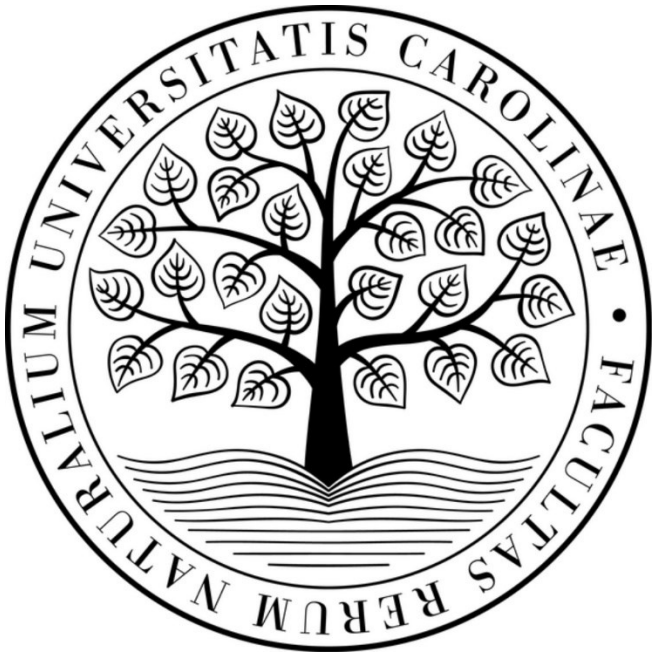
Univerzita Karlova, Přírodovědecká fakulta
Charles University, Faculty of Science

Studijní program: Botanika / Study programme: Botany



Mgr. Gabriela Šrámková

Evolution of the genus *Arabidopsis* in its centre of diversity

Evoluce v centru diversity rodu *Arabidopsis*

Disertační práce / Doctoral thesis

Školitel / Supervisor: prof. RNDr. Karol Marhold, DrSc.

Prague, 2022

Declaration

I hereby declare that I made this thesis independently, using the mentioned references. I have not submitted or presented any part of the thesis for any other degree or diploma.

Prohlášení

Prohlašuji, že jsem disertační práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Gabriela Šrámková

Author contribution statement

I hereby declare that I have substantially contributed to all papers included in the thesis. My contributions to particular papers are as follows:

CS I Kolář F, Lučanová M, Záveská E, **Fuxová G**, Mandáková T, Španiel S, Senko D, Svitok M, Kolník M, Gudžinskas Z, Marhold K (2016) Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae).
Fieldwork, lab work, paper writing – total contribution 10 %

CS II Kolář F*, **Fuxová G***, Záveská E*, Nagano AJ, Hyklová L, Lučanová M, Kudoh H, Marhold K (2016) Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*.
Fieldwork, labwork, data analysis, paper writing – total contribution 30 %

CS III Šrámková-Fuxová G*, Záveská E*, Kolář F, Lučanová M, Španiel S, Marhold K (2017) Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction
Design of the study, fieldwork, labwork, data analysis, paper writing – total contribution 45 %

CS IV Šrámková G, Kolář F, Záveská E, Lučanová M, Španiel S, Kolník M, Marhold K (2019) Phylogeography and taxonomic reassessment of *Arabidopsis halleri* – a montane species from Central Europe.
Design of the study, fieldwork, labwork, data analysis, paper writing – total contribution 80 %

CS V Padilla-García N*, **Šrámková G***, Kolář F, Šlenker M, Zeisek V, Záveská E, Clo J, Lučanová M, Rurane I, Marhold K (manuscript) Niche differentiation following whole-genome duplication? The importance of considering the evolutionary history of genetic lineages when assessing climatic niche evolution.
Design of the study, fieldwork, labwork, data analysis, paper writing – total contribution 35 %

* equal contribution

First two papers were published under my maiden name.

Acknowledgements

First of all, I would like to thank my supervisor Karol for his guidance and patience and for taking me on, even though we didn't know each other at the beginning. I really appreciate this opportunity and am grateful for the mentoring I received and the friendly environment I was able to be a part of and find my place.

I also want to thank a lot Eliška and Filip who "set me up" with Karol and introduced me to the secrets of evolution and its study and the idea that there are never enough collected populations. And of course to the other members of our Arabidopsis gang, past and present, it was a great ride!

I am very grateful to all my co-authors and colleagues who have contributed to this work in whatever way, directly or indirectly. Thank you all!

I am also indebted to all my friends in the department and adjacent groups who were always there when I needed to solve a problem or just hang out and talk. Thanks to Verča and Lenka for training in the lab, Fred, Kuba, Fanda and Hanka for the spectacular field trips, Lenka, Janča, Kristý, Bětka, Verča, Majda, Romča, Kuba, Pajarito, Pincík, Kuba, Tomík and Adam for the coffee breaks and friendly support, and especially Zuzi for her persistence and undying enthusiasm. I hope I will be able to return it to you all one day.

And finally, I would like to give a really big thank to my family, who have continued to support me for so long, without which I would not have been able to study. To my parents for the hours and hours of babysitting, to Václav for his endless patience, and to Martin for being my lucky charm.

Abstract

A prerequisite for addressing general questions concerning the evolution of intraspecific variability in space and time is the knowledge of the distribution of variability within the species' range. The development of molecular methods has been a major step forward, allowing various evolutionary questions to be addressed using natural populations of model species and their close relatives. Although wild relatives of *Arabidopsis thaliana* have long been in the focus of plant evolutionary biologists and molecular geneticists, the patterns of genetic diversity and phenotypic variation in their natural populations are often overlooked.

The present work focuses on some of the most studied model species in the Brassicaceae family, *Arabidopsis halleri* and the complex of *A. arenosa*, whose members are widely used to study ecology, physiology and evolution as well as the molecular basis of phytoremediation and parallel adaptation.

The study aimed to determine intraspecific variation at the ploidy level, to reveal phylogenetic relationships and the spatial distribution of genetic diversity across the range, and to propose a new taxonomic concept based on the detected intraspecific genotypic and phenotypic variation.

In order to accomplish this goal, we used DNA flow cytometry, several molecular methods (AFLP, SSR, cpDNA and single/low-copy gene sequencing, ddRADSeq, whole-genome sequencing), and multivariate morphometric methods, all based on dense population sampling across the distributional range of both groups.

In the solely diploid *Arabidopsis halleri*, we identified three major genetic lineages within Europe whose distributions were strongly correlated with major geographic barriers in the Central European mountain systems. Subsequent analysis of individual lineages revealed a further geographical distribution of the revealed diversity, resulting in five stable subgroups differing also on the basis of morphology, which allowed a new intraspecific classification of *A. halleri*.

The *Arabidopsis arenosa* species complex comprised three cytotypes, forming predominantly cytotype-uniform populations. Diploid and tetraploid cytotypes showed a predominantly parapatric distribution with three secondary contact zones. In the *A. arenosa* complex, five diploid and five tetraploid genetic lineages were found, with lineages with the same ploidy being geographically isolated (correlating with the biogeographic subdivision of Central Europe). The revealed intraspecific genetic lineages do not correlate with the current taxonomic concept of the *A. arenosa* species complex, which should be thoroughly re-evaluated.

Key Words

Arabidopsis halleri, Arabidopsis arenosa, phylogeography, genetic diversity, autopolyploidy, taxonomy

Abstrakt

Předpokladem pro řešení obecných otázek týkajících se vývoje vnitrodruhové variability v prostoru a čase je znalost rozložení variability v rámci areálu druhu. Velkým krokem kupředu byl rozvoj molekulárních metod, které umožnily řešení různých evolučních otázek za využití přirozených populací modelových druhů a jejich blízkých příbuzných. Ačkoli jsou volně žijící příbuzní *Arabidopsis thaliana* v centru pozornosti rostlinných evolučních biologů a molekulárních genetiků již dlouhou dobu, rozložení jejich genetické diverzity a fenotypové variability v přirozených populacích jsou často přehlíženy.

Předkládaná práce se zaměřuje na jedny z nejstudovanějších modelových druhů v čeledi Brassicaceae, druh *Arabidopsis halleri* a druhový komplex *A. arenosa*, jejichž příslušníci jsou hojně využíváni pro studium ekologie, fyziologie a evoluce i molekulárních základů fytoremediace nebo paralelní adaptace.

Cílem studie bylo zjistit vnitrodruhovou variabilitu na ploidní úrovni, odhalit fylogenetické vztahy a prostorové rozložení genetické diverzity v celém areálu výskytu a navrhnout nový taxonomický koncept založený na zjištěné vnitrodruhové variabilitě.

Výsledků bylo dosaženo pomocí DNA průtokové cytometrie, několika molekulárních metod (AFLP, SSR, sekvenování cpDNA a single/low-copy genů, ddRADSeq, celogenomové sekvenování) a mnohorozměrných morfometrických metod, to vše za použití souboru dat tvořeného populacemi z celého areálu rozšíření obou skupin.

U čistě diploidního druhu *Arabidopsis halleri* jsme v rámci Evropy identifikovali tři hlavní genetické linie, jejichž rozšíření silně korelovalo s hlavními geografickými bariérami v horách střední Evropy. Následná podrobnější analýza těchto linií odhalila celkem pět stabilních podskupin lišících se i na základě morfologie, což nám umožnilo navrhnout novou vnitrodruhovou klasifikaci druhu *A. halleri*.

V populacích druhového komplexu *Arabidopsis arenosa* byly nalezeny tři různé cytotypy, kdy naprostá většina populací byla cytotypově uniformní. Diploidní a tetraploidní cytotypy vykazovaly převážně parapatrické rozšíření se třemi sekundárními kontaktními zónami. V komplexu *A. arenosa* bylo nalezeno pět diploidních a pět tetraploidních genetických linií, přičemž linie se stejnou ploidií byly geograficky izolované (korelující s biogeografickým členěním střední Evropy). Odhalené genetické linie nekorelují se současným taxonomickým pojetím druhového komplexu *A. arenosa*, proto je nutné stávající koncept přehodnotit na základě zjištěných fylogenetických vztahů.

Klíčová slova

Arabidopsis halleri, *Arabidopsis arenosa*, fylogeografie, genetická diverzita, autopolyploidie, taxonomie

Contents

List of included papers

This thesis is based on the following publications:

I. Kolář F, Lučanová M, Záveská E, **Fuxová G**, Mandáková T, Španiel S, Senko D, Svitok M, Kolník M, Gudžinskas Z, Marhold K (2016) Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae), *Biological Journal of the Linnean Society*, Volume 119, Issue 3, Pages 673–688. https://doi.org/10.1111/bij.12479

II. Kolář F, **Fuxová G**, Záveská E, Nagano AJ, Hyklová L, Lučanová M, Kudoh H, Marhold K (2016) Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. Molecular Ecology, 25: 3929-3949. https://doi.org/10.1111/mec.13721

III. **Šrámková-Fuxová G**, Záveská E, Kolář F, Lučanová M, Španiel S, Marhold K (2017) Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction, *Botanical Journal of the Linnean Society*, Volume 185, Issue 3, Pages 321–342. https://doi.org/10.1093/botlinnean/box064

IV. **Šrámková G**, Kolář F, Záveská E, Lučanová M, Španiel S, Kolník M, Marhold K (2019) Phylogeography and taxonomic reassessment of *Arabidopsis halleri* – a montane species from Central Europe. *Plant Systematics and Evolution* 305, 885–898. https://doi.org/10.1007/s00606-019-01625-y
and (2020) Correction to: Phylogeography and taxonomic reassessment of *Arabidopsis halleri* – a montane species from Central Europe. *Plant Systematics and Evolution* 306, 60. https://doi.org/10.1007/s00606-019-01626-x

V. Padilla-García N, **Šrámková G**, Kolář F, Šlenker M, Zeisek V, Záveská E, Clo J, Lučanová M, Rurane I, Marhold K (manuscript prepared for submission) Niche differentiation following whole-genome duplication? The importance of considering the evolutionary history of genetic lineages when assessing climatic niche evolution.

The case studies are referred to by corresponding Roman numerals in the following text (CS I, CS II, CS III, CS IV, CS V).

PART A – GENERAL CHAPTERS

1 Introduction

Evolution represents a change in the variation of population characteristics that are heritable and change over generations through mutations. New alleles may encode novel functional traits and, together with gene flow, genetic drift and natural selection, trigger the speciation. This is, however, intermingled with additional processes, particularly in the plant kingdom. The polyploidization, hybridization and historical aspects enter the game and shape – until then simple – speciation.

The current state of biodiversity is the result of years of evolution and the traces are still detectable in the DNA of survivors. But the mere study of genetic information does not provide full-scale data on the distribution of populations and species. In order to reveal the impacts of evolutionary processes, it is necessary to study also geographical distribution of genetic variation and compare it to the phylogeny (Hewitt 2001; Avise 2009). Phylogenetic relationships among genetic lineages set in the biogeographical context, commonly known as phylogeography (introduced by Avise et al. 1987), can shed light on the species history and can help us to classify studied lineages into correct taxonomy. Such classification was previously frequently based only on phenotypic variation, which, however, results from both genotype and the influence of surrounding environmental factors and is therefore often misleading (Dubois 2003; Zink 2004). And a clear understanding of why and how lineages diversify and their valid taxonomy is required to efficiently manage and protect biodiversity for the future.

In an ideal situation, we would study populations from all corners of the species distribution, have the appropriate dating methods and geographical history of the area as well, and in conclusion, based on phylogeography, we would be able to assess the microevolutionary processes operating within the species and with subsequent extrapolation to explain macroevolutionary differences among species and higher taxa. Unfortunately, this scenario is a utopia, as generating this amount of data and their processing would be technically and financially impossible. Nevertheless, we should strive for the best possible conditions. Thanks to mathematicians and theoretical population geneticists we no longer need to model entire populations as coalescent theory allows us to use only the sample of alleles to model gene genealogies and estimate phylogeographic parameters (Wakeley 2009). Yet, we still need to obtain a representative sample of alleles of the studied complex if we want to receive representative results from the analytical tools. Therefore, the selection of sampled populations, as well as marker selection, plays a key role. Partial knowledge of the species distribution and incomplete sampling of populations are often a source of the incomplete framework, where part of the genetic variability can stay hidden (CS III, CS V; Avendaño et al. 2017), equally insufficient variability of molecular markers can lead to neglecting the deeper structure inside the dataset. This could result in postulating improper biogeographic patterns or evolutionary processes and further into incorrect taxonomic classifications.

1.1 Europe as one of the centres of biodiversity

The major impact on the distribution, variation and evolution of the current temperate organisms caused Pleistocene climate fluctuations (2.6–0.01 Mya; e.g. Hewitt 2000, 2004; Qiu et al. 2011). They induced the range shifts of all species, changed their connectivity and the amount and distribution of intraspecific genetic variation. Although the impact of climatic changes depends on latitude and topography and vary among different taxa, common phylogeographic trends can be found. It was hypothesised that temperate species migrated south to overcome climatically unfavourable periods and established themselves in glacial refugia, where they were more likely to survive in a more stable climate (Taberlet et al. 1998). Initial studies revealed the glacial survival in the three main southern European refugia - the Balkan, Iberian and Apennine Peninsulas - and the postglacial recolonization of previously periglacial (central Europe) and glacial areas (northern Europe and the central European mountains) to the north (Taberlet et al. 1998; Hewitt 1999, 2004; Stewart and Lister 2001; Tzedakis et al. 2013). Subsequent studies, based on increased sampling and detailed analyses of genetic variation at the population level, as well as paleoecological data, have also brought forward evidence of glacier survival in non-Mediterranean areas in Central and Eastern Europe (Fig. 1; Stewart and Lister 2001; Willis and Van Andel 2004; Bhagwat and Willis 2008; Provan and Bennett 2008; Juřičková et al. 2014), as in the Carpathian Mountains, which were only sparsely glaciated (Ronikier 2011) and hosted forest communities during the Last Glacial Maximum (LGM; Jankovská and Pokorný 2017). In plants, the majority of examples of Carpathian LGM survival can be found among the alpine (reviewed in Ronikier 2011) and montane taxa (Magri et al. 2006; Těšitel et al. 2009), but studies using a system spanning over a wide altitudinal range were missing.

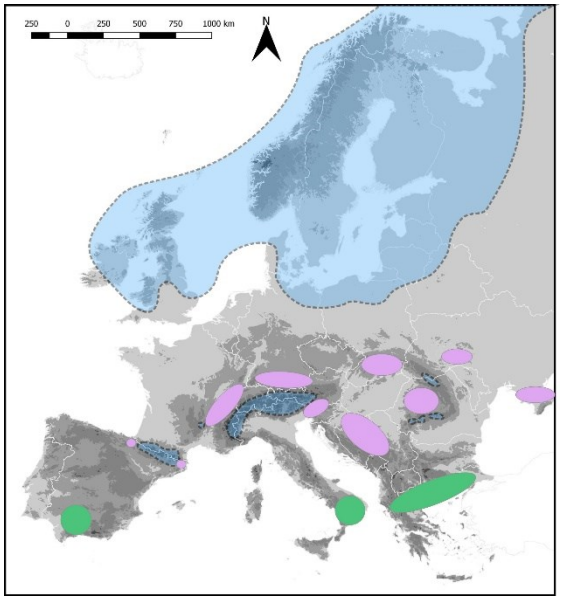


Fig. 1 General ideas about the general location of LGM refugia in the early Holocene. In *green* three major southern European glacial refugia; in *pink* northern "cryptic" glacial refugia; in *blue* the maximal extent of the continental ice sheet during LGM. Based on figures in Schmitt and Varga 2012.

Repeated cycles of glaciation-induced range shifts and isolation of different populations in refugia resulted in a modified possibility of interbreeding and accumulation of distinct mutations. The subsequent postglacial range expansion from southern refugia throughout the present species' distribution range also provided opportunities for secondary contacts of genetically differentiated lineages, creating sharply delineated contact zones where lineages have met but did not penetrate, or broader transition belts with intermingling lineages or admixed populations (Hewitt 1999; Stewart et al. 2010). By comparing different phylogeographic patterns it is possible to find such contact hybrid zones, which

tend to cluster along the Alps, the Pyrenees, western and eastern Central Europe and Central Scandinavia (Fig. 2; Taberlet et al. 1998; Hewitt 1999, 2004; Schmitt 2007). Similarly, also biogeographical barriers played a role during the recolonization and hybridization, but their presence is harder to reveal, as the genetic differentiation between geographically close populations could have historical reasons or simply insufficient marker resolution, therefore a comparison of numerous species' phylogeographic patterns is crucial.

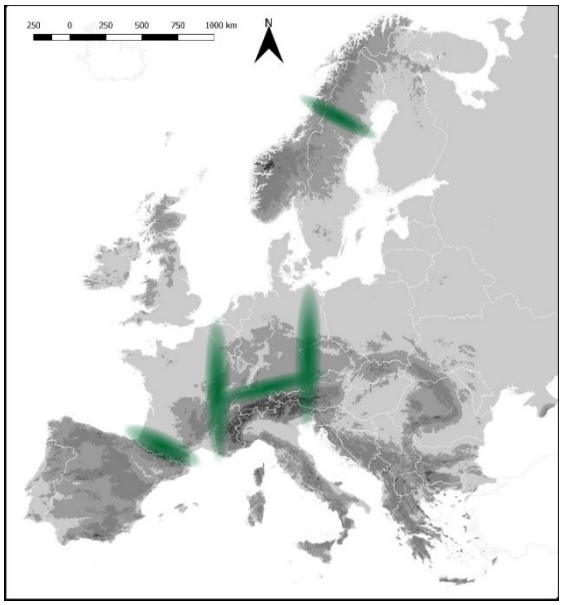


Fig. 2 The secondary contact and hybridization zones during the postglacial range expansion, in *green*. Modified from Schmitt 2007.

From this point of view the best-explored areas, profiting of a thorough study of plants of alpine and subalpine belts, are mountain systems (e.g. Schmitt 2009; Thiel-Egenter et al. 2009; Meirmans et al. 2011; Ronikier 2011; Ronikier and Zalewska-Gałosz 2014; Zozomová-Lihová et al. 2015; Kirchheimer et al. 2016; Knotek and Kolář 2018). In Europe, we can find a large number of high mountain systems varying in their size area and elevation. The island-like distribution of this harsh higher-altitude environment and its phylogeographic patterns has been studied over the last two decades, mainly for the species surviving glaciation within refugia in close proximity. Colonization and recolonization (from one or more refugia) resulted in harbouring higher genetic diversity on the genome level and endemism on the species level as well, creating biodiversity hotspots (e.g. Schmitt 2009; Schmitt and Besold 2010; Bálint et al. 2011; Mráz and Ronikier 2016). Each mountain system has its own most common pattern of genetic differentiation that is detectable in phylogeographic studies, when observed genetic lineages can be explained by several microrefugia in the perialpine areas serving as centres of dispersal. Although the distribution patterns differ from species to species, major phylogeographic trends are emerging. In the Alps, we mostly find the pattern of four or two genetic groups with the strong biogeographic border between the Western and Eastern Alps (Schönswetter et al. 2005; Thiel-Egenter et al. 2011), the Pyrenees host most often two genetic lineages within species (e.g. Kropf et al. 2002; Schmitt et al. 2006; Lauga et al. 2009) as well as the Carpathians (Mráz and Ronikier 2016), where multiple glacial refugia can be also

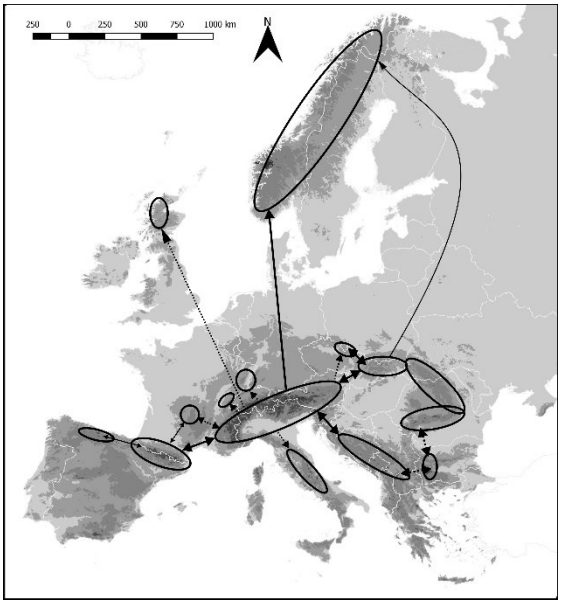


Fig. 3 Biogeographical links detectable between individual mountain ranges. *Bold arrows* indicate commonly observed sharing of identical genetic lineages; *solid arrows* show frequent pattern; *dotted arrows* indicate relatively few known cases; direction of arrows indicate the direction of exchange. Modified from Schmitt 2017.

found along the entire mountain arch (Thiel-Egenter et al. 2011; Ronikier et al. 2012; Taberlet et al. 2012). Individual mountain ranges also do not represent isolated genetic entities; multiple biogeographic links between them can be detectable as identical genetic lineages present in different areas (Fig. 3).

1.2 Speciation as an outcome of evolution

There is no doubt that evolution induces changes in populations, with various barriers entering the process, leading to population separation and increasing genetic isolation. This process can lead up to speciation, and because it is still evolving, the outcome – the species – is not easy to define. On the other hand, the term species is used frequently and is in the centre of cascading effect especially in the nature conservation – loss of populations of the species, loss of ecosystem diversity counted by the number of species, species survival, etc. The ability to identify individual species and use the correct taxonomic concepts is crucial for understanding the community structure and function (Knowlton and Jackson 1994; Zachos 2016). It is also important for many consequent disciplines, from the simple understanding of the subject of study and the possibility of inter-laboratory comparison of results, the position in phylogeny and the knowledge of relatives, to accurate conservation strategies, the correct selection of biological entities for ecological and behavioural studies, as well as phytomedical and chemical research.

Although the "species" is one of the most important taxonomical units in biology, its definition is still controversial. Species concepts not only define what a species is, but by the definition of what a species is, they also define the speciation. Therefore, the definition of the term "species" based on the different species concepts have been endlessly discussed and many species concepts have been proposed based on different aspects of the speciation, apparently often resulting in different conclusions (Agapow et al. 2004; Isaac et al. 2004). But whether we want it or not, taxonomic names allow us to assess species boundaries and phylogenetic relationships of taxa (Godfray 2007).

In plants, the species concept is challenged, and understanding the nature and genetics of the factors that restrict gene flow between species, and the conditions under which this isolation occurs, is complicated by two evolutionary processes – hybridization and polyploidization. Therefore, confession of the importance of semipermeable reproductive barriers plays a key role in defining species concepts in plants. Despite the ongoing hybridization and potential introgression of the genes (at least 25% of plant species; Mallet 2005), the species can differentiate and persist. Moreover, depending on the frequency of intermating and the fitness of resulting hybrids, the extensive gene flow may result in the extinction of taxa via genetic assimilation (e.g. Ayres et al. 2004; Genovart et al. 2012) or merging two taxa into a single lineage (e.g. Hegde et al. 2006; Taylor et al. 2006). In contrast, reduced fitness of hybrid progeny can result in forming a stable hybrid zone allowing introgression of some alleles (e.g. Brodin and Haas 2009; Pinto et al. 2019) and at least partial reproductive isolation of hybrids can result in the production of hybrid neospecies (Chapman and Burke 2007), both scenarios preserving parental taxa. An alternative way of creating immediate isolation of hybrid offspring is via chromosome doubling, polyploidization. Two general types of polyploids are defined – autopolyploids, which arise by the multiplication of one chromosome set, and allopolyploids, resulting from the fusion of structurally different chromosome sets (Tate et al. 2005). The later ones can generate new genetic variation that may be important for the adaptation of hybrid species to new habitats that are different from those of their

parents, thereby aiding the reproductive isolation of the hybrid from its parents (Rieseberg et al. 2003). Autopolyploids can, on the other hand, speciate without hybridization, simply by having multiple chromosome sets that cause problems during meiosis during backcrossing. This, however, also brings the advantage of the multiplication of alleles in the locus, which leads to the possibility of non-lethal accumulation of mutations with subsequent differential expression and, of course, ecological and evolutionary consequences. Even though the study of polyploidy has risen considerably during past decades (Soltis et al. 2014), the amount of autopolyploids is likely underestimated, because taxonomists often do not recognize autopolyploids as a separate species (Soltis et al. 2007). These taxonomically cryptic autopolyploid cytotypes may represent a substantial part of plant diversity, given the fact that half of the polyploids are autopolyploids, with only ~7% taxonomically described (Barker et al. 2016). On the other hand, the delimitation of newly emerged autopolyploid species (and the necessity of their taxonomic classification) is disputable because mechanisms of direct differentiation from parents, both phenological and genetic, are lacking.

1.2.1 Subspecies – the necessary taxonomical units?

If the delimitation of species is a complex process, delimiting the subspecies is even more difficult due to the expected ongoing gene flow. Subspecies is defined as "a population, or collection of populations, that appears to be a separately evolving lineage with discontinuities resulting from geography, ecological specialization, or other forces that restrict gene flow to the point that the population or collection of populations is diagnosably distinct" (Taylor et al. 2017b). Therefore, it is obvious that a subspecies can be only defined in relation to a species and at least two subspecies are needed for the idea to make sense. Yet subspecies were described among various taxa and play an important role in conservation and evolutionary studies (e.g. Haig et al. 2006; Phillimore and Owens 2006; Taylor et al. 2017a; Ramírez-Rodríguez and Amich 2019). Unfortunately, the focus on subspecies could misdirect conservation efforts, if existing subspecies are not genetically distinct, for example, if they were delimited only by arbitrarily chosen morphological characters (Zink 2004). To avoid similar discrepancies and different subjective definitions within various taxa, biological threshold needs to be established, evaluating both the lower (population vs. subspecies) and upper (subspecies vs. species) boundaries of a subspecies concept (Martien et al. 2017). The divergence between putative taxa within these boundaries need to be tested by an integrative approach to combine the results from several different analytical methods and data types (Padial et al. 2010; Tobias et al. 2010; Patten and Remsen 2017) and with the obtained knowledge the taxonomic frameworks can be set. The subspecies as a taxonomical unit represents an important stepping stone between populations and species and provides an opportunity to describe geographic patterns of variation in nature that are characterized as lineages in the scientific papers, but do not reach the "ordinary" world of government officials and environmental protection.

