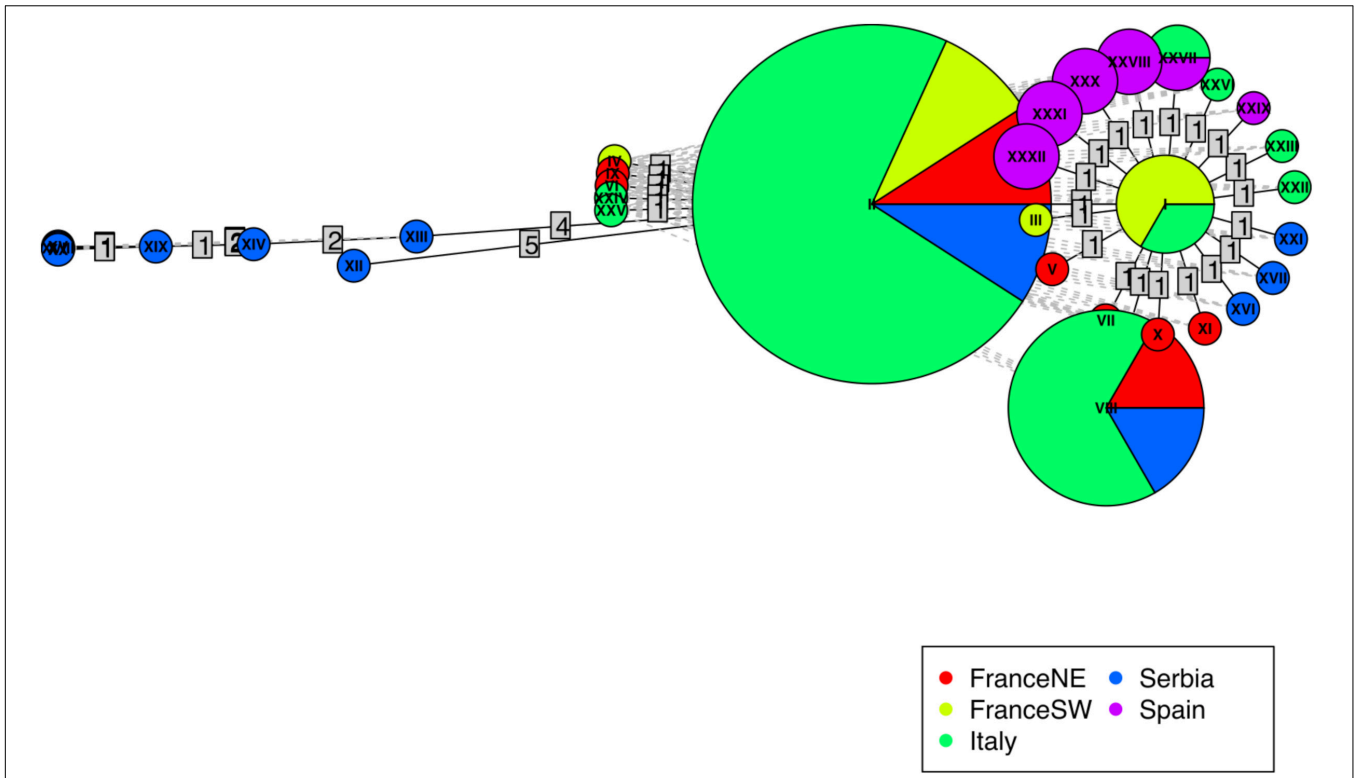


FIGURE 2 | Haplotype network of the 32 *Ulmus laevis* haplotypes computed using R package "pegas." Circles are colored according to the provenance of haplotype from different geographic areas. The size of the circles is proportional to the frequencies of each haplotype across all populations. Lines represent the mutational steps between haplotype sequences.

identified that gave rise to 32 haplotypes, and all the sampled populations showed private haplotypes.

Paleobotanical records witnessed the presence of *Ulmus* spp. in Italy (Mercuri et al., 2002; Amorosi et al., 2014) but, unfortunately, these reports cannot shed light on the past white elm distribution in Italy because, so far, it is not possible to distinguish *Ulmus* species. Our estimates of divergent time of *U. laevis* samples highlight that a subset of Serbian individuals started diverging from the Upper Miocene, while another subset formed a clade with all other provenances. Considering that this approach produces values regarding allelic divergence, while the species are always more recent, our tree is consistent with the divergent times calculated for the other Ulmaceae species (Zhang et al., 2017; Hou et al., 2020). The early split of above all Serbian individuals from other populations combined with the diversity that cluster I samples exhibit in more recent times highlights the role of the Serbian population as a reservoir for EWE genetic diversity and main source for recolonization. At the same time, the more recent divergence exhibited by samples from Italy and southwest France could suggest their long-term refugia isolation.

TABLE 3 | Values of G_{ST} suggested by Nei (1987) for all pairs of populations below the diagonal and N_{ST} above.

	France_SW	France_NE	Serbia	Italy	Spain
France_SW	–	0.008	0.308	−0.061	0.012
France_NE	0.023	–	0.353*	−0.033	0.346*
Serbia	0.030	−0.021	–	0.351*	0.339*
Italy	0.054	0.027	0.049	–	0.311
Spain	0.094	0.045	0.044	0.139	–
All samples G_{ST}	0.034				
All samples N_{ST}	0.078				

*Tests of phylogeographical signal (i.e., $N_{ST} > G_{ST}$): *$p < 0.05$.*

Considering that the mutation rate of organellar DNA is remarkably low (the mutation rate of cpDNA is in the order of 1.3×10^{-9}) (Wolfe et al., 1987), private haplotypes (P) can be considered an indicator of long-term residency which makes anthropogenic introductions unlikely (Wares et al., 2002). Interestingly, the higher number of P, as well as the highest effective number of haplotypes, was observed in Serbia. Nevertheless, private haplotypes were found also in all other areas, though at reduced frequency.

The high number of private haplotypes is reflected in the haplotype network that, alongside the star-like structure, has many tips. The central haplotype, haplotype I, common to France_SW and Italy, should represent the most ancient condition, according to the coalescent theory (Castelloe and Templeton, 1994), while more recently derived haplotypes should be at the tips. According to Posada and Crandall (2001), the haplotypic structure would predict Italy and France_SW as potential refugia but most of the haplotypes are close to

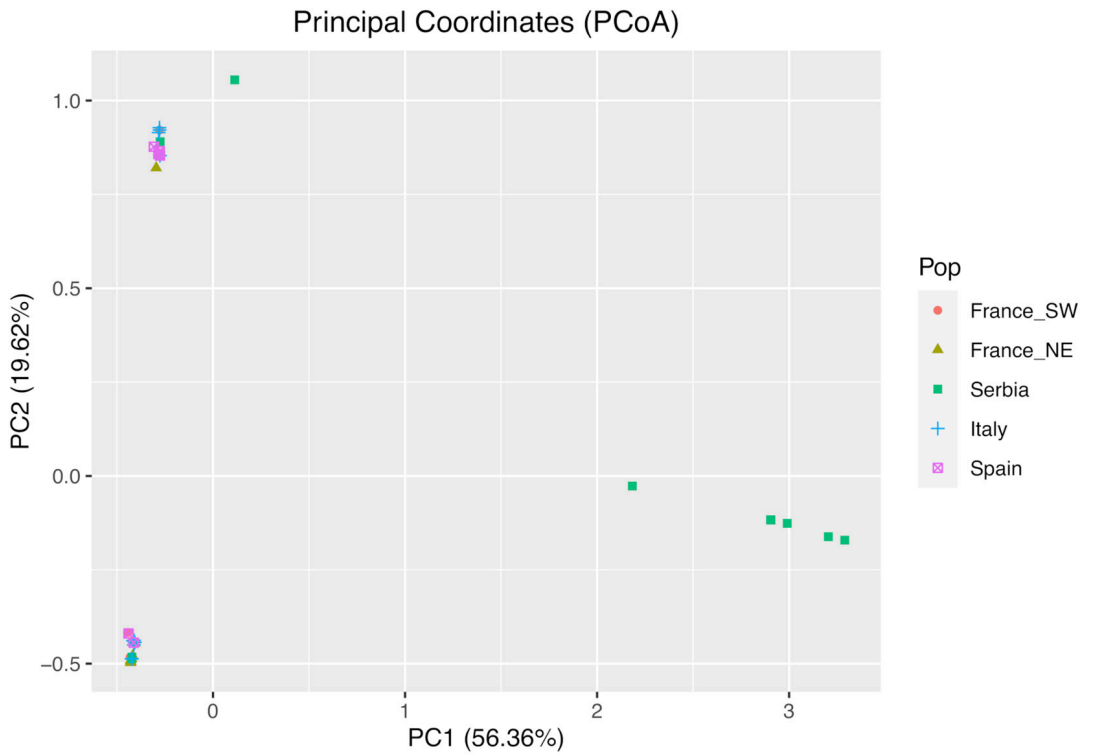


FIGURE 3 | Principal coordinates analysis (PCoA) showing variation between the 54 chloroplast (cp) genomes. Each point represents a haplotype. Geographical regions are marked by different colors.

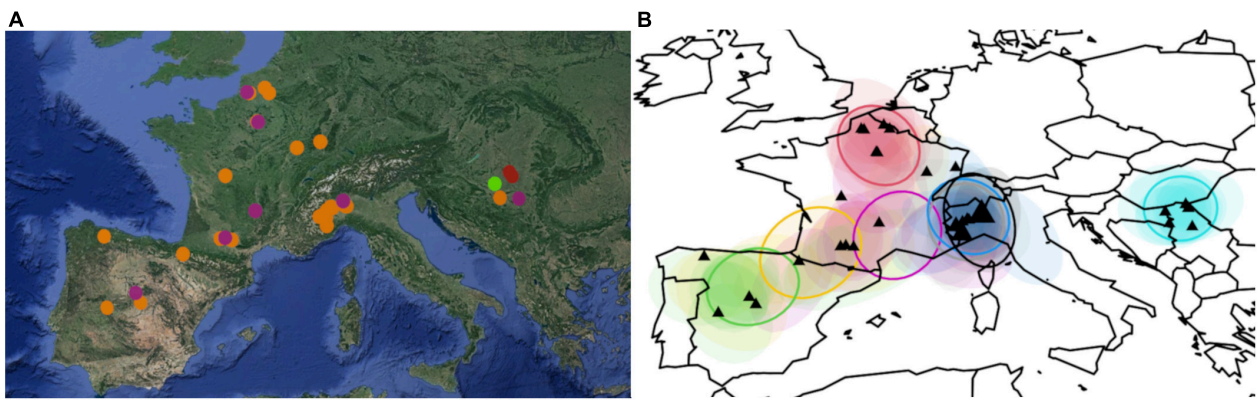


FIGURE 4 | (A) Map showing the location of the four genetic clusters identified by Bayesian analysis of population structure (BAPS) for $k = 4$, using the spatial model. Chart was generated by PhyloGeoViz (Tsai, 2011). Terrain image copyright by Google Earth. The colors represent the cluster to which the analyzed individuals belong based on BAPS assignment. **(B)** Cluster analysis built by the Bayesian phylogeographic and ecological clustering. Each colored areas indicate a different phylogeographic grouping.

haplotype I, differing for just one mutation; hence, this structure is not so strong. In fact, the haplotypes found in the Serbian population are distributed over the entire network, from close derivatives of ancestral haplotypes to most distant, and their frequent appearance is remarkable and in agreement with a long-term occurrence.

Genetic diversity among and within populations is expected to be higher in refugia than in recently colonized areas. Plant populations in refugia are recognized as those that exhibit high levels of genetic diversity and uniqueness (Petit et al., 2003; Carnaval et al., 2009). Haplotype diversity is high in all samples, but interestingly, Serbian population has higher nucleotide diversity (π = 2.439e-05). Thus, higher genetic diversity evidenced by the Serbian population could reflect the accumulation of nucleotide mutations that characterize the isolation of individuals belonging to a refugia population.

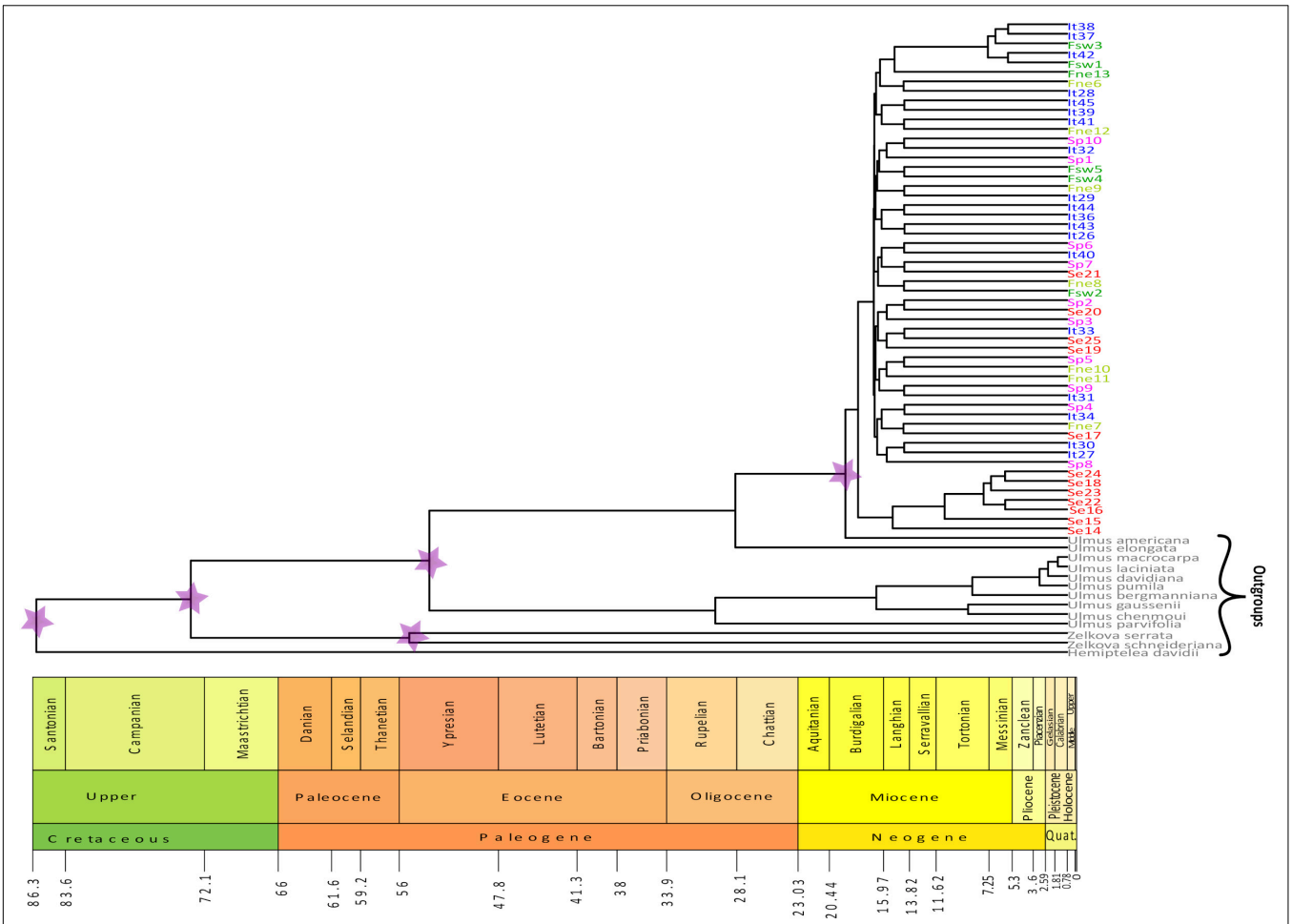


FIGURE 5 | Phylogenetic tree of *Ulmus laevis* individuals calculated using BEAST software. The tree includes 54 *U. laevis* samples from five populations (indicated by different colors: blue = Italy, dark green = France_SW, light green = France_NE, red = Serbia, pink = Spain) and 13 outgroups species within Ulmaceae family: *Ulmus americana* (NC_044473.1); *Ulmus elongata* (NC_046061.1); *Ulmus macrocarpa* (NC_032720.1); *Ulmus laciniata* (MT742253.1); *Ulmus davidiana* (NC_032718.1); *Ulmus pumila* (MW279236.1); *Ulmus bergmanniana* (NC_057601.1); *Ulmus gaussenii* (NC_037840.1); *Ulmus chenmoui* (NC_037758.1); *Ulmus parvifolia* (MT701612.1); *Zelkova serrata* (NC_040958.1); *Zelkova schneideriana* (NC_041074.1); and *Hydrangea davidii* (MK070168.1). The stars on nodes indicate the five points of age constraints on divergent time. Time axis shows the age in Mya. Geological time scale visualized with the R strap package.

The Balkan Peninsula might have provided suitable habitat for *U. laevis* during the Pleistocene climate oscillations. Multiple refugia have been inferred in the Balkan Peninsula for many animals and plants during the Pleistocene glacial periods (Ursenbacher et al., 2008; Bagnoli et al., 2016).

Our phylogeographic survey showed a low genetic differentiation due to the high level of haplotype diversity found within populations. Overall, N_{ST} was not significantly higher than G_{ST}; however, no negligible phylogeographic signal was detected in pairwise comparisons. In fact, N_{ST} of Serbian samples was significantly higher than G_{ST} of samples from Italy, Spain, and France_NE.

The Bayesian genetic clustering identified four groups, all of which present in Serbia, while only two were detected in the other sampled areas; these data together with the results of PCoA highlight the Balkan area as the one of the greatest diversity, which is possibly the main refuge area for the species. The confirmed absence of a clear structure is particularly significant considering the use of uniparentally inherited genetic markers. In contrast, the weak genetic structure within regions can be explained by a high rate of seed dispersal and long-range colonization of plastome variants (maternally inherited with seeds), while the overall observed genetic structuring could be originated by the accumulation of mutations in different refugia followed by recent population expansion. Consistent with these observations are the results of BSPs used for inference of coalescence patterns (Heller et al., 2013), which indicate a peculiarity of Serbian individuals attributable to the glacial refuge condition.

According to Fuentes-Utrilla et al. (2014), Spain can be considered a potential refuge, but in this view, Italy emerged as a previously unrecognized native population. If the appearance of new alleles following introductions can be ruled out because of the very low mutation rate and then two alternative hypotheses

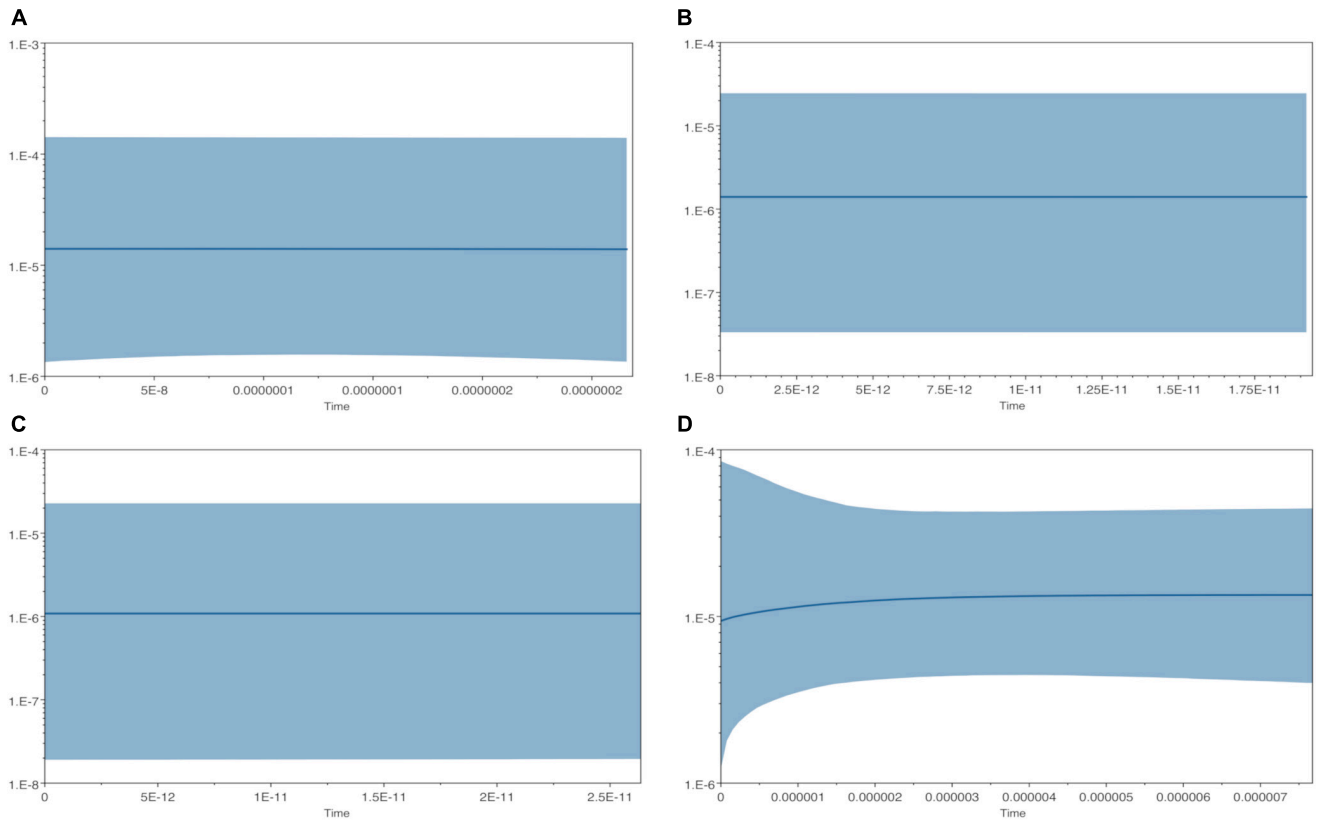


FIGURE 6 | Bayesian Skyline plots derived from cpDNA sequences of *Ulmus laevis*. The analysis was performed for each population separately. The *X*-axis is in units of million years ago (Mya) and the *Y*-axis is the population size, expressed in units of Neτ, the product of effective population sizes per generation length. The thick solid line represents the mean effective population size N_e, while the 95% highest posterior density (HPD) limits are shown by the blue area. **(A)** France_NE, **(B)** France_SW, **(C)** Italy, and **(D)** Serbia.

can be formulated, namely, the first one is that the observed data are the product of multiple introductions, but the sampling was not adequate to highlight the source; alternatively, we are observing the results of a composite demographic history, with small/cryptic refugia scattered at the margins of actual species distribution contributing to the actual diversity of haplotypes. Although we are dealing with a small-scale sample size, it was established that this should not interfere with the estimates of patterns of genetic diversity at a wide geographical range (Taberlet et al., 2012). Overall, the observed pattern of diversity could be explained with the existence of a principal refuge in the Balkans that represented the main source of post-glacial colonization from southeast to central and northern Europe. Based on our results, mainly the presence of private haplotypes, the Italian population may no longer be considered an anthropogenic introduction, but could be considered native and, consistently, with previous study for French (Timbal and Collin, 1999) and Spanish populations (Fuentes-Utrilla et al., 2014), all most likely originated from Serbian refuge.

Although the connection with Balkan refugia still needs deeper investigation, unfortunately, there are no warnings about the presence of white elms on the northeastern part of Italy, but long-range dispersal cannot be ruled out.

In support of our hypothesis, there are also arguments linked to the ecology of the species and the habitat transformation: most of the sampled trees were at the margin of high intensively cultivated areas close to rivers or scattered not far from such small stands. The area where we hypothesized a native presence of white elms, on the western side of the Po valley, experienced a high rate of population growth since the Neolithic Age. During the Roman Age, an extensive deforestation occurred, due to the assignment of land grants to war veterans, and continued in different ways till nowadays (Marchetti, 2002). The prolonged human presence highly impacted natural ecosystems by mainly extending urbanization and cultivated areas of agricultural crop. In that context, the introduction of white elms, species with no great economical relevance, especially at the bottom of the Alps, has no reasonable explanations. The Italian EWE population, which is conserving an important amount of genetic diversity, is composed of small, scattered populations at the margin of the distribution of species, and this pattern is in line with the evolution of the territory. Continued anthropogenic action along river courses could further limit the current populations, putting them at risk of extinction. This risk is a call for the maintenance of evolutionary processes through the conservation of genetic resources. Riparian forests, in whose ecosystem the EWE is

a distinctive species, have high biodiversity and an important ecological role, so any change in their ecological dynamics also impacts the population dynamics of trees and, thus, their genetic diversity (Guilloy-Froget et al., 2002). For this reason, the conservation of genetic resources should be done with a dynamic approach not only at the level of single species but also on the whole riparian ecosystem (Lefèvre et al., 2013).

The pattern of genetic diversity observed in *U. laevis* is largely congruent with other previously studied European species, such as *Hordelymus europaeus* and *Fagus sylvatica* and another riparian tree, *Alnus glutinosa* (Magri, 2008; Dvořková et al., 2010; Havrdová et al., 2015); in particular, it points to the existence of a main refugium in the Balkan Peninsula, while additional southern refugial rear-edge populations in Iberian and Italian Peninsulas represent current the relicts with unique haplotype diversity.

Although more intensive sampling of the European populations in combination with nuclear genotyping is needed to give a more complete picture of the phylogeography of *U. laevis*, this study, to the best of our knowledge, for the first time extended the population genetic analysis of EWE to Italian and Balkan populations, revealing probably the main refuge of the species during the last glaciation. In addition, we extended the natural range of *U. laevis* in Europe to northwestern Italy and highlighted their small size and fragmentation and also at risk. The risk comes more from population size and habitat erosion than from DED; anyway, proper management practices become important to preserve these small stands with both *in situ* and *ex situ* actions to preserve and hopefully restore riparian ecosystems.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: https://www.ncbi.nlm.nih.gov/genbank/, PRJNA777116.

AUTHOR CONTRIBUTIONS

FS and ASa conceived the study. ST, FS, IP, and AP conducted all laboratory and population genetics analyses. GB, FP, ASa, LG, and ASe collected the samples. ST and FS led the writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research has received funding from the Parco Lombardo della Valle del Ticino and ERSAF (Ente Regionale per i Servizi all'Agricoltura e alle Foreste) Lombardia.

ACKNOWLEDGMENTS

We thank Eric Collin (IRSTEA, France) and Martin Venturas (E.T.S.I. Montes, Universidad Politécnica de Madrid, Spain) for providing samples from France and Spain, respectively; we also thank Paolo Varese, Andrea Ebone, and Gian Paolo Mondino for sharing their distribution knowledge on EWE in Piedmont. This article is dedicated to the memory of Nenad Keca, Sumarski fakultet Univerziteta u Beogradu, a friend and an esteemed colleague. Nenad sampled *Ulmus laevis* in Serbia and, with his help, it was possible to shed light on the phylogeography of EWE.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.826158/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Amorosi, A., Antonioli, F., Bertini, A., Marabini, S., Mastronuzzi, G., Montagna, P., et al. (2014). The middle–upper pleistocene fronte section (Taranto, Italy): an exceptionally preserved marine record of the Last Interglacial. *Glob. Planet. Change* 119, 23–38. doi: 10.1016/J.GLOPLACHA.2014.04.007

Bagnoli, F., Tsuda, Y., Fineschi, S., Bruschi, P., Magri, D., Zhelev, P., et al. (2016). Combining molecular and fossil data to infer demographic history of Quercus cerris: insights on European eastern glacial refugia. *J. Biogeogr.* 43, 679–690. doi: 10.1111/JBI.12673

Bagnoli, F., Vendramin, G. G., Buonamici, A., Doulis, A. G., GonzÁlez-MartÍnez, S. C., La Porta, N., et al. (2009). Is *Cupressus sempervirens* native in Italy? An answer from genetic and palaeobotanical data. *Mol. Ecol.* 18, 2276–2286. doi: 10.1111/j.1365-294X.2009.04182.x

Bartolucci, F., Peruzzi, L., Galasso, G., Albano, A., Alessandrini, A., Ardenghi, N. M. G., et al. (2018). An updated checklist of the vascular flora native to Italy. *Plant Biosyst.* 152, 179–303. doi: 10.1080/11263504.2017.1419996

Bell, M. A., and Lloyd, G. T. (2015). strap: an R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence. *Palaeontology* 58, 379–389. doi: 10.1111/PALA.12142

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

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., et al. (2014). BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/JOURNAL.PCBI.1003537

Brasier, C. M. (2000). "Intercontinental spread and continuing evolution of the Dutch Elm disease pathogens," in *The Elms*, ed. C. P. Dunn (Boston, MA: Springer), 61–72. doi: 10.1007/978-1-4615-4507-1_4

Brewer, S., Cheddadi, R., de Beaulieu, J. L., Reille, M., Allen, J., Almqvist-Jacobson, H., et al. (2002). The spread of deciduous Quercus throughout Europe since the last glacial period. *For. Ecol. Manag.* 156, 27–48. doi: 10.1016/S0378-1127(01)00646-6

Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T., and Moritz, C. (2009). Stability predicts genetic diversity in the brazilian atlantic forest hotspot. *Science* 323, 785–789. doi: 10.1126/SCIENCE.1166955

Castelloe, J., and Templeton, A. R. (1994). Root probabilities for intraspecific gene trees under neutral coalescent theory. *Mol. Phylogenet. Evol.* 3, 102–113. doi: 10.1006/MPEV.1994.1013

Collin, E. (2002). *Strategies and Guidelines for the Conservation of the Genetic Resources of Ulmus spp*. 50–67. Available online at: https://hal.inrae.fr/hal-02580649 (accessed October 15, 2021).

Corander, J., Marttinen, P., Sirén, J., and Tang, J. (2008). Enhanced bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinforma* 91:539. doi: 10.1186/1471-2105-9-539

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 98:772. doi: 10.1038/nmeth.2109

Decocq, G., Bordier, D., Wattez, J.-R., and Racinet, P. (2004). A practical approach to assess the native status of a rare plant species: the controverse of *Buxus sempervirens* L. in northern France revisited. *Plant Ecol.* 1731, 139–151. doi: 10.1023/B:VEGE.0000026337.85794.FB

Dias, A., Giovannelli, G., Fady, B., Spanu, I., Vendramin, G. G., Bagnoli, F., et al. (2020). Portuguese Pinus nigra J.F. Arnold populations: genetic diversity, structure and relationships inferred by SSR markers. *Ann. For. Sci.* 773, 1–15. doi: 10.1007/S13595-020-00967-9

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45:e18. doi: 10.1093/nar/gkw955

Dvořková, H., Fér, T., and Marhold, K. (2010). Phylogeographic pattern of the European forest grass species Hordelymus europaeus: cpDNA evidence. *Flora Morphol. Distrib. Funct. Ecol. Plants* 205, 418–423. doi: 10.1016/J.FLORA.2009.12.029

Ebrahimi, A., Antonides, J. D., Pinchot, C. C., Slavicek, J. M., Flower, C. E., and Woeste, K. E. (2021). The complete chloroplast genome sequence of American elm (*Ulmus americana*) and comparative genomics of related species. *Tree Genet. Genomes* 17, 1–13. doi: 10.1007/S11295-020-01487-3

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491. doi: 10.1093/GENETICS/131.2.479

Franke, A., Gil, L., Rusanen, M., Aguiar, A., Collin, E., Harvengt, L., et al. (2004). Methods and progress in the conservation of elm genetic resources in Europe. *Investig. Agrar. Sist. y Recur. For.* 13, 261–272. doi: 10.5424/831

Fuentes-Utrilla, P., Venturas, M., Hollingsworth, P. M., Squirrell, J., Collada, C., Stone, G. N., et al. (2014). Extending glacial refugia for a European tree: genetic markers show that iberian populations of white elm are native relicts and not introductions. *Heredity* 112, 105–113. doi: 10.1038/hdy.2013.81

Gepts, P. (2006). Plant genetic resources conservation and utilization: the accomplishments and future of a societal insurance policy. *Crop Sci.* 46, 2278–2292. doi: 10.2135/CROPSCI2006.03.0169GAS

Gitzendanner, M. A., Soltis, P. S., Wong, G. K.-S., Ruhfel, B. R., and Soltis, D. E. (2018). Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *Am. J. Bot.* 105, 291–301. doi: 10.1002/AJB2.1048

Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238

Gugerli, F., Rüegg, M., and Vendramin, G. G. (2009). Gradual decline in genetic diversity in Swiss stone pine populations (*Pinus cembra*) across Switzerland suggests postglacial re-colonization into the Alps from a common eastern glacial refugium. *Bot. Helv.* 119, 13–22. doi: 10.1007/s00035-009-0052-6

Guilloy-Froget, H., Muller, E., Barsoum, N., and Hughes, F. M. R. (2002). Dispersal, germination, and survival of populus nigra L. (salicaceae) in changing hydrologic conditions. *Wetlands* 22, 478–488.

Halewood, M., Chiurugwi, T., Hamilton, R. S., Kurtz, B., Marden, E., Welch, E., et al. (2018). Plant genetic resources for food and agriculture: opportunities and challenges emerging from the science and information technology revolution. *New Phytol.* 217, 1407–1419. doi: 10.1111/NPH.14993

Havrdová, A., Douda, J., Krak, K., Vít, P., Hadincová, V., Zákravský, P., et al. (2015). Higher genetic diversity in recolonized areas than in refugia of *Alnus glutinosa* triggered by continent-wide lineage admixture. *Mol. Ecol.* 24, 4759–4777. doi: 10.1111/MEC.13348

Helfrich, P., Rieb, E., Abrami, G., Lücking, A., and Mehler, A. (2019). "Treeannotator: versatile visual annotation of hierarchical text relations," in *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, Miyazaki, 1958–1963.

Heller, R., Chikhi, L., and Siegismund, H. R. (2013). The confounding effect of population structure on bayesian skyline plot inferences of demographic history. *PLoS One* 8:e62992. doi: 10.1371/JOURNAL.PONE.0062992

Heuertz, M., Fineschi, S., Anzidei, M., Pastorelli, R., Salvini, D., Paule, L., et al. (2004). Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in Europe. *Mol. Ecol.* 13, 3437–3452. doi: 10.1111/j.1365-294X.2004.02333.x

Hoban, S., Campbell, C. D., da Silva, J. M., Ekblom, R., Funk, W. C., Garner, B. A., et al. (2021). Genetic diversity is considered important but interpreted narrowly in country reports to the convention on biological diversity: current actions and indicators are insufficient. *Biol. Conserv.* 261:109233. doi: 10.1016/J.BIOCON.2021.109233

Hou, H., Ye, H., Wang, Z., Wu, J., Gao, Y., Han, W., et al. (2020). Demographic history and genetic differentiation of an endemic and endangered *Ulmus lamellosa* (*Ulmus*). *BMC Plant Biol.* 20:526. doi: 10.1186/s12870-020-02723-7

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment Software Version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/MOLBEV/MST010

Koskela, J., Lefèvre, F., Schueler, S., Kraigher, H., Olrik, D. C., Hubert, J., et al. (2013). Translating conservation genetics into management: pan-European minimum requirements for dynamic conservation units of forest tree genetic diversity. *Biol. Conserv.* 157, 39–49. doi: 10.1016/J.BIOCON.2012.07.023

Kumar, P., Choudhary, M., Jat, B. S., Kumar, B., Singh, V., Kumar, V., et al. (2021). Skim sequencing: an advanced NGS technology for crop improvement. *J. Genet.* 100:38. doi: 10.1007/S12041-021-01285-3

Lanier, H. C., Massatti, R., He, Q., Olson, L. E., and Knowles, L. L. (2015). Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Mol. Ecol.* 24, 3688–3705. doi: 10.1111/MEC.13270

Lefèvre, F., Koskela, J., Hubert, J., Kraigher, H., Longauer, R., Olrik, D. C., et al. (2013). Dynamic conservation of forest genetic resources in 33 European Countries. *Conserv. Biol.* 27, 373–384. doi: 10.1111/j.1523-1739.2012.01961.x

Magri, D. (2008). Patterns of post-glacial spread and the extent of glacial refugia of European beech (*Fagus sylvatica*). *J. Biogeogr.* 35, 450–463. doi: 10.1111/J.1365-2699.2007.01803.X

Magri, D., Vendramin, G. G., Comps, B., Dupanloup, I., Geburek, T., Gömöry, D., et al. (2006). A new scenario for the quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytol.* 171, 199–221. doi: 10.1111/j.1469-8137.2006.01740.x

Manolopoulou, I., Hille, A., and Emerson, B. (2016). BPEC: an R package for bayesian phylogeographic and ecological clustering. *J. Stat. Softw.* 92, 1–32.

Marchetti, M. (2002). Environmental changes in the central Po Plain (northern Italy) due to fluvial modifications and anthropogenic activities. *Geomorphology* 44, 361–373. doi: 10.1016/S0169-555X(01)00183-0

Meirmans, P. G., and Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes* 4, 792–794. doi: 10.1111/j.1471-8286.2004.00770.x

Mercuri, A. M., Accorsi, C. A., and Bandini Mazzanti, M. (2002). The long history of Cannabis and its cultivation by the Romans in central Italy, shown by pollen records from Lago Albano and Lago di Nemi. *Veg. Hist. Archaeobotany* 114, 263–276. doi: 10.1007/S003340200039

Mittempergher, L., and La Porta, N. (1991). Hybridization studies in the Eurasian species of elm (*Ulmus* spp.). *Silvae Genet.* 40, 237–243.

Mona, S., Catalano, G., Lari, M., Larson, G., Boscato, P., Casoli, A., et al. (2010). Population dynamic of the extinct European aurochs: genetic evidence of a north-south differentiation pattern and no evidence of post-glacial expansion. *BMC Evol. Biol.* 10:83. doi: 10.1186/1471-2148-10-83

Morgante, M., Pfeiffer, A., Costacurta, A., and Olivieri, A. M. (1996). Molecular tools for population and ecological genetics in coniferous trees. *Phyt. Ann. Rei Bot.* 36, 129–138.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press. doi: 10.7312/NEI-92038

Niu, Y. T., Ye, J. F., Zhang, J. L., Wan, J. Z., Yang, T., Wei, X. X., et al. (2018). Long-distance dispersal or postglacial contraction? Insights into disjunction between Himalaya–Hengduan Mountains and Taiwan in a cold-adapted herbaceous genus, Triplostegia. *Ecol. Evol.* 8, 1131–1146. doi: 10.1002/ece3.3719

Paradis, E., Jombart, T., Schliep, K., Potts, A., Winter, D., and Paradis, M. E. (2015). *Package "Pegas" Title Population and Evolutionary Genetics Analysis System*.

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/BIOINFORMATICS/BTS 460

Pedreschi, D., García-Rodríguez, O., Yannic, G., Cantarello, E., Diaz, A., Golicher, D., et al. (2019). Challenging the European southern refugium hypothesis: species-specific structures versus general patterns of genetic diversity and differentiation among small mammals. *Glob. Ecol. Biogeogr.* 28, 262–274. doi: 10.1111/GEB.12828

Petit, R. J., Aguinagalde, I., De Beaulieu, J. L., Bittkau, C., Brewer, S., Cheddadi, R., et al. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300, 1563–1565. doi: 10.1126/science.1083264

Petit, R. J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., et al. (2002). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *For. Ecol. Manag.* 156, 49–74. doi: 10.1016/S0378-1127(01)00634-X

Pignatti, S., Guarino, R., and La Rosa, M. (2017). *Flora d'Italia*, 2nd Edn, Vol. 1-4. Bologna: Edagricole.

Pons, O., and Petit, R. J. (1996). Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144, 1237–1245. doi: 10.1093/genetics/144.3.1237

Porth, I., and El-Kassaby, Y. A. (2014). Assessment of the genetic diversity in forest tree populations using molecular markers. *Divers* 6, 283–295. doi: 10.3390/D6020283

Posada, D., and Crandall, K. A. (2001). Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* 16, 37–45. doi: 10.1016/S0169-5347(00)02026-7

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/SYSBIO/SYY032

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/MOLBEV/MSX248

Sacchetti, P., Tiberi, R., and Mittempergher, L. (1990). Preference of *Scolytus multistriatus* (Marsham) during the gonad maturation phase between two species of elm. *Redia* 73, 347–354.

Santini, A., Montaghi, A., Vendramin, G. G., and Capretti, P. (2005). Analysis of the Italian Dutch Elm disease fungal population. *J. Phytopathol.* 153, 73–79. doi: 10.1111/J.1439-0434.2004.00931.X

Scotti, I., Gugerli, F., Pastorelli, R., Sebastiani, F., and Vendramin, G. G. (2008). Maternally and paternally inherited molecular markers elucidate population patterns and inferred dispersal processes on a small scale within a subalpine stand of Norway spruce (Picea abies [L.] Karst.). *For. Ecol. Manag.* 255, 3806–3812. doi: 10.1016/J.FORECO.2008.03.023

Svejgaard Jensen, J. (2003). *EUFORGEN Technical Guidelines for Genetic Conservation and use for Lime (Tilia spp.)*. Rome: IPGRI.

Sytsma, K. J., Morawetz, J., Chris Pires, J., Nepokroeff, M., Conti, E., Zjhra, M., et al. (2002). Urticalean rosids: circumscription, rosid ancestry, and phylogenetics based on rbcL, trnL-F, and ndhF sequences. *Am. J. Bot.* 89, 1531–1546. doi: 10.3732/ajb.89.9.1531

Taberlet, P., Zimmermann, N. E., Englisch, T., Tribsch, A., Holderegger, R., Alvarez, N., et al. (2012). Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecol. Lett.* 15, 1439–1448. doi: 10.1111/ele.12004

Tamošaitis, S., Jurkšienė, G., Petrokas, R., Buchovska, J., Kavaliauskienė, I., Danusevičius, D., et al. (2021). Dissecting taxonomic variants within *Ulmus* spp. complex in natural forests with the aid of microsatellite and morphometric markers. *For* 12:653. doi: 10.3390/F12060653

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/NAR/GKX391

Timbal, J., and Collin, E. (1999). The *Ulmus laevis* Pallas in southern France – distribution and strategy for genetic resource conservation. *Fr. For. J.* 51, 593–604. doi: 10.4267/2042/5468

Toro, M. A., and Caballero, A. (2005). Characterization and conservation of genetic diversity in subdivided populations. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 1367–1378. doi: 10.1098/rstb.2005.1680

Tsai, Y. H. E. (2011). PhyloGeoViz: a web-based program that visualizes genetic data on maps. *Mol. Ecol. Resour.* 11, 557–561. doi: 10.1111/J.1755-0998.2010.02964.X

Ueda, K., Kosuge, K., and Tobe, H. (1997). A molecular phylogeny of Celtidaceae and Ulmaceae (Urticales) based on rbcL nucleotide sequences. *J. Plant Res.* 110, 171–178. doi: 10.1007/bf02509305

Unamba, C. I. N., Nag, A., and Sharma, R. K. (2015). Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. *Front. Plant Sci.* 6:1074. doi: 10.3389/fpls.2015.01074

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

Vakkari, P., Rusanen, M., and Kärkkäinen, K. (2009). High genetic differentiation in marginal populations of European White Elm (*Ulmus laevis*). *Silva Fenn.* 43, 185–196. doi: 10.14214/SF.205

Venturas, M., Fuentes-Utrilla, P., Ennos, R., Collada, C., and Gil, L. (2013). Human-induced changes on fine-scale genetic structure in *Ulmus laevis* Pallas wetland forests at its SW distribution limit. *Plant Ecol.* 214, 317–327. doi: 10.1007/s11258-013-0170-5

Wang, X. R., Wang, X. R., and Szmidt, A. E. (2001). Molecular markers in population genetics of forest trees. *Scand. J. For. Res.* 16, 199–220. doi: 10.1080/02827580118146

Wares, J. P., Goldwater, D. S., Kong, B. Y., and Cunningham, C. W. (2002). Refuting a controversial case of a human-mediated marine species introduction. *Ecol. Lett.* 5, 577–584. doi: 10.1046/J.1461-0248.2002.00359.X

Whiteley, R. (2004). *Quantitative and Molecular Genetic Variations in \emph{Ulmus laevis} Pall*.

Wickham, H. (2016). *ggplot2*. doi: 10.1007/978-3-319-24277-4

Wolfe, K. H., Li, W. H., and Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U.S.A.* 84, 9054–9058. doi: 10.1073/pnas.84.24.9054

Zhang, M. L., Wang, L., Lei, Y., and Sanderson, S. C. (2017). Cenozoic evolutionary history of Zelkova (Ulmaceae), evidenced from ITS, trnL-trnF, psbA-trnH, and rbcL. *Tree Genet. Genomes* 13, 1–10. doi: 10.1007/s11295-017-1182-4

Zhang, Q. Y., Deng, M., Bouchenak-Khelladi, Y., Zhou, Z. K., Hu, G. W., and Xing, Y. W. (2021). The diversification of the northern temperate woody flora – A case study of the Elm family (Ulmaceae) based on phylogenomic and paleobotanical evidence. *J. Syst. Evol* doi: 10.1111/jse.12720

Zuo, L., Zhang, S., Zhang, J., Liu, Y., Yu, X., Yang, M., et al. (2020). Primer development and functional classification of EST-SSR markers in *Ulmus* species. *Tree Genet. Genomes* 16:74. doi: 10.1007/S11295-020-01468-6

Zuo, L.-H., Shang, A.-Q., Zhang, S., Yu, X.-Y., Ren, Y.-C., Yang, M.-S., et al. (2017). The first complete chloroplast genome sequences of *Ulmus* species by de novo sequencing: genome comparative and taxonomic position analysis. *PLoS One* 12:e0171264. doi: 10.1371/JOURNAL.PONE.0171264

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 14 March 2022
doi: 10.3389/fpls.2022.822331

Stability in the South, Turbulence Toward the North: Evolutionary History of *Aurinia saxatilis* (Brassicaceae) Revealed by Phylogenomic and Climatic Modelling Data

Ivana Rešetnik[1], Eliška Záveská[2], Marin Grgurev[1], Sandro Bogdanović[3,4], Paolo Bartolić[5] and Božo Frajman[6]*

[1] Department of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia, [2] Institute of Botany, Czech Academy of Sciences, Prague, Czechia, [3] Department of Agricultural Botany, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia, [4] Centre of Excellence for Biodiversity and Molecular Plant Breeding, Zagreb, Croatia, [5] Department of Botany, Charles University, Prague, Czechia, [6] Department of Botany, University of Innsbruck, Innsbruck, Austria

OPEN ACCESS

Edited by:
Susann Wicke, Humboldt University of Berlin, Germany

Reviewed by:
Gábor Sramkó, MTA-DE Lendület Evolutionary Phylogenomics Research Group, Hungary
Jelena Mlinarec, University of Zagreb, Croatia
Jan W. Arntzen, Naturalis Biodiversity Center, Netherlands

***Correspondence:**
Ivana Rešetnik
ivana.resetnik@biol.pmf.hr

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

Received: *25 November 2021*
Accepted: *31 January 2022*
Published: *14 March 2022*

Citation:
Rešetnik I, Záveská E, Grgurev M, Bogdanović S, Bartolić P and Frajman B (2022) Stability in the South, Turbulence Toward the North: Evolutionary History of Aurinia saxatilis (Brassicaceae) Revealed by Phylogenomic and Climatic Modelling Data. Front. Plant Sci. 13:822331. doi: 10.3389/fpls.2022.822331

The Balkan Peninsula played an important role in the evolution of many Mediterranean plants and served as a major source for post-Pleistocene colonisation of central and northern Europe. Its complex geo-climatic history and environmental heterogeneity significantly influenced spatiotemporal diversification and resulted in intricate phylogeographic patterns. To explore the evolutionary dynamics and phylogeographic patterns within the widespread eastern Mediterranean and central European species *Aurinia saxatilis*, we used a combination of phylogenomic (restriction-site associated DNA sequencing, RADseq) and phylogenetic (sequences of the plastid marker *ndh*F) data as well as species distribution models generated for the present and the Last Glacial Maximum (LGM). The inferred phylogenies retrieved three main geographically distinct lineages. The southern lineage is restricted to the eastern Mediterranean, where it is distributed throughout the Aegean area, the southern Balkan Peninsula, and the southern Apennine Peninsula, and corresponds to the species main distribution area during the LGM. The eastern lineage extends from the eastern Balkan Peninsula over the Carpathians to central Europe, while the central lineage occupies the central Balkan Peninsula. Molecular dating places the divergence among all the three lineages to the early to middle Pleistocene, indicating their long-term independent evolutionary trajectories. Our data revealed an early divergence and stable *in situ* persistence of the southernmost, eastern Mediterranean lineage, whereas the mainland, south-east European lineages experienced more complex and turbulent evolutionary dynamics triggered by Pleistocene climatic oscillations. Our data also support the existence of multiple glacial refugia in southeast Europe and highlight the central Balkan Peninsula not only as a

cradle of lineage diversifications but also as a source of lineage dispersal. Finally, the extant genetic variation within *A. saxatilis* is congruent with the taxonomic separation of peripatric *A. saxatilis* subsp. *saxatilis* and *A. saxatilis* subsp. *orientalis*, whereas the taxonomic status of *A. saxatilis* subsp. *megalocarpa* remains doubtful.

Keywords: *Aurinia saxatilis*, demographic modelling, glacial refugia, *ndh*F, RAD sequencing, species distribution modelling

INTRODUCTION

The Balkan, Apennine and Iberian Peninsulas extending to the Mediterranean Basin represent European biodiversity hotspots, hosting high species richness and genetic diversity (Petit et al., 2003; Hewitt, 2011; Nieto Feliner, 2014; Gömöry et al., 2020). They were important Pleistocene refugia, from where biota expanded during interglacial periods, including the Holocene, and colonised central and northern Europe (Taberlet et al., 1998; Hewitt, 1999; Tzedakis et al., 2013). Especially, the Balkan Peninsula as part of the eastern Mediterranean is standing out in its role as a source for postglacial colonisation of Europe (Hewitt, 2011; Nieto Feliner, 2014). A lower impact of the Ice Ages coupled with climatic and topographic complexity of the Balkans facilitated spatiotemporal processes, enabling glacial survival and triggered a differentiation of biota. Some plants experienced stable long-term persistence without extensive differentiation and range expansion during the Holocene (e.g., *Aesculus hippocastanum*, Walas et al., 2019; *Euphorbia heldreichii*, Caković and Frajman, 2020), while other groups underwent enhanced population differentiation and speciation. Recent phylogeographic studies focusing on the southern Balkan Peninsula and adjacent Aegean Basin revealed the importance of this area for genetic diversification and speciation and indicated its marginal role in northward expansion, as seen in the *Alyssum montanum-repens* complex (Španiel et al., 2017), annual *Alyssum* species (Cetlová et al., 2021), *Campanula* (Crowl et al., 2015), *Cymbalaria* (Carnicero et al., 2020), the *Euphorbia verrucosa* alliance (Caković et al., 2021) and the genus *Nigella* (Jaros et al., 2018). Similarly, in *Veronica chamaedrys* (Bardy et al., 2010) and *Edraianthus graminifolius* (Surina et al., 2014), southern populations exhibited deeper and older differentiation not followed by lineage expansions. On the other hand, several more northern lineages successfully expanded from the Balkan Peninsula and colonised central Europe and areas beyond (Magri et al., 2006; Frajman and Oxelman, 2007; Bardy et al., 2010; Rešetnik et al., 2016; Đurović et al., 2017; Caković et al., 2021). In addition, multiple Balkan lineages migrated *trans*-Adriatically and colonised the Apennine Peninsula (e.g., Rešetnik et al., 2016; Frajman and Schönswetter, 2017; Falch et al., 2019) or expanded to the Carpathians and the Pontic area (e.g., Frajman and Oxelman, 2007; Puşcaş et al., 2008; Csergö et al., 2009; Ronikier, 2011; Stachurska-Swakoń et al., 2013; Đurović et al., 2017).

The initial simplistic scenario of exclusively southern peninsular refugia has been challenged because of growing evidence suggesting more complex patterns, including glacial survival in more northern European regions. 'Cryptic' northern refugia were not only inferred for cold-tolerant arctic, boreal, and subalpine species (e.g., Stachurska-Swakoń et al., 2013; Ronikier and Zalewska-Gałosz, 2014) but also for temperate species, such as *Arabidopsis arenosa* (Kolář et al., 2016), *Arabidopsis halleri* (Šrámková-Fuxová et al., 2017), *Cyclamen purpurascens* (Slovák et al., 2012), *Erythronium dens-canis* (Bartha et al., 2015), *Euphorbia ilirica* (Frajman et al., 2016), *Helleborus niger* (Záveská et al., 2021), *Hepatica transsilvanica* (Laczkó and Sramkó, 2020), and *Rosa pendulina* (Daneck et al., 2016). These studies imply that phylogeographic connections between the Balkan Peninsula and central Europe (Bardy et al., 2010; Slovák et al., 2012; Rešetnik et al., 2016; Záveská et al., 2021) are older than previously thought and not necessarily examples of postglacial range expansions but rather cases of local LGM survival in central European refugia.

Aurinia saxatilis (L.) Desv. is a species ranging from central Europe (northern Pannonia, Moravia) over the Carpathians to the central, eastern, and southern Balkan Peninsula and adjacent shores of western Asia Minor as well as the southern Apennine Peninsula (Persson, 1971). It is the most widely distributed of seven *Aurinia* species, which have, as yet, unresolved phylogenetic relationships (Rešetnik et al., 2013). It is an up to 30-cm tall evergreen perennial plant having rosette- and stem leaves with stellate indumentum and long branched inflorescence-bearing numerous small, bright yellow flowers. Seeds of this species are thin, rounded, and broadly winged, which can facilitate wind dispersal. It is a popular ornamental plant in gardens, with the common name Basket of Gold referring to its numerous bright yellow flowers. *Aurinia saxatilis* is a diploid species (Lysak et al., 2009), growing on calcareous rocky grounds and dry soils mainly as saxicolous chasmophyte-forming vegetation of walls, cliffs, and rocky places (Persson, 1971). Because of its morphological variability, it has been divided into three subspecies having partly overlapping distributions (Dudley, 1964; Persson, 1971; Akeroyd, 1993; Plazibat, 2009). Typical *A. saxatilis* is widespread in central and southeastern Europe, *A. saxatilis* subsp. *orientalis* (Ard.) T.R. Dudley is distributed in the Balkan Peninsula, southern Carpathians, and western Anatolia, whereas *A. saxatilis* subsp. *megalocarpa* (Hausskn.) T.R. Dudley is limited to the southern Apennine Peninsula and the Aegean region (including Crete and the west Anatolian coast).

To explore the evolutionary dynamics and phylogeographic patterns in *A. saxatilis*, here, we combine phylogenomics and species distribution modelling. First, using plastid *ndhF* sequences, we reconstruct phylogenetic relationships and establish a temporal outline of its diversification. Second, using genome-wide SNPs generated by restriction-site associated DNA sequencing (RADseq) coupled with demographic and species distribution modelling, we reveal range wide patterns of

genetic diversity, compare alternative scenarios of population divergences, and infer the species' evolutionary history. We further explore whether the distribution of *A. saxatilis* in eastern and central Europe is a consequence of post-glacial expansion from the southern Balkan refugium or if there is evidence for *in situ* glacial survival. Finally, we intersect the phylogenetic data with infraspecific entities and discuss taxonomic implications.

MATERIALS AND METHODS

Plant Material

We sampled 77 populations of *A. saxatilis* throughout its range (**Supplementary Figure 1**). In addition to herbarium specimens, a leaf material of one to five individuals per population was collected in silica gel. Voucher data and GenBank numbers are presented in **Supplementary Table 1**, and the geographic origin of the sampled populations is shown in **Supplementary Figure 1**. For one individual from 75 populations, we sequenced the plastid *ndh*F region, and for one to three individuals from 63 populations, RADseq was performed. Additionally, in plastid analyses we used six published GenBank sequences of *Aurinia* and 56 of other Alysseae (**Supplementary Table 2**). For RADseq, we used the published genome of *Odontarrhena argentea* (BioSample: SAMEA4639485) as a reference for mapping raw reads and as an outgroup.

Wet Lab Methods

Extraction of total genomic DNA was performed from silica gel-dried leaves using GenElute Plant Genomic DNA Miniprep Kit (Sigma–Aldrich) and following the manufacturer's instructions. PCR and sequencing of *ndhF* were performed as described in Rešetnik et al. (2013) using the primers 5F, 989R, 989F, 1703R, 1410F, and 2100R (Beilstein et al., 2006). After checking amplicons on 1% TBE-agarose gel, they were purified enzymatically using exonuclease I and shrimp alkaline phosphatase (SAP; Fermentas) following the manufacturer's instructions. Sequencing was carried out at Macrogen Europe using the same primers as for amplification. Single-digest RADseq libraries were prepared using the restriction enzyme *Pst*I (New England Biolabs) and a protocol adapted from Paun et al. (2016) and Záveská et al. (2021). Up to three individuals of each of 63 populations of *A. saxatilis* were sequenced on Illumina HiSeq at VBCF NGS Unit[1] as 100-bp single-end reads.

Plastid Data Analyses

Contigs of plastid sequences were assembled and edited, and the sequences were aligned using Geneious Pro 10.2.3[2]. Maximum parsimony (MP) and MP bootstrap (MPB) analyses were performed using PAUP 4.0b10 (Swofford, 2002). The most parsimonious trees were searched for heuristically with 1,000 replicates of random sequence addition, TBR swapping and MulTrees on. Swapping was performed on a maximum of 1,000 trees (nchuck = 1000). All characters were equally weighted and unordered. The dataset was bootstrapped using full heuristics, 1,000 replicates, TBR branch swapping, MulTrees option off, and random addition sequence with five replicates. *Berteroa incana*, *Berteroa obliqua*, *Galitzkya macrocarpa*, and *Lepidotrichum uechtritzianum* were used for rooting based on Rešetnik et al. (2013). Bayesian analyses were performed with the MrBayes 3.2.2 (Ronquist et al., 2012) applying the GTR + G substitution model proposed by the Akaike information criterion implemented in the MrModelTest (Nylander, 2004). The settings for the Metropolis-coupled Markov chain Monte Carlo process included two runs with four chains each, run simultaneously for 10,000,000 generations each, sampling trees every 1,000th generation using default priors. The posterior probability (PP) of the phylogeny and its branches was determined from the combined set of trees, discarding the first 2,500 trees of each run as burn-in. Convergence of the MCMC procedure was assessed further by calculating effective sample sizes (ESSs) with the programme Tracer 1.6.0 (Rambaut et al., 2014). We also analysed the plastid sequences using statistical parsimony as implemented in TCS (Clement et al., 2000) with the connection limit set to 95% and gaps (all being 1 bp long) treated as fifth character state, because all the chromatograms were unambiguous, thus enabling the usage of gap characters in the TCS analysis.

Divergence times were estimated using the plastid sequences and following dating approaches of Huang et al. (2020) and Walden et al. (2020). More precisely, we selected four accessions representing major lineages of *A. saxatilis* as revealed by analyses of the complete plastid dataset (see above) and aligned them to sequences representing other *Aurinia* species and different outgroup genera as used in Huang et al. (2020) and Walden et al. (2020), respectively. In preliminary analyses we analysed both datasets, applying different secondary calibration points from both studies. In the first case, we set the age of the clade corresponding to tribe Alysseae to 13.8–19.9 Mya (95% values of highest posterior densities, HPDs) according to Huang et al. (2020), whereas in the second case, we set the age of this clade to 12.78–21.03 Mya following Walden et al. (2020). As both analyses resulted in comparable age estimations for *Aurinia*, here, we more precisely present only the analysis following Huang et al. (2020). Dating analysis was performed using the BEAST 1.8.2 (Drummond et al., 2012), applying the birth-death speciation prior (Gernhard, 2008) and the GTR + G substitution model with estimated base frequencies, and a lognormal relaxed clock with a weakly informative prior on clock rate (exponential with mean 0.001). The prior age of the root was set to 17.1 million years with a normally distributed standard deviation of 1.45, which corresponds to the median age and roughly to 95% HPD interval 13.8–19.9 Mya of the corresponding node inferred in Huang et al. (2020). Two MCMC chains were run for 100 million generations, logging parameters every 10,000 generations. The performance of the analysis was checked with Tracer 1.6.0 (Rambaut et al., 2014); both the effective sample sizes and mixing were deemed appropriate. Maximum clade credibility (MCC) trees were produced and annotated with Tree Annotator (part of the BEAST package) after removing 10% burn-in and combining the two tree files with the LogCombiner (part of the BEAST package), and visualised with the FigTree 1.4.2 (Rambaut, 2014).

[1] http://www.vbcf.ac.at/ngs/

[2] https://www.geneious.com

RADseq: Variant Calling and Filtering

The raw reads were demultiplexed according to index barcodes using the BamIndexDecoder 1.03[3] and based on inline barcodes with process_radtags.pl implemented in the Stacks 1.35 (Catchen et al., 2011, 2013) with default settings. We used an Illumina short-read assembly of *O. argentea* (BioSample: SAMEA4639485) as a reference for mapping the raw reads of *A. saxatilis* samples with the mem algorithm of the BWA 0.7.5a-r405 (Li and Durbin, 2009). Mapping files were then sorted by reference coordinates, and read groups were added with the Picard Tools 2.0.1[4]. Realignments around indels were performed for each bam file using the Genome Analysis Toolkit 3.6-0-g89b7209 (McKenna et al., 2010). We conducted the realignments around indels to correct mapping errors made by genome aligners and to make read alignments more consistent in regions that contain indels. It follows the older GATK > v.4 best practice recommendation. It seems that this step is redundant, at least for RADseq data, but we kept it in our pipeline, similarly as we did in Záveská et al. (2019), or as applied in Brandrud et al. (2020). Then, ref_map.pl from Stacks was applied to the mapped bam files with the requirement of a minimum coverage to identify an allele (−m) of 4×. The programme Populations implemented in Stacks was used to export selected loci. Filtering on (i) number of heterozygous genotypes per locus, maximum observed heterozygosity (–max_obs_het) 0.65 to avoid combining paralogs within the same RAD locus, and (ii) missing data for a minimum of 20 out of 63 populations present with at least 20% of individuals from each population (−p 20 and −r 0.2 flags) were applied to all the datasets. For the first dataset, later used for ML phylogenetic reconstruction, multiple SNPs per locus were exported into full sequence phylip format using flags '—phylip,' '–phylip-var,' and '–phylip-var-all.' For Bayesian clustering, a set of RADseq loci was exported into structure format using the –structure flag, and the –write-single-snp flag was used to select only a single (first) SNP per fragment to minimise the chance of selecting linked loci. Finally, for species tree reconstruction *via* SNAPP and for preparation of site frequency spectra (SFS) for demographic modelling, we selected two to 23 individuals with lowest proportion of missing data from each fast STRUCTURE group (see below). For those individuals, we exported an SNP per locus in vcf format using the programme Populations using the –vcf flag and further filtered using the VCFtools 0.1.15 (Danecek et al., 2011) for a maximum of 40% of missing data and a minimum depth of coverage of 10×.

RADseq: Exploratory Analyses of SNP Data

To infer phylogenetic relationships among all 180 individuals of *A. saxatilis* and one outgroup sample of *O. argentea*, we computed a maximum likelihood (ML) phylogeny based on full sequences of 2,645 concatenated RADSeq loci using the RAxML 8.2.8 (Stamatakis, 2014) with random starting trees and the GTR + G nucleotide substitution model. The best-scoring ML tree was bootstrapped with the frequency-based stopping criterion (Pattengale et al., 2010).

For 180 individuals of *A. saxatilis*, the optimal grouping of populations was determined using the Bayesian clustering software for large SNP datasets and the fastSTRUCTURE v1.0 (Raj et al., 2014), and for each *K* from 1 to 10 with provided script structure.py using a simple prior. The optimal value of K was determined by the chooseK.py script packaged with the fastSTRUCTURE. The fastSTRUCTURE was used for the entire dataset and then for the three subsets representing three main groups inferred by the analysis of the entire dataset. Based on the same dataset, further filtered for maximum of 30% missing data, we plotted genetic distances among individuals by principal coordinate analysis (PCoA) calculated with the R package Adegenet 2.1.0 (Jombart, 2008; Jombart and Ahmed, 2011) based on Euclidean distances. To account for potential presence of hybridisation within the entire data set, we used the SplitsTree 4.12.6 (Huson and Bryant, 2006) to create a the NeighborNet based on Hamming distances (Hamming, 1950). The programme Populations in Stacks was used to estimate the number of private alleles and nucleotide diversity (π) per population.

To estimate divergence times, we used the species tree method SNAPP 1.5.1 (Bryant et al., 2012) in the BEAST 2.6.4 (Bouckaert et al., 2014) as described in Stange et al. (2018). We used a reduced data set containing 49 individuals proportionally sampled from five lineages identified by the RAxML and the fastSTRUCTURE, excluding the admixed east Balkan and North Macedonian populations (see Section "Results"). We constructed a new RAD data subset for these five lineages using the filtering parameters described above but requiring loci to be shared among all the 49 samples and further filtering of biallelic SNPs only. To scale the tree, we used a secondary calibration point obtained from the results of dating analysis based on the plastid data. In particular, the prior age of the root (i.e., diversification of *A. saxatilis*) was set to 2.2 Mya with a normally distributed SD of 0.9, which corresponds to the median age and 95% HPD interval of the corresponding node (see Section "Results"). To improve mixing and convergence of the model, we constrained the monophyly of two ingroup clades that were consistently resolved as monophyletic with strong support in RAxML analysis, in particular the Central European and East Balkan-Carpathian Group and the South Balkan-Apennine Group. We ran two independent MCMCs for 100,000 generations, discarding 30% of the generations as burn-in. Log and tree traces from the two runs were combined. We estimated the maximum clade credibility (MCC) tree from the posterior distribution using the TreeAnnotator 2.5.0.

RADseq: Demographic Modelling

To investigate alternative divergence scenarios of the three geographically adjacent groups of *A. saxatilis* revealed by the fastSTRUCTURE analysis (the Core Central Balkan Subgroup, the East Balkan-Carpathian Group without admixed populations, and the Central European Group, see Section "Results"), we used the diffusion approximation method of dadi to analyse two-dimensional (2D-JSFS) site frequency spectra (Gutenkunst et al., 2009). We used an established 2D analysis

[3]https://github.com/wtsi-npg/illumina2bam

[4]http://broadinstitute.github.io/picard

pipeline (Portik et al., 2017; Charles et al., 2018) and adapted publicly available python scripts[5] that define 2D models, perform model fitting, and execute plotting functions. As input data, we used a two-dimensional joint site frequency spectrum prepared *via* conversion of a vcf file to a folded SFS that included down-projection of the sampling size to maximise the number of sampled individuals while minimising levels of missing data for downstream multi-population comparisons and was performed using the easySFS tool[6]. We excluded all admixed individuals. The down-projection and filtering resulted in the following allele numbers: Core Central Balkan Subgroup, 24 alleles; East Balkan-Carpathian Group, 16 alleles; Central European Group, 18 alleles.

We applied a 2D analysis pipeline for pairwise comparison of two neighbouring groups of populations, i.e., the Core Central Balkan vs. the East Balkan-Carpathian Group and the East Balkan-Carpathian vs. the Central European Group, respectively. We tested whether one of the groups is a result of postglacial expansion from the other group, or if both groups survived at least the LGM in separate refugia. Potential scenarios for two groups, thus, represent events of (1) old (preglacial) vicariance, (2) old (preglacial) founder event, and (3) recent (postglacial) founder event as depicted in **Supplementary Figure 2A** *via* particular models (**Supplementary Figures 2B, 3**). Under the last two scenarios, founded groups are supposed to have experienced at least one bottleneck reflecting decrease in effective population size during the LGM followed by population expansion during the Holocene. However, detection of exponential growth in the last ten thousand years is challenging especially for organisms with long generation time (>10 years; Adams and Hudson, 2004; Elleouet and Aitken, 2018). Therefore, we chose to test simplified models, including either (i) no change in population size since the split, (ii) early stage of population expansion followed by constant population size, or (iii) continuously expanding population size (**Supplementary Figure 2B**). In a demographic context, we refer to genetic groups identified by the fastSTRUCTURE analyses as "populations" in accordance with recent studies (e.g., Portik et al., 2017; Charles et al., 2018).

To be able to differentiate the scenarios of a pre-glacial vs. a post-glacial split of the populations, we dated the splits between the groups using SNAPP (see above). For both pairwise comparisons, we tested eight island demographic models ('island' referring to the derived, recently founded population in contrast to the ancestral 'mainland' population; we use the more intuitive terms 'derived population' and 'ancestral founding population' in the following). These eight models represent the three main hypotheses and modifications with ancestral or recent migration included to explain additional features of the 2D-JSFS (**Supplementary Figure 3**). Models representing an old vicariance and a recent founder event were taken from Charles et al. (2018), and models of old founder events were designed in Záveská et al. (2021). For models in both the vicariance and founder event categories, we followed Charles et al. (2018) and included a variable "s" that defines the fraction of the ancestral population (nuA) founding each daughter population, where nuA*s represents the derived population and nuA*$(1-s)$ represents the ancestral founding population. We enforced an upper limit of 0.5 for *s*. In the optimization of the models, we followed Záveská et al. (2021). The optimised models were compared using Akaike information criterion (AIC), and the replicate with highest likelihood for each model was used to calculate AIC scores, ΔAIC scores and Akaike weights (ωi) (Burnham and Anderson, 2002). We did not transform parameters into biologically meaningful estimates, because our primary aim was to perform model selection.

Species Distribution Modelling

To infer the potential present and past distribution of *A. saxatilis*, distribution models were generated using different modelling algorithms implemented in the *biomod2* R package (Thuiller, 2003; Thuiller et al., 2009, 2016) following an ensemble forecasting approach. Occurrence data were obtained through extensive fieldwork in a period that ranged from 2005 to 2019. Localities within a distance of 1 km^2 were removed using the *spThin* R package (Aiello-Lammens et al., 2015) to account for spatial autocorrelation. The final dataset covering the whole geographic range of *A. saxatilis* consisted of 95 independent records, whereas the two subsets corresponding to two main genetic groups (see Section "Results" and **Supplementary Figure 1**) for which a sufficient number of geographic records was available included 38 (the South Balkan-Apennine Group) and 43 (the remainder of the area) records.

Eight climatic variables were selected based on their ecological relevance for *A. saxatilis* (**Supplementary Table 3**) and taken at a resolution of 30 arc-s (∼1 km) from the Chelsa Climate database (Karger et al., 2018). To account for potential multicollinearity between variables, variance inflation factor (VIF) and Pearson correlation coefficient were computed using the *usdm* R package (Naimi et al., 2014). Subsequently, three variables having a VIF factor greater than 10 or Pearson correlation coefficient (r) higher than 0.8 were excluded. The remaining five variables (temperature seasonality, bio04; mean daily mean air temperatures of the warmest quarter, bio10; precipitation amount of the wettest month, bio13; mean monthly precipitation amount of the driest quarter, bio17; mean monthly precipitation amount of the warmest quarter, bio18) were used for present and past species distribution modelling (SDM) of the total dataset as well as both subsets.

Projection of environmental space during the LGM period (approximately 22,000 years ago) was projected with the same set of variables available in the Chelsa climate PMIP3 dataset. Variables were calibrated according to seven different Paleoclimate Modelling Intercomparison Project (PMIP3) global circulation models (GCMs): NCAR-CCSM4, MRI-CGCM3, MPI-ESM-P, MIROC-ESM, CESS-FGOALS-g2, IPSL-CM5A-LR, and CNRM-CM5.

Species distribution modelling (SDM) was performed in *biomod2* by ensemble modelling. This method is based on the idea of "ensemble forecasting," a method in which many models are developed using different statistical techniques and later combined in one final ensemble model that is the best possible combination of all developed models (Araújo and New,

[5]https://github.com/dportik/dadi_pipeline
[6]https://github.com/isaacovercast/easySFS

2007). Species distribution models for each group were built using five modelling algorithms: Maximum Entropy (Maxent), Generalised Linear Model (GLM), Generalised Additive Model (GAM), Boosted Regression Trees (GBM), and Random Forest (RF). For each group and each algorithm, we ran 10 different models by cross-validation using a 70:30 data split, developing in total 50 models. Model performance was assessed based on the AUC value ("area under the receiver operating characteristic curve") in two steps. In the first step, the median AUC value of all the 50 models was calculated, and in the second step, the ensemble modelling process was run again but only for models whose AUC value was higher than the calculated median value. In this manner, we selected only models that performed best across all runs and all modelling techniques. The ensemble model calculated in step two was the final model that was used in ensemble forecasting during the LGM period but for each GCM separately. To obtain the final habitat suitability model for LGM, we calculated the average pixel value across all GCMs.

RESULTS

Plastid Sequence Phylogenies and Divergence Time Estimation

The plastid sequences of 75 *A. saxatilis* were 1,917 to 1,919 nucleotides long, after the exclusion of a section of ca. 70 bp between the first fragment (between primers 5F and 989R) and the second fragment (between primers 989F and 2100R). After the addition of six closely related *Aurinia* taxa and four outgroup sequences, the final alignment was 1,919-nucleotide long. Of 155 variable characters, 75 (3.9% of all the characters) were parsimony informative. A total of 2,499 best trees found in the parsimony analysis had a score of 188, and consistency index was 0.90 (0.82 after exclusion of uninformative characters). Bayesian and parsimony analyses resulted in largely congruent trees (**Figure 1A** and **Supplementary Figure 4**). *Aurinia* was resolved as monophyletic (MPB 100%, posterior probabilities, PP 1). In *Aurinia*, phylogenetic relationships were geography-correlated (**Figure 1B**) and several clades with good support (MPB > 59% and PP > 0.94) were resolved. One evolutionary lineage (MPB 77%, PP 1; violet in **Figure 1**), named hereafter the South Balkan-Apennine Group, included all accessions of *A. saxatilis* from the southern Balkan Peninsula (including the Aegean islands) and Italy, and the sympatric Greek endemic *Aurinia gionae* and *Aurinia moreana*. The second evolutionary lineage (MPB 90%, PP 1; brown in **Figure 1**), named hereafter the East Balkan-Carpathian Group, included all accessions of *A. saxatilis* from the eastern Balkan Peninsula, eastern Carpathians, and Apuseni Mts, and of allopatric *Aurinia leucadea* and *Aurinia sinuata*, both endemic to the Adriatic Basin. The third lineage (MPB 60%, PP 0.96; dark green in **Figure 1**), named hereafter the Central European Group, included all accessions of *A. saxatilis* from the northern Pannonian Basin, Western Carpathians, and Moravia and Bohemia. All other accessions of *A. saxatilis* distributed in the central Balkan Peninsula (light green in **Figure 1**) are named hereafter the Central Balkan Group; they were in basal polytomy (considering the support of MPB 55% and PP 0.52 as non-relevant) including central and southern Balkan *Aurinia corymbosa* and disjunctly distributed southeast Alpine and south Carpathian *Aurinia petraea*. In this group, the northwestern populations 368, 387, 405, 431, and 486 formed a clade with MPB 86% and PP 1, and the easternmost populations 177, 178, 434, and 435 formed a clade with MPB 63% and PP 0.95. In addition, some pairs of mostly geographically allied accessions had MPB > 59% and PP > 0.94 (see **Supplementary Figure 4** for details).

The phylogenetic relationships revealed in the trees were also reflected in the TCS haplotype network (**Figure 1C**): the Central Balkan Group was six mutational steps both from the South Balkan-Apennine Group and the East Balkan-Carpathian Group, whereas the Central European Group was only two mutational steps away from the first group. The South Balkan-Apennine Group was relatively uniform, with several populations sharing the same haplotype, with the most divergent being mostly those from the Aegean Basin, especially Crete (population 495). In all the other groups, haplotype diversity was higher, and most of the populations had their own haplotypes, and some of them were rather divergent (e.g., 173, 404, 405, 445, 488, and 460).

The phylogenetic relationships and inferred divergence times within the tribe Alysseae (**Figure 1D** and **Supplementary Figure 5**) corresponded to those inferred in Huang et al. (2020). *Aurinia* was monophyletic (PP 1) and originated 6.6 Mya (HPD 4–9.5), but started to diversify only 2.2 Mya (HPD 1.2–3.5) and continued to diversify throughout the Pleistocene. In *Aurinia*, the inferred relationships corresponded to the main groups inferred by phylogenetic analyses of the complete *Aurinia* dataset (see above), with the exception that a sister relationship between the South Balkan-Apennine Group (including *A. gionae* and *A. moreana*; PP 1) and all the other groups (PP 0.95) was inferred. In the latter clade, one lineage (PP 1) corresponded to the East Balkan-Carpathian Group (including *A. leucadea* and *A. sinuata*), and the other (PP 1) to the combined Central Balkan and Central European Groups.

RADseq Phylogenetic Relationships

Population Structure and Phylogenetic Relationships

The average number of high-quality reads per sample retained after demultiplexing and quality filtering was 0.68 million (*SD* = 0.21). The data have been deposited in the NCBI Short Archive (BioProject PRJNA761287, accessions number SRR15735925–SRR15736104).

A RAxML phylogenetic tree based on 27,791 SNPs (**Supplementary Table 4**) that was rooted using *O. argentea* resolved *A. saxatilis* as monophyletic (**Figure 2A** and **Supplementary Figure 6**). In *A. saxatilis*, four main clades with bootstrap support (BS) > 65% were identified. With the exception of the clade having BS 69%, they largely corresponded to the three main groups inferred by Bayesian population clustering based on 7,005 unlinked SNPs (**Supplementary Table 4**) using fastSTRUCTURE (colour codes in the first column in **Figure 2A**). The groups were allopatrically distributed (**Figure 2B**). With the exception of the strongly admixed

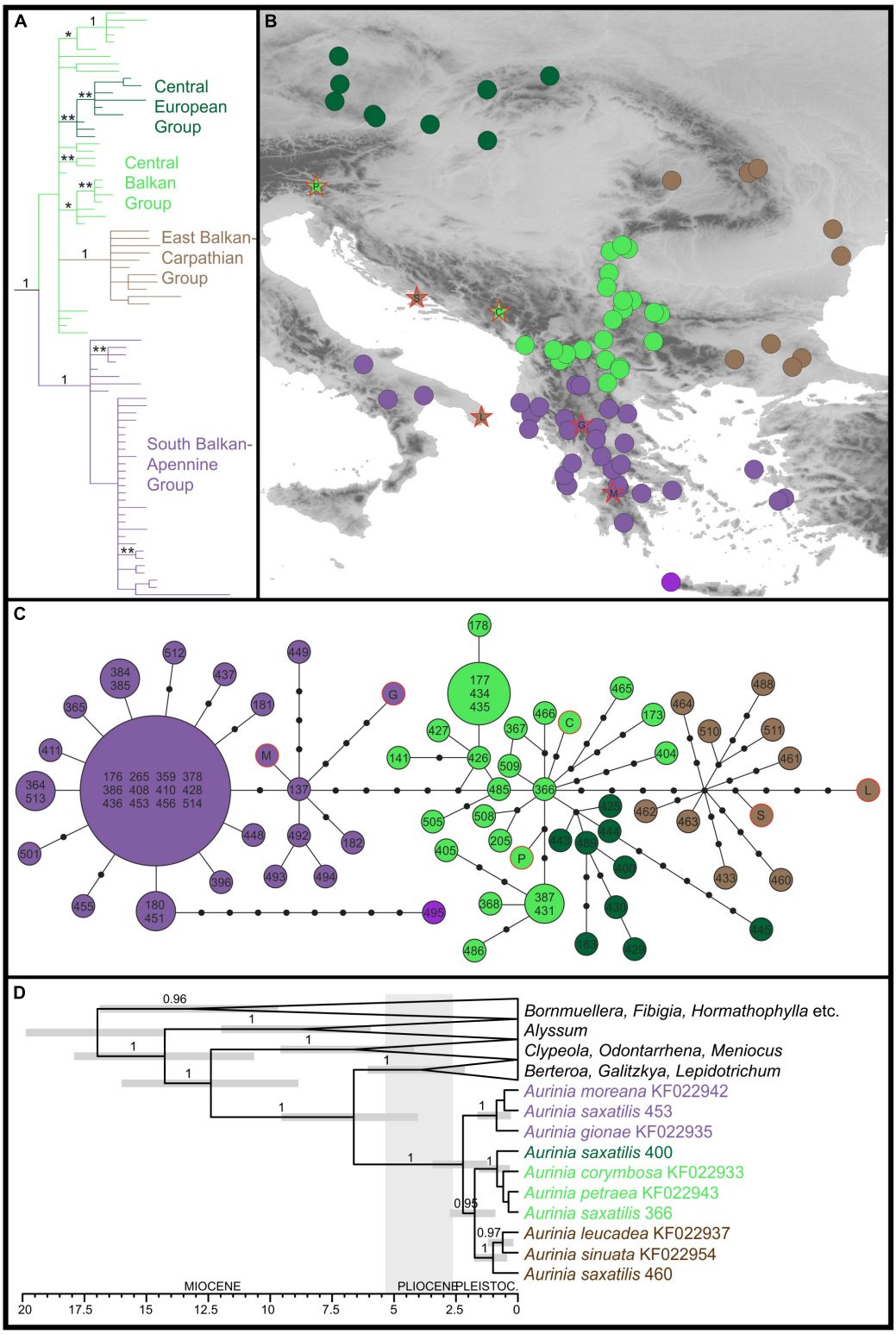


FIGURE 1 | Phylogenetic relationships inferred by plastid *ndh*F sequences of *Aurinia saxatilis* and related species. The colour codes indicate the main groups. **(A)** Simplified Bayesian consensus phylogram of the ingroup sequences (as shown in **Supplementary Figure 4**); numbers/asterisks above branches are PP values: *0.90–0.94; **0.95–0.99. **(B)** Geographic distribution of the main clade/haplotype groups inferred by the Bayesian analyses **(A)** and the statistical parsimony haplotype network **(C)**. **(D)** Simplified Bayesian chronogram indicating the time of divergences in *Aurinia* and in relation to related genera (for details, see **Supplementary Figure 5**). Population identifiers in **(C,D)** correspond to **Supplementary Table 1** and **Supplementary Figure 1**.