1.3 Genus Arabidopsis

Representative of the tribe Camelineae, the genus *Arabidopsis* as currently circumscribed comprises about fourteen species worldwide, including those that were previously classified within a separate genus *Cardaminopsis* (O'Kane and Al-Shehbaz 1997, Shimizu-Inatsugi et al. 2009, Kiefer et al. 2014 – Brassibase). The first step towards such classification was made in 1987 (Greuter et al. 1988) based on the proposal by Štěpánek (1983) when the name *Arabidopsis* Heynh. was conserved with the different type (*Arabidopsis thaliana*) and as a result of this, the genus comprised many different species based on the morphology of fruits and seeds (*Arabis*, *Braya*, *Cardaminopsis*, *Cymatocarpus*, *Halimolobus*, *Nasturiopsis* etc). Subsequently, after the first phylogeny of the group was published and morphological reassessment was done, only nine generally accepted species were left in the genus (O'Kane and Al-Shehbaz 1997) – the *Arabidopsis thaliana* (L.) Heynhold and its "wild relatives" – *A. arenosa* (L.) Lawalrée, *A. neglecta* (Schult.) O'Kane and Al-Shehbaz, *A. croatica* (Schott) O'Kane and Al-Shehbaz, *A. lyrata* (L.) O'Kane and Al-Shehbaz, *A. halleri* (L.) O'Kane and Al-Shehbaz, *A. cebennensis* (DC.) O'Kane and Al-Shehbaz, *A. pedemontana* (Boiss.) O'Kane and Al-Shehbaz and *A. suecica* (Fr.) Norrl. The diploid model species *Arabidopsis thaliana* is under intensive study for more than fifty years, being the first sequenced plant genome (*Arabidopsis* Genome Initiative 2000), making the genus a leading model in plant science. And thanks to the knowledge of the functions of its 27000 genes and 35000 coding proteins present on five chromosomes, it is more than convenient to study also wild populations of the other species from the genus. Unfortunately, the circumscription, taxonomy and evolutionary relationships among those wild relatives were and still are only poorly known (Koch and Matschinger 2007) and as a consequence, a vast number of experimental studies used a misleading taxonomy or an inaccurate phylogenetic framework. This not only hampers comparison of results among studies but there is also a risk that wrong conclusions may be drawn within a misleading evolutionary context (Koch et al. 2008). This concerns especially the evolutionary lineages of *A. arenosa* and *A. halleri*.

1.3.1 Arabidopsis halleri

The purely diploid taxon *Arabidopsis halleri* (2n=16, Kolník and Marhold 2006) is a clonal, self-incompatible and highly outcrossing perennial weed. Thanks to its ability to colonize contaminated sites with high content of heavy metals (Zn, Cd) and to absorb high amounts of these metals in leaf tissues, this species has become a key model for studying the genetic basis of heavy metal hyperaccumulation and phytoremediation (e.g. Willems et al. 2007; Pauwels et al. 2008, 2012; Roosens et al. 2008; Verbruggen et al. 2009; Krämer 2010; Bomblies and Weigel 2010; Stein et al. 2017; Stolpe et al. 2017; Preite et al. 2019). The species has diverse geographical distribution from lowlands to subalpine zones and represents rarely studied montane species. Despite the absence of polyploidy, high morphological variability can be observed, which in combination with geographical diversity was probably responsible for the unresolved taxonomy of this group.

The centre of diversity is in Central Europe, where four of the five taxa occur (Fig 4; Jones and Akeyrod 1993; Jalas and Suominen 1994; Kolník and Marhold 2006). *A. halleri* subsp. *halleri* (L.) O'Kane and Al-Shehbaz is the only one widely distributed throughout Central Europe. Other subspecies show geographically separated or isolated distribution. *A. halleri* subsp. *tatrica* (Pavl.) Kolník is an endemic of the Western Carpathians (Tatry Mts.), *A. halleri* subsp. *dacica* (Heuff.) Kolník occurs in the Eastern and Southern Carpathians and the Balkans, and *A. halleri* subsp. *ovirensis* (Wulfen) O'Kane and Al-Shehbaz in the Eastern Alps (Kolník and Marhold 2006; Koch et al. 2008). The fifth taxon – *A. halleri* subsp. *gemmifera* (Matsum.) O'Kane and Al-Shehbaz – occurs in eastern Russia, northeastern China, Korea, Japan and Taiwan, and is one of the ancestors of the polyploid *Arabidopsis kamchatica* (Shimizu-Inatsugi et al. 2009). The genetic structure of European populations was studied using different molecular markers (Pauwels et al. 2008, 2012; Wasowicz et al. 2015), unfortunately, these studies did not contain sufficient and balanced sampling, especially in the southern and south-eastern Carpathians and the Balkans (Kolník and Marhold 2006). This made it impossible to unambiguously confirm the genetic structure and assess the authenticity of published mountain subspecies. At the same time, the origin and age of populations of *A. halleri* subsp. *gemmifera* in Japan was not clarified, whether it is the result of an ancient vicariance or a modern long-distance dispersal.

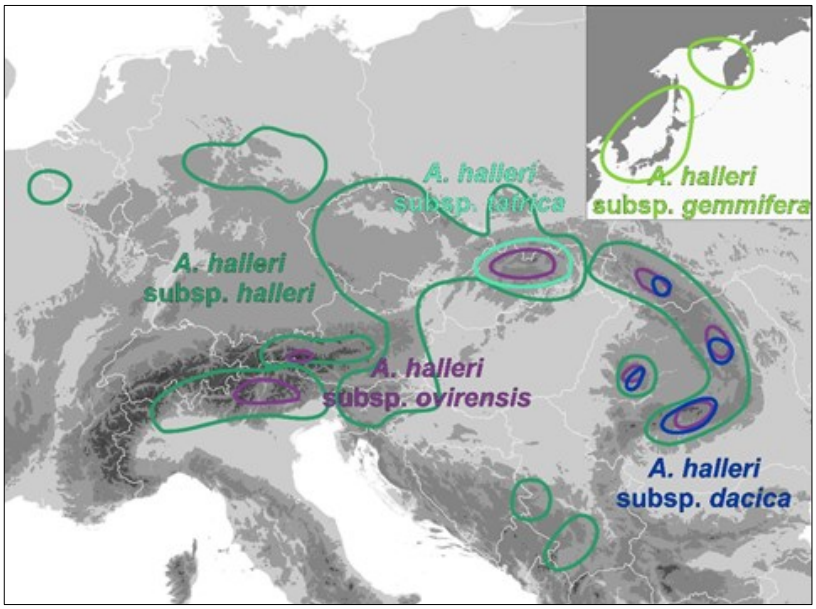


Fig. 4 Distribution of the described and undescribed taxa of the *Arabidopsis halleri* in Central Europe and East Asia (based on Měsíček 1970, Kolník and Marhold 2006).

1.3.2 Arabidopsis arenosa

Arabidopsis arenosa is a colline, montane and subalpine species complex currently comprising a group of poorly defined diploid (2n=16) and tetraploid (2n=32) taxa with the highest diversity in the Central Europe (and in the Carpathians in particular; Koch and Matschinger 2007), involving also triploids and aneuploids resulting from widespread hybridization events (Měsíček 1970; Kolník and Marhold unpubl.). Its genome structure was unknown, though a close resemblance to the n=8 karyotype of *A. lyrata* (Kuittinen et al. 2004; Lysak et al. 2006) could have been expected. Besides the polyploidy and hybridization, the situation was further complicated by a large but structured variation in morphology, ecological preferences (the individuals occupy a wide range of habitats from alpine regions to sandy coastal dunes, some populations spreading into man-made habitats) and breeding strategies (probably different levels of selfing vs outcrossing; Clauss and Koch 2006). The species complex consists of biennials and short-lived perennials.

Only few studies have been attempted to unravel the evolutionary history of the *Arabidopsis arenosa* complex (e. g. Měsíček 1970; Koch and Matschinger 2007; Schmickl et al. 2012; Arnold et al. 2015) and even fewer tried to solve the taxonomical concept based on the morphological data. The most thorough study of populations, but only from the Carpathian region, was undertaken by Měsíček (1970) based on the morphological and karyological data and resulted in recognition of several species and subspecies (at that time attributed to the genus *Cardaminopsis*) that were, however, never validly published and are kept as *nomina provisoria* (nom. prov.). This all resulted in inconsistent taxonomical solutions within different studies, where we can find up to five species and various subspecies (Fig. 5; e.g. Měsíček 1970; Novikova et al. 2016; Koch 2018), more or less divided by the ploidy. This could be a reason, why most of the evolutionary and experimental studies, including papers comprising this thesis, recognize only a single widely conceived species, *A. arenosa* (e.g. Yant et al. 2013; Arnold et al. 2015; Baduel et al. 2016) while the systematically oriented or local studies present more taxa (species or subspecies; e.g. Měsíček and Goliašová 2002; Schmickl et al. 2012).

Based on Měsíček (1970), the *Arabidopsis arenosa* (L.) Lawarlée s.str. is further differentiated into *A. arenosa* subsp. *arenosa* (2n=16/32), with distributional range in Central and Western Europe and lower altitudes of Scandinavia, and *A. arenosa* subsp. *borbasii* (Zapal.) O'Kane and Al-Shehbaz (2n=32), inhabiting mountain ranges in Central and Western Europe. Another di-tetraploid species of the complex is *Arabidopsis neglecta* (Schult.) O'Kane and Al-Shehbaz found in the Carpathian mountains, differentiated into subspecies based on the ploidy, *A. neglecta* subsp. *neglecta*, nom. prov. (2n=16) and *A. neglecta* subsp. *robusta,* nom. prov. (2n=32). *Arabidopsis petrogena* (A. Kern.) V.I. Dorof. is another species found in the lower altitudes of the Carpathians and Pannonian lowland, with subspecies *A. petrogena* subsp. *petrogena,* nom. prov. (2n=16) and *A. petrogena* subsp. *exoleta*, nom. prov. (2n=32). Within the complex, we encounter two more purely diploid species, *Arabidopsis carpatica* nom. prov. (2n=16), found on the limestones of the Western Carpathians, and *Arabidopsis nitida,* nom. prov. (2n=16), also found in the Carpathian mountain ranges, at middle to subalpine altitudes.

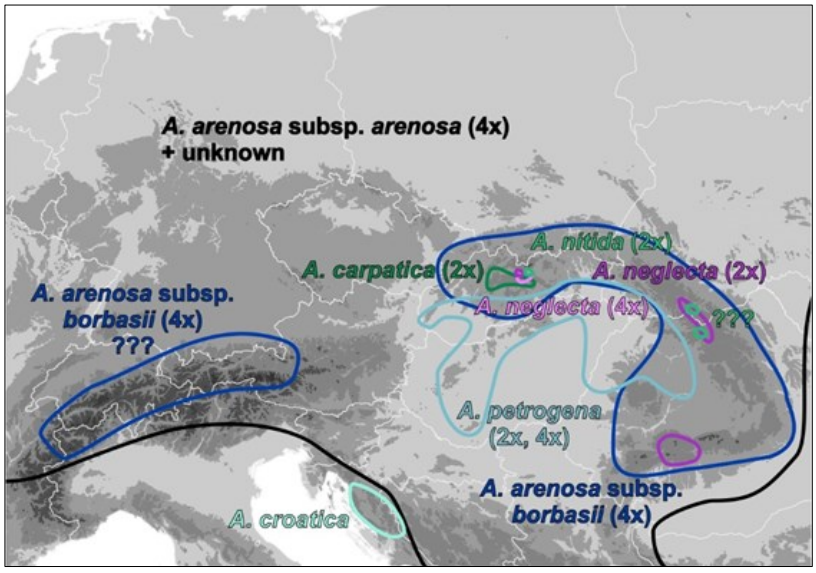


Fig. 5 Distribution and ploidy levels of the described and unde-scribed taxa of the *Arabidopsis arenosa* group in the Central Europe (based on Měsíček 1970).

2 Objectives

The main aim of the thesis is to obtain a detailed knowledge of the cytological and genetic structure of two wild relatives of *Arabidopsis thaliana* based on extensive sampling of the *Arabidopsis arenosa* complex and *A. halleri,* covering their entire natural distribution area. These two species have become very promising and eventually important model taxa for numerous genomic and genetic studies and experiments in recent years, however, their phylogenetic relationships and evolutionary histories were not sufficiently explored, as discussed in the previous chapter.

Therefore, the primary focus of presented case studies is to reveal the geographic distribution of distinct diploid/tetraploid genetic lineages (Case Study II, III, V) preceded by ploidy distribution investigation (Case study I), to infer the range-wide patterns of genetic diversity together with evolutionary history (Case study II, III) and finally to provide taxonomic treatment of intraspecific variation with the use of morphological characters (Case study IV).

Specifically, I have addressed these particular objectives with the following questions:

A) To investigate ploidy and genome size variation across the distributional ranges

- What is the geographic distribution of diploids/triploids/tetraploids within the *A. arenosa* complex and *A. halleri* in Central Europe?
- Is ploidy level variation present within populations? Are there any contact zones between distinct ploidies?
- How does the genome size vary within the diploids and tetraploids? If there are homoploid differences in genome size, do they correspond with the distinct genetic lineages?
- What are the ploidy level and putative area of origin of the plants of the *A. arenosa* complex currently spreading through man-made habitats in Central Europe?

B) To reveal phylogenetic relationships and range-wide patterns of genetic diversity

- What is the genetic variation across the ranges of both studied complexes?
- Do these genetic structures correlate with geography or ecology (e.g. for alpine vs. montane habitats)?
- How many polyploid lineages have originated within the *A. arenosa* complex in Central Europe? Did the tetraploids overcome their diploid ancestors in the occupied area and/or range of colonized habitats?

C) To reveal intraspecific morphological variation and propose a taxonomical concept based on it

- What is the modelling factor of morphological variation – genetics or environmental gradients? Are there any morphological characters defining the major genetic lineages within species?
- Do the revealed genetic lineages affect the current taxonomic concept?

3 Materials and methods

Plant material used in this study was collected in order to achieve homogeneous sampling across the entire distributional range of the *Arabidopsis arenosa* complex and *A. halleri* in Europe and to include all previously recognised European species and subspecies of the complexes. In addition, non-European subspecies of *A. halleri* was also included by sampling populations spanning most of its distribution in Japan. Sampling used in presented case studies represents the largest datasets published, which forms the most comprehensive basis for the study of genetic relationships within these taxa. Bearing in mind the future necessity of taxonomic conclusions and possible morphological variability caused by phenotypic plasticity, herbarium vouchers, environmental data (such as habitat and phytocenological relevés) and GPS coordinates were also sampled in addition to silica gel dried material. Here I present methods applied in the case studies that were used to achieve the objectives of this thesis.

The ploidy level of all individuals was assessed by **flow cytometry**, complemented with chromosome counts and homoploid genome size diversity analyses for some accessions (detailed protocol in CS I). The knowledge about the ploidy of each individual based on the combined knowledge of chromosome counts and relative genome size is crucial for further molecular analyses, whether for AFLP and SSR approach or optimizing NGS methods and obtaining genome-wide SNPs. When combined with the geographic structure, the obtained cytogeographic data from sufficiently large cytometric screens form a substantial part for complex evaluation of the diversity and dynamics of ploidy-mixed plant systems (Kron et al. 2007).

To reveal phylogenetic relationships, we employed several **molecular methods** providing a different level of particularity, leveraging the wide spectrum of molecular approaches available for the closely related model species *Arabidopsis thaliana* and the model systems themselves.

The first studies were based on simple, albeit the most suitable methods at that time. **AFLP** (amplified fragment-length polymorphism, CS III, CS IV), the DNA-fingerprinting technique allowed the detection of DNA polymorphisms across different genome regions (Vos et al. 1995; Meudt and Clarke 2007) and thus provided a large number of polymorphisms, it was suitable for cases of limited variability of other markers (Tribsch et al. 2002). Unfortunately, the method is not able to provide any information about the heterozygosity of studied populations and also its reproducibility is suboptimal. Therefore, co-dominant nuclear microsatellite markers (**SSR**) were amplified to overcome this disadvantage (CS II, CS III). SSRs exhibit a high degree of polymorphism due to a large number of alleles per locus (Valdés et al. 1993) and therefore they are suitable to assess genotypic variation within and among populations (Jarne and Lagoda 1996). This method was effectively used for detecting pathways of introduction/colonization, for comparison of the amount of genetic variation in native and introduced populations, in the evaluation of the role of introgression, homoploid hybridization, of the impact of different reproductive systems on the population structure and, despite its possible limits in the uncertain allele composition inference in polyploids, also for polyploidization (Dobeš et al. 2004; Durka et al. 2005; Jørgensen et al. 2008). We took the advantage of published primers (Clauss et al. 2002; Schmickl et al. 2011) and employed them, gaining diversity indices also at the population level. This method is highly reproducible and obtained datasets can be supplemented with new samples that are set into the context, which was used later for geographically or morphologically

problematic samples collected after performing the main study. In order to infer ancestral relationships, we employed **non-coding chloroplast DNA** sequences, the fundamental molecular markers widely employed in plant species-level phylogenetics, taxonomy and phylogeography. The cpDNA regions are inherited uniparentally and they lack recombination. In presented studies (CS III, CS IV) three variable plastid DNA regions were designed (based on Novikova et al. 2016) and the most variable was used to construct a plastid DNA haplotype network. As an additional marker to infer approximate divergence dates of *Arabidopsis halleri* and its genetic subgroups, a nuclear gene encoding the enzyme chalcone synthase (CHS) was used (CS III). **Single- or Low-copy nuclear markers** (Strand et al. 1997), such as CHS, are useful for the reconstruction of phylogeny at the generic and species level as well as for identifying parental donors of suspected hybrids or polyploids. Despite potential drawbacks that may occur (complex genetic architecture, orthologous genes) several studies successfully employed various single-copy loci for the phylogenetic inference within Brassicaceae (Bailey and Doyle 1999; Bailey et al. 2006).

Thanks to the rise of **high-throughput sequencing techniques** and the importance of the genus *Arabidopsis*, many wet-lab library preparation methods were optimized for the model system. Our team took the opportunity to learn different library-prep methods abroad from our international collaborators and employed NGS techniques as one of the first at the faculty.

The genome-wide single-nucleotide polymorphism (SNP) markers were obtained using **double-digest RADseq** (restriction-associated DNA sequencing; (Peterson et al. 2012)). This genotyping method samples target genomes at reduced complexity and allows identification of single-nucleotide polymorphisms (SNPs) at putatively homologous loci across many individuals without any prior genomic information on the studied taxa (Baird et al. 2008; Andrews et al. 2016). The RADseq approach was used to delineate relationships among populations and assess the strength and (a)symmetry of gene flow among species, lineages and cytotypes. The use of two enzymes instead of one (double-digest protocol) was selected to obtain sufficient resolution (tens of thousands of reliable SNPs) at lower costs (Peterson et al. 2012). Genome-wide SNPs obtained from the first double-digest-library-prep protocol in CS II were used to reconstruct genetic differentiation of diploid populations of *Arabidopsis arenosa*, their phylogenetic relationships and to test the admixed origin of the diploid Baltic group. Unfortunately, this protocol contained AT-rich enzymes and therefore also AT-rich genome regions were addressed. This resulted in a lower yield of variable SNPs and the gain:price ratio was unsatisfactory. Therefore, the second double-digest RADseq protocol was established (Wos et al. 2019; Knotek et al. 2020), CS V), utilizing only one restriction enzyme with a balanced AT:CG ratio (protocol in Wos et al. 2019). Obtained SNPs were used to reconstruct the evolutionary history of diploids and tetraploids of *A. arenosa* in European alpine-foothill populations, to test the consequences of polyploidization on genetic diversity and parallel alpine differentiation (Wos et al. 2019, Knotek et al. 2020).

In CS V the genetic structure of *Arabidopsis arenosa*'s tetraploids was investigated across the entire distributional range for the first time, containing, in addition to the RadSeq data, also data from **whole-genome sequencing (WGS)** protocol. WGS represents a powerful tool to address a wide range of questions (from phylogenetic reconstructions to genome-wide association studies) by sequencing the whole genomes of the targeted samples. The data obtained by this method represent a large pool of information researchable by different fields of interest; here we addressed the RADseq loci from the WGS data and incorporated them into the RADseq dataset.

I have successfully established those three different NGS protocols in the DNA lab of the Department of Botany, Charles University – two double-digest protocols from Hiroshi Kudoh's lab (University of Kyoto; CS II) and Kirsten Bomblies's lab (Harvard

University; Wos et al. 2019), and an efficient low-cost whole-genome sequencing protocol from Darren Heavens (Earlham Institute, Norwich; Perez-Sepulveda et al. 2020). This represented a breakthrough in the use of NGS methods at our Department and those protocols have been optimized by me for many other species since then.

Significant morphological differences between foothill and alpine populations (colour of petals, plant height), that led up to taxonomic differentiation of species and subspecies, indicated the need to reevaluate the morphological variation of the genus in addition to the genetic structure. Therefore, morphometric data obtained from herbarium vouchers were also collected during the project and the **multivariate morphometric methods** were carried out to examine the relative contributions of the environment, genetic structure and geographical distances of *A. halleri* (CS III) and to test the sufficiency of morphological differentiation of the genetic lineages of *A. halleri*, including the detection of the best characters for the identification of new intraspecific classification (CS IV). Both generative and vegetative traits were measured in order to cover the diversity known from the Brassicaceae family complicated by the parallel evolution of most of the morphological characters (Walden et al. 2020) and as both could be important in detecting eg. hybrids or ploidy levels (Macková et al. 2018). As the differential morphological characters within polyploid complexes in Brassicaceae are mostly of a quantitative nature, standard multivariate morphometric methods (cluster analyses, principal component analyses, principal coordinate analyses and discriminant analyses) were proved in numerous studies as most appropriate (Španiel et al. 2012, 2017).

4 Key results and discussion

The aim of this research was based on the lack of information about the wild species of an important model genus *Arabidopsis* from the Brassicaceae family. The genomic studies of *Arabidopsis thaliana* had suggested that its wild relatives would soon be on the rise when it comes to studying the role of genes and their expression, phytoremediation, or the effect of polyploidization *per se* on several factors. Therefore, the knowledge about relationships and the geographical structure of genetic lineages of *Arabidopsis halleri* and the *A. arenosa* complex was the first step in creating a clear phylogenetic and taxonomic framework that can be used in follow-up studies.

Given the need to resolve the evolutionary history of the *Arabidopsis arenosa* complex as a whole, we decided not to distinguish previously described taxa of this complex as evolutionary units and to use ploidy alone as a parameter for creating the studied datasets. In the following text, *A. arenosa* is considered as a broadly conceived species consisting of all the taxa presented in the chapter 1.3.1, built upon the monophyly of the whole group (Hohmann et al. 2014).

4.1 Ploidy level and genome size variation across the distributional ranges

In the first place, it was necessary to obtain a detailed knowledge of the geographic distribution of cytotypes. For ***Arabidopsis halleri***, this meant mainly validating the solely **diploid** status of this genus. All accessions from 136 populations including also *A. halleri* subsp. *gemmifera* from East Asia (Fig 6), confirmed this premise, except for one triploid individual found on the border of Switzerland and Italy. Diploid individuals showed only little variation in nuclear DNA content (CS III).

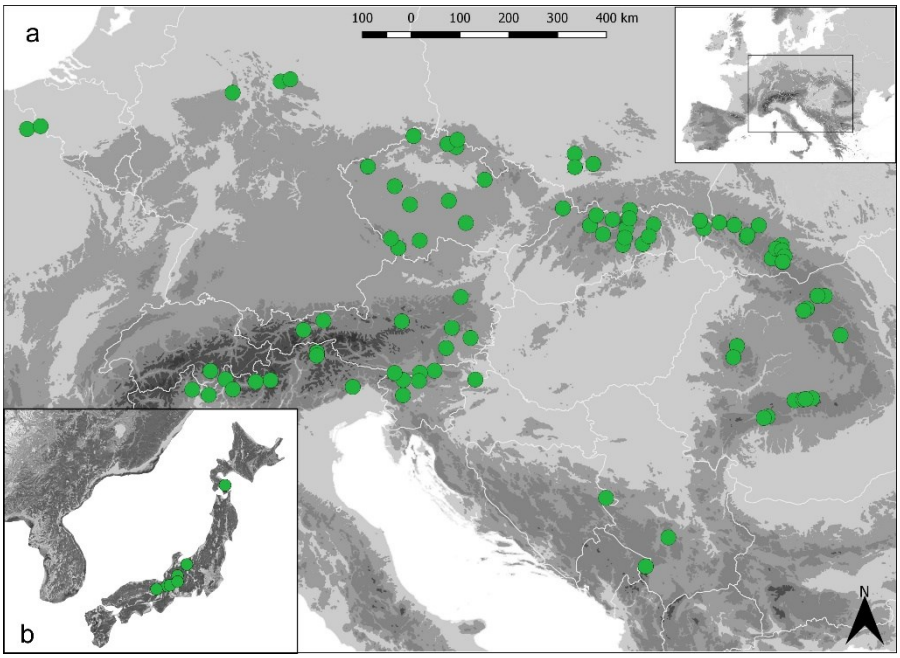


Fig. 6 Distribution of the 136 sampled populations of the *Arabidopsis halleri* **a** in Europe; **b** in Japan.

The situation in the ***Arabidopsis arenosa*** complex was much more interesting since it is a known diploid-autotetraploid group (chromosome counts by Měsíček 1970, Krendl and Polatschek published in Schmickl et al. 2012), especially in the Western Carpathians, where the coexistence of different cytotypes along the altitudinal gradient within different environments was reported (Měsíček and Goliašová 2002). In order to obtain sufficient cytogeographic data, individuals from 496 populations (the most complete sampling performed by our group until now) were ploidy-checked, resulting in **three different DNA ploidy levels** – diploid 2x, triploid 3x and tetraploid 4x (Fig 7; CS I, CS II, CS V). The representation of diploids and tetraploids was almost equal, the triploid cytotype was, on the other hand, extremely rare. The majority of populations was uniformly diploid or tetraploid, even in the two zones of cytotype co-occurrence in the Slovenian Forealps and the Carpathians. Mixed diploid-tetraploid populations were found mainly in the Western Carpathians, while the occurrence of triploid individuals was restricted to diploid populations. For the first time, the neglected areas of the Balkans, Southeastern Carpathians, Baltic coast and Scandinavia were sufficiently studied, revealing the surprising prevalence of diploids in some areas (the Pannonian basin, the Dinaric Alps), in contrast to another diploid-tetraploid complex of *Arabidopsis lyrata* subsp. *petraea* with only a few diploid populations (Schmickl and Koch 2011). For the first time, the distinct group of diploid populations along the southern Baltic Sea coast was reported (CS I), occupying ecologically distinct sandy coastal habitat. Tetraploids dominated the area of the Alps and to the north of them and west of the Carpathians, in the Hercynian Massif, and they also occupied the Scandinavian region.

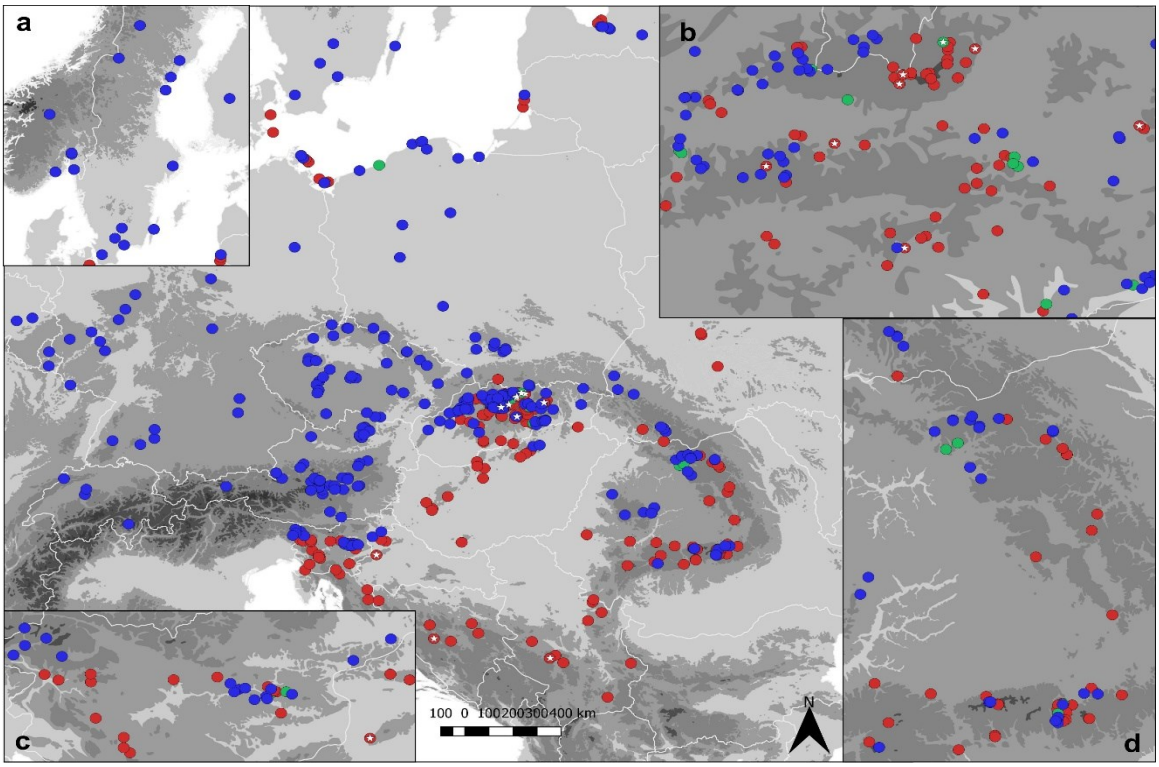


Fig. 7 Distribution and ploidy level of the 496 populations of the *Arabidopsis arenosa* species complex in Europe (*red* – diploid, *blue* – tetraploid, *green* – mixedploid, *asterisk* – triploid individuals observed in the population). **a** subset of Scandinavia; **b** detail of the Western Carpathian contact zone; **c** detail of the contact zone in the Slovenian Forealps; **d** detail of the South-Eastern contact zone.

Contact zones of diploid and tetraploid cytotypes of *A. arenosa* complex found in the Carpathians and the Slovenian Forealps (CS I) represent most probably **secondary contact zones**, where cytotypes regained the contact after previous spatial separation (Fig. 7; Petit et al. 1999). This hypothesis is supported by the separate distribution in other areas and the dominating intrapopulation cytotype uniformity. The uniformity within the contact zones could be preserved by demographic processes in detached populations, where the species preference for open habitats with low competition isolates the population and neutral evolution together with minority cytotype exclusion (Levin 1975) causes the dynamic changes in cytotype frequencies. In favour of this scenario speaks also the absence of long-distance dispersal adaptation and short lifespan with limited vegetative persistence (Kolář, Lučanová, personal observation). On the other hand, similar pattern could be found also in another Brassicaceae perennial *Alyssum montanum* (Španiel et al. 2011, 2012).

As **primary** different ploidy **contact zones** of the *Arabidopsis arenosa* complex, we may consider populations with **triploid** individuals that arise *de novo* from diploids via the fusion of reduced (n) and unreduced (2n) gametes (CS I). The hybridization of different ploidies seemed to be improbable because triploid individuals were found mostly in otherwise diploid populations (CS I; 6 triploids found in contact zones by Morgan et al. 2020) represented only 0,14% of sampled individuals). It is worth noting that the triploids reported in CS I were the first adult triploid *A. arenosa* plants ever published, as triploid individuals recorded before were karyologically investigated seedlings (Kolník, unpublished). This interesting discrepancy was further studied in detail by Morgan et al. (2021), as Měsíček (1970) presumed interploidal hybridization as a scenario for triploid production, but only on ex-situ germinated progeny. By interploidy and control intraploidy experimental crosses followed by extensive phenotyping, the ploidy variation in the progeny of interploidy crosses was found. Most of the triploid individuals arose from tetraploid mothers pollinated by diploid fathers, but all interploidy crosses resulted in lower germination rates and survival of young rosettes than both diploid and tetraploid control treatments and the pollen produced by triploid progeny was less fertile than the control, indicating that the triploid block acts as a strong reproductive isolation mechanism reducing the interploidal crossing (Köhler et al. 2010).

The gene flow from diploids to tetraploids detected by genetic studies (Monnahan et al. 2019) is mediated rather by the unreduced gamete production in diploid individuals (confirmed by Morgan et al. 2021 by a small number of tetraploid progeny) plays a role in forming genetic and cytotype diversity in contact zones and should be studied further.

The **genome size** of individuals from the *Arabidopsis arenosa* complex varies on the homoploid level within both diploids (up to 1.17-fold) and tetraploids (up to 1,21-fold), both inter and intrapopulation, having linear character and no spatial structure (CS I). Mean monoploid DNA content was similar among all three ploidy levels, but not identical (monoploid value of tetraploids was 7,6% higher than that of diploids, on average). Part of the variation could be caused by methodological bias (like seasonal variation, changing instruments or buffers, different tissues; Bainard et al. 2011), but this error was minimized by performing repeated analyses and by the presence of double peaks in analyses with selected individuals. Another scenario is the presence of aneuploidy and dysploidy, characteristic for some polyploid complexes of Brassicaceae (Marhold 2010; Mandáková et al. 2013) and also found in karyological analyses of *Arabidopsis* seedlings (Měsíček 1970, Kolník and Marhold, unpublished), or different intensity in genomic processes within species (e.g. unequal crossing over, non-coding DNA multiplication; Bennetzen et

al. 2005; Leitch and Leitch 2013). One way or another, the variation is most probably a result of neutral processes (Šmarda and Bureš 2010; Oliver et al. 2013) as we did not find any correlation with geographic, altitudinal or environment factors within contact zones (CS II; Knotek et al. 2020; Morgan et al. 2020; CSV).

4.2 Phylogenetic relationships and range-wide patterns of genetic diversity

To detect the range-wide patterns of genetic diversity of both species, we applied several molecular methods available, applicable and most suitable at the time. In the case of *Arabidopsis arenosa* complex, separate cytotype datasets were created based on the data from CS I, to reveal the phylogenetic relationships of diploid individuals in the first place (CS II) as the ancestors of polyploidization, and tetraploids were analysed subsequently (CS V).

For individuals of ***Arabidopsis halleri*** from European and Japanese populations, Bayesian clustering and PCoA of the whole AFLP dataset revealed the distinct position of the **Japanese lineage**, but the divergence level was comparable to that among European lineages. The **northern hemisphere disjunction** spanning the Palearctic region could be explained by two scenarios – recent human-mediated dispersal or ancient vicariance promoted by contacts taking place since the early Tertiary (Sanmartín et al. 2001). Based on the European haplotypes present in Japanese populations we suggest an intermediate scenario of long-distance dispersal from Europe to East Asia probably during the Pleistocene (CS III).

The European populations comprise increased levels of diversity in all markers as well as the majority of previously described subspecies (Hoffmann 2005; Hohmann et al. 2014), which indicates the ancestral potential of this area. Previous studies of *A. halleri* (Van Rossum et al. 2004; Pauwels et al. 2005, 2006, 2008, 2012; Kolník and Marhold 2006; Wasowicz et al. 2016) did not represent the entire European species range, completely neglecting the Balkan area, therefore, the phylogeographic conclusions were misleading. Our homogenous sampling pointed out the major genetic grouping into **three main European lineages**, whose distribution strongly correlated with major geographical boundaries in the Central European mountain systems, such as barriers separating the Alps and Hercynian mountains or the border between Western and Eastern Carpathians (Fig 8; e.g. Pawlowski 1970; Mráz and Ronikier 2016). The **Alpine** and **North-Western** lineages were addressed previously (Pauwels et al. 2005, 2012), but the **South-Eastern** lineage, located in the South and Eastern Carpathians and the Balkan Peninsula, was unknown, probably due to the undersampling of this area. The hierarchical approach was further applied in order to detect finer genetic substructuring of the main groups by carrying out a separate clustering analyses for each of them (CS IV). The stability of those subgroups was tested, revealing consistent stable grouping of populations only within Alpine and North-Western groups. We thus defined **five subgroups**: Western Alpine, Eastern Alpine, Hercynian, Western Carpathian and South-Eastern Carpathian (Fig. 11).

The genetic (and also morphological) differentiation present within the Alpine lineage in the Alps almost completely matches the “Brenner line” located in the area west of the Dolomites, one of the two lines phylogeographically delimiting the Eastern and Western Alps (Schönswetter et al. 2005; Thiel-Egenter et al. 2011; Smyčka et al. 2018).

The similar scenario is found within the North-Western lineage, where the Western Carpathian subgroup, exhibiting elevated levels of genetic diversity and proportion of rare genetic markers known for this area, is genetically and also morphologically well differentiated from the Hercynian subgroup. The main separation of the South-Eastern lineage from the rest of the dataset follows the border between the Western and Eastern Carpathians (Mráz and Ronikier 2016) and interestingly this lineage hosts the largest proportion of unique genetic diversity across all studied markers.

In terms of the separation of the main detected lineages, the applied admixture model suggested only minor admixture between populations of NW Carpathian group and Alpine group in the Austrian Tyrol and between NW Carpathian and SE Carpathian populations in central Slovakia, but together with observed ancestral haplotype sharing among lineages, we assume that old vicariance with incomplete lineage sorting and/or ancient gene flow was more probable than recent hybridization events and the revealed low level of admixture is caused by recent human-mediated spread followed by admixture.

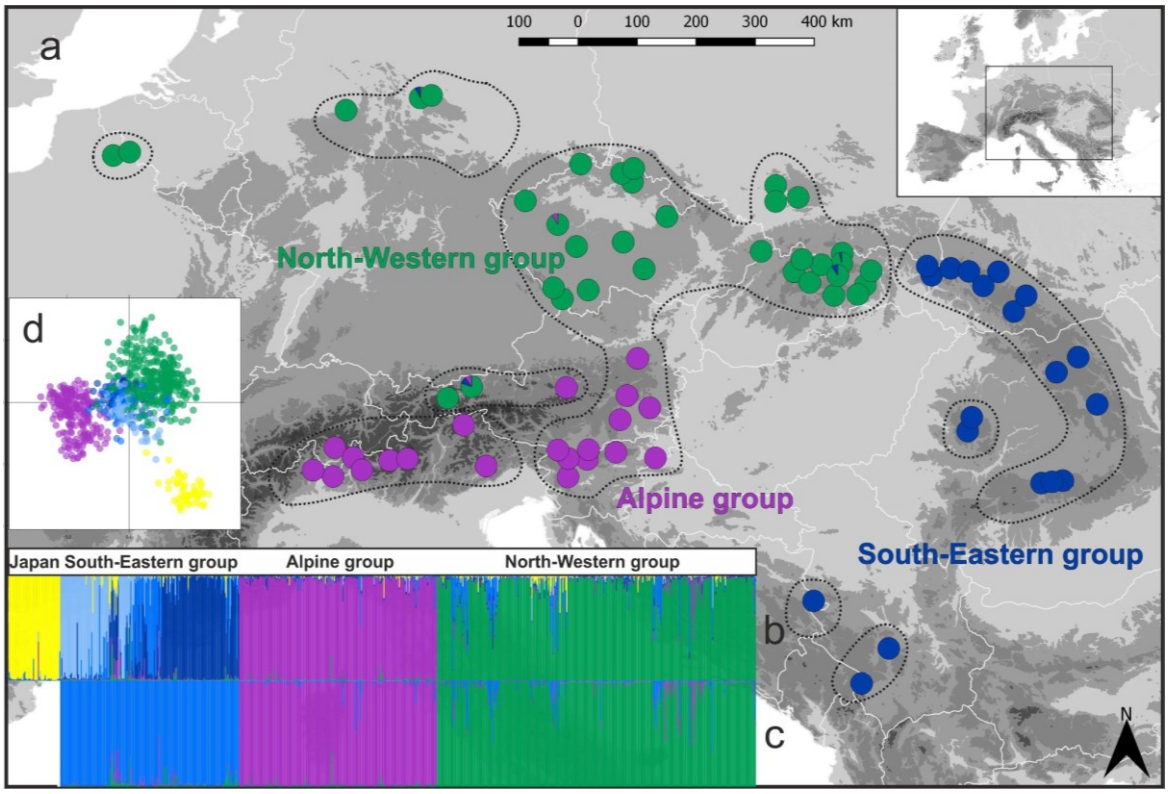


Fig. 8 Geographic distribution of major genetic groups of *Arabidopsis halleri* across its European range. **a** geographic distribution of sampled populations (colour pie charts reflect the proportion of individuals belonging to a particular STRUCTURE group); the *dotted line* denotes the borders of the distribution range of *A. halleri*; **b** cluster assignment of the AFLP phenotypes revealed by STRUCTURE for the complete Eurasian dataset; **c** cluster assignment of individuals revealed by separate STRUCTURE analyses of the European populations; **d** principal coordinate analyses based on Jaccard distances of the Eurasion dataset, colour coded according to their major ancestry inferred by the corresponding STRUCTURE analyses. Modified from CS III.

The first study examining the genetic structure of ***Arabidopsis arenosa*** polyploid species complex was based on the previous thorough cytological screening (CS I) and involved **diploid** individuals from pure as well as mixed populations. The optimal grouping of the populations was determined based on genome-wide SNPs and revealed the separation into **five groups** that corresponded with geography (Fig 9). The **Dinaric** and **Pannonian** groups were the most separated on the first and second PCoA axis,

leaving the **Western Carpathian** and **Southeastern Carpathian** groups in a closer relationship, but still distinguishable by Bayesian clustering.

The spatially isolated **Baltic** lineage showed admixture between W and SE groups and even while analysing only Carpathian and Baltic populations, individuals from the Baltic coast occupied an intermediate position in the PCoA ordination and admixed assignment in STRUCTURE. But this **disjunct outpost** hosts ecologically distinct stands such as chalk cliffs and grey sand dunes and we wanted to reveal also the origin of those postglacial colonizations. Therefore, the origin of those populations was investigated using three alternative scenarios for coalescent-based approximate Bayesian computing (ABC) further supporting the admixed origin of the Baltic group being genetically very close to

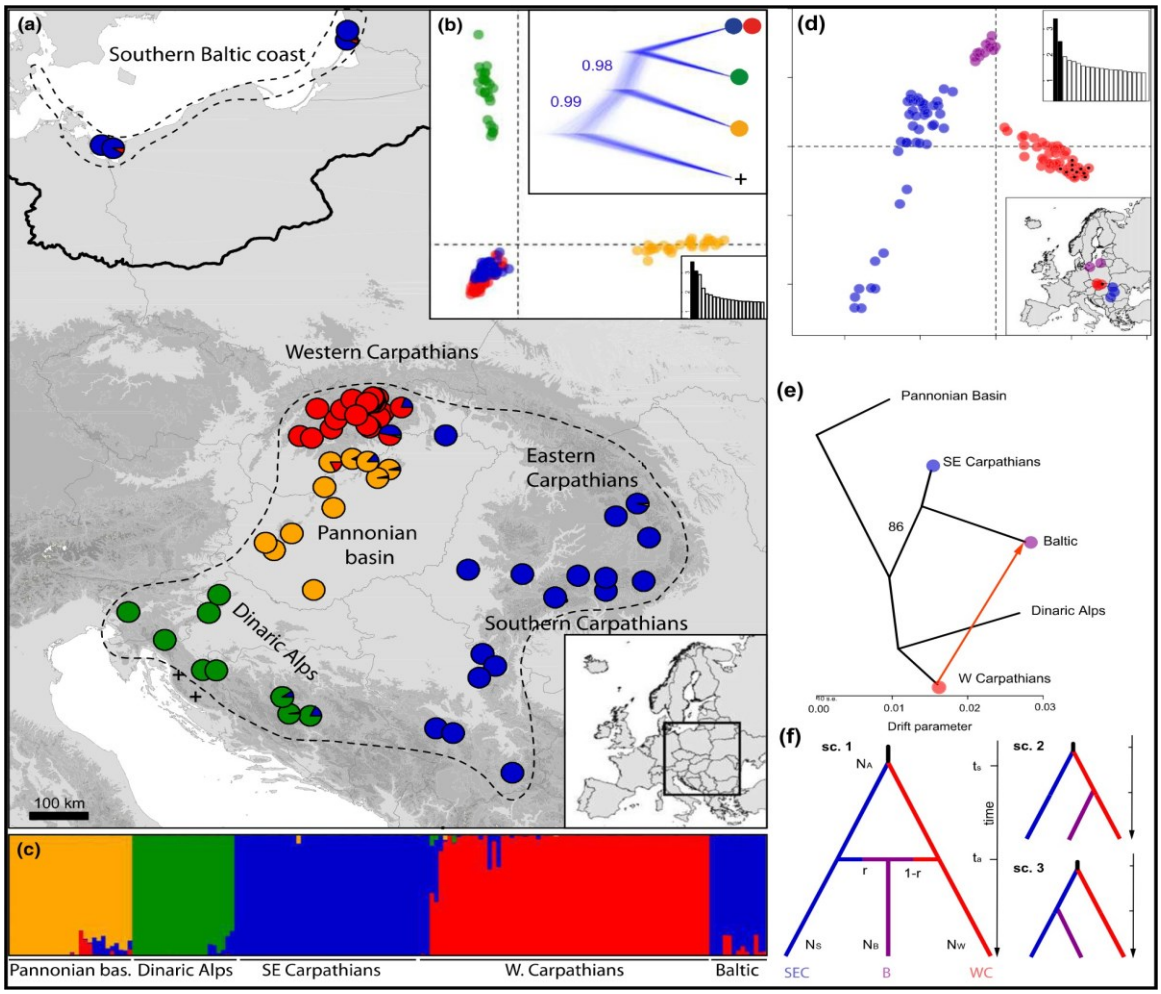


Fig. 9 Rangewide genetic differentiation of diploid individuals from *Arabidopsis arenosa* species complex (**a-c**) and reconstruction of the relationships among the Baltic and Carpathian populations (**d-f**). **a** geographical distribution of sampled populations (pie charts reflecting proportional assignment to a particular STRUCTURE group) and an outgroup, *A. croatica* (*black cross*), *bold line* indicates maximal extent of the continental ice sheet during the last glaciation, dashed line denotes borders of the distribution range of diploid cytotype; **b** principal coordinate analyses of the *A. arenosa* individuals (histogram shows proportional contribution in explaining variance by the first 20 aces) and species tree of the three most distinct groups inferred under multispecies coalescent analyses, rooted with *A. croatica*, posterior probabilities above the branches; **c** cluster assignment of the individuals revealed by STRUCTURE; **d** principal coordinate ordination of Carpathian and Baltic individuals; **e** Treemix maximum-likelihood graph showing relationships among main lineages of diploid *A. arenosa* species complex with one migration edge, individuals from high-altitude populations of W Carpathian group are marked by *black dot*; **f** three competing scenarios differing by the mode of origin of the Baltic populations simulated and tested in ABC framework with varying effective population sizes (N) and migration rate (r), scenario 1 was the most likely. Modified from CS II.

the W and SE Carpathian groups (Fig. 9f). The Baltic sea coastline was affected by the continental ice sheet and changes in global ocean levels (Björck 1995), creating the current landmass ca. 5700 years ago (Janke et al. 1993; Wohlfarth et al. 2008). This allows us to hypothesize that admixture occurred in the past when the W and SE lineages met and hybridized, and only then did the admixed individuals migrate north into the postglacial niches.

The genetic distribution of European **tetraploid** populations from the ***A. arenosa*** species complex have already been subjected to investigations by many studies, therefore, the basic knowledge about the gene pools and their approximate geographical distribution based on fragmented sampling is already known (Arnold et al., 2015; Monnahan et al. 2019). These studies supported a **single origin of tetraploids** in the W Carpathians followed by a subsequent spread across Europe probably accompanied and mediated by interploidy gene flow with local diploids.

In the subsequent study (CS V), we built on the largest homogeneous range-wide collections from both natural and anthropogenic habitats of tetraploids and we presented detailed information about the genetic clustering of 187 populations, including previously undersampled (SE Carpathians, Alps, Baltics, Hercynian massif) and completely unsampled (Scandinavia, western Europe) areas. Our SNP data confirmed the division into **five genetic clusters** (Fig 10), four of them corresponded to geographic regions and the last one, ruderal, occupying man-made habitats, clustered irrespective to geography. However, the lineages have shown a higher degree of admixture than in previous studies. In each lineage, we observed "pure" populations that have been assigned to one single genetic cluster whereas in others intermediate to higher levels of admixture were found (Fig 10e). This can be caused by our broad geographical sampling, which also included contact zones between two or more lineages, especially between the Alpine and Central European ones where we usually find populations showing the highest levels of admixture. The most admixed differentiated lineage is the **Central European**, which can be found from western (Belgium, Germany) to Central (Czech Republic, Austria) Europe. All populations assigned to this lineage show a high proportion of membership to other clusters, mainly to the ones including populations located in very close proximity, in the Eastern Alps. Indeed, many populations assigned to the **Alpine** lineage, located in the biogeographic region of the Eastern Alps (Schmitt 2017), also show high levels of admixture. All this evidence suggests hybridization events such as between tetraploid lineages or with local diploids (Monnahan et al. 2019). Another crucial element in those areas is partially sympatric occurrence of *Arabidopsis lyrata* known to hybridize with *A. arenosa* on tetraploid level (Schmickl and Koch 2011; Lafon-Placette et al. 2017; Monnahan et al. 2019), creating a well known hybrid zone in the eastern Austrian Alps (Schmickl et al. 2011). This interspecies gene flow could also play a role in the level of admixture of the present populations. The **Western Carpathian** lineage occupies well-known areas that have been described as glacial refugia for many species in these mountains and surrounding areas (also in Poland and Hungary; e.g. Thiel-Egenter et al. 2011; Alvarez et al. 2012; Ronikier et al. 2012). Tetraploids of *A. arenosa* originated in the Western Carpathians, which also represents one of the three contact zones between diploids and tetraploids (Monnahan et al. 2019; CS I). Many populations assigned to this lineage are "pure" while others show relatively low or intermediate levels of admixture. The second main contact zone between diploids and tetraploids of *A. arenosa* is located in the **Southeastern Carpathians**. The tetraploid lineage found in this area was previously

underrepresented in previous biogeographical studies. One single population was included by Arnold et al. (2015), while three populations were analysed by Monnahan et al. (2019). Our study includes up to 24 populations assigned to SE Carpathian lineage, thus being the first time that the distribution range of *A. arenosa* in SE Carpathians is fully covered. The genetic differentiation is high, compared to other lineages and admixture proportions are relatively low, which suggests a long period of strong isolation due to the biogeographical barrier separating W and SE Carpathians (Mráz and Ronikier 2016).

The **Ruderal** group represents the black horse in *A. arenosa*'s history and future. Despite the distribution of the lineage is not sharply geographically restricted, it is genetically well differentiated from the others and it is the only tetraploid lineage in *A. arenosa* that colonized higher latitudes of Central Europe and Scandinavia, yet still associated to anthropogenic niches, most likely expanding by migration along railway networks (Arnold et al. 2015). Populations assigned to this lineage show relatively low levels of admixture, but bear signs of introgression from the diploid *A. arenosa* lineage (Baltic diploid) and *A. lyrata* (Monnahan et al. 2019). In contrast, it has experienced huge niche expansion towards warmer and colder environmental conditions but contraction to drier areas with respect to the ancestral tetraploid W Carpathian lineage (CS V).

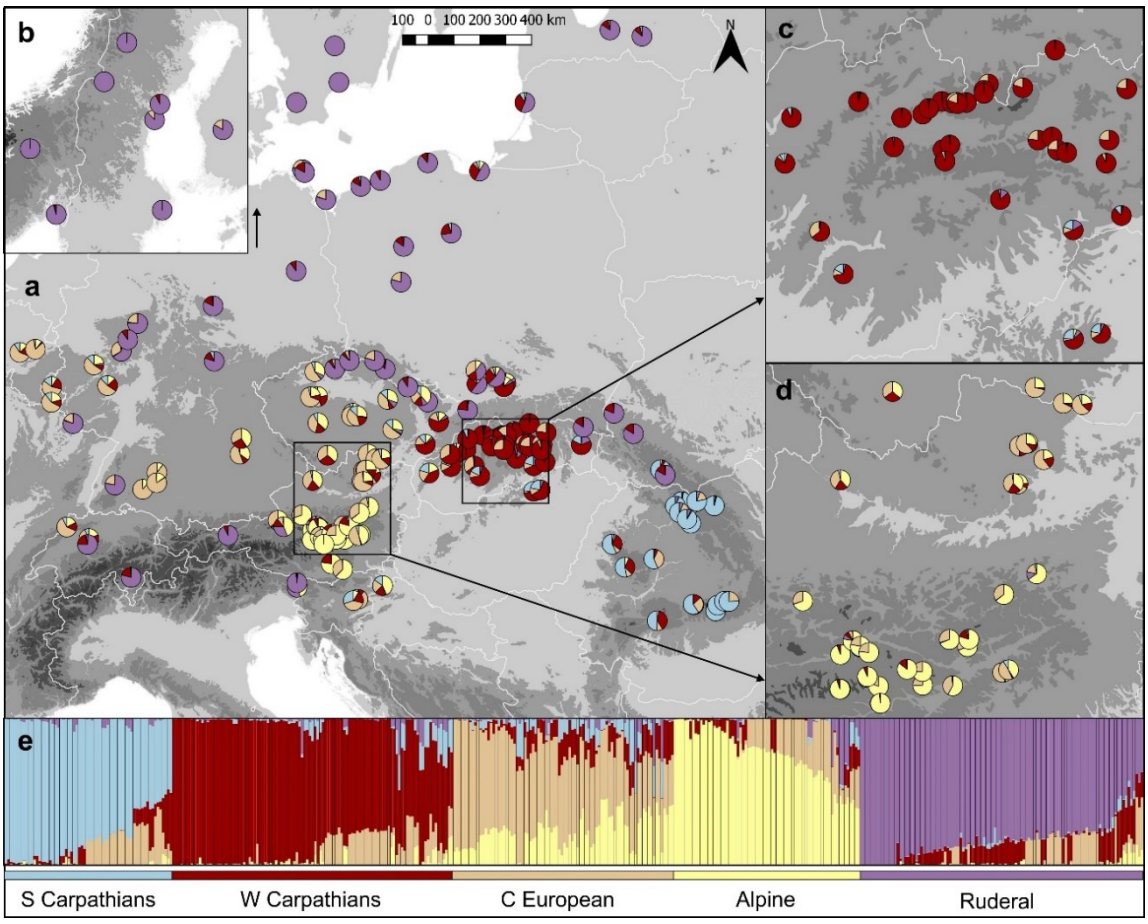


Fig. 10 Geographic distribution of tetraploid populations from the *Arabidopsis arenosa* complex. **a** geographical distribution of sampled populations (pie charts reflecting proportional assignment to a particular STRUCTURE group); **b** subset of Scandinavia; **c** detail of the Western Carpathian contact zone; **d** detail of the Austrian zone; **e** individual cluster assignment into different tetraploid lineages within the *A. arenosa* complex revealed by STRUCTURE (SE Carpathians – *light blue*, W Carpathians – *dark red*, C European – *light orange*, Alpine – *light yellow*, Ruderal – *purple*). Modified from CS V.

Apart from detecting the range-wide genetic structure of studied species, we also studied the plausible glacial persistence in **"cryptic" northern refugia** (Stewart and Lister 2001). The coordinated investigations of the sympatric *A. halleri* and the *A. arenosa* complex provided an opportunity to compare the evolutionary patterns and processes among closely related diploid vs. di-tetraploid complexes evolving within the same environmental context, widening the well documented glacial survival for cold-tolerant species (e.g. Tzedakis et al. 2013; Douda et al. 2014; Mandák et al. 2016). The alternative scenario to the one involving glacial survival in southern European refugia and post-glacial recolonization (Taberlet et al 1998; Hewitt 2004; Tzedakis et al. 2013) suggests glacial survival in the northern refugia such as in the Carpathians (Ronikier 2011).

In our data, we can find several lines of evidence suggesting the scenario of survival in northern refugia, such as populations with high genetic differentiation and proportion of rare fragments (corresponding to the areas of long-term persistence, e.g. Tribsch et al. 2002; Paun et al. 2008) in the areas of south and eastern foothills of Alps, Pannonian basin, the Western tCarpathians and multiple regions of South-Eastern Carpathians (CS II, CS III). The recent spread from the single pool (as from the classical southern glacial refugium) was disproved by relatively high differentiation of populations within each group in both species (CS II, CS III, Arnold et al. 2015). Our data support the existence of the northern glacial refugia known also from other species (Tzedakis et al. 2013) based on the fossil data (Birks and Willis 2008) and genetic structure of temperate plants and animals (Magri et al. 2006; Kotlik et al. 2006; Těšitel et al. 2009; Sommer and Zachos 2009). The most promising area for both *A. halleri* and diploid *A. arenosa* are the Carpathian foothills, as this area hosts their most diverse populations and was also affable habitat being inhabited by forest communities during the last glacial maxima (Juřičková et al. 2014; Jankovská and Pokorný 2017; CS II, CS III).

The establishment and success of polyploids are thought to often be facilitated by ecological niche differentiation from diploids reflecting a WGD-driven shift in important functional traits, which in turn strengthens prezygotic isolation between ploidies and contributes to polyploid speciation. It is crucial not only to know the ploidy level of polyploids (Duchoslav et al. 2020) but also the intraspecific genetic structure. In CS V we were trying to solve the mystery of inconsistent outcomes by studying **the role of niche differentiation** in the autopolyploid evolution of the *Arabidopsis arenosa* complex. Ecological niche modelling was used in a study conducted by Molina-Henao and Hopkins (2019), which concluded niche expansion but not the divergence of tetraploids of *A. arenosa*. These results contrast with other studies in which an absence of ecological niche differentiation was described at both the landscape (CS II) and intra-population scales (Wos et al. 2019), and in three contact zones independently (Morgan et al. 2020). We investigated differences among climatic niches of tetraploid lineages considering the observed inter-lineage admixture proportions covering up the whole geographic distribution area of the tetraploid cytotype for the first time. The study confirms that niche evolution of polyploids is detectable only when polyploid lineages are compared with their corresponding diploid ancestors, not globally, as both niche expansion and conservatism were found.

4.3 Taxonomical concept based on intraspecific variation

Arabidopsis species are one of the top model-species among the Brassicaceae to study ecology, physiology, molecular basis and evolution of phytoremediation or parallel adaptation e.g. Clauss and Koch 2006; Krämer 2010; Bohutínská et al. 2021 a, b, c; Konečná et al. 2021). Due to the inexplicit taxonomy or simply neglected phylogenetic relationships (as several species and subspecies are recognized within the taxa; chapter 1.3.1 and 1.3.2), many published studies could suffer from misleading phylogenetic and evolutionary context. Therefore, we accompanied our thorough phylogeographic studies with the environmental conditions of the population's locations and morphological data of sampled individuals as the revealed delimitation into several spatially and genetically well-defined groups of both species does not fully correlate with current taxonomic concepts.