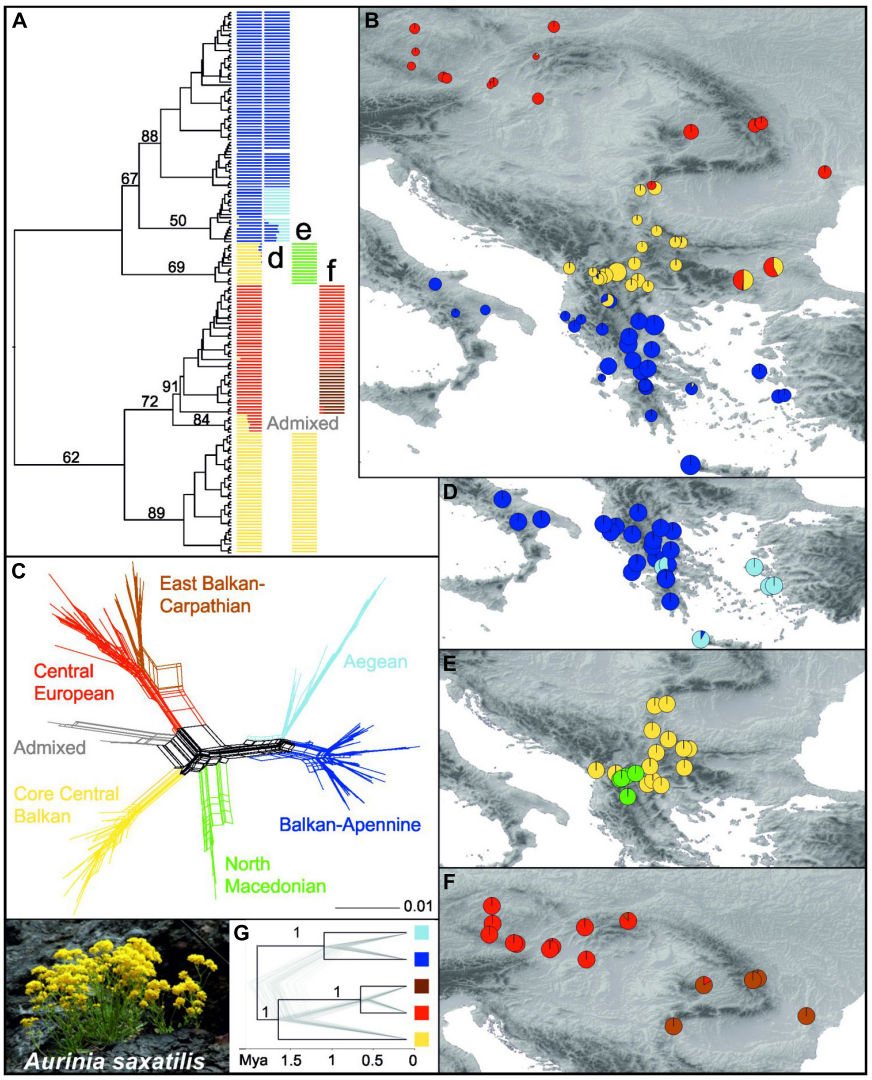


FIGURE 2 | Genetic structure in *Aurinia saxatilis* (photo by I. Rešetnik in the bottom left corner) based on restriction site-associated DNA sequencing (RADseq). **(A)** Simplified maximum likelihood tree as shown in **Supplementary Figure 6**, without outgroup and with bootstrap support of major clades above branches. The tree is complemented with barplot representations of the three main fastSTRUCTURE groups shown in **(B)** and their subgroups shown in **(D–F)**. In **(B)**, sizes of the pie charts reflect the number of private alleles in each population. **(C)** NeighborNet based on Hamming distances. **(G)** Simplified SNAPP analysis species tree with bootstrap support of major clades above branches, as shown in **Supplementary Figure 9**.

populations 433 and 488 between the red and the yellow clusters, and 428 between the blue and the yellow clusters, as well as the population 466 that was included in the red cluster, the RADseq clusters corresponded to the groups inferred with plastid sequences (see above; **Figure 1**), but the Central European Group and the East Balkan-Carpathian Group were united in the same cluster; therefore, for simplicity, we apply the same names of the groups as used for the plastid dataset. The South Balkan-Apennine Group (BS 67% in the RAxML tree) was divided into a subcluster composed of the Balkan and Apennine mainland populations (dark blue; Balkan-Apennine Subgroup), and a subcluster of populations from Crete and the Aegean islands (light blue; Aegean Subgroup; **Figure 2D**). The two subgroups were clearly divergent in NeighborNet (**Figure 2C**) and had BS 88 and 50% in the RAxML tree, respectively (**Figure 2A** and **Supplementary Figure 6**). The Central Balkan Group inferred with fastSTRUCTURE (yellow in **Figure 2B**) was split into two divergent lineages in the RAxML tree, with one having BS 89% and the other having BS 69% (**Figure 2A** and **Supplementary Figure 6**), which corresponded to two fastSTRUCTURE subclusters (**Figure 2E**), also clearly divergent in NeighborNet (**Figure 2C**): the North Macedonian Subgroup (green in **Figure 2E**) and the Core Central Balkan Subgroup (yellow in **Figure 2E**).

The combined East Balkan-Carpathian and Central European Groups (red in **Figure 2B**; BS 72% in the RAxML tree) included two strongly admixed populations, 488 and 433, from the eastern Balkan Peninsula that formed the earliest diverging lineage (BS

84%) within the group in the RAxML tree and were positioned intermediate between the Core Central Balkan Subgroup (yellow) and the remainder of the combined East Balkan-Carpathian and Central European Group (red–brown) in NeighbourNet (grey; **Figure 2C** and **Supplementary Figure 6**). These two combined groups were least divergent in NeighbourNet but divided by a fastSTRUCTURE analysis (brown and red; **Figure 2F**).

The scatterplot based on the principal coordinate analysis (PCoA) of 2,279 unlinked SNPs (**Supplementary Figure 7**) was highly congruent with the results of the RAxML, NeighbourNet, and fastSTRUCTURE analyses and revealed similar groupings of populations. Distances among the Core Central European Subgroup (yellow), combined East Balkan-Carpathian and Central European Groups (red-brown), and South Balkan-Apennine Group (blue) were similar, as revealed by the principal coordinates 1 and 2, while the admixed populations 433 and 488 from the eastern Balkan Peninsula and those of the North Macedonian Subgroup were intermediate among them; the latter were strongly divergent along the principal coordinate 3. The first three principal coordinates explained over 39% of the variation.

Although the total number of (both variable and invariant) sites recovered was comparable among the six subgroups (**Table 1**), the number of private alleles and nucleotide diversity (π) per population differed significantly among some of them (**Figure 2B** and **Supplementary Figure 8**). The lowest amounts of private alleles and lowest π values were observed in the Central European Group. The highest amount of private alleles was observed in the Aegean Subgroup, and highest π values were observed in the North Macedonian Subgroup (**Table 1** and **Supplementary Figure 8**).

Time-Calibrated Nuclear Data-Based Phylogeny

The species tree analysis SNAPP conducted for inference of a time-calibrated phylogeny was based on 1,914 unlinked SNPs and resulted in same relationships among the main lineages as in the RAxML tree but with higher support (PP 1) for all nodes (**Figure 2G** and **Supplementary Figure 9**). The onset of diversification of *A. saxatilis* was estimated at 1.9 Mya (95% HPD 0.6–2.9 Mya), followed by divergence between the Core Central Balkan Subgroup and the combined Central European and East Balkan-Carpathian Groups at 1.6 Mya (95% HPD 0.5–2.5 Mya). The Balkan-Apennine and Aegean subgroups diverged 1 Mya (95% HPD 0.3–1.6 Mya), followed by the divergence between the Central European Group and the East Balkan-Carpathian Group 0.6 Mya (95% HPD 0.2–0.9 Mya).

Demographic History

Detailed results of our demographic modelling are given in **Table 2**, **Supplementary Table 5**, and **Figure 3**. In both comparisons, the models of old divergence events fitted better to our data than the model of a recent founder event (**Supplementary Table 5**).

In comparison of the Core Central Balkan Subgroup and the East Balkan-Carpathian Group, we found a strong support (81% of the total model weight) for an old founder event with exponential size growth followed by a continuous asymmetric migration in the second epoch (ΔAIC = 4.18, ω_i = 0.81; **Figure 3A**, **Table 2**, and **Supplementary Table 5**). The proportion of ancestral founder population in the Carpathian Subgroup was estimated to be 0.1, and gene flow was inferred to be stronger from the East Balkan-Carpathian Group to the Core Central Balkan Subgroup (m_{12} = 0.23, m_{21} = 0.06; **Table 2**). The second best model (ω_i = 0.1) represents a similar scenario where the old founder event was replaced by a vicariance, while there was no migration in the first epoch, followed by an asymmetric migration in the second epoch.

In comparisons of the East Balkan-Carpathian and the Central European Groups, we performed model testing, so both groups acted either as ancestral population in founder event models or major fraction of the ancestral population in case of vicariance models. The best fitting model (ΔAIC = 2.02, ω_i = 0.42; **Figure 3B**) was a vicariance model with no migration (**Table 2** and **Supplementary Table 5**) where the East Balkan-Carpathian Group acted as a major fraction of the ancestral population (1 − s = 0.7). The second best model (ΔAIC = 2.3, ω_i = 0.15; **Supplementary Table 5**) suggested an old founder event of the Central European Group from the East Balkan-Carpathian Group, followed by an exponential size growth of the

TABLE 1 | Descriptors of the groups of populations of *Aurinia saxatilis* identified with restriction-site associated DNA sequencing (RADseq) data.

	Sites recovered	% polymorphic loci	Expected heterozygosity	Nucleotide diversity (π)	Private alleles
Central European Group	406119	3.4	0.0001	0.0001	31.3
East Balkan-Carpathian Group	361674.2	4.3	0.0001	0.0002	55.6
Core Central Balkan Subgroup	401131.1	4.8	0.0002	0.0003	48.3
North Macedonian Subgroup	425340.6	6.7	0.0003	0.0004	72.8
Balkan-Apennine Subgroup	421975.9	6.4	0.0002	0.0003	66.7
Aegean Subgroup	358335.2	4.5	0.0001	0.0002	81.7

The values are averaged over populations within each group.

TABLE 2 | Two best demographic models and unscaled parameter values for pairwise population comparisons.

population 1 vs. population 2	Model	Log-likelihood	ωi	Theta	nuA	nu1	nu2	T	T1	T2	*s*	m12	m21
Core Central Balkan Subgroup vs. East Balkan-Carpathian Group	founder_sec_contact_asym_two_epoch	−238.3	0.81	331.45	0.43	10.01	0.36		0.19	0.5	0.1	0.23	0.06
	vic_sec_contact_asym_mig	−240.4	0.10	171.1	1.4	0.03	1.05		0.43	0.68	0.4	0.09	0.05
East Balkan-Carpathian Group vs. Central European Group	vic_no_mig	−177.32	0.427	96.62	1.54	1.64	14.9	0.4			0.3		
	founder_anc_asym_two_epoch	−175.33	0.156	458.67	0.2	7.24	0.12		0.17	0.02	0.01	1.77	6.12

Detailed results of all the models tested are shown in ***Supplementary Table 5****. Abbreviations are as follows: founder_sec_contact_asym_two_epoch, old founder event with exponential size growth followed by asymmetric migration in the second epoch; vic_sec_contact_asym_mig, vicariance with no migration followed by asymmetric migration in the second epoch; vic_no_mig, vicariance with no migration; founder_anc_asym_two_epoch, old founder event with exponential size growth and continuous asymmetric migration followed by isolation;* ω_i*, Akaike weight; Theta [*$4N_{ref}\mu L$*, where L is the total length of the sequenced region from which single nucleotide polymorphisms (SNPs) were ascertained], effective mutation rate of the reference population (here corresponds to the ancestral population); nuA; effective population size of the ancestral population; nu1 and nu2, effective population sizes of populations 1 and 2, respectively, under the constant population size model; T, unscaled time of population split; T1, unscaled time between population split and ancient migration; T2, unscaled time between ancient migration and the present; s, fraction of the nuA running into population 2; m12, migration rate from population 2 to population 1; m21, migration rate from population 1 to population 2.*

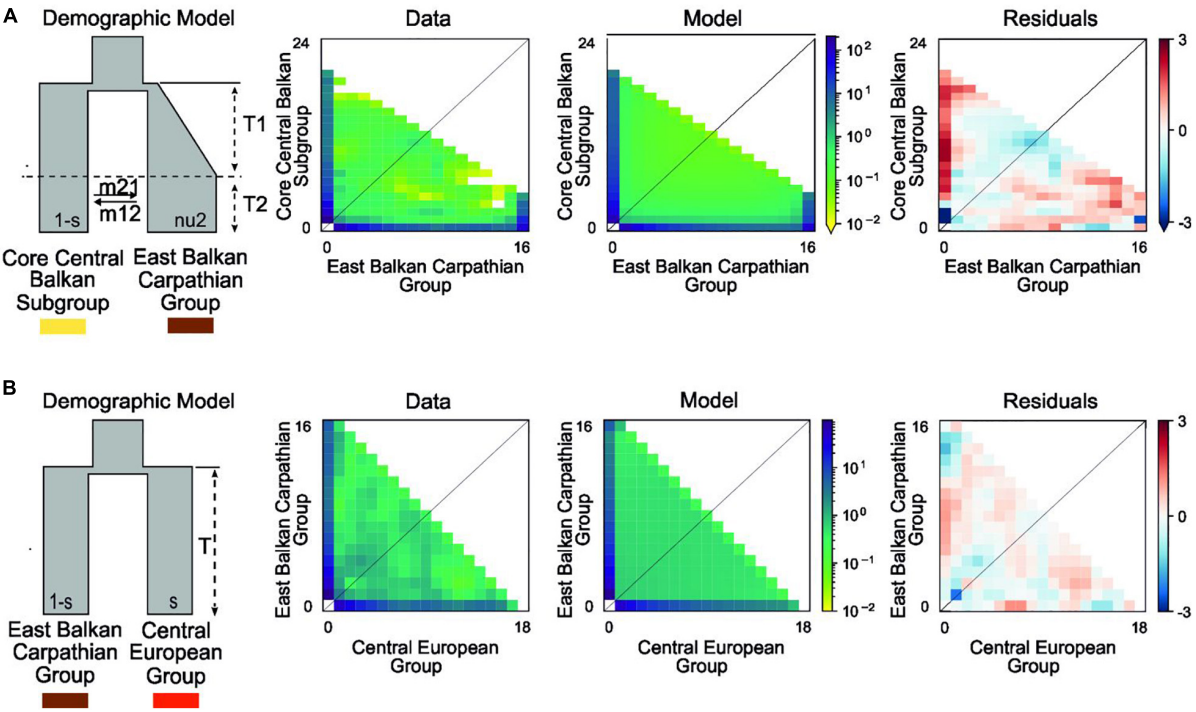


FIGURE 3 | Best-fitting demographic models for various phylogeographic groups of *Aurinia saxatilis* using a two-dimensional joint site frequency spectrum (2D-JSFS) between population sets including **(A)** the Core Central Balkan Subgroup (n = 24) and the East Balkan-Carpathian Group (n = 16) and **(B)** the East Balkan-Carpathian Group (n = 16) and the Central European Group (n = 18). A visual representation of the best-fit model is depicted, along with comparisons of the 2D-JSFS for data, model, and resulting residuals. For both comparisons, the best-fit model represents the scenario where the split is theoretically pre-glacial (see **Supplementary Figure 2** and Section "Introduction"). Additional models and parameter values are provided in **Table 2** and **Supplementary Table 5**.

Central European Group and continuous asymmetric migration ending with isolation.

Species Distribution Modelling

Mean AUC scores showed that the performance of all the models using current climate data was excellent (AUC > 0.95 in all cases; **Supplementary Table 6**). Moreover, sensitivity and specificity scores were always above 83, confirming the well-balanced performance of the developed models for current conditions. The present modelled habitat suitability agreed with the actual distribution of *A. saxatilis*, but some additional suitable areas with no known occurrences were indicated (**Figure 4A**). Additional regions identified as climatically highly suitable were the Black Sea coast and southerly adjacent regions of Asia Minor as well as the northwestern Apennine Peninsula and Sicily, and, to a lesser extent, the Adriatic Sea coastal area. SDMs hindcasted to the LGM revealed the southern parts of the Apennine and the Balkan Peninsula (including the Aegean

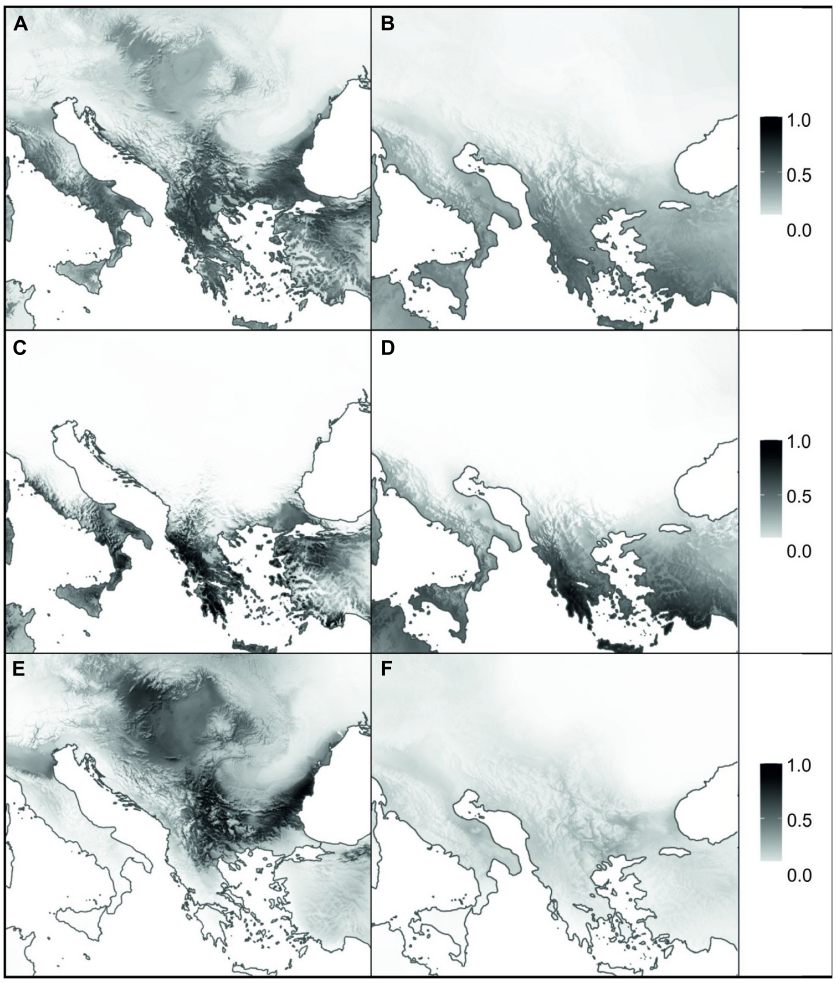


FIGURE 4 | Species distribution models for *Aurinia saxatilis* under current **(A,C,E)** and Last Glacial Maximum **(B,D,F)** climatic conditions. Panels **(A,B)** represent the models for complete distribution of *A. saxatilis*, whereas **(C–F)** are models for two distinct evolutionary lineages of *A. saxatilis*, namely the South Balkan-Apennine Group and the rest of the distribution, respectively (marked with blue and red symbols in **Supplementary Figure 1**). Grey shading corresponds to habitat suitability, ranging from 0 (white: no suitability) to 1 (black: maximum suitability).

islands and Crete) as well as Asia Minor as main areas of high suitability (**Figure 4B**).

Separate analyses of the South Balkan-Apennine Group indicated a long-term climatic stability and comparable area suitability for *A. saxatilis* at present and during the LGM (**Figures 4C,D**). On the other hand, the Central Balkan, East Balkan-Carpathian, and Central European Groups were predicted to have undergone substantial range contraction during the LGM compared to the present; there is limited support for potential refugia in southeastern Balkan-Pontic region and adjacent southern margins of the Carpathians (**Figures 4E,F**).

DISCUSSION

Our phylogenetic reconstructions coupled with species distribution modelling revealed that *A. saxatilis* had a long and dynamic history characterised by both vicariance and dispersal throughout Pleistocene, resulting in allo- to peripatrically distributed phylogeographic lineages. Both nuclear and plastid data suggested long-term isolation of the southern populations spanning the Eastern Mediterranean, from the southernmost Balkan Peninsula to the Aegean Basin and the southern Apennine Peninsula, with relative stability through time revealed by SDM. On the other hand, more northern inland populations experienced a turbulent history likely because of more severe impacts of Pleistocene climatic oscillations, triggering range contractions and secondary contacts with gene flow during range expansions. The results are more complex genetic patterns, which are partly incongruent between the plastid and nuclear data, although the main phylogeographic lineages match between the datasets. Our study, thus, corroborates the importance of the Eastern Mediterranean as a refugial area, enabling persistence of several lineages through time, but to a lesser extent its importance as a cradle of diversification (Mansion et al., 2009; Barres et al., 2013). In *A. saxatilis*, the cradle of diversification was the central

Balkan Peninsula from where, like in many other plant species (e.g., Magri et al., 2006; Bardy et al., 2010; Ronikier, 2011; Đurović et al., 2017; Kuzmanović et al., 2021), the Carpathians and Central Europe were colonised.

The molecular dating based on plastid sequences placed the origin of *Aurinia* in the Miocene–Pliocene boundary 6.6 Mya (95% HPD 4–9.5), with the onset of diversification of *A. saxatilis* dated to early Pleistocene 2.2 Mya (95% HPD 1.2–3.5; **Figure 1D**), or slightly later, based on dating analyses of the RADseq data (1.9 Mya, 95% HPD 0.6–2.9; **Figure 2G**). The diversification continued throughout the Pleistocene, and likely all main lineages diverged before the LGM, as the most recent divergence, i.e., the split between the Central European and the East Balkan-Carpathian Groups was dated to be 0.6 Mya (95% HPD 0.2–0.9). Various other temperate species exhibit deep divergences dating back to the mid Pleistocene, indicating that lineages can remain distinct during repeated glacial and interglacial intervals (e.g., Bardy et al., 2010; Záveská et al., 2021).

Long Term Persistence and Stability in the Eastern Mediterranean: Mountain Ridges and Sea Shaping Genetic Structure

Both nuclear-derived RADseq data and maternally inherited plastid sequences revealed a clear distinction of the southern populations of *A. saxatilis* from the southernmost Balkan Peninsula and the adjacent Aegean Basin, with a disjunction in the southern Apennine Peninsula (i.e., the South Balkan-Apennine Group). This area is characterised by a Mediterranean climate (Peel et al., 2007) and long-term climatic stability inferred by SDM, as areas suitable for *A. saxatilis* in this part of Europe did not differ considerably between the present and the LGM (**Figures 4A–D**), suggesting that the populations could persist in their locations throughout the Pleistocene and Holocene. A continuous persistent distribution in the mainland area with possible gene flow among populations resulted in a relatively uniform genetic structure, as a single haplotype is shared across multiple populations (**Figure 1C**) and relationships in the RADseq tree are poorly resolved in this group (**Figure 2** and **Supplementary Figure 6**). Outstanding in this respect is the mountainous Pindus range in central Greece, where populations harbour a larger number of private alleles (**Figure 2B**) and some populations also possess unique haplotypes (**Figure 1C**), suggesting local and weakly interconnected gene pools. As higher elevations were more affected by Pleistocene glaciations and the snow line on the Balkan Peninsula during the LGM was approximately 1,000 m lower than today (Turrill, 1929; Horvat et al., 1974), vertical migrations accompanied by range contractions and displacements leading to bottlenecks and hybridisation likely contributed to increased genetic peculiarity in this region. As *A. saxatilis* is typically found at mid-altitudes in deep valleys up to 1,000 m, higher mountain ridges can still act as a barrier to gene flow.

Contrary to the high genetic diversity in the mountainous areas over small distances, disjunct distant populations in the Apennine Peninsula are genetically depauperate and highly similar to the western Greek populations (**Figures 1**, **2**), which suggest their recent dispersal across the Adriatic Sea. In the RAxML tree of the RADseq data (**Supplementary Figure 6**), they diverged from the southern Balkan populations close to the tips of the tree, and two Apennine populations share the same haplotype with the southern Balkan populations (**Figure 1C**). *Aurinia saxatilis* is thus another example of recent, i.e., Pleistocene, *trans*-Adriatic dispersal and colonisation of the Apennine Peninsula (Frajman and Schönswetter, 2017; Falch et al., 2019; Garcia-Jacas et al., 2019; Janković et al., 2019) and corroborates the important role that the Balkans had not only for Central European diversity but also for the high diversity of the Apennine Peninsula.

The east- and southward expansion from mainland Greece to the Aegean archipelago and Crete happened much earlier, as the Aegean RADseq lineage, including a population from Attica, is the most early divergent in the South Balkan-Apennine Group (**Figure 2A** and **Supplementary Figure 6**), suggesting long isolation from the mainland populations and most likely recent migration back to Attica. As Crete was separated from the mainland since the Messinian Salinity Crisis (5.96–5.33 Mya; Poulakakis et al., 2015), i.e., much before the onset of diversification in *A. saxatilis*, the Crete and Aegean islands were likely colonised *via* long-distance dispersal of light-winged seeds. This migration was facilitated through land bridges during Pleistocene glaciations, enabling stepping-stone migrations. The expansion across the Aegean Basin was much more recent than the origin of the Aegean group, a scenario supported by a long branch and late diversification of this lineage in the RAxML tree (**Supplementary Figure 6**). However, whether the species persisted in isolation from the southern Balkan populations on one (or few) of the Aegean islands or in the westernmost Asia Minor, where *A. saxatilis* also occurs (Dudley, 1964), is impossible to answer because of our poor sampling in the Aegean basin. The colonisation of islands was likely connected with strong bottlenecks accompanied by genetic drift, as the Cretan population also harbours a high number of private alleles (**Figure 2B**) and a very divergent plastid haplotype (**Figure 1C**). The sea in the Aegeus, thus, likely had a similar effect on genetic structure as the high mountain peaks and ridges of the Pindus.

Turbulent Pleistocene History: From the Central Balkan Peninsula Over Carpathians to Central Europe

Contrary to the southern Balkan Peninsula that had high climatic stability through time, more northern populations of *A. saxatilis*, nowadays ranging to Central Europe, experienced a more turbulent history, as during the LGM (and probably during other glacials), climatically suitable areas decreased considerably (**Figure 4F**), triggering local extinctions and migrations. The main split between the south Balkan-Aegean populations and all other populations corresponds well with the boundary of the southern oro-Mediterranean and the northern temperate-continental climate (Horvat et al., 1974; Ronikier, 2011). Locally glaciated high mountain ranges of northern Greece could have acted and still act as a strong climatic barrier and a barrier to migration and gene flow of plants that are typically found at mid

altitudes (up to 1,000 m) or below, as in the case of *A. saxatilis*, the *Veronica chamaedrys* group (Bardy et al., 2010), *Edraianthus graminifolius* (Surina et al., 2014), and *Campanula versicolor* (Janković et al., 2019).

The divergence among the main lineages of *A. saxatilis* spanning from the central and eastern Balkan Peninsula over the Carpathians to Central Europe happened soon after the main genetic split discussed above. It is likely that climatic cooling and aridification that resulted in decrease of forested areas and increase in steppes (Svenning, 2003; Kirschner et al., 2020) enabled the rapid spread of *A. saxatilis* across southeast Europe in the early Pleistocene. This was followed by divergence in isolated refugia during glacial maxima, when climatically suitable areas diminished considerably as evidenced for LGM by SDM (**Figures 4E,F**). Smaller areas appropriate for species survival during the LGM (and likely other glacials) were more abundant in the central and eastern Balkans and southern margins of the Carpathians (**Figure 4**). Our demographic modelling (**Figure 3**) suggests that the east Balkan-Carpathian populations were derived from the central Balkan populations *via* an old founder event, which is consistent with the plastid data, where a long branch in the tree leads to the East Balkan-Carpathian Group from the polytomy of the Central Balkan populations (**Figure 1**). Colonisation of the Carpathians and the east Balkan-Pontic region from the central Balkan Peninsula was evidenced also in other plant groups (e.g., Frajman and Oxelman, 2007; Puşcaş et al., 2008; Csergö et al., 2009; Đurović et al., 2017).

Whereas the Central Balkan Group of *A. saxatilis* likely persisted glacial cycles *in situ* at lower altitudes, the East Balkan-Carpathian Group possibly had its refugium in the southern margin of the Carpathians and adjacent Pontic region. A similar genetic break, as inferred between the Central Balkan and East Balkan-Carpathian populations of *A. saxatilis,* is also seen in *Sesleria rigida* s.l. (Kuzmanović et al., 2013) and between vicariant sister species such as *Euphorbia niciciana* and *Euphorbia seguieriana* (Frajman et al., 2019) as well as *Campanula orbelica* and *Campanula alpina* (Ronikier and Zalewska-Gałosz, 2014), suggesting that several plants had glacial refugia divergent between the central Balkan Peninsula and the Carpathians.

Less clear is the origin of the Central European populations of *A. saxatilis* that diverged relatively late from the East Balkan-Carpathian Group, as suggested by the RADseq data (**Figure 2**), which best fit to a vicariant divergence (**Figure 3**), whereas the plastid data suggest an origin directly from the Central Balkan Group in a manner similar to that of the East Balkan-Carpathian Group (**Figure 1**). Independently of the origin of the Central European population, their high genomic similarity with the East Balkan-Carpathian populations and their late divergence indicated by the RADseq data (**Figure 2G**) suggest that they were connected throughout the majority of Pleistocene, and that they diverged before the LGM and, thus, likely had separate glacial refugia at least during the LGM. This is similar to the situation of *Cyclamen purpurascens* where one of the LGM refugia was suggested to be in the Tatra Mts (Slovák et al., 2012) and *Arabidopsis arenosa* for which Pannonian Basin and western Carpathians as northern refugia were postulated (Kolář et al., 2016). Enclaves of both coniferous and deciduous broadleaved trees, together with understory biota, existed in microclimatically favourable sites of central Europe (e.g., Magri et al., 2006; Kramp et al., 2009; Slovák et al., 2012), where *A. saxatilis* could have also survived the LGM, for example, in the Pannonian Basin where low climatic suitability was indicated by SDM (**Figure 4**).

As expected for a temperate species like *A. saxatilis*, its range increased in the face of climate warming in interglacials, and population groups isolated in glacial refugia came to secondary contacts, as evidenced by admixture patterns revealed by the RADseq data between the Central Balkan and the South Balkan-Apennine, as well as the Central Balkan and East Balkan-Carpathian Groups (**Figure 2**). Gene flow between the groups was evidenced not only by fastSTRUCTURE analyses but also by the intermediate position of some North Macedonian and eastern Balkan populations in NeighbourNet. Correspondingly, the high number of private alleles, high nucleotide diversity, and expected heterozygosity in the hybrid populations in North Macedonia and those in the eastern Balkans could be the result of the mixture of different genomic pools (**Figure 2B**, **Table 1**, and **Supplementary Figure 8**).

In summary, the central Balkan to central European populations of *A. saxatilis* had a turbulent demographic history highly influenced by Pleistocene climatic oscillations. They triggered bottlenecks and genetic drift during population contractions, and hybridisation during their expansions, resulting in complex genetic patterns, high plastid haplotype diversity and numerous unique haplotypes separated by multiple mutation steps (**Figure 1**) as well as slight incongruences between the plastid and genome-wide data. These processes contributed to high genetic diversity in the central parts of the Balkan Peninsula also in other plant groups (López-Vinyallonga et al., 2015; Đurović et al., 2017; Caković et al., 2021; Cetlová et al., 2021) and highlight this area not only as a cradle of lineage diversifications, but also as a source of lineage dispersals.

Taxonomic Implications

Despite the fact that other *Aurinia* species are nested in different clades of *A. saxatilis* in our plastid tree, either within sympatrically distributed populations as in the case of *A. gionae* and *A. moreana* or with allopatric populations being hundreds of kilometres away as in the case of *A. corymbosa*, *A. leucadea*, *A. petraea*, and *A. sinuata* (**Figure 1**), the nuclear data resolved *A. saxatilis* as monophyletic (ITS data in Rešetnik et al., 2013, preliminary analyses of RADseq data including all *Aurinia* species, Rešetnik et al., unpublished). Reasons for this incongruence between the nuclear and the plastid data could be incomplete lineage sorting, but also hybridisation and chloroplast capture (Rieseberg and Soltis, 1991) in the case of sympatric species.

Aurinia saxatilis is morphologically strongly polymorphic, and three subspecies were recognised (Dudley, 1964; Persson, 1971; Akeroyd, 1993; Plazibat, 2009). However, the quantitative characters used in subspecies delimitation (size, shape and pubescence of leaves, size and shape of siliculae, and length of styles) vary considerably among populations (Rešetnik, personal observations). Interestingly, the main genetic break inferred by both plastid and nuclear data in our study among the central and eastern Balkan, Carpathian, and central European

populations on one side and the southern Balkan-Apennine populations on the other correspond to *A. saxatilis* subsp. *saxatilis* and *A. saxatilis* subsp. *orientalis*, respectively. The main morphological characters distinguishing the two subspecies are related to the siliculae and rosette leaves: siliculae smaller and longer than wide with rounded apex, rosette leaves usually entire (*A. saxatilis* subsp. *saxatilis*), vs. siliculae larger and the same width and length or wider than longer, with emarginate or truncate apex, rosette leaves usually dentate to pinnatifid (*A. saxatilis* subsp. *orientalis*; Persson, 1971). Morphologically intermediate individuals occur mostly along the North Macedonian-Greek border (Persson, 1971), i.e., in the area where genetically admixed populations have been revealed by our study (**Figure 2B**).

The third subspecies, *A. saxatilis* subsp. *megalocarpa* is reported to be parapatric with *A. saxatilis* subsp. *orientalis*, with the former having larger quantitative characters of siliculae, style and seed wings (Dudley, 1964; Persson, 1971; Akeroyd, 1993). However, the measurements performed by Dudley (1964) and Persson (1971) are not congruent. They also partly disagree in distribution of *A. saxatilis* subsp. *megalocarpa*, for which Dudley (1964) indicated the Aegean islands and western Anatolia, whereas Persson (1971) also included southern Italy. The genetic divergence between the Balkan-Apennine Subgroup and the Aegean Subgroup in our RADseq data (**Figure 3**), thus, largely corresponds to Dudley's (1964) delimitation. More detailed morphological analyses with dense geographic sampling are needed to test the morphological divergence between both subspecies and generate precise data on their distributions and, thus, validate the taxonomic recognition of *A. saxatilis* subsp. *megalocarpa*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories: https://www.ncbi.nlm.nih.gov/genbank/, OK180986-OK181056, https://www.ncbi.nlm.nih.gov/, BioProject PRJNA761287, accessions number SRR15735925-SRR15736104.

AUTHOR CONTRIBUTIONS

IR conceived the study. IR and SB collected the majority of samples with additional samples provided by colleagues (listed in **Supplementary Table 1**). IR carried out the RADseq and PB plastid molecular laboratory work. IR and BF analysed the genetic data. EZ analysed the genomic data. MG produced the species distribution models. IR and BF wrote major parts of the manuscript, with exception of the parts about RADseq written by EZ. All authors read and edited the final version of the manuscript.

FUNDING

This study was financed by the Unity Through Knowledge Fund (project "Next generation systematics of the south-eastern European genus *Aurinia* (Brassicaceae): evolution and phylogeography of an intricate plant group" to IR), Croatian Science Foundation (project UIP-2017-05-2882 "Phylogeography and evolution of three ecologically divergent amphi-Adriatic plant groups" to IR), and Austrian Agency for International Cooperation and the Croatian Ministry of Science and Education (Austria-Croatia bilateral project 'HR17/2020' to BF and SB).

ACKNOWLEDGMENTS

We thank all the collectors listed in **Supplementary Table 1** for the help with collection of samples and Daniela Pirkebner for the help with RADseq library preparations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.822331/full#supplementary-material

REFERENCES

Adams, A. M., and Hudson, R. R. (2004). Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single nucleotide polymorphisms. *Genetics* 168, 1699–1712. doi: 10.1534/genetics.104.030171

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., and Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38, 541–545.

Akeroyd, J. R. (1993). "*Aurinia* (L.) Desv," in *Flora Europaea 1*, 2nd Edn, eds T. G. Tutin, V. H. Heywood, N. A. Burges, D. H. Valentine, and D. M. Moore (Cambridge: Cambridge University Press), 369–371.

Araújo, M. B., and New, M. (2007). Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47. doi: 10.1016/j.tree.2006.09.010

Bardy, K. E., Albach, D. C., Schneeweiss, G. M., Fischer, M. A., and Schönswetter, P. (2010). Disentangling phylogeography, polyploid evolution and taxonomy of a woodland herb (*Veronica chamaedrys* group, Plantaginaceae s.l.) in southeastern Europe. *Mol. Phylogenet. Evol.* 57, 771–786. doi: 10.1016/j.ympev.2010.06.025

Barres, L., Sanmartín, I., Anderson, C. L., Susanna, A., Buerki, S., Galbany-Casals, M., et al. (2013). Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). *Am. J. Bot.* 100, 867–882. doi: 10.3732/ajb.1200058

Bartha, L., Sramkó, G., Volkova, P. A., Surina, B., Ivanov, A. L., and Banciu, H. L. (2015). Patterns of plastid DNA differentiation in *Erythronium* (Liliaceae) are consistent with allopatric lineage divergence in Europe across longitude and latitude. *Plant Syst. Evol.* 301, 1747–1758. doi: 10.1007/s00606-014-1190-x

Beilstein, M. A., Al-Shehbaz, I. A., and Kellogg, E. A. (2006). Brassicaceae phylogeny and trichome evolution. *Am. J. Bot.* 93, 607–619. doi: 10.3732/ajb.93.4.607

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., et al. (2014). BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537

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

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

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*. New York, NY: Springer.

Caković, D., Cresti, L., Stešević, D., Schönswetter, P., and Frajman, B. (2021). High genetic and morphological diversification of the *Euphorbia verrucosa* alliance (Euphorbiaceae) in the Balkan and Iberian peninsulas. *Taxon* 70, 286–307.

Caković, D., and Frajman, B. (2020). Three tertiary *Euphorbia* species persisted in the forests of the Balkan Peninsula. *Plant Syst. Evol.* 306:50. doi: 10.1007/s00606-020-01672-w

Carnicero, P., Garcia-Jacas, N., Sáez, L., Constantinidis, T., and Galbany-Casals, M. (2020). Disentangling relationships among eastern Mediterranean *Cymbalaria* including description of a novel species from the southern Peloponnese (Greece). *Plant Syst. Evol.* 307:13. doi: 10.1007/s00606-020-01730-3

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., and Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3* 1, 171–182. doi: 10.1534/g3.111.000240

Cetlová, V., Zozomová-Lihová, J., Melichárková, A., Mártonfiová, L., and Španiel, S. (2021). Multiple drivers of high species diversity and endemism among *Alyssum* annuals in the Mediterranean: the evolutionary significance of the Aegean hotspot. *Front. Plant Sci.* 12:627909. doi: 10.3389/fpls.2021.627909

Charles, K. L., Bell, R. C., Blackburn, D. C., Burger, M., Fujita, M. K., Gvoždík, V., et al. (2018). Sky, sea, and forest islands: diversification in the African leaffolding frog *Afrixalus paradorsalis* (Anura: Hyperoliidae) of the lower Guineo- Congolian rain forest. *J. Biogeogr.* 45, 1781–1794. doi: 10.1111/jbi.13365

Clement, M., Posada, D., and Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1660. doi: 10.1046/j.1365-294x.2000.01020.x

Crowl, A. A., Visger, C. J., Mansion, G., Hand, R., Wu, H. H., Kamari, G., et al. (2015). Evolution and biogeography of the endemic *Roucela* complex (Campanulaceae: *Campanula*) in the Eastern Mediterranean. *Ecol. Evol.* 5, 5329–5343. doi: 10.1002/ece3.1791

Csergö, A. M., Schönswetter, P., Gyöngyvér, M., Deák, T., Boşcaiu, N., and Höhn, M. (2009). Genetic structure of peripheral, island-like populations: a case study of *Saponaria bellidifolia* Sm. (Caryophyllaceae) from the Southeastern Carpathians. *Plant Syst. Evol.* 278, 33–41. doi: 10.1007/s00606-008-0129-5

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and vcftools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Daneck, H., Fér, T., and Marhol, K. (2016). Glacial survival in northern refugia? Phylogeography of the temperate shrub *Rosa pendulina* L. (Rosaceae): AFLP vs. chloroplast DNA variation. *Biol. J. Linn. Soc.* 119, 704–718. doi: 10.1111/bij.12619

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Dudley, T. R. (1964). Synopsis of the genus *Aurinia* in Turkey. *J. Arnold Arbor.* 45, 390–400. doi: 10.5962/p.325004

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

Elleouet, J. S., and Aitken, S. N. (2018). Exploring approximate Bayesian computation for inferring recent demographic history with genomic markers in nonmodel species. *Mol. Ecol. Resour.* 18, 525–540. doi: 10.1111/1755-0998.12758

Falch, M., Schönswetter, P., and Frajman, B. (2019). Both vicariance and dispersal have shaped the genetic structure of Eastern Mediterranean *Euphorbia myrsinites* (Euphorbiaceae). *Perspect. Plant Ecol. Evol. Syst.* 39:125459. doi: 10.1016/j.ppees.2019.125459

Frajman, B., Graniszewska, M., and Schönswetter, P. (2016). Evolutionary patterns and morphological diversification within the European members of the *Euphorbia illirica* (*E. villosa*) group: one or several species? *Preslia* 88, 369–390.

Frajman, B., and Oxelman, B. (2007). Reticulate phylogenetics and phytogeographical structure of *Heliosperma* (Sileneae, Caryophyllaceae) inferred from chloroplast and nuclear DNA sequences. *Mol. Phylogenet. Evol.* 43, 140–155. doi: 10.1016/j.ympev.2006.11.003

Frajman, B., and Schönswetter, P. (2017). Amphi-Adriatic distributions in plants revisited: pleistocene trans-Adriatic dispersal in the *Euphorbia barrelieri* group (Euphorbiaceae). *Bot. J. Linn. Soc.* 185, 240–252. doi: 10.1093/botlinnean/box055

Frajman, B., Záveská, E., Gamisch, A., Moser, T., The STEPPE Consortium, and Schönswetter, P. (2019). Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Eurasian *Euphorbia seguieriana* s. l. (Euphorbiaceae). *Mol. Phylogenet. Evol.* 134, 238–252. doi: 10.1016/j.ympev.2018.10.046

Garcia-Jacas, N., López-Pujol, J., López-Vinyallonga, S., Janaćković, P., and Susanna, A. (2019). *Centaurea* subsect. *Phalolepis* in Southern Italy: ongoing speciation or species overestimation? Genetic evidence based on SSRs analyses. *System. Biodivers.* 17, 93–109. doi: 10.1080/14772000.2018.1549617

Gernhard, T. (2008). The conditioned reconstructed process. *J. Theor. Biol.* 253, 769–778. doi: 10.1016/j.jtbi.2008.04.005

Gömöry, D., Zhelev, P., and Brus, R. (2020). The Balkans: a genetic hotspot but not a universal colonization source for trees. *Plant Syst. Evol.* 306:5. doi: 10.1007/s00606-020-01647-x

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5:e1000695. doi: 10.1371/journal.pgen.1000695

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell Syst. Tech. J.* 29, 147–160.

Hewitt, G. M. (1999). Postglacial re-colonisation of European biota. *Biol. J. Linn. Soc.* 68, 87–112. doi: 10.1186/1471-2148-11-215

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

Horvat, I., Glavač, V., and Ellenberg, H. (1974). *Vegetation Südosteuropas*. Stuttgart: Fischer.

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

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

Janković, I., Satovic, Z., Liber, Z., Kuzmanović, N., Di Pietro, R., Radosavljević, I., et al. (2019). Genetic and morphological data reveal new insights into the taxonomy of *Campanula versicolor* s.l. (Campanulaceae). *Taxon* 68, 340–369. doi: 10.1002/tax.12050

Jaros, U., Tribsch, A., and Comes, H. P. (2018). Diversification in continental island archipelagos: new evidence on the roles of fragmentation, colonization and gene flow on the genetic divergence of Aegean *Nigella* (Ranunculaceae). *Ann. Bot.* 121, 241–254. doi: 10.1093/aob/mcx150

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

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., et al. (2018). Data from: climatologies at high resolution for the earth's land surface areas. *Dryad Digit. Rep.* doi: 10.5061/dryad.kd1d4

Kirschner, P., Záveská, E., Gamisch, A., Hilpold, A., Trucchi, E., Paun, O., et al. (2020). Long-term isolation of European steppe outposts boosts the biome's conservation value. *Nat. Commun.* 11:1968. doi: 10.1038/s41467-020-15620-2

Kolář, F., Fuxová, G., Záveská, E., Nagano, A. J., Hyklová, L., Lučanová, M., et al. (2016). Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. *Mol. Ecol.* 25, 3929–3949. doi: 10.1111/mec.13721

Kramp, K., Huck, S., Niketić, M., Tomović, G., and Schmitt, T. (2009). Multiple glacial refugia and complex postglacial range shifts of the obligatory woodland plant *Polygonatum verticillatum* (Convallariaceae). *Plant Biol.* 11, 392–404. doi: 10.1111/j.1438-8677.2008.00130.x

Kuzmanović, N., Comanescu, P., Frajman, B., Lazarević, M., Paun, O., Schönswetter, P., et al. (2013). Genetic, cytological and morphological differentiation within the Balkan-Carpathian *Sesleria rigida* sensu Fl. Eur. (Poaceae): a taxonomically intricate tetraploid-octoploid complex. *Taxon* 62, 458–472.

Kuzmanović, N., Lakušić, D., Frajman, B., Stevanovski, I., Conti, F., and Schönswetter, P. (2021). Long neglected diversity in the Accursed Mountains (western Balkan Peninsula): *Ranunculus bertisceus* is a genetically and morphologically divergent new species. *Bot. J. Linn. Soc.* 196, 384–406. doi: 10.1093/botlinnean/boab001

Laczkó, L., and Sramkó, G. (2020). *Hepatica transsilvanica* Fuss (Ranunculaceae) is an allotetraploid relictof the tertiary flora in Europe - Molecular phylogenetic evidence. *Acta Soc. Bot. Pol.* 89:8934. doi: 10.5586/asbp.8934

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

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

Lysak, M. A., Koch, M. A., Beaulieu, J. M., Meister, A., and Leitch, I. J. (2009). The dynamic ups and downs of genome size evolution in Brassicaceae. *Mol. Biol. Evol.* 26, 85–98. doi: 10.1093/molbev/msn223

Magri, D., Vendramin, G. G., Comps, B., Dupanloup, I., Geburek, T., Gömöry, D., et al. (2006). A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytol.* 171, 199–221. doi: 10.1111/j.1469-8137.2006.01740.x

Mansion, G., Selvi, F., Guggisberg, A., and Conti, E. (2009). Origin of Mediterranean insular endemics in the Boraginales: integrative evidence from molecular dating and ancestral area reconstruction. *J. Biogeogr.* 36, 1282–1296. doi: 10.1111/j.1365-2699.2009.02082.x

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., and Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography* 37, 191–203. doi: 10.1111/j.1600-0587.2013.00205.x

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

Nylander, J. A. A. (2004). *MrModeltest v2. Program Distributed by the Author*. Uppsala: Evolutionary Biology Centre, Uppsala University.

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R., Moret, B. M., and Stamatakis, A. (2010). How many bootstrap replicates are necessary? *J. Comput. Biol.* 17, 337–354. doi: 10.1089/cmb.2009.0179

Paun, O., Turner, B., Trucchi, E., Munzinger, J., Chase, M. W., and Samuel, R. (2016). Processes driving the adaptive radiation of a tropical tree (*Diospyros*, Ebenaceae) in New Caledonia, a biodiversity hotspot. *Syst. Biol.* 65, 212–227. doi: 10.1093/sysbio/syv076

Peel, M. C., Finlayson, B. L., and McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* 11, 1633–1644. doi: 10.5194/hess-11-1633-2007

Persson, J. (1971). Studies in the Aegean Flora XIX - Notes on *Alyssum* and some other genera of Cruciferare. *Bot. Not.* 124, 399–418.

Petit, R. J., Aguinagalde, I., De Beaulieu, J. L., Bittkau, C., Brewer, S., Cheddadi, R., et al. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300, 1563–1565. doi: 10.1126/science.1083264

Plazibat, M. (2009). A short synopsis of the tribe Alysseae (Brassicaceae) in Croatia with some taxonomic novelties. *Nat. Croat.* 18, 401–426.

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., et al. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol. Ecol.* 26, 5245–5263. doi: 10.1111/mec.14266

Poulakakis, N., Kapli, P., Lymberakis, P., Trichas, A., Vardinoyiannis, K., Sfenthourakis, S., et al. (2015). A review of phylogeographic analyses of animal taxa from the Aegean and surrounding regions. *J. Zoolog. Syst. Evol.* 53, 18–32.

Puşcaş, M., Choler, P., Tribsch, A., Gielly, L., Rioux, D., Gaudeul, M., et al. (2008). Post-glacial history of the dominant alpine sedge *Carex curvula* in the European Alpine system inferred from nuclear and chloroplast markers. *Mol. Ecol.* 17, 2417–2429. doi: 10.1111/j.1365-294X.2008.03751.x

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

Rambaut, A. (2014). *FigTree 1.4.2*. Available online at: http://tree.bio.ed.ac.uk/ (accessed November 10, 2020).

Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). *Tracer v 1.6*. Available online at: http://beast.bio.ed.ac.uk/Tracer (accessed November 10, 2020).

Rešetnik, I., Frajman, B., and Schönswetter, P. (2016). Heteroploid *Knautia drymeia* includes *K. gussonei* and cannot be separated into diagnosable subspecies. *Am. J. Bot.* 103, 1300–1313.

Rešetnik, I., Satovic, Z., Schneeweiss, G. M., and Liber, Z. (2013). Phylogenetic relationships in Brassicaceae tribe Alysseae inferred from nuclear ribosomal and chloroplast DNA sequence data. *Mol. Phylogenet. Evol.* 69, 772–786. doi: 10.1016/j.ympev.2013.06.026

Rieseberg, L. H., and Soltis, D. E. (1991). Phylogenetic consequences of cytoplasmicgene flow in plants. *Evol. Trends Plants* 5, 65–84.

Ronikier, M. (2011). Biogeography of high-mountain plants in the Carpathians: an emerging phylogeographical perspective. *Taxon* 60, 372–389.

Ronikier, M., and Zalewska-Gałosz, J. (2014). Independent evolutionary history between the Balkan ranges and more northerly mountains in *Campanula alpina* s.l. (Campanulaceae): genetic divergence and morphological segregation of taxa. *Taxon* 63, 116–131.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Slovák, M., Kučera, J., Turis, P., and Zozomová-Lihová, J. (2012). Multiple glacial refugia and postglacial colonization routes inferred for a woodland geophyte, *Cyclamen purpurascens*: Patterns concordant with the Pleistocene history of broadleaved and coniferous tree species. *Biol. J. Linn. Soc.* 105, 741–760.

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

Šrámková-Fuxová, G., Záveská, E., Kolář, F., Lučanová, M., Španiel, S., and 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. doi: 10.1093/botlinnean/box064

Stachurska-Swakoń, A., Cieślak, E., and Ronikier, M. (2013). Phylogeography of a subalpine tall-herb *Ranunculus platanifolius* (Ranunculaceae) reveals two main genetic lineages in the European mountains. *Bot. J. Linn. Soc.* 171, 413–428. doi: 10.1111/j.1095-8339.2012.01323.x

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stange, M., Sánchez-Villagra, M. R., Salzburger, W., and Matschiner, M. (2018). Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. *Syst. Biol.* 67, 681–699. doi: 10.1093/sysbio/syy006

Surina, B., Schneeweiss, G. M., Glasnović, P., and Schönswetter, P. (2014). Testing the efficiency of nested barriers to dispersal in the Mediterranean high mountain plant *Edraianthus graminifolius* (Campanulaceae). *Mol. Ecol.* 23, 2861–2875. doi: 10.1111/mec.12779

Svenning, J. C. (2003). Deterministic Plio-Pleistocene extinctions in the European cool-temperate tree flora. *Ecol. Lett.* 6, 646–653. doi: 10.1046/j.1461-0248.2003.00477.x

Swofford, D. L. (2002). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), ver. 4.0 beta 10*. Sunderland: Sinauer Associates.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G., and Cosson, J.-F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* 7, 453–464. doi: 10.1046/j.1365-294x.1998.00289.x

Thuiller, W. (2003). BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Glob. Chang. Biol.* 9, 1353–1362. doi: 10.1111/gcb.12728

Thuiller, W., Georges, D., Engler, R., and Breiner, F. (2016). *'biomod2': Ensemble Platform for Species Distribution Modelling*. Available online at: https://cran.r-project.org/web/packages/biomod2/biomod2.pdf (accessed June 11, 2021).

Thuiller, W., Lafourcade, B., Engler, R., and Araújo, M. B. (2009). BIOMOD-a platform for ensemble forecasting of species distributions. *Ecography* 32, 369–373. doi: 10.1111/j.1600-0587.2008.05742.x

Turrill, W. B. (1929). *The Plant-Life Of the Balkan Peninsula: A Phytogeographical Study*. Oxford: Clarendon.

Tzedakis, P. C., Emerson, B. C., and Hewitt, G. M. (2013). Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends Ecol. Evol.* 28, 696–704. doi: 10.1016/j.tree.2013.09.001

Walas, Ł, Ganatsas, P., Iszkuło, G., Thomas, P. A., and Dering, M. (2019). Spatial genetic structure and diversity of natural populations of *Aesculus hippocastanum* L. in Greece. *PLoS One* 14:e0226225. doi: 10.1371/journal.pone.0226225

Walden, N., German, D. A., Wolf, E. M., Kiefer, M., Rigault, P., Huang, X.-C., et al. (2020). Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. *Nat. Commun.* 11:3795. doi: 10.1038/s41467-020-1760.5-7

Záveská, E., Kirschner, P., Frajman, B., Wessely, J., Willner, W., Gattringer, A., et al. (2021). Evidence for glacial refugia of the forest understorey species *Helleborus niger* (Ranunculaceae) in the Southern as well as in the Northern Limestone Alps. *Front. Plant Sci.* 12:683043. doi: 10.3389/fpls.2021.683043

Záveská, E., Maylandt, C., Paun, O., Bertel, C., Frajman, B., Schönswetter, P., et al. (2019). Multiple auto-and allopolyploidisations marked the Pleistocene history of the widespread Eurasian steppe plant *Astragalus onobrychis* (Fabaceae). *Mol. Phylogenet. Evol.* 139:106572. doi: 10.1016/j.ympev.2019.106572

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 April 2022
doi: 10.3389/fpls.2022.873471

Genetic Variability in Balkan Paleoendemic Resurrection Plants *Ramonda serbica* and *R. nathaliae* Across Their Range and in the Zone of Sympatry

Maja Lazarević[1], Sonja Siljak-Yakovlev[2]*, Agathe Sanino[2], Marjan Niketić[3,4], Françoise Lamy[2,5], Damien D. Hinsinger[6], Gordana Tomović[1], Branka Stevanović[1], Vladimir Stevanović[4] and Thierry Robert[2,7]**

[1]Department of Plant Ecology and Phytogeography, Faculty of Biology, University of Belgrade, Belgrade, Serbia, [2]Ecologie Systématique Evolution, CNRS, AgroParisTech, Univ. Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France, [3]Natural History Museum, Belgrade, Serbia, [4]Serbian Academy of Sciences and Arts, Belgrade, Serbia, [5]Department of Biology, University of Versailles-Saint-Quentin, Versailles, France, [6]Département Biologie et Amélioration des Plantes, Polymorphisme des Génomes Végétaux, INRAE, Evry, France, [7]Biology Department, Sorbonne Université, Paris, France

OPEN ACCESS

Edited by
Gonzalo Nieto Feliner,
Real Jardín Botánico (RJB, CSIC), Spain

Reviewed by
Marcial Escudero,
Sevilla University, Spain
Eva Maria Temsch,
University of Vienna, Austria

***Correspondence:**
Maja Lazarević
majat@bio.bg.ac.rs
Sonja Siljak-Yakovlev
sonia.yakovlev@universite-paris-saclay.fr
Thierry Robert
thierry.robert@universite-paris-saclay.fr

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

Received: 10 February 2022
Accepted: 05 April 2022
Published: 28 April 2022

Citation:
Lazarević M, Siljak-Yakovlev S, Sanino A, Niketić M, Lamy F, Hinsinger DD, Tomović G, Stevanović B, Stevanović V and Robert T (2022) Genetic Variability in Balkan Paleoendemic Resurrection Plants Ramonda serbica and R. nathaliae Across Their Range and in the Zone of Sympatry. Front. Plant Sci. 13:873471. doi: 10.3389/fpls.2022.873471

The genus *Ramonda* includes three Paleoendemic and Tertiary relict species that survived in refugial habitats of the Balkan Peninsula (*R. nathaliae* and *R. serbica*) and the Iberian Peninsula (*R. myconi*). They are all "resurrection plants," a rare phenomenon among flowering plants in Europe. *Ramonda myconi* and *R. nathaliae* are diploids ($2n = 2x = 48$), while *R. serbica* is a hexaploid ($2n = 6x = 144$). The two Balkan species occur in sympatry in only two localities in eastern Serbia, where tetraploid potential hybrids ($2n = 4x = 96$) were found. This observation raised questions about the existence of gene flow between the two species and, more generally, about the evolutionary processes shaping their genetic diversity. To address this question, genetic markers (AFLP) and an estimate of genome size variation were used in a much larger sample and at a larger geographic scale than previously. The combination of AFLP markers and genome size results suggested ongoing processes of interspecific and interploidy hybridization in the two sites of sympatry. The data also showed that interspecific gene flow was strictly confined to sympatry. Elsewhere, both *Ramonda* species were characterized by low genetic diversity within populations and high population differentiation. This is consistent with the fact that the two species are highly fragmented into small and isolated populations, likely a consequence of their postglacial history. Within sympatry, enormous variability in cytotypes was observed, exceeding most reported cases of mixed ploidy in complex plant species (from $2x$ to $>8x$). The AFLP profiles of non-canonical ploidy levels indicated a diversity of origin pathways and that backcrosses probably occur between tetraploid interspecific hybrids and parental species. The question arises whether this diversity of cytotypes corresponds to a transient situation. If not, the question arises as to the genetic and ecological mechanisms that allow this diversity to be maintained over time.

Keywords: AFLP, gene flow, genetic diversity, genome size, Gesneriaceae, interploidy hybridization, mixed-ploidy zones, sympatry

INTRODUCTION

The Mediterranean basin is one of the richest regions in terms of biodiversity, with 23,000 to 25,000 plant species (Thompson, 2020). The Balkan Peninsula, as part of the Mediterranean Basin, one of the global hotspots of biodiversity (Myers et al., 2000), is one of the centers of species richness and endemism in Europe with more than 8,000 plant species, of which about 30% are endemic (Turill, 1929; Stevanović et al., 2007). Several geological, climatic, and biogeographical factors have been identified that have contributed to the richness and endemism of the flora of this region (Hewitt, 1996, 1999; Stevanović et al., 1999, 2007; Krstić et al., 2012; Kougioumoutzis et al., 2021). Polyploidization and interspecific hybridization have played an important role especially in the evolution of many plant taxa in the Balkan Peninsula, e.g., *Alyssum* (Španiel et al., 2017), *Cardamine* (Šlenker et al., 2021), *Centaurium* (Banjanac et al., 2017), *Cerastium* (Niketić et al., 2013, 2021), *Cotoneaster* (Bogunić et al., 2021), *Edraianthus* (Lakušić et al., 2009), *Knautia* (Frajman et al., 2015), *Luzula* (Bačič et al., 2016), *Sesleria* (Kuzmanović et al., 2013; Lazarević et al., 2015), *Sorbus* (Hajrudinović et al., 2015), *Veronica* (Padilla-García et al., 2018).

Polyploidization and hybridization are considered important forces in plant evolution (Wendel, 2000; Mallet, 2007; Soltis et al., 2009; Jiao et al., 2011; Van de Peer et al., 2021). Polyploids may differ from their diploid relatives in their morphology, phenology, physiology, biotic interactions, distribution, preferences for ecological niches, and invasiveness (e.g., Levin, 2002; Raabová et al., 2008; Maherali et al., 2009; te Beest et al., 2012; Weiss-Schneeweiss et al., 2013; Stutz et al., 2016; Decanter et al., 2020; Ulum et al., 2021). Many studies suggest better fitness of polyploids (Comai, 2005; Maherali et al., 2009; te Beest et al., 2012; Rey et al., 2017; Van de Peer et al., 2021), but opposite results can also be found in the literature (Nunvářová Kabátová et al., 2019). Polyploid individuals may be favored under stressful conditions, including major climatic or geological changes, and may provide an advantage in surviving and adapting to such conditions (Otto and Whitton, 2000; Comai, 2005; Weiss-Schneeweiss et al., 2013; Van de Peer et al., 2017, 2021; Decanter et al., 2020). Climate change may also lead to shifts in species distributions that bring previously separate species into contact and allow hybridization, including related species that have evolved to different ploidy levels (Vallejo-Marín and Hiscock, 2016).

There is increasing interest in the genus *Ramonda* as a model to study the role of interploidy gene flow in enhancing genetic and phenotypic diversity and the ecological factors that may contribute to the maintenance of cytotype diversity. The genus *Ramonda* Rich. (Gesneriaceae) includes three species—*R. myconi* (L.) Rchb., *R. nathaliae* Pančić & Petrović and *R. serbica* Pančić. All three species are considered relicts of the Tertiary (Košanin, 1921; Turill, 1929; Petrova et al., 2015), remnants of the tropical-subtropical flora from the time when the climate in southern Europe was warmer and more humid. Apart from being able to survive almost complete desiccation of their aboveground tissues in a state of anabiosis when habitat conditions are dry, and rehydrate when humidity becomes favorable again (Živković et al., 2005; Rakić et al., 2014), the three species also share similar habitats. They occur exclusively on north-facing rocky slopes of gorges, canyons, and mountain ravines, where conditions are wetter and more sheltered, allowing them longer active periods (Košanin, 1921; Rakić et al., 2014). However, *R. myconi* and *R. serbica* are found only on limestone, while *R. nathaliae* also inhabits ultramafic soils (Košanin, 1921; Stevanović and Stevanović, 1985; Lazarević et al., 2013; Rakić et al., 2013). *Ramonda myconi* and *R. nathaliae* are diploids with $2n = 2x = 48$ chromosomes and mean 2C values of 2.59 and 2.32 pg, respectively, while *R. serbica* is hexaploid with $2n = 6x = 144$ and mean 2C = 7.91 pg (Siljak-Yakovlev et al., 2008). Relatively high basic chromosome number ($x = 24$) indicates that the three species may be paleopolyploids (Rakić et al., 2014). Also, based on the similarity of monoploid genome size in *R. myconi* and *R. serbica*, the same authors hypothesized past existence of a common diploid ancestor with $2n = 24$ chromosomes from which later originated *R. nathaliae* on one side, and *R. myconi* and *R. serbica* on the other.

Despite ecological similarities, the ranges of three *Ramonda* species differ, with *R. myconi* inhabiting the NE part of the Iberian Peninsula, while *R. nathaliae* and *R. serbica* are endemics of the Balkan Peninsula. *Ramonda nathaliae* is distributed mainly in the Aegean watershed area (North Macedonia, N Greece and SW Serbia—one locality in SC Kosovo at the foothill of Mt. Šara) with small enclaves in E Serbia (the Black Sea watershed). In contrast, *R. serbica* is mainly present in the areas of the Adriatic watershed (Albania, S and SE Montenegro, W North Macedonia, SW Serbia—including W-SW Kosovo, NW Greece) and to a smaller extent in the Black Sea (SW, E and NE Serbia, NW Bulgaria) and the Aegean watersheds (one population in NW part of North Macedonia; **Figure 1A**, detailed map in Rakić et al., 2014). The range of the two species in the Balkans overlaps only in E Serbia in a narrow area of sympatry (**Figure 1A**) discovered only a few decades ago (Stevanović et al., 1986). In only two places in the nearby gorges of the Jelašnica and Nišava rivers (6–7 km apart), *R. nathaliae* and *R. serbica* can be found in the same macrohabitat, in the close proximity to each other opening possibility for hybridization. At the time of discovery of these two sites of sympatry, the authors did not notice any intermediate phenotypes that might indicate hybridization between the two species. However, thanks to a preliminary cytogenetic study at the two sites of sympatry, several individuals with intermediate chromosome number ($2n = 4x = 96$) and genome size (2C = 5.14 pg) were found at both locations (Siljak-Yakovlev et al., 2008). The presence of tetraploid plants has suggested the possibility of interploidy and thus interspecific hybridization. While two *Ramonda* species from the Balkan Peninsula can be clearly distinguished on the basis of leaf and flower characteristics (Pančić, 1874, 1884; Petrović, 1882), field and laboratory observations of tetraploid individuals have shown that it is very difficult to distinguish them from the parent species on the basis of morphological traits.

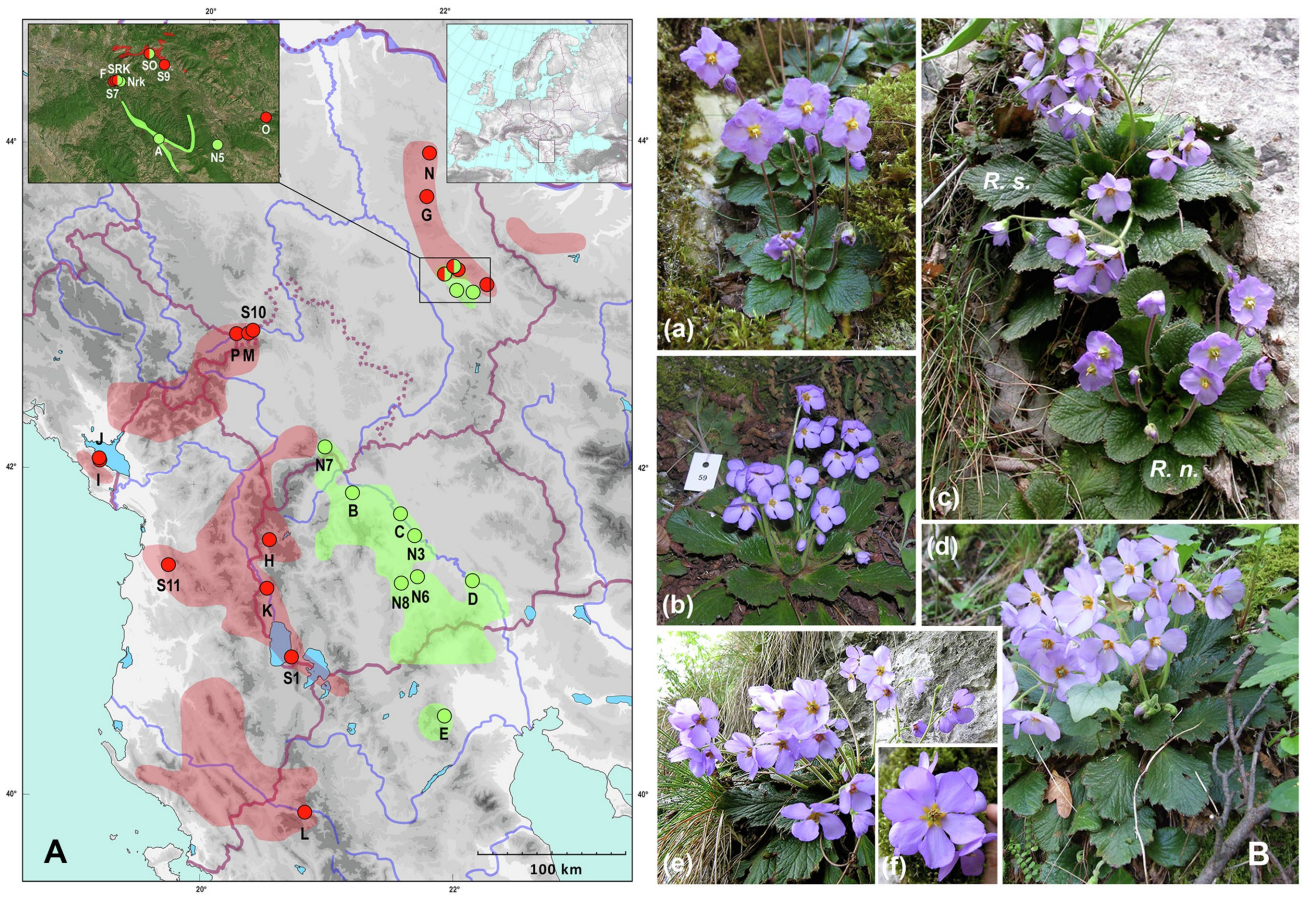


FIGURE 1 | (A) Geographic range of *R. nathaliae* and *R. serbica* and distribution map of investigated populations. Green color and green dots represent range and investigated populations of *R. nathaliae*. Red color and red dots represent range and investigated populations of *R. serbica*. Detailed distribution of two species in the zone of sympatry is presented in the upper left inset. Geographical position of the investigated area in Europe is presented in the upper right inset. SO—sympatry at Oblik, SRK—sympatry at Radovanski Kamen. See **Table 1** for detailed population legend. **(B)** *R. nathaliae* from population Divljana (a); *R. serbica* from population Radovanski Kamen (b); two species side by side in the sympatry Radovanski Kamen (c); tetraploid hybrid from Radovanski Kamen (d) and hybrid of higher ploidy level [probably backcross; (e)] showing some flowers with six petals (f).

Preliminary studies of flower morphology suggest greater similarity of tetraploid hybrids to *R. serbica* (**Figure 1B**) and very careful use of certain flower characters (the color of the anthers and the angle formed by the lines connecting the petal base and the points of maximum petal width) when attempting to distinguish tetraploid individuals from the canonical representatives of diploid *R. nathaliae* and hexaploid *R. serbica* (Lazarević et al., 2014).

Thus, the genetic status of the tetraploid individuals is unclear. Genetic analyses are needed to advance the study of possible hybridization between these two species. However, cytotype diversity in these two contact zones has not yet been thoroughly evaluated because initial studies were based on only a few individuals (Siljak-Yakovlev et al., 2008). Furthermore, if tetraploids are true hybrids, the question arises whether backcrosses between tetraploids and the two parental cytotypes are possible. Finally, the potential ability of these two species to hybridize raises the more general question of how genetically distinct they are from each other and what ecological and historical conditions have led to their differentiation. However, their genetic diversity has never been studied throughout their range up to date.

The aims of the present study were: (1) to investigate the variation in genome size within the two locations of sympatry between *R. nathaliae* and *R. serbica*, but also in several monospecific populations, to infer cytotype diversity; (2) to generate genetic profiles (using AFLP markers) of individuals within these two zones to test whether there is gene flow between the two species and to gain further insight into the genetic mechanisms that might explain the observed cytotype diversity; (3) to assess the genetic diversity and population structure of both species and their genetic differentiation using AFLP markers on a sample of monospecific populations covering most of their range in the Balkans. Knowledge of these genetic patterns is essential to provide a reference situation for the analysis of gene flow between the two species and especially in the two local areas of sympatry.

MATERIALS AND METHODS

Plant Material

Analyses included two species of the family Gesneriaceae from the Balkan Peninsula, *Ramonda nathaliae* and *R. serbica* from monospecific populations, and a significant sample of individuals from the only two sites of sympatry known to date, both located in E Serbia (**Table 1**; **Figure 1A**). For chromosome counting, live plants were collected in the field and grown at the Institute of Botany and Botanical Garden "Jevremovac" in Belgrade and at the Laboratoire d'Ecologie, Systématique et Evolution à l'Université Paris-Saclay. For genome size estimation (GS), fresh leaves were collected in the field, brought to the laboratory in zip bags and stored in the refrigerator at 4°C until analysis. Leaves used for DNA extraction were collected directly in the field and immediately placed in silica gel-filled bags. Voucher specimens are deposited in the herbarium of the Institute of Botany and Botanical Garden "Jevremovac," Faculty of Biology, University of Belgrade (BEOU).

Chromosome Preparation and Estimation of Nuclear DNA Content by Flow Cytometry

Mitotic chromosome plates were prepared for two populations (one of *R. nathaliae* and one of *R. serbica*) as an addition to previously published chromosome numbers. This was done using pretreatment with 0.002 M 8-hydroxyquinoline for 4.5 h at 16°C, followed by cold fixation in 3/1 (v/v) ethanol/acetic acid for 24–48 h, hydrolysis for 14 min at 60°C, staining with Schiff reagent (Feulgen and Rossenbeck, 1924) and applying classical squash technique. Two to five individuals per population were used for chromosome counting.

Genome size was estimated for the first time in ten monospecific populations (four of *R. nathaliae* and six of *R. serbica*) as an addition to previously published results, as well as in two localities where these species grow in sympatry. Total nuclear DNA amount was determined by flow cytometry according to Marie and Brown (1993) on fresh leaves from at least five individuals per monospecific population (or three individuals in only two populations) and from 213 individuals from two hybrid zones. *Pisum sativum* cv. Long Express (2C = 8.37 pg) and *Lycopersicon esculentum* cv. Roma (2C = 1.99 pg) were used as internal standards. Nuclei were stained using propidium iodide, and measurements were done by Elite ESP flow cytometer (Beckman-Coulter, Roissy, France) equipped with argon laser. Both methods are described in detail by Siljak-Yakovlev et al. (2008).

DNA Extraction and AFLP Analyses

AFLP analyses were performed by two independent molecular biology assays. A first experiment was designed to evaluate the genetic structure of populations of both species based on an initial sample of monospecific populations (16 populations in total—five of *R. nathaliae* and 11 of *R. serbica*, **Table 1**). This first analysis is referred to as the large-scale geographic study. The second AFLP experiment was above all designed to assess the hybridization potential of the two species when they occur in sympatry. Ten populations were studied: four monospecific populations of each species were included in this second analysis, as well as individuals collected at two sites of sympatry, Radovanski Kamen and Oblik (**Table 1**). Each individual from sites of sympatry was first identified as *R. nathaliae*, *R. serbica*, or "potential hybrid" based on its morphological characteristics (leaves and flowers). The material collected from individuals in sympatry consisted of leaves from 26 individuals potentially corresponding to *R. serbica*, 29 to *R. nathaliae*, and 40 individuals with ambiguous phenotypes at Oblik, and 11 individuals potentially corresponding to *R. serbica*, nine to *R. nathaliae*, and 35 individuals with ambiguous phenotypes at Radovanski Kamen. At each location of sympatry, sampling was conducted along a transect across the entire population. We systematically avoided sampling adjacent plants to eliminate the possibility of sampling clonal individuals since vegetative reproduction also occurs (Lazarević et al., 2013).

Fragments of leaves dried in silica gel weighting 21–26 mg were used for DNA extraction, which was performed using the NucleoSpin®96 Plant Kit (Macherey–Nagel) according to the manufacturer's instructions. AFLP analyses were performed according to the method described by Vos et al. (1995) with some modifications (see **Supplementary Material 1** for details).

Three primer combinations were selected for the full analysis (fluorescent dyes in parentheses): in both studies, EcoRI-AAC/MseI-CTC (FAM) and EcoRI-CAG /MseI-CTC (VIC) were used, whereas the third primer pair was different: EcoRI-ACA/MseI-CTC (NED) was used in the first AFLP experiment (monospecific populations only), and EcoRI-CAG/MseI-CTT (VIC) in the second experiment (hybridization study).

Amplified fragments were detected using the ABI Prism 3,100 sequencer on the Gentyane platform (INRAE, Clermont-Ferrand, France) and analyzed by GENESCAN 1.1 software (Applied Biosystems). DNA fragment detection results were indicated as present (1) or absent (0) by GeneMapper software version 4.0. Only unambiguous peaks in the size range of 50 to 500 bp were scored, with the minimum peak height set at 30. Each peak from each profile was checked by eye and its status changed accordingly. Finally, singletons (fragments that occur only in a single individual in the entire dataset) were removed to increase the reliability of the AFLP data.

To ensure the best possible reliability of AFLP markers, all samples from each AFLP experiment were managed in a single manipulation (same digestion, ligation, and PCR amplification mixtures and conditions for all samples). Two types of repeatability tests were performed. Four individuals, two from *R. nathaliae* and two from *R. serbica*, were used for the preliminary repeatability test. They were duplicated from the same amplification and then, genotyped with the same mixture of formamide and molecular weight marker. In this way, the reproducibility of the detection of the AFLP profiles was estimated (97.15%). In the second test, twenty individuals were amplified twice with two independent mixtures and the percentage of amplification reproducibility was 99.07%.

TABLE 1 | List of *Ramonda* populations included in the study. Dashed line represents the boundary between populations: used only for genome size estimation, or for two different molecular studies. In the last three columns, "1" indicates which analyses the individual populations were used for. Number of individuals included in molecular analyses is given in parentheses in the column AFLP.

Pop. code	Species	Locality	Altitude (m)	Latitude longitude	Geological substrate	Voucher number	Collectors	2*n*	2C	AFLP
N7	*R. nathaliae*	Serbia (Kosovo): Mt. Šara—Gotovuša	1,100	N 42°14′6.46″ E 21°4′30.14″	Limestone	21843 BEOU	Lazarević, M., Lazarević, P.		1	
N8	*R. nathaliae*	North Macedonia: Kozjak	1,610	N 41°23′56.72″ E 21°41′13.17″	Limestone	29886 BEOU	Niketić, M., Tomović, G.		1	
S11	*R. serbica*	Albania: Kruje	574	N 41°30′30.06″ E 19°47′45.13″	Limestone	32469 BEOU	Lakušić, D., Kuzmanović, N., Lazarević, P., Alegro, A.		1	
A	*R. nathaliae*	Serbia: Mt. Suva planina	1750	N 43°11′6.8″ E 22°10′0.7″	Limestone	20019 BEOU	Lakušić D.		1	1 (26)
B	*R. nathaliae*	North Macedonia: gorge of r. Treska, near Matka village	311.5	N 41°57′9.6″ E 21°17′51.5″	Limestone	31669 BEOU	Lazarević M., Sanino A., Siljak-Yakovlev S., Matevski V., Stevanović B., Stevanović V.		1	1 (26)
C	*R. nathaliae*	North Macedonia: gorge of r. Pčinja, near the road to Veles	266.1	N 41°49′18.9″ E 21°41′4.4″	Serpentine	31670 BEOU	Idem	1	1	1 (26)
D	*R. nathaliae*	North Macedonia: gorge of r. Vardar, near Demir Kapija	160	N 41°24′32.8″ E 22°15′42.1″	Limestone	31671 BEOU	Tomović G., Stevanović B., Stevanović V.			1 (26)
E	*R. nathaliae*	Greece: Mt. Vermion, gorge in the vicinity of village Sella, west of Noussa	*ca.* 1,000	N 40°35′5.4″ E 22°1′28.5″	Limestone	31668 BEOU	Niketić M.	1	1	1 (26)
F	*R. serbica*	Serbia: Radovanski Kamen, above the village Čukljenik	564	N 43°17′2.5″ E 22°4′16.4″	Limestone	31672 BEOU	Lazarević M., Lazarević P., Tomović G., Niketić M., Jović D., Petrović B.	1	1	1 (30)
G	*R. serbica*	Serbia: Mt. Rtanj	985	N 43°45′36.55″ E 21°55′40.01″	Limestone	31673 BEOU	Lazarević M., Lazarević P., Nešić D.		1	1 (8)
H	*R. serbica*	North Macedonia: gorge of r. Radika, Barička klisura	807	N 41°39′57.1″ E 20°37′3.6″	Limestone	Personal herbarium	Stevanović V., Tomović G.		1	1 (26)
I	*R. serbica*	Montenegro: gorge of r. Kroni e Murici	420	N 42°8′32.5″ E 19°13′9.9″	Limestone	19673 BEOU	Stevanović B., Stevanović V.	1	1	1 (26)
J	*R. serbica*	Montenegro: gorge of r. Kroni e Besit	250	N 42°9′19.7″ E 19°12′48.4″	Limestone	9880 BEOU	Stevanović B., Stevanović V.	1	1	1 (26)
K	*R. serbica*	North Macedonia: gorge of r. Crni Drim, near the village Lukovo	620	N 41°22′11.1″ E 20°35′57.6″	Limestone	2081 BEOU	Stevanović V., Tomović G.		1	1 (26)
L	*R. serbica*	Greece: Mt. Timfi, gorge of r. Vikos, near the village Vikos	*ca.* 650	N 40° 00′00″ E 20° 54′39.6″	Limestone	31674 BEOU	Niketić M., Tomović G.	1	1	1 (26)
M	*R. serbica*	Serbia: gorge of r. Mojstirska suhovara	920	N 42°55′39.6″ E 20°25′58.4″	Limestone	31675 BEOU	Lazarević M., Lazarević P., Pavićević D.		1	1 (26)
N	*R. serbica*	Serbia: gorge of r. Lazareva reka	310	N 44°1′38.1″ E 21°57′16.2"	Limestone	370/90	Lazarević M., Tomović G., Lakušić D.		1	1 (26)
O	*R. serbica*	Serbia: Čiflik	370	N 43°13′3″ E 22°25′6.2″	Limestone	31676 BEOU	Lazarević M., Lazarević P., Tomović G., Niketić M., Jović D., Petrović B.	1	1	1 (26)
P	*R. serbica*	Serbia: gorge of r. Godulja	750	N 42°55′27.4″ E 20°20′0.4″	Limestone	31677 BEOU	Lazarević M., Lazarević P., Pavićević D.		1	1 (5)
N3	*R. nathaliae*	North Macedonia: gorge of r. Vardar near Veles	165.8	N 41°41′15.2″ E 21°47′56.2″		33186 BEOU	Lazarević, M., Siljak-Yakovlev, S., Sanino, A., Matevski, V., Stevanović, B., Stevanović, V.		1	1 (25)
N5	*R. nathaliae*	Serbia: slopes of Mt. Suva planina near village Divljana in the vicinity of Bela Palanka	346.6	N 43°10′19.3″ E 22°18′10"	Limestone	20638 BEOU	Lazarević, M., Zlatković, B., Stevanović, V.	1	1	1 (17)

(Continued)

TABLE 1 | Continued

Pop. code	Species	Locality	Altitude (m)	Latitude longitude	Geological substrate	Voucher number	Collectors	*2n*	2C	AFLP
N6	*R. nathaliae*	North Macedonia: gorge of r. Raec near Kavadarci	295.7	N 41°26′14.1″ E 21°49′6.27″	Limestone	2063 BEOU	Lazarević, M., Šiljak-Yakovlev, S., Sanino, A., Matevski, V., Stevanović, B., Stevanović, V.		1	1 (25)
Nrk	*R. nathaliae*	Serbia: pure subpopulation from Radovanski Kamen	576.7	N 43°17′1.2″ E 22°4′23.5″	Limestone	20641 BEOU	Stevanović, V., Lazarević, M.	1	1	1 (16)
S1	*R. serbica*	North Macedonia: Galičica, Zli do	950	N 40°57′02.63″ E 20°48′02.73″	Limestone	Personal herbarium	Stevanović, B., Tomović, G., Stevanović, V.,			1 (24)
S7	*R. serbica*	Serbia: gorge of river Jelašnica in the vicinity of Niš	295.4	N 43°16′56″ E 22°3′42″	Limestone	20642 BEOU	Lazarević, M., Stevanović, V.	1	1	1 (24)
S9	*R. serbica*	Serbia: gorge of r. Nišava, 20 km east from Niš	308.5	N 43°18′38.1″ E 22°10′53.3″	Limestone	20639 BEOU	Lazarević, M., Zlatković, B., Stevanović, V.	1	1	1 (23)
S10	*R. serbica*	Serbia: gorge of r. Crna Reka in the vicinity of monastery St. Peter c. 8 km southwards from village Ribarići	1,100	N 42°56′46.3″ E 20°28′09.7″	Limestone	20623 BEOU	Lazarević, M., Lazarević, P., Pavićević, D.	1	1	1 (18)
Radovanski Kamen	Sympatry	Serbia: Radovanski Kamen	574.5	N 43°17′5.1″ E 22°4′16.6″	Limestone	20640 BEOU	Lazarević, M., Zlatković, B., Stevanović, V.	1	1	1 (55)
Oblik	Sympatry	Serbia: gorge of r. Nišava in the vicinity of village Ostrovica—Oblik	418.2	N 43°19′47.4″ E 22°8′44.9″	Limestone	20636 BEOU	Idem	1	1	1 (95)

Because the samples from the two AFLP experiments were managed separately, the two datasets were not combined for the population genetic analyses to ensure better reliability of the analyses.