Both species occupy a wide variety of substrates, climatic niches and habitats, spanning an elevational range over 2500 m, therefore, representing a suitable model for testing hypothesis concerning niche conservatism vs. shifts through evolutionary history (Pauwels et al. 2012; Schmickl et al. 2012; Hohmann et al. 2014). Although the data suggest a correlation between genetic structure and local bioclimatic conditions, suggesting ecological differentiation among major genetic lineages, this could be driven by the distinct geography of the groups.

The genetic divergence is not in fact the source of **morphological variation** as shown in CS III. For *Arabidopsis halleri,* the largest proportion of observed phenotypic plasticity was explained by ecological conditions. The presence of similar morphotypes in high-altitude habitats within populations in the Eastern Alps and the Southern Carpathians, the areas with the most distinct genetic relationships, represents the similar phenotypic effect as in diploid *A. arenosa*. In this case, the range-wide climatic niche differentiation was correlated with the genetic structure, not with the morphometric data. The major genetic lineages occupying distinct climatic niches along the altitude-related bioclimatic gradient and the most prominent niche shifts were found in the populations which expanded into alpine and northern coastal postglacial environments but were only weakly genetically differentiated.

Taking an advantage of revealed genetic structure complemented with morphological data, we presented a **taxonomic re-evaluation of *Arabidopsis halleri***, considering the historical taxonomic treatments (Fig 5; Jones and Akeyrod 1993; Jalas and Suominen 1994; Kolník and Marhold 2006). To provide a detailed phylogeographic structure with potential substructuring within the lineages, that could detect the structure proposed by current taxonomy, each major group was also analysed separately, revealing stable subgroups within the NW group and Alpine group, resulting in **five subgroups**, that were subsequently tested for morphological separation. The discriminant analyses of populations showed the morphological separation of five genetic sublineages, resulting in the proposed new intraspecific classification better reflecting the overall structure of the species, though their recognition based on single characters are not easy yet still sufficient on the subspecies level. The reliable differentiation can be done based on geographic origin (for details see Fig. 11), *Arabidopsis halleri* subsp. *halleri* spanning the area of Hercynian lineage, *A. halleri* subsp. *tatrica* for the populations from the Western Carpathians, *A. halleri* subsp. *ovirensis* in the Eastern Alps, *A. haller* subsp. *dacica* in the Southeastern Carpathians and newly proposed subspecies *A. halleri* subsp. *occidentalis* for the morphologically distinct populations of the Western Alps.

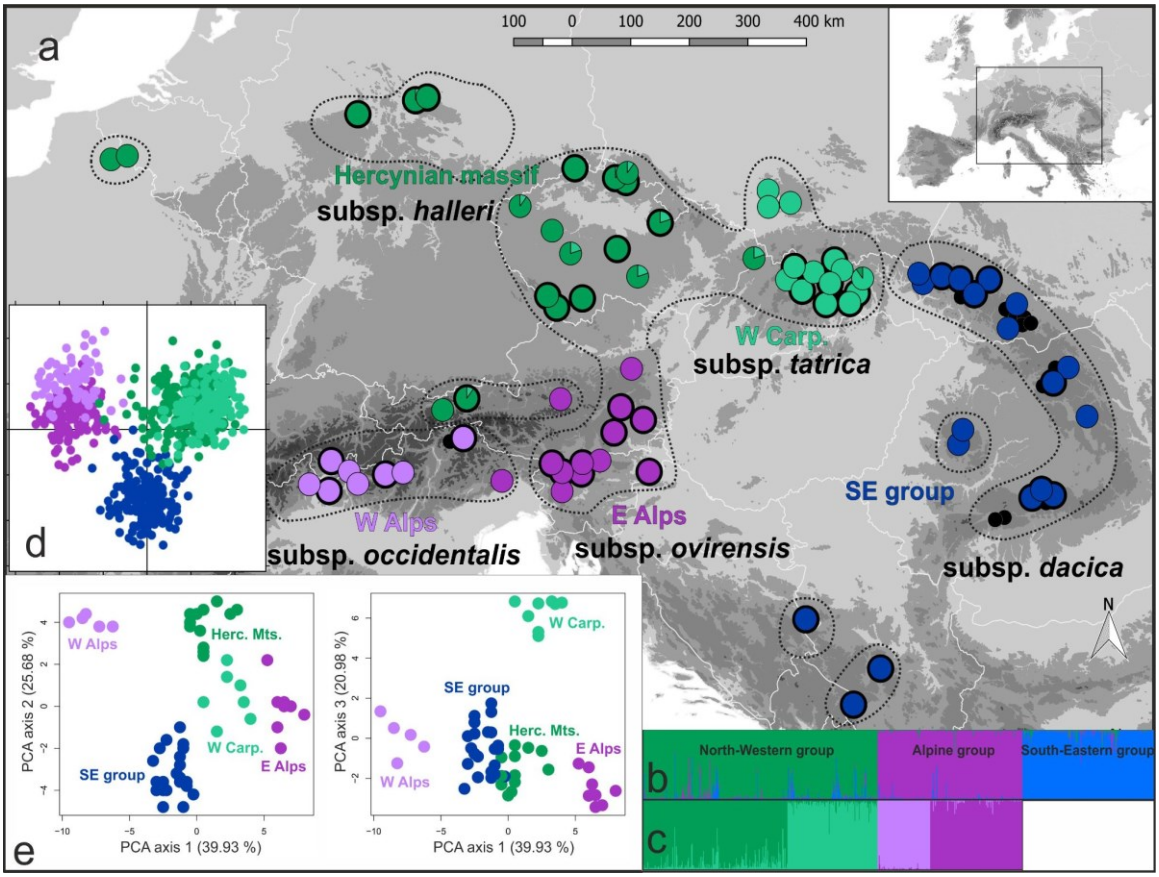


Fig 11 Taxonomic reassessment of *Arabidopsis halleri* based on the range-wide genetic differentiation. **a** colour pie charts reflecting the proportional assignment to particular STRUCTURE subgroups within each group, *thick borders* represent populations used also for morphological measurements, and *black dots* represent populations sampled for morphometry only), the *dotted line* denotes the borders of distribution range of *A. halleri*; **b** cluster assignment of the individual AFLP phenotypes revealed by STRUCTURE for the complete dataset; **c** cluster assignment of individuals revealed by separate STRUCTURE analyses of the North-Western and Alpine groups (analyses of the South-Eastern group did not result in consistent results); **d** principal coordinate analyses based on Jaccard distances between all AFLP phenotypes; the colour coding matches the assignment to the respective lineages as inferred by STRUCTURE; **e** morphological separation of the five European subgroups revealed by means of canonical discriminant analyses based on population means. Modified from CS IV.

The current taxonomy of ***Arabidopsis arenosa* complex** is even more problematic, harbouring the mixture of species and subspecies of different ploidy levels (Fig 4). Previously recognized diploid species *A. carpatica, A. neglecta* and *A. nitida* (Měsíček 1970) were clustered into the W Carpathian lineage, while the genetically most differentiated Dinaric and SE Carpathian diploid lineages were completely neglected in the previous taxonomic treatments. The only taxon fitting one of the described diploid groups is *A. petrogena*, which corresponds to the Pannonian lineage. The observed admixture in several areas suggests, that the barriers between groups are still permeable (unlike in *A. lyrata*; Leppälä et al. 2013), and the situation is even more complicated by the gene flow to the tetraploid lineages (Arnold et al. 2015; Monnahan et al 2019). The proper taxonomic assessment based on the morphological and experimental investigations is needed while taking into account also phylogenetic relationships between diploid and tetraploid groups, yet the main message of our studies is clear – employing *A. arenosa* as a model group should rely on the presented lineages, not the previous inaccurate taxonomical classification.

5 Conclusions and future prospects

During the past several years, wild relatives of *Arabidopsis thaliana* became an important model species for many scientific fields, e.g. heavy metal tolerance, adaptation for whole-genome duplication, parallel evolution. Therefore, the knowledge of their interspecific phylogenetic relationships is crucial for the correct interpretation of those studies. Here, I summarise the main contributions of the five case studies and the remaining questions to be answered.

Ploidy and genome size variation across the distributional ranges

- *Arabidopsis halleri* represents purely diploid taxon.
- Populations of the *A. arenosa* complex hosts three different cytotypes, whereas the minority triploid cytotype was found only within diploid populations.
- The vast majority of populations are cytotype uniform and diploid and tetraploid cytotypes exhibit a largely parapatric distribution. Diploids occupy mainly southeastern areas of Europe and spatially isolated localities along the southern shores of the Baltic Sea. Tetraploids dominate in the northwestern half of the species range.
- Three secondary contact zones were found in the Western Carpathians, northern Dinaric region and Southeastern Carpathians.
- The genome size of *A. arenosa* varies only slightly, probably as a result of neutral processes as we did not find any correlation with geographic, altitudinal or environmental factors within contact zones.

Phylogenetic relationships and range-wide patterns of genetic diversity

- We can find three main European genetic lineages (further differentiated into five subgroups) accompanied with one in East Asia within *Arabidopsis halleri* and five diploid and five tetraploid lineages in the *A. arenosa* complex, their distribution correlated with major geographical boundaries in the Central European biogeographic systems.
- The lineages of the same ploidy are geographically isolated, however in the areas of secondary contact of *A. arenosa*'s cytotypes gene flow from diploids to tetraploids was found. The tetraploids did not overcome their diploid ancestors in this area.
- Autotetraploid cytotypes of *A. arenosa* originated only once in the Western Carpathians and spread through most of Europe. Especially the Ruderal lineage, containing diagnostic alleles from W Carpathian and Baltic diploid lineages, has experienced huge niche expansion towards warmer and colder environmental conditions but contraction to drier areas with respect to the ancestral tetraploid W Carpathian lineage.

New taxonomical concept based on intraspecific variation

- Revealed intraspecific structures do not correlate with the current taxonomic concept of neither *Arabidopsis halleri* nor *A. arenosa*.

- The major factor of morphological variation in *A. halleri* is the environment as the largest proportion of observed phenotypic plasticity was explained by ecological conditions.
- The morphological separation of five genetically differentiated subgroups allowed a new intraspecific classification of *A. halleri*, even though the recognition based on single characters is not easy (yet still sufficient at the subspecies level). The reliable differentiation can be done based on geographic origin.

Looking at the obtained results, the next direction of our research is obvious. We are still lacking the taxonomic concept for the autopolyploid complex of *Arabidopsis arenosa*. The morphometric measurements were processed and are available, but before we will proceed to the taxonomic reassessment of this intricate group, the relationships between diploid and polyploid lineages need to be unravelled. To do so, the joint dataset based on the tetraploid and diploid dd-RADseq SNP data from the whole distributional range will be created and analysed and set into the context with the morphological findings. As crucial we see the problem of circumscription of the separate species or intraspecific taxa of different ploidies within the same area.

Another challenge represents the hybridization between tetraploids of the *Arabidopsis arenosa* complex and *A. lyrata*. Introgressive hybridization is now recognized as a widespread and fundamental evolutionary force (Sung et al. 2018), which can fuel adaptation (Arnold and Kunte 2017; Schmickl and Yant 2021). It might provide fertile ground for adaptive radiation either by enriching genetic variation in an initial hybridization event between two species that may then trigger radiation or by introducing adaptations that allow species of radiating lineages to occupy new niches and further diversify. Hybrid zones of *A. lyrata* and the *A. arenosa* complex are still active and thus provide a unique chance for studying microevolutionary processes related to introgression within this model genus. The on-going introgression has been reported for a hybrid zone in the eastern Austrian Fore-Alps (Schmickl and Koch 2011; Hohmann and Koch 2017), and this hybrid zone gives us the opportunity to study speciation mechanisms that involve gene flow in an immense detail. Taking the advantage of large sampling dataset of both species, including populations from hybrid zones, we will first address the hybrid zone range, modality and genetical structure along the environmental gradient and secondly the mechanisms of hybrid establishment, their competition with parental lineages and the question of putative interploidal gene flow direction.

References

Agapow P, Bininda-Emonds ORP, Crandall KA, et al (2004) The impact of species concept on biodiversity studies. Q Rev Biol 79:161–179. doi: 10.1086/383542

Alvarez N, Manel S, Schmitt T (2012) Contrasting diffusion of Quaternary gene pools across Europe: The case of the arctic-alpine *Gentiana nivalis* L. (Gentianaceae). Flora Morphol Distrib Funct Ecol Plants 207:408–413. doi: 10.1016/j.flora.2012.03.006

Andrews KR, Good JM, Miller MR, et al (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17:81–92. doi: 10.1038/nrg.2015.28

Arabidopsis Genome Iniciative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815. doi: 10.1038/35048692

Arnold B, Kim S, Bomblies K (2015) Single geographic origin of a widespread autotetraploid *Arabidopsis arenosa* lineage followed by interploidy admixture. Mol Biol Evol 32:1382–1395. doi: 10.1093/molbev/msv089

Arnold ML, Kunte K (2017) Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. Trends Ecol Evol 32:601–611. doi: 10.1016/j.tree.2017.05.007

Avendaño JE, Arbeláez-Cortés E, Cadena CD (2017) On the importance of geographic and taxonomic sampling in phylogeography: A reevaluation of diversification and species limits in a Neotropical thrush (Aves, Turdidae). Mol Phylogenet Evol 111:87–97. doi: 10.1016/j.ympev.2017.03.020

Avise JC (2009) Phylogeography: Retrospect and prospect. J Biogeogr 36:3–15. doi: 10.1111/j.1365-2699.2008.02032.x

Avise JC, Arnold J, Ball RM, et al (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst 18:489–522. doi: 10.1146/annurev.es.18.110187.002421

Ayres DR, Smith DL, Zaremba K, et al (2004) Spread of exotic cordgrasses and hybrids (*Spartina* sp.) in the tidal marshes of San Francisco Bay, California, USA. Biol Invasions 6:221–231. doi: 10.1023/B:BINV.0000022140.07404.b7

Baduel P, Arnold B, Weisman CM, et al (2016) Habitat-associated life history and stress-tolerance variation in *Arabidopsis arenosa*. 171:437–451. doi: 10.1104/pp.15.01875

Bailey CD, Doyle JJ (1999) Potential phylogenetic utility of the low-copy nuclear gene pistillata in dicotyledonous plants: comparison to nrDNA ITS and trnL intron in *Sphaerocardamum* and other Brassicaceae. Mol Phylogenet Evol 13:20–30. doi: 10.1006/mpev.1999.0627

Bailey CD, Koch MA, Mayer M, et al (2006) Toward a global phylogeny of the Brassicaceae. Mol Biol Evol 23:2142–2160. doi: 10.1093/molbev/msl087

Bainard JD, Husband BC, Baldwin SJ, et al (2011) The effects of rapid desiccation on estimates of plant genome size. Chromosom Res 19:825–842. doi: 10.1007/s10577-011-9232-5

Baird NA, Etter PD, Atwood TS, et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:1–7. doi: 10.1371/journal.pone.0003376

Bálint M, Domisch S, Engelhardt CHM, et al (2011) Cryptic biodiversity loss linked to global climate change. Nat Clim Chang 1:313–318. doi: 10.1038/nclimate1191

Barker MS, Arrigo N, Baniaga AE, et al (2016) On the relative abundance of autopolyploids and allopolyploids. New Phytol 210:391–398. doi: 10.1111/nph.13698

Bennetzen JL, Ma J, Devos K (2005) Mechanisms of recent genome size variation in flowering plants. Ann Bot 95:127–132. doi: 10.1093/aob/mci008

Bhagwat SA, Willis KJ (2008) Species persistence in northerly glacial refugia of Europe: A matter of chance or biogeographical traits? J Biogeogr 35:464–482. doi: 10.1111/j.1365-2699.2007.01861.x

Birks HJB, Willis KJ (2008) Alpines, trees, and refugia in Europe. Plant Ecol Divers 1:147–160. doi: 10.1080/17550870802349146

Björck S (1995) A review of the history of the Baltic Sea, 13.0-8.0 ka BP. Quat Int 27:19–

40. doi: 10.1016/1040-6182(94)00057-C
Bohutínská M, Alston M, Monnahan P, et al (2021a) Novelty and convergence in adaptation to whole genome duplication. Mol Biol Evol 38:3910–3924. doi: https://doi.org/10.1093/molbev/msab096
Bohutínská M, Handrick V, Yant L, et al (2021b) De novo mutation and rapid protein (co-)evolution during meiotic adaptation in Arabidopsis arenosa. Mol Biol Evol 38:1980–1994. doi: 10.1093/molbev/msab001
Bohutínská M, Vlček J, Yair S, et al (2021c) Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. Proc Natl Acad Sci U S A 118:. doi: 10.1073/pnas.2022713118
Bomblies K, Weigel D (2010) *Arabidopsis* and relatives as models for the study of genetic and genomic incompatibilities. Philos Trans R Soc B Biol Sci 365:1815–1823. doi: 10.1098/rstb.2009.0304
Brodin A, Haas F (2009) Hybrid zone maintenance by non-adaptive mate choice. Evol Ecol 23:17–29. doi: 10.1007/s10682-007-9173-9
Chapman MA, Burke JM (2007) Genetic divergence and hybrid speciation. Evolution (NY) 61:1773–1780. doi: 10.1111/j.1558-5646.2007.00134.x
Clauss MJ, Cobban H, Mitchell-Olds T (2002) Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaeae). Mol Ecol 11:591–601
Clauss MJ, Koch MA (2006) Poorly known relatives of *Arabidopsis thaliana*. Trends Plant Sci 11:449–59. doi: 10.1016/j.tplants.2006.07.005
Dobeš C, Mitchell-Olds T, Koch MA (2004) Intraspecific diversification in North American *Boechera stricta* (= *Arabis drummondii*), *Boechera* × *divaricarpa* , and *Boechera holboellii* (Brassicaceae) inferred from nuclear and chloroplast molecular markers—an integrat. Am J Bot 91:2087–2101. doi: 10.3732/ajb.91.12.2087
Douda J, Doudová J, Drašnarová A, et al (2014) Migration patterns of subgenus *Alnus* in Europe since the last glacial maximum: a systematic review. PLoS One 9:e88709. doi: 10.1371/journal.pone.0088709
Dubois A (2003) The relationships between taxonomy and conservation biology in the century of extinctions. C R Biol 326:9–21. doi: 10.1016/S1631-0691(03)00022-2
Duchoslav M, Jandová M, Kobrlová L, et al (2020) Intricate distribution patterns of six cytotypes of Allium oleraceum at a continental scale: niche expansion and innovation followed by niche contraction with increasing ploidy level. Front Plant Sci 11:. doi: 10.3389/fpls.2020.591137
Durka W, Babik W, Ducroz J-F, et al (2005) Mitochondrial phylogeography of the Eurasian beaver *Castor fiber* L. Mol Ecol 14:3843–3856. doi: 10.1111/j.1365-294X.2005.02704.x
Genovart M, Juste J, Contreras-Díaz H, Oro D (2012) Genetic and phenotypic differentiation between the critically endangered Balearic shearwater and neighboring colonies of its sibling species. J Hered 103:330–341. doi: 10.1093/jhered/ess010
Godfray HCJ (2007) Linnaeus in the information age. Nature 446:259–260. doi: 10.1038/446259a
Greuter W, Burdet. VM, Chaloner WG q, et al (1988) International code of botanical nomenclature. Adopted by the fourteenth International Botanical Congress. Berlin. July–August 1987 (Regnum vegetabile vol. 118)
Haig SM, Beever EA, Chambers SM, et al (2006) Taxonomic considerations in listing subspecies under the U.S. endangered species act. Conserv Biol 20:1584–1594. doi: 10.1111/j.1523-1739.2006.00530.x
Hegde SG, Nason JD, Clegg JM, Ellstrand NC (2006) The evolution of California's wild radish has resulted in the extinction of its progenitors. Evolution (N Y) 60:1187. doi: 10.1554/05-634.1
Hewitt GM (1999) Post-glacial re-colonization of European biota. Biol J Linn Soc 68:87–112. doi: 10.1006/bijl.1999.0332

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. Nature 405:907–913
Hewitt GM (2001) Speciation, hybrid zones and phylogeography - or seeing genes in space and time. Mol Ecol 10:537–49
Hewitt GM (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
Hoffmann MH (2005) Evolution of the realized climatic niche in the genus: *Arabidopsis* (Brassicaceae). Evolution (NY) 59:1425-1436. doi: 10.1111/j.0014-3820.2005.tb01793.x
Hohmann N, Koch MA (2017) An *Arabidopsis* introgression zone studied at high spatio-temporal resolution: interglacial and multiple genetic contact exemplified using whole nuclear and plastid genomes. BMC Genomics 18:810. doi: 10.1186/s12864-017-4220-6
Hohmann N, Schmickl R, Chiang TY, et al (2014) Taming the wild: resolving the gene pools of non-model *Arabidopsis* lineages. BMC Evol Biol 14:1–21. doi: 10.1186/s12862-014-0224-x
Isaac N, Mallet J, Mace G (2004) Taxonomic inflation: its influence on macroecology and conservation. Trends Ecol Evol 19:464–469. doi: 10.1016/j.tree.2004.06.004
Jalas J, Suominen J (1994) Atlas Florae Europaeae - distribution of vascular plants in Europe
Janke W, Kliewe H, Sterr H (1993) Holozäne Genese der Boddenküste Mecklenburg-Vorpommerns und deren künftige klimabedingte Entwicklung. In: Klimaänderung und Küste. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 137–152
Jankovská V, Pokorný P (2017) Forest vegetation of the last full-glacial period in the Western Carpathians (Slovakia and Czech Republic). Preslia 80:307–324
Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends Ecol Evol 11:424–429. doi: 10.1016/0169-5347(96)10049-5
Jones BMG, Akeyrod RJ (1993) *Cardamine*. In: Flora Europaea. Cambridge, pp 346–252
Jørgensen MH, Carlsen T, Skrede I, Elven R (2008) Microsatellites resolve the taxonomy of the polyploid *Cardamine digitata* aggregate (Brassicaceae). Taxon 57:882–892. doi: 10.1002/tax.573017
Juřičková L, Horáčková J, Ložek V (2014) Direct evidence of central European forest refugia during the last glacial period based on mollusc fossils. Quat Res 82:222–228. doi: 10.1016/j.yqres.2014.01.015
Kiefer M, Schmickl R, German DA, et al (2014) BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. Plant Cell Physiol 55:e3–e3. doi: 10.1093/pcp/pct158
Kirchheimer B, Schinkel CCFF, Dellinger AS, et al (2016) A matter of scale: apparent niche differentiation of diploid and tetraploid plants may depend on extent and grain of analysis. J Biogeogr 43:716–726. doi: 10.1111/jbi.12663
Knotek A, Kolář F (2018) Different low-competition island habitats in Central Europe harbour similar levels of genetic diversity in relict populations of *Galium pusillum* agg . (Rubiaceae). Biol J Linn Soc 125:491–507. doi: 10.1093/biolinnean/bly126
Knotek A, Konečná V, Wos G, et al (2020) Parallel alpine differentiation in *Arabidopsis arenosa*. Front Plant Sci 11:1–12. doi: 10.3389/fpls.2020.561526
Knowlton N, Jackson JBC (1994) New taxonomy and niche partitioning on coral reefs: jack of all trades or master of some? Trends Ecol Evol 9:7–9. doi: 10.1016/0169-5347(94)90224-0
Koch M a, Wernisch M, Schmickl R (2008) *Arabidopsis thaliana*'s wild relatives: an updated overview on systematics, taxonomy and evolution. Taxon 57:933–943
Koch MA (2018) The plant model system *Arabidopsis* set into an evolutionary, systematic and spatio-temporal context. J Exp Bot 70:55–67. doi: 10.1093/jxb/ery340
Koch MA, Matschinger M (2007) Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. Proc Natl Acad Sci 104:6272–6277
Köhler C, Mittelsten Scheid O, Erilova A (2010) The impact of the triploid block on the origin and evolution of polyploid plants. Trends Genet 26:142–148. doi:

10.1016/j.tig.2009.12.006
Kolník M, Marhold K (2006) Distribution, chromosome numbers and nomenclature conspect of *Arabidopsis halleri* (Brassicaceae) in the Carpathians. Biologia (Bratisl) 61:41–50
Konečná V, Bray S, Vlček J, et al (2021) Parallel adaptation in autopolyploid *Arabidopsis arenosa* is dominated by repeated recruitment of shared alleles. Nat Commun 12:. doi: 10.1038/s41467-021-25256-5
Kotlik P, Deffontaine V, Mascheretti S, et al (2006) A northern glacial refugium for bank voles (*Clethrionomys glareolus*). Proc Natl Acad Sci 103:14860–14864. doi: 10.1073/pnas.0603237103
Krämer U (2010) Metal hyperaccumulation in plants. Annu Rev Plant Biol 61:517–534. doi: 10.1146/annurev-arplant-042809-112156
Kron P, Suda J, Husband BC (2007) Applications of flow cytometry to evolutionary and population biology. Annu Rev Ecol Evol Syst 38:847–876. doi: 10.1146/annurev.ecolsys.38.091206.095504
Kropf M, Kadereit JW, Comes HP (2002) Late Quaternary distributional stasis in the submediterranean mountain plant *Anthyllis montana* L. (Fabaceae) inferred from ITS sequences and amplified fragment length polymorphism markers. Mol Ecol 11:447–463. doi: 10.1046/j.1365-294X.2002.01446.x
Kuittinen H, de Haan AA, Vogl C, et al (2004) Comparing the linkage Mmps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. Genetics 168:1575–1584. doi: 10.1534/genetics.103.022343
Lafon-Placette C, Johannessen IM, Hornslien KS, et al (2017) Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. Proc Natl Acad Sci 114:E1027–E1035. doi: 10.1073/pnas.1615123114
Lauga B, Malaval S, Largier G, Regnault-Rogert C (2009) Two lineages of *Trifolium alpinum* (Fabaceae) in the Pyrenees: evidence from random amplified polymorphic DNA (RAPD) markers. Acta Bot Gall 156:317–330. doi: 10.1080/12538078.2009.10516162
Leitch IJ, Leitch AR (2013) Genome size diversity and evolution in land plants. In: Plant Genome Diversity Volume 2. Springer Vienna, Vienna, pp 307–322
Leppälä J, Bokma F, Savolainen O (2013) Investigating incipient speciation in *Arabidopsis lyrata* from patterns of transmission ratio distortion. Genetics 194:697–708. doi: 10.1534/genetics.113.152561
Levin DA (1975) Minority cytotype exclusion in local plant populations. Taxon 24:35–43
Lysak MA, Berr A, Pecinka A, et al (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. Proc Natl Acad Sci 103:5224–5229. doi: 10.1073/pnas.0510791103
Macková L, Vít P, Urfus T (2018) Crop-to-wild hybridization in cherries — Empirical evidence from *Prunus fruticosa*. Evol Appl 11:1748–1759. doi: 10.1111/eva.12677
Magri D, Vendramin GG, Comps B, 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
Mallet J (2005) Hybridization as an invasion of the genome. Trends Ecol Evol 20:229–237. doi: 10.1016/j.tree.2005.02.010
Mandák B, Havrdová A, Krak K, et al (2016) Recent similarity in distribution ranges does not mean a similar postglacial history: a phylogeographical study of the boreal tree species *Alnus incana* based on microsatellite and chloroplast DNA variation. New Phytol 210:1395–1407. doi: 10.1111/nph.13848
Mandáková T, Kovařík A, Zozomová-Lihová J, et al (2013) The more the merrier: recent hybridization and polyploidy in *Cardamine*. Plant Cell 25:3280–3295. doi: 10.1105/tpc.113.114405
Marhold K (2010) IAPT/IOPB chromosome data 10. Taxon 59:1934–1938
Martien KK, Leslie MS, Taylor BL, et al (2017) Analytical approaches to subspecies

delimitation with genetic data. Mar Mammal Sci 33:27–55. doi: 10.1111/mms.12409
Meirmans PG, Goudet J, Gaggiotti OE (2011) Ecology and life history affect different aspects of the population structure of 27 high-alpine plants. Mol Ecol 20:3144–3155. doi: 10.1111/j.1365-294X.2011.05164.x
Měsíček J (1970) Chromosome counts in *Cardaminopsis arenosa* Agg. (Cruciferae). Preslia 42:225–248
Měsíček J, Goliášová K (2002) *Cardaminopsis* (C. A. Mey.) Hayek. In: Flora Slovenska. pp 388–415
Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends Plant Sci 12:106–117. doi: 10.1016/j.tplants.2007.02.001
Molina-Henao YF, Hopkins R (2019) Autopolyploid lineage shows climatic niche expansion but not divergence in *Arabidopsis arenosa*. Am J Bot 106:61–70. doi: 10.1002/ajb2.1212
Monnahan P, Kolář F, Baduel P, et al (2019) Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. Nat Ecol Evol 3:457–468. doi: 10.1101/411041
Morgan EJ, Čertner M, Lučanová M, et al (2020) Niche similarity in diploid-autotetraploid contact zones of *Arabidopsis arenosa* across spatial scales. Am J Bot 107:1–14. doi: 10.1002/ajb2.1534
Morgan EJ, Čertner M, Lučanová M, et al (2021) Disentangling the components of triploid block and its fitness consequences in natural diploid–tetraploid contact zones of *Arabidopsis arenosa*. New Phytol 232:1449–1462. doi: 10.1111/nph.17357
Novikova PY, Hohmann N, Nizhynska V, et al (2016) Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. Nat Genet 48:1077–1082. doi: 10.1038/ng.3617
O'Kane SLJ, Al-Shehbaz IA (1997) A Synopsis of *Arabidopsis* (Brassicaceae). Novon 7:323–327
Oliver KR, McComb JA, Greene WK (2013) Transposable rlements: powerful contributors to angiosperm evolution and diversity. Genome Biol Evol 5:1886–1901. doi: 10.1093/gbe/evt141
Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. Front Zool 7:16. doi: 10.1186/1742-9994-7-16
Patten MA, Remsen J V. (2017) Complementary roles of phenotype and genotype in subspecies delimitation. J Hered 108:462–464. doi: 10.1093/jhered/esx013
Paun O, Schönswetter P, Winkler M, Tribsch A (2008) Historical divergence vs. contemporary gene flow: Evolutionary history of the calcicole *Ranunculus alpestris* group (Ranunculaceae) in the European Alps and the Carpathians. Mol Ecol 17:4263–4275. doi: 10.1111/j.1365-294X.2008.03908.x
Pauwels M, Saumitou-Laprade P, Holl AC, et al (2005) Multiple origin of metallicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in central Europe: the cpDNA testimony. Mol Ecol 14:4403–4414. doi: 10.1111/j.1365-294X.2005.02739.x
Pauwels M, Frérot H, Bonnin I, Saumitou-Laprade P (2006) A broad-scale analysis of population differentiation for Zn tolerance in an emerging model species for tolerance study: *Arabidopsis halleri* (Brassicaceae). J Evol Biol 19:1838–1850. doi: 10.1111/j.1420-9101.2006.01178.x
Pauwels M, Willems G, Roosens N, et al (2008) Merging methods in molecular and ecological genetics to study the adaptation of plants to anthropogenic metal-polluted sites: implications for phytoremediation. Mol Ecol 17:108–119. doi: 10.1111/j.1365-294X.2007.03486.x
Pauwels M, Vekemans X, Godé C, et al (2012) Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). New Phytol 193:916–928. doi: 10.1111/j.1469-8137.2011.04003.x