Statistical Analyses and Population Genetics

Statistical analyses of genome size were performed in R4.1.2. For monospecific populations, each species was analyzed separately because genome size values differ greatly between the two *Ramonda* species. For small sample sizes, nonparametric tests were used (Fligner-Killeen test for homogeneity of rank variation and Kruskal–Wallis and Nemenyi tests for median comparisons). For pairwise comparisons, p values were corrected according to the Holm correction for multiple testing. Student t-test and Welch test (when homoscedasticity was rejected) were used for mean comparisons for larger sample sizes ($n > 15$).

AFLP-SURV 1.0 (Vekemans, 2002) was used to infer genetic diversity within populations. Allele frequencies were estimated from the presence or absence of fragment phenotypes using the Bayesian method described in Zhivotovsky (1999) with a non-uniform prior distribution of allele frequencies.

Genetic dissimilarity between individuals and populations was estimated using Shared Allele Distance (Chakraborty and Jin, 1993) with POPULATIONS 1.2.30 (Langella, 1999). Principal coordinate analysis (PCoA) was performed on the individual distance matrix. Neighbor-joining dendrograms were built using the population and individual distance matrices.

The distribution of genetic variation (between species, between populations, and within populations) was estimated by analysis of molecular variance (AMOVA) performed in Arlequin 3.5.1.2 (Excoffier and Lischer, 2010; http://cmpg.unibe.ch/software/arlequin35/Arlequin35.html). To test for correlations between geographic distance and genetic variance, a Mantel test was performed using the same software. The significance of Φ-statistics was tested based on the distributions of Φ-statistics calculated under the null hypothesis from 1,000 virtual samples obtained by random permutations of the original data set.

The genetic structure of the monospecific and the two populations of sympatry was studied using genetic model-based and non-genetic model-based methods. For the first type of analysis, Bayesian inference of genetic clusters and assignment of individuals was performed using STRUCTURE v.2.3.4 (Pritchard et al., 2000). STRUCTURE analyses were performed assuming admixture, correlated allele frequencies, and no prior population information. In a first step, 10 replicates were performed for each value of the genetic cluster (K), where K varied from 1 to 12, with a burn-in period of 50,000 iterations followed by a run length of 120,000 iterations of the Markov chain. The most probable number of groups was determined using the method of Evanno et al. (2005) implemented in Structure Harvester v.0.6.94 (Earl and vonHoldt, 2012). Then, 50 new replicates of the MCMC method were performed to determine the best K value. CLUMPP v.1.1.2 (Jakobsson and Rosenberg, 2007) was used to check for similarities between runs and average the membership

TABLE 2 | Chromosome numbers and nuclear DNA contents of investigated monospecific populations of *R. nathaliae* and *R. serbica* (*N* = number of samples for 2C DNA value).

Pop. Code	Species	Locality	*N*	2*n*	Ploidy level (*x*)	2C in pg and (Mbp)[a]	CV (%)	1Cx in pg and (Mbp)
A	*R. nathaliae*	RS: Mt. Suva planina	5		2	2.28 (2229)*	2.34	1.14 (1115)
B	*R. nathaliae*	MK: Matka	5		2	2.26 (2208)	1.50	1.13 (1104)
C	*R. nathaliae*	MK: Pčinja	5	48**	2	2.27 (2217)**	1.54	1.13 (1108)
E	*R. nathaliae*	GR: Mt. Vermion	5	48	2	2.28 (2231)*	1.93	1.14 (1116)
N3	*R. nathaliae*	MK: Vardar – Veles	5		2	2.35 (2295)**	0.50	1.17 (1147)
N5	*R. nathaliae*	RS: Divljana	5	48*	2	2.33 (2277)*	1.95	1.16 (1138)
N6	*R. nathaliae*	MK: Raec – Kavadarci	5		2	2.37 (2318)	0.88	1.18 (1159)
N7	*R. nathaliae*	RS (XK): Gotovuša	5		2	2.32 (2266)	2.17	1.16 (1133)
N8	*R. nathaliae*	MK: Kozjak	5		2	2.50 (2445)	0.92	1.25 (1222)
	R. nathaliae	Mean value	45		2	2.33 (2278)	3.36	1.16 (1139)
G	*R. serbica*	RS: Mt. Rtanj	5		6	7.76 (7587)	1.04	1.29 (1264)
H	*R. serbica*	MK: Radika	5		6	7.88 (7706)	0.77	1.31 (1284)
I	*R. serbica*	ME: Kroni e Murici	5	~150*	6	7.94 (7763)*	2.38	1.32 (1294)
J	*R. serbica*	ME: Kroni e Besit	5	144*	6	7.77 (7601)*	0.50	1.30 (1267)
K	*R. serbica*	MK: Crni Drim	5		6	7.95 (7772)	0.32	1.32 (1295)
L	*R. serbica*	GR: Mt. Timfi	5	144**	6	7.58 (7416)**	2.29	1.26 (1236)
M	*R. serbica*	RS: Mojstir	5		6	7.91 (7737)	2.49	1.32 (1289)
N	*R. serbica*	RS: Lazareva reka	5		6	7.58 (7413)*	1.37	1.26 (1236)
O	*R. serbica*	RS: Čiflik	5	144	6	7.79 (7619)	2.92	1.30 (1270)
P	*R. serbica*	RS: Godulja	5		6	7.61 (7446)**	6.28	1.27 (1241)
S7	*R. serbica*	RS: Jelašnica	5	144**	6	7.85 (7678)**	0.86	1.31 (1280)
S9	*R. serbica*	RS: Nišava	5	>144*	6	8.11 (7927)*	1.86	1.35 (1321)
S10	*R. serbica*	RS: Crna Reka	3	144*	6	8.32 (8136)*	1.82	1.39 (1356)
S11	*R. serbica*	AL: Kruje	3	-	6	7.76 (7592)	1.04	1.29 (1265)
	R. serbica	Mean value	66		6	7.83 (7660)	3.15	1.31 (1277)

[a]*Values in Mbp calculated based on pg = 978 Mbp (Doležel et al., 2003).*
**2C values and/or chromosome numbers from Siljak-Yakovlev et al. (2008);*
***2C values and/or chromosome numbers from Lazarević et al. (2013).*

probabilities. The final bar graphs showing the individual admixture coefficients were created using Structure Plot v.2.0 (Ramasamy et al., 2014).

In addition, and to evaluate the robustness of the classification used, a Discriminant Analysis of Principal Components (DAPC, Jombart et al., 2010) implemented in the adegenet package for R software was used for assigning each studied individual from the first study to a specific group based on its genetic characteristics.

Finally, in order to integrate cytogenetic data and genetic inferences, we provided distributions of ancestry coefficients values (*Q*-values) estimated by the Bayesian clustering analysis in sympatry and monospecific populations, according to ploidy levels inferred from genome size.

RESULTS

Chromosome Counts and Genome Size in Monospecific Populations

As expected, chromosome number showed that all individuals sampled in monospecific populations of *R. nathaliae* were diploid with $2n = 2x = 48$ chromosomes. The chromosome number of all individuals sampled in monospecific populations of *R. serbica* also confirmed that this species is hexaploid ($2n = 6x = 144$).

The average 2C DNA value was 2.33 pg in *R. nathaliae* and 7.83 pg in *R. serbica* (data from monospecific populations only; **Table 2**). Although differences in genome sizes between populations of *R. nathaliae* were significant (Kruskal–Wallis Chi-squared = 28.019, $df = 8$, p value < 1*e*-3), pairwise Nemenyi comparison tests yielded only three significant comparisons, all involving the population from Mt. Kozjak (N8) with slightly larger GS compared to other populations. There were also significant differences in genome size values between populations of *R. serbica* (Chi-squared = 40.405, $df = 13$, p value < 1*e*-3). However, this was the result of differences between populations with the smallest (Mt. Timfi and Lazareva Reka) and populations with the largest (Nisava and Crna Reka) genome size values.

Large-Scale Geographic Study of Genetic Diversity in Their Entire Range

Genetic Diversity Revealed With AFLP Markers

In the first molecular study, a total of 1,077 AFLP fragments were obtained with the three primer pairs. Once singletons excluded 163 fragments were present only in *R. nathaliae,* and 235 fragments were private to *R. serbica*. However, none of these fragments was fixed for alternative alleles in the two species (fragments present in all individuals of one species and absent in all individuals of the other species). Low frequency fragments dominated in both species. The number of fragments detected was moderately higher in *R. serbica* than in *R. nathaliae* (938 in the former and 866 in the latter), a trend expected due to the larger genome size of *R. serbica*.

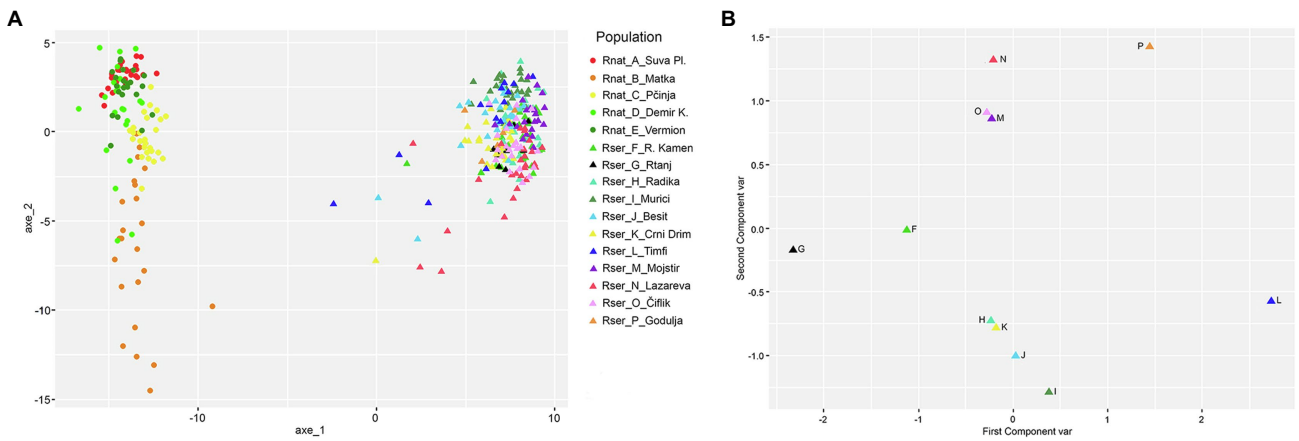


FIGURE 2 | Principal Coordinate analysis performed on the large geographical scale AFLP dataset using, the Shared Allele Distance computed on: **(A)** individuals of both *R. serbica* and *R. nathaliae* species (the first two axes represented 48.8 and 4.0% of the total variation, respectively); **(B)** on populations of *R. serbica* alone (the first two axes represented 52.2 and 30.0% of the total variation, respectively).

Within-population genetic diversity was moderately similar in *R. serbica* (mean $Hj = 0.105$) and *R. nathaliae* (mean $Hj = 0.092$; **Supplementary Table 1**). Very little variation from the expected heterozygosity was observed between populations of each species. The lowest genetic diversity was found in *R. nathaliae* from Mt. Suva Planina in E Serbia (A), and the highest values were found in *R. nathaliae* from Matka in North Macedonia (B) and *R. serbica* from Mt.Timfi in Greece (L).

Genetic Differentiation and Population Genetic Structure

AMOVA and PCoA analyses showed that most of the genetic diversity within the entire sample are explained by genetic differentiation between the two species, as would be expected in the case of truly independent evolutionary lineages (Φct = 0.718, see **Supplementary Table 2**; **Figure 2A**). The genetic differentiation of populations within species is also very high (Φsc = 0.318). However, AMOVA analyses for individual species showed that population genetic differentiation was much higher in *R. nathaliae* (Φst = 0.432) than in *R. serbica* (Φst = 0.244).

Total genetic diversity (Ht) was equally low in both species with values of Ht = 0.116. Among the populations of *R. nathaliae*, the lowest genetic distance was observed between the populations of Mt. Vermion (E) and Demir Kapija (D) and the highest between Mt. Suva Planina (A) and Matka (B) (**Supplementary Figure 1**). In *R. serbica*, a population from Mt. Timfi (L), at the southern edge of the species' range, and Godulja (P) in SW Serbia, were genetically the most distant from the others (**Supplementary Figure 1**). PCoA analysis of populations of *R. serbica* revealed a clear geographic pattern of genetic differentiation, contrasting northern populations from Serbia (G, F, P, N O, and M) on one side and southwestern populations from Greece, North Macedonia, and Montenegro (I, J, H, K, and L) on the other (**Figures 1A**, **2B**). The relationship between pairwise geographic distances between populations and genetic distances was not significant for *R. nathaliae* ($Z = r = 0.51$; $p = 0.131$), but was significant in *R. serbica* ($Z = r = 0.56$; $p < 1e\text{-}3$). However, this relationship is mainly due to the differentiation between northern and southern populations of this species. Overall, geographic distance between populations explained their genetic distances poorly, as can be clearly seen for northern populations G and F on the one hand, and M, N, O, and P on the other.

DAPC analysis also showed that the vast majority of individuals of both species could be unambiguously assigned to their population of origin (**Supplementary Figure 2**). There were only a few cases of assignment uncertainty or assignment to a different population in *R. serbica* between two populations in E Serbia (Radovanski Kamen (F) and Mt. Rtanj (G)) that were only 50 km apart, and also between two geographically very close populations (I and J), both from Mt. Rumija in Montenegro.

Interestingly, a few *R. serbica* individuals from several populations (F, J, K, L, and N) showed a slightly intermediate genetic profile between the two species (**Figure 2A**), although they were well assigned to their own species in the DAPC analysis.

Evidence for Hybridization in Two Sites of Sympatry

Genome Size Variability in the Two Sites of Sympatry

Individuals from both locations of sympatry showed tremendous but discontinuous variability in genome size (**Table 3**; **Figure 3**). A first group of individuals had a mean genome size of 2.32 pg (range between 2.22 and 2.51 pg). The two mean GS values of this group in Oblik and in Radovanski Kamen were not significantly different (Student-$t = 1.79$, df = 73, p value = 0.078). The GS of this group was also not significantly different from the mean genome size of *R. nathaliae* estimated from monospecific

TABLE 3 | Chromosome numbers, nuclear DNA contents and ploidy levels of individuals from sympatry between *R. nathaliae* and *R. serbica* (*N* = number of samples for 2C DNA value).

Locality	Taxon	Ploidy level (*x*)	2*n*	2C (pg)	CV (%)	Min–Max of 2C (pg)	*N*
RS: Oblik	*R. nathaliae*	2	–	2.34	2.00	2.27–2.45	30
RS: Radovanski Kamen	*R. nathaliae*	2	48*	2.31	2.81	2.22–2.51	45
RS: Oblik	*R. serbica*	6	–	7.95	2.01	7.72–8.29	23
RS: Radovanski Kamen	*R. serbica*	6	144*	8.00	4.38	7.48–8.45	18
RS: Oblik	Class I	4	96*	5.14	1.41	5.03–5.27	25
	Class II	–	–	6.12	–	–	1
	Class III	–	–	10.32	12.83	8.94–12.06	4
RS: Radovanski Kamen	Class I	4	96*	5.18	1.98	4.96–5.39	45
	Class II	–	–	6.05	5.25	5.52–6.55	14
	Class III	–	–	9.98	5.14	9.54–10.88	6
	Class IV	–	–	3.70	8.97	3.35–4.01	3

Individuals whose genome size did not correspond to those of either R. nathaliae or R. serbica were classified into four different classes.
**Chromosome numbers from Siljak-Yakovlev et al. (2008).*

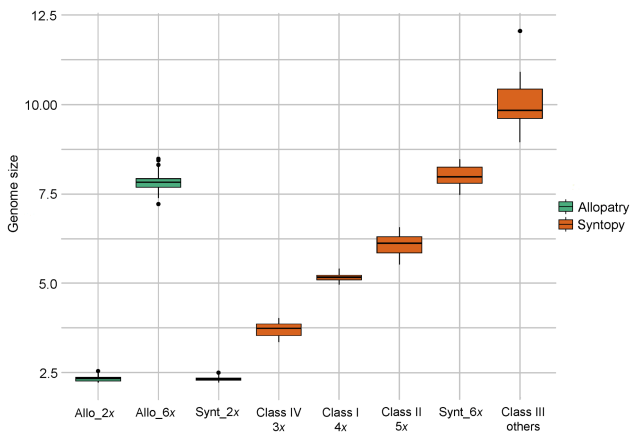


FIGURE 3 | Box-plot showing the distribution of genome size of each taxon and each genome class (according to **Table 3**).

populations (Student-*t* = 0.50, df = 118, *p* value = 0.615). Thus, the plants in this group certainly corresponded to diploid individuals.

A second group of individuals had a mean genome size of 7.98 pg (range between 7.48 and 8.45 pg). This group showed no significant difference between Oblik and Radovanski Kamen (Welch-t = −0.067, df = 22.53, p value = 0.96). Although the mean GS of this group at the two sites was significantly higher than the mean of the monospecific populations of *R. serbica* (Student-t = −2.851, df = 105, p value = 0.005), it remained within the range of variation of *R. serbica*. Pairwise comparisons showed that only the comparison with the Lazareva Reka population was at the threshold of significance (data not shown). Therefore, this group of individuals was likely hexaploid at both sites of sympatry.

However, other classes of genome size values suggest that there are several other ploidy levels at both sites of sympatry with genome composition contributed by both species. Indeed, a first class of individuals had a mean GS value of 5.17 pg (range between 4.96 and 5.39) (*cf.* Class I in **Table 3**). The individuals of this class in Oblik and Radovanski Kamen did not show any significant difference in their mean GS (Student-t = −1.65, df = 68, p value = 0.104). The mean GS of this entire class was very close, although significantly higher (Student-t = 7.79, df = 69, p value = 4.8e-11) than the expected value for a first-generation tetraploid interspecific hybrid (5.08 pg) calculated from the mean GS values of the two *Ramonda* species. This supported that these individuals are interspecific tetraploid hybrids. It is noticeable that a few individuals exhibited a combination of morphological characteristics of the two species (**Figure 1B**), suggesting that they may be first-generation interspecific hybrids. These plants were very abundant in our sample, representing 70 out of 214 individuals analyzed in the two sites of sympatry.

Finally, two other classes of GS values were also observed in the two sites of sympatry. Fifteen analyzed individuals had a higher genome size than the hypothetical tetraploid hybrids, ranging from 5.52 pg to 6.55 pg (*cf.* Class II in **Table 3**, **Figure 3**). In some of these individuals, the GS was close to what would be expected in a 5*x* interspecific hybrid. For example, a 5*x* individual with two *R. nathaliae* genome complements and three *R. serbica* genome complements would be expected to have a GS of about 6.27 pg. Thus, the plants in this Class II could correspond to individuals with mosaic genomes in which the proportion of the two species is different than in a first-generation tetraploid hybrid.

In both populations, a few individuals (10 in total) had a higher genome size than the hexaploid *R. serbica* (from 8.94 to 12.06 pg, *cf.* Class III in **Table 3**).

Finally, three individuals sampled at Radovanski Kamen were found to have an intermediate genome size between *R. nathaliae* and tetraploid interspecific hybrids (Class IV, **Table 3**). Their GS values were close to what would be expected in triploid individuals of *R. nathaliae*.

Taxonomic identification of individuals in monospecific populations of the two *Ramonda* species based on leaf and

flower morphological characters is always trivial and unambiguous (Siljak-Yakovlev et al., 2008; Lazarević et al., 2013, 2014). However, this was not the case for all individuals sampled at the two sites of sympatry (**Figure 1B**). For this reason, the GS values were also used to confirm or assign a taxonomic identification to be used in the analyses of the AFLP data (see below). In the following, all individuals in sympatry with the same genome size as *R. nathaliae* were classified as this species, and the same principle was applied to *R. serbica*. Other individuals are hereafter referred to as potential "hybrids" and assigned to their GS classes as indicated in **Table 3**.

Genetic Variability by AFLP Markers in the Two Sites of Sympatry

The following analyses included eight monospecific populations (four for each species) that were used as references in addition to the individuals sampled at the two sites of sympatry (**Table 1**). After removing singleton fragments, 1,263 AFLP fragments remained for analyses. None of the fragments were present in all individuals of one species and absent in all individuals of the other species, similarly to what was found in the large-scale geographic study. However, some fragments were highly abundant in one species and absent in the other, consistent with the data from the first set of populations (see above).

A strong contrast in AFLP fragment frequencies between *R. serbica* and *R. nathaliae* would have been expected because of the genome size and ploidy level of *R. serbica*. In fact, the distribution of the difference in fragment frequencies between the two species was clearly centered at zero (median = −8e-4; mean = 0.049) (**Supplementary Figure 3**). Thus, the dominant alleles at the AFLP markers (presence of the fragment) were almost equally distributed between the two species. Importantly, this result supports the assumption that in this study the dominance effect at AFLP markers was unlikely to cause bias in favor of either species in individual assignment in subsequent clustering analyses.

The strong genetic differentiation between the two species, already reported in the large-scale geographic study, was confirmed when only monospecific populations were considered (Φct = 0.548; **Figure 4A**). In contrast, it was noticeable that no genetic differentiation (Φct = −0.009) was observed between diploid and hexaploid individuals from the two sites in sympatry, as also evidenced by the strong mixing of individuals from both ploidy levels in the dendrogram (**Figure 4B**). The tetraploid plants also exhibited similar genetic variability as the diploid or hexaploid plants in sympatry as shown by the very strong mixing of the three ploidy levels in **Figure 4B**. However, clear and even strong genetic differentiation was observed between individuals from the two sites of sympatry for each ploidy level (Φst = 0.22 for diploids and Φst = 0.10 for hexaploids). Consistent with this result, clusters of individuals from each location of sympatry, regardless of their ploidy level, are clearly visible in **Figure 4B**.

Bayesian clustering yielded two genetic clusters as the solution that best fit the data, according to the Evanno test (**Supplementary Figure 4**) and the very clear plate obtained by the likelihood of the data at $K = 2$ (data not shown). This result was to be expected, since species differentiation is the factor that contributes most to the genetic structure of the population in the entire data set. Indeed, from the data of monospecific populations, it appears that the two genetic clusters correspond to one of each species (**Supplementary Figure 5**). DAPC analysis confirmed this result. Indeed, although $K = 3$ was the optimal solution (**Supplementary Figure 6**), one of those two clusters was composed only with one individual of the S7 monospecific population of *R. serbica* (**Supplementary Table 3**). As for the Bayesian clustering analysis, the two main clusters corresponded to *R. nathaliae* and *R. serbica*. We will therefore refer to "nathaliae" and "serbica" for these two clusters below. The distribution of individual ancestry coefficients showed two very different patterns between the monospecific populations and sites of sympatry (**Figure 5, Supplementary Figure 5**). Most individuals in monospecific populations had high ancestry values of the species where they belonged (mean ancestry of 0.93 for individuals of *R. serbica* to the genetic cluster "serbica" and 0.96 for individuals of *R. nathaliae* to the genetic cluster "nathaliae"; **Figure 5A**). Therefore, the AFLP data allowed unambiguous assignment of nearly all individuals sampled in monospecific populations to their own species, as already observed in the large-scale geographic study. Only three individuals of *R. serbica*, all from the gorge of Nišava river (S9), had an intermediate ancestry coefficient (**Figure 5A**), with the ancestry coefficient to the "serbica" genetic cluster ranging from 0.43 to 0.59. In comparison, the lowest value of ancestry to the "nathaliae" genetic cluster was up to 0.75 for individuals sampled in monospecific populations of *R. nathaliae*.

The two sites of sympatry showed a very different and more complex genetic pattern, similar to what was observed for genome size. The main difference with the monospecific populations was that there was much less agreement between the genome size values (and thus the inferred ploidy level) and the AFLP profiles. In fact, when 2*x* and 6*x* individuals were considered, the distribution of ancestry coefficients was bimodal for both cytotypes (**Figure 5B**). Surprisingly, 17 of the 35 diploid individuals had high ancestry coefficients to the hexaploid *R. serbica* (mean = 0.88; range: 0.68–0.99 for these 17 individuals). All but one of them were sampled in the Oblik site of sympatry. The remaining 18 diploid individuals had very high ancestry coefficients to *R. nathaliae*, as expected based on their assumed ploidy level (mean = 0.96; range: 0.83–1). In contrast, among hexaploid individuals, five of the total 30 individuals, all from Oblik, showed strong ancestry to *R. nathaliae*, contrary to what was expected based on their genome size.

The bimodal character of the distribution of the ancestry coefficient distribution was also observed for tetraploid individuals (**Figure 5B**), although the right side of the distribution was much flatter than that observed for monospecific *R. serbica* populations. This also testifies to the high genetic heterogeneity of tetraploid plants. The mean ancestry value of the 40 individuals belonging to the right side of the distribution was 0.80, very close to the value of 0.75 expected in the case of a first-generation hybrid resulting

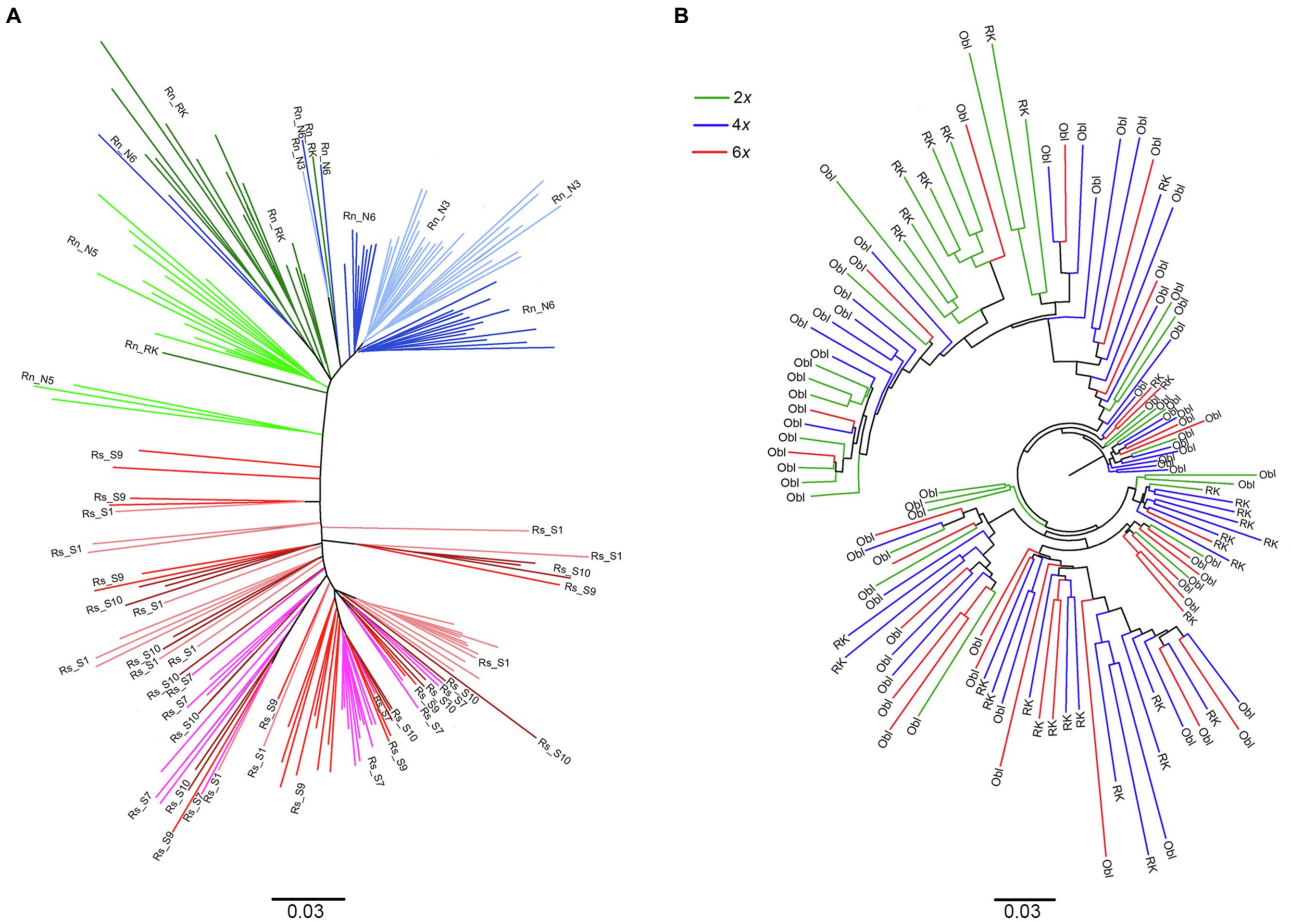


FIGURE 4 | Dendrogram built on inter-individual Shared Allele Distance using: **(A)** individuals of *R. serbica* and *R. nathaliae* sampled in allopatric monospecific populations (each color indicates a specific population whose code names are referred to in **Table 1**); **(B)** individuals from the two sites of sympatry [Oblik (Obl) and Radovanski Kamen (RK)]. All those individuals contributed to the interspecific gene flow study.

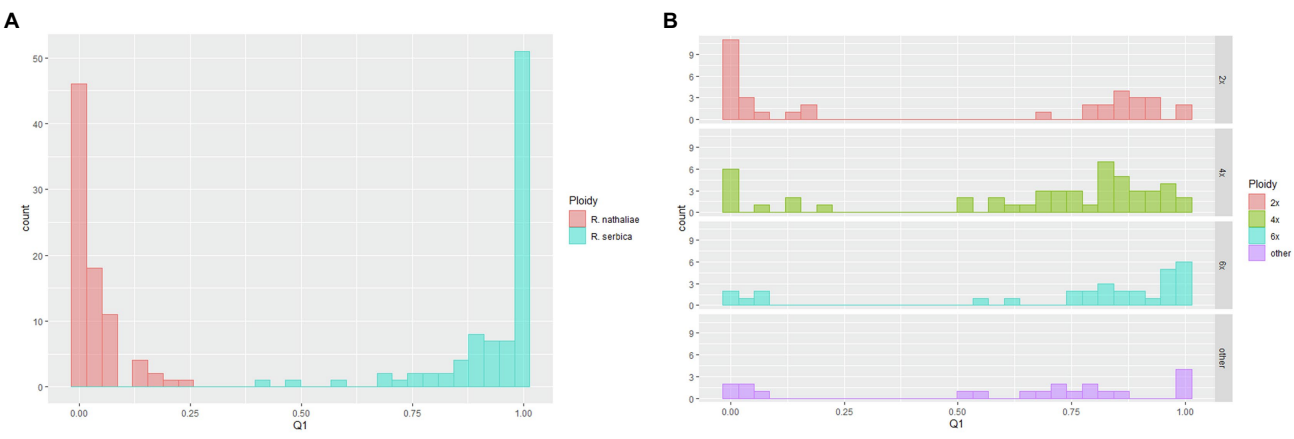


FIGURE 5 | Distribution of individual ancestry coefficients (inferences from the Bayesian clustering analysis) to the "serbica" genetic cluster for: **(A)** *R. nathaliae* (red) and *R. serbica* (blue) allopatric populations; **(B)** for individuals sampled in the two sites of sympatry, represented according to their genome size. "Other" refers to individuals that could not be undoubtedly assigned to a given ploidy level from their genome size.

from a cross between hexaploid *R. serbica* and diploid *R. nathaliae*. However, ten individuals, all from Oblik (**Supplementary Figure 5**), formed the left side of the distribution (i.e., individuals unambiguously assigned to *R. nathaliae* based on AFLP markers). Their mean ancestry value to "serbica" genetic cluster was 0.05 (range: 0.01–0.2), very close to the value obtained for *R. nathaliae* individuals from monospecific populations.

Apart from these three groups of individuals, eight individuals from Radovanski Kamen had rather mixed genetic profiles with a GS lower than *R. serbica* but higher (between 5.61 and 6.55 pg) than potential 4*x* individuals (*cf.* Class II in **Table 3**), but with a mean ancestry coefficient of 0.78 (range between 0.56 and 0.99) much closer to *R. serbica* than to *R. nathaliae*. Of the nine individuals in the Class III of genome size (ranging from 8.94 to 10.5 pg, **Table 3**), three were strongly assigned to the *R. serbica* genetic cluster, while two had a very high ancestry coefficient to *R. nathaliae*. This last observation suggests the existence of autopolyploid individuals with high ploidy level descended from *R. nathaliae*. Finally, the three individuals with a genome size between 3.35 and 4.01 pg (i.e., a GS between 2*x* and 4*x*) were unambiguously assigned to *R. nathaliae*.

It is noticeable that individual assignment posterior probabilities inferred by the DAPC analysis and ancestry coefficients inferred through the Bayesian approach were highly correlated (0.95 for all individuals in sympatry and 0.96 for the whole sample).

Both GS and AFLP patterns showed that more complex genomic structures than first-generation interspecific tetraploid hybrids are present at both sites of sympatry.

DISCUSSION

Chromosome Number, Genome Size, and Ploidy Level in Allopatric Populations

In the past, there were uncertainties concerning the chromosome number and ploidy levels of Balkan *Ramonda* species. Although Glišić (1924) found $n = 36$ to be the most frequent haploid chromosome number in *R. nathaliae*, later both Ratter (1963) and Contandriopoulos (1966) showed that *R. nathaliae* is a diploid with $2n = 2x = 48$ chromosomes, and this is confirmed by broad and detailed cytogenetic studies by Siljak-Yakovlev et al. (2008) and Lazarević et al. (2013). Still, some uncertainties remained about chromosome number and ploidy level in *R. serbica* based on the following reports: $2n = 72$ in *R. serbica* individuals collected in Sićevo Gorge in E Serbia (Glišić, 1924), $2n = 96$ in plants from Mt. Timfi—Vikos Gorge in Greece (Contandriopoulos, 1966), polyploid forms as a mixture of several chromosome numbers, from less than 72 to more than 144 chromosomes, with $2n = 72$ and $2n = 96$ being most common in only two populations in Albania (Daskalova et al., 2012). In this last study, five individuals per population were analyzed and hexaploid individuals were not detected. However, one of the populations analyzed for genome size in the present study is Kruje, the same population as in Daskalova et al. (2012), and only 6*x* individuals were found. In the last 15 years, Siljak-Yakovlev et al. (2008) and Lazarević et al. (2013), as well as the present study, have analyzed the genome size and/or chromosome number of a total of 133 individuals sampled from 16 monospecific populations of *R. serbica* throughout the species range. The vast majority of them were hexaploid individuals with $2n = 6x = 144$ chromosomes. Only in one population from Montenegro were two individuals found with a larger genome size, suggesting the possibility of the existence of 8*x* and even 10*x* individuals (Siljak-Yakovlev et al., 2008). All this indicates possible but rare alternative chromosome numbers in monospecific populations.

Genetic Diversity in *Ramonda* Species From the Balkan Peninsula

To date, very few studies have examined the genetic diversity of the two Balkan *Ramonda* species. In a study on the genetic variability of three populations of *R. serbica* from Albania and three populations of *R. nathaliae* from North Macedonia using RAPD markers, Daskalova et al. (2012) could not clearly distinguish the two species. More recently, Petrova et al. (2018) studied five populations of *R. serbica* in Bulgaria using ISSR markers. They showed that this species has low genetic variability and significant differentiation among populations. The present study is the first investigation of the genetic diversity of *R. nathaliae* and *R. serbica* as a whole, using the same molecular markers and including a large coverage of their respective ranges in the Balkan Peninsula. The number of AFLP markers is also high, and thus, we are confident that genetic diversity and structure patterns here reported are representative. The high level of genetic differentiation is in agreement with morphological data that have shown that the two species are broadly consistent across their range, with the exception of the two sites of sympatry (Lazarević et al., 2013, 2014). Therefore, it is clear that the two species actually correspond to two independent evolutionary lineages, suggesting a unique origin of hexaploid monospecific populations in their range. Moreover, both species showed a similar picture: a very high level of genetic differentiation between populations within species and very low signals of genetic admixture between populations of each species, except between a few pairs of geographically close populations.

What information about the evolutionary forces that shaped genetic diversity in these two species can be inferred from this pattern? There is no doubt that the dynamics and extent of the glacial periods in the mountains of the central part of the Balkan Peninsula contributed to the genetic and chorological differentiation of these two species. Glaciation in this region was much less pronounced than in other parts of Europe (Hughes et al., 2006), raising the possibility of existence of multiple and scattered refugia for plant and animal species survival in the Balkan region (Hewitt, 1996). Available data show that the snowline in different parts of the Mediterranean region, and especially in the Balkan Peninsula, extended to different heights, from down to

1,250–1,600 m in the highest mountains of the region (Menkovic et al., 2004; Hughes et al., 2006), to 270 m in some coastal parts of the Dinarides (Marjanac and Marjanac, 2004). There is a possibility that the Balkan *Ramonda* species were mountain plants that descended to refugia in the surrounding gorges and canyons during the Ice Age, where they are still predominantly found today (Košanin, 1921; Stevanović et al., 1986; Lazarević, 2012). Also, the occurrence of a series of lakes during the Miocene and Pliocene, which spread in different parts of the Balkan Peninsula (Krstić et al., 2012), may have had a significant impact on the distribution of plant species and contributed to the isolation of *Ramonda* populations from each other.

Thus, the strong genetic differentiation within species and between populations exhibited by the two *Ramonda* species is likely the result of long reproductive isolation of populations and strong drift effects during glacial periods, but possibly before as well (Krstić et al., 2012), as the geographic distribution of populations is highly fragmented. A similar pattern was found for *R. myconi* in N Spain and the central Pyrenees, corresponding to its entire range, with high genetic differentiation between regions (Dubreuil et al., 2008). This pattern led the authors to suggest that this species occurred in several different but genetically unrelated refugia in this area. Similarly, in another Tertiary relict species from the family Gesneriaceae *Haberlea rhodopensis* Friv. two distinct genetic clusters were identified in Bulgaria, and the existence of which was suggested to be a consequence of isolation between refugial populations (Petrova et al., 2014). Finally, Petrova et al. (2018) have shown that the genetic diversity of *R. serbica* populations in Bulgaria is very low. The authors hypothesize that this is the result of both post-glacial fragmentation and a particularly pronounced lack of genetic connectivity for the Bulgarian populations, which are on the periphery of the species range.

This pattern of variability is expected in long-living plants with geographically isolated populations and probably with limited seed dispersal, as is the case in *Ramonda* species (Picó and Riba, 2002; Lazarević et al., 2013; Petrova et al., 2014). The overall genetic diversity of the two studied species is low, similar to what was found for *R. myconi* in the Pyrenees (Dubreuil et al., 2008) and *H. rhodopensis* in Bulgaria and Greece (Petrova et al., 2014, 2015). In the post-glacial period, eventual dispersal outside refugial areas may also have led to repeated demographic bottlenecks, which should enhance or at least maintain genetic differentiation of populations. All the more, these two species still have a very patchy distribution and most populations are still spatially isolated. It is noticeable that the population of *R. nathaliae* with the lowest genetic diversity is found on Mt. Suva Planina, which is at the highest altitude of all populations of this species studied here. Living conditions are harsh there, and rosettes are much smaller than elsewhere (authors' personal observation). It is therefore likely that the very low genetic diversity of this population is the result of a very small effective size, due to the combined effects of small size and potential selection.

The current ranges of these two species do not overlap, and only two narrow areas of sympatry have been observed so far (Stevanović et al., 1986). This observation and the marked genetic differentiation between these two species may indicate both that the two species were separated prior to the ice ages and that their refugia were distant from each other. However, despite the high genetic differentiation between the two species, no specific markers (alternative alleles fixed in either species) were found among the hundreds of AFLP fragments revealed in this study. However, this result should be confirmed on the basis of orthologous sequence data. The absence of specific markers may be evidence of a relatively recent divergence between the two species, which is consistent with the results of Petrova et al. (2015) on Gesneriaceae species. Based on chloroplastic and ITS sequence polymorphisms, these authors inferred that the timing of the divergence between *R. nathaliae* and *R. serbica* may have occurred in the middle or last Pleistocene. This divergence is one of the youngest divergence times among Gesneriaceae species estimated in this study. This could also mean that the genus *Ramonda* is a relict of the Tertiary, but that the individual species are actually younger.

Interploidy Gene Flow

Polyploidy is common in plant species and is mainly favored by meiosis defects leading to the production of unreduced gametes (Brownfield and Köhler, 2011; Mason and Pires, 2015; Kreiner et al., 2017). Intrageneric variability in ploidy level is widespread in the literature, and situations have been described where different cytotypes coexist in large contact zones as well as in geographically much narrower areas. Therefore, the possibility of interploidy gene flow within the same species or between related species has been a focus of attention for many years. However, the vast majority of cases reported in the literature in natural species involved diploid and tetraploid cytotypes, whereas cases of contact zones between diploid and hexaploid plants have been reported less frequently (see Kolář et al., 2017 for a comprehensive review of reported cases). In *Aster amellus* L., for example, diploid and hexaploid cytotypes are usually found in separate locations, while a few parapatric and one mixed-ploidy population have been observed. Based on fine-scale cytotype screening, the authors concluded that there is no evidence for the existence of tetraploid plants (i.e., first-generation hybrids) probably as a consequence of the exclusion of minority cytotypes (Castro et al., 2012), as well as niche differentiation and shifts in flowering times between cytotypes (Raabová et al., 2008). One of the most comparable situations to that observed in Balkan *Ramonda* is the case of *Senecio carniolicus* Willd. (syn. *Jacobaea carniolica* (Willd.) Schkrank), an autopolyploid complex with diploid and hexaploid plants as the most widespread ploidy levels. In contrast to *Ramonda* species, a high number of mixed ploidy populations were found in this species (Sonnleitner et al., 2010). Tetraploid individuals were observed in restricted areas, which the authors believe may correspond to glacial refugia in the eastern Alps. Other ploidy levels ($3x$, $5x$, $7x$–$9x$) are also observed, but at very low frequency. The authors noted that there should be barriers limiting the production and/or maintenance of interploid hybrids, to which the distinction between $2x$ and $6x$ habitats contributes. Lower fitness of progeny from crosses

between diploids and polyploids also clearly contributes to limited gene flow between cytotypes (Sonnleitner et al., 2013).

It is necessary to emphasize that in the case of *Ramonda*, in contrast to the examples cited above, it is a question of two cytotypes (2*x* and 6*x*) belonging not to one but to two different species. This makes hybridization more difficult and the genome constitution of the hybrids extremely diverse and sometimes difficult to explain. The case of *Ramonda* documented in this article therefore stands out from most studies of mixed-ploidy populations. It is interesting because, first, it shows enormous variability in genome size at micro-level in the only sites of sympatry between two species. This variation in GS allows us to hypothesize that the observed ploidy levels are also diverse, ranging from at least 2*x* to 8*x* (or possibly up to 10*x*), including odd ploidy levels such as at least 3*x* and 5*x*. For example, the class III of genome size (range between 8.94 and 12.06 pg, **Table 3**) could include 7*x*, 8*x*, but also higher ploidy levels. Based on the mean 1C*x* values of *R. nathaliae* and *R. serbica*, one would expect GS values between 8.15 and 9.24 pg for 7*x* individuals. For 8*x* individuals, the expected range of variation would be between 9.32 and 10.56 pg. Therefore, we could not exclude the possibility that 9*x* and/or 10*x* ploidy values were also present in our sample.

Octoploid individuals may have arisen from whole genome duplication (WGD) of tetraploid individuals, resulting in potential stabilization of the hybrid genome, a phenomenon already suggested by several authors (reviewed in Glombik et al., 2020). In contrast, the extended study of genome size in the two Balkan *Ramonda* species reported here showed that there was no cytotype variability within each species in monospecific populations, except for two individuals of *R. serbica* from Montenegro that were presumably 8*x* and 10*x* (Siljak-Yakovlev et al., 2008). Aneuploidy could also contribute to the observed variation in GS in both sites of sympatry. However, these hypotheses remain to be confirmed by cytogenetic studies. Finally, genome rearrangements and transposition activation have been documented in artificial interspecific hybridization experiments (Szadkowski et al., 2011; Liu et al., 2014; Mason and Wendel, 2020). We cannot rule out that such phenomena also contributed to the large variation in genome size observed in both sites of sympatry.

Nevertheless, the proportion of cytotypes distinct from the canonical ploidy levels 2*x* and 6*x* (a total of 33% of 4*x* plants) appears to be very high in these two sites of sympatry compared with most previously reported studies (Kolář et al., 2017). Other classes of cytotypes were also not rare [for example, 6% of plants suspected of being 5*x* (referred as class II in **Table 3**)].

The existence of hybridization between the two species, *viz.* between 2*x* and 6*x* cytotypes, in each of the two sites of sympatry was clearly demonstrated by: (i) the presence of individuals with an intermediate genome size compatible with the expected value for F1 tetraploid hybrids given the genome sizes of the two parental species; (ii) the presence of tetraploid individuals combining morphological features of the two species; and (iii) the observation that diploid and hexaploid individuals at the two sites of sympatry differed genetically much less (based on AFLP markers) than in monospecific populations. Several factors make this hybridization between the two species in sympatry possible. Both species flower at the same time. Although the flowers are slightly different, they are most likely pollinated by the same pollinators (Lazarević et al., 2013). Our results showed therefore that pre- and/or post-zygotic barriers to hybridization that usually exist between different cytotypes and/or plant species (Vallejo-Marín and Hiscock, 2016) are not operative in two sites of sympatry (Mallet, 2007; Abdelaziz et al., 2021).

The joint distribution of cytotypes and ancestry coefficients estimated from AFLP data in the two sympatry sites revealed a very singular pattern, namely a bimodality of these coefficients for diploids and hexaploids, compared to that observed in allopatric monospecific populations, but also for tetraploids. This striking observation may help to better understand the evolutionary dynamics that took place at these two sites by suggesting hypotheses on the pathways by which cytotype diversity was achieved and on the processes that might have been involved in the maintenance of tetraploid individuals and that would explain their genetic constitution.

The high number of tetraploid plants and their high AFLP diversity, similar to that of diploid and hexaploid plants in both sites of sympatry, suggests that hybridization between the two species is not an uncommon event in these sites. Alternatively, it is possible that only a few founders of tetraploid hybrids are at the origin of all currently observed tetraploid plants and that preferential breeding among plants of this cytotype has provided the opportunity for genetic recombination between the genomes of *R. serbica* and *R. nathaliae*. This would result in a wide variety of genetic combinations from the two parental genomes, including the existence of genetic profiles similar to the parental types (i.e., *R. nathaliae* or *R. serbica*). This latter possibility would imply that synapsis occurs between the chromosomes of the two species, which remains to be proven. However, in that case, one would expect to observe a continuous distribution of ancestry coefficients within the tetraploid group, from "nathaliae" to "serbica" and all possibilities in between. Data reported here showed a different pattern with a clear bimodality in this distribution. The coexistence of divergent genomes in interspecific hybrids is considered as a major mechanism of molecular and phenotypic novelty. This is explained by the huge structural and functional variation caused both by the change in ploidy level, if any, and by novel genetic interaction effects generated between the divergent genomes that merge (see Nieto Feliner et al., 2020, for a review). This variation is a raw material on which selection can operate. For example, negative epistasis can contribute to post-zygotic barriers to interspecific gene flow through a decrease in fitness of interspecific F1 hybrids and/or their progenies (Taylor et al., 2009), limiting therefore genetic introgression into restricted regions of genomes (adaptive introgression, e.g., Marburger et al., 2019). It is therefore possible that, among tetraploid *Ramonda* interspecific hybrids, selection in favor of "parental" combinations of co-adapted genes has taken place, because of weak survival or fertility of individuals having a mixed genome composition. The AFLP patterns of the different classes of each cytotype also

strongly suggest that different pathways were involved in the production of individuals of the same ploidy level, the 4*x* first-generation hybrids, but also other more unusual and rare ploidy levels. Three putative triploid plants have been observed that show strong ancestry from *R. nathaliae*. This suggested that these triploids could have been produced by fertilization between an unreduced 2*x* gamete of *R. nathaliae* and either a haploid gamete of *R. nathaliae* or a 1*x* gamete with complete *R. nathaliae* genome complement from an interspecific tetraploid hybrid. The latter case would imply that triploids are restricted to sympatry. In the first case, the presence of some triploid individuals would also be expected in *R. nathaliae* populations, but this has not been reported so far. However, the cytotype variability of monospecific populations of *R. nathaliae* has always been based on only a few individuals, and it cannot be excluded that such triploid individuals exist but have not yet been identified. Indeed, despite the high sample size of the present study, only three potential triploids at sympatry were found, all at Radovanski Kamen. However, it is now known that the production of non-reduced gametes is highly variable between individuals within populations and that this ability is also influenced by environmental conditions (De Storme and Mason, 2014). In addition, mainly because of chromosome pairing during meiosis, triploids are often sterile or at least with greatly reduced fertility (Comai, 2005) and therefore rarely remain in natural populations. The above-described pathway to triploid plants opens another way to explain the existence of tetraploid plants with high ancestry coefficient to "nathaliae." One could hypothesize at least two explanations. First, it is theoretically possible to obtain such profiles in progeny resulting from fertilization between a reduced gamete of a diploid *R. nathaliae* and a triploid gamete of a triploid *R. nathaliae*. The triploid bridge is considered to be very important for interploid gene flow and polyploidization (Husband, 2004). Secondly, unreduced (2*x*) gametes of *R. nathaliae* may also be involved in the production of tetraploid individuals that possess only chromosomes of *R. nathaliae* by autopolyploidy. The existence of such tetraploid individuals could also open the possibility of obtaining 6*x* progeny with only *R. nathaliae* genomes, the presence of which is suggested by the AFLP data. This hypothesis deserves attention because such tetraploid "nathaliae" would have inherited the ability to produce unreduced gametes (i.e., 4*x* gametes) from their *R. nathaliae* diploid parent.

Therefore, the AFLP profiles of the tetraploid individuals may indicate the existence of an autopolyploid-allopolyploid mixture at both sites of sympatry. They also suggest that there could be some barriers to reproduction between the two tetraploid hypothetical types, since otherwise tetraploid individuals with a much higher genome admixture would arise in the progeny from such crosses, which was not observed. However, the few individuals in our sample that exhibit such an intermediate AFLP profile do not allow us to rule out this last possibility. If this is indeed the case, the tetraploid level could also be a bridge between the two species, reinforcing interploidy gene flow in sympatry.

Pentaploid individuals were also found in both places. They may have originated in different ways. For example, they may simply have come from a cross between a tetraploid (of whatever type) and a hexaploid, or even from the union of an unreduced 2*x* gamete from *R. nathaliae* and a 3*x* gamete from a hexaploid. These pentaploid individuals could also play a role as mediators of gene flow, but without forming an independent evolutionary unit (Tel-Zur et al., 2020; Peskoller et al., 2021).

Obviously, other pathways could be invoked to explain all patterns of GS and AFLP profiles observed in this study. For example, it has been observed (Lazarević et al., 2013) that hybrid individuals produced pollen with highly variable sizes, a well-known proxy for irregular meioses in plants. It is therefore possible that 4*x* plants could produce haploid gametes with one *R. serbica* complement, a pathway for the production of 2*x* progenies with the genetic feature of *R. serbica*, an intriguing pattern we observed in our data.

Further studies are therefore needed to gain a comprehensive and detailed understanding of the biological mechanisms that explain the cytotype and molecular diversity observed at the two sites of sympatry. Further studies should focus on finding evidence for the production of unreduced gametes, especially in *R. nathaliae*, but also in triploids and tetraploids. In particular, a more detailed genetic study of the genetic profiles of the plants found at the two sites of sympatry, based on large-scale sequence data, should be useful to assess the extent of genome mosaicism with respect to each cytotype. This would help to assess the extent of genetic introgression between the two species and the potential for recombination between their genomes. However, it is already clear that hybridization between the two species is occurring and that, at the very least, backcrossing (in which tetraploid hybrids give rise to pentaploids) is possible. Unfortunately, backcrosses with one or both parental species often result in the appearance of individuals whose morphology is difficult to distinguish from the parental species, and identification of hybrids based on morphology becomes very difficult or impossible (Lihová et al., 2007; Čertner et al., 2015; Tendal et al., 2018; Beirinckx et al., 2020), as seems to be the case with both sites of sympatry for *Ramonda* (Lazarević et al., 2014).

Our data (both cytotypes and AFLP markers) therefore suggest that the first-generation interspecific hybrids are viable and possibly fertile. It has already been shown that pollen grains from tetraploid individuals are viable, although somewhat less so than pollen from parental cytotypes in sympatry (Lazarević et al., 2013). On the other hand, seeds of tetraploid individuals are drastically smaller than those of parental species and have very low or no germination capacity (Lazarević et al., 2013), which is often the case in interspecific hybrids (Mallet, 2008; Schmickl and Yant, 2021). Interestingly, with the exception of *R. nathaliae* in Oblik, seeds of both parental cytotypes have shown lower germination rates compared to monospecific populations (Lazarević et al., 2013), which could be the consequence of a reduction in fitness caused by various factors such as genetic introgression or inclusion of pollen from the hybrid individuals in pollination processes (Jacquemyn et al., 2016). The large number of hybrid individuals observed in

sympatry suggests that they arise from regular gene flow between parental species in these sites, as shown by Abdelaziz et al. (2021) in two *Erysimum* species. If hybridization events between the two species were rare, competition between the cytotypes would have been expected to eliminate the hybrids due to their rarity (cytotype exclusion), unless there were counteracting factors (homogamy, enlargement of the hybrid niche, or heterosis). Vegetative reproduction may also contribute to their maintenance (Lazarević et al., 2013). It is also possible that the process of minority cytotype exclusion either does not work for some reason that should be investigated or has not yet eroded the diversity of cytotypes in the two sites of sympatry. This last argument could support the idea that the only two known secondary contact zones between the two species are recent and probably formed during repeated glaciations in the Pleistocene. Finally, the overall panorama of genetic and cytotype variability may indicate that a secondary contact zone between the two species emerged after a long period of isolation (Schmickl and Yant, 2021).

CONCLUSION

The present study shows the complex influence of various factors on the evolution of two representatives of the genus *Ramonda* in the Balkan Peninsula. The turbulent geological past of the Balkan Peninsula, together with climatic fluctuations throughout the Mediterranean, has influenced the distribution, ecology, speciation and genetic diversity of *Ramonda* species, leading to the formation of isolated populations in refugial habitats. During migrations favored by interglacial periods, a secondary contact zone between *R. nathaliae* and *R. serbica* formed in a narrow area in E Serbia. In this newly formed sympatry, an intense hybridization process takes place between the two species, leading to the occurrence of different ploidy levels (from $2x$ to $>8x$) and individuals with mixed morphological characteristics, suggesting that not only crosses between the parental species occur, but also backcrosses and possibly duplication of the whole genome. These two sites of sympatry are places for a huge genetic mixing. Moderate pollen fertility, but low seed germination and low dispersal potential suggest that the effects of hybridization are limited to the sites of sympatry. However, the rich mixture of ploidy levels shows that these relict species, which have preserved a specific survival strategy, are not evolutionary "dead ends" and raises the question of possible ecological and evolutionary consequences of this phenomenon. These two locations constitute therefore an open-air laboratory for evolutionary and interspecific recombination studies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

SS-Y, BS, VS, and TR designed the study. ML, SS-Y, BS, VS, TR, AS, MN, and GT collected plant material. ML and SS-Y performed cytometric measurements and chromosome counts. ML, AS, and FL performed molecular laboratory work and provided data for statistical analyses done by TR and DH. ML, TR, and SS-Y drafted the manuscript. All authors read, revised, and approved the final version of the manuscript.

FUNDING

This research was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (No. 451–03–68/2022–14/200178), the International Franco-Serbian "Pavle Savić" project, and the Federative Institute of Research (IFR 87 La Plante et son environnement) at Gif-Orsay. The l'Institut Français de Serbie provided grants to ML.

ACKNOWLEDGMENTS

We thank all colleagues who helped us to collect plant material for this study: Vlado Matevski, Bojan Zlatković, Predrag Lazarević, Dmitar Lakušić, Nevena Kuzmanović, Antun Alegro, Danko Jović, and Dragan Nešić. Sequencing was performed using the Gentyane platform (INRAE, Clermont-Ferrand, France). We thank Mickaël Bourge and Nicolas Valentin, and Spencer C. Brown and Olivier Catrice for their assistance on the Cytometry Platform, Institute of Integrative Biology of the Cell (I2BC), Gif-sur-Yvette (http://www.imagif.cnrs.fr). Finally, we thank two reviewers and the editor Gonzalo Nieto Feliner for their valuable comments and suggestions that considerably improved our manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.873471/full#supplementary-material

REFERENCES

Abdelaziz, M., Muñoz-Pajares, A. J., Berbel, M., Garcia-Muñoz, A., Gómez, J. M., and Perfectti, F. (2021). Asymmetric reproductive barriers and gene flow promote the rise of a stable hybrid zone in the Mediterranean high mountains. *Front. Plant Sci.* 12:687094. doi: 10.3389/fpls.2021.687094

Bačič, T., Frajman, B., and Dolenc Koce, J. (2016). Diversification of *Luzula* sect. *Luzula* (Juncaceae) on the Balkan Peninsula: a cytogenetic approach. *Folia Geobot.* 51, 51–63. doi: 10.1007/s12224-016-9235-2

Banjanac, T., Dragićević, M., Šiler, B., Gašić, U., Bohanec, B., Nestorović Živković, J., et al. (2017). Chemodiversity of two closely related tetraploid *Centaurium* species and their hexaploid hybrid: metabolomic search for high-resolution taxonomic classifiers. *Phytochemistry* 140, 27–44. doi: 10.1016/j.phytochem.2017.04.005

Beirinckx, L., Vanschoenwinkel, B., and Triest, L. (2020). Hidden hybridization and habitat differentiation in a Mediterranean macrophyte, the euryhaline genus *Ruppia*. *Front. Plant Sci.* 11:830. doi: 10.3389/fpls.2020.00830

Bogunić, F., Siljak-Yakovlev, S., Mahmutović-Dizdarević, I., Hajrudinović-Bogunić, A., Bourge, M., Brown, S. C., et al. (2021). Genome size, cytotype diversity and

reproductive mode variation of *Cotoneaster integerrimus* (Rosaceae) from the Balkans. *Plan. Theory* 10:2798. doi: 10.3390/plants10122798

Brownfield, L., and Köhler, C. (2011). Unreduced gamete formation in plants: mechanisms and prospects. *J. Exp. Bot.* 62, 1659–1668. doi: 10.1093/jxb/erq371

Castro, S., Loureiro, J., Procházka, T., and Münzbergová, Z. (2012). Cytotype distribution at a diploid – hexaploid contact zone in *Aster amellus* (Asteraceae). *Ann. Bot. London* 110, 1047–1055. doi: 10.1093/aob/mcs177

Čertner, M., Kolář, F., Schönswetter, P., and Frajman, B. (2015). Does hybridization with a widespread congener threaten the long-term persistence of the eastern alpine rare local endemic *Knautia carinthiaca*? *Ecol. Evol.* 5, 4263–4276. doi: 10.1002/ece3.1686

Chakraborty, R., and Jin, L. (1993). Determination of relatedness between individuals by DNA fingerprinting. *Hum. Biol.* 65, 875–895. PMID: 8300084

Comai, L. (2005). The advantages and disadvantages of being polyploidy. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711

Contandriopoulos, J. (1966). Contribution à l'étude caryologique des Gésneriacées d'Europe et de leur germination. 91 Congrès des Sociétés Savantes. *Rennes* 3, 271–280.

Daskalova, E., Dontcheva, S., Zekaj, Z., Bacu, A., Sota, V., Abdullai, K., et al. (2012). Initial determination of polymorphism and *in vitro* conservation of some *Ramonda serbica* and *Ramonda nathaliae* populations from Albania, Macedonia and Bulgaria. *Biotechnol. Biotec. Eq.* 26, 16–25. doi: 10.5504/50YRTIMB.2011.0004

De Storme, N., and Mason, A. (2014). Plant speciation through chromosome instability and ploidy change: cellular mechanisms, molecular factors and evolutionary relevance. *Curr. Plant Biol.* 1, 10–33. doi: 10.1016/j.cpb.2014.09.002

Decanter, L., Colling, G., Elvinger, N., Heiðmarsson, S., and Matthies, D. (2020). Ecological niche differences between two polyploid cytotypes of *Saxifraga rosacea*. *Am. J. Bot.* 107, 423–435. doi: 10.1002/ajb2.1431

Doležel, J., Bartoš, J., Voglmayr, H., and Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. *Cytometry* 51A, 127–128. doi: 10.1002/cyto.a.10013

Dubreuil, M., Riba, M., and Mayol, M. (2008). Genetic structure and diversity in *Ramonda myconi* (Gesneriaceae): effects of historical climate change on a preglacial relict species. *Am. J. Bot.* 95, 577–587. doi: 10.3732/ajb.2007320

Earl, D. A., and vonHoldt, 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

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Feulgen, R., and Rossenbeck, H. (1924). Mikroskopisch-chemischer Nachweis einer Nukleinsäure vom Typus der Thymonukleinsäure und die darauf beruhende elektive Färbung von Zellkernen in mikroskopischen Präparaten. *Hoppe-Seyler's Zeitschr. Physiol. Chem.* 135, 203–248. doi: 10.1515/bchm2.1924.135.5-6.203

Frajman, B., Rešetnik, I., Weiss-Schneeweiss, H., Ehrendorfer, F., and Schönswetter, P. (2015). Cytotype diversity and genome size variation in *Knautia* (Caprifoliaceae, Dipscoideae). *BMC Evol. Biol.* 15:140. doi: 10.1186/s12862-015-0425-y

Glišić, Lj. (1924). Development of the female x-generation and embryo in Ramonda. dissertation thesis. Belgrade (Yugoslavia): University of Belgrade.

Glombik, M., Bačovský, V., Hobza, R., and Kopecký, D. (2020). Competition of parental genomes in plant hybrids. *Front. Plant Sci.* 11:200. doi: 10.3389/fpls.2020.00200

Hajrudinović, A., Frajman, B., Schönswetter, P., Silajdžić, E., Siljak-Yakovlev, S., and Bogunić, F. (2015). Towards a better understanding of polyploidy *Sorbus* (Rosaceae) from Bosnia and Herzegovina (Balkan Peninsula), including description of a novel, tetraploid apomictic species. *Bot. J. Linn. Soc.* 178, 670–685. doi: 10.1111/boj.12289

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* 58, 247–276. doi: 10.1006/bijl.1996.0035

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.* 68, 87–112. doi: 10.1111/j.1095-8312.1999.tb01160.x

Hughes, P. D., Woodward, J. C., and Gibbard, P. L. (2006). Quaternary glacial history of the Mediterranean mountains. *Prog. Phys. Geography* 30, 334–364. doi: 10.1191/0309133306pp481ra

Husband, B. C. (2004). The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biol. J. Linn. Soc.* 82, 537–546. doi: 10.1111/j.1095-8312.2004.00339.x

Jacquemyn, H., van der Meer, S., and Brys, R. (2016). The impact of hybridization on long-term persistence of polyploid *Dactylorhiza* species. *Am. J. Bot.* 103, 1829–1837. doi: 10.3732/ajb.1600274

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

Jiao, Y., Wickett, N., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94. doi: 10.1186/1471-2156-11-94

Kolář, F., Čertner, M., Suda, J., Schönswetter, P., and Husband, B. C. (2017). Mixed-ploidy species: progress and opportunities in polyploidy research. *Trends Plant Sci.* 22, 1041–1055. doi: 10.1016/j.tplants.2017.09.011

Košanin, N. (1921). *La distribution geographique des deux especes de Ramondia du Balkan (in French)*. Belgrade: Academie des Sciences et Arts

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

Kreiner, J. M., Kron, P., and Husband, B. C. (2017). Evolutionary dynamics of unreduced gametes. *Trends Genet.* 33, 583–593. doi: 10.1016/j.tig.2017.06.009

Krstić, N., Savić, L., and Jovanović, G. (2012). The Neogene lakes on the Balkan land. *Annales Géologiques de la Péninsule Balkanique* 73, 37–60. doi: 10.2298/GABP1273037K

Kuzmanović, N., Comanescu, P., Frajman, B., Lazarević, M., Paun, O., Schönswetter, P., et al. (2013). Genetic, cytological and morphological differentiation within the Balkan – Carpathian *Sesleria rigida* sensu Fl. Eur. (Poaceae): a taxonomical intricate tetraploid – octoploid complex. *Taxon* 62, 458–472. doi: 10.12705/623.13

Lakušić, D., Rakić, T., Stefanović, S., Surina, B., and Stevanović, V. (2009). *Edraianthus*×*lakusicii* (Campanulaceae) a new intersectional natural hybrid: morphological and molecular evidence. *Plant Syst. Evol.* 280, 77–88. doi: 10.1007/s00606-009-0168-6

Langella, O. (1999). POPULATIONS version 1.2.30: population genetic software. CNRS UPR9034.

Lazarević, M. (2012). Cytogenetics, palynology and phylogeography of genus Ramonda (Gesneriaceae) in the Balkan Peninsula. dissertation thesis. [Belgrade (RS)]: University of Belgrade.

Lazarević, M., Kuzmanović, N., Lakušić, D., Alegro, A., Schönswetter, P., and Frajman, B. (2015). Patterns of cytotype distribution and genome size variation in the genus *Sesleria* Scop. (Poaceae). *Bot. J. Linn. Soc.* 179, 126–143. doi: 10.1111/boj.12306

Lazarević, M., Rakić, T., and Šinžar-Sekulić, J. (2014). Morphological differences between the flowers of *Ramonda serbica*, *R. nathaliae* and their hybrid. *Bot. Serb.* 38, 91–98.

Lazarević, M., Siljak-Yakovlev, S., Lazarević, P., Stevanović, B., and Stevanović, V. (2013). Pollen and seed morphology of resurrection plants from the genus *Ramonda* (Gesneriaceae): relationship with ploidy level and relevance to their ecology and identification. *Turk. J. Bot.* 37, 872–885. doi: 10.3906/bot-1209-58

Levin, D. A. (2002). *The Role of Chromosomal Change in Plant Evolution*. New York, NY: Oxford University Press.

Lihová, J., Kučera, J., Perný, M., and Marhold, K. (2007). Hybridization between two polyploid *Cardamine* (Brassicaceae) species in North-Western Spain: discordance between morphological and genetic variation patterns. *Ann. Bot. London* 99, 1083–1096. doi: 10.1093/aob/mcm056

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I. A. P., et al. (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* 5:3930. doi: 10.1038/ncomms4930

Maherali, H., Walden, A. E., and Husband, B. C. (2009). Genome duplication and the evolution of physiological responses to water stress. *New Phytol.* 184, 721–731. doi: 10.1111/j.1469-8137.2009.02997.x

Mallet, J. (2007). Hybrid speciation. *Nature* 446, 279–283. doi: 10.1038/nature05706

Mallet, J. (2008). Hybridization, ecological arces and the nature of species: empirical evidence for the ease of speciation. *Phil. Trans. R. Soc.* 363, 2971–2986. doi: 10.1098/rstb.2008.0081

Marburger, S., Monnahan, P., Seear, P. J., Martin, S. H., Koch, J., Paajanen, P., et al. (2019). Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* 10:5218. doi: 10.1038/s41467-019-13159-5

Marie, D., and Brown, S. (1993). A cytometric exercise in plant DNA histograms, with 2C values for 70 species. *Biol. Cell.* 78, 41–51. doi: 10.1016/0248-4900(93)90113-S

Marjanac, L., and Marjanac, T. (2004). "Glacial history of the Croatian Adriatic and coastal Dinarides" in *Quaternary Glaciations — Extent and Chronology. Part I: Europe*. eds. J. Ehlers and P. L. Gibbard (Amsterdam: Elsevier), 19–26.

Mason, A. S., and Pires, J. C. (2015). Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends Genet.* 31, 5–10. doi: 10.1016/j.tig.2014.09.011

Mason, A. S., and Wendel, J. F. (2020). Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. *Front. Genet.* 11:1014. doi: 10.3389/fgene.2020.01014

Menkovic, L., Markovic, M., Cupkovic, T., Pavlovic, R., Trivic, B., and Banjanac, N. (2004). "Glacial morphology of Serbia Yugoslavia, with comments on the Pleistocene glaciation of Montenegro, Macedonia and Albania" in *Quaternary Glaciations – Extent and Chronology. Part I: Europe*. eds. J. Ehlers and P. L. Gibbard (Amsterdam: Elsevier), 379–384.

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

Nieto Feliner, G., Casacuberta, J., and Wendel, J. F. (2020). Genomics of evolutionary novelty in hybrids and polyploids. *Front. Genet.* 11:792. doi: 10.3389/fgene.2020.00792

Niketić, M., Đurović, S. Z., Tomović, G., Schönswetter, P., and Frajman, B. (2021). Diversification within ploidy-variable Balkan endemic *Cerastium decalvans* (Caryophyllaceae) reconstructed based on genetic, morphological and ecological evidence. *Bot. J. Linn. Soc.*:boab037. doi: 10.1093/botlinnean/boab037

Niketić, M., Siljak-Yakovlev, S., Frajman, B., Lazarević, M., Stevanović, B., Tomović, G., et al. (2013). Towards resolving the systematics of *Cerastium* subsection *Cerastium* (Caryophyllaceae): a cytogenetic approach. *Bot. J. Linn. Soc.* 172, 205–224. doi: 10.1111/boj.12050

Nunvářová Kabátová, K., Kolář, F., Jarolímová, V., Krak, K., and Chrtek, J. (2019). Does geography, evolutionary history or ecology drive ploidy and genome size variation in the *Minuartia verna* group (Caryophyllaceae) across Europe? *Plant Syst. Evol.* 305, 1019–1040. doi: 10.1007/s00606-019-01621-2

Otto, S. P., and Whitton, J. (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* 34, 401–437. doi: 10.1146/annurev.genet.34.1.401

Padilla-García, N., Rojas-Andrés, B. M., López-González, N., Castro, M., Castro, S., Loureiro, J., et al. (2018). The challenge of species delimitation in the diploid-polyploid complex *Veronica* subsection *Pentasepalae*. *Mol. Phylogenet. Evol.* 119, 196–209. doi: 10.1016/j.ympev.2017.11.007

Pančić, J. (1874). *Flora Kneževine Srbije*. Belgrade: Državna štamparija.

Pančić, J. (1884). *Dodatak flori Kneževine Srbije*. Belgrade: Kraljevsko-srpska državna štamparija.

Peskoller, A., Silbernagl, L., Hülber, K., Sonnleitner, M., and Schönswetter, P. (2021). Do pentaploid hybrids mediate gene flow between tetraploi *Senecio disjunctus* and hexaploid *S. carniolicus* s. str. (*S. carniolicus* aggregate, Asteraceae). *Alpine Bot.* 131, 151–160. doi: 10.1007/s00035-021-00254-x

Petrova, G., Dzhambazova, T., Moyankova, D., Georgieva, D., Michova, A., Djilianov, D., et al. (2014). Morphological variation, genetic diversity and genome size of critically endangered *Haberlea* (Gesneriaceae) populations in Bulgaria do not support the recognition of two different species. *Plant Evol. Syst.* 300, 29–41. doi: 10.1007/s00606-013-0857-z

Petrova, G., Moyankova, D., Nishii, K., Forrest, L., Tsiripidis, I., Drouzas, A. D., et al. (2015). The European paleoendemic *Haberlea rhodopensis* (Gesneriaceae) has an Oligocene origin and a Pleistocene diversification and occurs in a long-persisting refugial area in southeastern Europe. *Int. J. Plant Sci.* 176, 499–514. doi: 10.1086/681990

Petrova, G., Petrov, S., and Möller, M. (2018). Low genetic diversity in small leading adge populations of the European paleoendemic *Ramonda serbica* (Gesneriacae) in Bulgaria. *Nord. J. Bot.* 36:01655. doi: 10.1111/njb.01655

Petrović, S. (1882). *Flora okoline Niša*. Belgrade: Kraljevsko-srpska državna štamparija.

Picó, F. X., and Riba, M. (2002). Regional-scale demography of *Ramonda myconi*: remnant population dynamics in a preglacial relict species. *Plant Ecol.* 161, 1–13. doi: 10.1023/A:1020310609348

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

Raabová, J., Fischer, M., and Münzbergová, Z. (2008). Niche differentiation between diploid and hexaploid *Aster amellus*. *Oecologia* 158, 463–472. doi: 10.1007/s00442-008-1156-1

Rakić, T., Ilijević, K., Lazarević, M., Gržetić, I., Stevanović, V., and Stevanović, B. (2013). The resurrection flowering plant *Ramonda nathaliae* on serpentine soil – coping with extreme mineral element stress. *Flora* 208, 618–625. doi: 10.1016/j.flora.2013.09.006

Rakić, T., Lazarević, M., Jovanović, Ž. S., Radović, S., Siljak-Yakovlev, S., Stevanović, B., et al. (2014). Resurrection plants of the genus *Ramonda*: prospective survival strategies – unlock further capacity of adaptation, or embark on the path of evolution? *Front. Plant Sci.* 4:550. doi: 10.3389/fpls.2013.00550

Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., and Naik, V. G. (2014). STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springerplus* 3:431. doi: 10.1186/2193-1801-3-431

Ratter, J. A. (1963). Some chromosome numbers in the Gesneriaceae. *Notes Roy. Bot. Gard. Edinburgh* 24, 221–229.

Rey, P. J., Manzaneda, A. J., and Alcántara, J. M. (2017). The interplay between aridity and competition determines colonization ability, exclusion and ecological segregation in the heteroploid *Brachypodium distachyon* species complex. *New Phytol.* 215, 85–96. doi: 10.1111/nph.14574

Schmickl, R., and Yant, L. (2021). Adaptive introgression: how polyploidy reshapes gene flow landscapes. *New Phytol.* 230, 457–461. doi: 10.1111/nph.17204

Siljak-Yakovlev, S., Stevanovic, V., Tomasevic, M., Brown, S. C., and Stevanovic, B. (2008). Genome size variation and polyploidy in the resurrection plant genus *Ramonda*: cytogeography of living fossils. *Environ. Exp. Bot.* 62, 101–112. doi: 10.1016/j.envexpbot.2007.07.017

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

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., et al. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348. doi: 10.3732/ajb.0800079

Sonnleitner, M., Flatscher, R., García, P. E., Rauchová, J., Suda, J., Schneeweiss, G. M., et al. (2010). Distribution and habitat segregation on different spatial scales among diploid, tetraploid and hexaploid cytotypes of *Senecio carniolicus* (Asteraceae) in the eastern Alps. *Ann. Bot. London* 106, 967–977. doi: 10.1093/aob/mcq192

Sonnleitner, M., Weis, B., Flatscher, R., García, P. E., Suda, J., Krejčiková, J., et al. (2013). Parental ploidy strongly affects offspring fitness in heteroploid crosses among three cytotypes of autopolyploid *Jacobaea carniolica* (Asteraceae). *PLoS One* 8:e78959. doi: 10.1371/journal.pone.0078959

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

Stevanović, V., Jovanović, S., Lakušić, D., and Niketić, M. (1999). "Characteristics of the flora of Serbia and its phytogeographical division" in *Red Data Book of Flora of Serbia 1 – Extinct and Critically Endangered Taxa*. ed. V. Stevanović (Belgrade: Ministry of Environment of the Republic of Serbia; Faculty of Biology, University of Belgrade; Institute for Protection of Nature of the Republic of Serbia), 393–399.

Stevanović, V., Niketić, M., and Stevanović, B. (1986). Sympatric area of the sibling and endemo-relict species *Ramonda serbica* Panč. and *R. nathaliae* Panč. Et Petrov. (*Gesneriaceae*) in Southeast Serbia (Yugoslavia). *Glasn. Inst. Bot. Bašte Univ. Beograd* 20, 45–55.

Stevanović, V., and Stevanović, B. (1985). *Asplenio cuneifolii-Ramondaetum nathaliae* - new chasmophytic community on serpentine rocks in Macedonia. *Bull. Nat. Hist. Mus. Belgrade Ser. B* 40, 75–87.

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

Stutz, S., Hinz, H. L., Konowalik, K., Müller-Schärer, H., Oberprieler, C., and Schaffner, U. (2016). Ploidy level in the genus *Leucanthemum* correlates with resistance to a specialst herbivore. *Ecosphere* 7:e01460. doi: 10.1002/ecs2.1460

Szadkowski, E., Eber, F., Huteau, V., Lode, M., Coriton, O., Jenczewski, E., et al. (2011). Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. *New Phytol.* 191, 884–894. doi: 10.1111/j.1469-8137.2011.03729.x

Taylor, S. J., Arnold, M., and Martin, N. H. (2009). The genetic architecture of reproductive isolation in Louisiana irises: hybrid fitness in nature. *Evolution* 63, 2581–2594. doi: 10.1111/j.1558-5646.2009.00742.xte

te Beest, M., Le Roux, J. J., Richardson, D. M., Brysting, A. K., Suda, J., Kubešová, M., et al. (2012). The more the better? The role of polyploidy in facilitating plant invasions. *Ann. Bot. London* 109, 19–45. doi: 10.1093/aob/mcr277

Tel-Zur, N., Mouyal, J., Zurgil, U., and Mizrahi, Y. (2020). In support of Winge's theory of hybridization followed by chromosome doubling. *Front. Plant Sci.* 11:954. doi: 10.3389/fpls.2020.00954

Tendal, K., Ørgaard, M., Larsen, B., and Pedersen, C. (2018). Recurrent hybridisation events between *Primula vulgaris*, *P. veris* and *P. elatior* (Primulaceae, Ericales) challenges the species boundaries: using molecular markers to re-evaluate morphological identifications. *Nord. J. Bot.* 36:e01778. doi: 10.1111/njb.01778

Thompson, J. D. (2020). *Plant Evolution in the Mediterranean: Insights for Conservation*. New York, NY: Oxford University Press.

Turill, W. B. (1929). *The Plant-Life of the Balkan Peninsula. A Phytogeographical Study*. Oxford: Clarendon Press.

Ulum, F. B., Hadacek, F., and Hörandl, E. (2021). Polyploidy improves photosynthesis regulation within the *Ranunculus auricomus* complex (Ranunculaceae). *Biology* 10:811. doi: 10.3390/biology10080811

Vallejo-Marín, M., and Hiscock, S. J. (2016). Hybridization and hybrid speciation under global change. *New Phytol.* 211, 1170–1187. doi: 10.1111/nph.14004

Van de Peer, Y., Ashman, T.-L., Soltis, P. S., and Soltis, D. E. (2021). Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* 33, 11–26. doi: 10.1093/plcell/koaa015

Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi: 10.1038/nrg.2017.26

Vekemans, X. (2002). *AFLP-SURV Version 1.0. Distributed by the author*. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. V. D., Hornes, M., et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414. doi: 10.1093/nar/23.21.4407

Weiss-Schneeweiss, H., Emadzade, K., Jang, T.-S., and Schneeweiss, G. M. (2013). Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenet. Genome Res.* 140, 137–150. doi: 10.1159/000351727

Wendel, J. F. (2000). Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249. doi: 10.1023/A:1006392424384

Zhivotovsky, L. A. (1999). Estimating population structure in diploids with multilocus dominant DNA markers. *Mol. Ecol.* 8, 907–913. doi: 10.1046/j.1365-294x.1999.00620.x

Živković, T., Quartacci, M. F., Stevanović, B., Marinone, F., and Navari-Izzo, F. (2005). Low-molecular weight substances in the poikilohydric plant *Ramonda serbica* during dehydration and rehydration. *Plant Sci.* 168, 105–111. doi: 10.1016/j.plantsci.2004.07.018

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 17 June 2022
doi: 10.3389/fpls.2022.842842

Reticulate Evolution in the Western Mediterranean Mountain Ranges: The Case of the *Leucanthemopsis* Polyploid Complex

Salvatore Tomasello[1] and Christoph Oberprieler[2]*

[1] Department of Systematics, Biodiversity and Evolution of Plants (With Herbarium), University of Göttingen, Göttingen, Germany, [2] Evolutionary and Systematic Botany Group, Institute of Plant Sciences, University of Regensburg, Regensburg, Germany

Polyploidization is one of the most common speciation mechanisms in plants. This is particularly relevant in high mountain environments and/or in areas heavily affected by climatic oscillations. Although the role of polyploidy and the temporal and geographical frameworks of polyploidization have been intensively investigated in the alpine regions of the temperate and arctic biomes, fewer studies are available with a specific focus on the Mediterranean region. *Leucanthemopsis* (Asteraceae) consists of six to ten species with several infraspecific entities, mainly distributed in the western Mediterranean Basin. It is a polyploid complex including montane, subalpine, and strictly alpine lineages, which are locally distributed in different mountain ranges of Western Europe and North Africa. We used a mixed approach including Sanger sequencing and (Roche-454) high throughput sequencing of amplicons to gather information from single-copy nuclear markers and plastid regions. Nuclear regions were carefully tested for recombinants/PCR artifacts and for paralogy. Coalescent-based methods were used to infer the number of polyploidization events and the age of formation of polyploid lineages, and to reconstruct the reticulate evolution of the genus. Whereas the polyploids within the widespread *Leucanthemopsis alpina* are autopolyploids, the situation is more complex among the taxa endemic to the western Mediterranean. While the hexaploid, *L. longipectinata*, confined to the northern Moroccan mountain ranges (north–west Africa), is an autopolyploid, the Iberian polyploids are clearly of allopolyploid origins. At least two different polyploidization events gave rise to *L. spathulifolia* and to all other tetraploid Iberian taxa, respectively. The formation of the Iberian allopolyploids took place in the early Pleistocene and was probably caused by latitudinal and elevational range shifts that brought into contact previously isolated *Leucanthemopsis* lineages. Our study thus highlights the importance of the Pleistocene climatic oscillations and connected polyploidization events for the high plant diversity in the Mediterranean Basin.

Keywords: Anthemideae, AllCoPol, BEAST2, *Leucanthemopsis*, polyploidy, reticulate evolution, western Mediterranean

OPEN ACCESS

Edited by:
Božo Frajman, University of Innsbruck, Austria

Reviewed by:
Judita Zozomova-Lihova, Institute of Botany (CAS), Slovakia
Santiago Martín-Bravo, Universidad Pablo de Olavide, Spain

****Correspondence:***
Salvatore Tomasello
salvatore.tomasello@uni-goettingen.de

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

Received: *24 December 2021*
Accepted: *18 May 2022*
Published: *17 June 2022*

Citation:
Tomasello S and Oberprieler C (2022) Reticulate Evolution in the Western Mediterranean Mountain Ranges: The Case of the Leucanthemopsis Polyploid Complex. Front. Plant Sci. 13:842842. doi: 10.3389/fpls.2022.842842

INTRODUCTION

Polyploidy is the presence of three or more complete chromosome sets in an organism. The frequency of whole-genome duplications (WGDs) is heterogenous throughout the tree of life, and it is particularly prominent in plants. The ancestor of all angiosperms was polyploid (Jiao et al., 2011) and further rounds of WGD occurred in different angiosperm lineages during the Cretaceous–Paleogene boundary (Vanneste et al., 2014). Approximately 15% of speciation events in angiosperms involved an increase in the number of complete chromosome sets and their further diversification resulted in 35% of extant angiosperms being polyploid (Wood et al., 2009). The study of polyploid complexes and polyploid formation is therefore fundamental to improving our understanding of plant evolution.

Polyploid species are not evenly distributed on the Earth, with polyploid frequencies increasing toward high latitudes (Rice et al., 2019). Additionally, polyploidy is relatively frequent in plants of the Mediterranean Basin, where it has been a driving force for diversification in several genera, e.g., *Campanula* L. (Crowl et al., 2017; Liveri et al., 2020), *Centaurea* L. (Garcia-Jacas et al., 2009; Moreyra et al., 2021), *Centaurium* Hill., (Mansion et al., 2005), *Leucanthemum* Mill., (Oberprieler et al., 2014), and *Veronica* L. (López-González et al., 2021). About 30% of plant species in the whole Mediterranean biome are polyploids (Rice et al., 2019), with similar estimates given for the European Mediterranean Basin (36%; Marques et al., 2018). In the particular case of the Iberian Peninsula, numbers are even higher and approximately 50% of all plant species, including angiosperms, are polyploids (Marques et al., 2018).

Polyploidy is often linked to hybridization (i.e., allopolyploidy). In contrast to autopolyploids, which contain elevated numbers of sub-genomes from an individual or from different individuals of the same species, allopolyploids derive from the merging of two or more chromosome sets from different species (Ramsey and Schemske, 1998; Van De Peer et al., 2017). Of the two processes, allopolyploidy has always been considered to be more common (Stebbins, 1950, Grant, 1981, Coyne and Orr, 2004; but see Ramsey and Schemske, 1998, 2002; Soltis et al., 2007). The reticulate character of allopolyploid evolution makes inference of phylogenies in those groups particularly difficult. Moreover, the increased number of chromosome sets exacerbates problems connected with gene paralogy, and the effect of stochastic factors intrinsic to evolution, such as incomplete lineage sorting (ILS), becomes more dramatic as a consequence of the increased effective population size. Disentangling between hybridization and ILS in (allo)polyploids is not a simple task (e.g., Avise, 1994; Rosewich and Kistler, 2000; Linder et al., 2003), and only a few methods have been proposed in the last decade that are capable of reconstructing reticulate evolution in polyploid complexes (Oxelman et al., 2017; Rothfels, 2021).

Among those methods, AlloppNET and AlloppMUL are capable of modeling polyploid evolution and producing phylogenetic networks and multi-labeled (MUL-)species trees in the presence of both ILS and hybridization (Jones et al., 2013; Jones, 2017b). Unfortunately, the model is computationally demanding and has been designed to infer phylogenies in relatively small datasets, including diploids and a single tetraploid species. Recently, Yan et al. (2021) extended a maximum parsimony method for inferring phylogenetic networks in the presence of ILS and hybridization (Yu et al., 2013) to complexes including polyploids. However, the method needs prior knowledge on the mode of polyploid formation (i.e., the number of reticulations must be specified in advance). Moreover, alleles of polyploid species/samples are not sorted into sub-genomes; a piece of information that might be needed for further and more specific analyses (e.g., species delimitation of polyploids, age estimation, and post-formation genome evolution).

Other methods utilize iterative approaches to sort alleles of polyploids into sub-genomes and then infer MUL-species trees and networks. Some of these methods use information from the plastid genome to determine which of the homeologs found in nuclear genes are from the maternal progenitor by looking at (in)congruence between the plastid phylogeny and the nuclear gene trees (Bertrand et al., 2015; Šlenker et al., 2021). Problems arise when dealing with ploidy levels higher than 4*x* (hexaploids may potentially comprise three different sub-genomes) or in cases when the maternal progenitor of a polyploid is extinct or not sampled, as often reported in polyploid complexes (Oberprieler et al., 2014; Karbstein et al., 2022). The program, AllCoPol (Lautenschlager et al., 2020) uses the number of deep coalescences to sort alleles into parental sub-genomes and does not need prior information on polyploid formations and/or maternal progenitors. It is based on the combinatoric approach described in Oberprieler et al. (2017), but the implementation of heuristics and machine-learning algorithms makes it applicable to larger datasets including high polyploids.

Leucanthemopsis (Giroux) Heywood is a small genus of the chamomile tribe of the daisy family (Compositae, Anthemideae). It is a polyploid complex including diploids, tetraploids, and hexaploids. According to the present taxonomy, the genus consists of ten species (Pedrol, 2019), eight of which are endemic to the Iberian Peninsula (**Figure 1**). The hexaploid, *L. longipectinata* (Font Quer) Heywood is a narrow endemic of the Rif Mountains of northern Morocco, whereas *L. alpina* (L.) Heywood is more widespread. The latter species is a polymorphic complex distributed in the south-western and central European alpine ranges (from the Pyrenees to the Carpathians) and comprises diploid, tetraploid, and hexaploid populations. Most of *Leucanthemopsis* species, especially the polyploid complexes (eight out of ten species), show a high degree of morphological polymorphism (**Figure 2**), and several infraspecific taxa have been described in the past (Heywood, 1975; Pedrol, 2019).

The onset of diversification processes in *Leucanthemopsis* has been estimated to be c. 4.4 million years (Ma) ago and the differentiation among the Iberian taxa was most likely within the last two million years (Tomasello et al., 2015). However, Tomasello et al. (2015) included only a few, mostly diploid samples, leaving phylogenetic relationships among Iberian taxa and temporal origin and diversification of polyploids unresolved.

In the present study, we therefore, aim at resolving the phylogenetic relationships within the western-Mediterranean genus *Leucanthemopsis*, including a comprehensive sampling

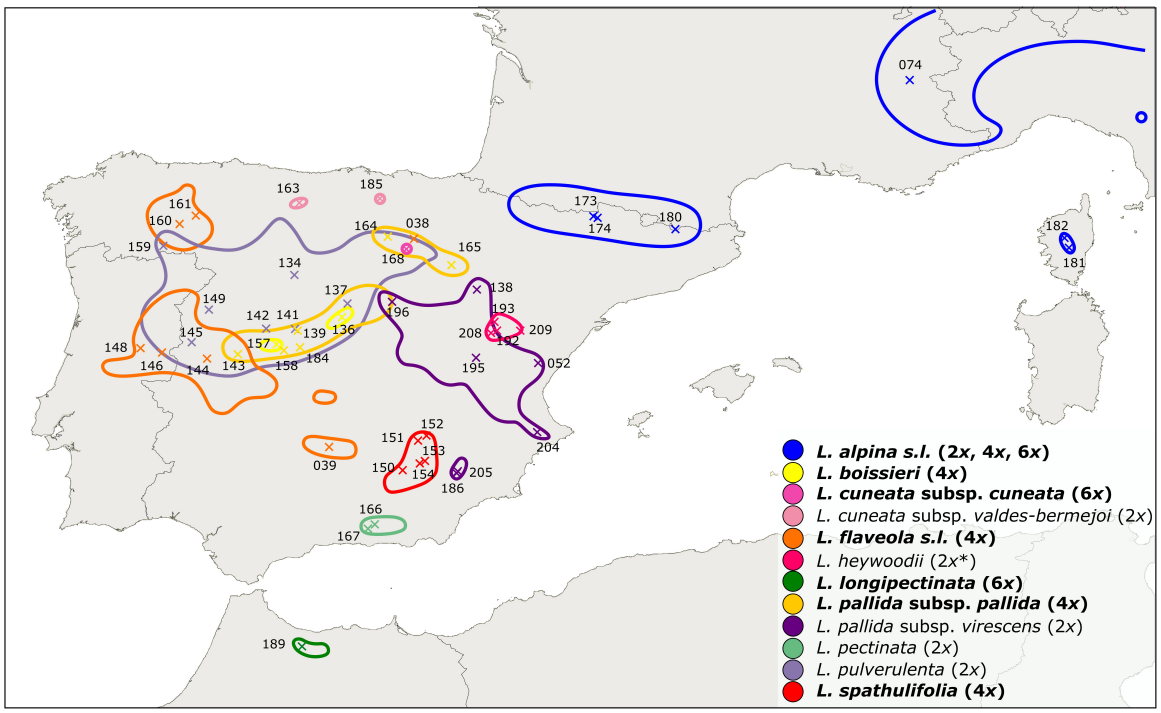


FIGURE 1 | Distribution map of *Leucanthemopsis* taxa in the western Mediterranean as inferred from literature and herbarium records (**Supplementary Figure 2**; **Supplementary Data Sheet 4**). Taxa in bold are polyploids or include polyploid populations and the corresponding ploidy levels follow the names. * Diploid ploidy of *L. heywoodii* was inferred by our flow-cytometric measurements, whereas Pedrol (2019) reported this taxon to be tetraploid. ×, geographic position of the populations included in the present study with corresponding IDs ("LPS" in **Supplementary Data Sheet 1**).

(all species described and more individuals per species) and using sequence information from a number of single/low copy nuclear genes as well as plastid regions. More specifically, we investigate the mode (auto- vs. allopolyploid), the frequency, and the temporal framework of the formation of polyploids and thus contribute to improving our understanding of the importance of polyploid evolution for the biodiversity of the Mediterranean Basin.

MATERIALS AND METHODS

Plant Material

Fifty-two accessions from members of the genus *Leucanthemopsis* were included in the present study. We used one to seven samples per taxon (17 taxa belonging to 10 *Leucanthemopsis* species). For population LPS168 of *L. cuneata* (Pau) Holub subsp. *cuneata,* two individuals were included, whereas from all other populations we used single individuals. Two additional accessions belonging to the sister genera *Prolongoa* Boiss. and *Hymenostemma* Kunze ex Willk. (Tomasello et al., 2015) were also included and used as outgroup in some of the analyses. Several accessions were collected by S. Tomasello in the summers of 2010 and 2011, silica-gel dried, and employed already in Tomasello et al. (2015), Tomasello and Oberprieler (2017), and Oberprieler et al. (2017). Additional samples were collected during the summers of 2015 and 2016 or sampled from herbarium specimens from the herbaria of the Real Jardín Botánico (MA), the Botanical Garden and Botanical Museum Berlin-Dahlem (B), and the Bavarian Natural History Collections Munich (M). This was done to include all taxa described by Heywood (1975) and Pedrol (2019), and to incorporate, when possible, material from *loci classici* or herbarium vouchers cited in the above-mentioned revisions. The importance of including topotypic material in phylogenetic studies involving intricate taxonomic groups has already been highlighted in other studies (Otero et al., 2021). A complete list of samples used in the present study is provided in the **Supplementary Data Sheet 1**. Herbarium vouchers for all the newly collected samples were deposited in the Herbarium Mediterraneum Panormitanum (PAL) and in some cases in the herbaria of the Real Jardín Botánico (MA), and of the University of Salamanca (SALA).

Flow Cytometry

For samples collected in the field and dried in silica-gel, ploidy was estimated via flow cytometry mostly using five accessions per population. For populations collected in "Puerto de Paramera" (LPS139-LPS141), where the diploid *L. pulverulenta* (Lag.) Heywood grows sympatrically with the tetraploid, *L. pallida* subsp. *pallida* (Miller) Heywood, we measured all collected and surveyed individuals (ten individuals per population). We, therefore, estimated the ploidy level for 40 populations belonging to 11 *Leucanthemopsis* taxa. Approximately 10% of the samples were re-analyzed and used

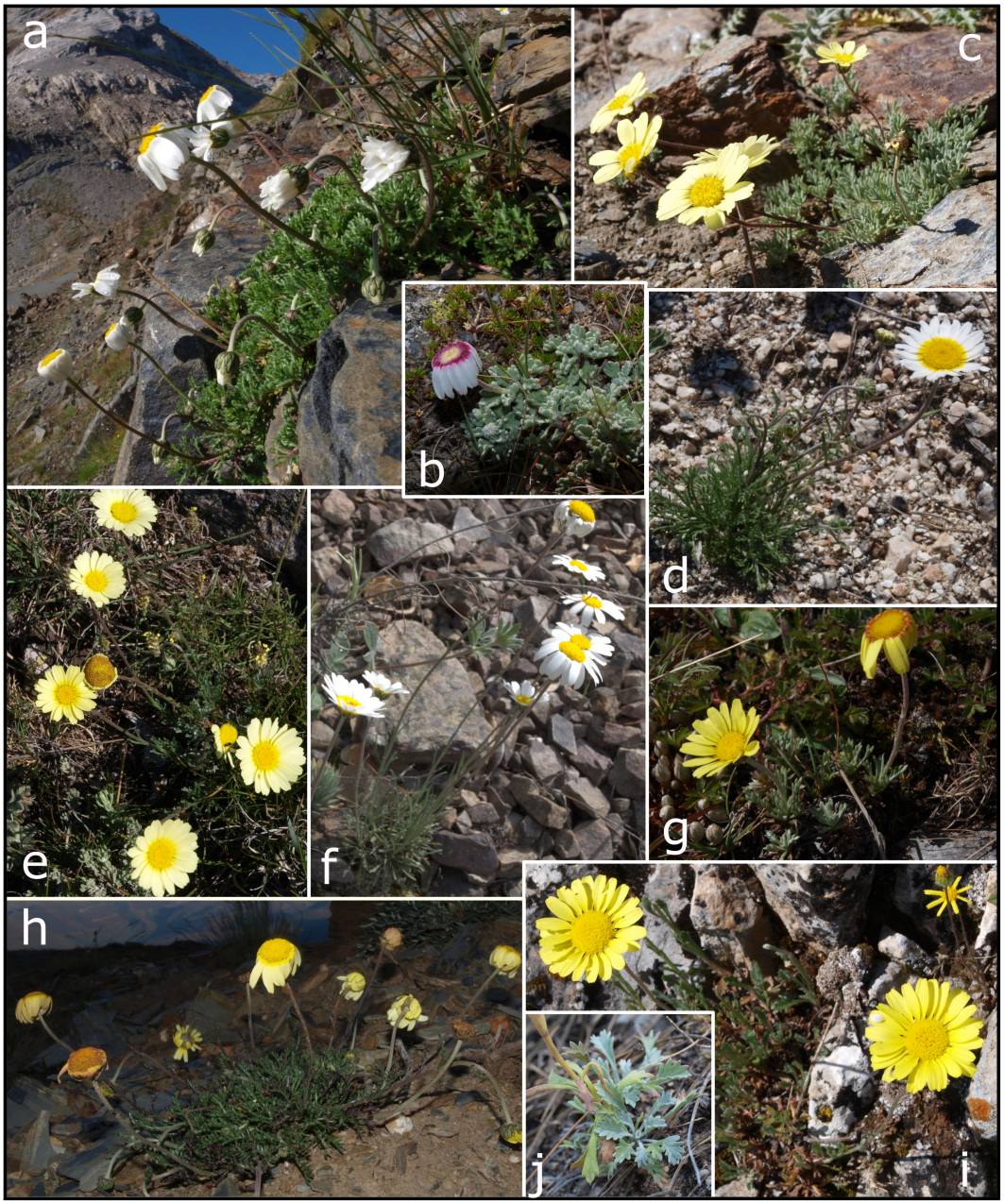


FIGURE 2 | Morphological diversity in *Leucanthemopsis*: **(a)** *L. alpina* subsp. *pyrenaica* (Port de Boucharo, Hautes-Pyrénées, France); **(b)** *L. alpina* subsp. *tomentosa* at Monte Renoso (Corsica, France). **(c)** *L. pectinata* at Veleta peak (Granada, Spain). **(d)** *L. pulverulenta* at Puerto de Villatoro (Avila, Spain). **(e)** *L. pallida* subsp. *pallida* at Puerto de Villatoro (Avila, Spain). **(f)** *L. pallida* subsp. *virescens* at Sierra de Vicort (Zaragoza, Spain). **(g)** *L. cuneata* subsp. *valdes-bermejoi* at Sierra del Brezo (Palencia, Spain). **(h)** *L. flaveola* at Peña Trevinca (Orense, Spain). **(i)** *L. spathulifolia* at Sierra de Cazorla (Jaén, Spain), and **(j)** detail of the spathulate basal leaves. Photos by Salvatore Tomasello.

as replicates. *Petunia hybrida* E.Vilm. cv. PxPc6 was used as an internal standard (2C = 2.85 pg; Marie and Brown, 1993). Approximately 0.5–1 cm^2 of leaf tissue of the standard and two- to three-fold tissue of the dehydrated sample material were employed. Nuclei were isolated by grinding the leaf material in Otto I buffer (Otto, 1992; Doležel and Göhde, 1995) and subsequently stained with 4′,6-diamidino-2-phenylindole (DAPI) in LB01 buffer (Doležel et al., 1989) modified with the supplement of ß-Mercaptoethanol (0.015 mM). Samples were measured on a PARTEC CyFlow® Space instrument (Partec GmbH, Münster, Germany). For each sample, at least 8,500 nuclei were counted. The results were processed using the FloMax® software (Partec GmbH, Münster, Germany). The ratio between the relative fluorescence of the sample nuclei and

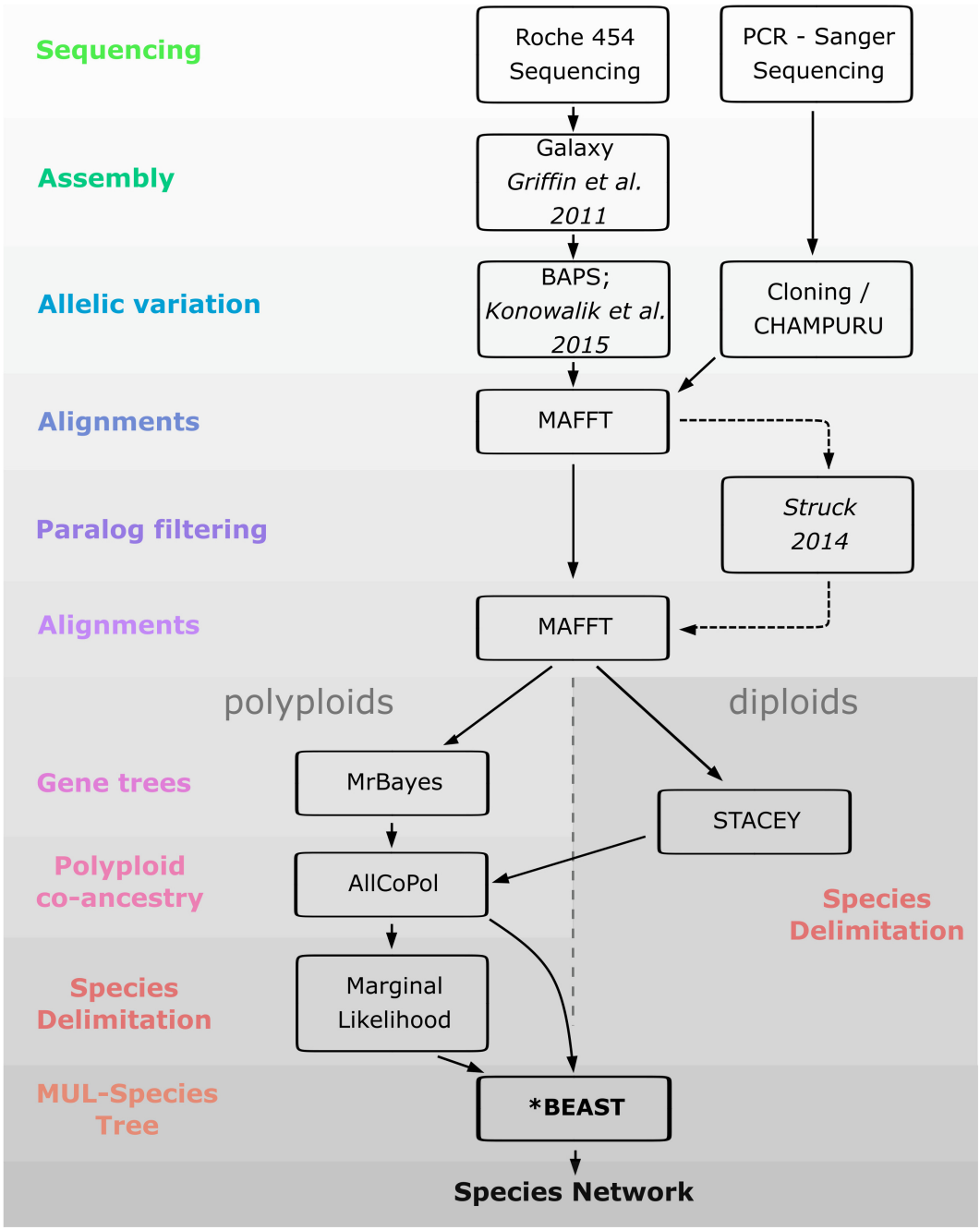


FIGURE 3 | Workflow of the analyses performed in the present study.

of those of the standard was used to estimate the ploidy of *Leucanthemopsis* accessions.

Conceptual Framework for the Phylogenetic Analyses

We used a mixed approach including Sanger sequencing and (Roche-454) high throughput sequencing of amplicons to gather allelic information from single-copy nuclear markers and sequences from two plastid regions. Nuclear regions were carefully tested for recombinants/PCR artifacts and for paralogy using gene phylogenies and looking at highly supported clades emerging from long branches and including samples from different taxa.

As a prerequisite for the reconstruction of polyploid networks, coalescent-based species delimitation approaches were first used to circumscribe diploid lineages. We then investigated the mode of formation of polyploids (auto- vs. allopolyploids)

by using allelic information along with methods capable to sort alleles in putative parental sub-genomes and infer species networks. Since taxon circumscription of polyploids of the Iberian Peninsula is controversial, we also estimated the number of times (allo)polyploidization took place independently by using marginal likelihood calculations for different taxon-circumscription hypotheses. Finally, we used coalescent-based methods to roughly estimate the age of the polyploid lineages and the temporal framework of reticulation events. The workflow of the analyses performed in the study is depicted in **Figure 3**.

DNA Extraction, Amplification, and Library Preparation

Deoxyribonucleic acid extracts were obtained using either the DNeasy Plant Mini Kit (Qiagen, Venlo, Netherland) in the laboratory of the CSIC "Real Jardín Botánico" in Madrid or a modified protocol based on the CTAB method by Doyle and Doyle (1987) at the Institute of Plant Sciences of Regensburg University.

For downstream phylogenetic analyses, two plastid markers (the intergenic spacer regions *psbA-trnH* and *trnC-petN*) and five single-copy nuclear markers (*B12*, *B20*, *C12*, *C16*, and *D35*; Chapman et al., 2007) were employed. The plastid spacer region *psbA-trnH* was amplified using the primers *psb*Af and *trn*Hr (Sang et al., 1997), whereas for the *trnC-petN* spacer region, we used the primers *trn*C (Demesure et al., 1995) and *pet*N1r (Lee and Wen, 2004). PCR amplifications were performed using the *Taq* DNA Polymerase Master-mix Red (Ampliqon/Biomol, Odense, Denmark) in a final volume of 12.5 μl, and according to the protocol suggestions of the manufacturer. The following temperature profile was employed: 2′ at 95°C, then 36 cycles of 30″ at 95°C, 60″ at 62°C, and 60″ at 72°C, with a final extension of 5′ at 72°C. PCR products were purified using Agencourt AMPure magnetic beads (Agencourt Bioscience Corporation, Beverly, MA, United States). Cycle sequencing was performed with the DTCS Sequencing Kit (Beckman Coulter, Fullerton, CA, United States), following the protocol suggested by the manufacturer. Sequences were analyzed on a CEQ 8000 sequencer (Beckman Coulter, Fullerton, CA, United States) and the obtained electropherograms were carefully checked for ambiguities using CHROMAS LITE v2.10[1] (Technelysium Pty Ltd., Tewantin, Australia) and/or FINCHTV v1.3.1 (Geospiza Inc., Seattle, United States). When necessary, we used the IUPAC codes to indicate single nucleotide polymorphisms (SNPs). In the electropherograms, a site was considered polymorphic when more than one peak was present and the weakest one reached at least 25% of the intensity of the strongest one (Fuertes Aguilar et al., 1999; Mansion et al., 2005).

Allelic variation for four of the five single-copy nuclear markers (all except for *D35*) was assessed via Roche-454 high throughput pyrosequencing. Prior to sequencing, amplicons went through two rounds of PCRs. In the first round, the Peqlab KAPAHiFi polymerase (Peqlab Biotechnologie GmbH, Erlangen, Germany) was used to reduce PCR errors as much as possible. The forward primers used for the amplifications were those from Chapman et al. (2007), tailed with a 29 bp-long M13-forward primer. The GS FLX Titanium Primer B sequence was added to the reverse primers. The reverse *C12* primer was designed to obtain *Leucanthemopsis* amplicons shorter than 350 bp (**Supplementary Table 1**). The PCRs were performed in a final volume of 15 μl, following the manufacturer's instructions and using the following "touch-down" program: 95°C for 5′; 20″ at 98°C, 30″ at 65–61°C, and 30″ at 72°C for 5 cycles; finally 35 cycles of 98°C for 20″, 60°C for 30″, and 72°C for 30″, with a final extension step of 72°C for 5′.

After purification of the PCR products, the second PCR round was performed to add sample-specific 4–5 bp long barcodes [i.e., Multiplex identifiers (MIDs)] to the amplicons. Therefore, the forward primers consisted of the following sequence combination: GS_FLX_Titanium_Primer_A–MID–M13-tail, while the reverse primer was always the GS_FLX_Titanium_Primer_B. A two-step PCR program was employed: 3′ at 95°C; 20 cycles of 20″ at 95°C, and 1′ at 68°C; with a final extension of 5′ at 72°C. After purification of PCR products, concentrations and fragment lengths were measured, and libraries were pooled in a way that allowed for an appropriate sequence coverage to detect all alleles in diploids, tetraploids, and hexaploids (as described in Oberprieler et al., 2017). Sequencing was performed on a Roche-454 pyrosequencing machine at Microsynth (Balgach, Switzerland).

The fifth single-copy nuclear marker, *D35* included a simple sequence repeat (SSR) or microsatellite motive. Because of this, even diploid accessions often showed alleles that differed in length. This allowed us to decrypt the different alleles by using forward and reverse sequence information as described by Flot et al. (2006) and to avoid next-generation sequencing or cloning. The amplicons of this marker were Sanger-sequenced and alleles were detected in diploids with the software CHAMPURU v1.0 (Flot, 2007), whereas in tetraploids, they were deciphered manually. Since this method is not applicable when more than four sequences of different lengths overlap in the electropherograms, we cloned the hexaploid accessions [two *L. alpina* subsp. *pyrenaica* (Vierh.) Tomasello & Oberpr., two *L. cuneata* subsp. *cuneata,* and one *L. longipectinata*]. For this purpose, we used the CloneJET PCR cloning kit (Fermentas, Waltham, MA, United States) according to the manufacturer's recommendations. Twenty-seven clones were picked for each accession, in order to have a probability of 0.95 to get sequence information for all possible alleles (refer to the formula proposed by Joly et al., 2006).

Assembly, Allele Detection, and Alignment of Markers

For those samples and markers sequenced on a Roche-454 sequencing machine, reads were assembled using R (R Development Core Team, 2008) and the Galaxy web portal (Giardine et al., 2005; Goecks et al., 2010) as described by Griffin et al. (2011). Reads were assigned to marker regions using forward-primer sequences and to accessions using the sample-specific MIDs. Reads with Phred scores below 20 in more than 10% of the nucleotide positions were discarded.

[1] http://technelysium.com.au/chromas.html

Potential alleles and/or chimerical sequences were identified using BAPS v5.2 (Corander et al., 2006, 2008; Cheng et al., 2011) and the procedure described by Konowalik et al. (2015). Accordingly, clusters with read numbers above the minimum expected for an allele, considering the ploidy of the sample and the total number of reads gained for the specific sample/marker combination (Oberprieler et al., 2017), were kept as possible allelic variants. By applying the "admixture based on mixture clustering" option in BAPS to the previously found clusters, reads of intermediate position, potentially resulting from alleged recombination during PCRs, were pinpointed and discarded after an additional visual inspection in ALIVIEW v1.18 (Larsson, 2014). Once alleles were identified, reads were collapsed into consensus sequences, applying a 20% threshold as a criterion to retain intra-allelic polymorphisms by scoring them as IUPAC-coded wobble nucleotide positions. Consensus sequences (alleles) were then marker-wise processed in ALIVIEW and aligned using MAFFT v6.833b (Katoh et al., 2002; Katoh and Toh, 2008).

Sequences from direct sequencing (e.g., from the plastid regions), cloning, or CHAMPURU (*D35*) were directly added to the alignments and aligned using MAFFT.

Paralog Filtering

Although the selected nuclear regions are supposed to be single-copy in members of Compositae (Chapman et al., 2007), we checked for putative paralog sequences in the alignment following a similar approach to the one described in Struck (2014). Following this approach, clades with high support values are first identified as paralogy suspects. Secondly, those clades with long stems and comprising taxa with independent *a priori* evidence of monophyly together with other taxa are considered paralogs.

Therefore, we inferred neighbor-net networks for each of the different nuclear regions using SPLITSTREE v4.14.6 (Huson and Bryant, 2006). We used the General Time Reversible (GTR) model to estimate the genetic distances with estimated site frequencies and maximum likelihood (equal rates of site variation). Then, we excluded from the alignments all sequences of the resulting networks forming well-supported clades separated by long branches and comprising sequences from different taxa.

Species Delimitation for the Diploid Taxa

Before proceeding with analyses including polyploid samples, we decided to apply coalescent-based species delimitation to the diploid accessions. We have done so because (i) diploid specimens need to be assigned to species (i.e., progenitor lineages) prior to allele co-ancestry analyses with polyploids (see "Gene Trees and Haplotype Network Inference"), (ii) we were not sure if diploid species or their numerous infraspecific taxa (Heywood, 1955, 1975) should be used as progenitor lineages, and (iii) species identification in some of the diploid taxa was considered problematic, especially concerning the diploid *Leucanthemopsis* populations growing in the central and eastern part of the Iberian Peninsula.

We used the BEAST2 (Bouckaert et al., 2019) package STACEY v1.2.1 (Jones, 2017a) to infer species boundaries among the diploid taxa described in *Leucanthemopsis*. STACEY uses a Bayesian approach to infer species delimitation and species phylogeny based on the multispecies coalescent model. It is one of the few model-based programs able to do so, demonstrating to outperform other methods while estimating the correct ultrametric species tree (Andermann et al., 2019) or inferring the best species delimitation scenario (Tomasello, 2018).

We used BEAUTI v2.6 (Bouckaert et al., 2019) to create an input file for STACEY. We included five nuclear loci and two plastid markers. Only diploid ingroup accessions were included in the analyses. During the inference, sequence substitution, clock- and gene-trees models were considered unlinked across loci. Sequence-substitution models were selected *a priori* in MODELTEST-NG (Flouri et al., 2015; Darriba et al., 2020) for each locus separately using the Bayesian Information Criterion (BIC). In the STACEY.xml input file, parameters of the substitution models were fixed to those found in MODELTEST-NG (**Supplementary Table 2**). The strict clock was enforced for all loci. The average clock rate of a random locus was fixed to one, while all other clock rates were estimated in relation to this locus. We set the "Collapse Height" to 1×10^{-4}. This parameter has no biological meaning, and values between 0.000001 and 0.0001 usually produce similar results in similar runtimes (Jones, 2017a). The "Collapse Weight" parameter was estimated using a normal prior distribution with mean = 0.95 and $\sigma = 1.0$. We assigned to the bdcGrowthRate prior a lognormal distribution ($M = 4.6$, $S = 1.5$), a gamma shape ($\alpha = 0.1$ and $\beta = 3.0$) to the popPriorScale prior, and a gamma prior ($\alpha = 1.0$, $\beta = 1.0$) to the relativeDeathRate.

The input files were run for 1×10^8 iterations, sampling every 1×10^4th generation. Two independent runs were performed to check the convergence among independent analyses. Convergence and ESS (values > 200) were checked in TRACER v1.7 (Rambaut et al., 2018). Output files containing the trees sampled in the two independent runs were combined using LOGCOMBINER v2.6 (Bouckaert et al., 2019), after discarding 10% of the sampled iterations as burn-in. The obtained file was processed with the "SpeciesDelimitationAnalyser"[2] (speciesDA.jar; Jones et al., 2015) using a "collapse height" of 1×10^{-4} and setting the similarity cut-off to 0.9. Finally, we produced a similarity matrix using a modified version of the R script provided by Jones et al. (2015).

Gene Trees and Haplotype Network Inference

In *D35*, the alignment region between positions 249 and 294, characterized by a microsatellite motive (see "Species Delimitation For the Diploid Taxa") producing non-informative homoplastic differences among sequences, was excluded from analyses. Indels in all marker alignments were coded as binary characters using the simple gap-coding method of Simmons and Ochoterena (2000) as implemented in the software program GAPCODER (Young and Healy, 2003). Bayesian inference (BI) phylogenetic analyses were performed in MRBAYES v.3.2.7 (Ronquist et al., 2012) for the plastid markers concatenated

[2]https://indriid.com/software.html

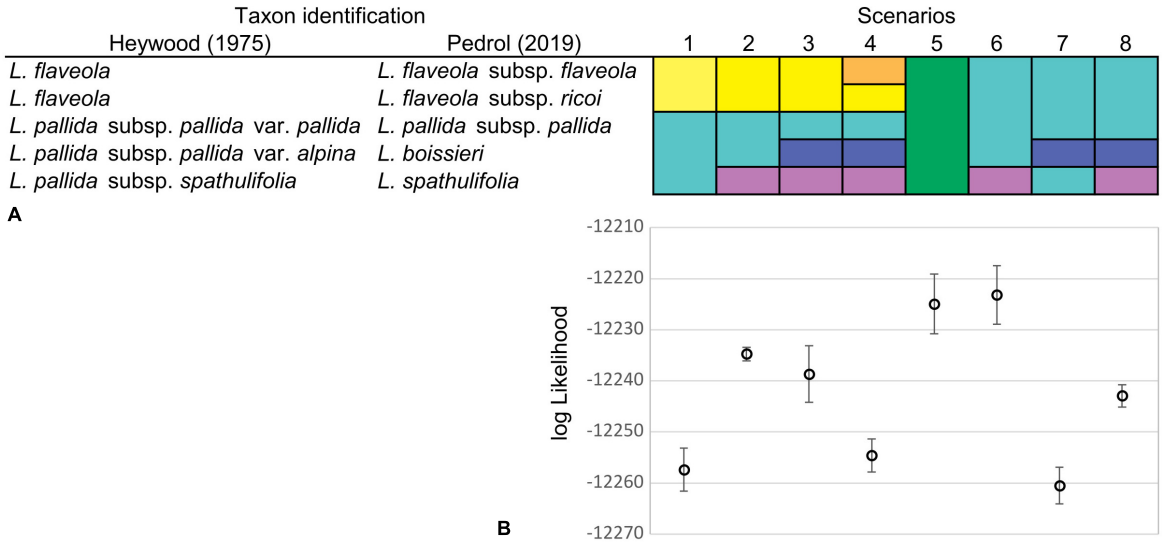


FIGURE 4 | (A) Species delimitation scenarios were investigated for the Iberian tetraploid taxa of *Leucanthemopsis*. Taxa recognized by Heywood (1975) and Pedrol (2019) were merged into different potential tetraploid lineages. Different colors in a scenario denote different lineages delimited (e.g., in scenario one, two species are delimited). **(B)** Means and standard deviations of marginal-likelihood estimates from five replicate analyses for each of the species classification scenarios of the Iberian tetraploid taxa. Marginal likelihoods were estimated via Path Sampling (PS) in the Model Selection package v1.3.4 of BEAST2.

TABLE 1 | Averaged fluorescent ratios (relative DNA content) and inferred ploidy for different taxa of *Leucanthemopsis* as obtained by flow-cytometric measurements.

Taxon	2*n*	*Leucanthemopsis/Petunia* ratio		Ploidy	Individuals	Populations
		Mean	SD			
L. alpina 2*x*	18	3.278	0.149	2*x*	21	4
L. alpina 4*x*	36	5.804	0.384	4*x*	10	2
L. alpina 6*x*	54	7.861	0.190	6*x*	10	2
L. boissieri	36	5.759	0.213	4*x*	10	2
L. cuneata subsp. *cuneata*	54	7.431	0.079	6*x*	5	1
L. cuneata subsp. *valdes-bermejoi*	18	3.420	0.106	2*x*	5	1
L. flaveola subsp. *flaveola*	36	5.847	0.072	4*x*	5	1
L. flaveola subsp. *ricoi*	36	5.781	0.120	4*x*	18	4
L. heywoodii	18*	3.307	0.066	2*x*	10	2
L. pallida subsp. *pallida*	36	6.227	0.265	4*x*	28	5
L. pallida subsp. *virescens*	18	3.477	0.342	2*x*	7	3
L. spathulifolia	36†	6.118	0.173	4*x*	25	5
L. pectinata	18	3.722	0.215	2*x*	10	2
L. pulverulenta	18	3.866	0.143	2*x*	42	7

2n, chromosome numbers are those reported in the literature (refer to ***Supplementary Data Sheet 1****); *Küpfer (1972) reported n = 9 for individuals ascribable to L. heywoodii (chromosome number expected for a diploid), whereas Pedrol (2019) considered the species being tetraploid; †Antunez-Matarrubia (1981) reported a single count of 2n = 38 for L. spathulifolia, all other counts found in the literature were 2n = 36. Detailed information on flow-cytometric measurements is provided in* ***Supplementary Data Sheet 2****.*

in a single alignment, and for each of the five nuclear regions separately. We used MODELTEST-NG and the Bayesian Information Criterion (BIC) to choose the best model of nucleotide substitution for each of the datasets (**Supplementary Table 3**). The analyses were run using two individual runs with three heated and one cold chain each, and a chain heating parameter of 0.2. The Metropolis-coupled Markov Chain Monte Carlo (MC^3) chains were executed for 10,000,000 generations, sampling trees every 1,000th generation. To check convergence between individual runs, we considered the average standard deviation of split frequencies (acceptable when less than 0.01) and compared likelihood values and parameter estimates in TRACER v.1.7. Trees sampled during the Bayesian search were finally used as input to estimate allele co-ancestry in polyploids (see "Allele Co-ancestry in Polyploids").

Additionally, plastid regions were used to reconstruct a haplotype network using the software POPART v.1.7[3] and the TCS network algorithm (Clement et al., 2002). Samples,

[3]http://popart.otago.ac.nz/index.shtml

LPS135-10 (*Prolongoa hispanica*) and LPS180-1 (*L. alpina*) were excluded due to two long deletions in *trnC-petN* introducing ambiguities in the haplotype network reconstruction.

Allele Co-ancestry in Polyploids

To assess allele co-ancestry in polyploid taxa, we used ALLCOPOL v.0.1.2 (Lautenschlager et al., 2020). This program uses allele permutations and infers the number of deep coalescences to assign alleles to diploid sub-genomes, as proposed by Oberprieler et al. (2017). ALLCOPOL uses heuristic approaches to overcome computational constraints encountered when increasing ploidy, number of genes, or samples.

Input for ALLCOPOL was produced by randomly choosing 500 gene trees from those sampled during the Bayesian search (excluding the first 1,000 trees) in each nuclear gene and in the concatenated plastid dataset. Gene trees (3,000 in total) were rooted using *Prolongoa*. During the heuristic search, we used the hill-climbing algorithm, with reinitialization (–u) and 1,000 iterations. We estimated allele co-ancestry for all polyploid taxa used in the different species-delimitation scenarios (see "Species Delimitation in Polyploids"). We ran the analyses twice for each taxon to check if the results were reproducible. In few cases, results were identical in terms of the number of extra lineages but slightly different concerning allele assignment. In these cases (which may indicate autopolyploidy), we calculated the probability of the gene trees under the two allele partition results using the CalcGTProb command (Yu et al., 2012) in PHYLONET (Than et al., 2008). We proceeded then with the allele co-ancestry receiving the best likelihood score.

Species Delimitation in Polyploids

Since unequivocal taxon delimitation among tetraploid taxa of the Iberian Peninsula can be problematic in *Leucanthemopsis* and since we wanted to estimate the number of times (allo)polyploidization took place independently (how many reticulation events happened), we calculated marginal likelihoods for different delimitation scenarios of the Iberian tetraploid taxa via the Path Sampling (PS) method (Baele et al., 2012). The tested classification hypotheses followed Heywood (1975) and Pedrol (2019) both at species and subspecies levels, including classifications treating *L. pallida* subsp. *pallida* and *L. flaveola* (Hoffmanns. & Link) Heywood as a single species, and a model treating all tetraploids as a single specific evolutionary unit. In total, eight different species-delimitation hypotheses were investigated (**Figure 4A**). For each classification model, we performed five independent runs and calculated the means and standard deviations of marginal-likelihood values. The marginal likelihood was estimated from 60 path steps, each run lasting for 5,000,000 generations and using a pre-burn-in of 10,000 generations. For these analyses, we used the MODELSELECTION package v.1.3.4 in BEAST2 applying a 10% burn-in.

Network Reconstruction and Age Estimation

We inferred a multilabel (MUL-)species tree assigning polyploid samples to taxa according to the species delimitation scenario having received the best marginal-likelihood score (see "Species Delimitation in Polyploids"). Subsequently, a species network was constructed by joining the leaves of the MUL-species tree (Lott et al., 2009; Oberprieler et al., 2017).

Input files for the BEAST analysis were prepared in BEAUTI v2.6 using the "*BEAST" template. We unlinked substitution models for all loci. Best substitution models were selected in MODELTEST-NG (see "Gene Trees and Haplotype Network Inference"; **Supplementary Table 3**). We incorporated into the analyses indel information, assigning to gap partitions a Mutation Death model. Clock and tree models of indel partitions were linked to those of the corresponding loci, while being kept unlinked across loci.

To be able to set the best clock model for each gene, we calculated marginal likelihoods via the PS method. The marginal likelihood was estimated from 100 path steps, each run for 1,000,000 generations with a pre-burn-in of 10,000 generations and a 10% burn-in. A difference of more than 3 log-likelihood units was considered as strong evidence for the acceptance of the relaxed (more parameter-rich) model against the strict clock (Kass and Raftery, 1995, Aydin et al., 2014). Since the relaxed clock was preferred in all cases (**Supplementary Table 4**), we used the "LogNormal Relaxed Clock" for all loci, giving to the ucldMean prior a lognormal distribution with an initial value of 0.00166 and 95% probability density ranging over two orders of magnitude (M = 0.013; S = 2.0). We have done so to give an informative prior to clock rates and assuming a standard substitution rate in plants of $5e^{-9}$ (Wolfe et al., 1989) and an average generation time of 3 years in *Leucanthemopsis* (considering that members of the genus are plurennial and not flowering during the first growing season). An exponential prior (mean = 1.0, offset = 0.0) was given to the ucldStdev.

For a better divergence time estimation, we decided to use a secondary calibration point besides the informative prior given to the clock rates (Tomasello et al., 2020). For this, we applied a normally distributed prior to the crown age of *Leucanthemopsis*, ranging between 2.9 and 5.9 Ma (mean = 4.39 Ma, σ = 0.91 Ma), which corresponds to the 95% highest posterior density (HPD) of the age estimated in Tomasello et al. (2015). We used the Calibrated Yule model as prior on the species tree. A broad uniform prior was applied to the birthRate (ranging from 0 to 1,000), whereas a 1/x prior (offset = 0.0) was given to the popMean. Two independent analyses were run for 4×10^8 generations, sampling every 20,000th generation. Effective sample size (ESS) and convergence between independent analyses were checked in TRACER v1.7. Results of the two analyses were merged using LOGCOMBINER v2.6 applying a burn-in of 10%. Finally, the remaining 18,000 trees were used to construct a maximum clade credibility tree with a posterior probability limit set to 0.5 and "Mean Heights" for node heights using TREEANNOTATOR v2.6 (Bouckaert et al., 2019).

A species network was subsequently generated from the MUL-species tree by joining leaves belonging to sub-genomes of the same polyploid lineage. Branches with posterior probability lower than 0.7 were collapsed in the species network. Intervals of 95% HPD of the age of a polyploid taxon were obtained by merging the

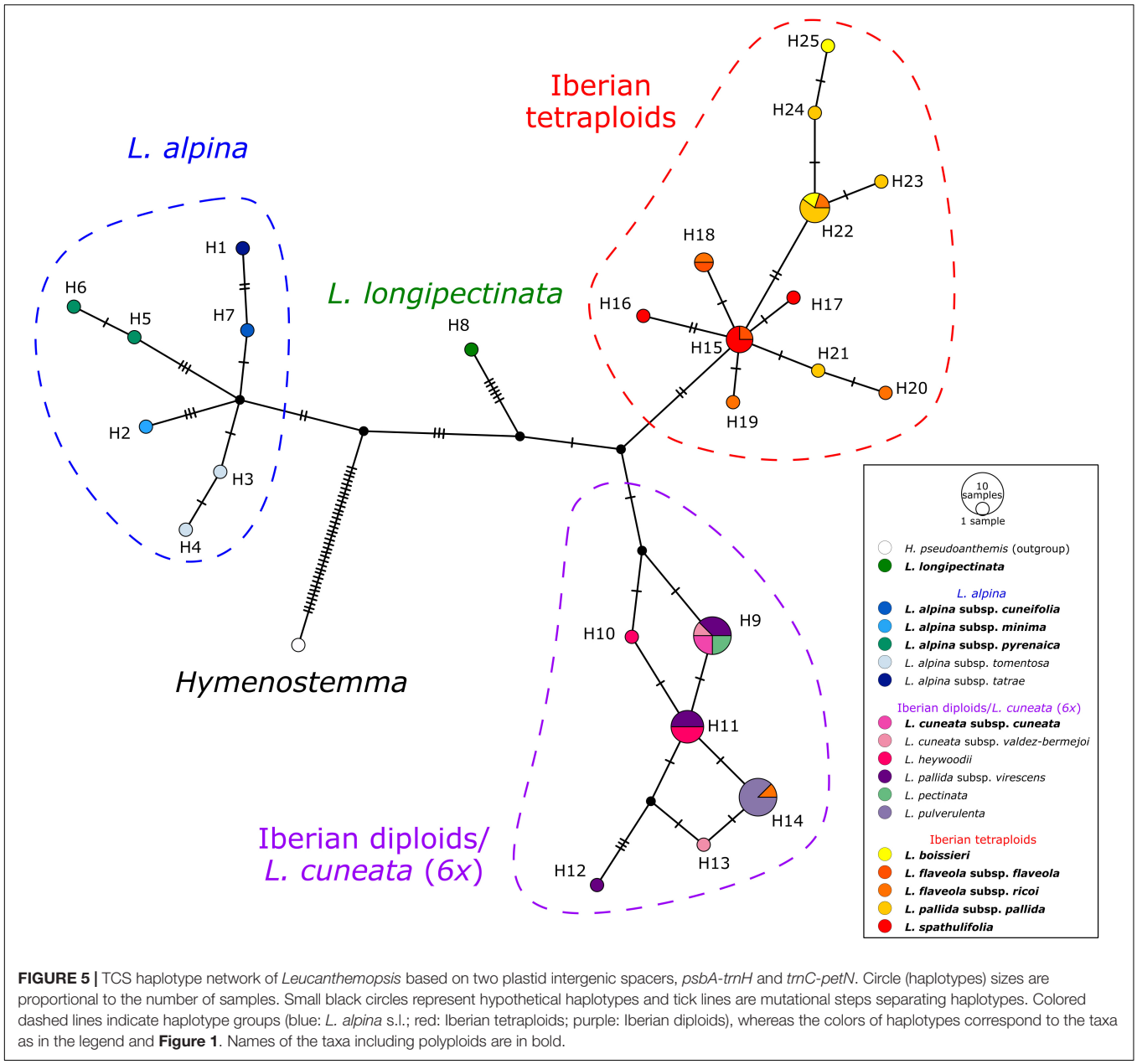


FIGURE 5 | TCS haplotype network of *Leucanthemopsis* based on two plastid intergenic spacers, *psbA-trnH* and *trnC-petN*. Circle (haplotypes) sizes are proportional to the number of samples. Small black circles represent hypothetical haplotypes and tick lines are mutational steps separating haplotypes. Colored dashed lines indicate haplotype groups (blue: *L. alpina* s.l.; red: Iberian tetraploids; purple: Iberian diploids), whereas the colors of haplotypes correspond to the taxa as in the legend and **Figure 1**. Names of the taxa including polyploids are in bold.

95% HPD of the estimated divergence times of the sub-genomes involved in its formation.

RESULTS

Genome Size and Ploidy Level

The quality of the flow-cytometric measurements in most of the cases was high, with average coefficient of variation (CV) values equal to 3.13% (±0.70) and 3.93% (±0.91) for the *Petunia* standard and the samples, respectively. Fluorescence ratios between replicate measurements varied on average by 0.02% (±0.06). Mean relative DNA content (fluorescent ratio between *Leucanthemopsis* and *Petunia* peaks) ranged from 3.27 to 3.86 in diploids, from 5.76 to 6.28 in tetraploids, and from 7.43 to 7.86 in hexaploids. In almost all cases, the ploidy of the measured samples corresponded to the expected one based on literature. **Table 1** reports averaged taxon values for the relative DNA content and ploidy. Detailed information on all measures is available in **Supplementary Data Sheet 2**.

Sequencing and Paralog Filtering

Library preparation and equimolar mixing worked well, and we did not have missing data for any sample, although the number of reads varied considerably across markers and accessions. In total 24,664 reads were obtained, and after quality trimming, about 83% of the reads were retained (20,429 reads; **Supplementary Data Sheet 3**). The number of

alleles per individual was in most of the cases not higher than expected based on the ploidy level of the samples and under the assumption of each nuclear region being single-copy (**Supplementary Data Sheet 3**). After paralog filtering, 11 consensus sequences (alleles) were deleted from *B12*, 21 from *B20*, five from *C16,* and one from *D35*.

Plastid Haplotype Network

The haplotype network obtained from the plastid spacers *psbA-trnH* and *trnC-petN* (**Figure 5**) showed four well-defined haplotype groups. The first group (blue in **Figure 5**) included accessions of *L. alpina*, the second (green) comprised the hexaploid Moroccan *L. longipectinata*. Further, the Iberian taxa were divided into two main haplotype groups, not corresponding to the present taxonomy, but rather to ploidy levels: one (red) was formed by only tetraploid accessions, whereas the other (purple) included mainly diploids. The hexaploid *L. cuneata* subsp. *cuneata* (from Sierra de Urbión, north-eastern Spain), had the same haplotype as the Iberian diploid *L. pectinata* (L.) G.López & C.E.Jarvis, some samples of *L. pallida* subsp. *virescens* (Pau) Heywood, and one of the two samples of *L. cuneata* subsp. *valdes-bermejoi* Pedrol. In the group containing only tetraploid taxa, haplotypes sampled in *L. spathulifolia* (J. Gay) Fern.Casas and a few samples of *L. flaveola* occupied a more proximal position in the network, whereas haplotypes found in *L. pallida* subsp. *pallida* and *L. boissieri* were more derived. Interestingly, one sample of *L. flaveola* had a haplotype otherwise found only in the diploid *L. pulverulenta* (H14 in **Figure 5**).

Species Delimitation in Diploids

"SpeciesDelimitationAnalyser" found a classification with four clusters (species) as the one with the highest frequency. The similarity matrix (**Figure 6**) also shows four distinct clusters, i.e., *L. pectinata*, *L. pulverulenta*, a cluster formed by the diploids of *L. alpina*, and a cluster including samples ascribable to *L. cuneata* subsp. *valdes-bermejoi*, *L. heywoodii* Pedrol, and *L. pallida* subsp. *virescens*. Individuals from one cluster had zero posterior probability of belonging to any other cluster.

Species Delimitation in Polyploids

The species-delimitation scenario producing the best marginal-likelihood scores was scenario 6 (log-likelihood mean = −12,222.28; SD = 4.36; **Figure 4B**), in which tetraploid samples from the Iberian Peninsula were divided into two specific evolutionary units: (1) *L. spathulifolia* and (2) all other tetraploid samples considered being a single species (**Figure 4A**). Scenario 5, in which all tetraploid Iberian taxa were treated as a single species, was the second-best, receiving log-likelihood values partially overlapping with those of Scenario 6 (**Supplementary Table 5**).

Species Network and Divergence Times

The MUL-species tree obtained from the *BEAST analyses showed a clear separation among three clades in *Leucanthemopsis*. The earliest diverging lineage was *L. longipectinata*, which is found to be the sister to a non-supported clade formed by all other taxa (**Supplementary Figure 1**). The other two supported lineages (clades) include members of *L. alpina* (posterior probability PP 0.89) and all Iberian taxa (PP 1.0), respectively. Relationships among taxa/sub-genomes within these two clades were only partially supported (**Supplementary Figure 1**).

The Moroccan hexaploid, *L. longipectinata* was inferred as being autopolyploid, with differentiations among the three sub-genomes having occurred presumably between 4.54 and 0.2 Ma ago. Additionally, tetra- and hexaploid populations of *L. alpina* were reconstructed as autopolyploid, whereas the Iberian hexaploid, *L. cuneata* subsp. *cuneata* was found to be formed by the contributions of (1) an early-branching lineage of the Iberian clade, (2) a lineage related to *L. pectinata* and the other diploids growing along the eastern part of the Iberian Peninsula (i.e., *L. cuneata* subsp. *valdes-bermejoi, L. heywoodii* and *L. pallida* subsp. *virescens*; **Figure 7**), and (iii) a lineage involved also in the formation of the Iberian tetraploids (i.e., *L. boissieri, L. flaveola,* and *L. pallida* subsp. *pallida*). In the origin of the Iberian tetraploids (including all the tetraploid taxa present in the Iberian Peninsula with the exclusion of *L. spathulifolia*) also the diploid *L. pulverulenta* took part. Finally, *L. spathulifolia* was formed by the contribution of a lineage related to *L. pectinata* and the complex including *L. cuneata* subsp. *valdes-bermejoi, L. heywoodii*, and *L. pallida* subsp. *virescens*, and (as in *L. cuneata* subsp. *cuneata*) an early-branching lineage of the Iberian clade. The formation of the Iberian polyploids was inferred to be the period between 2.87 and 0.59 Ma.

DISCUSSION

In the present study, we aimed at reconstructing the evolutionary history of the western Mediterranean genus *Leucanthemopsis* using comprehensive sampling and sequence information from single-copy nuclear genes and plastid markers. Both auto- and allopolyploidy were involved in the formation of polyploid taxa in the genus. In the Iberian Peninsula, allopolyploids originated during the early Pleistocene, probably thanks to latitudinal and elevational range shifts that brought isolated diploid *Leucanthemopsis* lineages into contact. Our study highlights the importance of the Pleistocene climatic oscillations and polyploidizations for the high plant diversity in the western Mediterranean Basin.

Flow-cytometric measurements were in line with previously reported ploidy levels based on chromosome counts except for *L. heywoodii*. We included four samples ascribable to this taxon (LPS192, LPS193, LPS208, and LPS209). Unfortunately, none of them was from the *locus classicus* (Sierra de Javalambre). However, all were from the distribution range of the species given by Pedrol (2019; i.e., Sierras de Gúdar, Albarracìn, Valdemeca, Javalambre). We determined the ploidy level of two populations (LPS208 and LPS209), which were found to be diploid. These two populations were visited and sampled by Juan Pedrol (among other botanists) in the summer of 2016. Additionally, population LPS209 (SALA158943) was identified by its collectors as *L. pulverulenta* subsp. *pseudopulverulenta* (Heywood) Heywood,

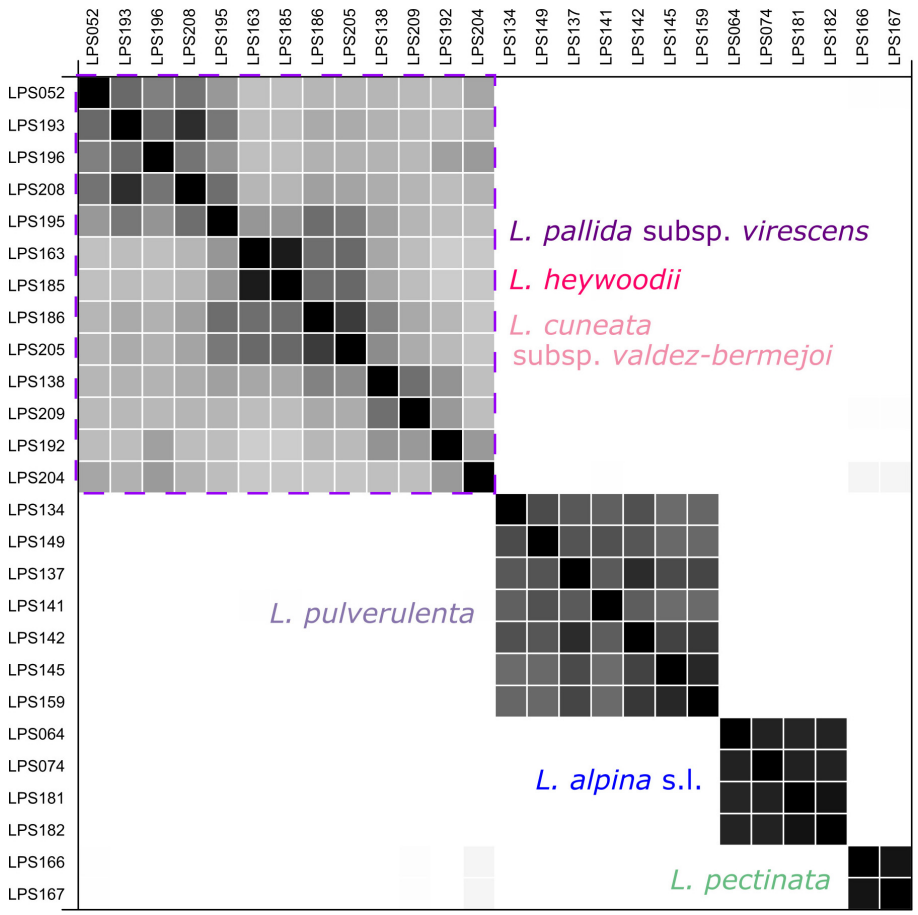


FIGURE 6 | Similarity matrix produced by the coalescent-based species-delimitation analyses (STACEY) for the diploid members of *Leucanthemopsis*. Posterior probabilities for pairs of individuals belonging to the same cluster (species) are shown in black (1.0) and white (0.0). The purple square encloses the cluster including samples of *L. cuneata* subsp. *valdes-bermejoi, L. heywoodii,* and *L. pallida* subsp. *virescens*. The colors correspond to those of **Figure 1**.

a taxon that was subsequently synonymized with *L. heywoodii* by Pedrol (2019). We do not know on which data the tetraploid report provided by Pedrol (2019) is based. To the best of our knowledge, the only chromosome count ever published from a *Leucanthemopsis* accession from the *locus classicus* of *L. heywoodii* was diploid (Küpfer, 1972). Therefore, we included the samples ascribable to this taxon in the species delimitation analyses of the diploid dataset.

Phylogenetic Relationships Among Diploid Members of *Leucanthemopsis*

The results from the STACEY analyses revealed four main genetic clusters (lineages), corresponding to *L. alpina, L. pectinata*, *L. pulverulenta*, and a cluster formed by *L. cuneata* subsp. *valdes-bermejoi*, *L. heywoodii,* and *L. pallida* subsp. *virescens* (**Figure 6**). The analyses did not support any differentiation within diploid *L. alpina*. *L. alpina* subsp. *tomentosa* (Loisel.) Heywood (LPS181 and LPS181 in the present study), endemic to the highest mountain peaks of Corsica, has been treated sometimes as an independent species (Loiseleur-Deslongchamps, 1807; De Candolle, 1837; Holub, 1977; Pignatti, 1979, 1982) due to its peculiar morphology and the isolated distribution range. In accordance with our previous study focused on this species complex (Tomasello and Oberprieler, 2017), however, our present results do not support such taxonomic treatment.

Within the cluster formed by *L. cuneata* subsp. *valdes-bermejoi, L. heywoodii,* and *L. pallida* subsp. *virescens*, a certain level of genetic structuring is visible, but without clear boundaries among the three taxa. The accessions of *L. cuneata* subsp. *valdes-bermejoi* (LPS163 and LPS185) received a high posterior probability of belonging together in the same species, but also show relatively high similarity with other samples of the cluster, especially with the *L. pallida* subsp. *virescens* populations collected in Pico Revolcadores (Murcia; i.e., LPS186 and LPS205). Otherwise, it is not possible to couple the observed phylogenetic patterns within the cluster with neither any morphological (e.g., white or yellow limb of ray florets) or ecological (growing on the siliceous or calcareous substrate) features nor with formerly recognized taxonomic units. As a consequence, this lineage is very heterogeneous as circumscribed in our STACEY analyses, and includes populations with yellow or white ray florets (i.e., the varieties *virescens* and *bilbilitanum* recognized

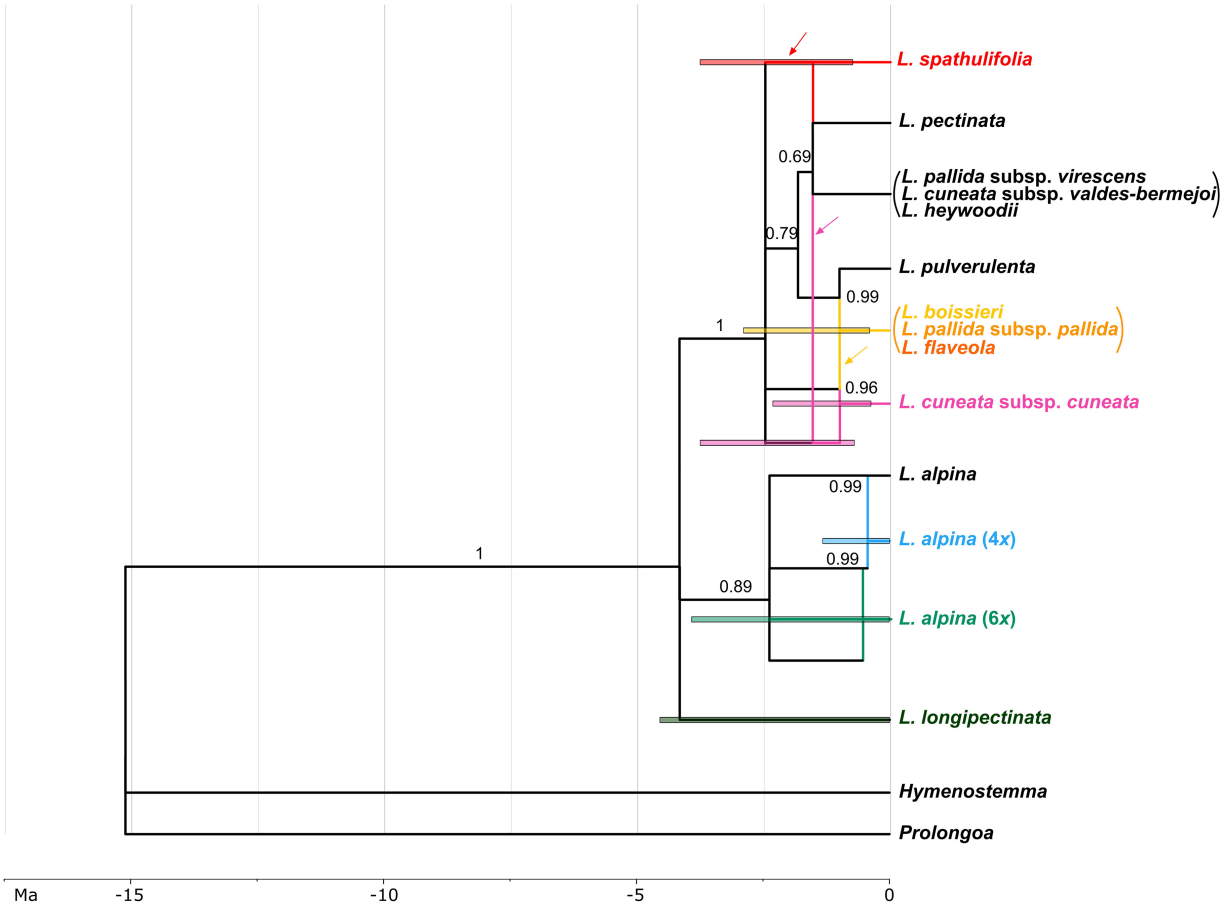


FIGURE 7 | Network representing the phylogenetic relationships and reticulate evolution in *Leucanthemopsis*. Colors for polyploids correspond to those in **Figure 1**. Taxa merged into lineages by the species delimitation analyses are within parentheses (the diploids *L. cuneata* subsp. *valdes-bermejoi*, *L. heywoodii*, and *L. pallida* subsp. *virescens,* and the tetraploid *L. boissieri*, *L. flaveola*, and *L. pallida* subsp. *pallida*, respectively). The numbers above branches are posterior probabilities obtained in the *BEAST analysis. Branches that had posterior probability below 0.69 in the multi-labeled species tree (**Supplementary Figure 1**) were collapsed in the species network. Edges representing the maternal progenitor of a polyploid as inferred in ALLCOPOL are indicated by arrows. Colored bars indicate 95% highest posterior density (HPD) intervals of the age estimates. For polyploids, 95% HPD of age estimates were obtained merging the 95% HPD of the divergence times of the sub-genomes contributing to the polyploidy formation (see "Materials and Methods" for more details). The time scale is expressed in millions of years (Ma).

by Heywood (1975) within *L. pallida* subsp. *virescens*), those growing in subalpine vegetation (*L. heywoodii*) as well as such from much lower altitudes (down to 800 m), mostly found on limestone, but in some cases also on siliceous outcrops [e.g., *L. pallida* subsp. *virescens* var. *bilbilitanum* sensu Heywood (1975)]. There have been attempts to acknowledge a such-circumscribed complex as a separate species (Valdés-Bermejo in Antunez-Matarrubia, 1981; Pérez-Romero et al., 2005). However, those attempts were not generally accepted later-on. Therefore, this complex surely deserves additional studies with a denser sampling and more powerful genetic markers.

In the MUL-species tree (**Supplementary Figure 1**) and in the species network (**Figure 7**), diploids of *L. alpina* are clearly separated from the Iberian diploids. In the clade constituted by Iberian diploids, *L. pulverulenta* is sister to the clade formed by *L. pectinata* and the lineages including all other Iberian diploids (see above). *L. pulverulenta* is a well-defined species, distinct from the other Iberian diploids in both morphological and ecological respects. In fact, this species has smaller leaves, both the basal and the stem ones being deeply pinnatisect, and ray florets always with white limbs. This species is a mesomediterranean element typical of sandy soils of the north-western and central Iberian *meseta* (Ladero and Velasco, 1978), whereas the other taxa are always, although to different extents, linked to mountain environments.

Origin of Polyploids

We used ALLCOPOL (Lautenschlager et al., 2020) to sort alleles to putative parental sub-genomes and infer the origin(s) of polyploids. This allowed us to calculate marginal likelihoods for different taxon assignments and estimate the number of independent polyploidization events. Šlenker et al. (2021) resolved the origin of the allotetraploid *Cardamine barbaraeoides* Halácsy using target enrichment data. Lautenschlager et al. (2020) and Šlenker et al. (2021) sorted alleles of the tetraploids into parental sub-genomes. To do so, the latter used the

position in the nuclear gene trees occupied by the homeologs of polyploids relative to the maternal progenitor. The authors demonstrated the scalability of their method to genomic data and provided the script "AlleleSorting."[4] However, the efficiency of the method needs to be proven (1) when information on the maternal progenitor is missing and (2) for ploidies higher than the tetraploid level. ALLCOPOL does not need the maternal progenitor of a polyploid to be sampled. Moreover, our relatively small dataset (up to the hexaploid level; a maximum 20 samples per polyploid lineage and no more than ten genomic regions) allowed the program to perform the analyses in a reasonable time.

The marginal-likelihood analyses supported the scenario with two independent origins of the Iberian tetraploids, i.e., one leading to *L. spathulifolia* and the other to all other Iberian tetraploids (Scenario 6 in **Figure 4A**). Also other pieces of evidence support a separate origin of *L. spathulifolia*, which possesses the least derived haplotypes in the group of the Iberian tetraploids (**Figure 5**). This taxon has a characteristic karyotype, with a pair of heteromorphic chromosomes, one of which is always smaller than the other (Fernandez-Casas, 1977; Antunez-Matarrubia, 1981), its leaf morphology is very peculiar, and pollen grains are smaller than in the other Iberian tetraploids (Antunez-Matarrubia, 1981). In addition, it grows strictly on limestone bedrock, whereas all other tetraploids are calcifuge (Heywood, 1975; Pedrol, 2019).

Concerning the origin of the Iberian tetraploids, the existence of an extinct diploid that contributed as a maternal progenitor in both tetraploid formations needs to be postulated. In fact, all Iberian tetraploids have a peculiar plastid type, well-distinct from those observed in other members of the genus (**Figure 5**). The same pattern is observable in the network reconstruction (**Figure 7**), although resolution is relatively low within this clade. This putatively extinct diploid must have diverged early from the other diploids in the Iberian clade. *L. spathulifola* was formed (apart from the above-mentioned extinct diploid) with the contribution of an ancestor of *L. pectinata* and the lineage including *L. cuneata* subsp. *valdes-bermejoi*, *L. heywoodii,* and *L. pallida* subsp. *virescens.* The affinity between *L. spathulifolia* and *L. pallida* subsp. *virescens* was highlighted by former botanists and even a possible autopolyploid origin of *L. spathulifolia* from the latter taxon has been hypothesized (Antunez-Matarrubia, 1981). The contribution of an *L. pallida* subsp. *virescens*-like progenitor would explain the morphological (this taxon bears slightly spathulate leaves) and ecological (both calcicole) affinity. Moreover, both taxa are mainly distributed in the eastern Iberian Peninsula, *L. pallida* subsp. *virescens* along the Iberian System and *L. spathulifolia* in the Baetic System. *L. spathulifolia* might have been formed during the past north-south species range shifts, when the diploid maternal progenitor came in contact with an *L. pallida* subsp. *virescens*-like progenitor in the Baetic System. Nevertheless, isolated populations ascribable to the lineage including *L. cuneata* subsp. *valdes-bermejoi, L. heywoodii,* and *L. pallida* subsp. *virescens* are still found not very distant from the species range of *L. spathulifolia* (e.g., LPS186 and LPS205 from Pico Revolcadores, Murcia).

The other tetraploidization event gave rise to *L. boissieri, L. flaveola*, and *L. pallida* subsp. *pallida*. *L. boissieri* was considered to be a variety of *L. pallida* subsp. *pallida* [i.e., *L. pallida* subsp. *pallida* var. *alpina* (Boiss. & Reuter) Heywood] until the most recent taxonomic treatment by Pedrol (2019). It differs from the typical variety by its white ray florets and the somewhat smaller overall habitus. It substitutes the typical variety above 2000 m in the highest peaks of the Central System. *L. flaveola* is considered to be the vicariant of *L. pallida* subsp. *pallida* toward the westernmost part of the Central System, in the mountain ranges between northern Portugal and the highest peaks between Galicia and Leon. The two taxa were often associated with each other and *L. flaveola* has sometimes been considered to be a subspecies of *L. pallida* (Ladero and Velasco, 1978). According to our analyses, *L. pulverulenta* was the other diploid progenitor, which is in line with the suggestion of Mariz (1891) that *L. pulverulenta* is a possible progenitor of *L. flaveola* and *L. pallida* subsp. *pallida*. Despite their different ecology, contact zones between *L. pulverulenta* and *L. pallida* subsp. *pallida* and/or *L. flaveola* are still observable along the Central Systems and, albeit only sporadically, hybridization might be ongoing to the present day (e.g., the *L. flaveola* accession LPS161 shares the haplotype with *L. pulverulenta*; **Figure 5**).

The hexaploid *L. cuneata* subsp. *cuneata*, a narrow endemic of the "Sierra de Urbión" in northern Spain, is reconstructed here as an allohexaploid (**Figure 7**). In this case, the maternal progenitor must have been from the lineage including *L. cuneata* subsp. *valdes-bermejoi*, *L. heywoodii,* and *L. pallida* subsp. *virescens*. It might be *L. cuneata* subsp. *valdes-bermejoi*, but this is difficult to confirm based on the haplotype network (**Figure 5**). It is likely that the maternal progenitor first hybridized with the extinct progenitor of the Iberian tetraploids, and then again with *L. pallida* subsp. *pallida* or *L. flaveola*. This hypothesis and the contribution of the former taxon as one of the progenitors of *L. cuneata* subsp. *cuneata* may receive some support from the classification proposed by Pau (1906), who first described this taxon as a variety of *L. pallida*.

The hexaploid, *L. longipectinata* clearly has an autopolyploid origin. It is the only representative of the genus in North Africa. According to the present temporal reconstructions, it diverged from the rest of the genus about 4 Ma ago. It is probably a descendant of a Tertiary *Leucanthemopsis* diploid confined to the North African mountain ranges, which gave rise to the actual hexaploid as a single extant taxon of this lineage that is limited to the high elevations of the northern Moroccan Rif Mountains. Alternatively, an allopolyploid origin within the African lineage followed by the extinction of all progenitors is also conceivable.

Additionally, polyploid members of *L. alpina* are of autopolyploid origin, too. This species is a polymorphic complex including diploid, tetraploid, and hexaploid populations. Whereas diploids are scattered in the western Alps, Corsica, the Dolomites, the western and the southern Carpathians, and the Dinaric Alps, the tetraploids are widespread across the distribution range of the whole species and the hexaploids are limited to the Central Pyrenees (Tomasello and Oberprieler,

[4]http://github.com/MarekSlenker/AlleleSorting

2017; Tomasello and Konowalik, 2020). As amply elucidated by Tomasello and Oberprieler (2017), the polyploid members have been formed recursively and independently in different places of the species range.

Spatio-Temporal Framework of Polyploidization Events in *Leucanthemopsis*

McCann et al. (2018) estimated the age of formation of allopolyploid taxa of *Melampodium* L. by using BEAST2 and treating sub-genomes of allopolyploids as "species." These were forced to be monophyletic with the diploid progenitors and splits were forced to be contemporaneous using the cross-bracing strategy (Shih and Matzke, 2013). In contrast to that study, we had no *a priori* knowledge of the diploid progenitors of the polyploids. Moreover, one of the diploid progenitors is missing in our dataset (the presumably extinct progenitor of all Iberian tetraploids). As a consequence, we have estimated the age of polyploid formation by joining (*a posteriori*) the 95% HPD of the splits contributing to the different sub-genomes of an allopolyploid. This results in broader and rougher age estimates of polyploid formations.

Based on our results, the formation of polyploids, particularly those in the Iberian clade, took place in the early Pleistocene (between 2.87 and 0.59 Ma; **Figure 7**; **Supplementary Figure 1**). The transition between Pliocene and Pleistocene is well known for the abrupt climate change causing the onset of the major northern hemisphere glaciations at approximately 2.7 Ma (Lisiecki and Raymo, 2005). Also in the Iberian Peninsula, a cyclic climate regime with warm and humid interglacials and cold and dry glacials established during the early Pleistocene (Altolaguirre et al., 2019). The onset of glacial–interglacial oscillations might have produced latitudinal and elevational species-range shifts that brought into contact the once isolated *Leucanthemopsis* lineages. Accordingly, *L. spathulifolia* could have originated during one of the southward shifts of a *L. virescens*-like lineage that met the maternal ancestor of the Iberian tetraploids in the Baetic System. The mountain ranges of eastern Spain (Baetic System, the "Serrania de Cuenca," the Iberian System until Urbión, and the Pyrenees) are well-known to have favored north-south migrational shifts of Oro-Mediterranean plant species (Rivas Martínez, 1969, 1973; Vargas, 2003).

The polyploidization event giving rise to the other Iberian tetraploids could have taken place between the maternal ancestor of all Iberian tetraploids and *L. pulverulenta*, presumably in the central and north-western Iberian Peninsula, where the latter taxon and most of the Iberian tetraploids still co-occur. This allo-polyploidization event must have been more recent, as evidenced by the younger divergence time (1.61–0.38 Ma ago) between *L. pulverulenta* and the sub-genome of the tetraploid derivatives (the second sub-genome in **Supplementary Figure 1**), placing the formation of this polyploid around the "Mid Pleistocene Transition" (∼0.8 Ma ago). If we assume that the maternal progenitor of the Iberian tetraploid was a montane or Oro-Mediterranean species, this could have met *L. pulverulenta* due to downward migrations during cold periods. The Mid-Pleistocene Transition is the time when high-amplitude of 100 ka glaciation cycles settled (Lisiecki and Raymo, 2005). In that period, glaciations became longer and more severe and alternation of cold and warm periods produced major latitudinal and elevational shifts in species-distribution ranges. Those shifts (and the resulting repeated isolation and collision events of diverging lineages) are considered one of the major forces driving plant speciation in southern Europe Quaternary (Tomasello et al., 2020; Kadereit and Abbott, 2021).

DATA AVAILABILITY STATEMENT

Sequences and raw reads presented for the first time in this study can be found under the ENA project PRJEB50763. For GenBank accession numbers of sequences already used in previous studies see **Supplementary Material**.

AUTHOR CONTRIBUTIONS

ST and CO conceived and designed the study and wrote the manuscript. ST collected plant material and performed the laboratory work. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the German Academic Exchange Service (DAAD), by a Ph.D. grant to ST from the Elite Network of Bavaria (Elitenetzwerk Bayern; https://www.elitenetzwerk.bayern.de/start), and by the SYNTHESYS project (http:/www.synthesys.info/) of the European Community (ES-TAF-1615).