Pawlowski B (1970) Remarques sur l'endémisme dans la flore des Alpes et des Carpates. Vegetatio 21:181–243
Perez-Sepulveda BM, Heavens D, Pulford C V, et al (2020) An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. bioRxiv 2020.07.22.200840. doi: 10.1101/2020.07.22.200840
Peterson BK, Weber JN, Kay EH, et al (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
Petit C, Bretagnolle F, Felber F (1999) Evolutionary consequences of diploid-polyploid hybrid zones in wild species. Trends Ecol Evol 14:306–311. doi: 10.1016/S0169-5347(99)01608-0
Phillimore AB, Owens IP. (2006) Are subspecies useful in evolutionary and conservation biology? Proc R Soc B Biol Sci 273:1049–1053. doi: 10.1098/rspb.2005.3425
Pinto BJ, Titus-McQuillan J, Daza JD, Gamble T (2019) Persistence of a geographically-stable hybrid zone in Puerto Rican dwarf geckos. J Hered 110:523–534. doi: 10.1093/jhered/esz015
Preite V, Sailer C, Syllwasschy L, et al (2019) Convergent evolution in *Arabidopsis halleri* and *Arabidopsis arenosa* on calamine metalliferous soils. Philos Trans R Soc B Biol Sci 374:20180243. doi: 10.1098/rstb.2018.0243
Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. Trends Ecol Evol 23:564–571. doi: 10.1016/j.tree.2008.06.010
Qiu YX, Fu CX, Comes HP (2011) Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. Mol Phylogenet Evol 59:225–244. doi: 10.1016/j.ympev.2011.01.012
Ramírez-Rodríguez R, Amich F (2019) Conservation assessment at a regional level: the study case of *Delphinium fissum* subsp. *sordidum* (Ranunculaceae), an endemic Iberian subspecies with disjunct distribution. Mediterr Bot 40:165–175. doi: 10.5209/mbot.60137
Rieseberg LH, Raymond O, Rosenthal DM, et al (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. Science (80-) 301:1211–1216. doi: 10.1126/science.1086949
Ronikier M, Mráz P, Ronikier M (2016) Biogeography of the Carpathians: evolutionary and spatial facets of biodiversity. Biol J Linn Soc 119:528–559
Ronikier M, Schneeweiss GM, Schönswetter P (2012) The extreme disjunction between Beringia and Europe in *Ranunculus glacialis* s. l. (Ranunculaceae) does not coincide with the deepest genetic split - A story of the importance of temperate mountain ranges in arctic-alpine phylogeography. Mol Ecol 21:5561–5578. doi: 10.1111/mec.12030
Ronikier M, 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. doi: 10.12705/631.4
Ronikier MM (2011) Biogeography of high-mountain plants in the Carpathians: An emerging phylogeographical perspective. Taxon 60:373–389
Roosens NHCJ, Willems G, Godé C, et al (2008) The use of comparative genome analysis and syntenic relationships allows extrapolating the position of Zn tolerance QTL regions from *Arabidopsis halleri* into *Arabidopsis thaliana*. Plant Soil 306:105–116. doi: 10.1007/s11104-007-9383-5
Sanmartín I, Enghoff H, Ronquist F (2001) Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biol J Linn Soc 73:345–390. doi: 10.1006/bijl.2001.0542
Schmickl R, Ehrich D, Brysting AK, et al (2011) Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). BMC Evol Biol 11:346. doi: 10.1186/1471-2148-11-346

Schmickl R, Koch MA (2011) *Arabidopsis* hybrid speciation processes. Proc Natl Acad Sci USA 108:14192–14197. doi: 10.5061/dryad.j5g76.1
Schmickl R, Paule J, Klein J, et al (2012) The evolutionary history of the *Arabidopsis arenosa* complex: Diverse tetraploids mask the Western Carpathian center of species and genetic diversity. PLoS One 7:e42691. doi: 10.1371/journal.pone.0042691
Schmickl R, Yant L (2021) Adaptive introgression: how polyploidy reshapes gene flow landscapes. New Phytol 230:457–461. doi: 10.1111/nph.17204
Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. Front Zool 4:1–13. doi: 10.1186/1742-9994-4-11
Schmitt T (2009) Biogeographical and evolutionary importance of the European high mountain systems. Front Zool 6:1–10. doi: 10.1186/1742-9994-6-9
Schmitt T (2017) Molecular biogeography of the high mountain systems of Europe: An overview. In: High Mountain Conservation in a Changing World. pp 63–74
Schmitt T, Besold J (2010) Upslope movements and large scale expansions: The taxonomy and biogeography of the *Coenonympha arcania-C. d arwiniana-C. gardetta* butterfly species complex. Zool J Linn Soc 159:890–904. doi: 10.1111/j.1096-3642.2009.00585.x
Schmitt T, Hewitt GM, Muller P (2006) Disjunct distributions during glacial and interglacial periods in mountain butterflies: *Erebia epiphron* as an example. J Evol Biol 19:108–113. doi: 10.1111/j.1420-9101.2005.00980.x
Schmitt T, Varga Z (2012) Extra-Mediterranean refugia: The rule and not the exception? Front Zool 9:1–12. doi: 10.1186/1742-9994-9-22
Schönswetter P, Stehlik I, Holderegger R, Tribsch A (2005) Molecular evidence for glacial refugia of mountain plants in the European Alps. Mol Ecol 14:3547–3555. doi: 10.1111/j.1365-294X.2005.02683.x
Shimizu-Inatsugi R, Lihová J, Iwanaga H, et al (2009) The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. Mol Ecol 18:4024–4048. doi: 10.1111/j.1365-294X.2009.04329.x
Šmarda P, Bureš P (2010) Understanding intraspecific variation in genome size in plants. Preslia 82:41–61
Smyčka J, Roquet C, Renaud J, et al (2018) Disentangling drivers of plant endemism and diversification in the European Alps - a phylogenetic and spatially explicit approach. Perspect Plant Ecol Evol Syst 30:31–40. doi: 10.1016/j.ppees.2017.08.003
Soltis DE, Soltis PS, Schemske DW, et al (2007) Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? Taxon 56:13–30. doi: 10.2307/25065732
Soltis DE, Visger CJ, Soltis PS (2014) The polyploidy revolution then…and now: Stebbins revisited. Am J Bot 101:1057–1078. doi: 10.3732/ajb.1400178
Sommer RS, Zachos FE (2009) Fossil evidence and phylogeography of temperate species: 'glacial refugia' and post-glacial recolonization. J Biogeogr 36:2013–2020. doi: 10.1111/j.1365-2699.2009.02187.x
Španiel S, Marhold K, Passalacqua NG, 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, Thiv M, Zozomová-Lihová J (2012) A new circumscription of *Alyssum montanum* ssp. *montanum* and *A. montanum* ssp. *gmelinii* (Brassicaceae) in Central Europe: molecular and morphological evidence. Bot J Linn Soc 169:378–402. doi: 10.1111/j.1095-8339.2012.01225.x
Španiel S, Marhold K, Zozomová-Lihová J (2017) The polyploid *Alyssum montanum* - *A. repens* complex in the Balkans: a hotspot of species and genetic diversity. Plant Syst Evol. doi: 10.1007/s00606-017-1470-3
Stein RJ, Höreth S, de Melo JRF, et al (2017) Relationships between soil and leaf mineral composition are element-specific, environment-dependent and geographically

structured in the emerging model *Arabidopsis halleri*. New Phytol 213:1274–1286. doi: 10.1111/nph.14219
Štěpánek J (1983) Proposal to conserve 2999 *Arabidopsis* Heynh. (Cruciferae) with a conserved type. Taxon 32:649. doi: 10.2307/1221753
Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. Trends Ecol Evol 16:608–613. doi: 10.1111/mec.13585
Stewart JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: Individualistic responses of species in space and time. Proc R Soc B Biol Sci 277:661–671. doi: 10.1098/rspb.2009.1272
Stolpe C, Krämer U, Müller C (2017) Heavy metal (hyper)accumulation in leaves of *Arabidopsis halleri* is accompanied by a reduced performance of herbivores and shifts in leaf glucosinolate and element concentrations. Environ Exp Bot 133:78–86. doi: 10.1016/j.envexpbot.2016.10.003
Strand AE, Leebens- Mack J, Milligan BG (1997) Nuclear DNA- based markers for plant evolutionary biology. Mol Ecol 6:113–118. doi: 10.1046/j.1365-294X.1997.00153.x
Sung C, Bell KL, Nice CC, Martin NH (2018) Integrating Bayesian genomic cline analyses and association mapping of morphological and ecological traits to dissect reproductive isolation and introgression in a Louisiana Iris hybrid zone. Mol Ecol 27:959–978. doi: 10.1111/mec.14481
Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol 7:453–464
Taberlet P, Zimmermann NE, Englisch T, 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
Tate J, Soltis DE, Soltis PS (2005) Polyploidy in plants. In: Gregory T (ed) The Evolution of the Genome. pp 371–426
Taylor BL, Archer FI, Martien KK, et al (2017a) Guidelines and quantitative standards to improve consistency in cetacean subspecies and species delimitation relying on molecular genetic data. Mar Mammal Sci 33:132–155. doi: 10.1111/mms.12411
Taylor BL, Perrin WF, Reeves RR, et al (2017b) Why we should develop guidelines and quantitative standards for using genetic data to delimit subspecies for data-poor organisms like cetaceans. Mar Mammal Sci 33:12–26. doi: 10.1111/mms.12413
Taylor JW, Turner E, Townsend JP, et al (2006) Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. Philos Trans R Soc B Biol Sci 361:1947–1963. doi: 10.1098/rstb.2006.1923
Těšitel J, Malinová T, Štech M, Herbstová M (2009) Variation in the *Melampyrum sylvaticum* group in the Carpathian and Hercynian region: two lineages with different evolutionary histories. Preslia 81:1–22
Thiel-Egenter C, Alvarez N, Holderegger R, et al (2011) Break zones in the distributions of alleles and species in alpine plants. J Biogeogr 38:772–782. doi: 10.1111/j.1365-2699.2010.02441.x
Thiel-Egenter C, Gugerli F, Alvarez N, et al (2009) Effects of species traits on the genetic diversity of high-mountain plants: A multi-species study across the Alps and the Carpathians. Glob Ecol Biogeogr 18:78–87. doi: 10.1111/j.1466-8238.2008.00421.x
Tobias JA, Seddon N, Spottiswoode CN, et al (2010) Quantitative criteria for species delimitation. Ibis (Lond 1859) 152:724–746. doi: 10.1111/j.1474-919X.2010.01051.x
Tribsch A, Schönswetter P, Stuessy TF (2002) *Saponaria pumila* (Caryophyllaceae) and the ice age in the European Alps. Am J Bot 89:2024–2033. doi: 10.3732/ajb.89.12.2024
Tzedakis PC, Emerson BC, Hewitt GM (2013) Cryptic or mystic? Glacial tree refugia in northern Europe. Trends Ecol Evol 28:696–704. doi: 10.1016/j.tree.2013.09.001
Valdés AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749. doi: 10.1093/genetics/133.3.737
Van Rossum F, Bonnin I, Fénart S, et al (2004) Spatial genetic structure within

a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy-metal-tolerant species. Mol Ecol 13:2959–2967. doi: 10.1111/j.1365-294X.2004.02314.x
Verbruggen N, Hermans C, Schat H (2009) Molecular mechanisms of heavy metal hyperaccumulation in plants. Phytoremediation Environ Pollut 181:759–776. doi: 10.4324/9781315161549
Vos P, Hogers R, Bleeker M, et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–14. doi: 10.1093/nar/23.21.4407
Wakeley J (2009) Coalescent theory: An introduction. Syst Biol 58:162–165. doi: 10.1093/schbul/syp004
Walden N, German DA, Wolf EM, et al (2020) Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. Nat Commun 11:1–12. doi: 10.1038/s41467-020-17605-7
Wasowicz P, Pauwels M, Pasierbinski A, et al (2015) Phylogeography of *Arabidopsis halleri* (Brassicaceae) in mountain regions of Central Europe inferred from cpDNA variation and ecological niche modelling. PeerJ 4:e1645. doi: 10.7287/peerj.preprints.1259v1
Willems G, Dräger DB, Courbot M, et al (2007) The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): An analysis of quantitative trait loci. Genetics 176:659–674. doi: 10.1534/genetics.106.064485
Willis KJ, van Andel TH (2004) Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. Quat Sci Rev 23:2369–2387. doi: 10.1016/j.quascirev.2004.06.002
Wohlfarth B, Bjorck S, Funder S, et al (2008) Quaternary of Norden. Episodes 31:73–81
Wos G, Mořkovská J, Bohutínská M, et al (2019) Role of ploidy in colonization of alpine habitats in natural populations of *Arabidopsis arenosa*. Ann Bot. doi: 10.1093/aob/mcz070
Yant L, Hollister JD, Wright KM, et al (2013) Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. Curr Biol 23:2151–2156. doi: 10.1016/j.cub.2013.08.059
Zachos FE (2016) Species Concepts in Biology. Springer International Publishing, Cham
Zink RM (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. Proc R Soc London Ser B Biol Sci 271:561–564. doi: 10.1098/rspb.2003.2617
Zozomová-Lihová J, Malánová-Krásná I, Vít P, et al (2015) Cytotype distribution patterns, ecological differentiation, and genetic structure in a diploid– tetraploid contact zone of *Cardamine amara*. Am J Bot 102:1380–1395. doi: 10.3732/ajb.1500052

PART B – CASE STUDIES

Paper I

Kolář F, Lučanová M, Záveská E, **Fuxová G**, Mandáková T, Španiel S, Senko D, Svitok M, Kolník M, Gudžinskas Z, Marhold K (2016) Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae), *Biological Journal of the Linnean Society*, Volume 119, Issue 3, Pages 673–688.

The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2016, **119**, 673–688. With 4 figures.

Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae)

FILIP KOLÁŘ[1,2], MAGDALENA LUČANOVÁ[2,1], ELIŠKA ZÁVESKÁ[1], GABRIELA FUXOVÁ[1], TEREZIE MANDÁKOVÁ[3], STANISLAV ŠPANIEL[1], DUŠAN SENKO[4], MAREK SVITOK[5,6], MARTIN KOLNÍK[7], ZIGMANTAS GUDŽINSKAS[8] and KAROL MARHOLD FLS[1,4]*

[1]*Department of Botany, Faculty of Science, Charles University in Prague, Benátská 2, CZ-128 01 Prague, Czech Republic*
[2]*Institute of Botany, Academy of Sciences of the Czech Republic, Zámek 1, CZ-252 43 Průhonice, Czech Republic*
[3]*Plant Cytogenomics Research Group, Central European Institute of Technology (CEITEC), Masaryk University, Kamenice 5, CZ-62500 Brno, Czech Republic*
[4]*Institute of Botany, Slovak Academy of Sciences, Dúbravská cesta 9, SK-845 23 Bratislava, Slovak Republic*
[5]*Department of Biology and General Ecology, Faculty of Ecology and Environmental Sciences, Technical University in Zvolen, T. G. Masaryka 24, SK-960 53 Zvolen, Slovak Republic*
[6]*Eawag Swiss Federal Institute of Aquatic Science and Technology, Department of Aquatic Ecology, Centre of Ecology, Evolution and Biogeochemistry, Seestrasse 79, CH-6047 Kastanienbaum, Switzerland*
[7]*Tematínska 4, SK-91501 Nové Mesto nad Váhom, Slovak Republic*
[8]*Nature Research Centre, Institute of Botany, Laboratory of Flora and Geobotany, Žaliųjų Ežerų Str. 49, LT-08406 Vilnius, Lithuania*

Received 4 May 2014; revised 18 August 2014; accepted for publication 8 December 2014

Detailed knowledge of the geographic distribution of cytotypes is a prerequisite for any experimental or molecular study of ploidy-variable plant systems. The *Arabidopsis arenosa* group, an intricate di-tetraploid complex from the plant model genus *Arabidopsis*, has remained largely neglected regarding the distribution and habitat associations of its cytotypes. Using flow cytometry, we conducted a large population-level cytological screen across the *A. arenosa* group range, involving more than 2900 individuals from 194 populations. We characterized a largely parapatric distribution of the diploid (Southeast Europe) and tetraploid (Northwest Europe) cytotypes with two contact zones – a narrow contact zone in the Slovenian Forealps and a diffuse contact zone across the Carpathians. In addition, a previously unknown isolated diploid lineage with distinct ecology was revealed from sandy areas of the southeastern Baltic coast. We also recorded several adult triploid individuals for the first time in wild *Arabidopsis arenosa*. Particularly in the Western Carpathians, the diploid and tetraploid populations are largely intermingled, and both cytotypes are spread along the whole lowland-alpine gradient of habitats, exhibiting no signs of ploidy-linked habitat differentiation. In contrast with the complexity at the landscape scale, the within-population cytological homogeneity and the rare occurrence of triploids indicate that the contact zone is rather stable.

KEYWORDS: contact zones – *Cruciferae* – environmental predictors – flow cytometry – habitat differentiation – polyploidy.

*Corresponding author. E-mail: karol.marhold@savba.sk

INTRODUCTION

The *Arabidopsis arenosa* group, a diploid-tetraploid species complex, represents one of the closest relatives of the prominent plant model *Arabidopsis thaliana* (Clauss & Koch, 2006). Polyploidization is a major diversification force in the complex, generating an intricate mixture of diploid populations and their tetraploid derivatives. Importantly, origin of the tetraploid populations solely from diploid representative(s) of the *A. arenosa* group is suggested by the cytotype distribution pattern, morphological similarities (Měsíček, 1970), close AFLP multilocus phenotypes (Schmickl *et al.*, 2012) and overall similarity in genome scans (Hollister *et al.*, 2012). The close relationships among diploid and tetraploid *A. arenosa* cytotypes represent a unique feature within *Arabidopsis*, as other wild polyploid members are of allopolyploid (hybrid) origin, based on more distantly related parents (*A. suecica*, Jakobsson *et al.*, 2006; *A. lyrata* subsp. *petraea*, Schmickl & Koch, 2011; *A. kamchatica*, Shimizu-Inatsugi *et al.*, 2009). The *Arabidopsis arenosa* group thus emerges as a highly promising system for addressing general questions on polyploidy in natural plant populations. Indeed, the first studies dealing with general evolutionary questions in this group have emerged recently, addressing the evolution of meiotic stability in polyploids (Hollister *et al.*, 2012; Yant *et al.*, 2013) and speciation processes (Jørgensen *et al.*, 2011; Schmickl & Koch, 2011).

The *A. arenosa* group comprises up to nine taxa (species or subspecies, partly still not formally described) spanning a wide ecological range from coastal sand dunes to high-alpine environments with a principal diversity centre most likely situated in the Carpathian Mountains in eastern Central Europe (Měsíček, 1998; Měsíček & Goliašová, 2002; Schmickl *et al.*, 2012). Available cytological data indicate that the Carpathian mountain arch harbours a complex mixture of diploid and tetraploid populations [chromosome counts by Měsíček (1970), F. Krendl and A. Polatschek (published in Schmickl *et al.*, 2012)]. In particular, the Western Carpathians appear to be a hotspot of ecological and taxonomic diversity of the whole species complex. There, populations of both diploid and tetraploid representatives of the *A. arenosa* group co-occur along the entire altitudinal gradient, from dry and warm steppes in the foothills (150 m a.s.l.) via shady rocks and screes on various substrates to alpine vegetation on the highest summits (2600 m a.s.l., Měsíček & Goliašová, 2002). This extensive cyto- and eco-geographical variation is remarkable both in general and particularly in the Carpathians, where the largest cytotype mixture of the *A. arenosa* group is found.

In the Carpathians, the few large-scale cytotype screens published to date are inconclusive with respect to general cytogeographic patterns. They range from near cytological homogeneity (*Vicia cracca*, Trávníček, Eliášová & Suda, 2010; *Alyssum montanum*, Španiel *et al.*, 2011) through the absence of geographical patterns and extensive intra-population cytotype mixture (*Phleum pratense* agg.; Perný *et al.*, 2008) to a relatively strong altitudinal differentiation (*Sesleria calcarea* – *S. tatrae* species complex, Lysak & Doležel, 1998; *Senecio jacobaea*, Hodálová *et al.*, 2007; *Pilosella officinarum*, Mráz *et al.*, 2008; *Knautia arvensis* agg., Kolář *et al.*, 2009). However, none of these studied species spans the entire altitudinal range of habitats.

A prerequisite for any ecological and/or molecular study of a ploidy-heterogeneous plant system is knowledge of the geographic distribution of cytotypes. Cytogeographic data complement phylogenetic and experimental data and serve as a foundation for addressing questions of frequency of polyploid formation, ecological differentiation of cytotypes, and the genetic background of polyploid evolution. For comprehensive evaluation of the true extent of diversity and dynamics of ploidy-mixed plant systems (e.g., detection of minority-represented cytotypes such as triploids), a sufficiently large and geographically wide flow cytometric screen is essential (Duchoslav, Šafářová & Krahulec, 2010; Trávníček *et al.*, 2011a,b; Krejčíková *et al.*, 2013, see Kron, Suda & Husband, 2007 for review). Despite an increasing interest in evolutionary, ecological, and genomic studies of the *A. arenosa* group, we still have only fragmentary knowledge on its karyological diversity and habitat associations. Most of the published records on the ploidy distribution are based on traditional low-throughput chromosome counting (allowing ploidy determination of a few individuals per population) and/or focus on the uniform tetraploid-inhabited regions of Western and Northern Europe (Měsíček, 1970; Schmickl *et al.*, 2012).

In this study, we employed a high-throughput technique for ploidy estimation – flow cytometry – complemented with chromosome counts to assess ploidy level and homoploid genome size diversity over the entire distribution range of the *A. arenosa* group. Considering the intricate and still unresolved relationships within this group, our study addressed only general patterns across the whole species complex and did not aim to resolve its internal taxonomic structure. Specifically, we addressed the following questions: (1) What is the pattern of ploidy distribution, especially of the so far undersampled diploids, and where are the cytotype contact zones located? (2) What is the ploidy level variation within populations? Are there any indications of recent polyploidization

events and/or inter-ploidy gene flow? (3) What is the level of variation in DNA content at the homoploid level and, if present, is this variation geographically structured? (4) Are there any indications for substantial niche differentiation between the cytotypes along large-scale environmental gradients (altitude, climatic niche, substrate, disturbance levels)? If so, is the differentiation stronger in the areas where both cytotypes co-occur in sympatry (Western Carpathians)?

MATERIAL AND METHODS

Field sampling

In total, 2963 individuals from 194 populations were collected across the entire range of the *Arabidopsis arenosa* group from 2011 to 2013. The sampling covered all currently recognised species and subspecies of the complex (except for the geographically, morphologically and ecologically distinct diploid stenoendemic *A. croatica*), namely *Arabidopsis arenosa* (L.) Lawalrée subsp. *arenosa*, *A. arenosa* subsp. *borbasii* (Zapał.) O'Kane & Al-Shehbaz, *A. carpatica* nom. prov., *A. neglecta* (Schult.) O'Kane & Al-Shehbaz subsp. *neglecta* nom. prov., *A. neglecta* subsp. *robusta* nom. prov., *A. nitida* nom. prov., *A. petrogena* (A. Kern.) V.I. Dorof. subsp. *petrogena* nom. prov., *A. petrogena* subsp. *exoleta* nom. prov. The above-mentioned provisional names on the level of species and subspecies were introduced in the genus *Cardaminopsis* by Měsíček (1970, 1998 and unpublished manuscript), but they were never validly published. Valid publication of these names requires further studies, and we are using them solely for a reference to other papers using this nomenclature (corresponding names are also used in the locality list in Table S1). Whenever possible, fresh tissues (preferably parts of stems with flowers) mostly from 1 to 20 (up to 51) individuals per population (15 individuals on average) were collected and placed in cold storage until flow cytometric evaluation. In selected populations, we also collected seeds for direct counts of chromosome numbers. We recorded GPS co-ordinates and altitude and characterized the environmental conditions of each site using the following parameters: habitat type, geological substrate and natural/anthropogenic character. Localities were considered anthropogenic only in cases of heavily human-disturbed or entirely human-created habitats (wall crevices, railway tracks, gravel deposits, etc.). Nevertheless, these taxa often colonise such sites as a result of accidental spreading from adjacent natural stands (e.g., road bank below a rock). To differentiate between such short-distance spontaneous colonization and long-distance anthropogenic spread, we further divided the anthropogenic stands into those close (less than approximately 1 km) to a natural habitat and those occupying purely anthropogenic habitats far from any potential natural locality (typically road banks and railway tracks). For locality details, see supplementary Table S1.

Flow cytometry

DNA ploidy level (Suda *et al.*, 2006) was inferred from nuclear DNA content determined by flow cytometry following the simplified two-step protocol (Doležel, Greilhuber & Suda, 2007). Approximately 10 square millimetres of fresh leaf tissue or one fresh petal from each plant to be analysed was chopped together with an appropriate volume of the internal reference standard (*Solanum pseudocapsicum*, 2C = 2.59 pg, Temsch, Greilhuber & Krisai, 2010; the same individual was used for all measurements) using a sharp razor-blade in a Petri dish containing 0.5 mL of ice-cold Otto I buffer (0.1 M citric acid, 0.5% Tween 20). The suspension was filtered through 42-μm nylon mesh and incubated for 10 min at room temperature. Isolated nuclei were stained with 1 mL of Otto II buffer (0.4 M $Na_2HPO_4 \cdot 12H_2O$) supplemented with 4,6-diamino-2-phenylindole (DAPI) at 4 μg mL^{-1} and β-mercaptoethanol at 2 μg mL^{-1}. After 1 min of incubation, the sample was run for 3000 particles in a Cyflow ML flow cytometer (Partec, Münster, Germany) equipped with the UV-led lamp. The histograms were evaluated with FloMax FCS 2.0 software (Partec, Münster, Germany). Fresh petals were preferred over vegetative parts for these analyses due to the absence of endopolyploidy (Galbraith, Harkins & Knapp, 1991). For petal samples, we analysed up to five individuals in a pooled sample to reduce the analysis costs and time demand. Our previous experiments showed that such practice enables reliable detection of minority cytotypes present even at a low proportion (20%). Nevertheless, each plant was separately re-analysed if mixed samples were suspected, peaks were asymmetrical, or the coefficient of variance of the *Arabidopsis* peak exceeded 5%. The same approach was applied for pooled leaf samples of tetraploids (a potential diploid or triploid individual would be clearly identified as an additional peak with lower fluorescence intensity); however, vegetative parts from diploid individuals were analysed individually in any case due to the presence of the tetraploid endopolyploid peak. In ten (5%) populations where fresh tissue was not available, we used samples dried with silica gel for ploidy estimation using the same protocol (see Table S1).