ACKNOWLEDGMENTS

We thank the late Peter Hummel and Anja Heuschneider for technical help in the molecular laboratory of CO's research group at the University of Regensburg. The technical help of Emilio Cano Cabezas in the molecular laboratory and of Charo Noya Santos in the herbarium of the Real Jardín Botánico (CSIC) in Madrid is gratefully acknowledged. We also thank Ulrich Lautenschlager for advice concerning AllCoPol analyses. Inès Alvarez and Pablo Vargas (Real Jardín Botánico) are thankfully acknowledged for their support during sampling in the summer 2011.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.842842/full#supplementary-material

REFERENCES

Altolaguirre, Y., Postigo-Mijarra, J. M. A., Barrón, E., Carrión, J. S., Leroy, S. A. G., and Bruch, A. A. (2019). An environmental scenario for the earliest hominins in the Iberian Peninsula: early pleistocene palaeovegetation and palaeoclimate. *Rev. Palaeobot. Palynol.* 260, 51–64. doi: 10.1016/j.revpalbo.2018.10.008

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

Antunez-Matarrubia, A. M. (1981). *Aportación al Conocimiento Taxonómico de las Especies Ibéricas de los Géneros Leucanthemopsis (Giroux) Heywood in Phalacrocarpum willk y Prolongoa boiss.* Master's thesis. Madrid: Universidad Complutense.

Avise, J. C. (1994). *Molecular Markers, Natural History and Evolution*. New York, NY: Chapman and Hall. doi: 10.1007/978-1-4615-2381-9

Aydin, Z., Marcussen, T., Ertekin, A. S., and Oxelman, B. (2014). Marginal likelihood estimate comparisons to obtain optimal species delimitations in *Silene* sect. *Cryptoneurae* (Caryophyllaceae). *PLoS One* 9:e106990. doi: 10.1371/journal.pone.0106990

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., and Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* 29, 2157–2167. doi: 10.1093/molbev/mss084

Bertrand, Y. J. K., Scheen, A.-C., Marcussen, T., Pfeil, B. E., de Sousa, F., and Oxelman, B. (2015). Assignment of homoeologs to parental genomes in allopolyploids for species tree inference, with an example from *Fumaria* (Papaveraceae). *Syst. Biol.* 64, 448–471. doi: 10.1093/sysbio/syv004

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al. (2019). BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650. doi: 10.1371/journal.pcbi.1006650

Chapman, M. A., Chang, J. C., Weisman, D., Kesseli, R. V., and Burke, J. M. (2007). Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). *Theor. Appl. Genet.* 115, 747–755. doi: 10.1007/s00122-007-0605-2

Cheng, L., Connor, T. R., Aanensen, D. M., Spratt, B. G., and Corander, J. (2011). Bayesian semi-supervised classification of bacterial samples using MLST databases. *BMC Bioinformatics* 12:302. doi: 10.1186/1471-2105-12-302

Clement, M., Snell, Q., Walke, P., Posada, D., and Crandall, K. (2002). "TCS: estimating gene genealogies," in *Proceedings 16th International Parallel and Distributed Processing Symposium (IEEE)*, (Lauderdale, FL), 7. doi: 10.1109/IPDPS.2002.1016585

Corander, J., Marttinen, P., and Mäntyniemi, S. (2006). A Bayesian method for identification of stock mixtures from molecular marker data. *Fish. Bull.* 104, 550–558.

Corander, J., Marttinen, P., Sirén, J., and Tang, J. (2008). Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9:539. doi: 10.1186/1471-2105-9-539

Coyne, J. A., and Orr, H. A. (2004). *Speciation*. Sunderland: Sinauer Associates.

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

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., and Flouri, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37, 291–294. doi: 10.1093/molbev/msz189

De Candolle, A. P. (1837). *Prodromus Systematis Naturalis Regni Vegetabilis, Sive, Enumeratio Contracta Ordinum Generum Specierumque Plantarum huc Usque Cognitarium, Juxta Methodi Naturalis, Normas Digesta. Pars Sexta*. Paris: Treuttel et Würtz.

Demesure, B., Sodzi, N., and Petit, R. J. (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* 4, 129–131. doi: 10.1111/j.1365-294x.1995.tb00201.x

Doležel, J., Binarová, P., and Lucretti, S. (1989). Analysis of nuclear DNA content in plant cells by flow cytometry. *Biol. Plant.* 31, 113–120. doi: 10.1007/BF02907241

Doležel, J., and Göhde, W. (1995). Sex determination in dioecious plants *Melandrium album* and *M. rubrum* using high-resolution flow cytometry. *Cytometry* 19, 103–106. doi: 10.1002/cyto.990190203

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaftissue. *Phytochem. Bull.* 19, 11–15.

Fernandez-Casas, J. (1977). Recuntos cromosómicos en plantas vasculares españolas. *Saussurea* 8, 33–55.

Flot, J. F. (2007). CHAMPURU 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. *Mol. Ecol. Notes* 7, 974–977. doi: 10.1111/j.1471-8286.2007.01857.x

Flot, J. F., Tillier, A., Samadi, S., and Tillier, S. (2006). Phase determination from direct sequencing of length-variable DNA regions. *Mol. Ecol. Notes* 6, .627–630. doi: 10.1111/j.1471-8286.2006.01355.x

Flouri, T., Izquierdo-Carrasco, F., Darriba, D., Aberer, A. J., Nguyen, L.-T., Minh, B. Q., et al. (2015). The phylogenetic likelihood library. *Syst. Biol.* 64, 356–362. doi: 10.1093/sysbio/syu084

Fuertes Aguilar, J., Rosselló, J. A., and Nieto Feliner, G. (1999). Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae). *Mol. Ecol.* 8, 1341–1346. doi: 10.1046/j.1365-294X.1999.00690.x

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

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15, 1451–1455. doi: 10.1101/gr.4086505

Goecks, J., Nekrutenko, A., Taylor, J., Afgan, E., Ananda, G., Baker, D., et al. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11:R86. doi: 10.1186/gb-2010-11-8-r86

Grant, V. (1981). *Plant Speciation*, 2nd Edn. New York, NY: Columbia University Press.

Griffin, P. C., Robin, C., and Hoffmann, A. A. (2011). A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. *BMC Biol.* 9:19–36. doi: 10.1186/1741-7007-9-19

Heywood, V. H. (1955). A revision of the Spanish species of *Tanacetum* L. Subsect. *Leucanthemopsis* Giroux. *Ann. Inst. Bot. A. J. Cavanilles* 12, 313–377.

Heywood, V. H. (1975). *Leucanthemopsis* (Giroux) Heywood — A new genus of the Compositae-Anthemideae. *Ann. del Inst. Botánico A. J. Cavanilles* 32, 175–187.

Holub, J. (1977). New names in Phanerogamae. *Folia Geobot. Phytotaxon.* 12, 293–311.

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

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916

Joly, S., Starr, J. R., Lewis, W. H., and Bruneau, A. (2006). Polyploid and hybrid evolution in roses east of the Rocky Mountains. *Am. J. Bot.* 93, 412–425. doi: 10.3732/ajb.93.3.412

Jones, G. (2017a). Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *J. Math. Biol.* 74, 447–467. doi: 10.1007/s00285-016-1034-0

Jones, G. (2017b). Bayesian phylogenetic analysis for diploid and allotetraploid species networks. *BioRxiv* doi: 10.1101/129361

Jones, G., Aydin, Z., and Oxelman, B. (2015). DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics* 31, 991–998. doi: 10.1093/bioinformatics/btu770

Jones, G., Sagitov, S., and Oxelman, B. (2013). Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. *Syst. Biol.* 62, 467–478. doi: 10.1093/sysbio/syt012

Kadereit, J. W., and Abbott, R. J. (2021). Plant speciation in the Quaternary. *Plant Ecol. Divers.* 14, 105–142. doi: 10.1080/17550874.2021.2012849

Karbstein, K., Tomasello, S., Hodač, L., Wagner, N. D., Marinček, P., Barke, B. H., et al. (2022). Untying Gordian knots: unraveling reticulate polyploid plant evolution by genomic data using the large *Ranunculus auricomus* species complex. *New Phytol.* doi: 10.1111/nph.18284

Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795.

Katoh, K., Misawa, K., Kuma, K. I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* 9, 286-98 doi: 10.1093/bib/bbn013

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

Küpfer, P. (1972). Cytotaxonomie et cytogéographie de quelques groupes d'orophytes du basin occidental de la Méditerranée et des Alpes. *C. R. Acad. Sc. Paris* 275, 1753–1756.

Ladero, M., and Velasco, A. (1978). Adiciones a la flora de los Montes de Toledo. *Ann. Inst. Botánico A. J. Cavanilles* 34, 497–519.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. doi: 10.1093/bioinformatics/btu531

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

Lee, C., and Wen, J. (2004). Phylogeny of panax using chloroplast trnC-trnD intergenic region and the utility of trnC-trnD in interspecific studies of plants. *Mol. Phylogenet. Evol.* 31, 894–903. doi: 10.1016/j.ympev.2003.10.009

Linder, C. R., Moret, B. M. E., and Nakhleh, L. (2003). "Network (reticulate) evolution: biology, models, and algorithms," in *Proceedings of The Ninth Pacific Symposium on Biocomputing (PSB)*.

Lisiecki, L. L., and Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* 20, PA1003. doi: 10.1029/2004PA001071

Liveri, E., Crowl, A. A., Mavrodiev, E., Yıldırım, H., Kamari, G., and Cellinese, N. (2020). Another piece of the puzzle, another brick in the wall: the inevitable fate of *Campanula* section *Quinqueloculares* (Campanulaceae: Campanuloideae). *Taxon* 69, 1239–1258. doi: 10.1002/tax.12372

Loiseleur-Deslongchamps, L. A. (1807). *Flora Gallica, seu Enumeratio Plantarum in Gallia Spontè Nascentium*. Paris: Matthaei Migneret.

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

Lott, M., Spillner, A., Huber, K. T., Petri, A., Oxelman, B., and Moulton, V. (2009). Inferring polyploid phylogenies from multiply-labeled gene trees. *BMC Evol. Biol.* 9:216. doi: 10.1186/1471-2148-9-216

Mansion, G., Zeltner, L., and Bretagnolle, F. (2005). Phylogenetic patterns and polyploid evolution within the Mediterranean genus *Centaurium* (Gentianaceae - Chironieae). *Taxon* 54, 931–950. doi: 10.2307/25065479

Marie, D., and Brown, S. C. (1993). A cytometric exercise in plant DNA histograms, with 2C values for 70 species. *Biol. Cell* 78, 41–51. doi: 10.1016/0248-4900(93)90113-S

Mariz, J. (1891). Subsídios para o estudo da flora Portuguesa - Compositae L. *Bol. da Soc. Broteriana* 9, 144–242.

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

McCann, J., Jang, T. S., Macas, J., Schneeweiss, G. M., Matzke, N. J., Novák, P., et al. (2018). Dating the species network: allopolyploidy and repetitive DNA evolution in American daisies (*Melampodium* sect. *Melampodium*, Asteraceae). *Syst. Biol.* 67, 1010–1024. doi: 10.1093/sysbio/syy024

Moreyra, L. D., Márquez, F., Susanna, A., Garcia-Jacas, N., Vázquez, F. M., and López-Pujol, J. (2021). Genesis, evolution, and genetic diversity of the hexaploid, narrow endemic *Centaurea*. *Diversity* 13:72. doi: 10.3390/d13020072

Oberprieler, C., Greiner, R., Konowalik, K., and Vogt, R. (2014). The reticulate evolutionary history of the polyploid NW Iberian *Leucanthemum pluriflorum* clan (Compositae, Anthemideae) as inferred from nrDNA ETS sequence diversity and eco-climatological niche-modelling. *Mol. Phylogenet. Evol.* 70, 478–491. doi: 10.1016/j.ympev.2013.10.013

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

Otero, A., Fernández-Mazuecos, M., and Vargas, P. (2021). Evolution in the model genus *Antirrhinum* based on phylogenomics of topotypic material. *Front. Plant Sci.* 12:631178. doi: 10.3389/fpls.2021.63117

Otto, F. J. (1992). "Preparation and staining of cells for high-resolution DNA analysis," in *Flow Cytometry and Cell Sorting*, (New York, NY: Springer Science & Business Media), 65–68. doi: 10.1007/978-3-662-02785-1_8

Oxelman, B., Brysting, A. K., Jones, G. R., Marcussen, T., Oberprieler, C., and Pfeil, B. E. (2017). Phylogenetics of allopolyploids. *Annu. Rev. Ecol. Evol. Syst.* 48, 543–577. doi: 10.1146/annurev-ecolsys-110316-022729

Pau, C. (1906). Sobre el "*Pyrethrum hispanicum*" de Willkomm. *Butlletí Inst. Catalana d'Història Nat.* 6, 88–93. doi: 10.2436/bichn.vi.8838

Pedrol, J. (2019). "*Leucanthemopsis*(Giroux) Heywood," in *Flora Iberica 16, 1907–1930*, eds S. Castroviejo, C. Aedo, M. Laínz, F. Muñoz Garmendia, G. Nieto Feliner, J. Paiva, et al. (Madrid: Real Jardín Botánico, CSIC).

Pérez-Romero, R., Pérez-Morales, C., Del Río, S., and Penas, Á. (2005). *Leucanthemopsis virescens* (Pau) R.Pérez-Romero, C. Pérez-Morales, S. del Río & Penas. A new combination. *Stud. Bot.* 24, 67–69.

Pignatti, S. (1979). Note critiche sulla Flora d'Italia. VI. Ultimi appunti miscellanei. *Giorn. Bot. Ital.* 113, 359–368.

Pignatti, S. (1982). *Flora d'Italia*. Bologna: Edagricole.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Ramsey, J., and Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* 29, 467–501. doi: 10.1146/annurev.ecolsys.29.1.467

R Development Core Team (2008). *R: A language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ramsey, J., and Schemske, D. W. (2002). Neopolyploidy in plants. *Annu. Rev. Ecol. Syst.* 33, 589–639. doi: 10.1146/annurev.ecolsys.33.010802.150437

Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabath, N., et al. (2019). The global biogeography of polyploid plants. *Nat. Ecol. Evol.* 3, 265–273. doi: 10.1038/s41559-018-0787-9

Rivas Martínez, S. (1969). "La vegetación de alta montaña española," in *V Simposio de Flora Europea (20–30 de Mayo de 1967): Trabajos y Comunicaciones* (Seville: Universidad de Sevilla), 53–80.

Rivas Martínez, S. (1973). Avance sobre una sintesi corologica de la Peninsula Iberica, Baleares y Canarias. *An. del Institudo Bot. A.J. Cavanilles* 30, 69–87.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Rosewich, U. L., and Kistler, H. C. (2000). Role of horizontal gene transfer in the evolution of fungi. *Annu. Rev. Phytopathol.* 38, 325–363. doi: 10.1146/annurev.phyto.38.1.325

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

Sang, T., Crawford, D. J., and Stuessy, T. F. (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am. J. Bot.* 84:1120. doi: 10.2307/2446155

Shih, P. M., and Matzke, N. J. (2013). Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12355-12360 doi: 10.1073/pnas.1305813110

Simmons, M. P., and Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49, 369–381. doi: 10.1093/sysbio/49.2.369

Šlenker, M., Kantor, A., Marhold, K., Schmickl, R., Mandáková, T., Lysak, M. A., et al. (2021). Allele sorting as a novel approach to resolving the origin of allotetraploids using hyb-seq data: a case study of the Balkan Mountain endemic *Cardamine barbaraeoides*. *Front. Plant Sci.* 12:659275. doi: 10.3389/fpls.2021.659275

Soltis, D. A., Soltis, P. S., Schemske, D. W., Hancock, J. F., Thompson, J. N., Husband, B. C., et al. (2007). Autopolyploidy in angiosperms: have we grossly underestimated thenumber of species? *Taxon* 56, 13–30. doi: 10.2307/25065732

Stebbins, G. L. (1950). *Variation and Evolution in Plants*. New York: Columbia University Press.

Struck, T. H. (2014). TreSpEx—detection of misleading signal in phylogenetic reconstructions based on tree information. *Evol. Bioinform.* 10, 51–67. doi: 10.4137/EBo.s14239

Than, C., Ruths, D., and Nakhleh, L. (2008). PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinform.* 9:322. doi: 10.1186/1471-2105-9-322

Tomasello, S. (2018). How many names for a beloved genus? – coalescent-based species delimitation in *Xanthium* L. (Ambrosiinae, Asteraceae). *Mol. Phylogenet. Evol.* 127, 135–145. doi: 10.1016/j.ympev.2018.05.024

Tomasello, S., Álvarez, I., Vargas, P., and Oberprieler, C. (2015). Is the extremely rare Iberian endemic plant species *Castrilanthemum debeauxii* (Compositae, Anthemideae) a "living fossil"? evidence from a multi-locus species tree reconstruction. *Mol. Phylogenet. Evol.* 82, 118–130. doi: 10.1016/j.ympev.2014.09.007

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

Tomasello, S., and Konowalik, K. (2020). On the *Leucanthemopsis alpina* (L.) Heywood growing in the Illyrian region. *PhytoKeys* 161, 27–40. doi: 10.3897/PHYTOKEYS.161.53384

Tomasello, S., and Oberprieler, C. (2017). Frozen ploidies: a phylogeographical analysis of the *Leucanthemopsis alpina* polyploid complex (Asteraceae, Anthemideae). *Bot. J. Linn. Soc.* 183, 211–235. doi: 10.1093/botlinnean/bow009

Van De Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi: 10.1038/nrg.2017.26

Vanneste, K., Maere, S., and Van de Peer, Y. (2014). Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130353. doi: 10.1098/rstb.2013.0353

Vargas, P. (2003). Molecular evidence for multiple diversification patterns of alpine plants in Mediterranean Europe. *Taxon* 52, 463–476.

Wolfe, K. H., Gouy, M., Yang, Y. W., Sharp, P. M., and Li, W. H. (1989). Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 86, 6201–6205. doi: 10.1073/pnas.86.16.6201

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., and Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13875–13879. doi: 10.1073/pnas.0811575106

Yan, Z., Cao, Z., Liu, Y., Ogilvie, H. A., and Nakhleh, L. (2021). Maximum parsimony inference of phylogenetic networks in the presence of polyploid complexes. *Syst. Biol.* 71, 706–720. doi: 10.1093/sysbio/syab081

Young, N. D., and Healy, J. (2003). GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinfo.* 4:6. doi: 10.1186/1471-2105-4-6

Yu, Y., Barnett, R. M., and Nakhleh, L. (2013). Parsimonious inference of hybridization in the presence of incomplete lineage sorting. *Syst. Biol.* 62, 738-51 doi: 10.1093/sysbio/syt037

Yu, Y., Degnan, J. H., and Nakhleh, L. (2012). The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet.* 8:e1002660. doi: 10.1371/journal.pgen.1002660

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fpls.2022.815379

From Western Asia to the Mediterranean Basin: Diversification of the Widespread *Euphorbia nicaeensis* Alliance (Euphorbiaceae)

*Valentina Stojilkovič[1,2], Eliška Záveská[1,3] and Božo Frajman[1]**

[1] Department of Botany, University of Innsbruck, Innsbruck, Austria, [2] Department of Biology, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia, [3] Institute of Botany of the Czech Academy of Sciences, Průhonice, Czechia

The Mediterranean Basin is an important biodiversity hotspot and one of the richest areas in the world in terms of plant diversity. Its flora parallels in several aspects that of the Eurasian steppes and the adjacent Irano-Turanian floristic region. The *Euphorbia nicaeensis* alliance spans this immense area from the western Mediterranean to Central Asia. Using an array of complementary methods, ranging from phylogenomic and phylogenetic data through relative genome size (RGS) estimation to morphometry, we explored relationships and biogeographic connections among taxa of this group. We identified the main evolutionary lineages, which mostly correspond to described taxa. However, despite the use of highly resolving Restriction Site Associated DNA (RAD) sequencing data, relationships among the main lineages remain ambiguous. This is likely due to hybridisation, lineage sorting triggered by rapid range expansion, and polyploidisation. The phylogenomic data identified cryptic diversity in the Mediterranean, which is also correlated with RGS and, partly, also, morphological divergence, rendering the description of a new species necessary. Biogeographic analyses suggest that Western Asia is the source area for the colonisation of the Mediterranean by this plant group and highlight the important contribution of the Irano-Turanian region to the high diversity in the Mediterranean Basin. The diversification of the *E. nicaeensis* alliance in the Mediterranean was triggered by vicariance in isolated Pleistocene refugia, morphological adaptation to divergent ecological conditions, and, to a lesser extent, by polyploidisation.

Keywords: Eurasian steppes, Mediterranean Basin, Irano-Turanian region, morphometry, phylogeny, polyploidy, RAD sequencing, taxonomy

OPEN ACCESS

Edited by:
Lisa Pokorny, Botanical Institute of Barcelona (CSIC), Spain

Reviewed by:
Sara Manafzadeh, ETH Zürich, Switzerland
Mario Fernández-Mazuecos, Autonomous University of Madrid, Spain
Paul Berry, University of Michigan, United States

****Correspondence:***
Božo Frajman
bozo.frajman@uibk.ac.at

Specialty section:
This article was submitted to Understanding Plant Diversity and Evolution in the Mediterranean Basin, a section of the journal Frontiers in Plant Science

Received: *15 November 2021*
Accepted: *05 May 2022*
Published: *23 June 2022*

Citation:
Stojilkovič V, Záveská E and Frajman B (2022) From Western Asia to the Mediterranean Basin: Diversification of the Widespread Euphorbia nicaeensis Alliance (Euphorbiaceae). Front. Plant Sci. 13:815379. doi: 10.3389/fpls.2022.815379

INTRODUCTION

The Mediterranean Basin is an important biodiversity hotspot and one of the richest areas in the world in terms of animal and plant diversity (Myers et al., 2000), harbouring 24,000 plant species, of which 60% are endemic (Greuter, 1991). Its rich biota is a result of a complex palaeogeologic and paleoclimatic history as well as current ecogeographical heterogeneity (Blondel and Aronson, 1999; Thompson, 2005; Blondel et al., 2010; Nieto Feliner, 2014). Because of its geological and climatic complexity, it provides an ideal area for studying biogeography and evolution (Ståhls et al., 2016). The western Mediterranean is geologically older (Krijgsman, 2002); however, the eastern

Mediterranean is considered to be more diverse (Nieto Feliner, 2014), thus acting as a reservoir for evolution of plants and as a cradle for the diversification of different lineages, which is, in particular, true for the Balkan Peninsula (Mansion et al., 2009; Roquet et al., 2009). Despite a high amount of studies dedicated to Mediterranean biodiversity, there are still gaps in our understanding how the Basin has come to be one of Earth's biodiversity hotspots (Nieto Feliner, 2014).

The Irano-Turanian (IT) floristic region, easterly adjacent to the Mediterranean, is another important hotspot of biodiversity, representing the meeting point of western and eastern floras of the Holarctic Kingdom, with many Mediterranean lineages having originated in the IT region (Zohary, 1973; Quézel, 1985; Manafzadeh et al., 2014, 2017). The western IT region is floristically richer than the eastern one, harbouring approximately 27,000 plant species, many of which are endemic (Takhtajan, 1986; Manafzadeh et al., 2017). Especially central Anatolia, a transition zone between Mediterranean and IT floras (Davis, 1971; Takhtajan, 1986), is characterised by high species endemism of presumably relatively recent origin (Davis and Hedge, 1975; Noroozi et al., 2019). In the same line, the easterly adjacent Armenian and Iranian plateaus, known for their heterogeneous flora rich in endemic genera and species (Hedge and Wendelbo, 1978), belong to the most active centres of speciation in the IT region (Boissier, 1867; Takhtajan, 1986; Mahmoudi Shamsabad et al., 2021).

North-easterly adjacent to the Eastern Mediterranean and western parts of the IT region, the second-largest continuous biome on Earth, the Eurasian steppes of the Circumboreal floristic region (Takhtajan, 1986) are spanning from Central and Eastern Europe (Pannonian and Pontic areas) to Central Asia (Lal, 2004; Wesche et al., 2016; Kirschner et al., 2020), where they extend to the IT region. They represent a large fraction of temperate grasslands (Coupland, 1993; Lavrenko and Karamysheva, 1993) and play an important ecological role, representing a complex array of microniches (Wilson et al., 2012) and serving as a global carbon sink (Lal, 2004). These grasslands, which are shaped by strongly seasonal climates (Peart, 2008), share several characteristics with the Mediterranean grasslands, which is reflected in the sharing of multiple plant genera and species across both biomes (Hamasha et al., 2012). In a recent study, Kirschner et al. (2020) have shown that there is a strong phylogeographic break between the azonal European steppes (roughly east of the Carpathians) and the zonal East European-Asian steppes in a number of plant and animal species.

One of the plant groups having the highest diversity in mountainous areas of the Mediterranean Basin and in the IT region, with some species inhabiting the steppes of Eurasia, is *Euphorbia* L. sect. *Pithyusa* (Raf.) Lázaro. With 50 species, this is one of the larger sections of *Euphorbia* subgen. *Esula* Pers. (Riina et al., 2013). Within this section, there is a lineage comprising c. 15 species distributed from Morocco and the Iberian Peninsula in the west to Central Asia in the east and referred to as the *E. barrelieri-nicaeensis-seguieriana* clade by Frajman et al. (2019). The onset of the diversification of this clade, with largely unresolved interspecific relationships, was dated as Pleistocene (Horn et al., 2014; Frajman and Schönswetter, 2017). Whereas Frajman and Schönswetter (2017) and Frajman et al. (2019) explored evolutionary relationships within the *E. barrelieri* Savi and *E. seguieriana* Neck. alliances, little is known about species delimitations and phylogenetic relationships within the *E. nicaeensis* alliance. *Euphorbia nicaeensis* All., described from Nice (France), is according to most recent taxonomic treatments (e.g., Flora Europaea, Radcliffe-Smith and Tutin, 1968; Med-Check list, Greuter et al., 1986) a morphologically variable and widespread species, distributed from Morocco and Iberia in the west to Anatolia and western Russia in the east. Radcliffe-Smith and Tutin (1968) claimed that "a number of, more or less, local populations can be recognised and have often been given a specific rank, but the distinctions between them are of a minor nature and do not seem to be clear-cut, but two subspecies can be recognised," namely southern European (Mediterranean) *E. nicaeensis* subsp. *nicaeensis* and eastern and central European *E. nicaeensis* subsp. *glareosa* (Pall. ex M. Bieb.). In contrast, Greuter et al. (1986) recognised six additional subspecies alongside subsp. *glareosa* and subsp. *nicaeensis*: two from Italy [*E. nicaeensis* subsp. *japygica* (Ten.) Arcangeli and subsp. *prostrata* (Fiori) Arrigoni] and four from Eastern Europe [*E. n.* subsp. *cadrilateri* (Prodan) Kuzmanov, subsp. *dobrogensis* (Prodan) Kuzmanov, subsp. *goldei* (Proh.) Greuter and Burdet and subsp. *stepposa* (Zoz) Greuter and Burdet]. Such a treatment with a broadly circumscribed *E. nicaeensis*, including a varying number of subspecies, was also implemented in most national floras and taxonomic treatments (e.g., Kuzmanov, 1979; Radcliffe-Smith, 1982; Micevski, 1998; Govaerts et al., 2000). In contrast, some, mostly earlier, authors separated the Mediterranean *E. nicaeensis* from the central/eastern European species, and different names have been applied for the latter, e.g., *E. glareosa* Pall. ex M. Bieb. (Beck-Mannagetta, 1920; Janković and Nikolić, 1972) or *E. pannonica* Host. (Hegi, 1966). Some authors (e.g., Prokhanov, 1949; Geltman, 2009, 2020) recognised even several species that were in the above-mentioned treatments mostly considered synonyms or subspecies: *E. glareosa*, *E. goldei* Prokh., *E. stepposa* Zoz, and *E. volgensis* Krysth. Some additional species, e.g., *E. erythrodon* Boiss. and Heldr. and *E. petrophila* C. A. Mey. from the Pontic and IT regions, were also included in the *E. nicaeensis* alliance by Radcliffe-Smith (1982) and Geltman (2020).

Phylogenetic studies of Internal Transcribed Spacer (ITS) sequences, including a limited number of samples, resolved the Mediterranean *E. nicaeensis* separated from the eastern European and Asian populations treated under *E. glareosa*, *E. pannonica*, *E. petrophila*, and *E. stepposa*. These species either formed a grade of accessions successively sister to the *E. nicaeensis* clade, or a poorly supported sister clade to the *E. nicaeensis* clade, which also included *E. hercegovina* Beck (Frajman and Schönswetter, 2011, 2017; Riina et al., 2013; Frajman et al., 2019). *Euphorbia hercegovina* is endemic to the western Balkan Peninsula between Montenegro and Bosnia and Herzegovina and was, in the past, considered closely related or conspecific with *E. barrelieri* Savi (e.g., Poldini, 1969; Greuter et al., 1986; Govaerts et al., 2000; Trinajstić, 2007; Geltman, 2009); however, Frajman and Schönswetter (2017) showed that it does not belong to the *E. barrelieri* group. In the most recent study

by Frajman et al. (2019), *E. macroclada* Boiss. (one individual sampled) was included in the same clade as *E. hercegovina* and *E. nicaeensis* in the nuclear dataset; plastid sequences, however, positioned *E. macroclada* close to the western Asian *E. cheiradenia* Boiss. (Riina et al., 2013).

Euphorbia macroclada is morphologically similar to *E. nicaeensis* and was also included in *E.* ser. *Nicaeenses* Prokh. by Prokhanov (1949). It is distributed predominately in Anatolia (Turkey) but extends its range to adjacent Syria, Iraq, Iran, and Armenia (Radcliffe-Smith, 1982). In a Restriction Site Associated DNA Sequencing (RAD) phylogeny based on limited sampling (Frajman et al., 2019), one population of *E. macroclada* was inferred as sister to three populations of *E. nicaeensis* (*E. hercegovina* was not included in the study), and one population of *E. glareosa* was sister to them. Thus, relationships between and within the *E. nicaeensis* and *E. glareosa* lineages still remain unresolved, and it is not clear how the evolutionary relationships are reflected in morphological differentiation and species delimitations or in their spatial distribution. In addition, it remains unknown what role the Mediterranean Basin, the IT region, and the Eurasian steppes of the Circumboreal region played in diversification of this widespread group. Finally, even if polyploidy seems to be rare in *E.* subgen. *Esula* (Frajman and Schönswetter, 2011; Riina et al., 2013), recent studies, including a dense geographic sampling, have revealed polyploidisation events in several groups of *Euphorbia* (e.g., Stevanoski et al., 2020; Caković et al., 2021). The diploid chromosome number ($2n = 18$) appears to be the most common within the *E. nicaeensis* alliance, but there is also a report of $2n = 56$ for *E. nicaeensis* in Spain and a recent report of a tetraploid ($2n = 36$) population of *E. macroclada* from Turkey (Rice et al., 2015; Genç and Kültür, 2020), indicating that polyploidisation might be one of the processes driving diversification within our study group.

The aim of our study is to disentangle phylogenetic relationships within the *E. nicaeensis* alliance and to determine its position within the *E. barrelieri-nicaeensis-seguieriana* clade. After identifying clear evolutionary lineages using next-generation RADseq as well as nuclear ribosomal ITS sequences, we explored biogeographic relationships within the *E. nicaeensis* alliance, specifically amongst *E. hercegovina*, *E. macroclada*, and *E. nicaeensis* (hereafter, the *E. nicaeensis* lineage) and their relation to *E. erythrodon*, *E. glareosa* s.l. (including *E. pannonica* and *E. stepposa*), and *E. petrophila* (hereafter, the *E. glareosa* lineage). We explored which geographic region served as a source for the Mediterranean taxa and how their diversity is distributed within this biome. Given the highest taxonomic diversity in the western IT region, we hypothesise that this region was the source for the Mediterranean taxa of the *E. nicaeensis* alliance, as evidenced in many other groups (Manafzadeh et al., 2017), and that Pleistocene glaciations had an important role in driving diversification of this lineage, in line with previous evidence (Nieto Feliner, 2014; Caković et al., 2021). This was the case also within the *E. verrucosa* L. group (Caković et al., 2021), which strongly overlaps with *E. nicaeensis*, both in distribution and habitats. In addition, we intersected the phylogenetic data with relative genome size (RGS) data and assessed the incidence of polyploidisation within our study group. Finally, by also intersecting morphometric data with phylogenetic patterns, we are able to discuss reasons for observed incongruences. By identifying cryptic diversity within the Mediterranean, we propose a revised taxonomic treatment, including the description of a new species and revised distribution data for the studied taxa.

MATERIALS AND METHODS

Plant Material

Plant material for molecular analyses and RGS estimation (silica-gel dried leaf material), as well as for morphometric analyses (herbarium vouchers), was collected in the field between 2006 and 2019. We additionally sequenced ITS from 18 herbarium specimens from the herbaria MA (16), WU (1) and the private herbarium of W. Gutermann (1); a herbarium specimen of the outgroup *E. humilis* Ledeb. from FRU was used for RADseq. In addition, 32 herbarium specimens from herbaria M (10) and W (22) were used for morphometric analyses. In total, 145 populations of *E. erythrodon* (one population), *E. glareosa* s.l. (29 populations), *E. hercegovina* (13 populations), *E. macroclada* (34 populations), *E. nicaeensis* (66 populations; including population 58 of *E. nicaeensis* subsp. *japygica*, which we name *E. japygica* hereafter for simplicity), and *E. petrophila* (2 populations), as well as 30 populations of 14 outgroup taxa were studied (**Figure 1**, **Supplementary Figure 1**, and **Supplementary Table 1**). For the focal taxa *E. hercegovina*, *E. macroclada*, and *E. nicaeensis*, our sampling covers their complete distributions. In the case of *E. hercegovina* and *E. nicaeensis*, material from areas corresponding to the type localities was included in most analyses, including RADseq (populations 77 and 40, respectively), whereas, in the case of *E. macroclada*, the population 81 is ca. 250 km away from the type locality. In the case of *E. glareosa* s.l., which includes different taxa as outlined in the introduction, we studied several samples likely belonging to different taxa (species or subspecies), but it was beyond the aims of this study to include type populations of all these taxa. The same was the case for *E. erythrodon* and *E. petrophila*.

Restriction Site Associated DNA Sequencing Library Preparation

Total genomic DNA was extracted from dried leaf tissue (ca. 10 mg) using the CTAB protocol described by Frajman and Schönswetter (2011) and purified with the NucleoSpin gDNA clean-Up kit (Macherey-Nagel, Düren, Germany). Single-digest RAD libraries were prepared using the restriction enzyme PstI (New England Biolabs, Ipswich, MA, United States) and a protocol adopted from Paun et al. (2015). Briefly, we started with 110 ng DNA per individual and ligated 100 mM P1 adapters to the restricted samples overnight at 16°C. Shearing by sonication was performed with an M220 Focused-ultrasonicator (Covaris) with settings targeting a size range of 200 to 800 bp and peak at 400 bp. A total of 115 individuals (65 individuals from 32 ingroup populations and 52 individuals from 26 populations of various

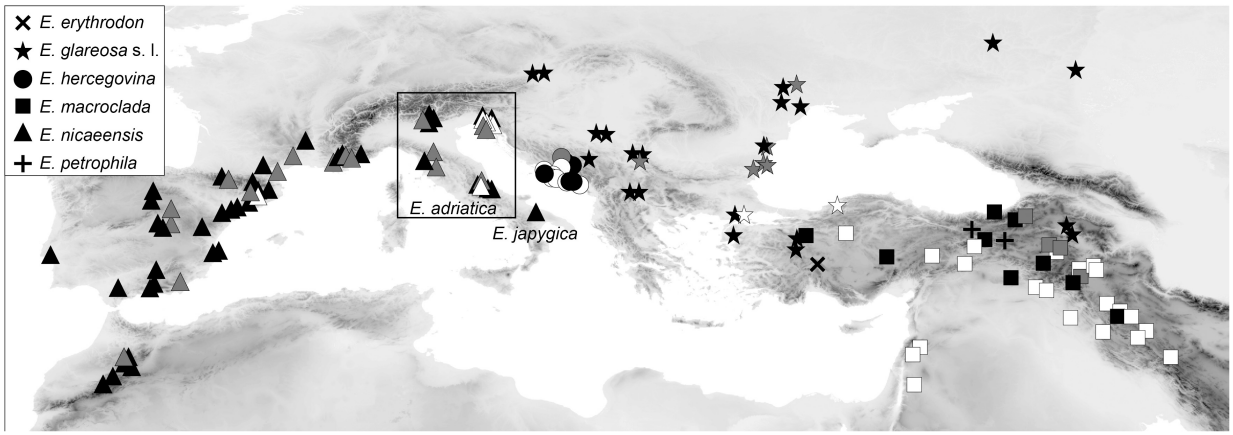


FIGURE 1 | Distribution of *Euphorbia nicaeensis* alliance populations sampled and used in this study. Black symbols indicate populations used in genetic (RADseq and/or ITS) and, mostly, also, in relative genome size (RGS) and morphometric analyses. Grey symbols indicate additional populations used in RGS and/or morphometric analyses; white symbols indicate populations used only in morphometric analyses. The corresponding population numbers are presented in **Supplementary Figure 1**, and details are given in **Supplementary Table 1**. *Euphorbia adriatica* and *E. japygica* have been previously included in *E. nicaeensis* but are, based on our results, independent species.

outgroup taxa from *E.* sect. *Pithyusa*; see **Supplementary Table 1** for details) were sequenced on an Illumina HiSeq 2500 at CSF Vienna[1] as 100 bp single reads using Illumina chemistry.

Identification of Restriction Site Associated DNA Sequencing Loci and Single Nucleotide Polymorphism Calling

The raw reads were quality filtered and demultiplexed according to individual barcodes using the Picard command line tool BamIndexDecoder[2] and the program "process_radtags.pl" implemented in Stacks v2.3d (Catchen et al., 2011, 2013). The RAD loci were further assembled, and Single Nucleotide Polymorphism (SNPs) were called using the "denovo_map.pl" pipeline in Stacks. A dataset used for subsequent phylogenetic reconstruction was built using a minimum × 5 coverage to identify a stack (-m 5), a maximum number of five differences between two stacks per locus per sample (-M 5), and a maximum number of five differences among loci to be considered as orthologous across multiple samples (-n 5). The parameter values (-m, -M, -n) were previously optimised for *E. seguieriana* and other species, here used as the outgroup, as described in Kirschner et al. (2020) and similarly applied by Frajman et al. (2019). The program "populations" in the Stacks package was used: (i) to identify samples with missing data > 28%, to exclude them from further analyses; (ii) to pre-filter loci by setting the maximum observed heterozygosity to 0.65 to process a nucleotide site at a locus (and prevent processing of paralogs); and (iii) to export an SNP dataset in vcf format. Starting with a vcf file, we further filtered out sites with depth of coverage < × 10 and sites with > 50% missing samples, using VCFtools v0.1.15 (Danecek et al., 2011). The filtered vcf file was converted to phylip format using the python script vcf2phylip (Ortiz, 2019) and used for downstream phylogenetic reconstructions.

Phylogenetic Analyses, Species Tree Inference, and Ancestral Area Estimation Based on the Complete RADseq Dataset

To infer phylogenetic relationships among all 115 individuals, we computed a maximum likelihood (ML) phylogeny using RAxML v8.2.8 (Stamatakis, 2014). Invariant sites were removed from the original phylip format using the script "deleteAlignColumn.pl"[3], and Felsenstein's ascertainment bias correction was further used to account for missing invariant sites as recommended by Leaché et al. (2015). Tree searches were done under a GTR model with categorical optimisation of substitution rates (ASC_GTRCAT; Stamatakis, 2014). *Euphorbia cheiradenia*, *E. kopetdaghi* (Prokh.) Prokh., *E. matritensis* Boiss., *E. minuta* Loscos and Pardo, and *E. polycaula* Boiss. and Hohen. were used for rooting based on Riina et al. (2013) and their placement in a preliminary NeighborNet constructed using SplitsTree4 v12.3 (Huson and Bryant, 2006). The best-scoring ML tree was bootstrapped using the frequency-based stopping criterion (Pattengale et al., 2010).

To infer a species tree and estimate divergence times, we applied SNAPP v1.5.1 (Bryant et al., 2012) implemented in BEAST v2.6.4 (Bouckaert et al., 2014) as described by Stange et al. (2018). We used a reduced data set containing 48 individuals that represented the 19 main lineages identified by RAxML, mainly corresponding to species; *E. nicaeensis* excepted, for which two lineages were resolved and treated independently in species tree inference. Two populations, corresponding to *E. erythrodon* 80 and *E. glareosa* 125, which Structure analyses showed to be introgressed (see section "Results"), were excluded from species tree inference. We constructed a new RAD data subset for these 19 lineages using the filtering parameters

[1]http://csf.ac.at/ngs/

[2]https://github.com/wtsi-npg/illumina2bam

[3]https://www.biostars.org/p/55555/

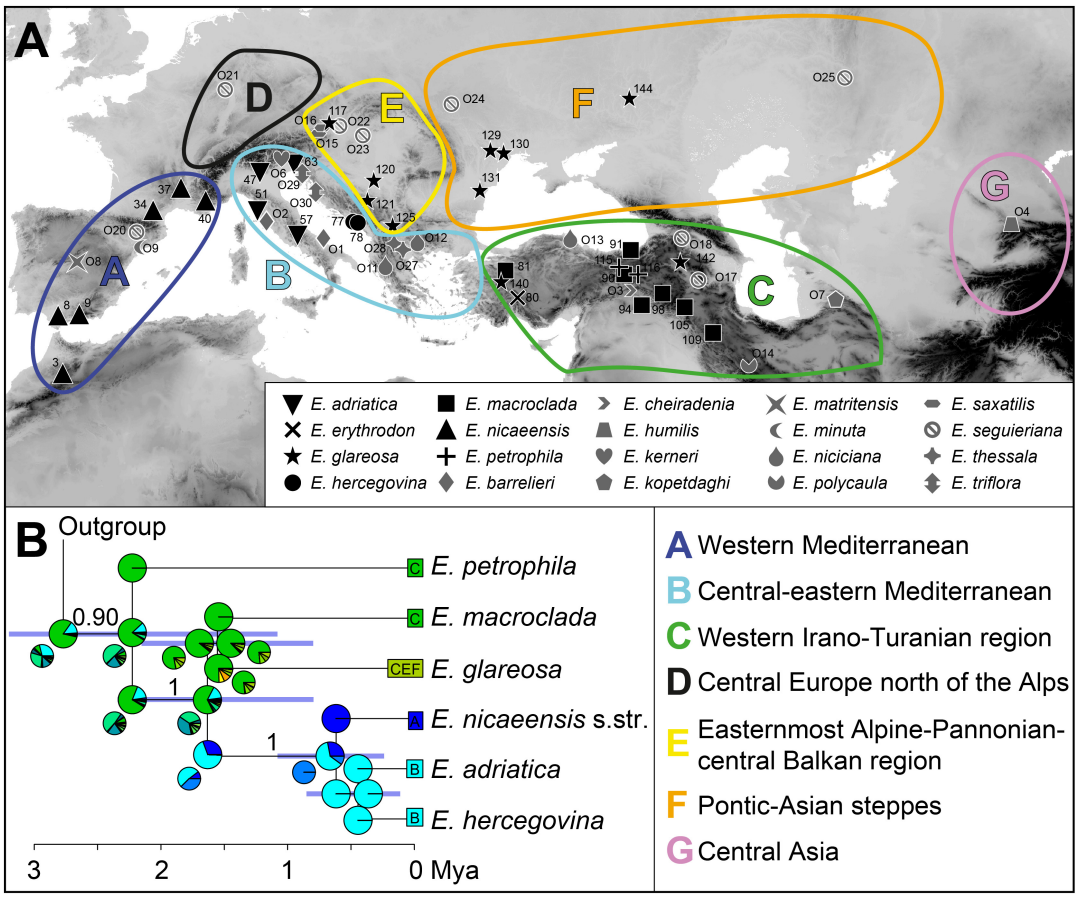


FIGURE 2 | (A) Geographic areas and geographic provenance of the investigated populations. **(B)** The time-calibrated species tree inferred with SNAPP and used in the biogeographic analysis with BioGeoBEARS. Numbers above branches are posterior probabilities > 0.80, and the horizontal bars correspond to 95% highest posterior densities (HPD) for the age estimates. Pie charts at each node show the marginal probabilities of alternative ancestral ranges obtained from the BioGeoBEARS analysis under the DEC + J model. In addition, smaller pie charts resulting from the DEC model are presented in cases where the reconstruction between the models differed. Colours in pie charts correspond to the geographic areas in A. The trees, including outgroup taxa, are presented in **Supplementary Figure 3**. Population numbers correspond to **Supplementary Table 1**.

described above, but requiring loci to be shared among all 19 lineages. We randomly selected one SNP per locus and further filtered for biallelic SNPs only, achieving 1,498 SNPs in total. To scale the tree, we used a secondary calibration by setting the prior age of the root to 4.25 Mya with a normally distributed SD of 1.7, which corresponds to the median age and 95% highest posterior density (HPD) interval of the split between *E. kopetdaghii* and *E. nicaeensis* in Horn et al. (2014). To improve mixing and convergence of the model, we constrained the monophyly of both ingroup and outgroup, as they were consistently resolved as monophyletic with strong support in our RAxML analysis. We ran two independent MCMCs for 100,000 generations, discarding 30% of the generations as burn-in. Log files were analysed using TRACER v1.6 to assess convergence and ensure that the effective sample size (ESS) for all parameters was > 200 (Rambaut et al., 2014). We estimated the maximum clade credibility (MCC) tree from the posterior distributions of both runs using TreeAnnotator v2.5.0 (Drummond and Rambaut, 2007).

We performed a biogeographic analysis with BioGeoBEARS (Matzke, 2013), using the MCC tree inferred with SNAPP, and defined seven geographic areas (**Figure 2A**). We combined the knowledge accrued on floristic patterns to designate geographic areas [e.g., that of Takhtajan (1986), largely adopted by Manafzadeh et al. (2017), for the IT region], while also considering more recent phylogeographic studies (e.g., Kirschner et al., 2020). The boundaries of the areas were also adapted to our phylogenetic results, i.e., the borders were drawn in the areas for which we identified main phylogenetic breaks. The Mediterranean region *sensu* Takhtajan (1986) was split into western Mediterranean (A) and central-eastern Mediterranean (B), following the main split within *E. nicaeensis* and distribution patterns of some co-occuring western Mediterranean species outlined by Caković et al. (2021). The third area corresponds to the western IT floristic region (C) *sensu* Manafzadeh et al. (2017), including Anatolia, the Caucasus, and the Armenian and Iranian plateaus, which largely corresponds to the distribution of *E. macroclada*, but also to that of *E. petrophila* and *E. erythrodon*.

It remains unclear whether the populations of *E. petrophila* from Krym (Area F) and Lesbos, in the Aegean area (Area B), are in the same lineage as Anatolian *E. petrophila*; therefore, we mapped this species only to the area C. We also included the *E. niciciana* O13 population south of the Black Sea, in Region C, as this species is mainly found in Anatolia and the here sampled population grouped with other samples from Anatolia in the study by Frajman et al. (2019). Region D (central Europe north of the Alps) included only a single population of *E. seguieriana*, which was in a clearly divergent clade in the study of Frajman et al. (2019). In contrast, the samples from the easternmost Alpine-Pannonian-central Balkan area were included in its own region E. The latter region includes the azonal steppes of Takhtajan's (1986) Circumboreal floristic region, since they were clearly divergent from the Pontic-Asian steppes (hereafter, Area F), for a number of plant and animal taxa, in Kirschner et al. (2020). The final area, central Asia (G), includes a single population of *E. humilis*, species that is isolated from all other members of the *E. barrelieri-nicaeensis-seguieriana* clade. We used six biogeographic models in a common likelihood framework: a likelihood version of Dispersal-Vicariance analysis (DIVALIKE; Ronquist, 1997), LAGRANGE Dispersal and Extinction Cladogenesis (DEC model, Ree et al., 2005; Ree and Smith, 2008), a likelihood version of BayArea (Landis et al., 2013), and an alternative version for each of the models that include founder-event speciation (+J). It was, however, claimed by Ree and Sanmartin (2018) that statistical comparisons of likelihoods between DEC and DEC+J are inappropriate (but see Matzke, 2013[4]). The maximum number of areas for each node was set to four, which is the maximum number of areas occupied by the extant taxa (Ronquist, 1996; Hilpold et al., 2014). Each terminal node in the tree was coded with the total distribution area of the taxon/lineage. We defined a dispersal probability matrix to determine the effect of geographic distance on dispersal ability. The rate of dispersal between adjacent areas was set to 1 (e.g., between the western and the eastern Mediterranean), between non-adjacent but geographically close areas to 0.75 (e.g., between the western Mediterranean and northern Europe), between more distant areas to 0.5 (e.g., between the western Mediterranean and Anatolia), and between the most distant areas to 0.25 (e.g., between the western Mediterranean and Central Asia), following Hilpold et al. (2014). After running the six models, we compared the results with a likelihood ratio test, applying the Akaike Information Criterion to select the best fit model.

Genetic Structure Within the *Euphorbia nicaeensis* Alliance Based on Restriction Site Associated DNA Sequencing Data

To explore in more detail the phylogenetic relationships and potential presence of gene flow within the *E. nicaeensis* alliance, we analysed a subset of nine main lineages, inferred from the complete dataset with RAxML, by using Bayesian clustering and SNAPP. The nine lineages corresponded to the monophyletic species *E. erythrodon*, *E. hercegovina*, and *E. macroclada*, and, for *E. glareosa*, *E. nicaeensis*, and *E. petrophila*, which appeared poly- or paraphyletic in the RAxML tree, two lineages were delimited in each case. This additional analysis was performed because, in the analysis of the complete dataset, where several distantly related lineages (e.g., that of *E. polycaula*) were included, the dataset included only 1,498 SNPs due to larger amounts of missing data across different lineages and filtering for no missing data needed for SNAPP analysis. When analysing only closely related species of the *E. nicaeensis* alliance, and including one less-distant outgroup (*E. triflora*), the amount of SNPs after filtering was much higher, and 3,000 unlinked SNPs were then used for the SNAPP analysis, resulting in better resolved relationships (i.e., better support in terms of PP values).

For the Bayesian clustering analysis, a set of RADseq loci was exported into Structure format using the –structure flag, while sampling loci shared by at least 10% of populations and 40% of individuals within those populations (using -p and -r flags). One SNP per locus (–write-single-snp flag) was selected to minimise the chance of selecting linked loci, resulting in 15,978 SNPs in total. For 65 individuals representing nine lineages of the *E. nicaeensis* alliance, the optimal grouping of populations was determined using fastSTRUCTURE v1.0 (Raj et al., 2014). The analysis was performed for K (number of groups), ranging from 1 to 10, with the script *structure.py*, using a simple prior. The optimal number of K was determined using the script *chooseK.py*; both scripts are part of the fastSTRUCTURE package.

The same dataset, but including *E. triflora* as the outgroup and further filtered for absence of missing data, was used as input for SNAPP analysis, resulting in 3,000 unlinked SNPs. The analysis was performed as described above for the entire dataset, but applying to the split between *E. triflora* and the ingroup a secondary calibration, set to 3.3 Mya with a normally distributed SD of 0.9, derived from estimating divergence times for the entire dataset. A topological constraint was applied to the monophyletic groups inferred by RAxML analysis for the entire dataset as outlined above.

Internal Transcribed Spacer Sequencing and Analyses of Sequence Data

Internal Transcribed Spacer sequencing, contig assembly and editing, and sequence alignment were performed as described by Frajman and Schönswetter (2011), with modifications described by Cresti et al. (2019). We sequenced ITS for one individual per population from 58 ingroup populations and included 12 ingroup and seven outgroup GenBank sequences selected based on Riina et al. (2013) and Frajman et al. (2019) (**Supplementary Table 1**). The final ITS alignment thus consisted of one sequence of *E. erythrodon*, 17 of *E. glareosa* s.l., six of *E. hercegovina*, one of *E. japygica*, ten of *E. macroclada*, 34 of *E. nicaeensis*, and one of *E. petrophila*, resulting in a much larger dataset with denser geographic sampling of the focal taxa, compared to the RADseq analysis. GenBank numbers are given in **Supplementary Table 1**.

Maximum parsimony (MP) and MP bootstrap (MPB) analyses were performed using PAUP v4.0b10 (Swofford, 2002) as described by Frajman et al. (2019). Bayesian analyses were

[4]https://osf.io/vqm7r/

performed using MrBayes 3.2.1 (Ronquist et al., 2012), applying the HKY+Γ substitution model proposed by the Akaike information criterion (AIC) implemented in MrAIC.pl v1.4 (Nylander, 2004) and the settings as in Frajman et al. (2019). The posterior probabilities (PP) of the phylogeny and its branches were determined from the combined set of trees, discarding the first 1,001 trees of each run as burn-in. TRACER v1.6 was used to assess convergence and ensure that the ESS for all parameters was > 200 (Rambaut et al., 2014). In addition, a NeighborNet was produced with ITS sequences of *E. erythrodon*, *E. glareosa* s.l., *E. hercegovina*, *E. japygica*, *E. macroclada*, and *E. nicaeensis*, which formed a poorly supported clade in the ITS tree, using SplitsTreev4.12.3 (Huson and Bryant, 2006).

Relative Genome Size Estimation

Relative genome size was measured using flow cytometry (FCM) as described by Suda and Trávníček (2006). Nuclei of silica gel dried material of *E. erythrodon* (one population), *E. glareosa* s.l. (19 populations), *E. hercegovina* (four populations), *E. japygica* (one population), *E. macroclada* (11 populations), *E. nicaeensis* (38 populations), and *E. petrophila* (two populations), as well as fresh leaves of a reference standard (*Bellis perennis* L., 2C = 3.38 pg; Schönswetter et al., 2007), were stained using 4',6-diamidino-2-phenylindole (DAPI). If the peaks of the reference standard and the sample overlapped, *Pisum sativum* cv. Kleine Rheinländerin (2C = 8.84 pg; Greilhuber and Ebert, 1994) was used instead. The RGS was estimated for one to five individuals per population (see **Supplementary Table 1**). A CyFlow space flow cytometer (Partec, GmbH, Münster, Germany) was used to record the relative fluorescence of 3,000 nuclei and FloMax software (Partec) to evaluate histograms and to calculate coefficients of variation (CVs) of both the standard and the sample peaks. The RGS was calculated as the ratio between the values of the mean relative fluorescence of the sample and the standard. For statistical analyses of RGS data, RStudio 1.2.5019 (RStudio Team, 2019, version R-3.6.1), with the visualisation package "ggplot2," was used. Scatter and box plots were produced for individual samples as well as for species and ploidy levels.

Morphometric Analyses

We performed morphometric analyses of 16 individuals of *E. glareosa* s.l., 20 *of E. hercegovina,* 32 of *E. macroclada*, and 45 individuals of *E. nicaeensis*. As we were interested in the differentiation among the species and not in the inter-population variability, we analysed one individual per population. The exception was narrowly distributed *E. hercegovina*, for which three to five individuals per population from three populations were analysed (**Supplementary Table 1**). Our sampling of the two other focal taxa, *E. macroclada* and *E. nicaeensis*, roughly corresponds to their respective range's size, whereas we only included a limited number of specimens of *E. glareosa* s.l. for comparison.

After initial inspection of 44 morphological characters, we measured or scored 33 characters that showed variation in the investigated taxa and calculated 15 ratios (**Table 1**). Stem characters were measured manually, whereas the leaf

TABLE 1 | Characters studied in the morphometric analyses of *Euphorbia glareosa, E. hercegovina, E. macroclada*, and *E. nicaeensis*.

No.	Stem
1	Stem length, cm
2	Stem width, cm
3	Stem glabrous/pubescent
	Pleiochasium
4	Number of terminar rays
5	Length of (the longest) terminal ray, cm
6	Number of branching of (the longest) terminal ray
	Axillary rays
7	Number of fertile axillary rays
8	Length of (the longest) fertile axillary ray, cm
	Middle stem leaf
9	Length of a middle stem leaf, cm
10	Width of a middle stem leaf, cm
11	Ratio Length of a middle stem leaf / Width of a middle stem leaf
12	Distance from the base to the widest part of a middle stem leaf, cm
13	Ratio Distance from the base to the widest part of a middle stem leaf / Length of a middle stem leaf
	Ray leaves
14	Length of a ray leaf, cm
15	Width of a ray leaf, cm
16	Ratio Length of a ray leaf/Width of a ray leaf
17	Distance from the base to the widest part of a ray leaf, cm
18	Ratio Distance from the base to the widest part of a ray leaf / Length of a ray leaf
	Raylet leaves
19	Length of a raylet leaf, cm
20	Width of a raylet leaf, cm
21	Ratio Length of a raylet leaf/Width of a raylet leaf
22	Distance from the base to the widest part of a raylet leaf, cm
23	Ratio Distance from the base to the widest part of a raylet leaf / Length of a raylet leaf
	Cyathium
24	Length of cyathial involucre, mm
25	Width of cyathial involucre, mm
26	Ratio Length of cyathial involucre/Width of cyathial involucre
27	Depth of gland emargination, mm
28	Length of cyathial gland, mm
29	Width of cyathial gland, mm
30	Ratio Depth of gland emargination/Length of cyathial gland
31	Ratio Length of cyathial gland/Width of cyathial gland
	Fruit
32	Fruit length, mm
33	Fruit width, mm
34	Ratio Fruit length/Fruit width
35	Distance from the base to the widest part of the fruit, mm
36	Ratio Distance from the base to the widest part of the fruit/Fruit length
37	Style length, mm
38	Fruit glabrous/pubescent/glandular
	Seed
39	Seed length, mm
40	Seed width, mm
41	Ratio Seed length/Seed width
42	Distance from the base to the widest part of a seed, mm
43	Ratio Distance from the base to the widest part of a seed/Seed length
44	Caruncle length, mm
45	Caruncle width, mm
46	Ratio Caruncle length/Caruncle width
47	Distance from the base to the widest part of caruncle, mm
48	Ratio Distance from the base to the widest part of caruncle/Caruncle length

characters were measured partly on scanned herbarium images using ImageJ (Abràmoff et al., 2004) or manually on actual herbarium specimens. All other characters (cyathium, fruit, and seed characters) were measured on images taken with a stereomicroscope Olympus SZX9 using the Olympus image analysis software analySIS pro. Fruits and seeds were developed only in a limited number of specimens. For *E. glareosa* s.l., we measured eleven fruits and nine seeds; for *E. macroclada*, 17 fruits and ten seeds; for *E. nicaeensis*, 18 fruits and 15 seeds, all from different populations, whereas, for *E. hercegovina*, we measured four fruits and four seeds from three populations. In addition, since not all characters were scorable in all individuals, we replaced the missing values in the final data matrix with the species' mean or mode, the latter in the following characters: number of terminal rays, number of branching of (the longest) terminal ray, and number of fertile axillary rays. Statistical analyses were performed using SPSS 24.0. Correlation among metric characters was tested employing Pearson and Spearman correlation coefficients, and one character from each character pair, yielding a correlation coefficient > 0.90, was excluded from further analyses. Box plot diagrams were produced for all characters in order to visualise and show the variation among four species. After standardisation to zero mean and one unit variance, principal component analysis (PCA) was performed. Subsequently, discriminant analysis (DA) was performed. The PCA and DA analyses were performed separately for (1) vegetative parts of the plants and cyathium characters, for (2) fruit, as well as (3) seed characters, in two steps: for *E. glareosa* s.l., *E. hercegovina*, *E. macroclada*, and *E. nicaeensis* and for closely related *E. hercegovina* and two phylogenetic lineages of *E. nicaeensis* (see section "Results"). Based on the morphometric data, we produced species descriptions and an identification key. Metric values presented there correspond to the 10th and 90th percentiles, supplemented by extreme values in parentheses.

RESULTS

Phylogenetic and Biogeographic Analyses Based on the Complete RADseq Dataset

The average number of raw single reads per sample retained after quality filtering was ca. 0.74 million (standard deviation, $SD = 0.3$). The RADseq data are available in the NCBI Short Read Archive (SRA; BioProject ID PRJNA761526, BioSample accessions SRR15817339-SRR15817453). The relationships inferred with RAxML, based on 18,059 SNPs (5,731 variable loci), reflected both the taxonomic entities (**Figure 3A** and **Supplementary Figure 2**) and the geographic structure (**Figure 3B**). The *E. barrelieri-nicaeensis-seguieriana* clade was monophyletic (bootstrap support, BS, 99%). The *E. barrelieri* group, including *E. barrelieri*, *E. kerneri* Huter, *E. saxatilis* Jacq., *E. thessala* (Formánek) Degen and Dörfl. and *E. triflora* Schott, Nyman and Kotschy (**Figure 3D**; BS, 100%), was sister to a clade (BS 99%), consisting of our study group taxa (BS 99%) and a sister lineage (BS 82%), consisting of *E. humilis* (BS, 100%), as well as *E. niciciana* Borbás and *E. seguieriana* Neck. (*E. seguieriana* group; BS, 100%; **Figure 3C**).

Within our study group *E. petrophila* 116 (**Figure 3E**) was sister to all other accessions (BS, 66%), which formed three lineages: the first corresponded to *E. petrophila* 115, the second comprised most accessions of *E. glareosa*s.l. (BS, 62%), and the third included all other accessions in a clade with BS of 100%. In this latter clade, *E. erythrodon* 80 and *E. glareosa* 125 were consecutive sisters to a clade with all other accessions (BS, 68%), consisting of *E. macroclada* (BS, 100%) and *E. nicaeensis*, including *E. hercegovina* (BS, 100%). Within the main clades, corresponding to *E. glareosa* s.l., *E. macroclada* and *E. nicaeensis* (including *E. hercegovina*), relationships mostly reflected geography (**Figure 3B**), notable exceptions being geographically distant populations 131 and 142 of *E. glareosa*s.l. that were grouped together in a clade with the BS of 100%, and the population 105 of *E. macroclada*, which was geographically amongst the phylogenetically more divergent populations. Populations of *E. nicaeensis* were included in two clades, of which the western Mediterranean clade (BS, 93%) was sister to a clade (BS, 72%), including *E. hercegovina* (BS, 83%) and the Apennine and northwesternmost Balkan populations of *E. nicaeensis* (BS, 91%). To increase readability, from here on, as well as in the figures and tables, we refer to the western Mediterranean populations as *E. nicaeensis*, to the Apennine-Balkan populations as *E. adriatica*, and to both together as *E. nicaeensis* s.l.

The time-calibrated species tree of the entire dataset (**Supplementary Figure 3** and **Figure 2B**) inferred similar main relationships as RAxML, but with certain differences. The *E. barrelieri* group (PP 1) included *E. barrelieri*, *E. kerneri*, *E. thessala*, and *E. triflora* (PP 1), but not *E. saxatilis*, which was resolved as sister (0.96 PP) to the *E. seguieriana* group (PP, 1). The *E. nicaeensis* alliance was weakly supported as monophyletic (PP, 0.90). Relationships among monophyletic (PP, 1) *E. adriatica*, *E. hercegovina*, and *E. nicaeensis* were unresolved, as was their relationship to *E. glareosa* and *E. macroclada*. They all formed a clade (PP, 1), a sister to *E. petrophila*. Relationships amongst the three main groups of the *E. barrelieri-nicaeensis-seguieriana* clade (including *E. humilis*) were poorly supported; the onset of their diversification dated back to 3.3 Ma (95% HPD, 1.8–4.8 Ma). The *E. nicaeensis* alliance putatively originated 2.8 Ma (95% HPD, 1.6–3.8 Ma); its diversification might have started in the early Pleistocene, 2.2 Ma (95% HPD, 1.1–3.2 Ma), and continued throughout the Pleistocene until *E. adriatica*, *E. hercegovina*, and *E. nicaeensis* might have diverged in the late Pleistocene, 0.4–0.6 Ma (95% HPD, 0.1–1.1 Ma).

As a result of comparison amongst six biogeographic models using BioGeoBears, the DEC + J model was selected as the best fitting. The best model not including the parameter +J (which was questioned by Ree and Sanmartin, 2018) was DEC. Detailed comparison of the models is given in **Supplementary Table 3**. Ancestral-area estimation with BioGeoBears (**Supplementary Figure 3** and **Figure 2B**) indicated a high uncertainty in the geographic origin of the entire *E. barrelieri-nicaeensis-seguieriana* clade but indicated with high probability that the MRCA of the *E. nicaeensis* alliance was distributed either throughout

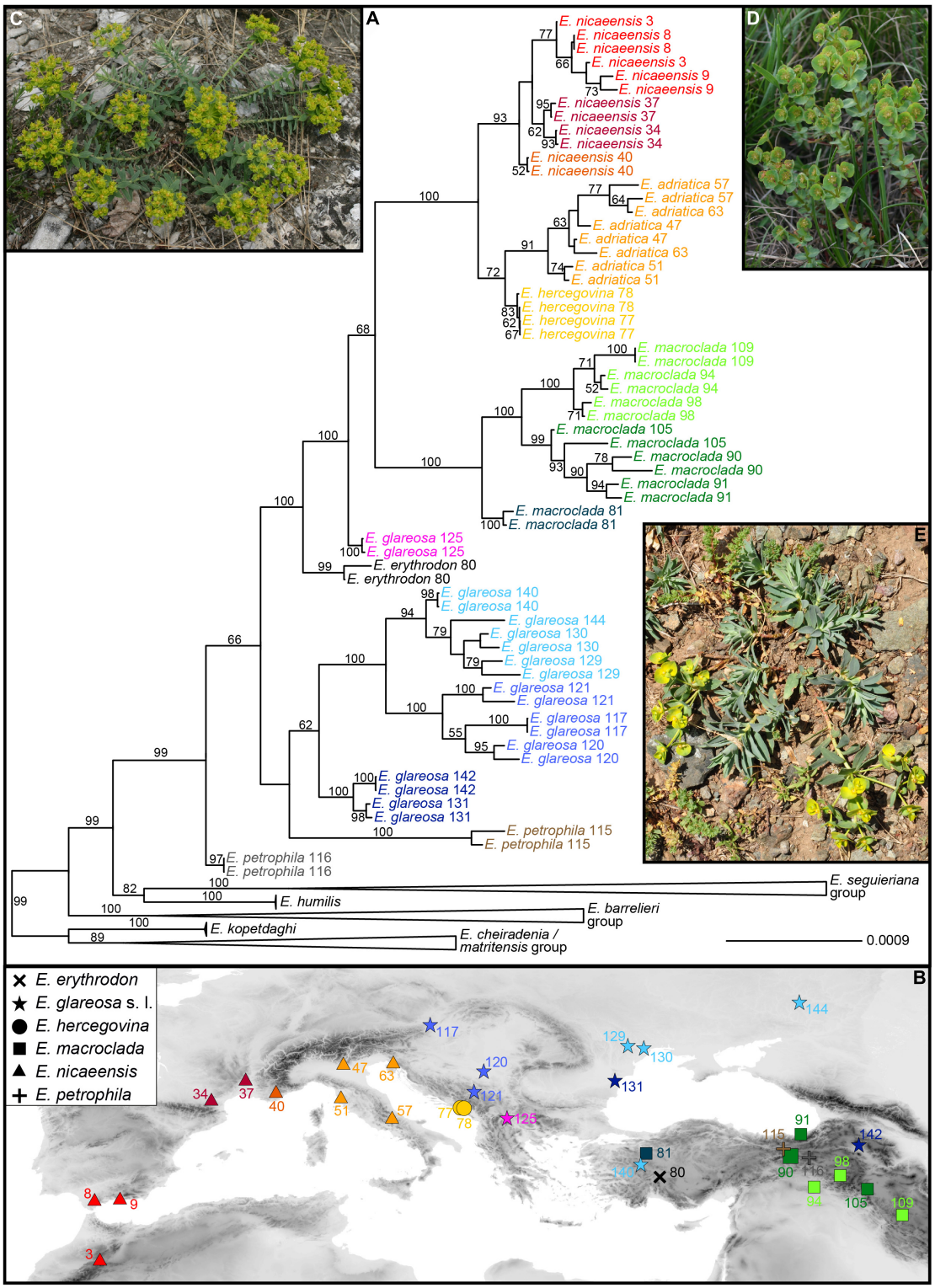


FIGURE 3 | (A) Phylogenetic relationships within the *Euphorbia nicaeensis* alliance and between this alliance and its closest relatives within *E.* sect. *Pithyusa*, as inferred by maximum likelihood analysis of RADseq loci, with indicated bootstrap values above 50%. **(B)** Geographic provenance of the investigated populations with colour coding as in **(A)**. Two outgroup species, *E. seguieriana* (from Austria; **C**) and *E. triflora* (from Slovenia; **D**) from the *E. seguieriana* and the *E. barrelieri* groups, and the ingroup *E. petrophila* (from Turkey; **E**) are shown *in situ*. Photos: B. Frajman **(C,D)**, C. Gilly **(E)**. Population numbers correspond to **Supplementary Table 1**. The tree, including outgroup taxa, is presented in **Supplementary Figure 2**.

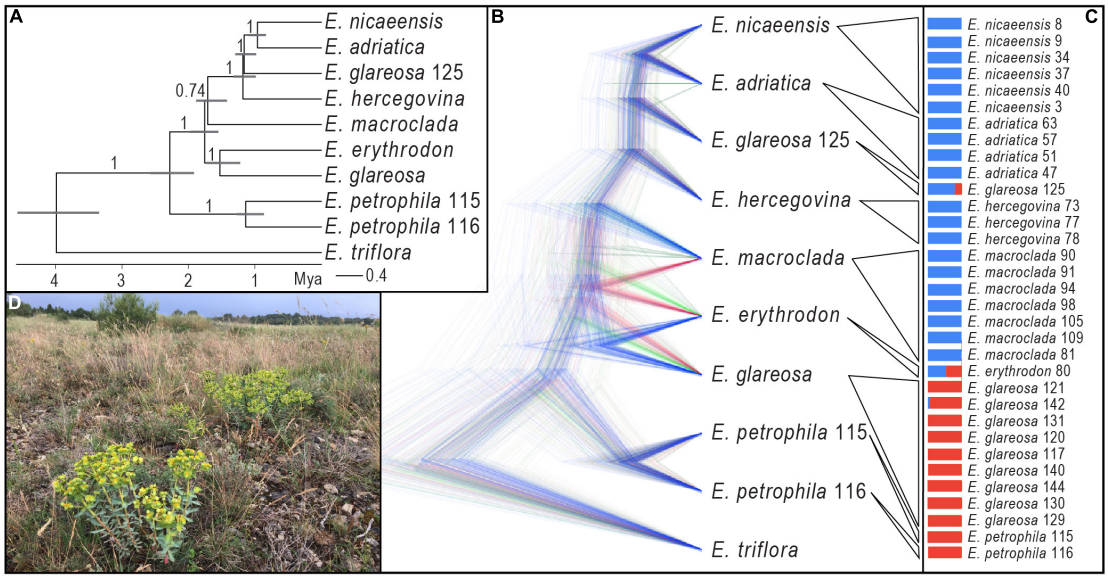


FIGURE 4 | Phylogenetic relationships within the *Euphorbia nicaeensis* alliance based on RADseq data. **(A)** The time-calibrated species tree inferred with SNAPP. Numbers above branches are posterior probabilities, and the horizontal bars correspond to 95% highest posterior densities (HPD) of the age estimates. **(B)** Alternative topologies visualised with DensiTree and represented by different colours. **(C)** Division of all populations into two groups (blue and red) with Bayesian clustering using fastSTRUCTURE. **(D)** *Euphorbia nicaeensis* in its natural habitat northwest of Carcassonne in France (Photo: B. Frajman). Population numbering corresponds to **Supplementary Table 1**.

Anatolia (area C; DEC + J) or Anatolia and the eastern Mediterranean (areas B and C; DEC). While *E. macroclada* and *E. petrophila* remained in Anatolia and adjacent regions of the IT, *E. glareosa* s.l. extended its distribution to the Pannonian and Pontic steppic areas, and the MRCA of the *E. nicaeensis* group dispersed further west into the Mediterranean Basin, where it diverged, giving rise to the western Mediterranean *E. nicaeensis* and the eastern Mediterranean *E. adriatica* and *E. hercegovina*.

Phylogenetic Analyses Based on the Reduced RADseq Dataset for the *Euphorbia nicaeensis* Alliance

The species tree of the *E. nicaeensis* alliance with *E. triflora* as the outgroup based on 3,000 SNPs (**Figure 4A**) revealed a slightly different topology compared to the analyses of the complete dataset. The monophyly of the *E. nicaeensis* alliance was well supported (1 PP), and the onset of its diversification was estimated at 2.3 Ma (95% HPD, 1.9–2.5 Ma). Contrary to the analyses of the complete dataset with RAxML, *E. petrophila* was monophyletic (PP, 1) and was a sister to all other ingroup taxa. *Euphorbia erythrodon* was supported as a sister to *E. glareosa* (PP1; the *E. glareosa* group), and both were a sister (PP 1) to a poorly supported clade (PP, 0.74; the *E. nicaeensis* group), comprising all other taxa and population 125 of *E. glareosa*. The split between the *E. glareosa* group and the *E. nicaeensis* group was dated at 1.8 Ma (95% HPD, 1.6–2 Ma). Within the *E. nicaeensis* group, *E. macroclada* was a sister to all other species that formed a monophyletic group (PP, 1), with the onset of the diversification dated at 1.2 Ma (95% HPD, 1–1.3 Ma). In the latter group, *E. nicaeensis* (**Figure 4D**) was a sister to *E. adriatica* (PP, 1), both were a sister to *E. glareosa* 125 (PP, 1), and all three were sisters to *E. hercegovina* (PP, 1).

Bayesian clustering (**Figure 4C**) split the *E. nicaeensis* alliance into two clusters, one including *E. adriatica*, *E. hercegovina*, *E. macroclada*, and *E. nicaeensis*, and the other one, *E. petrophila*, and most populations of *E. glareosa*. Two populations were admixed between both clusters, namely, *E. erythrodon* 80 and *E. glareosa* 125. A conflicting position of *E. erythrodon*, between *E. glareosa* and *E. macroclada*, was also suggested by the occurrence of alternative topologies in the SNAPP trees, as visualised with DensiTree (**Figure 4B**); otherwise, the topology corresponded to the MCC tree (**Figure 4A**).

Internal Transcribed Spacer Phylogenies

Of 715 characters included, 33 (4.6%) were parsimony informative; consistency and retention indices were 0.87 and 0.96, respectively. The trees inferred by parsimony and Bayesian analyses (**Figure 5A**) were largely congruent, with the exception of relationships within the ingroup, which remained unresolved with parsimony. In general, the ingroup (excluding *E. petrophila*) was poorly supported as monophyletic (PP, 0.78; MPB, 52%) and included two unsupported clades in the Bayesian tree, the first corresponding to *E. glareosa*s 0.l (PP, 0.55) and the second to the *E. nicaeensis* clade (PP, 0.86). This second clade included a polytomy, including *E. adriatica, E. hercegovina*, *E. japygica*, *E. nicaeensis*, and an accession of *E. glareosa* 125 from North Macedonia, as well as a clade (PP, 0.96), with *E. macroclada* and the only accession of *E. erythrodon*. *Euphorbia petrophila* was positioned amongst the outgroup taxa.

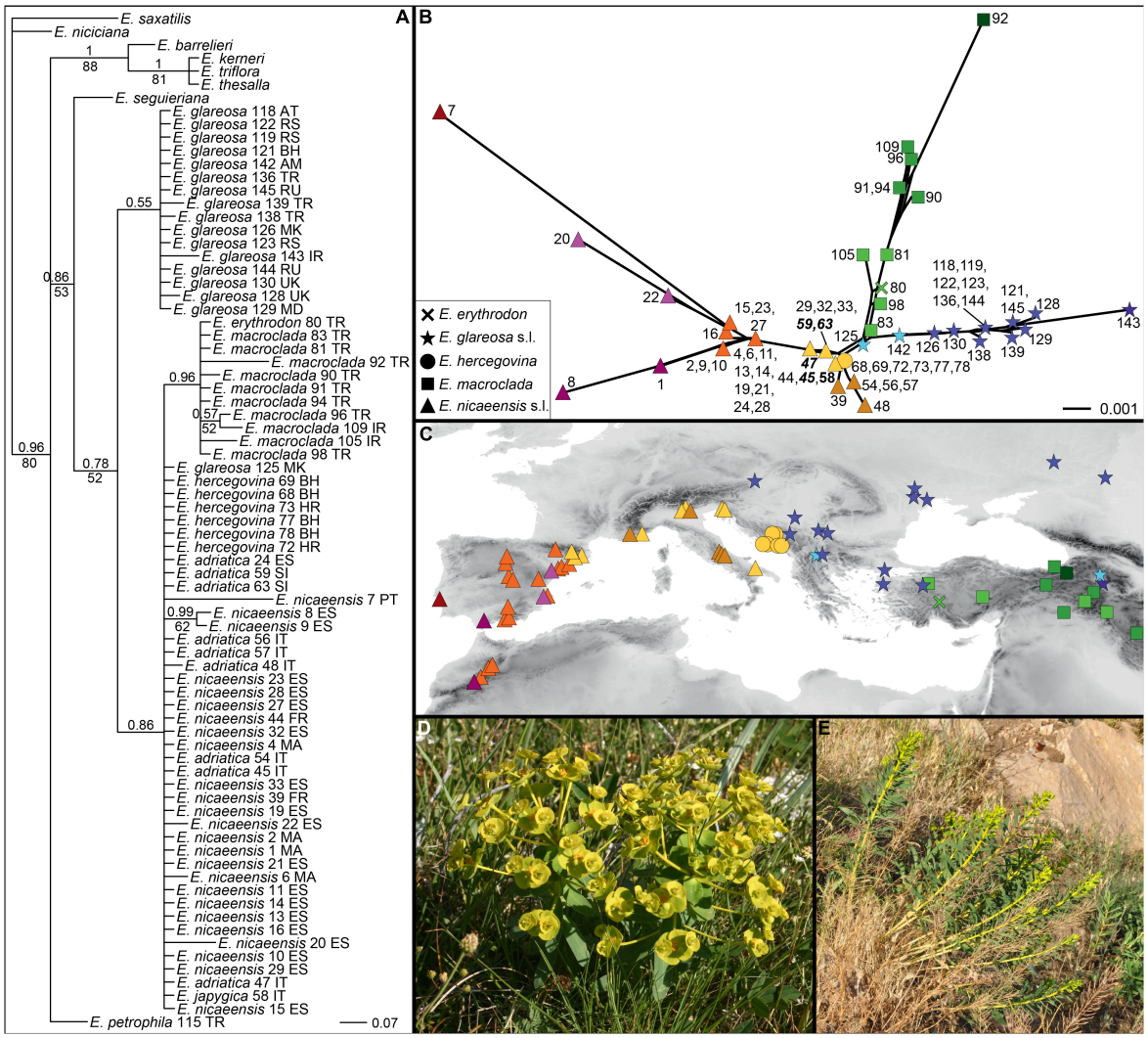


FIGURE 5 | Phylogenetic relationships within the *Euphorbia nicaeensis* alliance and between this alliance and its closest relatives within *E.* sect. *Pithyusa* as inferred by Internal Transcribed Spacer (ITS) sequences. **(A)** Bayesian consensus phylogram with posterior probabilities > 0.50 above, and parsimony bootstrap values > 50% below branches; country codes follow the accession names. **(B)** NeighborNet and **(C)** geographic position of the ITS ribotype groups revealed by the NeighborNet and indicated by different colours. Population numbers in **(A,B)** are presented in **Supplementary Figure 1** and in **Supplementary Table 1**. Population numbers of populations that are based on our revised taxonomic treatment belong to *E. adriatica* and are in **(B)** in bold italics (45, 47, 59, 63), and the one corresponding to *E. japygica* (58) is in bold. **(D)** *Euphorbia adriatica* from Italy and **(E)** *E. macroclada* from Turkey in their natural habitats. Photos: B. Frajman **(D)**, C. Gilly **(E)**.

The NeighborNet (**Figure 5B**) revealed a geographic structure (**Figure 5C**) in the variation of ITS sequences. Accessions of *E. adriatica* (**Figure 5D**), *E. hercegovina*, *E. japygica*, and *E. nicaeensis* from the central Mediterranean were positioned in the centre of the network (yellow, brown). Three main splits branched from the central ribotypes, corresponding to the western Mediterranean *E. nicaeensis* (orange, red, violet), *E. glareosa* s.l. (blue), and *E. macroclada* (**Figure 5E**), including *E. erythrodon* (green).