For genome size estimation, one individual per selected population (see Table S1) was run on a CyFlow SL flow cytometer (Partec, Münster,

Germany) equipped with a green (532 nm) solid-state laser. The sample preparation followed the methodology described above, with the only modification being that the stain solution consisted of Otto II buffer enriched with propidium iodide and RNase (both at 50 μg mL^{-1}) and β-mercaptoethanol at 2 μg mL^{-1}. The analyses were run for 5000 particles. We applied the following stringent criteria to obtain precise and stable estimates of genome size: (i) only analyses with the coefficient of variation of the sample peak below 3% were taken into account, (ii) each sample was measured at least three times on different days to minimise potential random instrumental drift (Doležel & Bartoš, 2005), and (iii) the between-day variation was defined to not exceed 3%; otherwise, the most remote value was discarded and the sample was re-analysed. The reliability of flow cytometric measurements (i.e., between-plant differences) was repeatedly confirmed in simultaneous runs of *Arabidopsis* accessions with distinct genome sizes (Greilhuber, 2005).

Chromosome preparations

Plants for chromosome counts were selected such that they covered the entire sampling area. Plants were grown from seeds in plastic Petri dishes on sieved potting soil in a phytotron with long day illumination (16 h light at 20 °C, 8 h dark at 15 °C). Young inflorescences were fixed in ethanol/acetic acid (3 : 1, v/v) fixative for 24 h at 4 °C. The fixative was replaced with 70% ethanol, and the material was stored at –20 °C until further use. Chromosome spreads were prepared as described by Mandáková, Marhold & Lysak, (2014). Slides were examined under phase contrast for the presence of suitable mitotic metaphase spreads. Selected preparations were stained with 2 mg mL^{-1} DAPI in Vectashield anti-fade mounting medium (Vector Laboratories, Burlingame, CA, USA) and photographed using an Olympus BX-61 epifluorescence microscope and a CoolCube camera (MetaSystems, Altlussheim, Germany). Individual images were processed with Photoshop CS software (Adobe Systems, San Jose, CA, USA).

Data analyses

Spatial segregation of cytotypes across the entire range (except for the spatial outlier Scandinavian populations) and separately within the Western Carpathian contact zone was analysed using the Mantel test implemented in the ade4 R package (Dray & Dufour, 2007). A correlation coefficient (r_M) was calculated for: (i) the matrix of mutual geographic distances among populations; and (ii) the binary matrix of ploidy levels, and it was compared to the distribution of coefficients obtained from matrices generated by random rearrangements (9999 permutations) of the original matrices. Only the majority ploidy level of the population was considered (i.e., rare triploid cytotypes were omitted). In addition, Mantel tests were used for testing the spatial autocorrelation of homoploid genome size by comparing a matrix of geographic distances with genome size distance matrix for a particular cytotype (diploid and tetraploid accessions were analysed separately).

Differences among the cytotypes in associations with anthropogenic stands and geological substrates (assessed only for non-anthropogenic populations) were assessed using the chi-squared test in contingency tables (*P*-values were assessed using 200 replicates). General linear models were used for testing the association of cytotypes with altitude as well as for the relationships among homoploid genome size and the following environmental predictors: (non)anthropogenic character of the original habitat, altitude, and substrate type (the last one only for natural localities). Unless stated otherwise, all analyses were performed in R 2.15.2 (R Development Core Team, 2013).

To capture the interrelationship of environmental predictors and ploidy level in sufficiently detailed scale, it was necessary to use background climatic and landscape data, which are long-term averages and provide seasonal variability. Primary data layers that included air temperature, solar radiation, and terrain (elevation, horizon) were obtained from the SolarGIS data, version 1.9 (the high-resolution climate database operated by GeoModel Solar, Bratislava, Slovakia). Data on air temperature at 2 m (in °C) were derived from the Climate Forecast System Reanalysis and Global Forecast System databases (National Centers for Environmental Prediction, Suitland, Maryland, USA) for the period from 1994 to 2011, recalculated to 15-minute values. The data were spatially enhanced to 1-km resolution to reflect variability induced by high-resolution (dissected) terrain. Solar radiation was calculated from the satellite and atmospheric data. The sources were: (i) Meteosat First and Second Generation (PRIME and Indian Ocean Data Coverage Regions, European Organisation for the Exploitation of Meteorological Satellites, Darmstadt, Germany) in 15-min or 30-min values, (ii) outputs from the Monitoring Atmospheric Composition and Climate (European Centre for Medium-Range Weather Forecasts, Reading, UK) for the decade from 2003 to 2013, and (iii) atmospheric models from Global Forecast System database (National Oceanic and Atmospheric Administration, Silver Spring, Maryland, USA) for the period from 1994 to 2013. Solar radiation represents annual (total) and monthly long-term averages of global irra-

diation: (i) without (global horizontal irradiation, GHI), and (ii) with impinging on local terrain accounting for the slope and azimuth of the terrain (GTI) (in $kWhm^{-2}$) and annual (total) and monthly long-term averages of photosynthetically active radiation (PAR) (400–700 nm in $kWhm^{-2}$). Monthly long-term averages of precipitation were obtained from WorldClim, version 1.4 (Hijmans *et al.*, 2005). For the purpose of this study, the hourly data on air temperature and solar radiation were integrated into long-term monthly averages. These averages were further spatially enhanced by disaggregation, based on the correlation between terrain altitude and climatic variables. The disaggregated monthly and yearly averages created from this reanalysis were validated against selected ground measurements (from the meteostations flagged with quality codes 2, 3, 6, 7; see list of quality codes from the National Climatic Data Center). Based on disaggregation and validation, which was calculated individually for each pixel (smallest grid unit), these data (rasters) in the GIS (Geographic Information System) environment represent annual trends, seasonality and extremes for particular areas. Morphometry of the terrain (terrain slope, terrain azimuth) was developed via elevation [altitude above sea level; source SRTM3 data (The Shuttle Radar Topography Mission, available at http://srtm.usgs.gov/) up to the latitude 60°N]. We calculated distances from the Equator (northing) and the prime meridian (easting) in kilometres to account for spatial gradients and autocorrelation. For these calculations, we used PostGIS/PostgreSQL, version 1.5.1, released under the GNU/GPL license.

Distribution of the major ploidy levels (diploids and tetraploids) was modelled using generalized linear models (GLM) with binomial error distribution and the logit link function (i.e., logistic regression). A range of GIS-derived data was used as environmental explanatory variables (see Table S2 for a complete list of variables and abbreviations of variable names). Northing and easting were used as spatial predictors to detect possible geographic gradients. Prior to the analyses, distribution of variables and correlations among them were assessed. To avoid a multicollinearity, elevation was excluded from modelling due to its strong correlation with mean annual temperature ($r = -0.94$). Intrinsically strong positive correlations were found among monthly values and annual summary characteristics of temperature (Fisher weighted mean $r = 0.98$), precipitation ($r = 0.82$), GHI ($r = 0.85$), GTI ($r = 0.92$) and PAR ($r = 0.85$); thus, only annual characteristics were pre-selected for further analyses. However, annual GTI, GHI and PAR were highly correlated with each other ($r = 0.98$). Consequently, only PAR was employed as a predictor in the analyses because this quantity is intuitively understandable and is a frequently used measure of radiation. The remaining variables did not show considerable skewness or intercorrelations and were used in the modelling procedure as predictors (see Table S3 for a list of predictors). Separate GLMs were built for the whole dataset and the Western Carpathian contact zone. Initially, full models were fitted to the data, including all spatial and environmental predictors. The full models were simplified following backward stepwise deletion associated with likelihood-ratio tests. Only those variables for which the conditional effect was significant at $\alpha = 5\%$ were retained in the final models. Spatial correlograms were used to check for autocorrelation in the residuals of the final models. Because the final models showed significant positive autocorrelation at short distances, the data were re-fitted using generalized mixed effect models (GLMM) (Dormann *et al.*, 2007) to prevent biased estimates of model coefficients and the inflation of type I errors. GLMMs with Gaussian spatial correlation structure were fitted using penalised quasi-likelihood (Venables & Ripley, 2002). Final GLMMs are presented graphically as a series of effect plots (Fox, 2003). The ability of the final models to discriminate between sites with diploids and those with tetraploids was assessed by means of classification tables (cut-off value: 0.5) and Somers' Dxy rank correlations (Newson, 2006) between observed incidences of cytotypes and predicted probabilities.

RESULTS

Ploidy level variation and cytogeography

Three different DNA ploidy levels (diploid – $2x$, triploid – $3x$, and tetraploid – $4x$) were detected among 2963 individuals from 194 populations belonging to the *A. arenosa* group (Fig. 1). The tetraploid individuals [1588 (54%) individuals in 107 (55%) populations] only slightly prevailed over their diploid counterparts [1369 (46%) individuals in 88 (45%) populations]. The triploid cytotype was extremely rare (six individuals, 0.2%) and it was in all cases represented by a single individual each in otherwise diploid populations. Despite cytotype co-occurrence in several areas and a large within-population sampling (15 individuals per population were sampled on average), the vast majority of the populations (96%) were detected as cytotype uniform, i.e., either diploid or tetraploid. Only a single di-tetraploid mixed-ploidy population was found in the Tatry Mts. (Western Carpathians, AA170) in addition to diploid-triploid mixtures recorded at six sites across the diploid cytotype range (see Table S1, for locality details). Chromosome counts confirmed the estimated ploidy levels and revealed $2n = 2x = 16$ in 17 accessions from

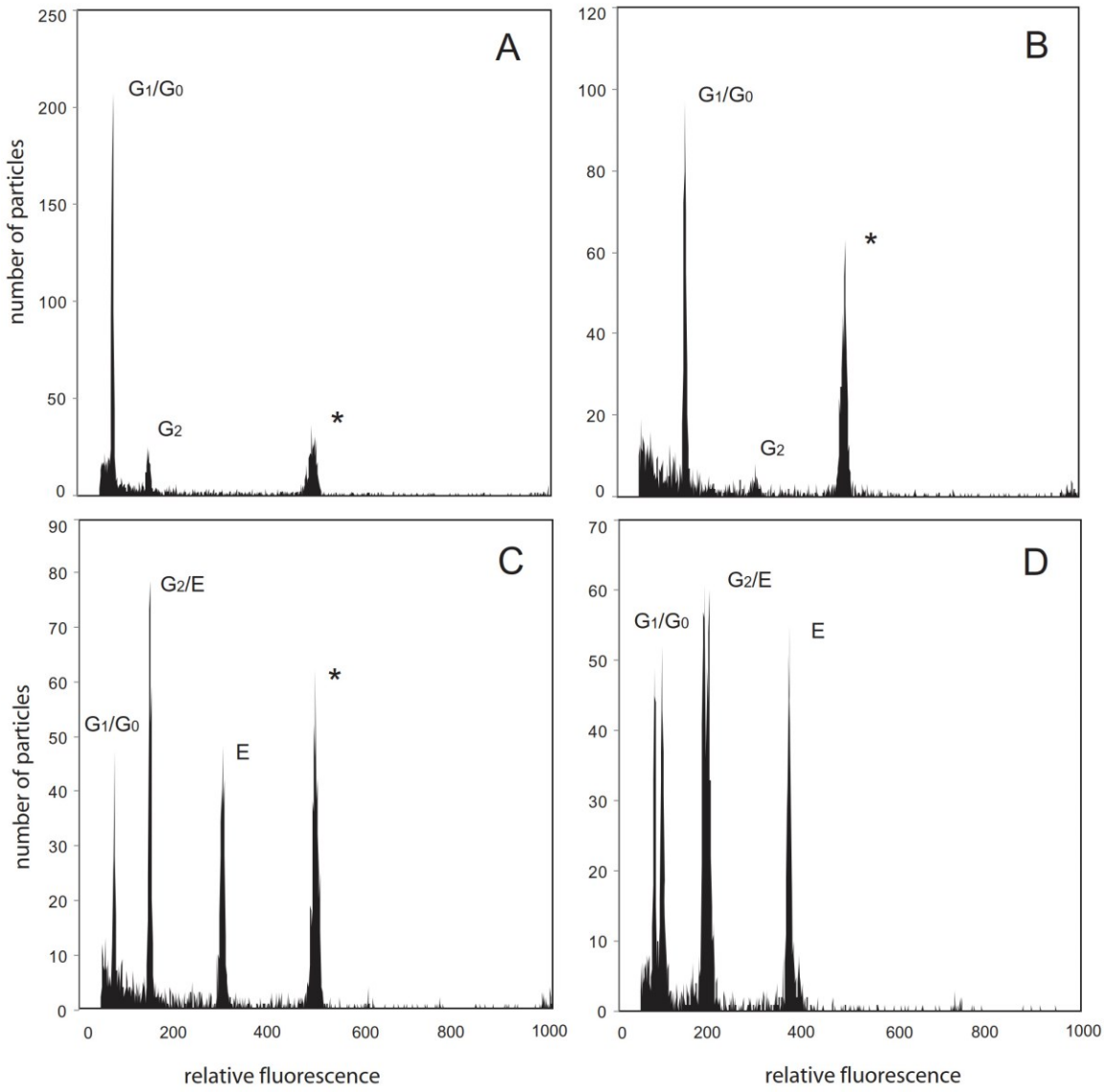


Figure 1. Flow cytometric histograms of suspensions of DAPI-stained nuclei isolated from diploid (A, C, D) and tetraploid (B) accessions of the *Arabidopsis arenosa* group. A + C, Analysis of nuclei of identical diploid individuals (pop. AA084) isolated from either fresh petal (A) or stem leaf (C). B, Pooled sample of five tetraploid individuals (pop. AA117, nuclei isolated from fresh petal tissue). D, Simultaneous analysis of two diploid accessions from pop. AA090 documenting within-population divergence in nuclear DNA contents (difference in fluorescence intensity, 14%; nuclei from both samples were simultaneously isolated, stained, and analysed). Letters denote peaks of nuclei corresponding to different phases of the cell cycle (G_0–G_2) and/or levels of endopolyploidy (E); the internal standard *Solanum pseudocapsicum* used in analyses A-C is marked by an asterisk.

the Carpathians (AA018, AA023, AA070, AA084, AA090, AA091, AA123, AA157), Dinaric Alps (AA054, AA124, AA125, AA126, AA127, AA128), Pannonian lowland (AA110), and southern Baltic coast (AA153, AA200) and $2n = 4x = 32$ in 10 accessions from the Carpathians (AA015, AA067, AA082, AA087, AA088), southern and eastern Alps (AA049, AA149), southern Poland (AA059), Scandinavia (AA181) and Luxem-

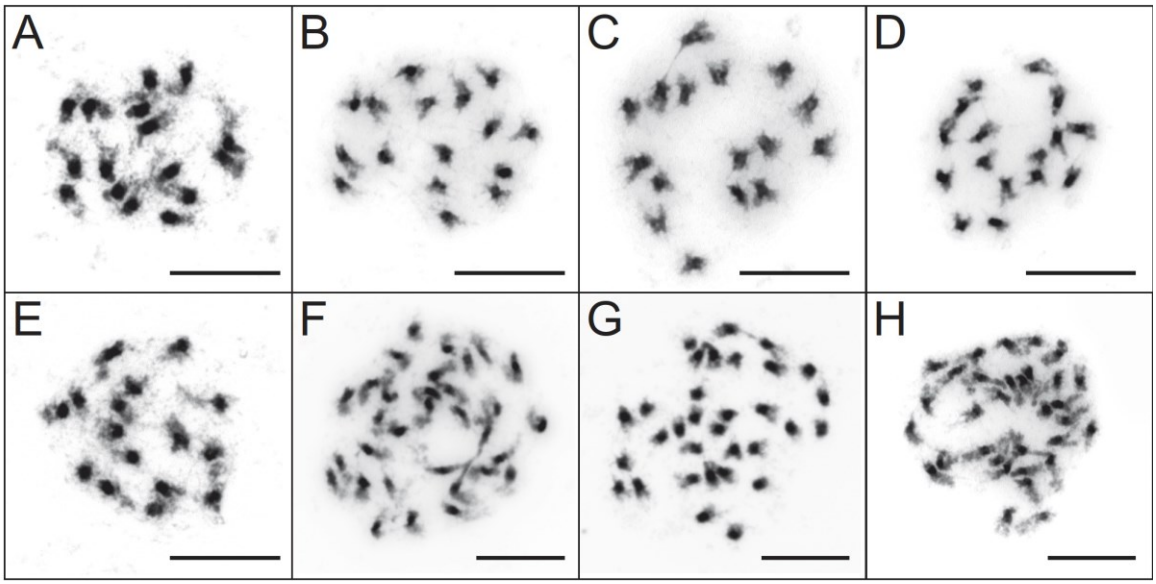


Figure 2. DAPI-stained mitotic chromosome spreads from flower bud tissue of the *Arabidopsis arenosa* group. A, *A. arenosa* s.l. AA200 (Lithuania, coastal sands; $2n = 2x = 16$). B, *A. arenosa* s.l. AA124 (Serbia, dry rocks; $2n = 2x = 16$). C, *A. arenosa* s.l. AA070 (Romania, dry rocks; $2n = 2x = 16$). D, *A. neglecta* AA084 (Slovakia, alpine scree; $2n = 2x = 16$). E, *A. carpatica* AA023 (Slovakia, limestone outcrop in middle altitudes; $2n = 2x = 16$). F, *A. arenosa* AA181 (Norway, secondary gravel; $2n = 4x = 32$). G, *A. neglecta* subsp. *robusta* AA087 (Slovakia, alpine rocks; $2n = 4x = 32$). H, *A. petrogena* subsp. *exoleta* AA082 (Romania, limestone rocks; $2n = 4x = 32$). See Table S1 for locality details. Scale bars = 10 μm.

bourg (AA190) (Fig. 2, Table S1). Neither dysploidy, aneuploidy nor accessory chromosomes were observed in the karyologically investigated accessions.

Diploid and tetraploid cytotypes exhibited a largely parapatric distribution; a weak but significantly non-random spatial differentiation of cytotypes was also supported by the Mantel test ($r_M = 0.06$, $P < 0.001$). Tetraploids dominate in the northwestern half of the *A. arenosa* group range (Scandinavia, Germany, Alps, Hercynian massif) whereas diploids occupy mainly southeastern areas (most of the Carpathians, Pannonian basin, Dinaric Alps, Fig. 3). In addition, four spatially isolated diploid populations were found along southern shores of the Baltic Sea. They grew exclusively in coastal sand dunes and in adjacent open forests and thus occupied distinct environments from their spatially closest tetraploid counterparts that were found exclusively in human-disturbed habitats (Table S1). Natural populations of both cytotypes meet at the landscape scale in two contact zones, a smaller and rather abrupt one situated in Slovenia (less than 100 km wide) and a large and diffuse zone across the Carpathian mountain arch (Fig. 4). In the Romanian Carpathians, the tetraploids occupy the northern half of the Eastern Carpathians and the Apuseni Mts., whereas diploids dominate in Southern Carpathians and in the southern half of the Eastern Carpathians (the only exceptions in this area are two tetraploid populations, AA065 and AA067, occupying alpine screes and a limestone canyon, respectively). In contrast, in the Western Carpathians, the diploid and tetraploid populations were largely spatially intermingled throughout the landscape (Fig. 4) although the cytotypes still exhibited weak but significant spatial associations (Mantel test, $r_M = 0.06$, $P = 0.013$).

Homoploid differentiation in DNA content

In addition to ploidy variation, the accessions of the *A. arenosa* group also exhibited a considerable variation in DNA content at the homoploid level as the di- and tetraploid accessions varied 1.17-fold and 1.21-fold, respectively. Nevertheless, this range included two diploid individuals and one tetraploid individual with abruptly higher genome sizes (9–13% higher than the average, see Fig. S1). After exclusion of these three individuals, the variation dropped to 1.12-fold and 1.14-fold in diploids and tetraploids, respectively. Homoploid genome size was not spatially structured, as evidenced by non-significant Mantel tests ($r_M = -0.11$, $P = 0.88$ and $r_M = 0.07$, $P = 0.19$, for diploid and tetraploid accessions, respectively). In addition, a comparable 1.14-fold difference was found among five individuals from one exceptionally highly

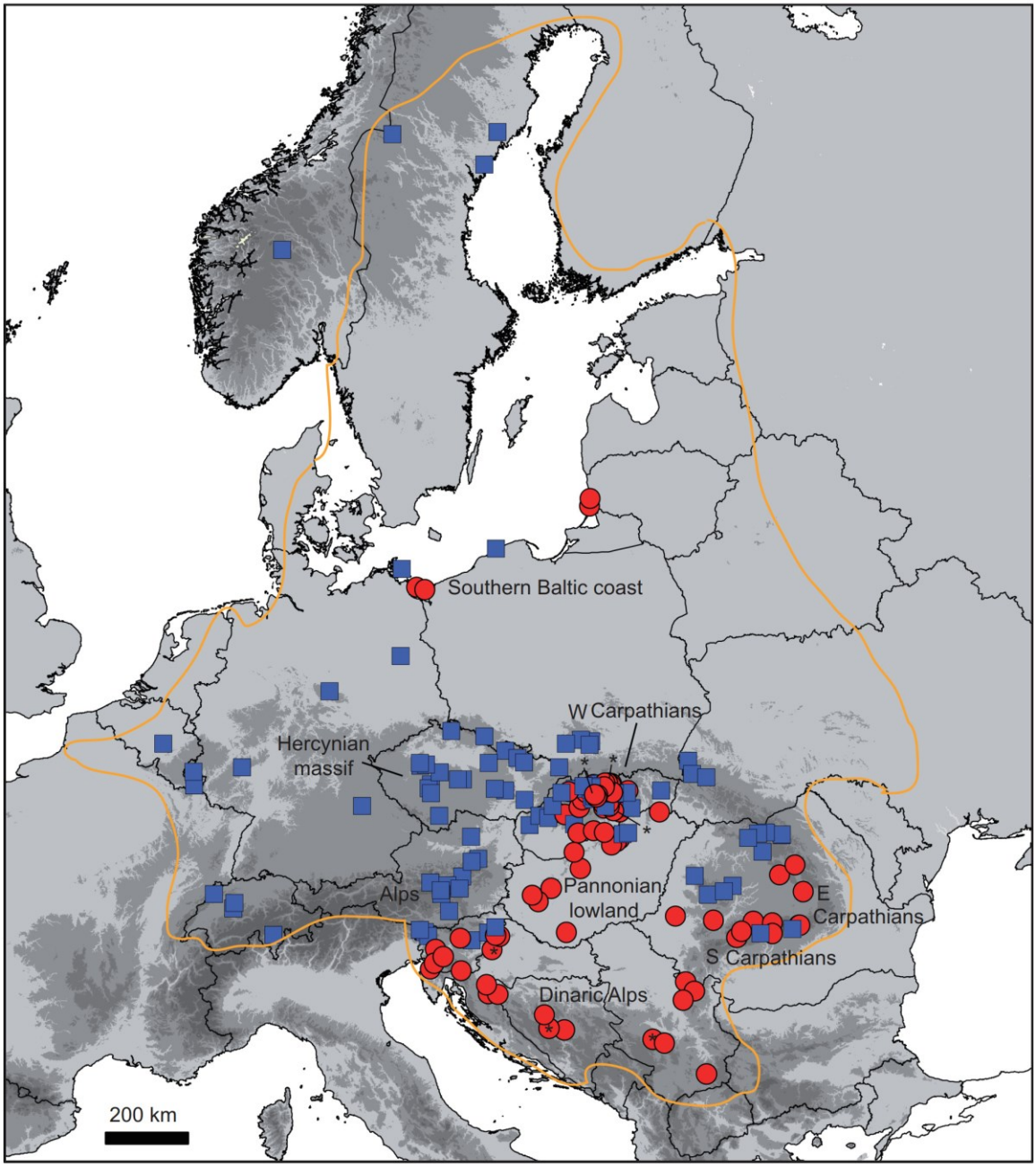


Figure 3. Distribution and ploidy level of the 194 studied populations of the *Arabidopsis arenosa* group in Europe (red – diploid, blue – tetraploid, asterisk – triploid, 2963 individuals investigated in total). The continuous distribution range of the whole species complex is marked by the orange outline (following Hoffmann, 2005).

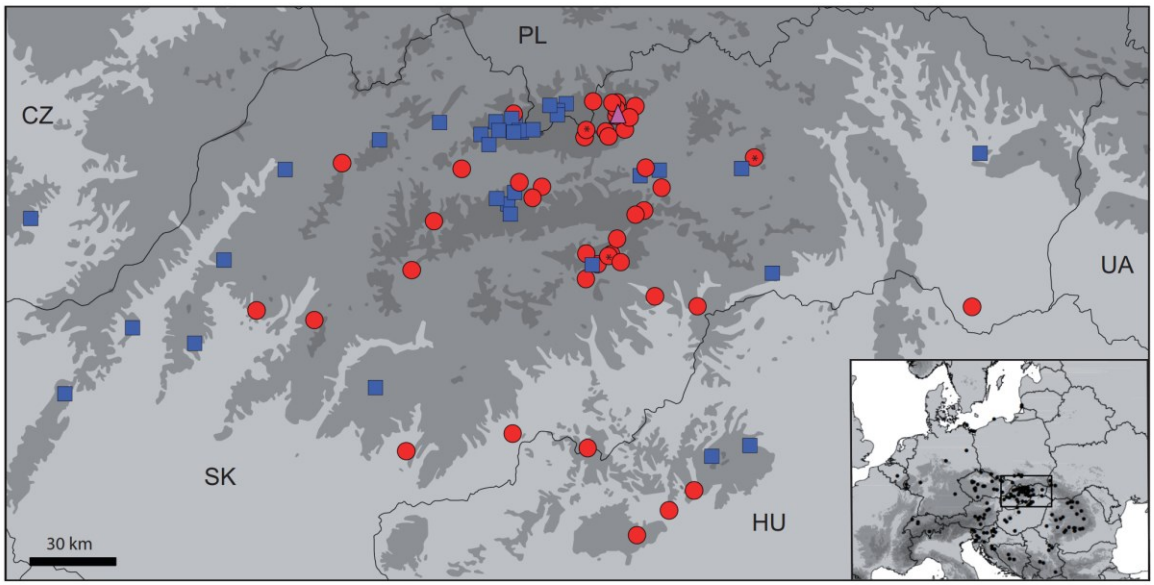


Figure 4. Detail of the contact zone of cytotypes of the *Arabidopsis arenosa* group in the Western Carpathians (red circle – diploid, blue square – tetraploid, pink triangle – mixed di-tetraploid population, asterisk – triploid; based on 1374 individuals from 79 populations).

variable population from the alpine zone of the Western Carpathians (pop. AA090, see also Fig. 1D). The genome size variation was correlated neither with (non)-anthropogenic habitat character ($F_{1,31}$ = 0.92, P = 0.34 in tetraploids; diploids were not tested due to the negligible proportion of anthropogenic populations, see the next section) nor with altitude, substrate type and/or their interaction ($F_{3,30}$ = 0.61, P = 0.613 and $F_{3,20}$ = 1.37, P = 0.28 in diploids and tetraploids, respectively). Mean monoploid DNA content (after exclusion of the individuals with exceptionally high values) was similar among all three ploidy levels, though it was not entirely identical (average ratio to internal standard divided by ploidy level was 0.068, 0.070, and 0.073 for 2*x*, 3*x*, and 4*x*, respectively: the tetraploid value was on average 7.6% higher than that of diploids).

Niche differentiation

Distribution of ploidy levels through the entire investigated area was significantly correlated to a west/east gradient (easting), total annual PAR and total annual precipitation (Table 1). Probability of tetraploid occurrence decreased toward the east and also with increasing PAR and precipitation (Fig. S2). Considering contact zone data, only a south/north gradient appeared significant (Fig. S3). Generally, the cytotypes occupied somewhat different climatic niches as revealed by the moderately high discriminatory power of the models. Nevertheless, the particular climatic factors strongly reflected by the spatial gradients and only two environmental predictors (total annual PAR and total annual precipitation) significantly improved the GLMM with incorporated geographical predictor in the entire *A. arenosa* group area. No environmental predictors were shown to be significant in the contact zone (Table 1).

Almost no significant differences in substrate requirements and/or altitudinal ranges of the diploids or tetraploids were detected, whether across the entire area or in the densely sampled zone of sympatry in the Western Carpathians. The only exception was a significant association of tetraploids with anthropogenic stands (Table 2). Although both cytotypes were able to grow in habitats created or disturbed by man in close proximity to the natural stands (14 vs. 10 localities for tetraploids and diploids, respectively), the tetraploids were significantly more frequent (22 vs. 4 localities) in anthropogenic stands distant from natural localities, i.e., showing stronger potential for anthropogenic spread. Nevertheless, this difference was not apparent within the Western Carpathian contact zone because tetraploids occupied the anthropogenic stands in other parts of the distributional range (mainly in the northern part, i.e., Scandinavia, Poland, northern Germany, and northern Czech Republic).

Table 1. Summary of the final logistic generalized mixed effect models (GLMMs) testing the effects of spatial gradients (northing, easting) and the environmental correlates (total annual PAR, total annual precipitation; only those with the conditional effect significant at $\alpha = 5\%$ are presented) on the distribution of diploid and tetraploid populations of the *Arabidopsis arenosa* group in the entire range of the group and in the Western Carpathian contact zone

	Whole model				Model parameters				
Data set	χ^2	*P*	class (%)	Somers' Dxy (95% CL)	Predictor (unit)	B	SE	$\chi^2_{(1)}$	*P*
Whole data set	36.7	<0.001	71.2	0.59 (0.45, 0.71)	Easting (km)	–0.0036	0.0007	23.67	<0.001
					Total annual PAR ($kWh.m^{-2}$)	–0.0196	0.0038	27.62	<0.001
					Total annual precipitation (mm)	–0.0014	0.0007	4.47	0.035
Contact zone	4.17	0.041	59.0	0.29 (0.05, 0.54)	Northing (km)	0.0123	0.0061	4.17	0.041

Characteristics of the final models: χ^2, test statistics; *P*, probabilities; class and Somers' Dxy, classification success. Characteristics of particular parameters: B, estimates of model coefficients; SE, standard error of estimates; $\chi^2_{(1)}$ and *P*, results of likelihood-ratio tests.

Table 2. Differences among diploid and tetraploid populations of the *Arabidopsis arenosa* group from the entire distribution area and from the contact zone in the Western Carpathians in the investigated habitat characteristics (significant results are in bold)

	2*x*	4*x*	Test
Anthropogenic stands+			
Whole range	74/14	36/70	**$\chi^2 = 8.19$, $P^* = 0.003$**
W Carpathians only	37/8	26/8	$\chi^2 = 0.40$, $P^* = 0.579$
Anthropogenic spread†+			
Whole range	84/4	84/22	**$\chi^2 = 10.88$, $P^* = 0.001$**
W Carpathians only	42/2	34/0	$\chi^2 = 1.59$, $P^* = 0.510$
Geology (calcareous vs. siliceous-neutral + volcanic)			
Whole range	46/37	46/35	$\chi^2 = 0.01$, $P^* = 1$
W Carpathians only	25/17	23/11	$\chi^2 = 0.53$, $P^* = 0.490$
Geology (calcareous vs. siliceous-neutral vs. volcanic)			
Whole range	46/25/12	45/31/4	$\chi^2 = 4.6$, $P^* = 0.111$
W Carpathians only	25/9/8	23/8/3	$\chi^2 = 1.59$, $P^* = 0.505$
Altitude			
Whole range	710 m (1–1950 m)	716 m (1–2269 m)	$F(1,192) = 0.008$, $P = 0.927$
W Carpathians only	845 m (161–1950 m)	970 m (251–2031 m)	$F(1,77) = 1.14$, $P = 0.289$

*The *P*-value was estimated using 2000 simulations.
†Only purely anthropogenic habitats far from any potential natural locality were considered as a distinct factor level in this analysis (see Methods).
+Number of positive and negative cases are before and after slash, respectively.

DISCUSSION

We present the first large-scale evaluation of within- and among-population cytotype diversity of the *Arabidopsis arenosa* group, an important ploidy-variable species complex from the plant model genus. Our study extends the knowledge of cytotype distribution across the range of this group particularly by: (i) expanding sampling efforts to mostly neglected regions (the Balkans, Carpathians, Baltic coast), (ii) a thorough sampling in the zone of spatial contact between cytotypes (Carpathians), and (iii) by a substantial

increase of total sample size [over 2900 individuals in total, on average 15 per population in the current dataset vs. 730 and 273 individuals, on average five and two per population, in the previous surveys by Měsíček (1970) and Schmickl *et al.* (2012), respectively]. In addition, we present an overview of genome size variation within each ploidy level of the *A. arenosa* group.

Geography correlates with ploidy level distribution, but not with homoploid DNA content variation

Globally, the diploid and tetraploid populations of the *A. arenosa* group exhibit a parapatric distribution with two zones of cytotype spatial overlap, in the Slovenian Forealps and in the Carpathians. Although tetraploids were the prevailing cytotype, the diploid cytotype spans through more than one third of the total area, which is much larger than previously assumed (see Fig. 3). In addition, the diploid populations are relatively common in some areas, occupying a variety of habitats, and in certain regions such as the Pannonian basin and the Dinaric Alps, they represent the only cytotype. This is in strong contrast to another di-tetraploid member of the genus in Central Europe, *Arabidopsis lyrata* subsp. *petraea*, which is represented by a few diploid populations isolated in cryptic Holocene refugia and by the locally more common hybridogenous tetraploid cytotype (Polatschek, 1966; Schmickl & Koch, 2011). The differentiation into diploid-dominated southern vs. tetraploid-dominated northern (partly even formerly glaciated) regions represents a common cytogeographic pattern in the European flora that most likely reflects environmental changes during past climatic oscillations (Ehrendorfer, 1980; van Dijk & Bakx-Schotman, 1997; Weiss-Schneeweiss *et al.*, 2013).

In addition, a previously unknown and ecologically distinct group of diploid populations has been found along the southern Baltic Sea coast, in the previously glaciated region at least 600 km from the closest diploid populations in the Western Carpathians. The recent introduction of these populations is not likely because the *A. arenosa* diploids generally do not show long-distance spreading in man-made habitats (see Table 2) and because the Baltic diploids exclusively occupy natural coastal sandy areas (searches in PR, PRC, W, and WU herbaria, plants from such habitats were found likely to occur from eastern Denmark to Estonia, F. Kolář, unpublished). Considering the large areas currently unfavourable for *A. arenosa* survival in northern Central Europe (forested or cultivated flatlands), a long-term isolation of the Baltic diploids from the main diploid range is probable, at least since the earlier phases of the Holocene. The presence of several geographically distinct and ecologically variable groups of diploid populations (at least two disjunct areas, with a wide range of habitats along a 0–2600 m altitudinal gradient) implies that their tetraploid derivatives, possibly combining several of the distinct diploid gene pools, should show considerable levels of genetic variation. Schmickl *et al.* (2012) did, indeed, detect large genetic variation among tetraploid populations of this group (even in the previously glaciated areas) and attributed it to the combined effects of several periglacial refugia, the absence of large bottlenecks and possible introgression from other sympatric *Arabidopsis* species. We hypothesise that recurrent origins of tetraploids from distinct gene pools and/or subsequent 2x → 4x introgression might have added another level of complexity to the *A. arenosa* group. In conclusion, the complicated cytogeographic pattern together with the most likely intricate internal sub-structuring of the species complex requires careful consideration in any ecological, genetic or genomic study employing taxa of the *A. arenosa* group as a model.

In addition to distinct ploidy levels, the plants studied here also exhibited a small but still considerable variation in genome size within each cytotype (up to 1.21-fold). The observed differences in DNA content might represent a combination of several causes of both biological and methodological origin. First, aneuploidy is usually responsible for larger abrupt differences in genome size (Roux *et al.*, 2003; Šmarda & Bureš, 2006), and it also appears to be a plausible explanation for the exceptionally high DNA content values detected in both diploid and tetraploid accessions of the *A. arenosa* group (Fig. S1). Both aneuploidy and dysploidy is not rare in Brassicaceae and may be almost a rule in certain polyploid complexes such as those of the genus *Cardamine* [*Cardamine pratensis* group, Urbanska-Worytkiewicz & Landolt, (1974), Marhold (1994), Mandáková *et al.* (2013); or *C. yezoensis* and related taxa, Marhold *et al.* (2010)]. In addition, high levels of aneuploidy were also observed in karyological analyses of *Arabidopsis* seedlings (Měsíček, 1970; M. Kolník and K. Marhold, unpublished). Second, different intensity of genomic processes, such as non-coding repetitive DNA proliferation, unequal crossing-over and illegitimate recombination, are considered major causes for gradual homoploid variation in DNA content within a species (Devos, Brown & Bennetzen, 2002; Bennetzen, Ma & Devos, 2005; Leitch & Leitch, 2013). Finally, methodological bias resulting from instrumental shifts and the influence of secondary metabolites could not be ruled out as we analysed different tissues (leaf, stem or petal) of plants that originated from ecologically distinct sites, collected in different parts of the season. Recent investigations have shown that, for instance, seasonal variation, choice of particular

instrument or isolation buffer could result in up to 10% variation in fluorescence intensities (Bainard *et al.*, 2011). However, we checked for artificial shifts by performing repeated analyses of the same accession on at least three different days, keeping the between-day variation below 3%. In addition, we demonstrated the genuine basis of the larger genome size differences by the presence of double peaks in simultaneous analyses of the individuals with distinct genome size values (which is considered to be the best evidence for true genome size differentiation, Greilhuber, 2005; Fig. 1D).

Small genuine differences in DNA contents are usually explained either as a result of neutral processes (random within- and across-population fluctuations or random accumulation of changes in spatially isolated areas/genetic lineages, Šmarda & Bureš, 2010; Oliver, McComb & Greene, 2013) or as an evolutionary constraint imposed by the surrounding environment and/or biological traits of the organism (e.g., rapid lifecycle and various traits relate to invasiveness; Greilhuber & Leitch, 2013). Our data favour the first, neutral scenario because we found no correlation of genome size in the entire *A. arenosa* group with any major geographic, altitudinal or environmental gradient. In contrast, a geography-correlated > 10% variation in genome size has been recently found among Swedish genome-sequenced accessions of *A. thaliana*, but the selective background for such variation remains unconfirmed (Long *et al.*, 2013).

High cytogeographical complexity in the Carpathians contrasts with intrapopulation cytotype uniformity

Spatial relationships between cytotypes within species can be categorised as sympatric, parapatric or allopatric, depending on whether they are geographically intermixed, adjacent or disjunct, respectively. When polyploids first arise, they necessarily occur in sympatry with their diploid progenitors. Subsequent cytotype expansion or retreat results in parapatric or allopatric distributions. Two types of ploidy contact are recognised depending on their evolutionary history (Petit, Bretagnolle & Felber, 1999): (i) primary, when polyploids arise *de novo* from local diploids/lower polyploids, and (ii) secondary, when different cytotypes regain contact after a phase of spatial separation. The *Arabidopsis arenosa* group most likely combines both scenarios at different spatio-temporal scales. The mixed diploid-triploid populations could be regarded as the primary cytotype contacts in which triploids originated recurrently via union of reduced (n) and unreduced ($2n$) gametes of the diploid. The alternative scenario, of triploid origin via inter-ploidal hybridization (favoured by Měsíček, 1970), seems improbable in light of our cytogeographic data. In all cases, only a single triploid plant was found in otherwise purely diploid populations; moreover, such populations were mostly found in exclusively diploid-inhabited areas (e.g., in the Dinaric Alps). Since the advance of large-scale ploidy screening studies enhanced by flow cytometry, the occurrence of odd cytotypes within multiple ploidy species is more the rule than the exception (Husband, Baldwin & Suda, 2013), and rare (auto)triploids have been found even in otherwise purely diploid species (Slovák *et al.*, 2009; Dušková *et al.*, 2010). Our records represent the first adult triploid individuals of *A. arenosa* detected in the wild. The extremely low frequency of adult triploids in our dataset (0.2%) in contrast with rather frequent triploid incidence in karyologically investigated seedlings (M. Kolník, unpublished) indicate strong yet still incomplete selection against the triploid progeny. Formation of viable triploid individuals in natural populations is an important prerequisite for incipient autopolyploid speciation (via triploid bridge, Husband, 2004) and thus shows important evolutionary potential for recurrent polyploidization within the *A. arenosa* group.

The two large areas of the diploid and tetraploid cytotype contact in the Carpathians and the Slovenian Forealps most likely represent secondary contact zones. This is indicated by the prevailing separate distribution of the cytotypes in the remaining areas and the intrapopulation cytotype uniformity (only one di-tetraploid population was found throughout the area studied). We will further discuss the origin and dynamics only of the sufficiently sampled zone in the Western Carpathians. This area hosts a complex landscape mosaic of spatially intermingled diploid and tetraploid populations that is in striking contrast with the within-population ploidy uniformity. Interestingly, both cytotypes occupy various substrates and climatic niches, and they occur from the lowland steppes up to high-alpine habitats. The absence of altitudinal differentiation is particularly interesting because it has been the only trend found repeatedly among the other investigated Carpathian taxa to date (Lysak & Doležel, 1998; Hodálová *et al.*, 2007; Mráz *et al.*, 2008). In addition, no general trend in cytotype-specific associations with geological substrates has been detected, although substrate specificity represents a major driver of plant spatial distributions and is also the principal speciation trigger among European mountain plants (Alvarez *et al.*, 2009; Moore & Kadereit, 2013) as well as in *Arabidopsis* (Hunter & Bomblies, 2010; Schmickl & Koch, 2011). Collectively, we argue that ecological factors appear to play a minor role in the cytotype segregation; instead, random processes such as

colonization history and genetic drift should be taken into account.

The marked prevalence of the cytotype-pure populations even within the Carpathian contact zone could be attributed to the demographic processes in the presumably strongly isolated populations. Both diploid and tetraploid populations of the *A. arenosa* group prefer open primary habitats with low competition, such as rocks, screes, sparse grasslands, and subalpine stands (Holocene cryptic refugia, Birks & Willis, 2008, see Table S1 for details on occupied habitats). In such sites isolated from each other, the processes of neutral evolution (random fluctuations in cytotype frequencies) complemented with frequency-dependent selection against the rare cytotype (i.e., minority cytotype exclusion; Levin, 1975) could have occurred, ultimately leading to cytologically pure populations even from the hypothetical ploidy-mixed populations. Such a scenario involving dynamic changes in cytotype frequencies is supported by the short lifespan of the studied plants, which have no special adaptations for long-distance dispersal and very limited clonal growth and vegetative persistence (F. Kolář, M. Lučanová, personal observation). In contrast with *Arabidopsis*, other plant systems in the Western Carpathians investigated at comparable detail exhibit frequent within-population cytotype mixtures. Nevertheless, in both cases, the plants are long-living clonal perennials either with frequent asexual reproduction (*Pilosella officinarum*, Mráz *et al*., 2008) or preferring sites under strong human impact (*Phleum pratense* agg., Perný *et al*., 2008). However, another example of the almost complete absence of cytotype-mixed populations comes from the Brassicaceae family; although diploid, tetraploid and rare hexaploid populations of perennial *Alyssum montanum* are spatially intermingled on a large scale in Central Europe, they are cytotype uniform (Španiel *et al*., 2011, 2012).

It should be noted that other evolutionary processes such as recurrent *in situ* polyploidization and/or local adaptation may also have contributed to the observed pattern in certain areas, and further detailed molecular investigations are needed. For example, the spatially isolated occurrence of tetraploids (admixed in the only ploidy-mixed population AA170) among purely diploid populations suggests a local autopolyploid origin. In summary, current evidence suggests that areas with co-occurring diploid-tetraploid *A. arenosa* represent a rather stabilized secondary contact zone, at least on a coarse spatial scale.

Large niche overlap among cytotypes

Polyploidy can have a profound effect on various morphological, anatomical and physiological plant traits that further translate into distinct ecological requirements of cytotypes (reviewed in Levin, 2002). However, the general validity of shifts in climatic niche of diploids and their polyploid relatives has been recently questioned because no correlation was found in the majority of the thoroughly investigated closely related diploid–(auto)polyploid species groups (Glennon, Ritchie & Segraves, 2014). Our results further support the latter opinion because we found mostly no association or only a weak association between ploidy level and the environment in the *Arabidopsis arenosa* group. With the exception of higher tendency of tetraploids for spreading across anthropogenic stands, both cytotypes occur virtually along the entire range of habitats occupied by the species complex. Both cytotypes could be found on calcareous and acidic substrates, and both span from lowlands to alpine habitats. The climatic niche of the cytotypes is also largely similar, with the only differences caused by spatially correlated factors, reflecting the prevailing non-overlapping distribution ranges of the cytotypes. The absence of polyploidy-linked extension of realized climatic niches has previously been suggested for *Arabidopsis*, although dramatic changes in the realized climatic niche contributed to the evolution of the whole genus (Hoffmann, 2005). In addition, no traces of selection towards the ecological separation have been found: the levels of ecological differentiation were comparable in the areas where the cytotypes co-occur (Western Carpathians) and throughout the distribution area.

Nevertheless, it should be emphasised that our study focused on the *Arabidopsis arenosa* group as a whole, and some genetic lineages with distinct ecological and/or geographical associations may be found *within* each cytotype. For example, the ecologically and partly also morphologically distinct populations on railway tracks and other secondary habitats that prevail in northern Europe (but reach as far as southern Germany and Switzerland) might represent such distinct lineages, thus explaining the observed overall preference of tetraploids for anthropogenous stands.

ACKNOWLEDGEMENTS

We are indebted to Jana Bayerová, Kateřina Černá, Ondřej Černý, Martin Hanzl, Jindřich Chrtek, Michael Jutzi, Klára Kabátová, Zuzana Khodlová, Eva Kolářová, Petr Koutecký, Jaromír Kučera, Radek Lučan, Lenka Macková, Jana Malinská, Pavol Mered'a, Clemens Pachschwöll, Jana Smatanová, Jan Suda, Milan Štech, Pavel Trávníček, Tomáš Urfus, and Judita Zozomová for help with sampling. The research was supported by the Czech Science Foundation (grant no. P506/12/0668), the Slovak Research and Development Agency (APVV; grant no. APVV-

0139-12), and by the European Social Fund and the state budget of the Czech Republic (CZ.1.07/2.3.00/30.0022 to S.Š., CZ.1.07/2.3.00/30.0037 to T.M.). Part of the calculations were performed in the Computing Centre of the Slovak Academy of Sciences using the supercomputing infrastructure acquired in projects ITMS 26230120002 and 26210120002 (Slovak infrastructure for high-performance computing) supported by the Research & Development Operational Programme funded by the ERDF. This paper was presented at the interdisciplinary symposium 'Biogeography of the Carpathians: Evolution of Biodiversity in a Spatiotemporal Context' held in September 2013 in Kraków, Poland. The authors thank the organizers of the symposium for the invitation.

REFERENCES

Alvarez N, Thiel-Egenter C, Tribsch A, Holderegger R, Manel S, Schönswetter P, Taberlet P, Brodbeck S, Gaudeul M, Gielly L, Küpfer P, Mansion G, Negrini R, Paun O, Pellecchia M, Rioux D, Schüpfer F, Van Loo M, Winkler M, Gugerli F, IntraBioDiv Consortium. 2009. History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters* **12:** 632–640.

Bainard JD, Husband BC, Baldwin SJ, Fazekas AJ, Gregory TR, Newmaster SG, Kron P. 2011. The effects of rapid desiccation on estimates of plant genome size. *Chromosome Research* **19:** 825–842.

Bennetzen JL, Ma J, Devos KM. 2005. Mechanisms of recent genome size variation in flowering plants. *Annals of Botany* **95:** 127–132.

Birks HJB, Willis KJ. 2008. Alpines, trees, and refugia in Europe. *Plant Ecology & Diversity* **1:** 147–160.

Clauss MJ, Koch MA. 2006. Poorly known relatives of *Arabidopsis thaliana*. *Trends in Plant Science* **11:** 449–459.

Devos KM, Brown JKM, Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Research* **12:** 1075–1079.

Doležel J, Bartoš JAN. 2005. Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany* **95:** 99–110.

Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* **2:** 2233–2244.

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30:** 609–628.

Dray S, Dufour AB. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* **22:** 1–20.

Duchoslav M, Šafářová L, Krahulec F. 2010. Complex distribution patterns, ecology and coexistence of ploidy levels of *Allium oleraceum* (Alliaceae) in the Czech Republic. *Annals of Botany* **105:** 719–735.

Dušková E, Kolář F, Sklenář P, Rauchová J, Kubešová M, Fér T, Suda J, Marhold K. 2010. Genome size correlates with growth form, habitat and phylogeny in the Andean genus *Lasiocephalus* (Asteraceae). *Preslia* **82:** 127–148.

Ehrendorfer F. 1980. Polyploidy and distribution. In: Lewis WH, ed. *Polyploidy. Biological relevance*. New York: Plenum Press, 45–60.

Fox J. 2003. Effect displays in R for generalised linear models. *Journal of Statistical Software* **8:** 1–27.

Galbraith DW, Harkins KR, Knapp S. 1991. Systemic endopolyploidy in *Arabidopsis thaliana*. *Plant Physiology* **96:** 985–989.

Glennon KL, Ritchie ME, Segraves KA. 2014. Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecology Letters* **17:** 574–582.

Greilhuber J. 2005. Intraspecific variation in genome size in angiosperms: identifying its existence. *Annals of Botany* **95:** 91–98.

Greilhuber J, Leitch IJ. 2013. Genome size and the phenotype. In: Greilhuber J, Doležel J, Wendel JF, eds. *Plant genome diversity, vol. 2*. Vienna: Springer, 323–344.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25:** 1965–1978.

Hodálová I, Grulich V, Horová L, Valachovič M, Marhold K. 2007. Occurrence of tetraploid and octoploid cytotypes in *Senecio jacobaea* ssp. *jacobaea* (Asteraceae) in Pannonia and the Carpathians. *Botanical Journal of the Linnean Society* **153:** 231–242.

Hoffmann MH. 2005. Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution* **59:** 1425–1436.

Hollister JD, Arnold BJ, Svedin E, Xue KS, Dilkes BP, Bomblies K. 2012. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genetics* **8:** e1003093.

Hunter B, Bomblies K. 2010. Progress and promise in using Arabidopsis to study adaptation, divergence, and speciation. *The Arabidopsis Book / American Society of Plant Biologists* **8:** e0138.

Husband BC. 2004. The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal of the Linnean Society* **82:** 537–546.

Husband BC, Baldwin SJ, Suda PJ. 2013. The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In: Greilhuber J, Doležel J, Wendel JF, eds. *Plant genome diversity, vol. 2*. Vienna: Springer, 255–276.

Jakobsson M, Hagenblad J, Tavaré S, Säll T, Halldén C, Lind-Halldén C, Nordborg M. 2006. A unique recent origin of the allotetraploid species *Arabidopsis suecica*: evidence from nuclear DNA markers. *Molecular Biology and Evolution* **23:** 1217–1231.

Jørgensen MH, Ehrich D, Schmickl R, Koch MA, Brysting AK. 2011. Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). *BMC Evolutionary Biology* **11:** 346.

Kolář F, Štech M, Trávníček P, Rauchová J, Urfus T, Vít P, Kubešová M, Suda J. 2009. Towards resolving the *Knautia arvensis* agg. (Dipsacaceae) puzzle: primary and secondary contact zones and ploidy segregation at landscape and microgeographic scales. *Annals of Botany* **103:** 963–974.

Krejčíková J, Sudová R, Lučanová M, Trávníček P, Urfus T, Vít P, Weiss-Schneeweiss H, Kolano B, Oberlander K, Dreyer LL, Suda J. 2013. High ploidy diversity and distinct patterns of cytotype distribution in a widespread species of *Oxalis* in the Greater Cape Floristic Region. *Annals of Botany* **111:** 641–649.

Kron P, Suda J, Husband BC. 2007. Applications of flow cytometry to evolutionary and population biology. *Annual Review of Ecology, Evolution, and Systematics* **38:** 847–876.

Leitch IJ, Leitch AR. 2013. Genome size diversity and evolution in land plants. In: Greilhuber J, Doležel J, Wendel JF, eds. *Plant genome diversity, vol. 2*. Vienna: Springer, 307–322.

Levin DA. 1975. Minority cytotype exclusion in local plant populations. *Taxon* **24:** 35–43.

Levin DA. 2002. *The role of chromosomal change in plant evolution*. Oxford: Oxford University Press.

Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, Platzer A, Zhang Q, Vilhjálmsson BJ, Korte A, Nizhynska V, Voronin V, Korte P, Sedman L, Mandáková T, Lysak MA, Seren Ü, Hellmann I, Nordborg M. 2013. Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics* **45:** 884–890.

Lysak MA, Doležel J. 1998. Estimation of nuclear DNA content in *Sesleria* (Poaceae). *Caryologia* **51:** 123–132.

Mandáková T, Kovařík A, Zozomová-Lihová J, Shimizu-Inatsugi R, Shimizu KK, Mummenhoff K, Marhold K, Lysak MA. 2013. The more the merrier: recent hybridization and polyploidy in *Cardamine*. *The Plant Cell* **25:** 3280–3295.

Mandáková T, Marhold K, Lysak MA. 2014. The widespread crucifer species *Cardamine flexuosa* is an allotetraploid with a conserved subgenomic structure. *New Phytologist* **201:** 982–992.

Marhold K. 1994. Chromosome numbers of the genus *Cardamine* L. (Cruciferae) in the Carpathians and Pannonia. *Phyton (Horn)* **34:** 19–41.

Marhold K, Kudoh H, Pak JH, Watanabe K, Španiel S, Lihová J. 2010. Cytotype diversity and genome size variation in eastern Asian polyploid *Cardamine* (Brassicaceae) species. *Annals of Botany* **105:** 249–264.

Měsíček J. 1970. Chromosome counts in *Cardaminopsis arenosa* agg.(Cruciferae). *Preslia* **42:** 225–248.

Měsíček J. 1998. *Cardaminopsis*. In: Marhold K, Hindák F, eds. *Zoznam nižších a vyšších rastlin Slovenska – checklist of non-vascular and vascular plants of Slovakia*. Bratislava: Veda, 395–396.

Měsíček J, Goliašová K. 2002. *Cardaminopsis* (C. A. Mey.) Hayek. In: Goliašová K, Šípošová H, eds. *Flóra Slovenska*. Bratislava: Veda, 388–415.

Moore AJ, Kadereit JW. 2013. The evolution of substrate differentiation in *Minuartia* series *Laricifoliae* (Caryophyllaceae) in the European Alps: in situ origin or repeated colonization? *American Journal of Botany* **100:** 2412–2425.

Mráz P, Šingliarová B, Urfus T, Krahulec F. 2008. Cytogeography of *Pilosella officinarum* (Compositae): altitudinal and longitudinal differences in ploidy level distribution in the Czech Republic and Slovakia and the general pattern in Europe. *Annals of Botany* **101:** 59–71.

Newson R. 2006. Confidence intervals for rank statistics: Somers' D and extensions. *Stata Journal* **6:** 309–334.

Oliver KR, McComb JA, Greene WK. 2013. Transposable elements: powerful contributors to angiosperm evolution and diversity. *Genome Biology and Evolution* **5:** 1886–1901.

Perný M, Kolarčik V, Majeský L, Mártonfi P. 2008. Cytogeography of the *Phleum pratense* group (Poaceae) in the Carpathians and Pannonia. *Botanical Journal of the Linnean Society* **157:** 475–485.

Petit C, Bretagnolle F, Felber F. 1999. Evolutionary consequences of diploid–polyploid hybrid zones in wild species. *Trends in Ecology & Evolution* **14:** 306–311.

Polatschek A. 1966. Cytotaxonomische Beiträge zur Flora der Ostalpenländer, I. *Österreichischen Botanischen Zeitschrift* **113:** 1–46.

R Development Core Team. 2013. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Roux N, Toloza A, Radecki Z, Zapata-Arias FJ, Doležel J. 2003. Rapid detection of aneuploidy in *Musa* using flow cytometry. *Plant Cell Reports* **21:** 483–490.

Schmickl R, Koch MA. 2011. *Arabidopsis* hybrid speciation processes. *Proceedings of the National Academy of Sciences USA* **108:** 14192–14197.

Schmickl R, Paule J, Klein J, Marhold K, Koch MA. 2012. The evolutionary history of the *Arabidopsis arenosa* complex: diverse tetraploids mask the western Carpathian center of species and genetic diversity. *PLoS ONE* **7:** e42691.

Shimizu-Inatsugi R, Lihová J, Iwanaga H, Kudoh H, Marhold K, Savolainen O, Watanabe K, Yakubov VV, Shimizu KK. 2009. The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. *Molecular Ecology* **18:** 4024–4048.

Slovák M, Vít P, Urfus T, Suda J. 2009. Complex pattern of genome size variation in a polymorphic member of the Asteraceae. *Journal of Biogeography* **36:** 372–384.

Šmarda P, Bureš P. 2006. Intraspecific DNA content variability in Festuca pallens on different geographical scales and ploidy levels. *Annals of Botany* **98:** 665–678.

Šmarda P, Bureš P. 2010. Understanding intraspecific variation in genome size in plants. *Preslia* **82:** 41–61.

Španiel S, Marhold K, Filová B, Zozomová-Lihová J. 2011. Genetic and morphological variation in the diploid–polyploid *Alyssum montanum* in Central Europe: taxonomic and evolutionary considerations. *Plant Systematics and Evolution* **294:** 1–25.

Španiel S, Marhold K, Thiv M, Zozomová-Lihová J. 2012. A new circumscription of *Alyssum montanum* ssp. *montanum* and *A. montanum* ssp. *gmelinii* (Brassicaceae) in Central Europe: molecular and morphological evidence. *Botanical Journal of the Linnean Society* **169:** 378–402.

Suda J, Krahulcová A, Trávníček P, Krahulec F. 2006. Ploidy level versus DNA ploidy level: an appeal for consistent terminology. *Taxon* **55:** 447–450.