Relative Genome Size

Relative genome size values revealed clearly different DNA-ploidy levels (Suda and Trávníček, 2006) within *E. glareosa* s.l. and *E. petrophila*, whereas only diploids were found within *E. erythrodon*, *E. hercegovina*, *E. macroclada*, and *E. nicaeensis* s.l.; the single population sample of *E. japygica* was polyploid (**Figure 6** and **Supplementary Figure 4**). The mean RGS for the single population sample of *E. erythrodon* was 0.71 and that of *E. macroclada* ranged between 1.02 and 1.23. In *E. hercegovina*, RGS ranged between 1.07 and 1.15. The RGS of *E. nicaeensis* s.l. ranged between 1.10 and 1.50 and was discretely distributed; the populations from the Apennine Peninsula and the northwesternmost Balkan Peninsula (*E. adriatica*) had RGS between 1.10 and 1.19, and those from west of the Alps (including the Maritime Alps in France) between 1.29 and 1.50. The only RGS-polyploid population from the *E. nicaeensis* lineage was

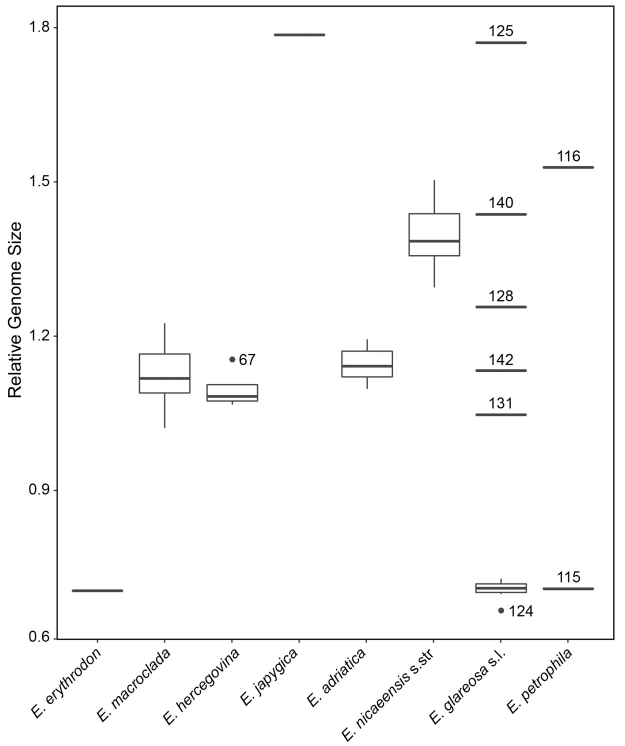


FIGURE 6 | Relative genome size (RGS) variation in the *Euphorbia nicaeensis* alliance. Outliers putatively belonging to the same ploidy level as the majority of the samples are presented as dots, whereas putatively ploidy-divergent outliers are presented as lines, including their population numbers, which correspond to **Supplementary Figure 1** and **Supplementary Table 1**.

that of *E. japygica* with RGS of 1.79. RGS-diploid populations of *E. glareosa* s.l. had RGS between 0.67 and 0.73, whereas the putatively polyploid populations had divergent RGS values, ranging between 1.05 and 1.77 (population, 131: RGS, 1.05; 142: 1.13; 128: 1.26; 140: 1.44; 125: 1.77). The two populations of *E. petrophila* had RGS of 0.71 (likely diploid) and 1.53 (likely tetraploid).

Morphometry

The character-states for all morphological characters, including ratios, are presented in **Supplementary Table 2**. Box plot diagrams of important differential characters are shown in **Supplementary Figure 5**. The correlation coefficients exceeded 0.9 in two character pairs: Length of a middle stem leaf/Distance from the base to the widest part of a middle stem leaf and Fruit length/Fruit width, and, thus, the characters Distance from the base to the widest part of a middle stem leaf and Fruit width were excluded from the PCA and DA analyses.

The PCA scatter plot (first three components explaining 28.41, 12.06, and 8.65% of the total variation) based on vegetative and cyathium characters showed a weak separation of *E. macroclada* from *E. nicaeensis* s.l., *E. glareosa* s.l., and *E. hercegovina*, along the first component but a big overlap among all four species on the second component (**Figure 7A**). The characters, which contributed most to the separation along the first component, i.e., those having the highest component scores (between 0.65 and 0.86) were stem length, stem width, length of the longest terminal ray, length of the longest fertile axillary ray, length of a middle stem leaf, width of a middle stem leaf, length of a ray leaf, width of a ray leaf, length of a raylet leaf, and width of a raylet leaf. The DA scatter plot (**Figure 7B**) showed a clear separation of both *E. macroclada* and *E. hercegovina* from the other two taxa (*E. glareosa* s.l. and *E. nicaeensis* s.l.) along the first factor (Wilks' Lambda = 0.019, χ^2 = 378.856, df = 90, $p < 0.001$) and an overlap amongst all four species along the second factor (Wilks' Lambda = 0.138, χ^2 = 187.835, df = 58, $p < 0.001$). Variables with the highest discriminant loadings on the first factor were stem width, length of a ray leaf, length of a raylet leaf, width of a raylet leaf, width of cyathial involucre, ratio length of cyathial involucre/width of cyathial involucre, depth of gland emargination, length of cyathial gland, width of cyathial gland, and ratio length of cyathial gland/width of cyathial gland. BoxPlots (**Supplementary Figure 5**) revealed that *E. macroclada* is the largest of the studied species, resulting in higher measurement values of characters indicated as important in PCA and DA analyses. On the other hand, *E. hercegovina* has the smallest and narrowest leaves of all four taxa; the cyathial involucre is narrowest in *E. glareosa* s.l., which, consequently, has the highest ratio between length and width of the cyathial involucre.

For fruit characters, the PCA (first three components explaining 42.30, 29.18, and 16.64% of the total variation) showed a weak separation of *E. macroclada* from all other species along the first component but a strong overlap along the second component (not shown). The characters contributing the most to the separation along the first component, i.e., those having the highest component scores (between 0.8 and 0.99), were fruit length, distance from the base to the widest part of the fruit and style length. Similarly, the DA scatter plot (first factor: Wilks' Lambda = 0.142, χ^2 = 85.856, df = 18, $p < 0.001$; second factor: Wilks' Lambda = 0.667, χ^2 = 17.801, df = 10, $p = 0.058$) showed a separation of *E. macroclada* from *E. glareosa* s.l., *E. hercegovina*, and *E. nicaeensis* s.l. along the first factor but an overlap amongst all four species along the second factor, where a slight trend separating *E. glareosa* s.l. and *E. nicaeensis* s.l. could be observed (**Figure 7C**). The characters fruit length and distance from the base to the widest part of the fruit had the highest discriminant loadings on the first factor. BoxPlots (**Supplementary Figure 5**) confirmed that *E. macroclada* had the largest fruits, resulting in larger measurement values for most fruit characters. They also indicated that *E. glareosa* s.l. can have smaller, especially narrower fruits compared to *E. nicaeensis* s.l., whereas the values of *E. hercegovina* overlapped with those of *E. nicaeensis* s.l.

The PCA for seed characters (first three components explaining 46.81, 20.41, and 15.03% of the total variation) showed an overlap between *E. nicaeensis* s.l., *E. hercegovina*, and *E. glareosa* s.l., and a weak separation of *E. macroclada* along the first component, whereas, along the second component, all species overlapped strongly (not shown). The characters contributing the most to the separation along the first component, i.e., those having the highest component scores

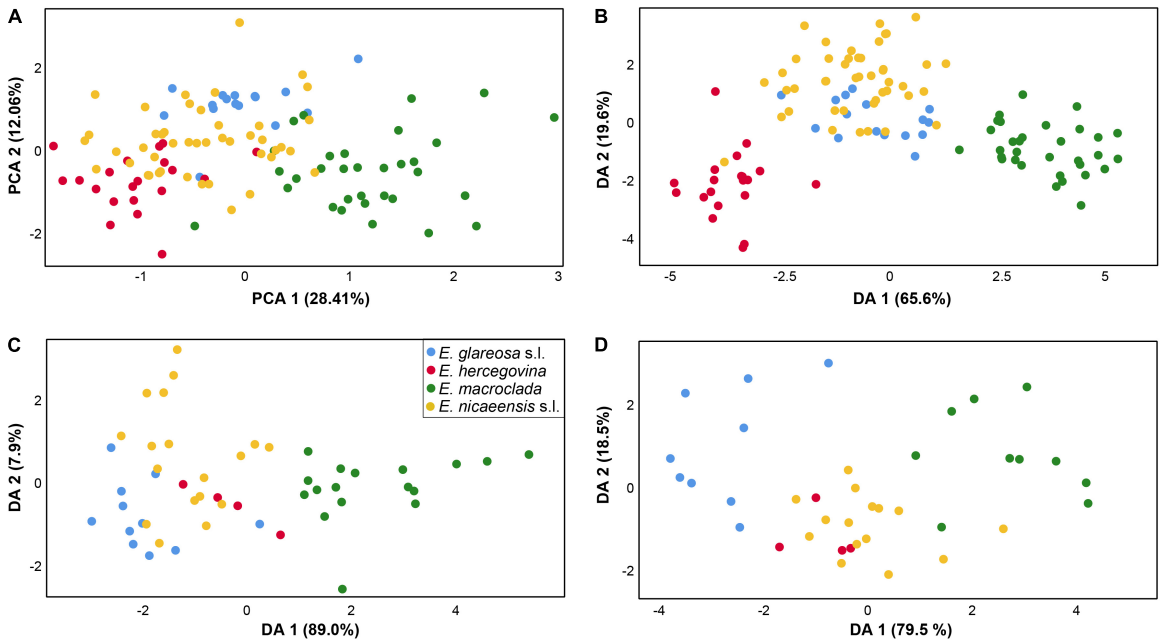


FIGURE 7 | Morphological differentiation amongst *Euphorbia glareosa* s.l. (blue), *E. hercegovina* (red), *E. macroclada* (green), and *E. nicaeensis* s.l. (yellow). Principal component analysis (PCA; **A**) and discriminant analyses (DA) based on **(B)** 21 metric and nine ratio vegetative and cyathium characters, **(C)** four metric and two ratio fruit characters, and **(D)** six metric and four ratio seed characters.

(between 0.65 and 0.97), were seed length, seed width, distance from the base to the widest part of a seed, caruncle length, and caruncle width. The DA scatter plot (first factor: Wilks' Lambda = 0.089, χ^2 = 72.600, df = 30, $p < 0.001$; second factor: Wilks' Lambda = 0.460, χ^2 = 23.310, df = 18, $p = 0.179$) showed an overlap between *E. nicaeensis* s.l. and *E. hercegovina* and a slight separation of *E. glareosa* s.l. and *E. macroclada* along the first factor, and an overlap amongst all taxa along the second factor (**Figure 7D**). Variables with highest discriminant loadings on the first factor were seed length, ratio seed length/seed width, ratio of distance from the base to the widest part of a seed/seed length, caruncle length, caruncle width, and ratio caruncle length/caruncle width.

The PCA and DA analyses, including only closely related *E. adriatica*, *E. hercegovina*, and *E. nicaeensis*, showed a separation amongst the three taxa that were most pronounced in the vegetative and cyathium characters (**Figure 8A,B**) and less pronounced in fruit and seed characters (**Figure 8C,D**). The PCA scatter plot (first three components explaining 23.72, 13.74, and 10.76% of the total variation) based on vegetative and cyathium characters showed a slight trend in separation of the three taxa but also a strong overlap (**Figure 8A**). On the other hand, the DA scatter plot (**Figure 8B**) showed a clear separation amongst the three taxa. Variables with the highest discriminant loadings on the first factor (Wilks' Lambda = 0.030, χ^2 = 162.458, df = 60, $p < 0.001$), which clearly separated *E. hercegovina* from *E. nicaeensis*, whereas *E. adriatica* was intermediate, were length of (the longest) terminal ray, number of fertile axillary rays, length of (the longest) fertile axillary ray, length of a middle stem leaf, width of a middle stem leaf, ratio length of a middle stem leaf/width of a middle stem leaf, width of a ray leaf, ratio length/width of a ray leaf, distance from the base to the widest part of a ray leaf, ratio of distance from the base to the widest part of a ray leaf/length of a ray leaf, length of a raylet leaf, width of a raylet leaf, ratio length/width of a raylet leaf, distance from the base to the widest part of a raylet leaf, ratio of distance from the base to the widest part of a raylet leaf/length of a raylet leaf, width of cyathial involucre, and ratio length of cyathial involucre/width of cyathial involucre. Variables with highest discriminant loadings on the second factor (Wilks' Lambda = 0.240, χ^2 = 66.352, df = 29, $p < 0.001$), which separated *E. adriatica* from the other two species, were stem glabrous/pubescent, length of (the longest) terminal ray, width of a middle stem leaf, ratio length of a middle stem leaf/width of a middle stem leaf, length of a ray leaf, width of a ray leaf, distance from the base to the widest part of a ray leaf, length of a raylet leaf, width of a raylet leaf, ratio length/width of a raylet leaf, length of cyathial involucre, width of cyathial involucre, ratio length of cyathial involucre/width of cyathial involucre, depth of gland emargination, length of cyathial gland, ratio depth of gland emargination/length of cyathial gland, and ratio length of cyathial gland/width of cyathial gland.

The PCAs of fruit and seed characters, respectively, showed a high overlap of *E. adriatica*, *E. hercegovina*, and *E. nicaeensis* (not shown), whereas the DAs (**Figure 8B,C**) indicated a discrimination trend between *E. nicaeensis* from *E. adriatica* and *E. hercegovina* along the first factors (fruit characters: Wilks' Lambda = 0.175, χ^2 = 26.134, df = 14, $p = 0.025$; seed characters: Wilks' Lambda = 0.204, χ^2 = 19.089, df = 18, $p = 0.386$). Variables with highest discriminant loadings on these factors were fruit

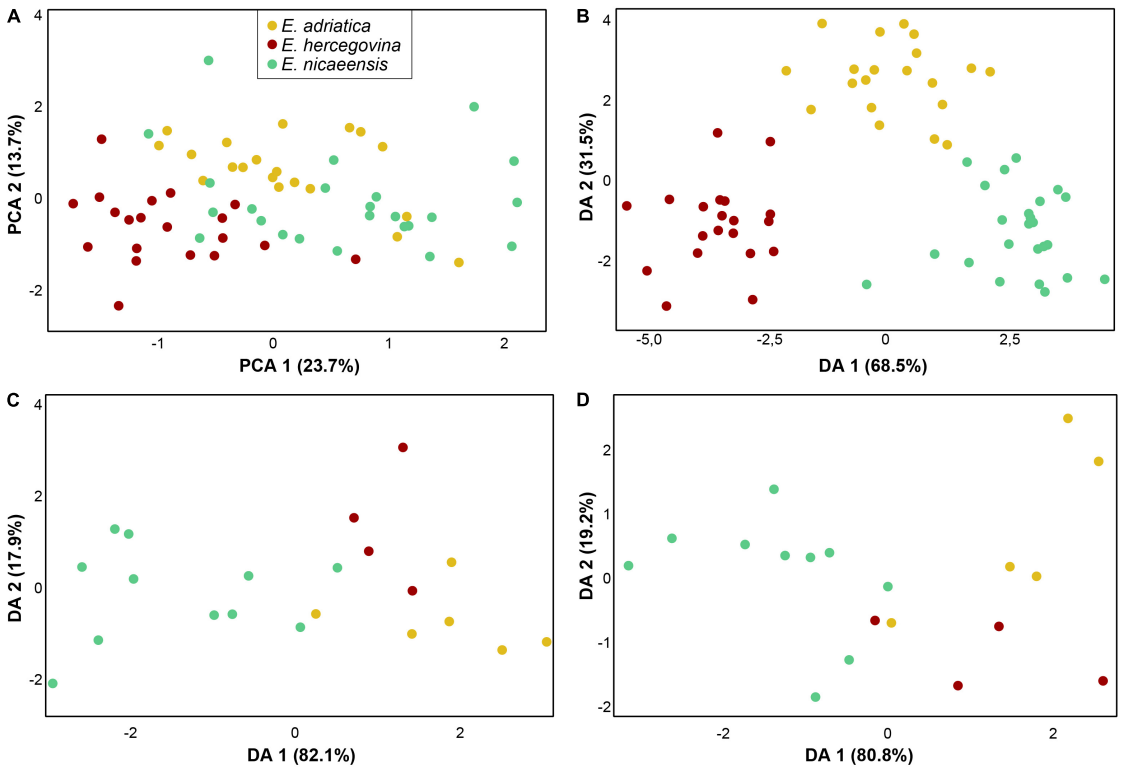


FIGURE 8 | Morphological differentiation amongst *Euphorbia adriatica* (yellow), *E. hercegovina* (red), and *E. nicaeensis* (green). Principal component analysis (PCA; **A**) and discriminant analyses (DA) based on **(B)** 21 metric and nine ratio vegetative and cyathium characters, **(C)** five metric and two ratio fruit characters, and **(D)** six metric and four ratio seed characters.

length, fruit width, ratio fruit length/fruit width, seed length, seed width, ratio seed length/seed width, distance from the base to the widest part of a seed, ratio of distance from the base to the widest part of a seed/seed length, caruncle length, and distance from the base to the widest part of caruncle.

DISCUSSION

RADseq, originally developed for intraspecific phylogeographic studies (McCormack et al., 2013; Lemmon and Lemmon, 2013), allowed us to establish a clear phylogenetic hypothesis regarding the origin of and the relationships within the *E. nicaeensis* alliance. The RADseq data clearly inferred the *E. nicaeensis* alliance as monophyletic (**Figures 2**, **3**) and, alongside the RGS and morphometric data, helped us describe a new species, *E. adriatica* (see below), which is, together with *E. erythrodon*, *E. glareosa* s.l., *E. hercegovina*, *E. japygica, E. macroclada*, *E. nicaeensis*, and *E. petrophila*, a member of this alliance.

Phylogeographic Origin and Diversification of the Mediterranean Taxa

Within the Mediterranean Basin, diversification patterns revealed by different methods indicate complex evolutionary pathways and cryptic divergence that went unnoticed by earlier taxonomists. Biogeographic analyses (**Figure 2B**) support the Western Asian origin of the Mediterranean lineage, which diversified *in situ* likely as a result of Pleistocene climatic oscillations, accompanied by adaptation to different habitats and polyploidisation. Thus, our study underlines the importance of the IT floristic region as a source area for many Mediterranean lineages (Manafzadeh et al., 2017). Unexpectedly, one of the main genetic breaks revealed by RADseq (**Figures 3**, **4**), accompanied also by a clear divergence in RGS (**Figure 5**), falls within the seemingly continuous distribution of *E. nicaeensis* s.l. and separates the populations west of the Alps from those south of the Alps, which we describe as a new species, *E. adriatica* (see below). Given the Pleistocene diversification within the *E. nicaeensis* alliance (**Figures 2B**, **4A**), It is likely that the inferred phylogeographic pattern—also reflected in RGS and morphological divergence (**Figures 6**, **8**)—is a result of survival in isolated Pleistocene glacial refugia in the western (Iberian Peninsula) and the central/eastern Mediterranean (Apennine and Balkan peninsulas). All three peninsulas are renowned as important glacial refugia, where distinct genetic lineages persisted through the Pleistocene and Quaternary climatic fluctuations (Bilton et al., 1998; Hewitt, 1999, 2000, 2011; Petit et al., 2003; Nieto Feliner, 2014; Cresti et al., 2019).

Interestingly, the western lineage (*E. nicaeensis*) largely corresponds in its range to *E. flavicoma* DC. from the *E. verrucosa* alliance, which was also suggested to have had its Pleistocene

refugium in the Iberian Peninsula (Caković et al., 2021). *Euphorbia flavicoma* and *E. nicaeensis* are ecologically similar, inhabiting Mediterranean scrublands, dry and warm grasslands, and open forests. From the putative Pleistocene refugium in the Iberian Peninsula, both species extended their ranges to southern France, where their eastward migration was likely obstructed by the Alps. Similar moderate expansion out of Iberia has also been observed in *Arabis scabra* L. (Koch et al., 2020), *Pinus pinaster* Aiton (Bucci et al., 2007), and *Quercus suber* L. (Magri et al., 2007). Congruent expansion patterns of different warm-adapted taxa have likely been influenced by climatic factors, which prevented more extensive dispersals out of Iberia (Caković et al., 2021). In westernmost Europe, Mediterranean climate is, nowadays, prevalent in the Iberia, and is restricted to southernmost France (Peel et al., 2007).

Compared to *E. nicaeensis*, Apennine-Balkan *E. adriatica* exhibits a smaller RGS and is ecologically divergent, usually thriving in mesophilic submediterranean grasslands and scrublands in the Apennine Peninsula, the southern outskirts of the Alps, and the northwestern Balkan Peninsula. This species likely had its glacial refugium in the Apennine Peninsula, from where it dispersed to the southern margin of the Alps, which would have acted as a prominent biogeographic barrier for northward migration of Apennine biota (Hewitt, 2000). The third Mediterranean species, separated by a distribution gap of 300 km from *E. adriatica*, is the Balkan endemic *E. hercegovina*. It grows in open dolomitic grasslands and pine forests with a submediterranean character in the central Balkan Peninsula, where it likely had its Pleistocene refugium, and from where it did not spread considerably. It is morphologically clearly divergent and also has a divergent RGS. Contrary to many other examples of disjunctly distributed amphi-Adriatic lineages, in which the distribution area in the Balkan Peninsula exceeds in its size that in the Apennine Peninsula [see Frajman and Schönswetter (2017) and Falch et al. (2019) for reviews], this is not the case for *E. adriatica* and *E. hercegovina*. Partly incongruent relationships amongst the three species, resolved by different analytical approaches of the RADseq data outlined in the Results, are accompanied by different patterns of morphological and RGS divergence. Whereas the RGS data support a closer relationship between *E. adriatica* and *E. hercegovina* (**Figure 6**), morphological data point to *E. adriatica* and *E. nicaeensis* as more similar (**Figure 8**).

In addition to the Pleistocene glaciations, which fragmented the range of the Mediterranean *E. nicaeensis* lineage and likely triggered its divergence in three glacial refugia, adaptation to different substrates and climatic conditions, as well as polyploidisation in the southern Apennine Peninsula, contributed to diversification of the *E. nicaeensis* group. Specifically, the only analysed population of *E. japygica* from the southern Apennine Peninsula is DNA-polyploid (**Figure 6**). This, alongside the morphological differentiation reported by Fenu et al. (2016) and the lack of overlap in distribution with *E. adriatica* warrant recognition of this taxon at the species level, as originally proposed by Tenore (1830). Further studies, including more populations, are, however, needed to clarify the status of this taxon.

Morphological Diversification Reflects Ecology Rather Than Phylogenetic Relationships

Morphological diversification only partly follows evolutionary trajectories in the *E. nicaeensis* alliance. The discordant patterns likely result from adaptation to similar habitats within divergent phylogenetic lineages, on the one hand, and to different habitats within the same evolutionary lineages, on the other hand. The traditional, morphology-based taxonomic treatments largely do not reflect evolutionary relationships. *Euphorbia glareosa* s.l. and *E. nicaeensis* s.l., which are morphologically similar (**Figure 7**) and were often considered conspecific (e.g., Radcliffe-Smith and Tutin, 1968; Kuzmanov, 1979; Radcliffe-Smith, 1982; Greuter et al., 1986; Govaerts et al., 2000), are, in fact, phylogenetically clearly divergent. Either adaptation to similar environments has triggered a parallel evolution of similar morphological traits in both lineages or the overall similarity of *E. glareosa* s.l. and *E. nicaeensis* s.l. was inherited from a shared common ancestor. It is well-known that the European steppes share several characteristics with the Mediterranean grasslands, which is reflected in multiple shared taxa and similar adaptations across both biogeographic regions (Peart, 2008; Hamasha et al., 2012). There is much variability in morphological traits connected to a habit (plant and leaf size) both within *E. nicaeensis*, but, especially, within *E. glareosa* s.l., which is reflected in the description of several intraspecific taxa (e.g., Kuzmanov, 1979; Greuter et al., 1986; Geltman, 2009, 2020). Whether, in the latter case, the morphological variability reflects evolutionary entities or, rather, adaptation to divergent habitats requires further investigation and is beyond the scope of this study.

Similarly, morphologically distinct *E. hercegovina*, which was earlier considered to belong to *E. barrelieri* (Hayek, 1924; Kuzmanov, 1963; Poldini, 1969; Greuter et al., 1986; Govaerts et al., 2000; Geltman, 2009), is nested within (**Figure 3A**) or closely related to *E. nicaeensis* s.l. (**Figure 4**). Its morphological divergence is likely a result of adaptation to dolomitic substrates, which are also typical for taxa of the *E. barrelieri* group (Frajman and Schönswetter, 2017). The soil derived from dolomitic bedrock is shallow and dry, resulting in extreme growing conditions. Plants growing in such habitats have to be tolerant of high magnesium and low-moisture levels, leading to the strong selective pressures that such extreme habitats impose (Ware, 1990; Noss, 2012).

Similarly, *E. erythrodon* growing on mountain ridges and screes has a dwarf prostrate habit, with stems that rarely exceed 7 cm (Radcliffe-Smith, 1982), which is likely an adaptation to mountain habitats (Larcher et al., 2010; Körner, 2012; Gehrke et al., 2016). A superficially similar case is provided by *E. seguieriana* subsp. *loiseleurii* (Rouy) Greuter and Burdet, which occurs in the windswept summit area of Mt. Ventoux in the French Provence, and exhibits a similar dwarf habit as an adaptation to this habitat. Whereas this latter taxon is nested within *E. seguieriana* and does not deserve taxonomic recognition (Frajman et al., 2019), *E. erythrodon* is phylogenetically distinct. However, the single studied population is resolved as intermediate between *E. glareosa* s.l. and

E. macroclada (**Figures 4B,C**); further studies, including additional populations, are needed to confirm our preliminary findings. Finally, the more robust habit and bigger size of all organs in *E. macroclada*, as compared to *E. nicaeensis* s.l. (**Supplementary Figure 5**), are, possibly, a consequence of the former taxon mostly growing in deeper clay soils over siliceous and sandstone substrates, which are more humid and nutrient-rich than the better-drained calcareous substrates, on which *E. adriatica* and *E. nicaeensis* usually occur (Frajman, personal observation).

Altogether, our results demonstrate how heterogeneous environments can outweigh a phylogenetic signal, resulting in taxonomic treatments not reflecting evolutionary pathways. Also, in other plant lineages, It has been shown that heterogeneous environments have contributed to the high diversity of the Mediterranean (e.g., Frajman and Oxelman, 2007; Du Pasquier et al., 2017; Đurović et al., 2017).

Higher Phylogenetic and Relative Genome Size Diversity in Irano-Turanian and Steppic Areas as Compared to the Mediterranean Basin

The Mediterranean Basin acted as a cradle for the diversification of the *E. nicaeensis* lineage, where the phylogenetic relationships clearly reflect geographical distribution of the taxa. The diversification of the closely related taxa from the easterly adjacent IT and Eurasian steppic regions, in contrast, was more turbulent and geographically less coherent, resulting in multiple, clearly divergent sympatric lineages as indicated by RADseq data (**Figures 2**, **3**), as well as in the polyploidisation events indicated by the RGS data (**Figure 5**). Whereas we only detected a single polyploid population (58) from southern Italy within *E. nicaeensis* s.l. (two further polyploid populations with $2n = 56$ have been reported from Spain; Vilatersana and Bernal, 1992), we recorded multiple RGS-divergent populations within *E. glareosa* s.l. and *E. petrophila*. They possibly result from several polyploidisation events, even if other factors influencing changes in RGS (Pellicer et al., 2018) cannot be ruled out in the absence of chromosome counts.

The species from the IT region that appears most closely related to the Mediterranean taxa is *E. macroclada*, distributed from Anatolia to the Armenian, Iranian, and Syrian highlands (**Figure 1**; Radcliffe-Smith, 1982). Less clear is the phylogenetic position of narrowly distributed *E. erythrodon*, which is limited to mountain ridges and screes of southwestern and central Anatolia. It was resolved as a sister to the *E. nicaeensis* lineage by the complete RADseq dataset (**Figure 3**) and is nested within *E. macroclada* in the ITS phylogeny (**Figure 5**). The RADseq data limited to the *E. nicaeensis* alliance indicate its close relation to *E. glareosa* s.l. (**Figure 4A**), but also evidence admixture with sympatric *E. macroclada* (**Figure 4B**). A close relationship with *E. glareosa* s.l. is further supported by their similar RGS (**Figure 6**).

The RGS data (**Figure 6**) further indicate that a substantial increase in genome size (GS) likely happened in the common ancestor of *E. adriatica*, *E. hercegovina*, *E. macroclada*, and *E. nicaeensis*. Without chromosome counts, it is impossible to say whether this increase was due to polyploidisation. Since the two published chromosome counts of *E. macroclada* and most counts of *E. nicaeensis* are $2n = 18$, which are the same as those of diploid *E. glareosa* s.l. and outgroup *E. niciciana* and *E. seguieriana* (Rice et al., 2015), we hypothesise that the increase in GS in the *E. nicaeensis* lineage was not due to polyploidisation but likely due to other processes. Alongside polyploidy, accumulation of retrotransposons and other repetitive elements is considered main factors of GS increase in angiosperms (Pellicer et al., 2018), e.g., leading to 2-fold increase in GS in the wild rice relative *Oryza australiensis* (Piegu et al., 2006). Genç and Kültür (2020), recently, have also published a tetraploid chromosome count from *E. macroclada*, whereas our RGS data only indicated polyploidisation in *E. glareosa* and *E. petrophila* from Turkey. Additional studies with more complete taxonomic and denser geographic sampling are needed to display how important role polyploidisation played for the diversification of this group in the IT region.

Most of the studied populations of *E. glareosa* s.l., with the exception of population 125 from North Macedonia, formed a monophyletic group in the RADseq data, closely related to *E. erythrodon* and *E. macroclada* (**Figures 2**, **3**). In the ITS tree, these populations formed a poorly supported clade (**Figure 5A**), while, in the ITS NeighborNet, they were positioned along the same split, where population 125 was at the basis of this split, close to the populations of *E. hercegovina* and *E. adriatica* (**Figure 5B**). Population 125 from North Macedonia had the highest RGS of all studied populations, indicating its polyploid origin, which is likely responsible for its divergent phylogenetic placement separated from the other populations of *E. glareosa*. As all geographically close Balkan populations clearly belong to *E. glareosa* s.l.—the closest diploid population, 126, being only 25 km away—we also included population 125 in this species. Alternatively, based on the strikingly similar RGS of the population 58 of *E. japygica* from southern Italy (**Figure 6**), we cannot exclude their common origin, as population 125 also has hairy fruits, which is the most important character distinguishing *E. japygica* from *E. nicaeensis* s.l. (Tenore, 1830). Biogeographic connections between southern Italy and the Balkan Peninsula have been evidenced in several other plant groups [see Frajman and Schönswetter (2017) for a review]. Further studies, including chromosome counts and extensive sampling in southern Italy, as well as in geographically intermediate Albania, where populations belonging to the *E. nicaeensis* alliance are only known from three localities in the north of the country (Barina, 2017), are needed.

In the same line, the polyploid origin of population 116 of *E. petrophila* is likely the reason for its divergence from the diploid population 115, as inferred by the complete RADseq data (**Figure 3A**). On the other hand, in the analyses of the *E. nicaeensis* alliance dataset, both populations were sisters to all other ingroup accessions (**Figure 4**). Within *E. glareosa* s.l., additional divergent RGS values of some populations scattered throughout the entire distribution area have been observed (**Figure 5**). It remains unclear whether this is

due to tetraploidisation, followed by genome downsizing that differentially acted in geographically distant populations, or if accumulation of repetitive elements (Pellicer et al., 2018) is responsible for the observed pattern. Further studies, with extended geographic and taxonomic sampling, are needed to disentangle the diversification patterns within *E. glareosa* s.l., which is an assemblage of around ten described taxa (Prokhanov, 1949; Kuzmanov, 1979; Radcliffe-Smith, 1982), and to establish its relationships with *E. petrophila* and the Anatolian narrow endemics *E. pestalozzae* Boiss. and *E. pisidica* Hub.-Mor. and M. S. Khan (Radcliffe-Smith, 1982).

Partly Incongruent Patterns Inferred by Different Analyses of the Restriction Site Associated DNA Sequencing Data

Different analytical approaches of the RADseq data resulted in partly incongruent patterns outlined in the sections "Results" and "Discussion" above, which is often the case in phylogenetic analyses of such data (Wagner et al., 2020; Cai et al., 2021; Rose et al., 2021; Hühn et al., 2022). Both biological as well as methodological factors can be responsible for the observed incongruences. Having in mind the group's diversification onset in the Pleistocene, and the vast areas (from northwest Africa and Iberia to Central Asia) that it inhabits, rapid range expansions resulting in incomplete lineage sorting and secondary contacts amongst recently diverged lineages can be responsible for some of the observed incongruences (Cai et al., 2021; Rose et al., 2021). Genetic admixture evidenced for *E. erythrodon* and the population 125 of *E. glareosa* (**Figure 4C**) was another possible cause underlying incongruences. The population 125 had, along with *E. japygica*, the highest genome size of all samples (**Figure 6**) and may be of allopolyploid origin. The polyploid nature of some samples can strongly influence phylogenetic inference. Incongruence between two differently analysed datasets was thus observed in *Salix*, where one of the analysed species was of allopolyploid origin (Wagner et al., 2020). However, for merely reconstructing relationships, which is the focus of our study, the approach used in the present study has been shown to be valuable even in plant groups with high incidence of polyploidy (Záveská et al., 2019).

In addition, topological differences between the two different RADseq datasets (**Figure 4** vs. **Figure 5**) may be attributable also to the various amounts of loci shared across the sampled taxa. In the complete dataset (**Figure 4**), the limited amount of loci shared (1,498 SNPs) across the broad range of species sampled had too little information to resolve relationships at deeper nodes. On the other hand, the 3,000 loci shared across species limited only to the *E. nicaeensis* alliance better resolved relationships amongst them, and the inferred phylogenies better reflected the morphological (e.g., grouping of *E. adriatica* and *E. nicaeensis*, as well as both populations of *E. petrophila* with strong support) and the RGS data (e.g., grouping of *E. erythrodon* with *E. glareosa*). This underlines that RADseq data are better suited to infer relationships amongst most closely related taxa, compared to deeper evolutionary nodes.

Taxonomic Considerations and a Revised Taxonomic Treatment

The combination of phylogenetic, RGS, and morphometric data allows us to propose a revised taxonomic treatment of the *E. nicaeensis* alliance, resolving some long-standing uncertainties about species delimitations within this alliance, but also introducing new questions that will need to be answered in the future. In **Figure 9**, we graphically present the relations within the *E. nicaeensis* alliance, partly based on our study and partly outlined in the introduction (for *E. glareosa* s.l.). The main taxonomic outcomes of our study can be summarised in the following four points. (1) *Euphorbia nicaeensis*. l. and *E. glareosa* s.l. are phylogenetically divergent and geographically allopatric lineages with distinct RGS and should, despite their morphological similarity, be treated as distinct species and not as subspecies, as proposed by Radcliffe-Smith and Tutin (1968) and Greuter et al. (1986). (2) *Euphorbia macroclada* is a distinct species distributed in the IT region, closely related to the Mediterranean *E. nicaeensis* lineage, but including traces of genomic imprint shared with *E. erythrodon* and *E. glareosa*. (3) The Mediterranean *E. nicaeensis* lineage is an assemblage of three allopatric, phylogenetically divergent, and morphologically different (although with overlapping character states) groups of populations with distinct RGS that deserve recognition at the species level, namely, the western Mediterranean *E. nicaeensis*, the central-eastern Mediterranean *E. adriatica*, and the eastern Mediterranean *E. hercegovina*. In addition, we preliminary treat the southern Italian populations as a distinct species, *E. japygica*, but further studies are needed to explore whether all populations belonging to this taxon are (a) polyploid, (b) morphologically and phylogenetically distinct from *E. adriatica*, and (c) share a most recent common ancestor with Balkan populations in Albania and North Macedonia. (4) *Euphorbia erythrodon*, *E. glareosa* s.l., and *E. petrophila*, which thrive in the Eurasian steppes and the southerly adjacent IT region, need to be further studied based on an extensive geographic and taxonomic sampling (including also *E. pestalozzae* and *E. pisidica*) to (a) disentangle phylogenetic relationships and morphological differentiation amongst the populations and different taxa described from this area, (b) explore the incidence of polyploidy within this group, and (c) provide a revised taxonomic treatment.

Identification Key

High variability and overlap of morphological characters amongst the taxa of the *E. nicaeensis* alliance hinder a construction of a reliable identification key based solely on morphology. Especially, the overlap between *E. nicaeensis* s.l. and *E. glareosa* s.l. is considerable, whereas *E. hercegovina* and *E. macroclada* are more divergent. Moreover, absence or presence of horns on the cyathial glands, traditionally used to discriminate between *E. glareosa* s.l. and *E. macroclada*, respectively (e.g., Prokhanov, 1949; Radcliffe-Smith, 1982; Geltman, 2020), proved not to be a reliable character, as several examined specimens of *E. glareosa* s.l. also had horn-like appendages. Also the characters given in Flora Europaea

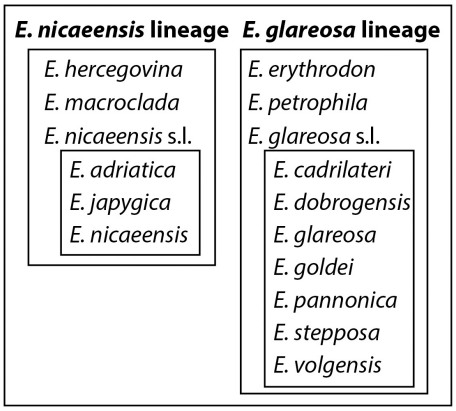


FIGURE 9 | Relations amongst taxa within the *E. nicaeensis* alliance as a result of the outcomes of this study, combined with previous treatments within *E. glareosa* s.l., which were not addressed in this study.

(Radcliffe-Smith and Tutin, 1968) to distinguish between *E. glareosa* s.l. and *E. nicaeensis* s.l., i.e., the number of rays and capsule size, have proved to be highly overlapping between the two taxa (see species descriptions below). In the following key, several morphological characters, therefore, overlap, but a combination of several characters with geographic data makes unambiguous identification possible.

1 Robust erect plants, (19) 30–62 (70) cm high, with (3) 4–5 (6)-mm thick stems. Cauline leaves (3) 4–6.5 (8) × (0.5) 0.7–1.6 (2) cm. Cyathial glands with two, often lobate or multifid horns. Fruits (5.9) 6.1–8.4 (8.8) × (3.1) 4.1–5.7 (6.2) mm. Seeds (2.9) 3.1–4 (4.2) × (1.9) 2–2.5 (2.6) mm. *West Asia*.....***E. macroclada.***

1* Less robust, decumbent to erect plants, (4) 10–40 (45)-cm high, with (1) 2–4-mm thick stems. Cauline leaves (0.7) 1.5–5 (5.5) × (0.2) 0.4–1.1 (1.5) cm. Cyathial glands without or with horns; horns sometimes lobate, never multifid. Fruits (3.6) 4–5.5 (6) × (2.2) 2.4–3.8 (4.2) mm. Seeds (2) 2.5–3.1 (3.4) × (1.2) 1.5–2 (2.1) mm..... **2.**

2 Decumbent plants, (6.5) 8–17 (22)-cm high, with 1–2-mm thick stems. Cauline leaves (0.7) 1–2 (2.7)-cm long, (1.1) 1.6–4.1 (5.4) times longer than wide. *Central Balkan Peninsula*..... ***E. hercegovina.***

2* Decumbent to erect plants, (5) 11–40 (45)-cm high, with (1) 2–4-mm thick stems. Cauline leaves (1.7) 2–5 (5.5)-cm long, (2.3) 3.2–6.2 (6.8) times longer than wide..... **3.**

3 Terminal rays 5–9 (11), the longest (2) 2.5–5.5 (7)-cm long. Cauline leaves 1.8–4 (4.5) × (0.3) 0.4–0.8 cm, (3.8) 4.6–6.2 (6.8) times longer than wide. Fruits (3.6) 3.7–4.4 (4.6) mm × (2.3) 2.4–3.1 mm. Seeds, 2.6–2.7 mm × 1.6–1.9 (2) mm, (1.3) 1.4–1.6 (1.7) times longer than wide. *Apennine Peninsula*, *northern Balkan Peninsula*..... **4.**

3* Terminal rays 5–12 (13), the longest (1.2) 3–8.5 (10)-cm long. Cauline leaves (1.7) 2.3–5 (5.5) cm × (0.5) 0.6–1.2 (1.5) cm, (2.3) 3.2–5.5 (6.5) times longer than wide. Fruits (3.6) 4.1–5.7 (6) mm × (2.2) 3–4.1 (4.2) mm. Seeds (2) 2.2–3.1 (3.4) mm × 1.2–2.1 mm, (1.4) 1.5–1.9 (2) times longer than wide. *Western Mediterranean (excl. Apennine Peninsula), Central and East Europe, West Asia*..... **5.**

4 Fruits glabrous. *Northern and* central *Apennine Peninsula*, *northern Balkan Peninsula*..... ***E. adriatica***

4* Fruits pubescent. *Southern Apennine Peninsula* ..... ***E. japygica.***

5 Fertile axillary rays (1) 2–9 (10), the longest (1.2) 2–7.2 (8.1)-cm long. Cauline leaves (1.7) 2.2–3.8 (5) cm × (0.5) 0.6–1 cm. Cyathial glands usually bigger, (0.7) 0.8–1.5 (1.7) mm × (0.9) 1.1–1.9 (2.1) mm. Fruits (4) 4.3–5.7 (6) mm × (2.7) 3.1–4.1 (4.2) mm, 1.3–1.5 (1.6) times longer than wide. Styles (0.9) 1.9–2.2-mm long. Seeds, 2.7–3.1 (3.3) mm × 1.8–2.1 mm. *Western Mediterranean*..... ***E. nicaeensis***

5* Fertile axillary rays (2) 4–14 (15), the longest (3.2) 3.9–8.9 (12.2)-cm long. Cauline leaves (2.4) 2.5–5 (5.3) × (0.5) 0.6–1.2 (1.5) cm. Cyathial glands usually smaller, 0.7–1 mm × 1–1.5 mm. Fruits (3.6) 4.1–4.6 (5.5) mm × (2.2) 2.4–3.1 (3.3) mm, 1.4–1.9 times longer than wide. Styles (1.1) 1.3–1.7-mm long. Seeds, (2) 2.2–3.1 (3.4) mm × (1.2) 1.3–1.7 (1.8) mm. *Central and East Europe, West Asia.* .....***E. glareosa.***

Taxonomic Treatment

(1) ***Euphorbia adriatica*** **Stojilkovič, Záveská, and Frajman, sp. nov.** Type: Holotype: "Flora of Slovenia, Primorska, Kras: south of the road Senožeče – Senadole, 1.5 km west of Senožeče; 550 m; 14°0'38″ E 45°43'9″ N; dry meadow. 16 August 2021 *V. Stojilkovič and B.* Frajman 16939 (W0164201[5]). Isotypes in IB 113154, LJU, FI018954, ZA 62967 and 62969).

= *E. seguieriana* var. *prostrata* Fiori, Nuov. Fl. Italia 2: 183. 1926 ≡*E. nicaeensis* subsp. *prostrata* (Fiori) Arrigoni, Inform. Bot. Ital. 12: 140. 1980 (publ. 1981). — Type: Flora Italica – Herbarium Adr. Fiori: "Prov. di Firenze, Impruneta ai Sassi neri, solo serpentinoso, 315 m," 4 June 1911, *Adr[iano] Fiori s.n.* (FI002664!).

Note: *Euphorbia nicaeensis* subsp. *prostrata* has been described from serpentine outcrops in Italy, and dwarf individuals from these populations are, in our opinion, merely ecotypes adapted to this specific substrate. A similar adaptation to serpentines, partly at the same localities, has been observed also in *Euphorbia spinosa* L. (Stevanoski et al., 2020).

Diagnosis: Euphorbia adriatica differs from closely related *E. hercegovina* in being more robust in all vegetative characters, i.e., mostly having higher and thicker stems and bigger leaves, but smaller fruits. Compared to *E. nicaeensis*, *E. adriatica* often has relatively longer leaves compared to their width and smaller fruits with shorter styles as well as smaller seeds. It also has glabrous fruits, whereas *E. japygica* has pubescent fruits.

Description: Glabrous and glaucous perennial, (7) 11–37 (40)-cm high, with (1) 2–3-mm thick stems. Terminal rays

[5]https://w.jacq.org/W0164201

5–9 (11), the longest (2) 2.5–5.5 (6.7)-cm long, 1–2 times dichotomously branched. Fertile axillary rays 2–7 (10), the longest (2) 2.6–6.1 (7.2)-cm long. All leaves with entire margin. Cauline leaves linear oblanceolate (1.8) 1.9–4.1 (4.4) cm × (0.3) 0.4–0.8 cm, (3.8) 4.6–6.2 (6.8) times longer than wide, widest at (0.6) 0.7–0.8 of their length, with cuneate base, and acute apex. Ray leaves broadly ovate, (0.7) 0.9–1.8 cm × 0.5–1.1 (1.2) cm, (1) 1.1–2.1 (2.7) times longer than wide, widest at 0.4–0.6 (0.7) of their length. Raylet leaves broadly ovate to reniform, 0.6–1 (1.3) cm × 0.8–1.4 (1.5) cm, (0.6) 0.7–0.8 (0.9) times longer than wide, widest at (0.1) 0.2–0.5 (0.5) of their length, with cordate base, and obtuse apex. Cyathial involucre campanulate, (1.6) 1.8–2.6 (2.7) mm × (1.4) 1.7–2.7 (3.2) mm, (0.8) 0.9–1.4 (1.6) times longer than wide. Cyathial lobes usually pubescent. Cyathial glands obovate-truncate, (0.6) 0.7–1.3 (1.9) mm × (0.9) 1–1.8 (2.3) mm, (0.4) 0.6–0.8 (1) times longer than wide, usually with two lobate horns, with emargination/horn length (0.1) 0.2–0.5 mm. Fruits glabrous, pruinose-papillose, broadly ovate, (3.6) 3.7–4.4 (4.6) mm × (2.3) 2.4–3.1 mm, 1.3–1.6 (1.8) times longer that wide, styles (0.7) 1.3–1.6-mm long. Seeds ovoid, smooth, yellowish, brownish or greyish, 2.6–2.7 mm × 1.6–1.9 (2) mm, (1.3) 1.4–1.6 (1.7) times longer than wide. Caruncle conical, (0.5) 0.6–0.7 mm × 0.8–1 mm, 0.6–0.8 times longer than wide.

Distribution: central and northern Apennine Peninsula to the southern margin of the Alps (Italy), northwest Balkan Peninsula (Croatia: Istria and Kvarner, western Slovenia).

Habitat: submediterranean grasslands, scrublands, open forests, and rocky outcrops, mostly over calcareous substrate but also on serpentine.

Etymology: We name this species after the Adriatic Sea, on both sides of which it is distributed.

(2) ***Euphorbia hercegovina*** **Beck** in Glasn. Zemaljsk. Muz. Bosni Hercegovini 32: 95. 1920 ≡ *E. barrelieri* var. *hercegovina* (Beck) Hayek in Repert. Spec. Nov. Regni Veg. Beih. 30: 133. 1924 ≡ *E. barrelieri* subsp. *hercegovina* (Beck) Kuzmanov in Izv. Bot. Inst. (Sofia) 12: 120. 1963 ≡ *Tithymalus barrelieri* subsp. *hercegovinus* (Beck) Soják, Cas. Nár. Mus., Odd. Prír. 140: 170. 1972. — Lectotype (Geltman, 2009, p. 184): [Bosnia and Herzegovina] "Hercegovina, auf dem Leotar," 7 June 1894, *B[eck] s.n.* (PRC 456036!)[6].

Description: Glabrous and glaucous perennial, (6) 8–17 (22)-cm high, with 1–2-mm thick stems. Terminal rays (4) 5–7 (8), the longest (1.2) 1.9–6.3 (7.9)-cm long, 1–2 times dichotomously branched. Fertile axillary rays (1) 3–6, the longest (1.7) 2.2–6.6 (9.2)-cm long. All leaves with entire margin. Cauline leaves linear-oblanceolate, (0.7) 1.1–2.1 (2.7) cm × (0.2) 0.4–1.1 (1.3) cm, (1.1) 1.6–4.1 (5.4) times longer than wide, widest at (0.5) 0.6–0.8 (0.9) of their length, with cuneate base and obtuse apex. Ray leaves ovate-lanceolate to obovate-lanceolate, (0.8) 0.9–1.4 (1.7) cm × 0.6–1.2 (2) cm, (0.7) 0.8–1.5 (2) times longer than wide, widest at (0.2) 0.3–0.5 (0.6) of their length. Raylet leaves broadly ovate, 0.5–0.9 cm × (0.7) 0.8–1.5 (1.7) cm, (0.4) 0.5–0.7 (0.8) times longer than wide, widest at 0.2–0.3 (0.4) of their length, with cordate base and rounded to obtuse apex. Cyathial involucre campanulate, (1.3) 1.6–2.5 mm × (1.5) 1.7–2.5 (2.6) mm, (0.6) 0.8–1.2 times longer than wide. Cyathial lobes mostly pubescent. Cyathial glands obovate-truncate, (0.6) 0.7–1.4 (1.5) mm × (1) 1.2–1.9 (2.2) mm, 0.4–0.9 (1.1) times longer than wide, usually with two horns, with emargination/horn length of 0.2–0.4 (0.5) mm. Fruits glabrous, pruinose-papillose, broadly ovate, (4.6) 4.7–5.2 mm × (2.8) 2.9–3.7 (3.8) mm, 1.4–1.6 (1.7) times longer than wide; styles, 1.4–1.6-mm long. Seeds ovoid to ellipsoid, smooth, yellowish-brown or grey, 2.5–2.8 mm × 1.6–1.8 (1.9) mm, (1.4) 1.5–1.6 times longer than wide. Caruncle conical, 0.6–0.7 mm × (0.7) 0.8–0.9 (1) mm, 0.7–0.8 (0.9) times longer than wide.

Distribution: central Balkan Peninsula (Bosnia and Herzegovina, Croatia, Montenegro).

Habitat: on dolomitic substrate in rocky and gravely grasslands, garrigues, and open scrublands, pine forests.

(3) ***Euphorbia japygica*** **Ten.**, Fl. Napol. 4: 266. 1830 ≡ *E. nicaeensis* var. *japygica* (Ten.) Nyman, Consp. Fl. Eur.: 653. 1881 ≡ *E. nicaeensis* subsp. *japygica* (Ten.) Arcang., Comp. Fl. Ital.: 620. 1882 ≡ *E. seguieriana* var. *japygica* (Ten.) Fiori, Fl. Italia 2: 286. 1901 ≡ *Tithymalus nicaeensis* subsp. *japygicus* (Ten.) Soják, Cas. Nár. Mus., Odd. Prír. 140: 174. 1972. — Type: not designated (not in FI, NAP nor RO).

Note: According to Fenu et al. (2016), *E. japygica* differs from *E. adriatica* in the fruits being hairy, whereas, in the latter, the fruits are glabrous. The only specimen of the former taxon included in our morphometric study (No. 58), in addition to having hairy fruits, also had larger fruits and longer styles than any measured specimen of *E. adriatica*, but additional morphometric studies are needed to clarify the morphological divergence between both taxa to generate a description for *E. japygica* and clarify its taxonomic status; its treatment as species in this paper should thus be seen as preliminary. It should also be examined whether all populations of this taxon are polyploid and whether the Albanian and some North Macedonian populations treated as *E. nicaeensis* (Barina, 2017) or *E. glareosa* (Micevski, 1998) belong to this species. Also, a type specimen needs to be designated, as there is no original material available in herbaria FI, NAP, and RO, where Tenore's specimens are deposited. In addition, despite citing "Flora napolitana tav. 232. A" in description of *E. japygica* (Tenore, 1830, p. 266), on the illustration nr. 232, *E. esuloides* Ten. is depicted, and there are no further indications (e.g., in indices of Flora Napolitana) of existence of an illustration of *E. japygica* that could potentially serve as a lectotype.

Distribution: southern Apennine Peninsula (Italy: Basilicata, Puglia, doubtful in Campania; Fenu et al., 2016).

Habitat: arid grasslands and garrigues up to 1,000 m.

(4) ***Euphorbia macroclada*** **Boiss.**, Diagn. Pl. Orient., sér. 1, 5: 54. 1844 ≡ *Tithymalus macrocladus* (Boiss.) Klotzsch et Garcke in Abh. Königl. Akad. Wiss. Berlin 1859: 97.1860. — Lectotype (Khan, 1964, p. 119): [Turkey], "Denisleh [Denizli]

[6]https://prc.jacq.org/PRC456036

ad collibus argillosis," June 1842, *Boissier s.n.*, (G-BOIS barcode G00754301 – image!).

= *E. schizoceras* Boiss., Diagn. Pl. Orient., sér. 1, 5: 54. 1844 ≡ *Tithymalus schizoceras* (Boiss.) Klotzsch et Garcke in Abh. Königl. Akad. Wiss. Berlin 1859: 98. 1860 ≡ *E. tinctoria* Boiss. et Huet var. *schizoceras* (Boiss.) Boiss., in DC., Prodr. 15, 2: 166. 1862 — Lectotype (Geltman, 2015, p. 130): "Kurdistan, Berg Gara," 3 August [1841], *Th. Kotschy570* (G-BOIS barcode G00754277 – image!; isolectotypes: BM 000951553, G00313291, G00390389, G00421116 – image!, K 001080072, LE 01071163, 01071180).

= *E. lorentii* Hochst. in J. A. Lorent, Wanderungen im Morgenland. . .: 344. 1845. —Type, unknown (not at TUB!). Locality indicated in the protologue: [Syria] "bei Latakia."

= *E. syspirensis* K. Koch in Linnaea, 21: 727. 1848 ≡ *Tithymalus syspirensis* (K. Koch) Klotzsch et Garcke in Abh. Königl. Akad. Wiss. Berlin 1859: 97. 1860. — Type, unknown. Locality indicated in the protologue: "Im Gaue Sber auf Porphyr und Kieselschiefer, c. 4,000' hoch."

= *E. damascena* Boiss., Diagn. Pl. Orient., sér. 1, 12: 113. 1853 ≡ *Tithymalus damascenus* (Boiss.) Klotzsch et Garcke in Abh. Königl. Akad. Wiss. Berlin, 1859: 96. 1860. — Lectotype (Geltman, 2020): "Syria, Damasci collis," May–July 1846, *E. Boissier* (G-BOISS barcode G00754311 – image!).

= *E. noeana* Boiss. in DC., Prodr. 15, 2: 166. 1862. Pro syn.: "in pl. Noé exs."

= *E. tinctoria* Boiss. et Huet, in DC., Prodr. 15, 2: 166. 1862. — Lectotype (Geltman, 2006: 164): [Turkey], "Elmali, Gémichem quine dans les ravins," 9 July 1860, *Bourgeau598*," (G-BOISS barcode G00754304 – image!; isolectotypes: G00390388, G-DC G00313297 – image!, MPU014638 – image!).

= *E. macroclada* var. *aceras* Hand.-Mazz. in Ann. K. K. Naturhist. Hofmus. 26: 140. 1912. — Lectotype (designated here): [Mesopotanien-Expedition des naturwissenschaftl. Orientvereines in Wien; Turkey], "Kurdistania occidentalis: Taurus Cataonicus. Inter urbem Malatja et vicum Kjachta, in lapidosis et glareosis inter Kory et Furendscha, substrato calcareo, 1,200–1,900 m," 19 July 1910, *H. F. Handel-Mazzetti2492*. (WU 046588 — image!).

Description: Glabrous or pubescent perennial, (18) 29–62 (69)-cm high, with (3) 4–5 (6)-mm thick stems. Terminal rays (3) 5–8 (9), the longest (1.6) 4.1–9.7 (12)-cm long, 1–2 (3) times dichotomously branched. Fertile axillary rays (1) 4–13 (15), the longest (2.7) 4.9–10.7 (12.9)-cm long. All leaves with entire margin. Cauline leaves lanceolate to oblanceolate, (3) 4–6.3 (7.6) × (0.6) 0.7–1.6 (1.9) cm, (2.9) 3.6–6.2 (7) times longer than wide, widest at (0.4) 0.5–0.7 (0.8) of their length, with cuneate base and acute apex. Ray leaves broadly ovate to obovate, (1.1) 1.2–2.6 (3.6) cm × (0.8) 1–1.8 (3) cm, (0.8) 0.9–1.7 (3) times longer than wide, widest at 0.2–0.5 (0.6) of their length. Raylet leaves broadly ovate to reniform, 0.7–1.2 (1.8) cm × 1–1.9 (2.6) cm, (0.5) 0.6–0.8 (1.1) times longer than wide, widest at (0.1) 0.2–0.4 (0.5) of their length, with cordate base and rounded to mucronate apex. Cyathial involucre campanulate, (1.4) 1.8–3.1 (3.5) mm × (2.2) 2.4–3.6 (3.8) mm, (0.4) 0.6–1.2 (1.3) times longer than wide. Cyathial lobes usually densely pubescent. Cyathial glands obovate-truncate, (0.7) 1–2 (2.4) mm × (1.3) 1.5–2.7 (3.3) mm, (0.5) 0.6–0.9 (1.1) times longer than wide, usually with two, often lobate or multifidi, horns, with gland emargination/horn length (0.1) 0.2–0.7 (0.8) mm. Fruits glabrous or pubescent, pruinose-papillose, broadly ovate, (5.9) 6.1–8.4 (8.8) mm × (3.1) 4.1–5.7 (6.2) mm, (1.2) 1.4–1.6 (1.9) times longer than wide, styles (1.3) 1.6–2.5 (2.8)-mm long. Seeds ovoid to ellipsoid, smooth, white, yellow or brown, (2.9) 3.1–4 (4.2) mm × (1.9) 2–2.5 (2.6) mm, (1.2) 1.4–1.9 (2) times longer than wide. Caruncle conical, (0.8) 0.9–1 mm × 1.1–1.2 (1.3) mm, 0.7–0.9 times longer than wide.

$2n$ = 18 (Lessani and Chariat-Panahi, 1979; Chariat-Panahi et al., 1982; Fasihi et al., 2016), 36 (Genç and Kültür, 2020).

Distribution: Anatolia (Turkey), Armenian Highlands (Armenia), Iranian Plateau (Iran), Levant (Iraq, Israel, Jordan, Lebanon, Syria); Irano-Turanian element.

Habitat: stony steppes and scrubland, semideserts, fallow fields, mostly in mountaneous areas.

(5) ***Euphorbia nicaeensis*** **All.**, Fl. Pedem. 1: 285. 1785 ≡ *Galarhoeus nicaeensis* (All.) Haw., Syn. Pl. Succ.: 144. 1812 ≡ *Tithymalus nicaeensis* (All.) Klotzsch and Garcke in Abh. Königl. Akad. Wiss. Berlin, 1859: 89. 1860. ≡ *Esula nicaeensis* (All.) Fourr. in Ann. Soc. Linn. Lyon, n.s., 17: 150. 1869 ≡ *Euphorbia seguieriana* var. *nicaeensis* (All.) Fiori in A. Fiori and al., Fl. Italia 2: 286. 1901. – Lectotype (designated here): Herbarium Allioni, *Euphorbia nicaeensis*. (TO, s.n. – image!).

Note: There is only a single specimen of *E. nicaeensis* in the herbarium of Allioni at TO. Despite the fact that there were no locality data stated on the label, we considered the specimen as a part of the original material and selected it here as a lectotype. The locality cited in the protologue is the following: [France], In Comitatu Nicaeensi [Nice] inter Cimie [Cimiez], and la Trinità [La Trinité].

= *E. nicaeensis* var. *aurasiaca* Maire in Bull. Soc. Hist. Nat. Afrique N. 31: 40. 1940. – Lectotype (designated here): B. Balansa, Pl. D'Algerie, 1853 "Lambčse, dans les bois," 16 July 1853, *Balansa1005* [sub *E. luteola*] (MPU004312 – Image!).

= *E. demnatensis* Coss. ex Batt. in J. A. Battandier and L. C. Trabut, Fl. Algérie, Dicot.: 802. 1890 ≡ *E. nicaeensis* var. *demnatensis* (Coss. ex Batt.) Maire in Mém. Soc. Sci. Nat. Maroc 7: 178. 1924. ≡ *E. nicaeensis* subsp. *demnatensis* (Coss. ex Batt..) Breistr. in?. – Lectotype (designated here): Socété dauphinoise, 1883, "Djebel Bouachfal, prov. de Demnat (Maroc)," 3 August 1882, *Ibrahim 4007* (P00540548 – Image!).

= *E. dasycarpa* Coss. ex Batt. in J. A. Battandier and L. C. Trabut, Fl. Algérie, Dicot.: 802. 1890 ≡ *E. nicaeensis* All. var. *dasycarpa* (Coss. ex Batt. and Trab.) Maire in Mém. Soc. Sci. Nat. Maroc 7: 178. 1924. – Lectotype (designated here): "In monte Djebel Afougueur ad austro-occidentem Urbis Maroc, Imperio Maroccano," 1 June 1876, *Ibrahim* (P05546285 – Image!).

= *E. nicaeensis* var. *hebecarpa* DC. in J. B. A. M. de Lamarck and A. P. de Candolle, Fl. Franç., ed. 3, 5: 363. 1815. – Lectotype (designated here): "Euphorbia myrsinites L., Pyr. or.

[= Pyrénées orientales]." 1814, *J. Coder 226.* (G00313242 – Image!). Note: despite the fact that de Candolle labelled this specimen as "Euphorbia nicaeensis γ" and, in the protologue, γ corresponds to var. *salzmanii*, the collector of the specimen "Coder" corresponds to the indication in the protologue for *E. nicaeensis* var. *hebecarpa*. Moreover, the fruits of the specimen are slightly pilose ("capsulis pilosiusculis").

= *E. nicaeensis* var. *oleifolia* DC. in J. B. A. M. de Lamarck and A. P. de Candolle, Fl. Franç., ed. 3, 5: 363. 1815. – Lectotype (designated here): "Euphorb. oleafolia Gouan. Castelnau," 1807, *Dufour s.n.* (G00313234 – Image!).

= *E. nicaeensis* var. *salzmanii* DC. in J. B. A. M. de Lamarck and A. P. de Candolle, Fl. Franç., ed. 3, 5: 363 (1815). – Lectotype (designated here): "Euphorbia nicaeensis var. involllanceol. Gravels," 1810, *Salzman s.n.* (G00313235 – Image!). Note: despite the fact that de Candolle labelled this specimen as "Euphorbia nicaeensis δ" and, in the protologue, δ corresponds to var. *hebecarpa*, the locality "Gravels" written on the label corresponds to "Grabels prés Montpellier" indicated in the protologue for *E. nicaeensis* var. *salzmanii*. Moreover, Salzman is also indicated as a collector in the protologue, and the fruits of the specimen are glabrous ("capsulis glabris").

= *E. nicaeensis* subsp. *hispanica* Degen and Hervier in Bull. Acad. Int. Géogr. Bot. 17: 205. 1907. ≡ *E. nicaeensis* var. *hispanica* (Degen and Hervier) Cuatrec. in Trab. Mus. Ci. Nat. Barcelona 12: 354. 1929. – Lectotype (designated by Font Garcia et al., 1997): "Barranco del Rio Segura, lieux arides et calcaries, 1,500 m," July 1906, *Reverchon 1162* (MA-01-00075510 – Image!).

Description: Glaborus or pubescent perennial, (5) 15–33 (39)-cm high, with 2–4-mm thick stems. Terminal rays (5) 6–12 (13), the longest (1.2) 2.8–7.1 (8.1)-cm long, 1–2 times dichotomously branched. Fertile axillary rays (1) 2–9 (10), the longest (1.2) 2–7.2 (8.1)-cm long. All leaves with entire margin. Cauline leaves linear-oblanceolate, (1.7) 2.2–3.8 (5) cm × (0.5) 0.6–1 cm, (2.3) 3.1–5.4 (6.5) times longer than wide, widest at 0.6–0.8 of their length, with cuneate base and acute apex. Ray leaves broadly ovate to obovate, (0.7) 1.2–2 (3) cm × (0.5) 0.9–1.5 (1.6) cm, (0.8) 1–2.2 (2.6) times longer than wide, widest at (0.2) 0.3–0.6 (0.7) of their length. Raylet leaves broadly ovate to reniform, (0.5) 0.7–1.2 (1.3) cm × (0.4) 1–1.6 (1.7) cm, (0.4) 0.6–1.1 (1.9) times longer than wide, widest at (0.1) 0.2–0.4 (0.5) of their length, with cordate base and rounded to obtuse apex. Cyathial involucre campanulate, (1.4) 1.7–3 (3.2) mm × (1.3) 1.6–3 (3.3) mm, (0.7) 0.8–1.4 (1.7) times longer than wide. Cyathial lobes are usually pubescent. Cyathial glands obovate-truncate, (0.7) 0.8–1.5 (1.7) mm × (0.9) 1.1–1.9 (2.1) mm, (0.4) 0.6–0.9 (1) times longer than wide, usually with two, often lobate, horns, with emargination/horn length (0.1) 0.2–0.5 (0.6) mm. Fruits glabrous or pubescent, pruinose-papillose, broadly ovate, (4) 4.3–5.7 (6) mm × (2.7) 3.1–4.1 (4.2) mm, 1.3–1.5 (1.6) times longer than wide, styles (0.9) 1.9–2.2-mm long. Seeds, ovoid, smooth, yellowish-brown or grey, 2.7–3.1 (3.3) mm × 1.8–2.1 mm, 1.5–1.7 (1.8) times longer than wide. Caruncle conical, (0.5) 0.6–0.8 mm × (0.8) 0.9–1.1 mm, 0.6–0.8 times longer than wide.

$2n = 18$ (Perry, 1943; Löve, 1978). The chromosome count of $2n = 56$ by Vilatersana and Bernal (1992) is likely erroneous, as it is their count of $2n = 40$ for *E. seguieriana* (see Frajman et al., 2019), and it might be due to inappropriate fixation used in the study (R. Vilatersana, written communication to B. Frajman, 2.1.2018).

Distribution: northern Algeria and Morocco, Iberian Peninsula (Spain and Portugal) and southern France.

Habitat: mountainous areas of the western Mediterranean Basin.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repositories and accession numbers can be found in the article/**Supplementary Material**. The RADseq data are available in the NCBI Short Read Archive (SRA; BioProject ID PRJNA761526, BioSample accessions SRR15817339-SRR15817453) and the ITS sequences in GenBank (GenBank numbers are listed in **Supplementary Table 1**).

AUTHOR CONTRIBUTIONS

BF conceived and designed the study, collected the plant material, coordinated the lab work, performed the data analyses, and wrote substantial parts of the manuscript. VS performed parts of the lab work, morphometric measurements, data analyses, and wrote substantial parts of the manuscript. EZ coordinated the lab work connected to RAD sequences and performed all analyses of RADseq data, wrote corresponding parts of the manuscript, and commented on other parts of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was financially supported by the Tiroler Wissenschaftsförderung (TWF, Grant 0404/1642 to BF), a Synthesis Grant (ES-TAF-6700 to BF), and by the France-focus and the Italy-centre of the University of Innsbruck.

ACKNOWLEDGMENTS

We thank the curators of the herbaria BCN, BEO, BEOU, BOZ, FI, FRU, G, HFLA, M, MA, MPU, NAP, P, PRC, RO, TO, TUB, W, WHB, and WU abbreviations according to (http://sweetgum.nybg.org/science/ih/), as well as W. Gutermann for providing herbarium material and information about the types and all collectors listed in **Supplementary Table 1** for help with the collection of samples. D. Pirkebner, L. Silbernagl,

and A. Dudaš helped with laboratory work. L. Silbernagl and A. Dudaš prepared some figure drafts. C. Gilli provided some photos, and G. Nieto Feliner, D. Geltman, and R. Riina helped with the literature survey, and the last two gave feedback on the revised taxonomic treatment of *E. nicaeensis*. P. Schönswetter improved the previous version of this manuscript. We are grateful to M. Bodner, P. Daniel Schlorhaufer, M. Imhiavan, and their colleagues from the Botanical Gardens of the University of Innsbruck for successfully cultivating our living collection of *Euphorbia*.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.815379/full#supplementary-material

REFERENCES

Abràmoff, M. D., Magalhães, P. J., and Ram, S. J. (2004). Image processing with image. *J. Biophotonics* 11, 36–41.

Barina, Z. (2017). *Distribution Atlas of Vascular Plants in Albania*. Budapest: Hungarian Natural History Museum.

Beck-Mannagetta, G. (1920). Flora Bosne, Hercegovine i bivšeg Sandžaka Novog Pazara 2. *Glas. Zem. Muz. Bosne Herceg.* 32, 395–410.

Bilton, D. T., Mirol, P. M., Mascheretti, S., Fredga, K., Zima, J., and Searle, J. B. (1998). Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proc. Royal Soc. B Biol. Sci.* 265:1402. doi: 10.1098/rspb.1998.0423

Blondel, J., and Aronson, J. (1999). *Biology and Wildlife of the Mediterranean Region*. New York, NY: Oxford University Press.

Blondel, J., Aronson, J., Bodiou, J.-Y., and Boeuf, G. (2010). *The Mediterranean Basin – Biological Diversity in Space and Time*. Oxford: Oxford University Press.

Boissier, P. (1867). *Flora Orientalis*. Geneva: Basel: H. Georg.

Bouckaert, R. R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537

Bryant, D., Bouckaert, R. R., Felsenstein, J., Rosenberg, N. A., and Choudhury, R. A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* 29, 1917–1932. doi: 10.1093/molbev/mss086

Bucci, G., Gonzalez-Martinez, S. C., Le Provost, G., Plomion, C., Ribeiro, M. M., Sebastiani, F., et al. (2007). Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. Revealed by chloroplast microsatellite markers. *Mol. Ecol.* 16:10. doi: 10.1111/j.1365-294X.2007.03275.x

Cai, L., Xi, Z., Moriarty Lemmon, E., Lemmon, A. R., Mast, A., Buddenhagen, C. E., et al. (2021). The perfect storm: gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient Angiosperm clade. *Malpighiales. Syst. Biol.* 70, 491–507. doi: 10.1093/sysbio/syaa083

Caković, D., Cresti, L., Stešević, D., Schönswetter, P., and Frajman, B. (2021). High genetic and morphological diversification of the *Euphorbia verrucosa* alliance (Euphorbiaceae) in the Balkan and Iberian peninsulas. *Taxon* 70, 286–307. doi: 10.1002/tax.12427

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3* 1, 171–182. doi: 10.1534/g3.111.000240

Catchen, J. M., Hohenlohe, P., Bassham, S., Amores, A., and Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354

Chariat-Panahi, M. S., Lessani, H., and Cartier, D. (1982). Etude caryologique de quelques espèces de la flore l'Iran. *Rev. Cytol. Biol. Veg. Bot.* 5, 189–197.

Coupland, R. T. (1993). "Overview of the grasslands of Europa and Asia," in *Ecosystems of the world: Eastern Hemisphere and résumé*, Vol. 8B, ed. R. T. Coupland (Amsterdam: Elsevier), 1–2.

Cresti, L., Schönswetter, P., Peruzzi, L., Barfuss, M. H. J., and Frajman, B. (2019). Pleistocene survival in three *Mediterranean refugia*: origin and diversification of the Italian endemic *Euphorbia gasparrinii* from the *E. verrucosa* alliance (Euphorbiaceae). *Bot. J. Linn. Soc.* 189, 262–280. doi: 10.1093/botlinnean/boy082

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Davis, P. H. (1971). "Distribution patterns in Anatolia with particular reference to endemism," in *In Plant Life of South-West Asia*, ed. P. H. Davis (Edinburgh: Royal Botanic Garden), 15–27.

Davis, P. H., and Hedge, I. C. (1975). The flora of Turkey: past, present and future. *Candollea* 30, 331–351.

Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.

Du Pasquier, P.-E., Jeanmonod, D., and Naciri, Y. (2017). Morphological convergence in the recently diversified *Silene gigantea* complex (Caryophyllaceae) in the Balkan Peninsula and south-western Turkey, with the description of a new subspecies. *Bot. J. Linn. Soc.* 183, 474–493.

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

Falch, M., Schönswetter, P., and Frajman, B. (2019). Both vicariance and dispersal have shaped the genetic structure of Eastern Mediterranean *Euphorbia myrsinites* (Euphorbiaceae). *Perspect. Plant Ecol. Evol. Syst.* 39:125459. doi: 10.1016/j.ppees.2019.125459

Fasihi, J., Zarre, S., Azani, N., and Salmaki, Y. (2016). Karyotype analysis and new chromosome numbers of some species of *Euphorbia* L. (Euphorbiaceae) in Iran. *Iran. J. Bot.* 22, 65–71.

Fenu, G., Bacchetta, G., Bernardo, L., Calvia, G., Citterio, S., Foggi, B., et al. (2016). Global and regional IUCN red list assessments: 2. *Ital. Bot.* 2, 93–115. doi: 10.3897/italianbotanist.2.10975

Font Garcia, J., Vilar Sais, L., Villar Pérez, L., Castroviejo, S., Vargas, P., Frost-Olsen, P., et al. (1997). Notulae taxinomicae, chorologicae, nomenclaturales, bibliographicae aut philologicae in opus "Flora iberica" intendentes. *Anales Jardín Bot. Madrid* 55, 189–200.

Frajman, B., and Oxelman, B. (2007). Reticulate phylogenetics and phytogeographical structure of *Heliosperma* (Sileneae, Caryophyllaceae) inferred from chloroplast and nuclear DNA sequences. *Mol. Phylogenet. Evol.* 43, 140–155. doi: 10.1016/j.ympev.2006.11.003

Frajman, B., and Schönswetter, P. (2011). Giants and dwarfs: molecular phylogenies reveal multiple origins of annual spurges within *Euphorbia subg. Esula. Mol. Phylogenet. Evol* 61, 413–424. doi: 10.1016/j.ympev.2011.06.011

Frajman, B., and Schönswetter, P. (2017). Amphi-adriatic distributions in plants revisited: pleistocene trans-Adriatic dispersal in the *Euphorbia barrelieri* group (Euphorbiaceae). *Bot. J. Linn. Soc.* 185, 240–252.

Frajman, B., Záveská, E., Gamisch, A., Moser, T., The Steppe Consortium and Schönswetter, P. (2019). Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Eurasian Euphorbia segueriana s. l. (Euphorbiaceae). *Mol. Phylogenet. Evol.* 134, 238–252. doi: 10.1016/j.ympev.2018.10.046

Gehrke, B., Kandziora, M., and Pirie, M. D. (2016). The evolution of dwarf shrubs in alpine environments: a case study of *Alchemilla* in Africa. *Ann Bot.* 117, 121–131. doi: 10.1093/aob/mcv159

Geltman, D. V. (2006). Lectotypificatio nominum specierum et taxorum intraspecificorum nonnullorum in genere *Euphorbia* L. (Euphorbiaceae). *Novosti Sist. Vyssh. Rast.* 38, 162–164.

Geltman, D. V. (2009). Spurges (*Euphorbia* L., Euphorbiaceae) of the boreal Eurasia. I. Section Paralias Dumort [In Russian with English summary]. *Novosti Sist. Vyssh. Rast.* 41, 166–191.

Geltman, D. V. (2015). Typification of some specific and infraspecific names in *Euphorbia* (Euphorbiaceae). *Novosti Sist. Vyssh. Rast.* 46, 126–133.

Geltman, D. V. (2020). A synopsis of *Euphorbia* (Euphobiaceae) for the Caucasus. *Novosti Sist. Vyssh. Rast.* 51, 43–78.

Genç, I., and Kültür, Ş (2020). Karyological analysis of twelve Euphorbiaspecies from Turkey. *Caryologia* 73, 13–19.

Govaerts, R., Frodin, D. G., and Radcliffe-Smith, A. (2000). *Checklist and bibliography of Euphorbiaceae 2*. Kew: Royal Botanic Gardens.

Greilhuber, J., and Ebert, I. (1994). Genome size variation in Pisum sativum. *Genome* 37, 646–655. doi: 10.1139/g94-092

Greuter, W. (1991). Botanical diversity, endemism, rarity and extinction in the Mediterranean area: an analysis based on the published volumes of Med - Checklist. *Bot. Chron.* 10, 63–79.

Greuter, W., Burdet, H. M., and Long, G. (1986). *Med-Checklist 3*. Geneve: Med-Checklist trust of OPTIMA.

Hamasha, H. N., von Hagen, B., and Röser, M. (2012). Stipa (Poaceae) and allies in the Old World: molecular phylogenetics realigns genus circumscription and gives evidence on the origin of American and Australian lineages. *Plant Syst. Evol.* 298, 351–367.

Hayek, A. (1924). Prodromus florae peninsulae Balcanicae. *Repert. Spec. Nov. Regni Veg.* 30, 1–160.

Hedge, I. C., and Wendelbo, P. (1978). Patterns of distribution and endemism in Iran. *Notes R. Bot. Gard. Edinb.* 36, 441–464.

Hegi, G. (1966). *Illustrierte Flora von Mitteleuropa 5.1: Linaceae-Violaceae. 2*. Hamburg: Paul Parey Verlag.

Hewitt, G. M. (1999). Postglacial re-colonisation of European biota. *Biol. J. Linn. Soc.* 68, 87–112. doi: 10.1186/1471-2148-11-215

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405, 907–913. doi: 10.1038/35016000

Hewitt, G. M. (2011). "Mediterranean peninsulas: the evolution of hotspots," in *Biodiversity Hotspots*, eds F. E. Zachos and J. C. Habel (Berlin: Springer), 123–147.

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

Horn, J. W., Xi, Z., Riina, R., Peirson, J. A., Yang, Y., Dorsey, B. L., et al. (2014). Evolutionary bursts in Euphorbia (Euphorbiaceae) are linked with photosynthetic pathway. *Evolution* 68, 3485–3504. doi: 10.1111/evo.12534

Hühn, P., Dillenberger, M. S., Gerschwitz-Eidt, M., Hörandl, E., Los, J. A., Messerschmid, T. F. E., et al. (2022). How challenging RADseq data turned out to favor coalescent-based species tree inference, a case study in Aichryson (Crassulaceae). *Mol. Phylogenet. Evol.* 167:107342. doi: 10.1016/j.ympev.2021.107342

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

Janković, M., and Nikolić, V. (1972). "*Euphorbia* L," in *Flore de la Republique Socialiste de Serbie 3*, ed. M. Josifović (Beograd: Srpska Akademija nauka i umetnosti), 538–567.

Khan, M. S. (1964). Taxonomic revision of Euphorbia in Turkey. Notes Royal Bot. Gard. *Edinburgh* 25, 71–161.

Kirschner, P., Záveská, E., Gamisch, A., Hilpold, A., Trucchi, E., Paun, O., et al. (2020). Long-term isolation of European steppe outposts boosts the biome's conservation value. *Nat. Commun.* 11:1968. doi: 10.1038/s41467-020-15620-2

Koch, M. A., Möbus, J., Klöcker, C. A., Lippert, S., Ruppert, L., and Kiefer, C. (2020). The Quaternary evolutionary history of Bristol rock cress (Arabis scabra, Brassicaceae), a Mediterranean element with an outpost in the north-western Atlantic region. *Ann. Bot.* 126:1. doi: 10.1093/aob/mcaa053

Körner, C. (2012). *Alpine Treelines*. New York, NY: Springer.

Krijgsman, W. (2002). The Mediterranean: mare Nostrum of the earth sciences. *Earth Planet. Sci. Lett.* 205, 1–12.

Kuzmanov, B. (1979). "Euphorbia," in *Flora Reipublicae popularis Bulgaricae 7*, ed. B. Kuzmanov (Sofia: Academia scientiarum Bulgaricae), 118–177.

Kuzmanov, B. A. (1963). Taksonomično proučavanje na vidovite ot rod *Euphorbia* L., rasprostraneni v Balgarija. *Bull. Inst. Botan. Sofia* 12, 101–186.

Lal, R. (2004). Carbon sequestration in soils of Central Asia. *Land Degrad. Dev.* 15, 563–572.

Landis, M., Matzke, N., Mooer, B., and Huelsenbeck, J. (2013). Bayesian analysis of biogeography when the number of area is large. *Syst. Biol.* 62, 789–804. doi: 10.1093/sysbio/syt040

Larcher, W., Kainmüller, C., and Wagner, J. (2010). Survival types of high mountain plants under extreme temperatures. *Flora* 205, 3–18.

Lavrenko, E. M., and Karamysheva, Z. V. (1993). "Steppes of the former Soviet Union and Mongolia," in *Ecosystems of the world 8B: natural grasslands: Eastern Hemisphere and résumé*, ed. R. T. Coupland (Amsterdam: Elsevier), 3–59.

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

Lemmon, E. M., and Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* 44, 99–121. doi: 10.1146/annurev-ecolsys-110512-135822

Lessani, H., and Chariat-Panahi, S. (1979). In IOPB chromosome number reports LXV. *Taxon* 28, 635–636.

Löve, A. (1978). IOPB Chromosome Number Reports LXII. *Taxon* 27, 519–535.

Magri, D., Fineschi, S., Bellarosa, R., Buonamici, A., Sebastiani, F., Schirone, B., et al. (2007). The distribution of *Quercus suber* chloroplast haplotypes matches the palaeogeographical history of the western Mediterranean. *Mol. Ecol.* 16:24. doi: 10.1111/j.1365-294X.2007.03587.x

Mahmoudi Shamsabad, M., Moharrek, F., Assadi, M., and Nieto Feliner, G. (2021). Biogeographic history and diversification patterns in the Irano-Turanian genus *Acanthophyllum* s.l. (Caryophyllaceae). *Plant Biosyst.* 155, 425–435.

Manafzadeh, S., Salvo, G., and Conti, E. (2014). A tale of migrations from east to west: the Irano-Turanian floristic region as a source of Mediterranean xerophytes. *J. Biogeogr.* 41, 366–379.

Manafzadeh, S., Staedler, Y. M., and Conti, E. (2017). Visions of the past and dreams of the future in the Orient: the Irano-Turanian region from classical botany to evolutionary studies. *Biol. Rev.* 92, 1365–1388. doi: 10.1111/brv.12287

Mansion, G., Selvi, F., Guggisberg, A., Conti, E., and Vegetale, B. (2009). Origin of Mediterranean insular endemics in the Boraginales: integrative evidence from molecular dating and ancestral area reconstruction. *J. Biogeogr.* 36, 1282–1296.

Matzke, N. J. (2013). Probabilistic historical biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* 5, 242–248. doi: 10.21425/F5FBG19694

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., and Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenet. Evol.* 66, 526–538. doi: 10.1016/j.ympev.2011.12.007

Micevski, K. (1998). *Flora na Republika Makedonija (The flora of the Republic of Macedonia)*. Skopje: Macedonian Academy of Sciences and Arts.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., de Fonseca, G. A. B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858.

Nieto Feliner, G. (2014). Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspect. Plant Ecol. Evol. Syst.* 16, 265–278.

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

Noss, R. F. (2012). *Forgotten Grasslands of the South: Natural History and Conservation*. Washington, DC: Island Press.

Nylander, J. A. A. (2004). *MrAIC. pl. 1.4. 3. Program Distributed by the Author*.

Ortiz, E. M. (2019). *vcf2phylip v2.0: Convert a VCF Matrix Into Several Matrix Formats for Phylogenetic Analysis*. doi: 10.5281/zenodo.2540861

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R., Moret, B. M., and Stamatakis, A. (2010). How many bootstrap replicates are necessary? *J. Comput. Biol.* 17, 337–354. doi: 10.1089/cmb.2009.0179

Paun, O., Turner, B., Trucchi, E., Munzinger, J., Chase, M. W., and Samuel, R. (2015). Processes driving the adaptive radiation of a tropical tree (Diospyros, Ebenaceae) in New Caledonia, a biodiversity hotspot. *Syst. Biol.* 65, 212–227. doi: 10.1093/sysbio/syv076

Peart, B. (2008). *Life in a Working Landscape: Towards a Conservation Strategy for The World's Temperate Grasslands: Compendium of Regional Templates on the Status of Temperate Grasslands. Conservation and Protection*. Vancouver, BC: IUCN/WCPA.

Peel, M. C., Finlayson, B. L., and McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* 11, 1633–1644. doi: 10.5194/hess-11-1633-2007

Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes* 9:88. doi: 10.3390/genes9020088

Perry, B. A. (1943). Chromosome numbers and phylogenetic relationships in the Euphorbiaceae. *Am. J. Bot.* 30, 527–543.

Petit, R. J., Aguinagalde, I., De Beaulieu, J. L., Bittkau, C., Brewer, S., Cheddadi, R., et al. (2003). Glacial Refugia: hotspots but not melting pots of genetic diversity. *Science* 300, 1563–1565. doi: 10.1126/science.1083264

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., et al. (2006). Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* 16, 1262–1269. doi: 10.1101/gr.5290206

Poldini, L. (1969). Kritische Bemerkungen über die Euphorbia saxatilis-triflora-kerneri Verwandschaft. *Acta Bot. Croat.* 28, 317–328.

Prokhanov, Y. I. (1949). "Genus 856. *Euphorbia* L," in *Flora SSSR*, eds B. K. Shishkin and E. G. Bobrov (Moskva-Leningrad: Akademii Nauk SSSR), 233–378.

Quézel, P. (1985). "Definition of the Mediterranean region and the origin of its flora," in *Plant Conservation in the Mediterranean Area*, ed. C. Gomez-Campo (Dordrecht: W. Junk), 9–24.

Radcliffe-Smith, A. (1982). "*Euphorbia* L," in *Flora of Turkey*, Vol. 7, ed. P. H. Davis (Edinburgh: Edinburgh University Press), 571–630.

Radcliffe-Smith, A., and Tutin, T. G. (1968). "*Euphorbia* L," in *Flora Europaea 2*, eds T. G. Tutin, V. H. Heywood, D. M. Moore, D. H. Valentine, S. M. Walters, and D. A. Webb (Cambridge: Cambridge University Press), 213–226.

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

Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). *Tracer v1.6*. Available online at: http://beast.bio.ed.ac.uk/tracer (accessed October 12, 2021).

Ree, R. K., and Sanmartin, I. (2018). Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *J. Biogeogr.* 45, 741–749. doi: 10.1111/jbi.13173

Ree, R. H., Moore, B. R., Webb, C. O., and Donoghue, M. J. (2005). A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59, 2299–2311. doi: 10.1111/j.0014-3820.2005.tb00940.x

Ree, R. H., and Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57, 4–14. doi: 10.1080/10635150701883881

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

Riina, R., Peirson, J. A., Geltman, D. V., Molero, J., Frajman, B., Pahlevani, A., et al. (2013). A worldwide molecular phylogeny and classification of the leafy spurges, *Euphorbia subgenus* Esula (Euphorbiaceae). *Taxon* 62, 316–342.

Ronquist, F. (1996). *DIVA ver. 1.1. Computer Manual*. Uppsala: Uppsala University.

Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* 46, 195–203. doi: 10.1093/sysbio/46.1.195

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Roquet, C., Sanmartín, I., Garcia-Jacas, N., Sáez, L., Susanna, A., Wikström, N., et al. (2009). Reconstructing the history of Campanulaceae with a Bayesian approach to molecular dating and dispersal-vicariance analyses. *Mol. Phylogenet. Evol.* 52, 575–587. doi: 10.1016/j.ympev.2009.05.014

Rose, J. P., Toledo, C. A. P., Moriarty Lemmon, E., Lemmon, A. R., and Sytsma, K. J. (2021). Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus Polemonium (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Syst. Biol.* 70, 162–180. doi: 10.1093/sysbio/syaa049

RStudio Team (2019). *RStudio: Integrated Development for R*. Boston, MA: RStudio Inc. Available online: http://www.rstudio.com/ (accessed April 20, 2021).

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

Ståhls, G., Vujić, A., Petanidou, T., Cardoso, P., Radenković, S., Ačanski, J., et al. (2016). Phylogeographic patterns of Merodon hoverflies in the Eastern Mediterranean region: revealing connections and barriers. *Ecol. Evol.* 6, 2226–2245. doi: 10.1002/ece3.2021

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stange, M., Sánchez-Villagra, M. R., Salzburger, W., and Matschiner, M. (2018). Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. *Syst. Biol.* 67, 681–699. doi: 10.1093/sysbio/syy006

Stevanoski, I., Kuzmanović, N., Dolenc Koce, J., Schönswetter, P., and Frajman, B. (2020). Disentangling relationships between amphi-Adriatic *Euphorbia spinosa* and Balkan endemic *E. glabriflora* (Euphorbiaceae). *Bot. J. Linn. Soc.* 194, 358–374. doi: 10.1093/botlinnean/boaa032

Suda, J., and Trávníček, P. (2006). Estimation of relative nuclear DNA content in dehydrated plant tissues by flow cytometry. *Curr. Protoc. Cytom.* Chapter 7:Unit7.30. doi: 10.1002/0471142956.cy0730s38

Swofford, D. L. (2002). *Phylogenetic Analysis Using Parsimony*. Sunderland, MA: Sinauer Associate.

Takhtajan, A. (1986). *Floristic Regions of the World*. Berkeley, CA: University of California Press.

Tenore, M. (1830). *Flora Napolitana 2*. Napoli: Stamperia Francese.

Thompson, J. D. (2005). *Plant Evolution in the Mediterranean*. Oxford: Oxford University Press.

Trinajstić, I. (2007). Nomenklaturno-taksonomska i korološka razmatranja o vrsti Euphorbia hercegovina G. Beck. *Hrvatska Misao (Sarajevo)* 30, 82–88.

Vilatersana, R., and Bernal, M. (1992). Mediterranean chromosome number reports – 2. *Flora Mediterr.* 2, 249–255.

Wagner, N. D., He, L., and Hörandl, E. (2020). Phylogenomic relationships and evolution of polyploid Salix species revealed by RAD sequencing data. *Front. Plant Sci.* 11:1077. doi: 10.3389/fpls.2020.01077

Ware, S. (1990). Adaptation to substrate – and lack of it – in rock outcrop plants: sedum and Arenaria. *Amer. J. Bot.* 77:8. doi: 10.2307/2444581

Wesche, K., Ambarlý, D., Kamp, J., Török, P., Treiber, J., and Dengler, J. (2016). The Palearctic steppe biome: a new synthesis. *Biodivers. Conserv.* 25, 2197–2231.

Wilson, J. B., Peet, R. K., Dengler, J., and Pärtel, M. (2012). Plant species richness: the world records. *J. Veg. Sci.* 23, 796–802.

Záveská, E., Maylandt, C., Paun, O., Bertel, C., Frajman, B., The Steppe Consortium et al. (2019). Multiple auto- and allopolyploidisations marked the Pleistocene history of the widespread Eurasian steppe plant *Astragalus onobrychis* (Fabaceae). *Mol. Phylogenet. Evol.* 139:106572. doi: 10.1016/j.ympev.2019.106572

Zohary, M. (1973). *Geobotanical Foundations of the Middle East (Volume I and II)*. Amsterdam: Gustav Fischer Verlag, Stuttgart and Swets & Zeitlinger.

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 July 2022
doi: 10.3389/fpls.2022.695958

Convergent Morphological Evolution in *Silene* Sect. *Italicae* (Caryophyllaceae) in the Mediterranean Basin

Yamama Naciri[1,2†], Zeynep Toprak[1,2,3†], Honor C. Prentice[4], Laetitia Hugot[5], Angelo Troia[6], Concetta Burgarella[7], Josep Lluis Gradaille[8] and Daniel Jeanmonod[1,2]*

[1] Unité Systématique et Médiation, Conservatoire et Jardin botaniques de Genève, Geneva, Switzerland, [2] Plant Systematics and Biodiversity Laboratory, Department of Botany and Plant Biology, University of Geneva, Geneva, Switzerland, [3] Department of Molecular Biology and Genetic, Faculty of Sciences, University of Dicle, Diyarbakir, Turkey, [4] Department of Biology, Lund University, Lund, Sweden, [5] Conservatoire botanique national de Corse, Office de l'Environment de la Corse, Corte, France, [6] Dipartimento di Scienze e Tecnologie Biologiche, Chimiche e Farmaceutiche, Università degli Studi di Palermo, Palermo, Italy, [7] Department of Organismal Biology, Uppsala University, Uppsala, Sweden, [8] Jardí Botànic de Sóller, Palma, Spain

OPEN ACCESS

Edited by:
Gonzalo Nieto Feliner, Real Jardín Botánico (RJB, CSIC), Spain

Reviewed by:
Andrea Sánchez Meseguer, Royal Botanical Garden (CSIC), Spain
Farzaneh Jafari, University of Tehran, Iran
Mario Mairal, Stellenbosch University, South Africa

***Correspondence:**
Yamama Naciri
Yamama.Naciri@ville-ge.ch

[†] These authors have contributed equally to this work and share first authorship

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

Received: 15 April 2021
Accepted: 06 June 2022
Published: 12 July 2022

Citation:
Naciri Y, Toprak Z, Prentice HC, Hugot L, Troia A, Burgarella C, Gradaille JL and Jeanmonod D (2022) Convergent Morphological Evolution in Silene Sect. Italicae (Caryophyllaceae) in the Mediterranean Basin. Front. Plant Sci. 13:695958. doi: 10.3389/fpls.2022.695958

Recent divergence can obscure species boundaries among closely related taxa. *Silene* section *Italicae* (Caryophyllaceae) has been taxonomically controversial, with about 30 species described. We investigate species delimitation within this section using 500 specimens sequenced for one nuclear and two plastid markers. Despite the use of a small number of genes, the large number of sequenced samples allowed confident delimitation of 50% of the species. The delimitation of other species (e.g., *Silene nemoralis*, *S. nodulosa* and *S. andryalifolia*) was more challenging. We confirmed that seven of the ten chasmophyte species in the section are not related to each other but are, instead, genetically closer to geographically nearby species belonging to *Italicae* yet growing in open habitats. Adaptation to chasmophytic habitats therefore appears to have occurred independently, as a result of convergent evolution within the group. Species from the Western Mediterranean Basin showed more conflicting species boundaries than species from the Eastern Mediterranean Basin, where there are fewer but better-delimited species. Significant positive correlations were found between an estimation of the effective population size of the taxa and their extent of occurrence (EOO) or area of occupancy (AOO), and negative but non-significant correlations between the former and the posterior probability (PP) of the corresponding clades. These correlations might suggest a lower impact of incomplete lineage sorting in species with low effective population sizes and small distributional ranges compared with that in species inhabiting large areas. Finally, we confirmed that *S. italica* and *S. nemoralis* are distinct species, that *S. nemoralis* might furthermore include two different species and that *S. velutina* from Corsica and *S. hicesiae* from the Lipari Islands are sister species.

Keywords: species tree, species delimitation, ITS, *trn*H-*psb*A, *trn*S-*trn*G, chasmophyte, adaptation, coalescence

INTRODUCTION

Silene L. is the largest genus in the family Caryophyllaceae (Greuter, 1995; Melzheimer, 1988; Bittrich, 1993) with about 870 species described around the world (Jafari et al., 2020). The genus is native to all continents apart from Australia and Antarctica, and has two main hotspots of diversity. The first, hosting more than half of the known species, is located in the Mediterranean area and the second is found in South-West Asia (Zohary, 1973; Polunin, 1980; Greuter, 1995). There are also two main centers of endemism in the Mediterranean area: Turkey and Greece, with 45 and 38% of endemic species, respectively (Coode and Cullen, 1967; Davis, 1971; Greuter, 1997; Trigas et al., 2007; Yıldız et al., 2009; Yıldız and Çırpıcı, 2013).

According to the latest revisions, the genus *Silene* is divided into three subgenera; *S.* subg. *Silene*, *S.* subg. *Behenantha* (Otth) Endl. and *S.* subg. *Lychnis* (L.) Greuter (Greuter, 1995; Jafari et al., 2020). The delimitation of the first two subgenera is supported by several earlier phylogenetic studies (e.g., Oxelman et al., 2001; Petri and Oxelman, 2011; Aydin et al., 2014a). Uncertainties, however, exist at lower taxonomic levels and the many co-existing sectional, subsectional, group and series rankings illustrate the taxonomic complexity of the genus. *Silene* section *Siphonomorpha* Otth is an example of this complexity within subg. *Silene*. The section has had a range of different taxonomic circumscriptions, and Naciri et al. (2017) showed that section *Siphonomorpha*, as defined by Coode and Cullen (1967) and Yıldız and Çırpıcı (2013), is not supported phylogenetically. Naciri et al. (2017) therefore suggested extending *Siphonomorpha* to include members of 13 previously defined sections (*Saxifragoideae*, *Coronatae*, *Tataricae*, *Chloranthae*, *Barbeyanae*, *Nanosilene*, *Otites*, *Koreanae*, *Brachypodae*, *Graminiformes*, *Longitubulosae*, *Dianthoidea*, *Holopetalae*). Within the "extended *Siphonomorpha*" group of Naciri et al. (2017), three sections were found to be monophyletic, with high support: *Paradoxae* Greuter with four species (possibly five with *S. schwarzenbergerii*), *Italicae* (Rohrb.) Schischk., with 31 species (including, for example, *S. italica* (L.) Pers. and *S. nemoralis* Waldst. and Kit.), and *Siphonomorpha* sensu Greuter (1995) with nine species, including *S. nutans* (**Table 1**). A fourth clade, *Giganteae,* comprising a unique species [*S. gigantea* (L) L.], still awaits a formal sectional description (Du Pasquier, 2016; Du Pasquier et al., 2017). *Italicae*, *Paradoxae*, *Siphonomorpha* and *Giganteae* are genetically close to each other but their exact phylogenetic relationships are still not well-resolved (Naciri et al., 2017; Jafari et al., 2020). Although the analysis by Jafari et al. (2020) covers a high number of *Silene* species and specimens, and further extended the definition of the *Siphonomorpha* section, the study only included around half of the species from *Italicae*, *Paradoxae*, *Siphonomorpha* and *Giganteae*. In the present study, we use the classification of Naciri et al. (2017), because the focus of the study is put on interspecific relationships within the section *Italicae*, rather than on the relationships between *Italicae* and other sections. *Italicae*, *Giganteae*, *Siphonomorpha* s.s and *Paradoxae* therefore refer to the sections defined in Naciri et al. (2017). Our study includes 73% of all species from *Italicae*, *Paradoxae*, *Siphonomorpha* and *Giganteae* (84, 80, 22, and 100%, respectively).

The species within *Italicae* are distributed around the Mediterranean area. All the species studied so far are diploids, with a chromosome number of $2n = 24$ (Jeanmonod, 1984a,b, 1985a; Chater et al., 1993; Yıldız and Çırpıcı, 2013) and they are hermaphrodite to gynodioecious. The western species of the section are distributed from Morocco to Tunisia in North Africa and from Portugal to Italy in Europe (Jeanmonod, 1984a,b, 1985a,b). The eastern species grow from the Balkan Peninsula, Greece, Western Turkey and Cyprus to Syria and, possibly, Lebanon (Du Pasquier et al., 2017). Despite several recent studies (Yıldız and Çırpıcı, 2013; Du Pasquier et al., 2017; Naciri et al., 2017) the speciation and diversification patterns of the western species of *Italicae* are still under discussion. Moreover, the evolution of the chasmophyte species within the group is poorly understood. Jeanmonod (1984a, 1984b) suggested that the seven chasmophytic species of the Western Mediterranean Basin [*S. mollissima* (L.) Pers., *S. velutina* Loisel., *S. tomentosa* Otth, *S. hicesiae* Brullo and Signor., *S. hifacensis* Rouy, *S. andryalifolia* Pomel and *S. auriculifolia* Pomel] were all derived from a chasmophytic ancestor that diverged from the ancestor of the remaining *Italicae* species during, or shortly after, the Messinian Crisis, around 5 million years ago (MYA; Gautier et al., 1994). These taxa are all characterized, among other traits, by woody and long-lived roots, numerous sterile basal rosettes (that are present at anthesis), and condensed inflorescences in a verticillaster with many flowers and short internodes. Jeanmonod (1984a, 1984b) referred to these seven western chasmophytes as the *Mollissima* group. Three additional chasmophytic species were subsequently added to this group: *S. badaroi* Brestr. (= *S. tyrrhenia* Jeanm. and Boquet; Jeanmonod and Bocquet, 1983) in South France and Italy, *S. oenotriae* Brullo in South Italy (Brullo, 1998) and *S. gazulensis* A. Galán et al. in Spain (Galán de Mera et al., 1999). Du Pasquier et al. (2017) reported that *S. gigantea*, the sole member of *Giganteae*, which is a sister group to *Italicae*, started to diversify as recently as 1.2 MYA. The results of Naciri et al. (2017) revealed that the diversification of *Italicae* was somewhat older than that of *Giganteae*—but less than 5 MYA which contradicts Jeanmonod (1984a,b) hypothesis about the divergence time of the *Mollissima* group.

Another explanation for the morphological similarity of the *Mollissima* species is that selection within harsh, rocky habitats has led to convergent evolution, with selection for similar morphologies within different lineages. Du Pasquier et al. (2017) showed that the chasmophytic characters associated with *S. gigantea* ssp. *gigantea* (i.e., flowers condensed in verticillasters and rosette leaves remaining green throughout the flowering period), appeared several times in the course of evolution. Đurović et al. (2017) showed within the *Silene saxifraga* alliance that similar morphologies were found in different, geographically shaped lineages that did not coincide with taxonomic identities. Among other processes, they highlighted the role of convergent morphological evolution to explain this pattern. Consequently, the similarities among the species of the *Mollissima* group may be explained either by a common ancestry or by convergent evolution, and resolution of the

TABLE 1 | List of the sampled species, their chasmophytic status, their distribution area and the number of specimens sequenced for each.

Clade	Species	Chasmophyte	Nb	Countries (Regions)
Italicae	*S. italica* (L.) Pers.	No	78	France, Italy, Greece, Turkey, Syria, Lebanon, Iran
	S. nemoralis Waldst. and Kit.	No	54	Germany, Austria, Romania, Check Republic, Slovakia, Spain, France, Italy, Serbia
	S. badaroi Brestr.	**Yes**	3	Southern France, Italy
	S. mellifera Boiss. and Reut.	No	11	Spain
	S. coutinhoi Rothm. and P. Silva	No	3	Spain
	S. gazulensis A. Galán and al.	**Yes**	3	Spain
	S. fernandezii Jeanm.	No	4	Southern Spain
	S. hifacensis Rouy	**Yes**	35	Spain (Balearic Islands, Valencia)
	S. mollissima (L.) Pers.	**Yes**	13	Spain (Balearic Islands)
	S. tomentosa Otth	**Yes**	6	Gibraltar
	S. andryalifolia Pomel	**Yes**	31	Northern Algeria, Northern Morocco, South Spain
	S. patula Desf.	No	53	Northern Algeria, Morocco, Tunisia
	S. rosulata Soy.-Will. and Godr.	No (on sands)	1	Algeria, Morocco
	S. auriculifolia Pomel	**Yes**	1	Algeria
	S. longicilia (Brot.) Otth	No	3	Portugal
	S. rothmaleri P. Silva	No	-	Portugal
	S. hicesiae Brullo and Signor.	**Yes**	17	Italy (Lipari Islands)
	S. oenotriaee Brullo	**Yes**	2	Southern Italy
	S. nodulosa Viv.	No	8	France (Corsica), Italy (Sardinia)
	S. velutina Loisel.	**Yes**	28	France (Corsica), Italy (Sardinia)
	S. spinescens Sibth. and Sm.	No	13	Greece (Peloponnesus, Pyrrhus)
	S. sieberi Fenzl	No	15	Greece (Crete)
	S. goulimyi Turrill	No	9	Greece (South Central Peloponnesus)
	S. damboldtinana Greuter and Melzh.	No	11	Northern Greece, South-eastern Albania, Southern Macedonia
	S. cythnia (Halácsy) Walters	No	9	Greece (Aegean islands)
	S. galateae Boiss.	No	5	Cyprus
	S. niederi Boiss.	No	5	North-western Greece
	S. viscariopsis Bornm.	No	-	Macedonia
	S. waldsteinii Griseb.	No	-	Greece, Bulgaria
	S. splendens Boiss.	No	-	Southwestern Turkey
	S. astartes Blanche ex Boiss.	No	-	Lebanon
Giganteae	*S. gigantea* (L.) L.	-	43	Greece, Crete, Turkey, Bulgaria
Paradoxae	*S. paradoxa* L.	-	12	France, Italy, Greece, Croatia, Bosnia, Macedonia
	S. aristidis Pomel	-	9	Algeria
	S. fruticosa L.	-	5	Sicily, Peloponnesus, Cyprus
	S. sessionis Batt.	-	-	Algeria
	S. confertiflora Chowdhuri	-	2	Turkey, Syria
Siphonomorpha	*S. nutans* L.	-	4	Spain, France, Germany, Italy, Switzerland
	S. viridiflora L.	-	4	Corsica, Romania, Bulgaria, Greece
	S. berthelotiana Webb ex H. Christ	-	-	Canary Islands
	S. bourgaei Webb ex H. Christ.	-	-	Canary Islands
	S. lagunensis Chr. Sm. ex Link	-	-	Canary Islands
	S. nocteolens Webb and Berthel.	-	-	Canary Islands
	S. ponogocalyx (Svent.) Bramwell	-	-	Canary Islands
	S. tamaranae Bramwell	-	-	Canary Islands
	S. sabinosae Pit.	-	-	Canary Islands
Total	33 species		500	

Nb, number of sequenced samples; species that belong to one of the studied groups but that could not be included in the analysis due to a lack of material are indicated with "-."

phylogenetic relationships among all the species of the section *Italicae* is required to be able to understand the evolution of the chasmophytic taxa.

An additional interesting feature of the species of *Italicae* is related to their population sizes and distributional ranges. Naciri et al. (2017) showed, on the basis of a relatively small

sample size, that the species-boundaries within the western group of taxa appear to be less distinct than within the eastern group of species. The western species have, on average, wider distributional ranges than the eastern ones and it may be assumed that this correlates with higher average population effective sizes. The larger population sizes within species the more incomplete lineage sorting may be expected—blurring species boundaries (Naciri and Linder, 2015) and providing a possible explanation for the apparently fuzzier delimitation of the western species.

The present study is based on 500 specimens from *Italicae*, *Paradoxae*, *Siphonomorpha* and *Giganteae*, as defined in Naciri et al. (2017), sequenced for two plastid intergenic spacers and one nuclear ribosomal marker. The sequence data were analyzed in a Bayesian framework using the Multi-Species Coalescent (MSC) approach and we characterized each species' distributional range using known occurrences from herbarium specimens. The study had three central aims: (1) Which relationships do the species of the *Mollissima* group have within *Italicae*—do they cluster into a single group, or do they represent unrelated species whose morphological resemblance is due to convergent evolution in rocky habitats? (2) Are species boundaries indeed less distinct in the Western Mediterranean Basin (WMB) than in the Eastern Basin (EMB) using a large sample of species and individuals? (3) Does incomplete lineage sorting (ILS) explain the observed pattern of phylogenetic variation?

MATERIALS AND METHODS

Plant Material and Molecular Data

A total of 500 samples providing a general coverage of the distributional areas of the studied groups was used. The main sampling effort focused on *Italicae* species with 421 specimens over 500 (84.2%). The remaining specimens belong to *Giganteae* (43; 8.6%), *Paradoxae* (28; 5.6%) and *Siphonomorpha* s.s. (8; 1.6%) as outgroups. The DNA samples included in the study were extracted from the Geneva herbarium (G) specimens, except for *S. hicesiae, S. hifacensis*, *S. velutina* and *S. mollissima* for which leaf samples preserved in silica gel were used. **Table 1** provides a list of the species included in the study and the number of individuals that were sequenced for each of them. For widespread species, individuals were selected from different populations so as to represent the taxon's geographic range. When available, existing information obtained in the laboratory of the Conservatoire et Jardin botaniques de Genève (CJBG) on intraspecific genetic variation was used. We selected individuals that represent the range of variation within the species of *Silene patula* (Naciri et al., 2010) and *S. paradoxa*, *S. aristidis* and *S. fruticosa* (Leuzinger et al., 2015). The results of Prentice et al. (2003) were used to select the individuals that would be subsequently sequenced in *S. hifacensis*. Additional information about voucher specimens preserved in G, is presented in **Supplementary Table 1** together with Genbank numbers.

DNA sequencing was carried out as described in Naciri et al. (2010) and Naciri et al. (2017). The plastid DNA intergenic regions *trn*H-*psb*A and *trn*S-*trn*G, and the nuclear ribosomal Internal Transcribed Spacer (ITS) were sequenced for all individuals in the CJBG laboratory and the resulting sequence data set included no missing data (**Supplementary Table 1**). The locus *trn*H-*psb*A was selected as it is one of the most diverse plastid ones in Angiosperms and was accordingly suggested as a suitable barcode for plants in addition to *mat*K and *rbc*L that are less diverse within species (Hollingsworth et al., 2011). The plastid *trn*S-*trn*G was also shown to be informative and polymorphic in many species of *Silene* and was used extensively in our laboratory (Naciri et al., 2010; Leuzinger et al., 2015; Du Pasquier et al., 2017 among others). The sequences produced and used for the first time (522) represented 34.8% of the total included in this study (*trn*H-*psb*A: 43.8%; *trn*S-*trn*G: 46.0%; ITS: 14.6%).

Sequence Alignments

Multiple alignments were performed using MUSCLE implemented in Geneious version 6.1.8. (available from https://www.geneious.com) with the default options. For the three loci, we first performed the alignment on a small subset of sequences (100), then the remaining sequences were added to the existing alignment using MAFFT version 7 (Katoh et al., 2002; Katoh and Standley, 2013). Each set of the alignment was visually checked to make sure that nucleotides were consistently aligned, and alignments were manually adjusted if necessary. The two plastid loci *trn*S-*trn*G and *trn*H-*psb*A were concatenated in Geneious. Indels were coded using the approach of Simmons and Ochoterena (2000), applying the principle of "simple gap coding," that consists in accounting for indels by replacing them in the alignment and imposing transversions instead (see Naciri et al., 2017). Substitution models were chosen on the basis of the AICc model selection criterion (Akaike's information criterion) by running the software jModelTest2 0.1 0.10 v20160303 (Darriba et al., 2012) for both ITS and the concatenated plastid dataset. For the latter dataset jModeltest was used before and after indel coding to check for its impact on substitution model.

Phylogenetic Analyses Under the Multi-Species Coalescent

We used the program STACEY implemented in BEAST2 (Jones et al., 2015; Jones, 2017) to estimate the species tree under the multispecies coalescent model, with no *a priori* assignment of sequences to the putative species (Drummond and Rambaut, 2007; Drummond et al., 2012; Bouckaert et al., 2014[1]).

Two input files were prepared using the STACEY template (Jones et al., 2015; Jones, 2017) implemented in BEAUTi2 (Bouckaert et al., 2014) including all the genes but with different priors. In order to minimize possible convergence problems, a starting tree obtained from the ITS data matrix by running the analyses for 50 million iterations in BEAST2, was manually added into both input xml files. For the first one, we set a general time reversible model with a gamma distribution prior and invariable sites (GTR + I + G) as the substitution model for the ITS region. For the concatenated plastid loci, we set a three-parameter model with a gamma distribution, and invariable sites (TPM1uf + I + G), with 4.0 as the "gamma category count" and

[1]https://www.beast2.org/

TABLE 2 | Composition of the 16 final groups within *Italicae*, ranked from top to down in **Figure 1** and SDA-clusters in **Figure 2**, with their associated ThetaS as an estimation of θ, extend of occurrence (EOO) and area of occupancy (AOO) in km^2 as computed in GeoCAT.

Italicae	SDA-clusters	Species	Nb	Regions	PP*	EOO	EOOcor	AOO	ThetaS
Group 1	1, 2 and 3	*S. patula* + ***S. andryalifolia*** + *S. rosulata* + ***S. auriculifolia*** + ***S. tomentosa*** + ***S. gazulensis***	95	Western Mediterranean Basin	0.94	752′383	547′117 (27%)	600	3.512
Group 2	4	*S. coutinhoi*	3	Western Mediterranean Basin	0.50	250′593	250′593 (-)	588	1.333
Group 3	5	*S. mellifera*	10	Western Mediterranean Basin	0.99	271′209	264′561 (2%)	2′104	1.061
Group 4	6	***S. hifacensis*** + ***S. mollissima***	48	Western Mediterranean Basin	0.16	15′125	4′354 (71%)	172	0.901
Group 5	7	*S. nodulosa*[1] + *S. nemoralis*[2]	22	Western Mediterranean Basin	0.37	531′420	267′733[£] (50%) 259′007[$] (51%)	216	1.646
Group 6	8	*S. nemoralis*[3]	35	Eastern Europe	0.83	554′004	554′004 (-)	56	0.486
Group 7	9	*S. fernandezii*	5	Western Mediterranean Basin	0.94	842	842 (-)	96	0.480
Group 8	10	*S. nodulosa*[4]	5	Western Mediterranean Basin	0.99	5′920	5′579 (6%)	228	0.960
Group 9	11	***S. hicesiae*** + ***S. velutina***	45	Western Mediterranean Basin	0.94	2′004	2′004 (-)	132	2.287
Group 10	12	*S. longicilia*	3	Western Mediterranean Basin	0.96	11′625	10′474 (10%)	260	1.333
Group 11	13	*S. italica* + *S. damboldtiana* + ***S. badaroi*** + ***S. oenotriae***	94	Western and Eastern Mediterranean Basin	0.81	4′248′212	2′838′218 (33%)	3,740	2.932
Group 12	14	*S. cythnia*	9	Eastern Mediterranean Basin	0.92	26′865	2′462 (91%)	192	0.368
Group 13	15	*S. goulimyi*	9	Eastern Mediterranean Basin	1.00	915	915 (-)	40	0.000
Group 14	16	*S. niederi*	5	Eastern Mediterranean Basin	1.00	24′592	22′020 (10%)	80	0.000
Group 15	17	*S. spinescens* + *S. sieberi*	28	Eastern Mediterranean Basin	0.98	80′938	25′860 (68%)	368	2.827
Group 16	18	*S. galateae*	5	Eastern Mediterranean Basin	1.00	2′112	2′112 (-)	36	0.000

EOOcor were corrected for maritime areas, with the correction applied within brackets.
The PP values associated to the clades of **Figure 1.*
[1]S. nodulosa from Sardinia, Italy (3 specimens); [2]S. nemoralis from France, Italy, Spain and Bosnia (19 specimens); [3]S. nemoralis from Austria, Romania and Slovakia (35 specimens);[4]S. nodulosa from Corsica, France (5 specimens).
[§]Corrected EOO excluding Corsica which is taken into account in Group 8 with the Corsican S. nodulosa.
The chasmophyte species are given in bold.

1.0 as the "shape." We used an uncorrelated lognormal relaxed clock for both regions (mean 1.0; standard deviation 1.25) and a gamma distribution with default options on their standard deviations. The mean substitution rate of ITS was fixed to 1.0. For the plastid loci, the ploidy level was set to half (1.0) that of ITS (2.0). We set the species tree prior with the default settings. A log normal prior (mean 4.6, standard deviation 2.0) was selected for the species tree growth rate (bdcGrowthRate.t:Species). The collapse weight parameter that controls the number of possible clusters in the data set (ω; see Jones et al., 2015; Jones, 2017) was set to a Beta distribution (alpha and beta = 1.0) and a lognormal prior (mean = −7.0; standard deviation = 2.0) was set on the "popPriorScale" for population sizes. Substitution parameters for the plastid and nuclear markers were specified with an exponential distribution using the default options. A Beta prior with alpha = 1.0 and beta = 8.0 was set for the relative death parameter "relativeDeathRate.t:Species". The second input file differed from the first one by the mean and standard deviation (1.0; 1.25, respectively) of the lognormal prior set on the "popPriorScale" for population sizes and gamma distributions specified with the default values for the substitution parameters.

The two input files were analyzed on several personal computers due to restrictions for long runs on the University cluster. For both input files, three independent chains were run for lengths of 2.1 billion generations each, by logging every 50,000 tracelog, and 2,000,000 trees. For each of the two input files the resulting tree files and log files were combined using LogCombiner after discarding the first 10% of the iterations as burn-in. We used the program Tracer version v1.6.0 (Rambaut and Drummond, 2007; Rambaut et al., 2013) to evaluate the convergence of each parameter in each run. The combined tree files were processed with Tree Annotator version 2.4.4 (Bouckaert et al., 2014). The species tree as well as the individual gene trees were visualized using the FigTree version 1.4.3 (Rambaut, 2009[2]).

Species Delimitation

After the STACEY analyses, the species tree file was processed with the Species Delimitation Analyzer tool (SDA, see Jones et al., 2015; Jones, 2017) to estimate the content of the minimal SDA-clusters, using the following parameters: no burn-in as it was already taken into account when combining the three independent runs, 1.10^{-4} collapse height, and 1.0 similarity cut off. SDA generates a two dimensional similarity matrix, where posterior probabilities of sequence pairs belonging to the same minimal cluster are estimated. To visualize the similarity matrix and the associated SDA-clusters, we used the R code provided by Jones et al. (2015) and Jones (2017) modified by Simon Crameri. This code allows for automated pairwise similarity matrix sorting, labeling, and line drawing. We used an automatic threshold of PP = 0.01 to delineate adjacent SDA-clusters considered as putative species. SDA-clusters therefore generally contain all individuals that fall within the same clade and that have pairwise

[2]http://beast.bio.ed.ac.uk

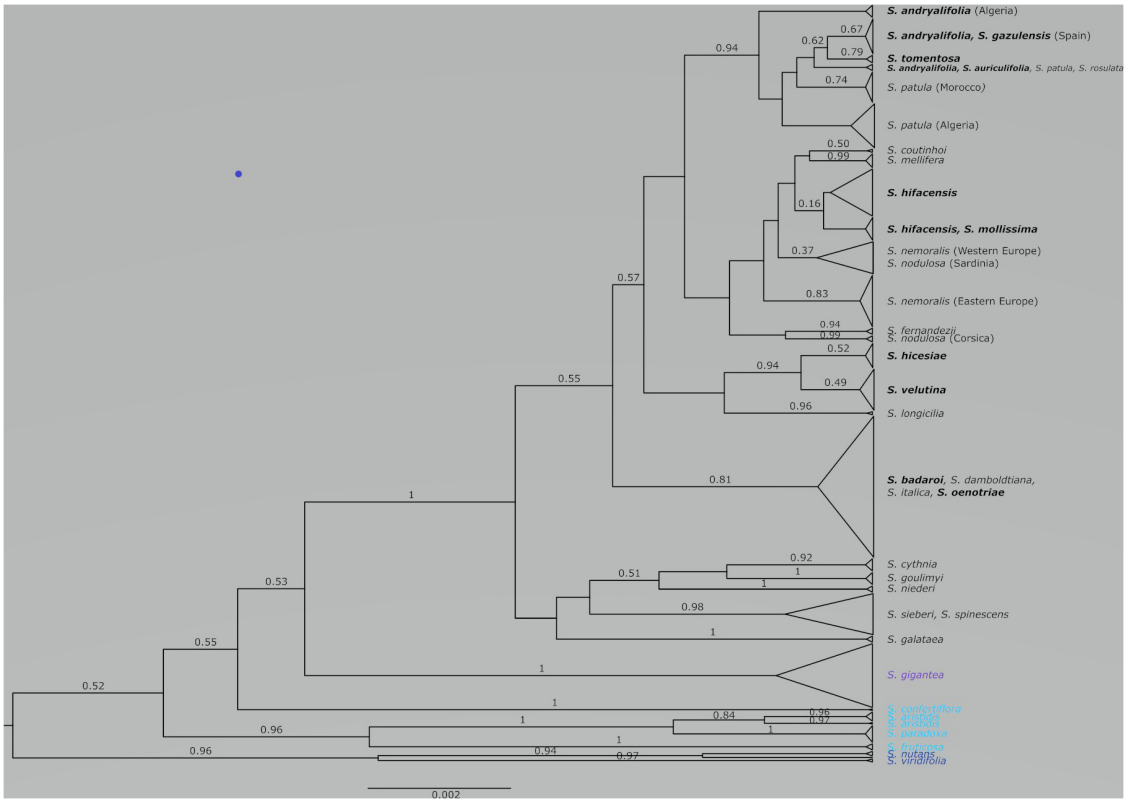


FIGURE 1 | Species tree obtained from the combination of three runs using STACEY on two linked plastid regions (*trn*H-*psb*A and *trn*S-*trn*G) and one nuclear marker (ITS). With some exceptions, posterior probabilities that are equal or higher than 0.50 are given above the corresponding branches. The scale at bottom is given in substitutions/site. The chasmophytic species are indicated in bold. The species belonging to *Italicae* are shown in black whereas the species included in *Giganteae*, *Paradoxae* and *Siphonomorpha* s.s. are shown in violet, blue and dark blue, respectively.

PP-values higher than 0.01 to belong to the same cluster. We then adjusted the number of final clusters (called Groups hereafter) to take into account the fact that SDA-clusters may represent genetic structure within species as well as interspecific differentiation [see Sukumaran and Knowles (2017)]. As recommended by Jones et al. (2015), we used a combination of clade supports in the species tree and of the pairwise similarity matrix to aid further considerations about species memberships. The program was run in R version 3.3.1 (R Development Core Team, 2014) and the code is available on GitHub repository[3].

Estimation of Distributional Ranges

Species or Group distributional ranges were estimated using the extent of occurrence (EOO) and area of occupancy (AOO). EOO is defined as the area covered by the minimum convex polygon encompassing all the known sites of occurrence of a given taxon. It might include large areas of obviously unsuitable habitats. The AOO of a taxon is defined as the area within its EOO for which its presence is certified by a record of occurrence according to a given grid. The measure reflects the fact that a taxon will not usually occur throughout its EOO but is, however, highly dependent on the sampling effort. EOO and AOO for each Group were estimated using the program GeoCat (Bachman et al., 2011), using a square mesh of 2 km^2. We used curated georeferenced specimens taken from the GBIF repository[4] as input data in GeoCat (see **Supplementary Files**). When the Groups contained two or more species, the areas of the species were calculated for all species together. In order to consider only terrestrial surfaces and avoid biases among continental species and those with a very fragmented insular distribution, the EOO of the concerned species was corrected to remove marine areas within the distributional range as follows: the polygon containing all the occurrences (EOO) produced in GeoCAT was superimposed on the official European administrative boundaries data from Eurostat (statistical office of the European Union[5]). The percentage of marine areas was then accurately estimated and the corrected EOO was reduced accordingly. The correction was necessary for two thirds of the groups and are given in **Table 2**. The correlation between the log of distributional ranges (either EOO or AOO) and posterior probabilities (PP) of the clades that best corresponded to each Group was then tested using Pearson's product-moment as well as Spearman and Kendall rank

[3]https://github.com/scrameri/smtools/tree/master/SpeciesDelimitation

[4]https://www.gbif.org/

[5]https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries

correlation tests (Sokal and Rohlf, 1995) in R (R Development Core Team, 2014).

Estimations of Effective Sizes

Species or Group effective population sizes (N_e) were estimated using the metrics θ, a measure of genetic diversity, which equals $2N_e\mu$ for haploid markers, where μ is the mutation rate (Watterson, 1975). The calculation of θ assuming a constant mutation rate across taxa therefore gives access to an estimation of N_e. Using the software Arlequin version 3.5 (Excoffier and Lischer, 2010), we computed θ for each Group using the chloroplast data and the Kimura-2 distance to exclude indels. We used ThetaS as an estimation of θ as it is suitable for non-recombining DNA obtained from the observed number of segregating sites among chloroplast haplotypes.

RESULTS

Sequence Data and Phylogenetic Analyses

The alignment for the 500 ITS rDNA sequences was comprised of 756 nucleotide characters. For the plastid spacers, *trn*H-*psb*A and *trn*S-*trn*G, the alignments represented totals of 210 and 473 nucleotides, respectively. For each of the two input files, the three separate STACEY runs did not converge adequately for all parameters but when combined, the resulting file converged with ESS values higher than 200 for all parameters for the first input file, and higher than 300 for all parameters for the second input file (see **Supplementary Table 2**). Therefore, the combined runs obtained for the second input file were used in the following.

Gene Trees

Apart from a few clades, the ITS and the pooled plastid gene trees showed congruent results at high taxonomic levels in the combined STACEY analyses. The ITS and the pooled plastid gene trees indeed revealed clades that largely correspond with *Italicae*, *Giganteae, Siphonomorpha* and *Paradoxae*. However, the plastid gene tree showed a much better agreement with the species tree topology than the ITS tree, and had much higher supports for most of the internal branches (**Supplementary Figures 1**, **2**). In the ITS gene tree, the only species within *Italicae* that are recovered as highly supported clades are *S. goulimyi*, *S. niederi*, *S. cythnia* and *S. galataea*. In the chloroplast gene tree, *S. patula* forms a highly supported clade with the four chasmophytic species from North Africa and Spain *S. andryalifolia*, *S. auriculifolia*, *S. tomentosa* and *S. gazulensis* as well as with *S. rosulata*. Similarly *S. hifacensis* and *S. mollissima* are clustered with *S. nemoralis* from Eastern Europe with a high support (PP = 0.96).

Species Tree for *Italicae, Giganteae, Paradoxae* and *Siphonomorpha*

The combined species tree obtained from the second input file in STACEY is illustrated in **Figure 1**. *Italicae* (PP = 1), *Giganteae* (PP = 1) and *Siphonomorpha* (PP = 0.96) are recovered as highly supported clades. Within *Paradoxae*, *S. aristidis* (PP = 0.84), *S. fruticosa* (PP = 1), and *S. paradoxa* (PP = 1) are clustered together with a high support (PP = 0.96) but *S. confertiflora* (PP = 1) was external to the cluster. There was poor support for the branches giving rise to *Italicae*, *Giganteae*, *Siphonomorpha* and *Paradoxae* (PP ranging between 0.52 and 0.55).

Species Tree and Species Delimitation Within *Italicae*

Figure 2 presents the result of the species delimitation analyzer based on the species tree in **Figure 1**. The darker the squares, the higher the pairwise posterior probability (PP) that two individuals belong to the same SDA-cluster. The vertical and horizontal lines delineate sets of individuals that have a non-null probability of belonging to the same SDA-cluster. Within this analysis (**Figure 2**), 25 SDA-clusters were recovered in total, of which 18 belong to *Italicae*. Overall, twelve SDA-clusters (4, 5, 9, 12, 14, 15, 16, 18, 21, 22, 23, 24) corresponded, each, to a single species (36% of all analyzed species). Five species (15% of all species) were split into more than one group (nine SDA-clusters) whereas seven SDA-clusters (2, 6, 7, 11, 13, 17, 25) were composed of two to six species (60% of all species). For subsequent analyses we adjusted the number of clusters using the results of **Figures 1**, **2**, keeping in mind that SDA-clusters might reflect genetic differentiation within species. Within *Italicae*, SDA-clusters 1 to 3 were merged into Group 1 (**Table 2**). Indeed *S. andryalifolia* and *S. patula* both fall into SDA-clusters 1–2 and 2–3, respectively. It is also clear from **Figure 2** that SDA-clusters 1 and 2 share genetic affinities with SDA-cluster 3 with non-null pairwise posterior probabilities (PP) values. Outside *Italicae*, SDA-clusters 19 and 20 both contain *S. gigantea* specimens and should be considered as a single Group. For *Italicae*, the 18 SDA-clusters therefore correspond to 16 Groups (see **Table 2**).

Group 1 (PP = 0.94) contains all the North African species in addition to *S. tomentosa* from Gibraltar (PP = 0.79) and *S. andryalifolia* from both North Africa and southern Spain. Group 4 (PP = 0.16) includes *S. hifacensis* from the Balearic Islands and Alicante and *S. mollissima* which is restricted to the Balearic Islands. Group 5 (PP = 0.37) contains the Sardinian specimens of *S. nodulosa* in addition to the Western specimens of *S. nemoralis* from France, Italy and Spain. The Eastern *S. nemoralis* constitute a single group (Group 6; PP = 0.83), whereas the Corsican *S. nodulosa* are clustered in Group 8 with high support (PP = 0.99). Group 9 contains *S. hicesiae* from the Lipari Islands and *S. velutina* from Corsica (PP = 0.94) and Group 11 is comprised of *S. italica*, *S. damboldtiana*, *S. badaroi* (= *S. tyrrhenia*) and *S. oenotriae* (PP = 0.81). All the other groups represent single species (**Table 2**).

Within *Italicae*, the relationships between species and/or Groups remain unresolved.

Chasmophytic Species

Within the 10 chasmophytic species of *Italicae*, those from mainland sites are typically grouped with species of open habitats within the same area. For example, *S. oenotriae*, found on inland cliffs in South Italy and *S. badaroi* (= *S. tyrrhenia*) in

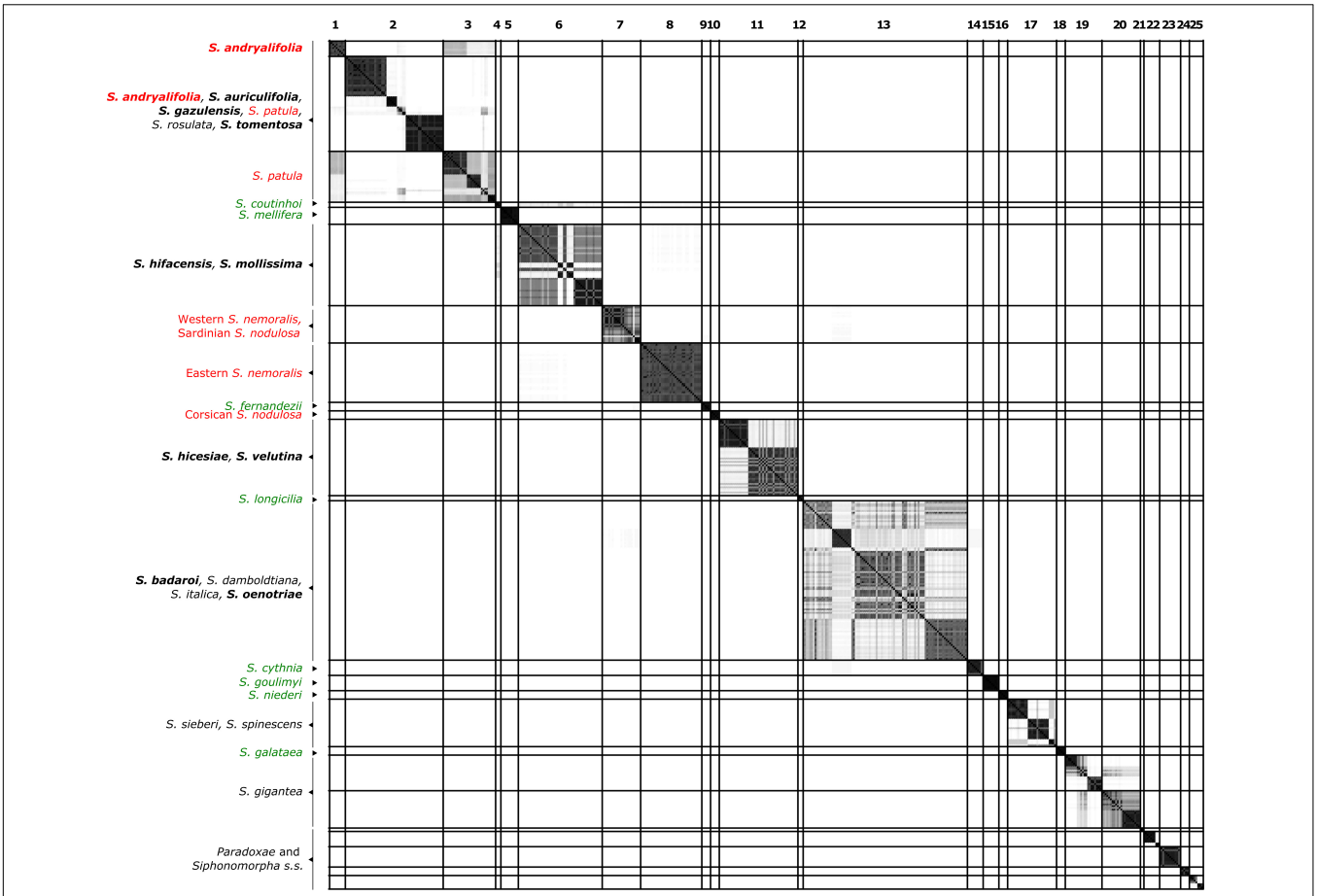


FIGURE 2 | Similarity matrix showing the posterior probability that two individuals belong to the same Multi-Species Coalescent cluster (MSCC) according to the species tree of **Figure 1**. A black square indicates a posterior probability of 1, while no color indicates a posterior probability of 0. Specimens are given in the same order as in **Figure 1**. The lines delimit 26 clusters that are comprised of individuals with pairwise posterior probabilities higher than 0.01. Species found in different clusters are indicated in red, while well-delimited species are colored in green. Species in black are those found with at least one additional species in the same clade. The chasmophytic species are indicated in bold. Here, Sect. *Paradoxae* includes *S. paradoxa*, *S. confertiflora*, *S. fruticosa* and *S. aristidis* and Sect. *Syphonomorpha* s.s. includes *S. nutans* and *S. viridiflora*.

southern France and northwestern Italy on coastal cliffs and rocks, both cluster with *S. italica* which is widely distributed from Spain to Iran and with *S. damboldtiana* from Greece, Albania and Macedonia (Group 11, PP = 0.81; **Table 2**). Similarly, *S. auriculifolia* from Algeria, *S. tomentosa* from Gibraltar, *S. andryalifolia* from North Africa and Southern Spain and *S. gazulensis* from Spain cluster with *S. patula* growing inland in Morocco and Algeria (Group 1, PP = 0.94; **Table 2**). The former group also includes *S. rosulata* which grows on coastal sands in North Africa.

The situation is, however, different for chasmophytic species that grow on islands. *S. velutina* from Corsica and *S. hicesiae* from the Lipari Islands cluster well together (PP = 0.94; **Table 2**) but display no clear relationship to other species inhabiting open habitats, such as *S. nodulosa* the unique other species of *Italicae* present in Corsica. The two species *S. mollissima* (Minorca and Majorca in the Balearic Islands) and *S. hifacensis* (Ibiza in the Balearic Islands and Alicante County in Spain) are grouped in a poorly supported cluster (Group 4, PP = 0.16; **Table 2**) and their relationships to other mainland species are unclear (**Figure 1** and **Supplementary Figures 1**, **2**).

TABLE 3 | Pairwise correlations between the area of occupancy (AOO), the extent of occurrence (EOOCor), the posterior probabilities (PP) and ThetaS as an estimation of θ for the Groups of **Table 2**.

		log(AOO)	Log(EOOCor)	ThetaS
PP	Pearson	-0.111^{ns}	-0.222^{ns}	-0.064^{ns}
	Kendall's tau	-0.233^{ns}	-0.300^{ns}	-0.350^{ns}
	Spearman's rho	-0.350^{ns}	-0.406^{ns}	-0.456^{ns}
log(AOO)	Pearson	–	0.674**	0.667**
	Kendall's tau	–	0.629*	0.583**
	Spearman's rho	–	0.533**	0.779***
log(EOO-Cor)	Pearson			0.539*
	Kendall's tau			0.417*
	Spearman's rho			0.629*

*^{ns} for $P > 0.05$; * for $P > 0.05$; ** for $P < 0.01$ and *** for $P < 0.001$.*

Species Distributions and Effective Sizes

The correlation between the PP values and the log of the areas of occupancy and that of the corrected range extent of occurrence [log(AOO)/log(EOOCor), respectively] were computed for the 16 Groups as defined above and listed in **Table 2**. The maps for the 16 Groups are given in **Supplementary Figure 3** and the corresponding csv files as **Supplementary Files 1–5**. The correction for maritime areas reduced the EOO by 2% (*S. mellifera*) to 91% (*S. cythnia;* **Table 2**). For AOO, the Pearson's product-moment correlation and the two rank tests were negative and non-significant (cor = −0.111, P = 0.682; Kendall's rank correlation tau = −0.233, P = 0.228; Spearman's rank correlation rho = −0.350, P = 0.184). For EOOCor, even more negative but still non-significant correlation coefficients were obtained (cor = −0.222, P = 0.408; tau = −0.300, P = 0.116; rho = −0.406, P = 0.120). The PP values were also negatively correlated with θ estimates (cor = −0.064, P = 0.813; tau = −0.350, P = 0.064; rho = −0.456, P = 0.078). AOO and EOOCor were positively and significantly correlated with θ estimates (**Table 3**).

DISCUSSION

Well Defined Sections but Unclear Relationships Among Them

Clear morphological characterizations of *Italicae*, *Paradoxae*, *Giganteae* and *Siphonomorpha* s.s. have been given by Chowdhuri (1957), Jeanmonod and Mascherpa (1982), Greuter (1995) and Du Pasquier (2016), among others. Eleven species belonging to these four groups (*S. italica, S. nemoralis, S. tomentosa, S. andryalifolia, S. gazulensis, S. rothmaleri, S. waldsteinii*, *S. paradoxa*, *S. fruticosa, S. gigantea* and *S. nutans*) were included in one of the most complete published phylogenies of the Caryophyllaceae based on ITS and five plastid loci (Greenberg and Donoghue, 2011). However, the genetic relationships among these species, as well as the relationships among the different sections remained unresolved in that study. In the latest phylogeny of the genus *Silene*, based on ITS and the plastid loci *rps*16, Jafari et al. (2020) increased the number of species analyzed for the same groups to 22, i.e., ten for *Italicae* (*S. italica, S. nemoralis* from Austria*, S. nodulosa* from Corsica*, S. longicilia, S. andryalifolia, S. hifacensis, S. damboldtiana, S. cythnia, S. viscariopsis, S. waldsteinii*), three for *Paradoxae* (*S. paradoxa*, *S. aristidis*, *S. fruticosa*), and eight for *Siphonomorpha* s.s. in addition to *S. gigantea* with a similar conclusion—no clear relationships emerged among the groups. The present study, however, confirms the consistency of *Italicae*, *Giganteae* and *Siphonomorpha* and, to a lesser extent, *Paradoxae*, as defined by Naciri et al. (2017). *Silene confertiflora* is always recovered as a highly supported clade in our analysis (**Figure 1**). Nevertheless the relationships of this species with sections *Italicae* or *Giganteae* is poorly understood and its placement remains unresolved in our study.

Despite the fact that a much greater number of specimens and species were analyzed in the present study compared to Naciri et al. (2017), the relationships among the four sections remain poorly understood. The low resolution among these sections may reflect a rapid diversification that occurred within a short period of time. In our study, the unclear relationships among sections cannot be attributed to a low analysis convergence because all the parameters gave ESS values above 300 (**Supplementary Table 2**). The parameters that displayed the lowest ESS values and were the latest to converge are the prior (ESS = 302), the birth-death-collapse model (ESS = 310) and the number of clusters (ESS = 345). These parameters are among the ones known to converge sometimes barely with STACEY (Jones, 2017). The birth-death-collapse prior, and the number of clusters are related to populations and species dynamics, which are difficult to estimate, but should not impact the support of deep branches. Jones (2017) pointed out that the birth-death-collapse prior depends strongly on the number of clusters, which might explain that they both exhibit among the lowest ESS values. In our analyses, the number of clusters was not constrained in the priors but the recovered SDA-clusters nevertheless mostly agreed with already described species. Another explanation for the ESS of the prior remaining among the lowest ones could be due to the dataset lacking enough information, i.e., with a too low number of substitutions. Disentangling the two explanations is difficult, however, and including data from more genes, using NGS techniques for instance (Toprak et al., unpublished results) might improve the observed pattern.

A Wide Range of Support for Species Boundaries Within *Italicae*

In STACEY, gene trees and species trees are estimated simultaneously. This is not the case for other methods that require a fixed guide tree and then estimate the fit of that particular tree with the proposed gene trees (Aydin et al., 2014b). Under STACEY, any change to the species tree, or to the different gene trees during the MCMC move, may improve or decrease the concordance between the gene trees and the species tree. The way STACEY produces the species tree, however, minimizes the influence of the prior settings on the outcomes of the analyses. In contrast to programs such as *BEAST, STACEY uses no prior knowledge about samples' assignment to species—reducing the impact of any prior knowledge (or any assignment error) on the results. Our analyses ended up with a variety of results—from genetically well-delimited species to species that present fuzzy genetic boundaries, or species that are divided into different lineages. Generally speaking, our analyses led to a better overall resolution than other studies such as Greenberg and Donoghue (2011) for instance, who used more genes than we did (ITS, *mat*K, *ndh*F, *trn*L-*trn*F, *trn*Q-*rps*16, and *trn*S-*trn*fM). This might be due to several factors: first *trn*H-*psb*A is one of the most diverse chloroplast gene among angiosperms but was seldom used in other *Silene* studies. We furthermore coded the indels which increases the informativeness of the locus. Second, the MSC approach takes into account that genes might be incongruent among them because they have their own independent histories and signal, which might not reflect organismal phylogeny. This means that part of the incongruence among genes is considered as "normal" under the MSC and STACEY deals with that when

estimating the species tree. Third, in this paper, the support is given for putative species not for individual genes. Finally, the fact that we take into account the intraspecific diversity might also explain why we obtained a better resolution.

Within the *Italicae* section, AOO and EOO as well as ThetaS were used as proxies for estimating effective population sizes (N_e). Significant positive correlations were found among those three parameters suggesting that species with low effective sizes indeed inhabit smaller ranges overall. However, we could not confirm that these lower effective sizes translate into higher PP-values as all correlations with PP were negative, as expected, but still non-significant. This lack of significance despite high correlations (as the one between PP and ThetaS found to be −0.456) might have different origins: first, a low test power as only 16 pairs of observations were taken into account; second, the fact that EOO and AOO are imperfect estimates of the species range and occupancy as they depend of the species knowledge and the collection records that can be found in public repositories such as GBIF; and third, the fact that ThetaS is also an imperfect estimation of the species effective size as it assumes a constant mutation rate across Groups and was designed for populations rather than for species or group of species. It would be interesting to test for similar correlations on a higher number of species to confirm on *Silene* the theoretical prediction that small effective population sizes lead to more rapid and complete lineage sorting (Naciri and Linder, 2015).

The Widely Distributed Species *S. italica*, *S. patula* and *S. nemoralis*

Within *Italicae*, three species are widely distributed: *S. italica* (EOO = 3,025,801 km^2), *S. patula* (EOO = 443,117 km^2) and *S. nemoralis* (EOO = 418,617 km^2). *Silene nemoralis* is placed into two different groups that agree well with geography (Group 5 from France, Italy, Spain and Bosnia including *S. nodulosa* from Sardinia and Group 6 from Austria, Romania and Slovakia) that both have moderate to low supports (PP = 0.37 and 0.83, and ThetaS = 1.646 and 0.486, respectively). The relationships between Groups 5 and 6 and other taxa are not well assessed, but the split of *S. nemoralis* into two genetic entities suggests either a high level of incomplete lineage sorting within a single species or the existence of two putative species that diverged recently, one in the Western Mediterranean Basin and the other in Eastern Europe. The assignment of *S. nemoralis* to two different clusters is consistent with the fact that the species has previously been subdivided, on the basis of morphological characters, into three different taxa: two western (*S. crassicaulis* Willk. and Costa in the Pyrenees and *S. nemoralis* var *pedemontana* Burnat and Barbey in the French Maritime Alps and Northern Italy) and one (*S. nemoralis s.str.*) restricted to Eastern Europe (Jeanmonod, 1985b). Therefore we could consider that *S. nemoralis* is the species actually found in Eastern Europe whereas *S. crassicaulis* is that of the Western Mediterranean Basin.

Unlike *S. nemoralis*, *S. italica* and *S. patula* each fall within a single group that includes several species, with moderate to high clade supports (PP = 0.81 and PP = 0.94, respectively). For *S. patula*, Group 1 includes 5 other species (*S. andryalifolia*, *S. auriculifolia*, *S. tomentosa*, *S. gazulensis* and *S. rosulata*), whereas for *S. italica*, Group 11 also includes *S. damboldtiana*, *S. oenotriae* and *S. badaroi*. These two groups include chasmophytic and non-chasmophytic species and their ThetaS estimations are the two highest (3.512 and 2.932, respectively).

Du Pasquier et al. (2017) reported that *S. gigantea* started to diversify around 1.2 MYA. According to **Figure 1**, similar or even younger ages may be inferred for the diversification of Groups 1 and 11. Therefore, if high effective population sizes can be assumed in taxa that, in addition, have had a recent history of diversification, it can be hypothesized that the diffuse species-boundaries within those groups are unlikely to be resolved by simply analyzing a higher number of genes.

The Chasmophytic Species of the *Mollissima* Group: A Case of Morphological Convergence

The *Mollissima* group as defined by Jeanmonod (1984b) is clearly polyphyletic and a common ancestor exclusively for all its species is not found. The switch between open rocky and chasmophytic habitats appears to have occurred several times within *Italicae*: all the chasmophytic species of Southern Spain (*S. andryalifolia*, *S. gazulensis*, *S. tomentosa*) and Northern Africa (*S. auriculifolia* and *S. andryalfolia*) share a common ancestor with *S. patula*, assessed by the high support found for the clade comprising all five species (Group 1; PP = 0.94). This is the first time that such a relationship between chasmophytic species and non-chasmophytic species has been confirmed with molecular data within *Italicae*. Our results also show for the first time that *S. velutina* and *S. hicesiae* diverged from the same ancestor, as assessed by the strong support of the clade including both species (Group 9; PP = 0.94). For that species pair, it is unclear which non-chasmophytic taxon is their closest relative. Similarly, the two species *S. mollissima* and *S. hifacensis* also share a clear genetic similarity according to **Figure 2**, although their group displays no support (Group 4, PP = 0.16). As for the preceding case, it is difficult to link them clearly with any other species although they show some genetic affinities with *S. coutinhoi* (**Figure 2**). The remaining two chasmophytic species (*S. oenotriae* and *S. badaroi*) that were not formally included in the *Mollissima* group of Jeanmonod (1984b) are clearly related to *S. italica* (Group 11, PP = 0.81). The *Mollissima* group has therefore no evolutionary significance in terms of shared ancestry but rather comprises species that seems to have independently adapted to chasmophytic conditions. Cultivation experiments are needed to assess the extent to which, at least in some cases, the morphological differentiation in chasmophytic species may partly reflect phenotypic plasticity. For example, *S. oenotriae* and *S. badaroi* have similar inflorescence characteristics to *S. italica*. When cultivated, *S. badaroi* (= *S. tyrrhenia*) displayed only attenuated characteristics of a chasmophytic plant (Jeanmonod and Bocquet, 1983), suggesting that a degree of plasticity may be involved in the morphological differentiation between the chasmophyte and the non-chasmophyte *S. italica*.

Overall, our results suggest that chasmophytic species, or at least some of them, have derived from geographically neighboring species that occur in open habitats. Chasmophytic species are usually found in restricted rocky places or cliffs and as small populations that are highly susceptible to genetic drift. In the case of *S. oenotriae* and *S. badaroi*, the process of speciation appears to be ongoing and the fact that their distributional ranges overlap with that of *S. italica* suggests that gene flow might remain between the chasmophytes and the non-chasmophyte. Both chasmophytes share the same single plastid haplotype (A1B1) and almost identical ITS sequences to that of *S. italica*. Depending on the species concept used, the two chasmophytic species could be regarded as ecotypes of *S. italica* rather than species. Our results suggest that the ecological divergence may have predated the reproductive isolation of the taxa (Coyne and Orr, 2004; Schemske, 2010). This conclusion would match that of Karrenberg et al. (2018) who demonstrated that in the speciation process between *S. latifolia* and *S. dioica*, extrinsic barriers associated with adaptive ecological divergence had more importance than intrinsic postzygotic barriers in building effective reproductive isolation.

The case of *S. patula*, and its associated chasmophyte taxa is more complex. Within Group 1, the relationships among the different clades are not resolved. From **Figure 2**, it is however noticeable that *S. patula* from Algeria has some genetic affinities with *S. andryalifolia* and *S. auriculifolia* from the same country, whereas *S. andryalifolia* from Spain is clustered with *S. tomentosa* and *S. gazulensis* from Spain. Naciri et al. (2010) showed that Eastern Algeria hosts most of the genetic diversity within *S. patula*, while the Moroccan populations are homogeneous. Genetic homogeneity within Morocco was attributed to a rapid and recent colonization of the country from Algeria, with associated plastid haplotype surfing (Klopfstein et al., 2006). According to this scenario, *S. andryalifolia* from North Africa would have diverged *in situ* before crossing the Straits of Gibraltar, and subsequently giving rise to other chasmophytic species in the Iberian Peninsula. The case of the restricted endemic from Oran, *S. auriculifolia*, is difficult to resolve, as only one accession could be included in the analyses. The species exhibits a single plastid haplotype (A15B16) that is also found in two North Moroccan specimens of *S. patula* with which it clusters, together with one specimen of *S. rosulata* from Alger and one specimen of *S. andryalifolia* from Tizi Ouzou.

Although the whole group of North-African and Iberian species is well supported (PP = 0.94; Group 1), unresolved relationships with the remaining species make it difficult to establish the relationships between this group and other species of *Italicae*. The unresolved relationships for the remaining chasmophytic species (*S. velutina*, *S. hicesiae*, *S. hifacensis*, *S. mollissima*) do not allow suggestions about divergence scenarios. However, the two examples detailed above (*S. italica* and *S. patula* and their allied species), suggest that the chasmophytic form corresponds to a morphological convergence driven by independent adaptation to cliffs or rocky habitats. A similar morphological convergence is also found in *Paradoxae* where *S. aristidis* and *S. sessionis* display chasmophytic characteristics while *S. paradoxa* does not (Leuzinger et al., 2015). Confirming convergent evolution among species from different areas will require the identification of the genes associated with the chasmophytic form and their comparison among species.

Plasticity and morphological convergence are common in *Silene*, perhaps explaining the difficulty encountered when trying to find infra-generic classifications (Naciri et al., 2017; Jafari et al., 2020). For instance, Du Pasquier et al. (2017) highlighted a strong morphological convergence leading to the chasmophytic form in different genetic lineages all described under *S. gigantea* subsp. *gigantea*. Similarly, Đurović et al. (2017) showed that morphological convergence partly explains why taxonomic delineation based on morphology does not match with genetics in the *Silene saxifraga* alliance. As in *S. andryalifolia* or *S. nemoralis* the genetic differentiation is rather geographically than morphologically structured in that group.

Silene italica and *S. nemoralis* Are Two Different Species

Silene nemoralis is one of the few species within *Italicae* (with *S. damboldtiana*) that is biennial. This character discriminates it from *S. italica* with which it has been often lumped. The Flora of Italy (Pignatti, 1982) used to consider *S. nemoralis* as a subspecies of *S. italica* however, in the recent second edition of the Flora (Pignatti et al., 2017), the two taxa are now treated as separate species. Our results clearly confirm that *S. nemoralis* and *S. italica* can be regarded as different species. Their distributional ranges do not overlap, and the species occupy different habitats. *Silene nemoralis* grows in the forest understory and in cool sites whereas *S. italica* is found in open and drier habitats. Because the relationships within *Italicae* are unresolved, it is difficult to say whether the ancestor of *Italicae* was perennial or biennial. A better resolved phylogeny based on more data would be needed to conclude on the ancestral state of perennials versus biennials within *Italicae*.

A High Level of Incomplete Lineage Sorting

This study highlights the impact of incomplete lineage sorting in recently diverging lineages (Naciri and Linder, 2020). For both gene trees (and in the resulting species tree) *Italicae*, *Giganteae*, *Paradoxae* and *Siphonomorpha* s.s. are recovered as highly supported clades but the resolution within the clades is much higher with the chloroplast genes than with ITS. This can be related to the respective effective sizes of the two types of genes—which is twice as high for ITS compared to the chloroplast (a diploid nuclear gene versus haploid chloroplast genes). Therefore, as observed here, more incomplete lineage sorting is expected for ITS than for the chloroplast genes (Naciri and Linder, 2015). Another explanation for conflicting results between ITS and the plastid markers might be the possible incomplete ITS concerted evolution mode after recent hybridization events (Nieto Feliner and Rosselló, 2007) leading to unresolved topologies. In our case, ITS gave almost no resolution at the species level, except for three species of the Eastern Mediterranean basin that have PP values of 1, restricted distribution ranges and no within-species diversity as captured

by ThetaS. This is in contradiction with previous finding in *Silene* (Rautenberg et al., 2010) and with the general expectation that nuclear genes will provide better assignment to species than chloroplast genes (Naciri et al., 2012)—one of the reasons why ITS has been considered for species assignment in addition to the chloroplast barcoding genes (China Plant BOL Group et al., 2011; Hollingsworth et al., 2011). In the case of *Italicae*, it is anticipated that a completely resolved phylogeny might be difficult to obtain, even with a large number of genes, as a result of the confounding effects of large effective population sizes for some species and a recent and fast diversification history, both resulting in a high level of incomplete lineage sorting that blurs species delimitation and relationships (Naciri and Linder, 2015). The relationships among species might be difficult to recover because of the recent radiation of *Italicae* in the Mediterranean region. This hypothesis will, however, have to be tested using a higher number of genes using NGS technologies.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

DJ and YN designed the project. YN produced the data. YN and ZT analyzed the data and wrote the manuscript. HP, DJ, CB, and AT revised the different versions of the manuscript. HP, LH, CB, AT, and JG provided samples for the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work has been supported by the Claraz Foundation to YN. ZT was supported by a grant from the Scientific and Technological Research Council of Turkey (TUBITAK BIDEB ID: 667895). Open access funding was provided by the University of Geneva.

ACKNOWLEDGMENTS

We would like to thank Mats Töpel for allowing us to use the Cluster Albiorix for a previous version of this manuscript and for maintaining it; Laurent Gautier for allowing us to sample from herbarium specimens at G; Nicolas Fumeaux for his help in the G herbarium; Bengt Oxelman and Graham Jones for their theoretical advice; Simon Crameri for the help with the visualization of similarity matrix and making his program available in R; John Cortes for providing samples of *S. tomentosa;* Régine Niba for her help with lab work. We are very grateful to Pascal Martin for his help with the EOO corrections and the maps and to Charles Pouchon for the help with the figures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.695958/full#supplementary-material

Supplementary Figure 1 | Gene tree obtained from the combination of three runs using STACEY on two linked plastid regions (*trn*H-*psb*A and *trn*S-*trn*G). Posterior probabilities that are equal or higher than 0.50 are given above the corresponding branches. The scale at bottom is given in substitutions/site. The chasmophytic species are indicated in bold. For some species, the geographic origin is indicated between brackets. The species belonging to *Italicae* are shown in black whereas the species included in *Giganteae*, *Paradoxae* and *Siphonomorpha* s.s. are shown in violet, blue and dark blue, respectively.

Supplementary Figure 2 | Gene tree obtained from the combination of three runs using STACEY on and one nuclear marker (ITS). Posterior probabilities that are equal or higher than 0.50 are given above the corresponding branches. The scale at bottom is given in substitutions/site. The chasmophytic species are indicated by bold. The species belonging to *Italicae* are shown in black whereas the species included in *Giganteae*, *Paradoxae* and *Siphonomorpha* s.s. are shown in violet, blue and dark blue, respectively.

Supplementary Figure 3 | Distribution maps with the estimations of extents of occurrence (EOO) delimited by a blue line and the occurrences given as red dots according to GeoCAT for the 16 Groups of **Table 2** and the files given as Supplementary data.

Supplementary Table 1 | List of the specimens used in this study together with their collection and lab numbers, the collector name, country of origin and Genbank number accessions for the plastid *trn*H-*psb*A, *trn*S-*trn*G and the nuclear ribosomal Internal Transcribed Spacer (ITS).

Supplementary Table 2 | ESS values for the different parameters in the combined STACEY analysis.

Supplementary File 1 | Coordinates obtained from GBIF for Groups 1, 2, and 3.

Supplementary File 2 | Coordinates obtained from GBIF for Groups 4, 5, 6, and 7.

Supplementary File 3 | Coordinates obtained from GBIF for Groups 8, 9, and 10.

Supplementary File 4 | Coordinates obtained from GBIF for Group 11.

Supplementary File 5 | Coordinates obtained from GBIF for Groups 12, 13, 14, 15, and 16.

REFERENCES

Aydin, Z., Ertekin, A. S., Långström, E., and Oxelman, B. (2014a). A new section of *Silene* (Caryophyllaceae) including a new species from South Anatolia, Turkey. *Phytotaxa* 178, 98–112.

Aydin, Z., Marcussen, T., Ertekin, A. S., and Oxelman, B. (2014b). Marginal likelihood estimate comparisons to obtain optimal species delimitations in *Silene* sect. Cryptoneurae (Caryophyllaceae). *PLoS One* 9:e116266. doi: 10.1371/journal.pone.0106990

Bachman, S., Moat, J., Hill, A. W., de la Torre, J., and Scott, B. (2011). Supporting red List threat assessments with GeoCAT: geospatial conservation assessment tool. ZooKeys 150, 117–126.

Bittrich, V. (1993). "Caryophyllaceae," in *The Families and Genera of Vascular Plants*, Vol. 2, ed. K. Kubitzki (Berlin: Springer-Verlag), 206–236.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comp. Biol.* 10:4. doi: 10.1371/journal.pcbi.1003537

Brullo, S. (1998). *Silene oenotriae* (Caryophyllaceae): a new species from S Italy. *Nordic J. Bot.* 17, 649–652.

Chater, A. O., Walters, S. M., and Akeyrod, J. R. (1993). "*Silene L*," in *Flora Europaea*, 2nd Edn, Vol. 1, eds T. G. Tutin, N. A. Burges, A. O. Chater, J. R. Edmonson, V. H. Heywood, D. M. Moore, et al. (Cambridge: Cambridge University Press), 191–218.

China Plant BOL Group, Li, D.-Z., Gao, L.-M., Li, H.-T., Wang, H., Ge, X.-J., et al. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19641–19646. doi: 10.1073/pnas.1104551108

Chowdhuri, P. K. (1957). Studies in the genus *Silene*. *Notes R. Bot. Gard. Edinburgh* 22, 221–278.

Coode, M. J. E., and Cullen, J. (1967). "*Silene L*," in *Flora of Turkey and the East Aegean Islands*, ed. P. H. Davis (Edinburgh: Edinburgh University Press), 179–242.

Coyne, J. A., and Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109

Davis, P. H. (1971). "Distribution patterns in Turkey with particular reference to Anatolia," in *Plant Life of South West Asia*, eds P. H. Davis, P. C. Harper, and J. C. Hedge (Edinburgh: The Botanical Society of Edinburgh), 15–27.

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214

Du Pasquier, P.-E. (2016). *Etude Taxonomique et Phylogéographique du Groupe du Silene italica (Caryophyllaceae) en Méditerranée Orientale*. Ph.D. thesis. Geneva: University of Geneva.

Du Pasquier, P.-E., Jeanmonod, D., and Naciri, Y. (2017). Morphological convergence in the recently diversified *Silene gigantea* complex (Caryophyllaceae) in the Balkan Peninsula and South-Western Turkey, with the description of a new subspecies. *Bot. J. Linn. Soc.* 183, 474–493.

Đurović, S., Schönswetter, P., Marjan, N., Tomoviæ, G., and Frajman, B. (2017). Disentangling relationships among the members of the *Silene saxifraga* alliance (Caryophyllaceae): phylogenetic structure is geographically rather taxonomically segregated. *Taxon* 66, 343–364.

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Galán de Mera, A., Cortés, J. E., Vicente Orellana, J. A., and Morales Alonso, R. (1999). *Silene gazulensis* sp. nov. (Caryophyllaceae): un nuevo endemismo delentorno delestrecho de Gibraltar. *Acta Bot. Malacitana* 24, 237–241.

Gautier, F., Clauzon, G., Suc, J. P., Cravatte, J., and Violanti, D. (1994). Age and duration of the Messinian salinity crisis. *C.R. Acad. Sci. Paris (IIA)* 318, 1103–1109.

Greenberg, A. K., and Donoghue, M. J. (2011). Molecular systematics and character evolution in Caryophyllaceae. *Taxon* 60, 1637–1652.

Greuter, W. (1995). Studies in Greek Caryophylloideae: *Agrostemma*, *Silene*, and *Vaccaria*. *Willdenowia* 25, 105–142.

Greuter, W. (1997). "*Silene L.*," in *Flora Hellenica*, eds A. Strid and K. Tan (Königstein: Koeltz Scientific Books), 239–323.

Hollingsworth, P. M., Graham, S. W., and Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS One* 6:e19254. doi: 10.1371/journal.pone.0019254

Jafari, F., Zarre, S., Gholipour, A., Eggens, F., Rabeler, R. K., and Oxelman, B. (2020). A new taxonomic backbone for the infrageneric classification of the species-rich genus *Silene* (caryophyllaceae). *Taxon* 69, 337–368.

Jeanmonod, D. (1984a). Révision de la section Siphonomorpha Otth du genre *Silene L.* (Caryophyllaceae) en Méditerranée occidentale. III: aggrégatitalica et espèces affines. *Candollea* 39, 549–639.

Jeanmonod, D. (1984b). Révision de la section Siphonomorpha Otth du genre *Silene L.* (Caryophyllaceae) en Méditerranée occidentale. II: le groupe du *S. mollissima*. *Candollea* 39, 195–259.

Jeanmonod, D. (1985a). Révision de la section Siphonomorpha Otth du genre *Silene L.* (Caryophyllaceae) en Méditerranée occidentale. V: synthèse. *Candollea* 40, 35–56.

Jeanmonod, D. (1985b). Révision de la section Siphonomorpha Otth du genre *Silene L.* (Caryophyllaceae) en Méditerranée occidentale. IV: speciescaeterae. *Candollea* 40, 5–34.

Jeanmonod, D., and Bocquet, G. (1983). *Silene tyrrhenia* Jeanmonod & Bocquet sp. nova (Caryophyllaceae) – une nouvelle espèce bien connue. *Candollea* 38, 297–308.

Jeanmonod, D., and Mascherpa, J.-M. (1982). Révision de la section Siphonomorpha Otth du genre *Silene L.* (Caryophyllaceae) en Méditerranée occidentale. *Candollea* 37, 497–523.

Jones, G. (2017). STACEY: algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *J. Math. Biol.* 74, 447–467. doi: 10.1007/s00285-016-1034-0

Jones, G., Aydin, Z., and Oxelman, B. (2015). DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics*. 31, 991–998. doi: 10.1093/bioinformatics/btu770

Karrenberg, S., Liu, X., Hallander, E., Favre, A., Herforth-Rahme, J., and Widmer, A. (2018). Ecological divergence plays an important role in strong but complex reproductive isolation in campions (*Silene*). *Evolution* 73, 245–261. doi: 10.1111/evo.13652

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transformation. *Nucl. Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

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

Klopfstein, S., Currat, M., and Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Mol. Biol. Evol.* 23, 482–490. doi: 10.1093/molbev/msj057

Leuzinger, M., Naciri, Y., Du Pasquier, P.-E., and Jeanmonod, D. (2015). Molecular diversity, phylogeography and relationships of the Sileneparadoxa group of section Siphonomorpha (Caryophyllaceae). *Plant Syst. Evol.* 301, 265–278.

Melzheimer, V. (1988). "*Silene L.*," in *Flora Iranica*, ed. K. H. Rechinger Academische (Graz: Verlagsanatalt).

Naciri, Y., and Linder, P. (2015). Species identification and delimitation: the dance of the seven veils. *Taxon* 64, 3–16.

Naciri, Y., and Linder, H. P. (2020). The genetics of evolutionary radiations. *Biol. Rev.* 95, 1055–1072.

Naciri, Y., Cavat, F., and Jeanmonod, D. (2010). *Silene patula* (Siphonomorpha, Caryophyllaceae) in North Africa: a test of colonisation using chloroplast markers. *Mol. Phylogenet. Evol.* 54, 922–932. doi: 10.1016/j.ympev.2009.11.015

Naciri, Y., Caetano, S., and Salamin, N. (2012). Plant DNA Barcodes and the influence of gene flow. *Mol. Ecol. Res.* 12, 575–580. doi: 10.1111/j.1755-0998.2012.03130.x

Naciri, Y., Du Pasquier, P.-E., Lundberg, M., Jeanmonod, D., and Oxelman, B. (2017). A phylogenetic circumscription of *Silene* sect. Siphonomorpha (Caryophyllaceae) in the Mediterranean basin. *Taxon* 66, 91–108.