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

Trávníček P, Dočkalová Z, Rosenbaumová R, Kubátová B, Szeląg Z, Chrtek J. 2011b. Bridging global and microregional scales: ploidy distribution in *Pilosella echioides* (Asteraceae) in central Europe. *Annals of Botany* **107:** 443–454.

Trávníček P, Eliášová A, Suda J. 2010. The distribution of cytotypes of *Vicia cracca* in Central Europe: the changes that have occurred over the last four decades. *Preslia* **82:** 149–163.

Trávníček P, Kubátová B, Čurn V, Rauchová J, Krajníková E, Jersáková J, Suda J. 2011a. Remarkable coexistence of multiple cytotypes of the *Gymnadenia conopsea* aggregate (the fragrant orchid): evidence from flow cytometry. *Annals of Botany* **107:** 77–87.

Urbanska-Worytkiewicz K, Landolt E. 1974. Biosystematic investigations in *Cardamine pratensis* L. s.l. 1. Diploid taxa from Central Europe and their fertility relationships. *Berichte des Geobotanischen Institutes der Eidgenössischen Technischen Hochschule, Stiftung Rübel* **42:** 42–139.

Van Dijk P, Bakx-Schotman T. 1997. Chloroplast DNA phylogeography and cytotype geography in autopolyploid *Plantago media*. *Molecular Ecology* **6:** 345–352.

Venables WN, Ripley BD. 2002. *Modern applied statistics with S*, 4th edn. New York: Springer.

Weiss-Schneeweiss H, Emadzade K, Jang TS, Schneeweiss GM. 2013. Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenetic and Genome Research* **140:** 137–150.

Yant L, Hollister JD, Wright KM, Arnold BJ, Higgins JD, Franklin FCH, Bomblies K. 2013. Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Current Biology* **23:** 2151–2156.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Locality details on sampled populations of the *Arabidopsis arenosa* group.
Table S2. Full list of spatial and environmental variables.
Table S3. List of spatial and environmental predictors used in modelling of cytotype distribution.
Figure S1. Distribution of nuclear DNA content values.
Figure S2. Significant partial relationships between cytotype distribution and its predictors in populations across the whole area of the *Arabidopsis arenosa* group.
Figure S3. Relationship between cytotype distribution and latitude in the contact zone in Western Carpathians.

Paper II

Kolář F, **Fuxová G**, Záveská E, Nagano AJ, Hyklová L, Lučanová M, Kudoh H, Marhold K (2016) Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. Mol Ecol, 25: 3929-3949.

MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 3929–3949 doi: 10.1111/mec.13721

Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*

FILIP KOLÁŘ,*†‡[1] GABRIELA FUXOVÁ,†[1] ELIŠKA ZÁVESKÁ,§[1] ATSUSHI J. NAGANO,¶**†† LUCIE HYKLOVÁ,† MAGDALENA LUČANOVÁ,†‡ HIROSHI KUDOH¶ and KAROL MARHOLD†‡‡

*Natural History Museum, University of Oslo, PO Box 1172 Blindern, Oslo NO-0318, Norway, †Department of Botany, Faculty of Science, Charles University in Prague, Prague CZ-128 01, Czech Republic, ‡Institute of Botany, The Czech Academy of Sciences, Průhonice CZ-252 43, Czech Republic, §Institute of Botany, University of Innsbruck, Innsbruck AT-6020, Austria, ¶Center for Ecological Research, Kyoto University, Kyoto JP-520-2113, Japan, **Faculty of Agriculture, Ryukoku University, Shiga JP-612-8577, Japan, ††JST PRESTO, Saitama JP-332-0012, Japan, ‡‡Institute of Botany, Slovak Academy of Sciences, Bratislava SK-845 23, Slovak Republic

Abstract

Quaternary climatic oscillations profoundly impacted temperate biodiversity. For many diverse yet undersampled areas, however, the consequences of this impact are still poorly known. In Europe, particular uncertainty surrounds the role of Balkans, a major hotspot of European diversity, in postglacial recolonization of more northerly areas, and the Carpathians, a debatable candidate for a northern 'cryptic' glacial refugium. Using genome-wide SNPs and microsatellites, we examined how the interplay of historical processes and niche shifts structured genetic diversity of diploid *Arabidopsis arenosa*, a little-known member of the plant model genus that occupies a wide niche range from sea level to alpine peaks across eastern temperate Europe. While the northern Balkans hosted one isolated endemic lineage, most of the genetic diversity was concentrated further north in the Pannonian Basin and the Carpathians, where it likely survived the last glaciation in northern refugia. Finally, a distinct postglacial environment in northern Europe was colonized by populations of admixed origin from the two Carpathian lineages. Niche differentiation along altitude-related bioclimatic gradients was the main trend in the phylogeny of *A. arenosa*. The most prominent niche shifts, however, characterized genetically only slightly divergent populations that expanded into narrowly defined alpine and northern coastal postglacial environments. Our study highlights the role of eastern central European mountains not only as refugia for unique temperate diversity but also sources for postglacial expansion into novel high-altitude and high-latitude niches. Knowledge of distinct genetic substructure of diploid *A. arenosa* also opens new opportunities for follow-up studies of this emerging model of evolutionary biology.

Keywords: approximate Bayesian computation, Arabidopsis, niche differentiation, phylogeography, RADseq

Received 4 December 2015; revision received 25 May 2016; accepted 1 June 2016

Introduction

The glacial–interglacial cycles of the Quaternary greatly influenced patterning of temperate diversity around the globe (e.g. Hewitt 2000, 2004; Qiu *et al.* 2011). Our

Correspondence: Filip Kolář, Fax: +47 22 85 1835;
E-mail: filip.kolar@gmail.com
[1]Equal contribution.

current knowledge about the impact of Quaternary historical processes on structuring intraspecific genetic diversity is, however, skewed towards a few well-studied areas. In Europe, a wealth of information has been accumulated particularly for the flora and fauna of alpine and arctic habitats (e.g. Schönswetter *et al.* 2005; Thiel-Egenter *et al.* 2011; Eidesen *et al.* 2013; Nägele & Hausdorf 2015) and southern European refugia (Weiss & Ferrand 2007; Nieto Feliner 2014), based on traditional high-resolution molecular markers such as organellar DNA, AFLPs and/or microsatellites. The typical historical scenario for temperate European biodiversity involves glacial survival in southern European refugia (the Iberian, Apennine and Balkan Peninsulas) and postglacial recolonization of the formerly periglacial (central Europe) and glaciated areas (northern Europe and central European mountain ranges), following the northwards-retreating ice sheet (Taberlet *et al.* 1998; Hewitt 2004; Tzedakis *et al.* 2013). Nevertheless, we are still far from the complete picture on historical evolution of European biota, mainly due to lack of comparable data from more easterly regions, leaving remarkable controversies about the role of several key regions of the (south-)eastern Europe in glacial survival and postglacial recolonization. In particular, while the Balkan Peninsula is undoubtedly a hotspot of European diversity both at species and intraspecific levels (Griffiths *et al.* 2004) and a 'classical' glacial refugium, it is still uncertain to which extent, it served as a source of postglacial recolonization of temperate Europe. While this scenario is valid for some species (e.g. Taberlet *et al.* 1998; Hewitt 1999; Magri *et al.* 2006; Havrdová *et al.* 2015), other studies based on dense sampling in the area and/or high marker resolution showed the Balkans may represent rather a harbour of its endemic diversity while populations from central, eastern and northern Europe experienced distinct evolutionary history (reviewed in Schmitt & Varga 2012). The newly emerging alternative scenario suggests glacial survival of temperate elements in more northerly areas (that were closer to the continental ice sheets) such as the Carpathian mountains, which were only scarcely glaciated (Ronikier 2011) and hosted forest communities through last glacial maximum, LGM (Jankovská & Pokorný 2008). Although the Carpathians are a hot candidate for 'cryptic' northern refugium (Stewart & Lister 2001; Willis & van Andel 2004; Provan & Bennett 2008), recently supported by directly dated fossils of temperate species (Juřičková *et al.* 2014), existence of such northen LGM refugia is still controversial in general (Tzedakis *et al.* 2013). Importantly, the contribution of phylogeography to this debate is still scarce and largely biased towards animal examples (Babik *et al.* 2004; Kotlík *et al.* 2006; Fijarczyk *et al.* 2011; Zieliński *et al.* 2013; Wielstra *et al.* 2015). In plants, some indications of Carpathian LGM survival come from alpine (reviewed in Ronikier 2011) and montane plants (Magri *et al.* 2006; Těšitel *et al.* 2009), but no study so far addressed this question using a system spanning over wide altitudinal range, which can provide additional important clues to the role of niche shifts in glacial survival and recolonization.

To address the role of historical processes and niche shifts in the evolution of eastern European flora, we applied genome-wide single-nucleotide polymorphism (SNP) markers, for the first time in a plant system from this area, in a rangewide study of diploid cytotype of *Arabidopsis arenosa*. This still poorly known member of *Arabidopsis*, remarkable for the natural occurrence of diploid and autotetraploid populations and striking ecological diversity, represents an emerging model system for understanding evolution through genome duplication and local adaptation (Yant *et al.* 2013; Wright *et al.* 2015). Taxonomic treatment of *A. arenosa* is still controversial, varying from recognizing a single species by most of the evolutionary and experimental studies (e.g. Yant *et al.* 2013; Arnold *et al.* 2015; Baduel *et al.* 2016) up to nine taxa (species or subspecies, partly not formally described) by systematically oriented and/or local studies (e.g. Měsíček & Goliašová 2002; reviewed by Schmickl *et al.* 2012). As our results indicate monophyly of the whole group (see also Hohmann *et al.* 2014) but contradict the traditional delimitations of the internal taxa, we will consistently refer to all populations as *A. arenosa*, for the sake of simplicity, and we will return to the species concept problem in the final part of the Discussion. Native range of the diploid *A. arenosa* cytotype covers most of the eastern temperate Europe, spanning a 2500-m altitudinal gradient from coastal habitats to alpine stands and a 1600-km latitudinal gradient from the submediterranean Balkan Peninsula to previously glaciated Baltic Sea coast (Kolář *et al.* 2015a). In contrast to other *Arabidopsis* species, little is still known about the rangewide patterns of genetic variation and evolutionary history of *A. arenosa*, particularly of its diploid cytotype. The single rangewide genetic study available to date, based on AFLP and plastid DNA markers, identified areas with higher genetic diversity but failed to discriminate main lineages within the group (Schmickl *et al.* 2012). Further phylogeographical inference using plastid data was hampered by extensive haplotype sharing among *Arabidopsis* species and also among regions within *A. arenosa* (Clauss & Koch 2006; Koch & Matschinger 2007; Schmickl *et al.* 2012). A recent successful application of restriction-associated DNA (RADseq) markers revealed a single autopolyploid origin of the widespread tetraploid populations (Arnold *et al.* 2015). Limited sampling focused on tetraploids (20 populations in

total, only six diploid), however, provided limited information on the genetic structure of the diploid cytotype. In fact, although tetraploid *Arabidopsis arenosa* has been subjected to molecular genetic studies for the last two decades (e.g. Kamm *et al.* 1995; Comai *et al.* 2000; Madlung *et al.* 2002), the diploid cytotype has been mostly neglected. Only very recently, a study of Wright *et al.* (2015) revealed selection in diploids on meiosis-related genes driven by several processes, one of which was possibly related to environmental temperature. This surprising outcome demonstrates the need for a rangewide assessment of genetic structure and niche diversity in *A. arenosa*, which may in turn allow addressing further evolutionary questions of general significance.

Here, we examine the genetic structure of diploid *Arabidopsis arenosa* across its entire distributional range and test for genetic correlates within its wide ecological niche. First, using genome-wide SNPs and nuclear microsatellites, we reconstruct phylogenetic relationships, reveal rangewide patterns of genetic diversity and infer the evolutionary history of the diploid cytotype. Specifically, we ask whether Balkan Peninsula and/or the Carpathians acted as glacial refugia for the species and which area provided postglacial recolonizers for the species' northernmost disjunct outpost in the formerly glaciated Baltic Sea coast. Then, we ask whether this genetic structure corresponds with major ecological gradients across sites occupied by diploid *A. arenosa* and interpret the observed discrepancies in the light of the evolutionary history of the whole group. Finally, based on our findings, we outline further prospects for integrative experimental and ecological studies in *Arabidopsis arenosa*.

Materials and methods

We collected leaf material of ~10 individuals from 64 natural populations of diploid *Arabidopis arenosa* in 2011–2013 across the entire range of its diploid cytotype and checked ploidy level of all collected individuals using flow cytometry, as described in Kolář *et al.* (2015a) (for locality details see Table S1, Supporting information). Two populations (three individuals each) of the closely related (Hohmann *et al.* 2014) diploid species *A. croatica* were also collected for SNP genotyping. To efficiently screen for both intra- and interpopulation genetic variation across the extensive rangewide sampling, we employed two multilocus markers providing complementary information at different levels of sampling: (i) genome-wide single-nucleotide polymorphisms (SNPs) for reconstructing among-population and among-lineage relationships (1–4 individuals per population, two on average) and (ii) microsatellites for inferring intrapopulation genetic diversity parameters and among-population differentiation (4–10 individuals per population, nine on average).

SNP genotyping

A subset of 177 individuals was genotyped for SNPs using double-digest RADseq (Peterson *et al.* 2012). Genomic DNA was digested by two restriction enzymes, *Bgl* II and *Nde* I, and the corresponding adapters were simultaneously ligated at 37 °C for 16 h. The reaction mixture consisted of 20 ng of genomic DNA, 5 units of *Bgl* II (NEB), 5 units of *Nde* I (NEB), 1× NEB buffer2 (NEB), 1× BSA (NEB), 0.2 microM *Bgl* II adapter, 0.2 microM *Nde* I adapter, 1 mM ATP (Takara) and 300 units of T4 DNA ligase (Enzymatics). The ligation product was purified by the AMpureXP (Beckman Coulter) according to the manufacturer's instructions. One-tenth of the purified DNA was used in the PCR enrichment with the KAPA HiFi HS ReadyMix (KAPA biosystems) (see Appendix S6, Supporting information for sequences of adaptors and primers used). Approximately 350-bp fragment of the PCR product was selected by E-Gel size select 2% (Life technologies). Single-end 50-bp and index sequence of the library was sequenced by HiSeq2500 (Illumina) with the TRUSEQ v3 chemistry.

Raw data processing, variant calling and filtration

Raw reads were demultiplexed, quality trimmed (>30 Phred quality score) and mapped using STAMPY version 1.0.23 (Lunter & Goodson 2011) on a repeat-masked genome of *Arabidopsis lyrata* v. 1.0.25 (Hu *et al.* 2011). Postmapping alignment processing was performed using Picard Tools. The GENOME ANALYSIS TOOL KIT v3.3.0 (GATK) (McKenna *et al.* 2010) was used for realignment around indels (*IndelRealigner* tool) and for simultaneous SNP discovery and genotyping (*HaplotypeCaller* and *GenotypeGVCFs*) following the recommended best practice, accepting only genotypes with confidence score higher than the calling threshold of the corresponding variant site (www.broadinstitute.org/gatk/). GATK performs SNP discovery and probabilistic genotype calling across all samples simultaneously, which is a more accurate than individual-based SNP calling (Nielsen *et al.* 2011). Using GATK (*VariantFiltration* and *SelectVariants*) and VCFTOOLS v0.1.14 (Danecek *et al.* 2011), we retained only bi-allelic sites that mapped to nuclear chromosome scaffolds with a minimum mapping quality of 40, which did not show mapping quality bias for the reads supporting the nonreference allele (ensured by keeping only variants with mapping quality rank sum test value above −12.5) and which were present in at least 50% of our individuals at a

sequencing depth of 8× or greater. In addition, we excluded potentially paralogous sites by excluding regions in which eight diploid whole-genome-sequenced *A. arenosa* individuals (Yant *et al.* 2013) were heterozygous in more than two positions within a < 2-kb region (following Arnold *et al.* 2015). We considered unlikely for all eight diploids from two distinct populations to be heterozygous at three or more sites within a gene or intergenic segment according to Hardy–Weinberg equilibrium. Finally, we also removed all variable sites with allele, which was uniform across our *A. arenosa*/*A. croatica* sample but was different from that in the *A. lyrata* reference. To reduce linkage among SNPs in our data set for STRUCTURE, SNAPP and TREEMIX analyses, we randomly selected one SNP per each 50-bp region corresponding to one RADlocus; for STRUCTURE analyses, we also removed alleles present only once in the entire data set (singletons).

Exploratory analyses of SNP data

We determined the optimal grouping of the populations using Bayesian clustering in STRUCTURE v2.3.2 (Pritchard *et al.* 2000). The analyses were performed separately for (i) the entire data set of *A. arenosa* (*A. croatica* excluded; 2313 SNPs, one random SNP per RAD locus, 11% of missing data), (ii) genetically close Carpathian + Baltic populations; 2323 SNPs, one random SNP per RAD locus, 11% of missing data) and (iii), in order to identify finer substructuring of the data, also separately for each of the groups identified by the analysis of the entire data set (1515–2137 SNPs, one random SNP per RAD locus, 11–12% of missing data in all data sets). The admixture model with uncorrelated allele frequencies was used. Ten replicate runs for *K* (number of groups) ranging from 1 to 10 were carried out using a burn-in of 100 000 iterations followed by 1 000 000 additional MCMC iterations. We identified the optimal number of groups as the value of *K* where the increase in likelihood started to flatten out, the result of replicate runs was similar, and the clusters were nonempty. Additionally, we employed the delta *K* criterion reflecting the differences in likelihood of runs at different *K* (Evanno *et al.* 2005). Further, we displayed genetic distances among individuals using principal coordinate analysis based on Euclidean distance (PCoA, replacing the missing values, 12% in total, by average allele frequency for that locus, in total 10 955 SNPs for entire data set and 8949 SNPs for the Carpathian + Baltic populations) calculated in R package ADEGENET v1.4-2 (Jombart 2008). As PCoA might be sensitive to handling the problems of missing data, we also analysed the same sets of individuals for SNPs that were present in min 90% individuals (3179 SNPs in the complete data set and 2468 SNPs in the Carpathian + Baltic data set, 2% missing data in both cases), but the pattern along the major three axes remained stable (not shown).

We estimated a species tree of the major genetic groups in a multispecies coalescent framework using SNAPP v2.2.0 (Bryant *et al.* 2012). To check for the effects of the a priori group delimitation on the tree topologies, we analysed the same data sets assigned into four (*A. croatica* and three genetically most distant groups of *A. arenosa* as identified by PCoA: Pannonian, Dinaric and Carpathian + Baltic groups) and six (the latter group further subdivided into W Carpathian, SE Carpathian and Baltic groups) groups, respectively. Due to large computational demands of the program, we analysed a subsample comprising one randomly selected individual per each nonadmixed (according to STRUCTURE) population, except for two samples per *A. croatica* population (53 individuals in total). Two different subsamples comprising the same populations but different individuals were analysed to check for consistency (2313 and 2213 SNPs, respectively, <5% of missing data). We initially ran two analyses with different theta priors to allow for different current and ancestral population sizes: (i) mean theta prior of 0.043 (corresponding with previous estimates for tetraploid *A. arenosa*, Hollister *et al.* 2012) (gamma distribution, alpha = 1.5, beta = 35) and (ii) mean theta prior = 0.1 (gamma distribution, alpha = 12, beta = 110 prior for large population sizes); the remaining parameters were left at defaults. Analysis with different priors produced the same topology, but the latter lead to higher likelihood and posterior (not shown), we thus further report the estimates using the latter prior settings. The analyses were checked for convergence using TRACER v1.6, making sure that Bayesian runs reached an effective sample size >200 after burn-in. We visualized the posterior distribution of species trees using DENSITREE v2.2.0. Finally, because of lack of reliable calibrations, we recalculated the estimated divergence times by mutation rate estimated for *A. arenosa*: 3.7×10^{-8} substitutions/site/generation (Arnold *et al.* 2015). As estimates of divergence times without external calibration should be interpreted with a caution, we strictly limit our interpretations onto rejecting the very recent (Holocene) divergence of the major lineages.

Finally, we searched for admixture among the five major groups of diploid *A. arenosa* using TREEMIX v1.12 (Pickrell & Pritchard 2012). Considering the groups as populations, we constructed a maximum-likelihood population graph from allelic frequencies of 2413 loci (<5% of missing data) and allowed for one migration edge to see the principal admixture event among the five groups. The trees were bootstrapped by 1000 replicates.

Evolutionary hypothesis testing

The mode of origin of Baltic populations from the W Carpathian and SE Carpathian groups was inferred in a coalescence framework using approximate Bayesian computation (ABC, Beaumont 2010) calculated in DIYABC v2.1 (Cornuet *et al.* 2014). The three groups were treated as populations to force the coalescence of individuals within each lineage. The data set comprised 2487 SNPs (<5% of missing data) and 52 individuals: we included all ten Baltic individuals and a subset of 24 and 18 individuals from the W and SE Carpathian groups, respectively, that were genetically and geographically closest to the Baltic ones (see Appendix S6, Supporting information for details). Three competing scenarios were compared: the Baltic group was modelled either as originating from an admixture of both Carpathian groups (scenario 1) or splitting from the W or SE Carpathian group (scenarios 2 and 3, respectively). The time of origin of the Baltic group was set as postdating the divergence of the two Carpathian lineages; population size changes were allowed for each population (for detailed prior settings, see Appendix S6, Supporting information). A total of 500 000 data sets were generated for each scenario; that is, 1 500 000 simulations were performed in total. The scenarios were compared using two approaches: one by directly counting the frequency of the various scenarios among the most similar simulated data sets (direct estimate approach; Miller *et al.* 2005) and one by doing a logistic regression of each scenario probability for the most similar simulated data sets on the deviations between simulated and observed summary statistics (Fagundes *et al.* 2007). In these two comparisons, 0.1% and 1% simulated data closest to the observed values were used, respectively. Finally, we evaluated the confidence of our scenario choice by simulating 1000 pseudo-observed data sets drawn from parameter prior distribution (replacing original summary statistics by discriminant scores of a linear discriminant analysis, Estoup *et al.* 2012) under two scenarios alternative to the selected scenario 1 and measured the proportions of times our selected scenario 1 had the highest posterior probability. Summing up these values provided estimate of a type II error, that is the probability of deciding for the preferred scenario when it is not true.

Microsatellite data analyses

To assess the levels of intrapopulation diversity and interpopulation differentiation, we genotyped 14 unlinked microsatellite loci in 570 individuals in the same 64 *A. arenosa* populations as in those used for SNP genotyping. The loci were previously employed in population genetic studies of *A. arenosa* and *A. lyrata* (Clauss *et al.* 2002; Schmickl & Koch 2011) (see Appendix S6, Supporting information for details on amplification protocol).

For each population, we calculated observed (Ho) and expected heterozygosity (He), Nei's unbiased estimator for gene diversity (Hs, equation 7.39 of Nei 1987) and average number of alleles (allelic richness, computed through rarefaction on the small sample size of minimum seven individuals, 1000 permutations) in MSA v4.05 (Dieringer & Schlötterer 2003). The inbreeding coefficient (F_{IS}) was inferred simultaneously with estimating frequency of null alleles in a Bayesian framework in INEST v2.0 (Chybicki & Burczyk 2009). For each population of at least nine individuals, we calculated posterior distributions of F_{IS} based on an individual inbreeding model by performing 500 000 MCMC iterations, sampling every 1000th generation and discarding first 10% of generations as a burn-in). Finally, we calculated a frequency-downweighted marker index (DW; Schönswetter & Tribsch 2005), hereafter termed also 'rarity', on a presence–absence matrix of alleles using the R script *AFLPdat* (Ehrich 2006). The DW value is expected to be higher in populations that harbour a high number of rare alleles, that is alleles with low frequency in the total data set, generally indicating long-term local persistence of such populations in contrast to recent immigration and/or long-distance dispersal (e.g. Paun *et al.* 2008). Differences in diversity and DW indices among the major genetic groups, as identified by the SNP data, were tested by one-way ANOVA (R package *stats*) and, separately, between ecologically divergent foothill and high-altitude populations of the W Carpathian group by a permutation two-sample test (999 permutations, R package *perm*).

We inferred the differentiation among populations in the entire data set and within each of the major genetic groups separately using the fixation index F_{ST} (Weir & Cockerham 1984) and its standardized extension for multi-allelic states (G_{ST}, Hedrick 2005) calculated in MSA. In addition, we also quantified the partitioning of genetic variation within and among populations in the entire data set as well as within each of the major genetic groups by analysis of molecular variance (AMOVA) calculated in *pegas* v0.8 (Paradis 2010). Hierarchical AMOVA was used to estimate the levels of genetic differentiation among the major groups. Finally, we tested for a significant correlation among matrices of genetic (Nei 1972) and Euclidean geographical distances among populations (isolation by distance) using a Mantel test in *adegenet*. We performed the same set of tests also for populations of the W Carpathian group alone, comparing ecologically highly divergent populations from high-altitude subalpine vs. 'normal' foothill

habitats (see Table S1, Supporting information for their delimitation).

Niche differentiation

We searched for concordance among genetic structure of the SNP-genotyped populations and environmental conditions of their original locations. We inferred among-population genetic distances from SNP data (Nei 1972) in *adegenet* and calculated Euclidean ecological distances for the same populations from parameters that were recorded either in situ during collections (calcareous vs. neutral/siliceous substrates, altitude) or derived from Worldclim data (Hijmans *et al.* 2005), that is climatic factors (19 bioclimatic variables) and topography (average slope inclination in the ~1 km radius). First, we tested for the overall correlation among the matrices of genetic and ecological distances based on the complete data set as well as on each of the major genetic groups separately using Mantel test (999 permutations, function *mantel.randtest* from package ADE4 v1.6-2 Dray & Dufour 2007). In order to further explore contribution of individual environmental predictors, we performed constrained analysis of principal coordinates (CAP), which is a multivariate extension of multiple regression based on distances between objects (Anderson & Willis 2003). We calculated the CAP models on the matrix of genetic distances by *capscale* procedure in *vegan 2.3-0* (Oksanen *et al.* 2013) and tested (500 permutations) for effects of altitude, substrate preferences, topography (slope inclination) and two composite climatic parameters (precipitation and temperature). These latter two parameters were represented by scores on first axis of a separate unconstrained ordination (principal component analysis) of the eleven temperature-linked and eight precipitation-linked bioclimatic variables, respectively. Along with addressing a complete model encompassing all five environmental variables, we also estimated the maximum amount of genetic variation that could be attributed to marginal effects and unique contributions of each environmental variable (considering them either as a single constraint or after partialling out the effects of other variables, respectively) in separate CAP analyses. We performed this set of CAP analyses for the entire data set as well as separately for populations belonging to each of the four major genetic groups (Baltic group was not analysed due to only four genotyped populations).

We further investigated the role of principal gradients in ecological differentiation among the major genetic groups using a larger set of 117 sites occupied by diploid *A. arenosa* (termed here 'ecologically screened populations'). All these populations have been sampled, georeferenced and cytotyped by means of flow cytometry by our team in years 2011–2015 (Kolář *et al.* 2015a and additional sampling, see the complete list in Table S1, Supporting information). The nongenotyped populations were assigned to the same major genetic group as their closest SNP-genotyped counterpart (see Table S1, Supporting information), what was feasible due to mostly allopatric distribution of the major genetic lineages (Fig. 1). First, we displayed the principal trends in variation in the bioclimatic variables using principal component analysis (*prcomp* function in R *stats*). Second, using linear and classificatory discriminant analyses, we tested for the differences in climatic niche preferences of populations assigned to the major genetic groups as they were identified by the SNP data. We ran the linear and classificatory discriminant analyses in R using functions *cca* of *vegan* and *lda* of MASS v7.3-29 (Venables & Ripley 2002), respectively, both wrapped in MORPHOTOOLS v1.01 (Koutecký 2014) based on eight uncorrelated bioclimatic variables, which passed the stepwise forward selection (calculated by *ordistep* function in *vegan*). Third, we used contingency tables to examine the differences among the groups in geological substrate preferences. Finally, we applied multinomial logistic regression calculated in NNET v7.3-9 (Venables & Ripley 2002) to compare the occurrence of the genetic groups in areas with different topography (slope inclination). We did not condition for effects of spatial autocorrelations in our models, as we addressed differentiation in the observed (realized) niches of the mostly allopatric lineages, which are by their nature determined by spatially correlated environmental conditions. We, however, separately tested for the correlation among matrices of the same, SNP-based, genetic and geographical distances (isolation by distance) using the Mantel test. All analyses were performed in R v3.0.2.

Results

In total, 241 M Illumina reads passing the quality threshold (on average 1.36 M per individual) were used for mapping and variant calling, yielding ~2500 variable RAD loci passing our filtering criteria, with average coverage of 51× per site per sample. The total number of SNPs used in the analyses varied from 1515 to 10 955 depending on the data set analysed (entire data set vs. subgroups) and whether random thinning to one SNP per RAD locus was applied.

Grouping of populations

First, we identified major genetic grouping of the 64 *A. arenosa* populations sampled rangewide (see Table S1, Supporting information for locality details) using the SNP data. Bayesian clustering using STRUCTURE