Nieto Feliner, G., and Rosselló, J. A. (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol. Phylogenet. Evol.* 44, 911–919. doi: 10.1016/j.ympev.2007.01.013

Oxelman, B., Lidén, M., Rabeler, R. K., and Popp, M. (2001). A revised generic classification of the tribe Sileneae (Caryophyllaceae). *Nordic J. Bot.* 20, 743–748. doi: 10.1111/j.1756-1051.2000.tb00760.x

Petri, A., and Oxelman, B. (2011). Phylogenetic relationships within *Silene* (Caryophyllaceae) section *Physolychnis*. *Taxon* 60, 953–968.

Pignatti, S. (1982). *Flora d'Italia*. Bologna: Edagricole.

Pignatti, S., Guarino, R., and La Rosa, M. (2017). *Flora d'Italia*, 2nd Edn, Vol. 2. Milan: Edagricole–New Business Media.

Polunin, O. (1980). *Flowers of Greece and the Balkans. A Field Guide*. Oxford: Oxford University Press.

Prentice, H. C., Malm, J. U., Mateu-Andrés, I., and Segarra-Moragues, J. G. (2003). Allozyme and chloroplast DNA variation in island and mainland populations of the rare Spanish endemic, *Silene hifacensis* (Caryophyllaceae). *Conserv. Genet.* 4, 543–555.

Rambaut, A. (2009). *FigTree v.1.4*. Available online at: http://evomics.org/resources/software/molecular-evolution-software/figtree/ (accessed June 15, 2021).

Rambaut, A., and Drummond, A. J. (2007). *Tracer v1.5*. Available online at: http://beast.bio.ed.ac.uk/Tracer/ (accessed June 15, 2021).

Rambaut, A., Suchard, M. A., Xie, W., and Drummond, A. J. (2013). *Tracer v1.6*. Available online at: http://beast.bio.ed.ac.uk/ (accessed June 15, 2021).

Rautenberg, A., Hathaway, L., Oxelman, B., and Prentice, H. C. (2010). Geographic and phylogenetic patterns in *Silene* section Melandrium (Caryophyllaceae) as inferred from chloroplast andnuclear DNA sequences. *Mol. Phylogenet. Evol.* 57, 978–991. doi: 10.1016/j.ympev.2010.08.003

R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Schemske, D. W. (2010). Adaptation and the origin of species. *Am. Nat.* 176, S4–S25.

Simmons, M. P., and Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* 49, 369–381.

Sokal, R. R., and Rohlf, F. J. (1995). *Biometry: The principles and Practice of Statistics in Biological Research*, 3rd Edn. New York, NY: W. H. Freeman.

Sukumaran, J., and Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proc. Natl. Acad. Sci. U.S.A.* 114, 1607–1612. doi: 10.1073/pnas.1607921114

Trigas, P., Iatrou, G., and Karetsos, G. (2007). Species diversity, endemism and conservation of the family Caryophyllaceae in Greece. *Biodivers. Conserv.* 16, 357–376.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7, 256–276. doi: 10.1016/0040-5809(75)90020-9

Yıldız, K., and Çırpıcı, A. H. (2013). Taxonomic revision of *Silene* (Caryophyllaceae) sections Siphonomorpha, Lasiostemones, Sclerocalycinae, Chloranthae, Tataricae and *Otites* in Turkey. *Turkish J. of Bot.* 37, 191–218.

Yıldız, K., Mınarecı, E., and Çırpıcı, A. H. (2009). Karyotypic study on *Silene*, section Lasiostemones species from Turkey. *Caryologia* 62, 134–141.

Zohary, M. (1973). *Geobotanical Foundations of the Middle East*. Stuttgart: Gustav Fischer Verlag.

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.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/fpls.2023.1116496

OPEN ACCESS

EDITED BY
Robert Philipp Wagensommer,
Free University of Bozen-Bolzano, Italy

REVIEWED BY
Emanuele Del Guacchio,
University of Naples Federico II, Italy
Eduardo Cires,
University of Oviedo, Spain

*CORRESPONDENCE
Božo Frajman
bozo.frajman@uibk.ac.at
Johannes Wessely
johannes.wessely@univie.ac.at

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

RECEIVED 05 December 2022
ACCEPTED 16 January 2023
PUBLISHED 16 February 2023

CITATION
Faltner F, Wessely J and Frajman B (2023) Phylogenetic data reveal a surprising origin of *Euphorbia orphanidis* (Euphorbiaceae) and environmental modeling suggests that microtopology limits its distribution to small patches in Mt. Parnassus (Greece). *Front. Plant Sci.* 14:1116496. doi: 10.3389/fpls.2023.1116496

Phylogenetic data reveal a surprising origin of *Euphorbia orphanidis* (Euphorbiaceae) and environmental modeling suggests that microtopology limits its distribution to small patches in Mt. Parnassus (Greece)

Felix Faltner[1], Johannes Wessely[2*] and Božo Frajman[1*]

[1]Department of Botany, University of Innsbruck/Universität Innsbruck, Sternwartestrasse, Innsbruck, Austria, [2]Department of Botany and Biodiversity Research, University of Vienna, Rennweg, Vienna, Austria

The Mediterranean Basin is one of the most biodiverse areas in the world, harboring 25,000 plant species, of which 60% are endemic. Some of them have narrow distributions, such as *Euphorbia orphanidis*, which is only known from alpine screes on Mt. Parnassos in Greece. Its exact distribution in this mountain was, however, poorly known, and its phylogenetic origin was also unclear. We performed extensive field work in Mt. Parnassos and could register *E. orphanidis* only in five patches of limestone screes in the eastern part of this mountain range, emphasizing its very narrow distribution, which is likely limited by topography influencing water availability as indicated by environmental modeling. We also registered 31 accompanying species and thus characterized its habitat. Using nuclear ribosomal internal transcribed spacer and plastid *ndhF–trnL* and *trnT–trnF* sequences, we show that it belongs to *E.* sect. *Patellares*, despite not having connate raylet leaves typical for this section, and not to *E.* sect. *Pithyusa* as previously suggested. The relationships among the species of *E.* sect. *Patellares* are poorly resolved, suggesting their simultaneous divergence that dated to the late Pliocene, which coincided with the establishment of the Mediterranean climate. The relative genome size of *E. orphanidis* is in the range of that for the other members of *E.* sect. *Patellares*, suggesting that it is diploid. Finally, we performed multivariate morphological analyses to generate a comprehensive description of *E. orphanidis*. Based on its narrow distribution and the anticipated negative impact of global warming, we consider this species endangered. Our study demonstrates how microrelief can limit the distribution of plants in topographically heterogeneous mountain environments and likely plays an important, yet neglected, role in shaping the distribution patterns of plants in the Mediterranean Basin.

KEYWORDS

endemism, endangered species, environmental modelling, Mediterranean basin, morphometry, phylogeny, taxonomy

1 Introduction

The Mediterranean Basin is one of the richest areas in the world in terms of animal and plant diversity and is considered one of the 25 global biodiversity hotspots (Myers et al., 2000). Owing to its geological and climatic complexity conferring a unique mosaic of habitats, it hosts approximately 25,000 plant species, of which 60% are endemic (Cuttelod et al., 2008; Nieto Feliner, 2014). One of the most biodiverse countries in the Mediterranean is Greece, which is positioned at the crossroads between Europe and Asia, harboring roughly 5,800 species of vascular plants. Of these, 1,462 (22%) are endemic (Dimopoulos et al., 2013), with the majority being restricted to small areas (Georghiou and Delipetrou, 2010). Their highest richness is in the Peloponnesus (468 taxa), followed by Crete and Karpathos (395 taxa), and Sterea Ellas, positioned just north of the Golf of Corinth, with 368 taxa (Dimopoulos et al., 2013). Local endemics in the coastal areas and the islands are predominantly distributed at low elevations (0–600 m a.s.l.), whereas those in the continental Greece occur mostly at higher elevations (Georghiou and Delipetrou, 2010). Many of them are linked to rock crevices and screes (Tan et al., 2001), which are particularly sensible to human disturbances (Panitsa et al., 2021).

Different factors, such as topographic, geologic, and climatic conditions, influence the distribution of endemic species (Fois et al., 2017; Zangiabadi et al., 2021). Narrow endemics are generally less stress tolerant than their widespread relatives and have specific ecological requirements. They often grow on steep slopes with high rock cover, such as screes, and in more open vegetation than their widespread congeners (Lavergne et al., 2004). Topographical variability that is particularly complex in mountainous regions can influence climatic and other plant growing conditions that strongly influence species distribution at a local scale (Whittaker and Levin, 1977; Ackerly et al., 2010; Boehm et al., 2021; Man et al., 2022). Such distinct habitats are often associated with restricted ranges of environmental conditions that affect species survival, but the exact causes limiting the distribution of narrow endemics are still poorly understood (Manne and Pimm, 2001; Munson and Sher, 2015).

One of the narrow endemics from high elevations in Sterea Ellas in Greece is *Euphorbia orphanidis* Boiss (Dimopoulos et al., 2013). It was first collected by Theodorus Orphanides near Lugari on Mt. Parnassos in 1852 and distributed in 1854 as *E. hohenackeri* within Flora Graeca Exiccata (no. 407). As the name *E. hohenackeri* Hochst. et Steud. was earlier applied for another species, it was renamed *E. orphanidis* by Boissier (1859). Since then, it has only been collected (see *Specimina visa*) and reported (Quezel, 1964; Garnweidner, 1986) a few times from calcareous rocks and screes on Mt. Parnassos, but despite its narrow range, its exact distribution remains unknown (Parnassos National Park Authority, personal communication). It is considered to grow from the upper part of the montane coniferous forests dominated by the Greek endemic *Abies cephalonica* Loudon and *Pinus nigra* J.F. Arnold (1,500 m) to the alpine zone (2,100 m; Aldén, 1986). Along with *E. orphanidis*, many other rare and endemic species, such as *Astragalus parnassia* Boiss., *Colchicum parnassicum* Sart.Orph. & Heldr., *Convolvulus parnassicus* Boiss. & Orph., *Paeonia parnassica* Tzanoud., and *Scutellaria rupestris* subsp. *parnassica* (Boiss.) Greuter & Burdet, occur in Mt. Parnassos. This mountain has an outstanding position among the Greek mountains and is regarded as a hotspot of endemic taxa, with several of them being critically endangered (Kougioumoutzis et al., 2021). The high concentration of rare endemics is a result of topological heterogeneity and the geographical position of Mt. Parnassos between the biodiversity hotspots of Peloponnesus and Pindos Mountains (Kougioumoutzis et al., 2021) and led to the foundation of the National Park Parnassos in 1938 (Tsitsoni et al., 2015). However, despite its limited distribution, *E. orphanidis* itself is not a protected species in Greece.

Euphorbia orphanidis is a glabrous, glaucous, prostrate to ascending perennial with an extensive, branched, fleshy rhizome and stems rising up to 15 cm (Radcliffe-Smith and Tutin, 1968). Based on its morphology, it was included in *Euphorbia* sect. *Paralias* Dumort. subsect. *Conicocarpae* (Prokh.) Prokh. by Geltman (2009) and correspondingly in *Euphorbia* sect. *Pithyusa* (Raf.) Lázaro in the most recent taxonomic treatment by Riina et al. (2013), but its phylogenetic position remains unknown. The latter section includes 50, mainly glabrous and glaucous species, often growing on rocky calcareous grounds, with a high diversity in the Mediterranean (Riina et al., 2013).

Considering the unknown phylogenetic position as well as the poorly known distribution of *E. orphanidis*, our aims were to investigate the evolutionary origin and to determine the systematic position of this species. In addition, we gathered exact data about its distribution and ecology and thus provide the information needed to design future conservation strategies as well as to disentangle environmental factors that possibly limit its distribution. To achieve the first aim, we (1) sequenced nuclear ribosomal internal transcribed spacer (ITS) and plastid *trnT–trnF* (*trnTF* in the following) and *ndhF–trnL* to infer its phylogenetic position with a dense sampling of closely related species, (2) estimated its relative genome size (RGS) using flow cytometry and compared it to the RGS of closely related species, and (3) determined its morphological characteristics using multivariate morphometrics. To achieve the second aim, we (4) performed extensive field work on Mt. Parnassos in 2019 and 2020 and mapped the occurrences of *E. orphanidis*, (5) determined its accompanying species, and (6) performed environmental niche modeling by applying different topographic predictors that likely determine the distribution range of *E. orphanidis* on Mt. Parnassos. We finally (7) assessed the conservation status of this narrow endemic following the IUCN criteria.

2 Materials and methods

2.1 Plant material

The plant material of *E. orphanidis* for RGS estimation and molecular and morphometric analyses was sampled from seven localities on Mt. Parnassos in 2020. For molecular and RGS analyses, leaf material was collected in silica gel, and herbarium vouchers were pressed. The morphological characteristics of *E. orphanidis* were studied on the 18 specimens that we collected (Supplementary Table S1) as well as on seven specimens from six collections deposited in the herbaria G, M, W, and WU, indicated by asterisks following the herbarium IDs in the "Taxonomic treatment" section. In addition, 13 outgroup species from the same section (109

populations) as well as 27 species from 10 closely related sections were included in the phylogenetic analyses; species from the same section were also subjected to RGS analyses (Supplementary Tables S1, S2). The ITS and *ndhF–trnL* sequences of most of accessions besides *E. orphanidis* were taken from previous studies (mostly from Pahlevani and Frajman, 2023), whereas most of the *trnTF* sequences were generated for this study (Supplementary Tables S1, S2; Figure 1).

2.2 Field work: Mapping the distribution of *E. orphanidis* and recording accompanying species

We visited Mt. Parnassos in September 2019 and 2020 and mapped the presence of *E. orphanidis* across the mountain range. We visited the localities given in the literature (Quezel, 1964; Garnweidner, 1986) as well as those from the herbarium labels. After acquiring knowledge about the ecology of this species, we systematically screened the majority of the screes in the eastern part of the mountain range as well as some in the central part indicated by Garnweidner (1986) and mapped the presence (or absence) of the species. In addition, we registered the accompanying species of *E. orphanidis* that were identified using Strid (1986), Strid and Tan, (1986), and Papiomytoglou et al. (2021).

2.3 DNA extraction, sequencing, and analyses of sequence data

Extraction of total genomic DNA and sequencing were performed for ITS and *trnTF* as described by Frajman and Schönswetter (2011) and for *ndhF–trnL* by Pahlevani and Frajman (2023). The sequencing was carried out at Eurofins Genomics (Ebersberg, Germany). Contigs were assembled and edited, and sequences were aligned using Geneious Pro 5.5.9 (Kearse et al., 2012). As the preliminary analyses showed that *E. orphanidis* belongs to *E.* sect. *Patellares* (Prokh.) Frajman and not to *E.* sect. *Pithyusa*, we sampled several accessions of different species from the former section. In total, 115 ITS, 73 *ndhF–trnL*, and 41 *trnTF* sequences were used, partly from previous studies and partly newly generated (Supplementary Tables S1, 2). Maximum parsimony (MP) and MP bootstrap (MPB) analyses were performed using PAUP 4.0b10 (Swofford, 2002). The most parsimonious trees were searched for heuristically with 100 replicates of random sequence addition, TBR swapping, and MulTrees on. The swapping was performed on a maximum of 1,000 trees (nchuck = 1,000). All characters were equally weighted and unordered. The data set was bootstrapped using full heuristics, 1,000 replicates, TBR branch swapping, MulTrees option off, and random addition sequence with five replicates. The Bayesian analyses were performed using MrBayes 3.2.1 (Ronquist et al., 2012), applying the GTR+Γ substitution model proposed by the Akaike information criterion implemented in MrAIC.pl 1.4 (Nylander, 2004) for all datasets. Values for all parameters, such as the shape of the gamma distribution, were estimated during the analyses. The settings for the Metropolis-coupled Markov chain Monte Carlo process included four runs with four chains each (three heated ones using the default heating scheme) and run simultaneously for 10,000,000 generations each, with sampling trees every 1,000th generation using default priors. The posterior probabilities (PP) of the phylogeny and its branches were determined from the combined set of trees, discarding the first 1,001 trees of each run as burn-in. Tracer 1.6 (Rambaut et al., 2014). was used to assess convergence and mixing, which were appropriate. In addition, a NeighborNet was produced with ITS sequences of *E.* sect. *Patellares*, using SplitsTree 4.12.3 (Huson and Bryant, 2006).

Divergence times were estimated with BEAST 1.8.2 (Drummond et al., 2012) using a pruned ITS alignment containing 22 accessions of *E.* sect. *Patellares* and 15 accessions of its sister clade inferred with the analysis of the entire ITS dataset. Birth–death speciation prior (Gernhard, 2008) and GTRI+Γ substitution model with estimated base frequencies were used for the phylogeny inference. A lognormal relaxed clock with a weakly informative prior on the clock rate

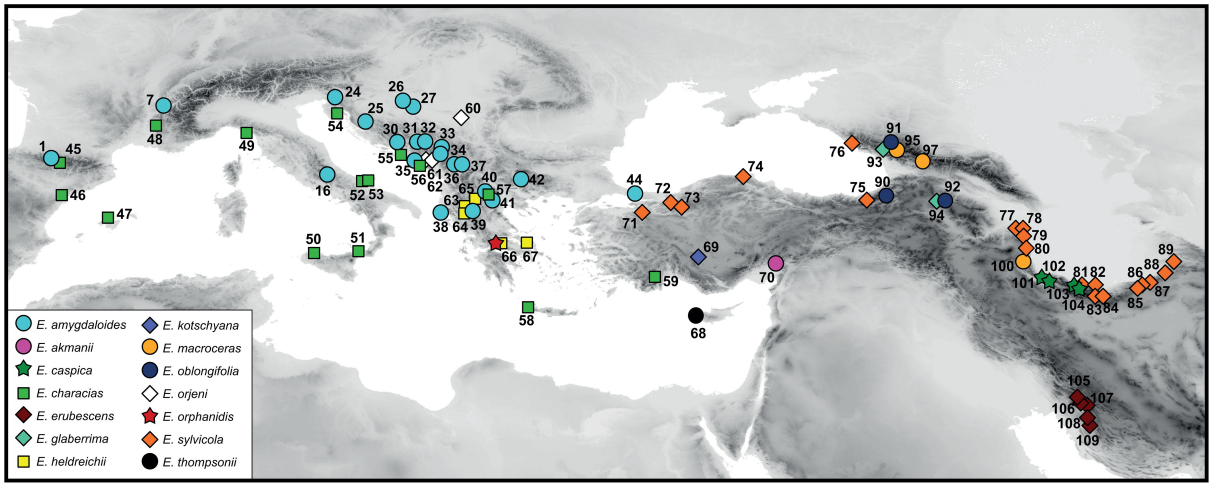


FIGURE 1
Distribution of populations of *Euphorbia* sect. *Patellares* used in phylogenetic and partly relative genome size analyses. The population numbers correspond to Supplementary Table S1 and are not indicated for *Euphorbia orphanidis*.

(exponential with mean 0.001) was applied. The prior age of the root was set to 17.7 million years with a normally distributed standard deviation of 2.3, which corresponds to the median age and 95% highest posterior densities (HPD) interval of the corresponding node (17.7 My; HPDs, 13.2–22.6; *i.e.*, the split between *E.* sect. *Patellares* and its sister clade) obtained from the dating analysis by Horn et al. (2014). Two independent MCMC chains were run for 10,000,000 generations, saving trees and parameters every 1,000 generations. The performance of the analysis was checked with Tracer 1.6 (Rambaut et al., 2014); both the effective sample sizes (ESS >200) and mixing were appropriate. The log and tree files from both runs were combined using Log Combiner (part of the BEAST package) after discarding 10% of each run as burn-in. The maximum clade credibility tree was then produced and annotated with Tree Annotator (part of the BEAST package) and visualized with FigTree 1.4.2 (Rambaut, 2014).

2.4 Relative genome size measurements

The RGS of 25 individuals of *E. orphanidis* from five localities were measured with a CyFlow space flow cytometer (Partec, GmbH, Münster, Germany) using 4′,6-diamidino-2-phenylindole and the reference standard *Bellis perennis* L. (2C = 3.38 pg; Schönswetter et al., 2007) following Suda and Trávníček (2006) and the modifications described by Cresti et al. (2019). In addition, the RGS values of 68 populations belonging to eight species from the same section as *E. orphanidis* were included in the analyses for comparison. The RGS data were visualized in RStudio 1.2.5001 (RStudio Team, 2019) by utilizing R-3.4.0 and the package "ggplot2".

2.5 Morphometric analyses

A total of 24 individuals from different collections (those collected by us are shown in Supplementary Table S1 and those from other collections are marked with an asterisk in the "Taxonomic treatment" section) were analyzed morphometrically. A total of 36 metric characters were measured or scored; based on them, 17 ratio characters were calculated (see the species description below). Plant height, stem, leaf, and ray characters were measured/scored manually. All the other characters (cyathium, fruit, and seed characters) were measured on microscopic images taken with a stereomicroscope Olympus SZX9 using the Olympus image analysis software analySIS pro.

2.6 Environmental modeling

As the presence/absence data of *E. orphanidis* were unbalanced (seven presences and 37 absences), we used a resampling approach to evaluate the potential influence of topographic predictors. We therefore resampled as many absences as presences from the dataset 10,000 times. For these samples, we derived the mean of all predictors and evaluated if the mean of the presences was outside the 95% confidence interval (see Figure 2). As topographic predictors, we used the Terrain Ruggedness Index (TRI), Topographic Position Index (TPI; Wilson et al., 2007), roughness, slope, aspect, and flow direction (flowdir), *i.e.*, the direction of the greatest drop in elevation. TRI is a measure of ruggedness, expressing the amount of elevation difference between adjacent cells. TPI, on the other hand, is a measure for terrain classification. It compares the elevation of a cell to the mean elevation of the surrounding cells. Hence, a negative value represents a cell that is lower than its surroundings (valleys). We did not evaluate climatic predictors due to the small extent covered by the occurrence data; for such small regions, the available climate data consists of measurements of the nearest climate station, which is statistically downscaled *via* correlations to altitude. Therefore, climatic variation in the data would only reflect elevational variation.

3 Results

3.1 ITS and plastid *ndhF–trnL* and *trnTF* phylogenies

The ITS sequences of *E. orphanidis* were 693, and the alignment was 739 characters long. A total of 191 characters (25.8%) were parsimony-informative. The homoplasy index was 0.37 (0.43 after exclusion of uninformative characters), and the retention index was 0.86. A total of 100,000 most parsimonious trees were found, and their score was 545. The Bayesian and maximum parsimony reconstructions resulted in congruent topologies when considering clades with MPB >70% and PP >0.95 (Figure 3A; Supplementary Figure S1). *Euphorbia* sect. *Patellares* was monophyletic (PP 1, MPB 100%), and within the section single accessions of *E. akmanii* I. Genç & Kültür, *E. kotschyana* Fenzl and *E. thompsonii* Holmboe and multiple accessions of *E. characias* L. were included in a basal polytomy, in which several monophyletic clades, mostly corresponding to species, were also included. One of them corresponded to *E. orphanidis* (PP 1, MPB 99%) and the others to *E. heldreichii* Orph. ex Boiss. (PP 1, MPB 92%) and *E. orjeni* Beck (PP 0.99, MPB 70%), which together formed a poorly supported clade in the parsimony tree (MPB 67%) as well as *E. amygdaloides* L. (PP 0.99, MPB 81%), *E. erubescens* Boiss. (PP 1, MPB 83%), *E. sylvicola* Pahlevani & Frajman (PP 1, MPB 99%), and a clade (PP 0.99, MPB 76%) including *E. caspica* Frajman & Pahlevani in a basal polytomy and a clade (PP 1, MPB 79%) of *E. glaberima* K. Koch, *E. macroceras* Fisch. & C.A. Mey, and *E. oblongifolia* (K. Koch) K. Koch. The clade including *E. caspica*, *E. glaberima, E. macroceras*, and *E. oblongifolia* was resolved as sister to a clade (MPB 57%) including all other species of *E.* sect. *Patellares* by parsimony, but not Bayesian analysis. Consistent with the ITS tree, the NeighbourNet of *E.* sect. *Patellares* (Figure 3B) was star-like, with *E. characias, E. kotsyana*, and *E. thompsonii* in the center of the star and *E. amygdaloides*, *E. akmanii*, *E. erubescens*, *E. heldreichii*, *E. orjeni*, *E. orphanidis*, and *E. sylvicola* in the terminal parts of their own star-rays. *Euphorbia glaberrima*, *E. macroceras*, and *E. oblongifolia* were all in the terminal part of a ray, in which *E. caspica* was positioned in the central part.

The topology of the ITS chronogram (Figure 4) was congruent with that of the ITS tree (Figure 3A), *i.e.*, there was a polytomy of several clades mostly corresponding to species since all other clades resolved in the tree had low support (PP < 0.8). *Euphorbia* sect.

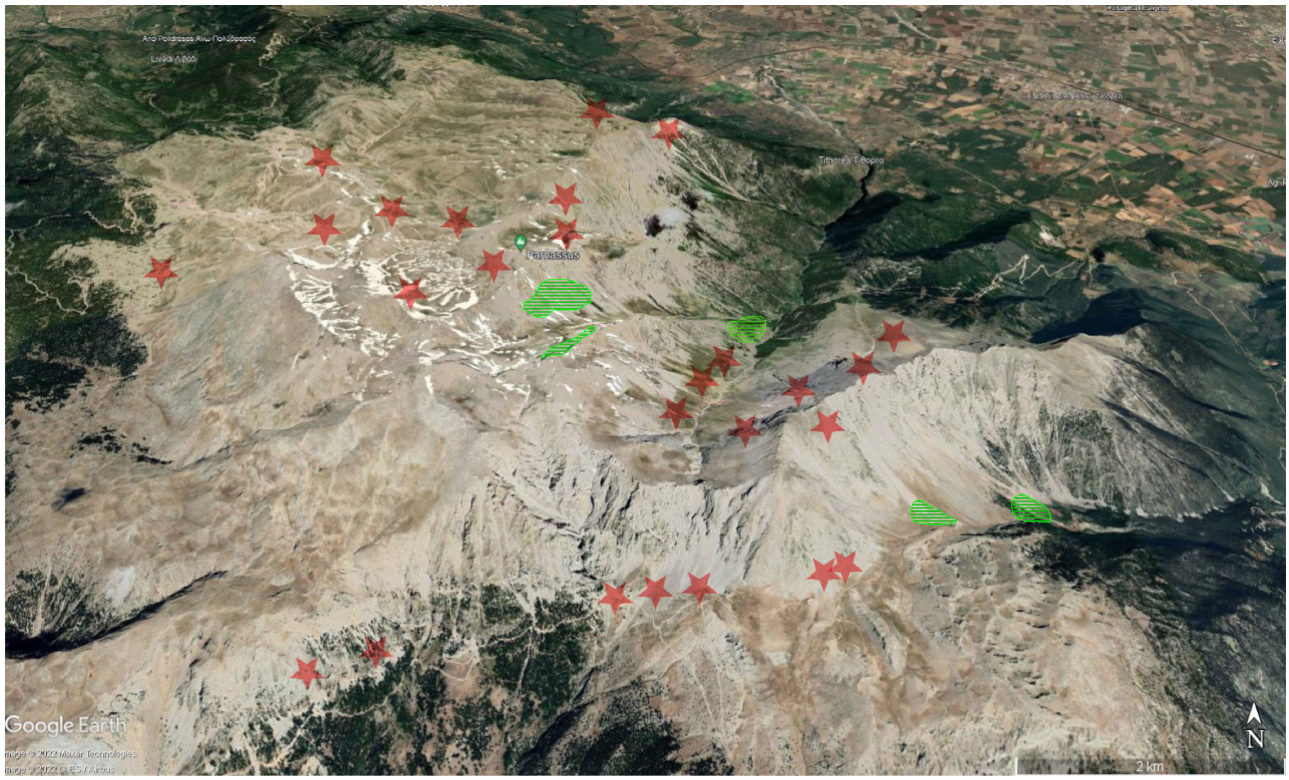


FIGURE 2
Distribution of *Euphorbia orphanidis* on Mt. Parnassus (green), with indicated areas as to where the species was searched for but not found (red stars). The map was adopted from Google Earth .

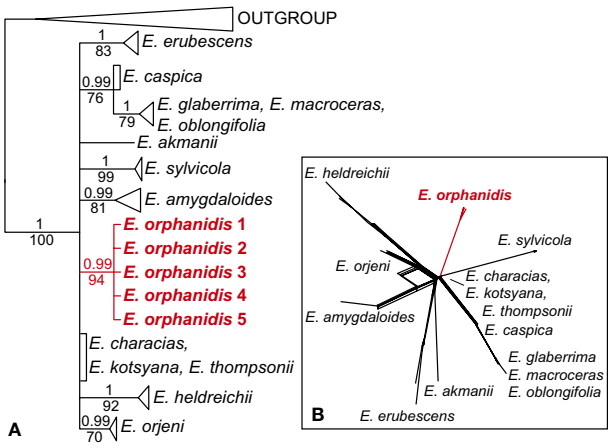


FIGURE 3
Phylogenetic relationships inferred from internal transcribed spacer sequences among the members of *Euphorbia* sect. *Patellares*. **(A)** Bayesian consensus phylogram; the numbers above the branches are posterior probability values >0.5, while those below the branches are maximum parsimony bootstrap values >50%. The clades of all species including multiple accessions, except that of *Euphorbia orphanidis*, are collapsed, and the complete tree is shown in Supplementary Figure S1. **(B)** NeighborNet.

Patellares originated in the Miocene, 17.9 Ma (HPDs: 13.7–22.2 Ma), whereas the onset of its diversification was dated to the late Pliocene, 2.8 (1.3–5.5) Ma. This was also the time of the origin of all main lineages (mostly corresponding to species) within *E.* sect. *Patellares*, whereas their diversification was dated to the Pleistocene.

The *ndhF–trnL* sequences of *E. orphanidis* were 911, and the alignment was 1,229 characters long. A total of 130 characters (10.6%) were parsimony-informative. The homoplasy index was 0.19 (0.29 after exclusion of uninformative characters), and the retention index was 0.89. In total, 100,000 most parsimonious trees were found, and their score was 349. The Bayesian and maximum parsimony reconstructions resulted in congruent topologies (Figure 5A). *Euphorbia* sect. *Patellares* was monophyletic (PP 1, MPB 79%), the relationships within the section were poorly resolved, and several species appeared polyphyletic. In the basal polytomy, several accessions of *E. characias* (three grouped together in a clade) and one accession each of *E. heldreichii* and *E. thompsonii* were included, along with a clade (PP 0.99, MPB 68%) including all other accessions. In this clade, relationships were also poorly resolved, with several accessions in a basal polytomy (considering the clades with support values PP 0.64 and 0.82 as non-relevant) or in small clades including only two accessions, like that of *E. erubescens* (PP 1, MPB 96%). There were three bigger clades in this polytomy. The first (PP 1, MPB 84%) included *E. caspica* (PP 0.96, MPB 66%) and a poorly supported clade (PP 0.82, MPB 56%) with one accession of *E. characias* from westernmost Anatolia and a clade (PP 0.98, MPB 62%) including *E. orphanidis*. The second (PP 0.99, MPB 51%) included *E. macroceras* and *E. sylvicola*, and the third (PP 1, MPB 87%) consisted of *E. amygdaloides*, *E. heldreichii*, *E. orjeni*, and *E. sylvicola*.

10.3389/fpls.2023.1116496

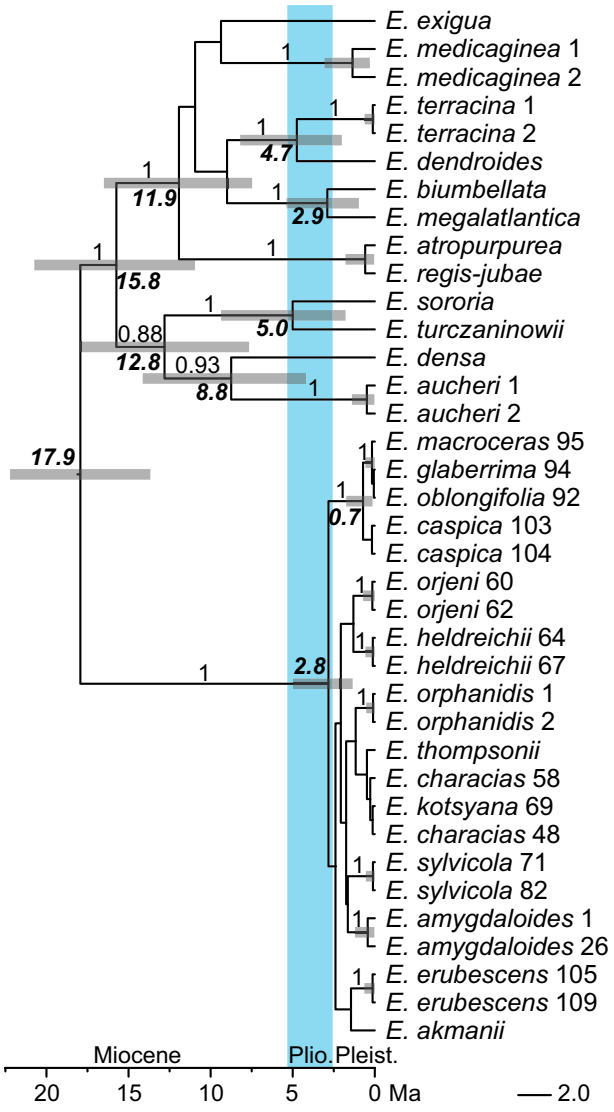


FIGURE 4
Bayesian consensus chronogram (Maximum Clade Credibility tree) inferred from internal transcribed spacer sequences. The numbers above the branches are posterior probability values >0.90, the bold numbers associated with the nodes indicate the median crown group ages in millions of years of the clades diversifying at those nodes, and the bars correspond to 95% highest posterior density of the age estimates. The population numbers of *Euphorbia* sect. *Patellares* correspond to the data in Supplementary Table S1 and Figure 1, while those of the outgroup sections correspond to the data in Supplementary Table S2.

The *trnTF* sequences of *E. orphanidis* were 1,486, and the alignment was 1,869 characters long. A total of 97 characters (5.2%) were parsimony-informative. The homoplasy index was 0.09 (0.18 after exclusion of uninformative characters), and the retention index was 0.91. Moreover, 97,000 most parsimonious trees were found, and their score was 262. The Bayesian and maximum parsimony reconstructions resulted in congruent topologies (Figure 5B) that mostly corresponded to the topology of the *ndhF–trnL* tree, although some differences could be observed. Contrary to the *ndhF–trnL* tree, *E. caspica* was not in the same clade as *E. orphanidis*, and in addition to population 59 of *E. characias* from Turkey, population 52 of *E. characias* from Italy was also in this clade (PP 1, MPB 88%); the latter population was in the basal polytomy of *E.* sect. *Patellares* along with some other accessions of *E. characias* and one each of *E. heldreichii* and *E. thompsonii* in the *ndhF–trnL* tree, if considering PP 0.84 non-relevant.

3.2 Relative genome size

The relative genome size of *Euphorbia orphanidis* varied between 1.675 and 1.703, which was in the range of the RGS of *E. amygdaloides* (1.571–1.873), *E. characias* (1.623–1.910), and *E. heldreichii* (1.619–1.734), whereas *E. caspica*, *E. macroceras*, *E. orjeni*, and *E. sylvicola* had a deviating RGS (Figure 6; Supplementary Figure S2; Supplementary Table S1).

3.3 Distribution of *E. orphanidis* on Mt. Parnassos

Euphorbia orphanidis is distributed in partly mobile to stabilized, mostly south-easterly (one north-easterly) exposed calcareous scree slopes on Mt. Parnassos (Figures 2, 7). Despite visiting and inspecting several scree slopes scattered around the Parnassos Mountain Range, we found the species only in five localized patches south-southeast of the main summit Liakoura (2,457 m). All localities are situated on the slopes of two main valleys separated by the ridge of Mavra Litharia (2,327 m) and Koukos (2,235 m), *i.e.*, the valley of Velitsa stream southwest of Tithorea and the valley west of Davlia above the monastery Moni Ierousalim. The localities are positioned between 1,500 and 2,300 m, *i.e.*, above the timberline, which is, in this part of the mountain range, formed by *Abies cephalonica*. Two patches in the former valley are in the area of Chouni south of the main summit Liakoura and one in the side valley above Tsares (slopes of Psilo Kotroni). In the latter valley, one patch is situated just above the timberline in the area of Kanalia, whereas the other is just below Akrino Nero. In all localities, we noted the presence of *E. orphanidis* only in patches of stabilized but open screes with medium-sized screes—in areas with bigger- or smaller-sized screes, the species does not occur.

We could not find *E. orphanidis* in the area around the Ski Club Athens reported by Garnweidner (1986), and no other collectors collected or reported it from this part of Mt. Parnassos. In this area, we only found the morphologically similar *E. deflexa* and noted the absence of suitable habitats for *E. orphanidis*; therefore, we deem this record erroneous. E. Garnweidner (written communication to B. Frajman on 22.9.2020) confirmed that the report of *E. orphanidis* for the summit region of Lyakoura (Garnweidner, 1986) actually was erroneous and corresponded to *E. deflexa*; therefore, both other reports from the very same day published by Garnweidner (1986), *i.e.*, the one from a karstic depression in the area of the Ski Club Athens and that from the area of the newer ski resort on the northern flanks of Mt. Parnassos are likely erroneous as well. Along the same line, we were not able to locate the historical locality "Lugari", from which the species was described, as this toponym is not mentioned in the maps and also not familiar to the national park authorities.

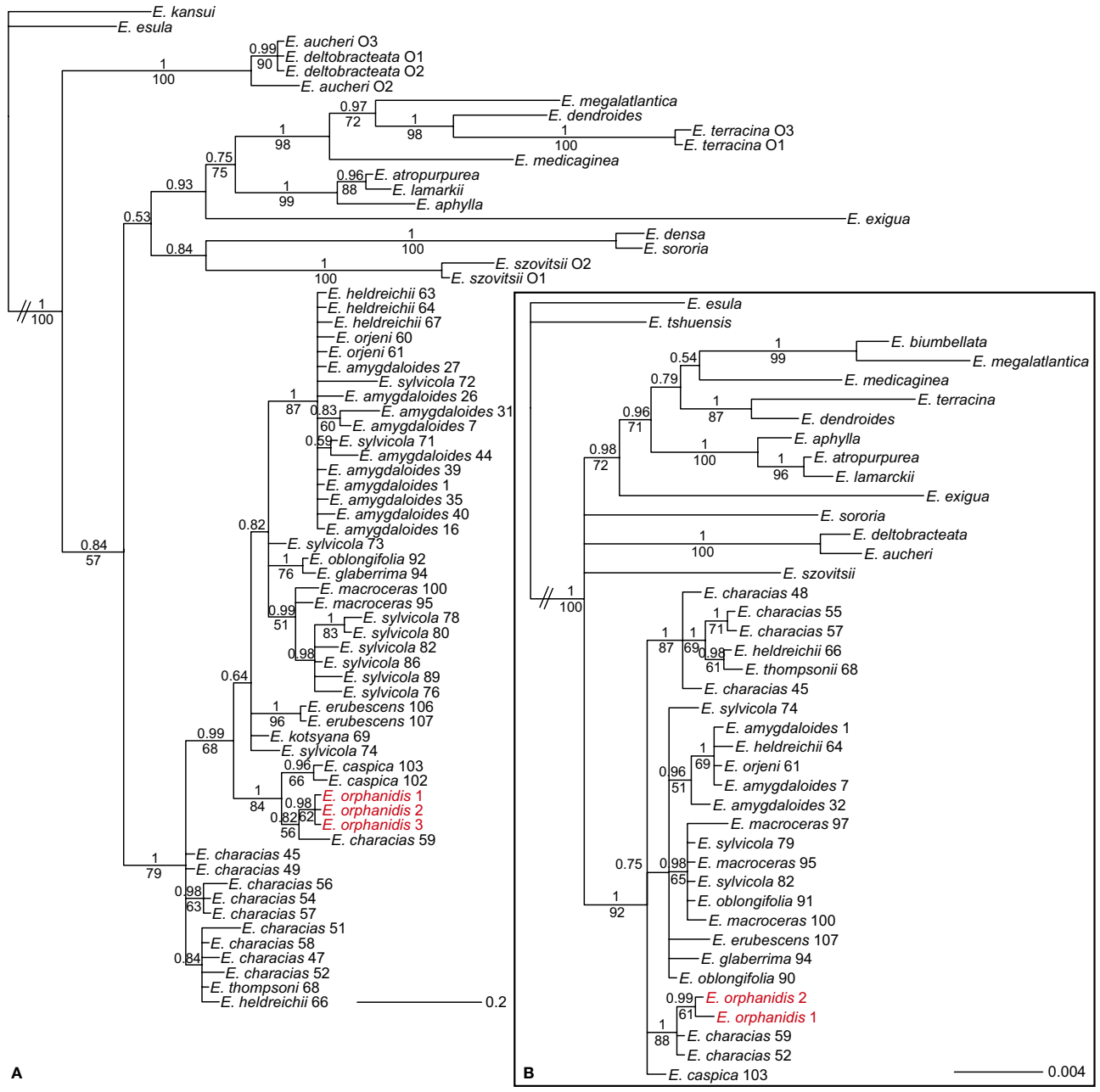


FIGURE 5
Bayesian consensus phylograms of plastid *ndhF–trnL* **(A)** and *trnTF* **(B)** sequences showing the phylogenetic relationships among the members of *Euphorbia* sect. *Patellares* and outgroup species. The numbers above the branches are posterior probability values >0.5, while those below the branches are maximum parsimony bootstrap values >50%. The population numbers of *E.* sect. *Patellares* correspond to the data in Supplementary Table S1 and Figure 1, while those of the outgroup sections correspond to the data in Supplementary Table S2.

"Likeri", a glacial cirque just northeast below the main summit, sounds similar to Lugari and is characterized by extensive screes, but we did not find *E. orphanidis* there.

3.4 Ecological characterization of *E. orphanidis* including environmental niche modeling

We registered 31 species accompanying *E. orphanidis* (Table 1). Two of them are endemic to Stereá Elláda, six to Greece, and many to the Balkan Peninsula, whereas some are more widespread. Among them, the most commonly found species in several locations along with *E. orphanidis* were *Digitalis laevigata* subsp. *graeca*, *Drypis spinosa*, *Festuca spectabilis*, *Scrophularia lanciniata*, and *Sideritis raeseri*.

From all the topographic predictors analyzed, only the Topographic Position Index (TPI) was significantly different (lower) for the sites were *E. orphanidis* was present in relation to the sites where it was absent (Figure 8). This indicates that the localities where *E. orphanidis* was recorded have a concave topography, *i.e.*, they are at the bottom (small valleys) of scree fields.

10.3389/fpls.2023.1116496

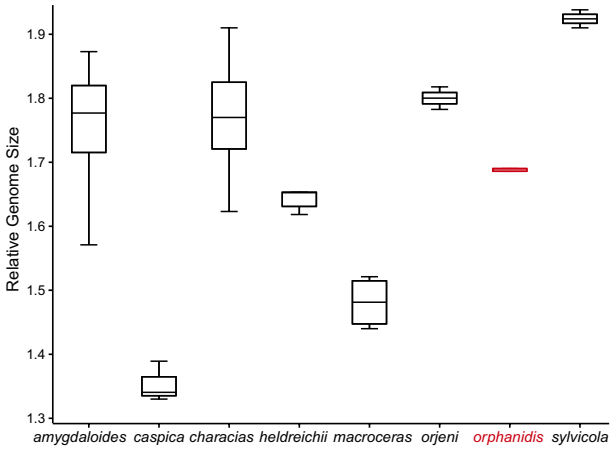


FIGURE 6
Relative genome size (RGS) variation in *Euphorbia orphanidis* and other species of *Euphorbia* sect. *Patellares*. The boxplots are based on the RGS population mean values presented in Supplementary Table S1. A scatterplot representing the mean values and standard deviation of all investigated populations is shown in Supplementary Figure S2.

4 Discussion

4.1 *Euphorbia orphanidis* belongs to *E. sect. Patellares* and not *E. sect. Pithyusa*

Our phylogenetic data based on nuclear and plastid sequences clearly show that *E. orphanidis* belongs to *E.* sect. *Patellares* (Prokh.) Frajman and not to *E.* sect. *Pithyusa* as suggested by Geltman (2009) and Riina et al. (2013) based on the species' morphology. *Euphorbia orphanidis*, along with 17 other species of *E.* sect. *Patellares* (Pahlevani and Frajman, 2023), likely originated in the late Tertiary. Whereas the divergence between *E. amygdaloides* and *E. characias* was dated to the Miocene/Pliocene boundary 5.6 Ma (Horn et al., 2014), our dating analysis rather inferred a younger divergence between these two species as well as among most other species of the section, probably due to a more complete taxon sampling in our study. This divergence was dated to the late Pliocene 2.8 Ma (Figure 4), and it is likely that most of the species of *E.* sect. *Patellares*, including *E. orphanidis*, originated in the Pliocene. The aridification that started 9–8 Ma and resulted in the establishment of the Mediterranean climate in the Pliocene 3.2–2.8 Ma (Suc, 1984) triggered the fragmentation of previously forested areas (Milne and

FIGURE 7
Euphorbia orphanidis in its natural habitat on Mt. Parnassos with parts of the inflorescences showing cyathial glands and fruits.

TABLE 1 Accompanying species of *Euphorbia orphanidis* on the calcareous screes on Mt. Parnassos (Greece).

Taxon	Distribution
Aethionema saxatile subsp. *creticum* Boiss. & Heldr.	Widespread (East Aegean islands, Greece, Kriti, Turkey)
Anthemis spruneri Boiss. & Heldr.	Greece
Thliphthisa purpurea subsp. *apiculata* (Sm.) P.Caputo & Del Guacchio.	Widespread (Balkan Peninsula and Turkey)
Carum heldreichii Boiss.	Greece
Centaurea musarum Boiss. & Orph.	Greece
Cerastium candidissimum Correns.	Greece
Cicer incisum (Willd.) K.Malý.	Greece, Iran, Kriti, Lebanon-Syria, North Caucasus, Transcaucasus, Turkey
Dactylis glomerata subsp. *hispanica* (Roth) Nyman.	Widespread
Daphne oleoides Schreb.	Widespread
Digitalis laevigata subsp. *graeca* (Ivanina) K.Werner.	Balkan Peninsula (Greece, Bulgaria)
Drypis spinosa L.	Balkan Peninsula and Italy
Euphorbia deflexa Sm.	Albania, Greece, Kriti
Euphrasia salisburgensis Funck ex Hoppe.	Widespread
Festuca spectabilis Bertol.	Balkan Peninsula and Italy
Lactuca viminea (L.) J.Presl & C.Presl.	Widespread
Linaria peloponnesiaca Boiss. & Heldr.	Balkan Peninsula
Lysimachia serpyllifolia Schreb.	Greece
Marrubium velutinum Sm.	Greece
Melica ciliata L.	Widespread
Morina persica L.	Widespread
Nepeta parnassica Heldr. & Sartori.	Greece and Albania
Odontites linkii Heldr. & Sartori ex Boiss.	Cyprus, East Aegean islands, Greece, Kriti
Pterocephalus pterocephala (L.) Dörfl.	Greece
Petrosedum ochroleucum (Chaix) Niederle.	Widespread
Ranunculus brevifolius Ten.	Balkan Peninsula, Italy and Turkey
Sanguisorba minor Scop.	Widespread
Scrophularia heterophylla subsp. *laciniata* (Waldst. & Kit.) Maire & Petitm.	Balkan Peninsula (Balkan Peninsula)
Sempervivum marmoreum Griseb.	Widespread
Senecio thapsoides DC.	Balkan Peninsula
Sideritis raeseri Boiss. & Heldr.	Balkan Peninsula
Thamnosciadium junceum (Sm.) Hartvig.	Greece
Valantia aprica (SM.) Tausch.	Albania, Greece, Kriti
Verbascum graecum Heldr. & Sartori.	Balkan Peninsula and Turkey

The distribution data are from POWO (2022) and Papiomytoglou et al. (2021), and the nomenclature follows POWO (2022). The distribution data are given for the species and within the parentheses for the infraspecific taxon listed.

Abbott, 2002), which likely caused the speciation in *E.* sect. *Patellares* that predominately includes mesophilous forest species (Riina et al., 2013). Many Tertiary species found a shelter in the Mediterranean Basin that is recognized as one of the most important refugia for species of this ancient flora (Thompson, 2005) and as a biodiversity hotspot with an exceptional role in the preservation of unique species and genetic diversity (Myers et al., 2000; Nieto Feliner, 2014). *Euphorbia orphanidis* certainly holds a special position among these species given its very limited distribution.

Euphorbia orphanidis is glabrous and glaucous (Figure 7), which fits well to other species of *E.* sect. *Pithyusa*. In addition, ecologically, growing in open, dry, and rocky habitats, it more resembles the members of *E.* sect. *Pithyusa*. In contrast, the majority of species belonging to *E.* sect. *Patellares* are mesophilous and grow in forests

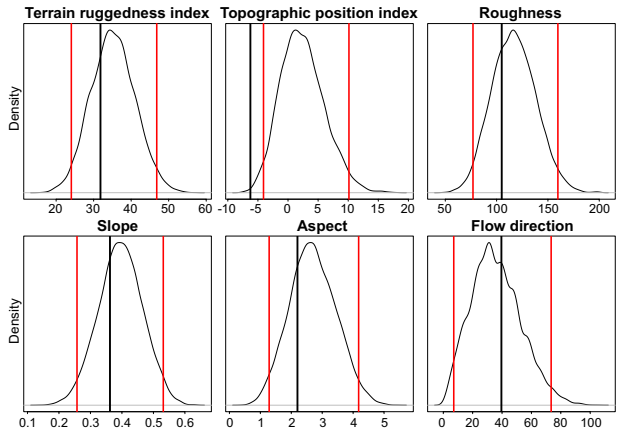


FIGURE 8
Density distribution of the mean topographic predictors of randomly distributed points (repeated 10,000 times). The red lines depict the 95% confidence interval of these random samples, while the black vertical lines depict the mean predictor values of the populations of *Euphorbia orphanidis*.

and scrublands (Riina et al., 2013; B. Frajman, personal observations). The leaf venation in *E. orphanidis* is not prominent but clearly pinnate, which is characteristic for *E.* sect. *Patellares*, whereas members of *E.* sect. *Pithyusa* have palmate leaf venation. Cyathial glands are semilunate to trapezoid with two horns and thus resemble other species of *E.* sect. *Patellares*, even if some members of *E.* sect. *Pithyusa* also have similar glands. Furthermore, the ovoid trilobed capsules fit better to *E.* sect. *Patellares*, whereas the members of *E. sect. Pithyusa* mostly have conical capsules. However, *E. orphanidis* is the only species of *E.* sect. *Patellares* which does not have connate raylet leaves, a feature that was considered a synapomorphy for this section (Frajman and Schönswetter, 2011; Riina et al., 2013) and likely got lost in *E. orphanidis*. Finally, the RGS of *E. orphanidis* also lies in range of the genome size of other members of *E.* sect. *Patellares* (Figure 6), which indicates that it is a diploid species and likely has 20 chromosomes, a typical number for this section (Frajman and Schönswetter, 2011; Riina et al., 2013; Rice et al., 2015); despite several trials, we could not establish the chromosome number for *E. orphanidis*, as no nuclei in appropriate division phase were found.

Contrary to the clear position of *E. orphanidis* within *E.* sect. *Patellares*, its precise origin within this section remains unclear. The ITS tree (Figure 3) is unresolved, and most of the species, including *E. orphanidis*, form their own clades in a polytomy. The NeighbourNet is star-like, suggesting a more or less simultaneous divergence among the species. In addition, the plastid phylogenetic trees (Figure 5) are largely unresolved but suggest a close relationship of *E. orphanidis* with populations of *E. characias* from the eastern Mediterranean. Even if plastid phylogenies in *Euphorbia* subgen. *Esula* mostly do not follow species boundaries, are commonly geography-correlated, and the populations of different species can share the same or similar haplotypes (Hand et al., 2015; Frajman et al., 2016; Frajman and Schönswetter, 2017), a closer relationship of *E. orphanidis* with *E. characias* appears plausible. *Euphorbia characias* is a widespread Mediterranean species that mostly occurs at lower altitudes compared with *E. orphanidis* but can also reach altitudes of 2,000 m (Lafranchis and Sfikas, 2009; B. Frajman, personal observations); we did not observe this species in Mt. Parnassos, but roughly 50 km away in the Corinthian Bay (B. Frajman, pers. observ.). Within *E.* sect. *Patellares*, *E. characias* is the most thermophilous species, often growing in dry stony garrigues. In this respect, it is the species from *E.* sect. *Patellares* that is ecologically most similar to *E. orphanidis*. Another species from this section that partly ecologically resembles *E. orphanidis* is a high-elevation ecotype of *E. heldreichii* Orph., an endemic species of the southern Balkan Peninsula that mostly grows in thermophilous lowland forests but can extend its range to the alpine belt (Caković and Frajman, 2020). We have also registered it in gravelly open grassland at 2,000 m in Mt. Parnassos. *Euphorbia orphanidis* also differs from both species and other members of *E.* sect. *Patellares* habitually, as it is much smaller than the other species. This is a trait that is generally considered typical of narrow endemics (Lavergne et al., 2004), and our data corroborate this hypothesis.

4.2 Narrow distribution of *E. orphanidis* is limited by microtopography and is of high conservation concern

Despite considerable efforts to find *E. orphanidis* in scree habitats scattered across the Parnassos Mountain range, we have only found it in five patches positioned between 1,500 and 2,300 m, in two valleys draining towards the east and positioned south-southeast of the main summit Liakoura (2,457 m). Topological heterogeneity is likely the most important factor influencing the distribution of *E. orphanidis* on Mt. Parnassos. As carunculate seeds of *Euphorbia* species are being dispersed by ants (Espadaler and Gómez, 1997) and several *Euphorbia* species, also those from *E.* sect. *Patellares*, have wide distributions, it seems unlikely that dispersal limitation is the reason for the limited distribution of *E. orphanidis*, but rather its specific ecology. TPI, on the other hand, gives a possible explanation. The only topographic predictor that was significantly different (lower) at sites were *E. orphanidis* is present compared with the screes where we could not find it was TPI. This indicates that the species only thrives in the concave small valleys towards the bottom of scree fields, where snow brought by avalanches accumulates. These sites are therefore likely more humid due to longer snow cover and less water runoff compared with the steeper scree areas above. On the other hand, more flat areas in the surroundings mostly have denser vegetation (B. Frajman, personal observations), which also limits the distribution of weak competitors such as *E. orphanidis*.

Quezel (1964) described an association of *Sclerochorton* (*i.e.*, *Thamnosciadium*) *junceum* and *Euphorbia deflexa* from the screes between 1,600 and 2,100 m on Mt. Parnassos and Mt. Giona, including further character species as *Freyera parnassica* Boiss. et Heldr., *Galium apiculatum* Sm. (= *Asperula purpurea* subsp. *apiculata*), *Nepeta sibthorpii* Benth subsp. *parnassica* (Heldr. & Sart.) (= *Nepeta parnassica*), as well as *Cicer ervoides* Brign. (= *Cicer incisum*). Based on the differential species *E. orphanidis*, *Festuca spectabilis* subsp. *affinis* (Hack.) Hack., and *Chaenorhinum minus* (L.) Lange, he proposed a local sub-association from Mt. Parnassos, especially from the Gourna region, which further corroborates our results indicating that *E. orphanidis* grows in very specific ecological conditions on Mt. Parnassos that do not occur

on the neighboring Mt. Giona, thus likely limiting its distribution also in a broader area, not only in Mt. Parnassos.

Small and genetically depauperate populations with narrow ecological niche are expected to undergo significant reductions in the near future (Theodoridis et al., 2018). Narrow endemic species often grow on slopes with high rock cover and open vegetation (Lavergne et al., 2004), similar to the habitats of *E. orphanidis* and many other Greek endemics. A total of 890 angiosperms of the Greek flora grow in screes, and 46.6% of them are endemic to Greece (Panitsa et al., 2021). Eight of 31 recorded accompanying species of *E. orphanidis* are likewise Greek endemic (Table 1), and three of them (including *E. orphanidis*) are endemic to Sterea Elladas (Panitsa et al., 2021). Narrow endemics growing in screes are generally less stress tolerant than their widespread relatives (Lavergne et al., 2004) and are considered weak competitors, thus intolerant to human disturbances (Panitsa et al., 2021).

As Mediterranean high-mountain plants are expected to face extreme heatwaves and summer droughts caused by climate change, which will not only influence the survival and fitness but also trigger changes in the reproduction and regeneration of these plants (Giménez-Benavides et al., 2017), and *E. orphanidis* is not well adapted to drought based on our environmental modeling, we can anticipate a continuing decline in the area of occupancy and extent and quality of habitat of this species in the future decades. This, along with its current area of occupancy being less than 10 km^2 (Figure 2) and its occurrence only in one mountain range (Mt. Parnassos) at no more than five locations, suggests that *E. orphanidis* is endangered (EN) following criterion B of the IUCN classification (IUCN, 2012).

4.3 Identification key to *Euphorbia* species on Mt. Parnassos

Euphorbia orphanidis can be confused with other *Euphorbia* species occurring on Mt. Parnassos (see above; cf. Garnweidner, 1986), where they partly grow in similar habitats and co-occur on the same plots, especially *E. deflexa*. To avoid their misidentifications in the future, we provide an identification key for all *Euphorbia* species that we registered above the timberline on Mt. Parnassos, including photographs of their seeds (Figure 9).

1 Plant upright, procumbent to ascending, not glaucous-papillose, (30)35–60(80)cm high, leaves (3)4.5–7(10) cm long, raylet leaves connate……*E. heldreichii*

1* Plant prostrate, decumbent to ascending, glaucous-papillose, up to 30(4) cm high, leaves up to 2.5(3.5) cm long, raylet leaves free……2

2 Caespitose plant with numerous prostrate to decumbent stems, (2)5–15(25) cm long, usually forming dense mats, without axillary rays. Cauline leaves dense, small, (1)2–5(8) × (1)2–3(5) mm. Terminal rays 2–3. Capsules 2.5–3.5 × 3.5–4.0 mm, with two wings on each keel. Seeds shallowly pitted……*E. herniariifolia*

2* Mostly larger plant not forming dense mats, usually with some axillary rays. Leaves bigger, longer than (3)5 mm and broader than (2)3 mm. Terminal rays mostly more than 3. Capsules often larger (but not in *E. deflexa*) not winged on valves. Seeds smooth, vermiculate-rugose or reticulate-pitted……3

3 More robust plant, with densely leafy stems. Leaves sessile, cuspidate or mucronate, rather thick and fleshy, (10)15–20 (35) × (4)7–12(18) mm. Nectarial glands with dilated, often weakly lobed horns. Capsule (4)5–6(7) × 6 mm. Seeds 3.5–4.5 × 2–2.5 mm, vermiculate or rarely smooth……*E. myrsinites*

3* Less robust plant, with less densely leafy stems. Leaves shortly petiolate, not cuspidate or mucronate, not thick and fleshy, (2)5–12(15) × (1.5) 3–6(7) mm. Horns of nectarial glands not dilated, long and slender. Capsule (2.8)3.0–4.5(4.8) × (3)3.5–4.5(5) mm. Seeds 2–2.5 × 1.5–1.8 mm, reticulate-pitted or smooth……4

FIGURE 9
Seeds of five *Euphorbia* species growing on Mt. Parnassus (above, ventral view; below, dorsal view).

4 Plant ascending to erect, without fleshy long rhizomes. Leaves (2)5–10(15) × (1.5)3–4(6) mm. Capsule (2.8)3.0–3.2(3.5) × 3 mm. Seeds reticulate-pitted, 2 × 1.5 mm……*E. deflexa*

4* Plant prostrate to ascending, with fleshy long rhizomes. Leaves 8–13 × (2.5)3–5.7(7) mm. Capsule (3.1)3.4–4.7(4.8) × (3)3.3–4.9(5) mm. Seeds smooth 2.5 × 1.8 mm……*E. orphanidis*

4.4 Taxonomic treatment

Euphorbia orphanidis Boiss., 1859, Diagn. Pl. Orient. Nov., Ser. 2, 4: 89. ≡ *Tithymalus orphanidis* (Boiss.) Soják, 1972, Čas. Nár. Mus., Odd. Prír. 140: 175. Lectotype (Aldén, 1986): [Greece] "Flora Graeaca Exsiccata No. 407: *Euphorbia hohenackeri* Boiss. et Orphan, nov. sp., in monte Parnassi prope Lugari (rara), Fl. Jun–Jul, alt. 5000–6000', 4/16 Jul 1854, *Theodorus G. Orphanides* (G00754290-photo)!. Isolectotypes: W 0031038!*, WU 078186!*; photos: G00754364!, HAL0118607!, K000911958!, L0151758!, P00655601!, P00655602! S 13-12930!, WAG0004323!; LE (according to Geltman, 2009).

Other original material (syntypes): Flora Graeca Exsiccata No. 407: Habit in m. Parnassi regione superiore, rara. Alt. 5000´–6000´. 16. Jul. 18??, *Theodorus G. Orphanides* (G00398586!*); Parnass, *Orphan.* (JE00002895-photo)!; De Heldreich Herbarium Graecum normale No. 344: Inter lapides mobiles, in reg. media m. Parnassi (supra Acrino-nero), alt. 4000´-4500´, Aug. 1855, *F. Giucciardi* (G00398585!*, WU 0078185!*; photos: P00655603!, P00655604!.De Heldreich Flora Graeca Exiccata 2967: *Euphorbia hohenackeri* Boiss. et Orphan., In m. Parnassi reg. media, Aug. 1855, *J. Guicciardi* (M0274982!*; photos: C10011249!, G00754289!, K000911957!, L0151757,LD1033460! P00655605)!; Herbarium Willkommii: *Euphorbia hohenackeri* Boiss. et Orphan., In monte Parnasso ad 3–6000', 16.7.1854, *Orphanides* (W0102305!*); *Euphorbia hohenackeri* Boiss. et Orphan., M. Parnassos, Jul. m. (MO1911900-photo!; neither the collector nor collection year is listed on the label; therefore, it is not certain if this specimen is a part of the original material).

Description: Glaucous and glabrous, prostrate to ascending perennial with extensive, branched, fleshy rhizome and several elongated stems. Plant (11)14–26(31) cm high, stems (5)7-20(24) cm long and (1)1.5–3(3) mm thick, with 1–7(11) axillary shoots. Middle stem leaves petiolate, obovate, entire, 8–13 × (2.5)3–5.7(7) mm, (1.3)1.9–3.4(3.7) times longer than wide, widest at (4)5.3–9(11) of the length, with a narrow basis and an obtuse apex. Terminal rays 3–10 (11), (7)10–35(80) mm long, 0–6 times dichotomous. Ray-leaves oblong-oblanceolate to ovate-oblong (4)5–8 × (4.5)7.1–12.9(18) mm, (0.6)1.1–1.9(2.2) times wider than long, widest at 0.1–0.4 of their length. Raylet-leaves rhombic-deltate to reniform, obtuse, occasionally emarginate, (2.5)3–6.5(7) × (4.5)5.5–9.5(11) mm, (1)1.2–2(2.2) times wider than long, widest at (0.5)0.7–3(3.5) of their length. Cyathial involucre (1.1)1.2–2(2.1) × (0.9)1–2(2.5) mm, (0.8)0.9–1.1 times longer than wide. Nectarial glands yellow, 1.5–1.6 × 1.4–1.6 mm, with two, 0.7–1 mm long, occasionally bifid horns. Capsule deeply sulcate, smooth, (3.1)3.4–4.7(4.8) × (3)3.3–4.9(5) mm, 0.9-1 times as long as wide, widest at (0.02)0.06–0.3 of the length. Style 0.10–0.13 mm long. Seeds smooth, broadly ovoid and dark grey 2.5 × 1.8 mm, 1.4 times longer than wide, widest at 0.1 of the length. Caruncle (0.6)0.7–0.9 × 0.9–1.3 mm, 1.2–1.6 times wider than long (Figures 7, 10).

Distribution and habitat. Endemic to the eastern part of Mt. Parnassos in Central Greece (Figure 2). Partly mobile calcareous scree slopes above the timberline from 1,500 to 2,300 m (Figure 7).

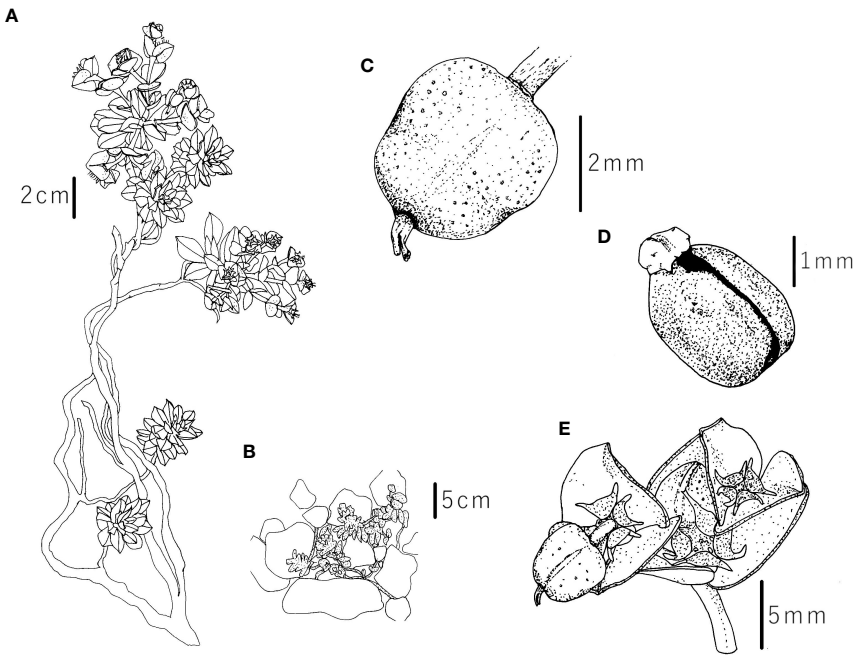


FIGURE 10
Iconography of *Euphorbia orphanidis*: **(A)** whole plant, **(B)** plant in its habitat, **(C)** capsule, **(D)** seed, and **(E)** inflorescence with fruit and nectarial glands. Drawings by (F) Faltner.

Conservation status. Endangered (EN) according to IUCN (2012) criterion B2.

Additional specimens seen: Parnassus – ascent to Gourna, May 1862, *John Stuart Mill.* (photo: K000911959!, P00736822)!; Iter Graecum: In lapidosis regionis abietinae Mt. Parnassi, loco Gurna dicto, rare, 15. Jul 1888, *D. Halacsy* (WU 0126727)!; Flora Graeca: Mt. Parnassos, NE-ENE of Arachova, Scree and gravel on a slope facing SE, 1800-1950 m, 06.07.1975, *Lars Åke Gustavsson 6717* (ATHU 41726 photo)!; Flora Graeca: Sterea Ellas, Nom. Viotias, Ep. Levadhiasi, Mt. Parnassos, SE part c.7 km ENE of Arachova, screes, c. 1700 m, 25.08.1982, *P. Hartvig, R. Franzen & K. I. Christensen 10420*, (G00398584!*); Flora Hellenica: Nom. Viotias, Ep. Levadias, Mt. Parnassos, SE part c.7 km ENE of Arachova, shaded rocks – in scree, 1750–1950 m. Lat: 38°31'N Long: 22°49' E, 25.08.1982, *Hartvig* et al. *10420* (ATHU 59759 photo!)

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

BF conceived and designed the study, performed field work, coordinated the lab work, performed data analyses, and wrote substantial parts of the manuscript. FF performed field work, parts of the lab work, morphometric measurements, identification of species, and data analyses and wrote substantial parts of the manuscript and used its earlier version as his master thesis. JW performed environmental modeling, wrote corresponding parts of the manuscript, and commented on other parts of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was financially supported by a Knoll-Widmung fund (Austrian Academy of Sciences) to BF and a KWA stipend of the University of Innsbruck to FF.

Acknowledgments

Collecting permits were issued by the National Park Parnassos and the Greek Ministry for Environment and Energy. We thank all friends and colleagues who helped in collecting the samples listed in Supplementary Tables S1, S2 as well as the herbarium curators from the herbaria G, M, W, and WU that provided herbarium material for morphometric analyses and designation of type specimens. Many thanks to G. Barbarigos, E. Garnweidner, E. Liveri, D. Marmygkas, A. Strid, P. Trigas, and A. Zikos for their help with the locality data and herbarium images. We are grateful to M. Magauer, D. Pirkebner and B. Pernfuss, who performed or supervised the laboratory work, and also to J. Amann, A. Dudaš, S. Đurović, P. Gostner, L. Mayr, S. Natterer, L. Silbernagl, V. Steinkasserer, V. Stojilković, and C. Zudrell who measured the relative genome size. Many thanks to W. Kofler for the help with taking seed micrographs. We are grateful to M. Bodner, M. Imhiavan, and their colleagues from the Botanical Gardens of the University of Innsbruck for successfully cultivating our living collection of *Euphorbia*.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1116496/full#supplementary-material

References

Ackerly, D., Loarie, S., Cornwell, W., Weiss, S., Hamilton, H., Branciforte, R., et al. (2010). The geography of climate change: Implications for conservation biogeography. *Divers. Distrib.* 16, 476–487. doi: 10.1111/j.1472-4642.2010.00654.x

Aldén, B. (1986). "*Euphorbia* L." in *Mountain flora of Greece 1*. Ed. A. Strid (Cambridge: Cambridge University Press).

Boehm, A. R., Hardegree, S. P., Glenn, N. F., Reeves, P. A., Moffet, C. A., and Flerchinger, G. N. (2021). Slope and aspect effects on seedbed microclimate and germination timing of fall-planted seeds. *Rangeland Ecol. Managem.* 75, 58–67. doi: 10.1016/j.rama.2020.12.003

Boissier, E. (1859). *Diagnoses plantarum novarum orientalium, ser. 2/4.* (Lipsiae & Parisiis: (B. Herrmann and J.-B. Baillière).

Caković, D., and Frajman, B. (2020). Three tertiary *Euphorbia* species persisted in the forests of the Balkan peninsula. *Plant Syst. Evol.* 306, 1–12. doi: 10.1007/s00606-020-01672-w

Cresti, L., Schönswetter, P., Peruzzi, L., Barfuss, M. J. H., and Frajman, B. (2019). Pleistocene survival in three Mediterranean refugia: Origin and diversification of the Italian endemic *Euphorbia gasparrinii* from the *E. verrucosa* alliance (Euphorbiaceae). *Bot. J. Linn. Soc* 189, 262–280. doi: 10.1093/botlinnean/boy082

Cuttelod, A., García, N., Abdul Malak, D., Temple, H., and Katariya, V. (2008). "The Mediterranean: A biodiversity hotspot under threat," in *The 2008 review of the IUCN red list of threatened species*. Eds. J.-C. Vié, C. Hilton-Taylor and S. N. Stuart (Gland, Switzerland: IUCN).

Dimopoulos, P., Raus, T. H., Bergmeier, E., Constantinidis, T. H., Iatrou, G., Kokkini, S., et al. (2013). Vascular plants of Greece: An annotated checklist. *Englera* 31, 1–372.

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075

Espadaler, X., and Gómez, C. (1997). Soil surface searching and transport of *Euphorbia characias* seeds by ants. *Acta Oecolog.* 18, 39–46. doi: 10.1016/S1146-609X(97)80079-3

Fois, M., Fenu, G., Cañadas, E. M., and Bacchetta, G. (2017). Disentangling the influence of environmental and anthropogenic factors on the distribution of endemic vascular plants in Sardinia. *PloS One* 12, 1–14. doi: 10.1371/journal.pone.0182539

Frajman, B., Graniszewska, M., and Schönswetter, P. (2016). Evolutionary patterns and morphological diversification within the European members of the *Euphorbia illirica* (*E. villosa*) group: one or several species? *Preslia* 88, 369–390.

Frajman, B., and Schönswetter, P. (2011). Giants and dwarfs: Molecular phylogenies reveal multiple origins of annual spurges within *Euphorbia* subg. *Esula*. *Mol. Phylogenet. Evol.* 60, 413–424. doi: 10.1016/j.ympev.2011.06.011

Frajman, B., and Schönswetter, P. (2017). Amphi-Adriatic distributions in plants revisited: Pleistocene trans-Adriatic dispersal in the *Euphorbia barrelieri* group (Euphorbiaceae). *Bot. J. Linn. Soc* 185, 240–252. doi: 10.1093/botlinnean/box055

Garnweidner, E. (1986). Florenliste der Exkursion der Bayerischen botanischen Gesellschaft 1983 nach Griechenland. *Ber. Bayer. Bot. Ges.* 57, 121–136.

Geltman, D. V. (2009). Spurges (*Euphorbia* l., euphorbiaceae) of the boreal Eurasia. i. section *Paralias* Dumort [In Russian with English summary]. *Novosti Sist. Vyssh. Rast.* 41, 166–191.

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

Gernhard, T. (2008). The conditioned reconstructed process. *J. Theor. Biol.* 253, 769–778. doi: 10.1016/j.jtbi.2008.04.005

Giménez-Benavides, L., Escudero, A., García-Camacho, R., García-Fernández, A., Iriondo, J. M., Lara-Romero, C., et al. (2017). How does climate change affect regeneration of Mediterranean high-mountain plants? an integration and synthesis of current knowledge. *Plant Biol.* 20, 50–62. doi: 10.1111/plb.12643

Hand, R., Hadjikyriakou, G., Christodoulou, C., and Frajman, B. (2015). Multiple origins of dendroid shrubs in the eastern Mediterranean *Euphorbia hierosolymitana* group (Euphorbiaceae) with description of a new species, *Euphorbia lemesiana*, from Cyprus. *Bot. J. Linn. Soc* 179, 295–307. doi: 10.1111/boj.12319

Horn, J. W., Xi, Z., Riina, R., Peirson, J. A., Yang, Y., Dorsey, B. L., et al. (2014). Evolutionary bursts in *Euphorbia* (Euphorbiaceae) are linked with photosynthetic pathway. *Evolution* 68, 3485–3504. doi: 10.1111/evo.12534

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

IUCN (2012). *IUCN red list categories and criteria: version 3.1. 2nd edition* (Gland and Cambridge: IUCN).

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kougioumoutzis, K., Kokkoris, I. P., Panitsa, M., Strid, A., and Dimopoulos, P. (2021). Extinction risk assessment of the Greek endemic flora. *Biology* 10, 195. doi: 10.3390/biology10030195

Lafranchis, T., and Sfikas, G. (2009). *Flowers of Greece* (Oberreifenberg: Koeltz).

Lavergne, S., Thompson, J. D., Garnier, E., and Debussche, M. (2004). The biology and ecology of narrow endemic and widespread plants: a comparative study of trait variation in 20 congeneric pairs. *Oikos* 107, 505–518. doi: 10.1111/j.0030-1299.2004.13423.x

Manne, L. L., and Pimm, S. L. (2001). Beyond eight forms of rarity: which species are threatened and which will be next? *Anim. Conserv.* 4, 221–230. doi: 10.1017/S1367943001001263

Man, M., Wild, J., Macek, M., and Kopecký, M. (2022). Can high-resolution topography and forest canopy structure substitute microclimate measurements? bryophytes say no. *Sci. Total Environ.* 821, 153377. doi: 10.1016/j.scitotenv.2022.153377

Milne, R. I., and Abbott, R. J. (2002). The origin and evolution of tertiary relict floras. *Adv. Bot. Res.* 38, 281–314. doi: 10.1016/S0065-2296(02)38033-9

Munson, S. M., and Sher, A. A. (2015). Long-term shifts in the phenology of rare and endemic rocky mountain plants. *Am. J. Bot.* 102, 268–1276. doi: 10.3732/ajb.1500156

Myers, N., Mittermeier, R. A., Mittermeier, C. G., de Fonseca, G. A. B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858.

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

Nylander, J. A. A. (2004). *MrAIC. pl. 1.4. 3*. Uppsala: Uppsala University.

Pahlevani, A. H., and Frajman, B. (2023). Widespread, but less than assumed: Populations of *Euphorbia amygdaloides* (Euphorbiaceae) from Western Asia cryptic two new species. *Perspect. Plant Ecol. Evol. Syst*. doi: 10.1016/j.ppees.2023.125717

Panitsa, M., Kokkoris, I. P., Kougioumoutzis, K., Kontopanou, A., Bazos, I., Strid, A., et al. (2021). Linking taxonomic, phylogenetic and functional plant diversity with ecosystem services of cliffs and screes in Greece. *Plants* 10, 992. doi: 10.3390/plants10050992

Papiomytoglou, V., Antonopoulos, Z., Goula, K., Zografidis, A., Kipopoulos, L., Kofinas, G., et al. (2021). *Greekflora.gr. mediterraneo*.

POWO (2022). *Plants of the world online* (Kew: Facilitated by the Royal Botanic Gardens).

Quezel, P. (1964). Végétation des hautes montagnes de la grèce méridionale. *Vegetatio* 12, 289–385.

Radcliffe-Smith, A., and Tutin, T. G. (1968). "*Euphorbia* L." in *Flora europaea*, vol. 2 . Eds. T. Tutin, V. H. Heywood, N. A. Burges, D. M. Moore, D. H. Valentine, S. M. Walters and D. A. Webb (Cambridge: Cambridge University Press).

Rambaut, A. (2014) *FigTree 1.4.2*. Available at: http://tree.bio.ed.ac.uk/.

Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014) *Tracer v1.6*. Available at: http://beast.bio.ed.ac.uk/tracer.

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

Riina, R., Peirson, J. A., Geltman, D. V., Molero, J., Frajman, B., Pahlevani, A., et al. (2013). A worldwide molecular phylogeny and classifcation of the leafy spurges, *Euphorbia* subgenus *Esula* (Euphorbiaceae). *Taxon* 62, 316–342. doi: 10.12705/622.3

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

RStudio Team (2019). *RStudio: Integrated development environment for r. RStudio* (Boston, MA: PBC).

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

Strid, A. (1986). *Mountain flora of Greece 1* (Cambridge: Cambridge University Press), 852.

Strid, A., and Tan, K. (1986). *Mountain flora of Greece 2* (Cambridge: Cambridge University Press), 974.

Suc, J.-P. (1984). Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature* 307, 429–432.

Suda, J., and Trávníček, P. (2006). Estimation of relative nuclear DNA content in dehydrated plant tissues by flow cytometry. *Curr. Protoc. Cytom.* Chapter 7:Unit7.30. doi: 10.1002/0471142956.cy0730s38

Swofford, D. L. (2002). *PAUP*: Phylogenetic analysis using parsimony (*and other methods)* (Sunderland, MA: Sinauer Associates).

Tan, K., Iatrou, G., and Johnsen, B. (2001). *Endemic plants of greece. the peloponnese* (Copenhagen: GADS), 35–47.

Theodoridis, S., Patsiou, T. S., Randin, C., and Conti, E. (2018). Forecasting range shifts of a cold-adapted species under climate change: are genomic and ecological diversity within species crucial for future resilience? *Ecography* 41, 1357–1369. doi: 10.1111/ecog.03346

Thompson, J. D. (2005). *Plant evolution in the Mediterranean* (Oxford: Oxford University Press).

Tsitsoni, T., Tsaprounis, N., Varvarigos, G., Lanara, T., Koukou, E., and Margaritopoulou, V. (2015). *Discover Parnassos* (Amfiklia: Parnassos National Park Management body).

Whittaker, R. H., and Levin, S. A. (1977). The role of mosaic phenomena in natural communities. *Theoret. Popul. Biol.* 12, 117–139. doi: 10.1016/0040-5809(77)90039-9

Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C., and Grehan, A. J. (2007). Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Mar. Geodesy* 30, 3–35. doi: 10.1080/01490410701295962

Zangiabadi, S., Zaremaivan, H., Brotons, L., Mostafavi, H., and Ranjbar, H. (2021). Using climatic variables alone overestimate climate change impacts on predicting distribution of an endemic species. *PloS One* 16, e0256918. doi: 10.1371/journal.pone.0256918

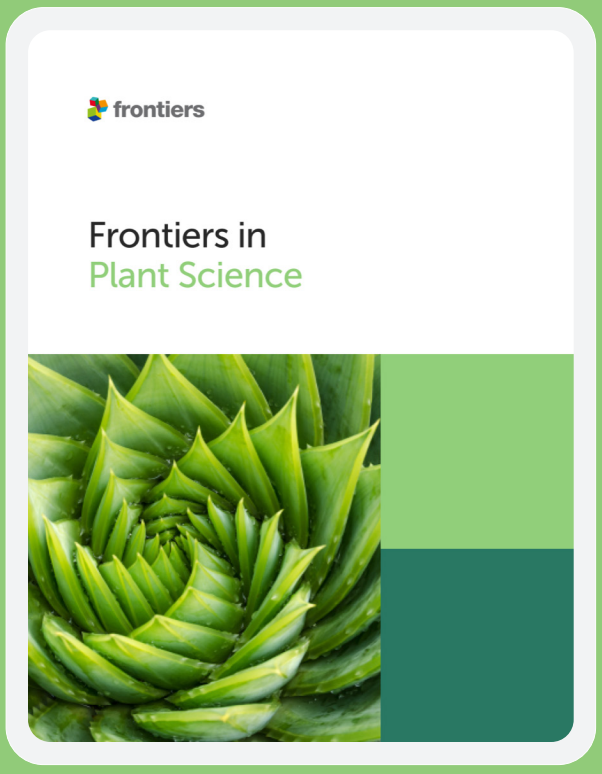
Frontiers in
Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact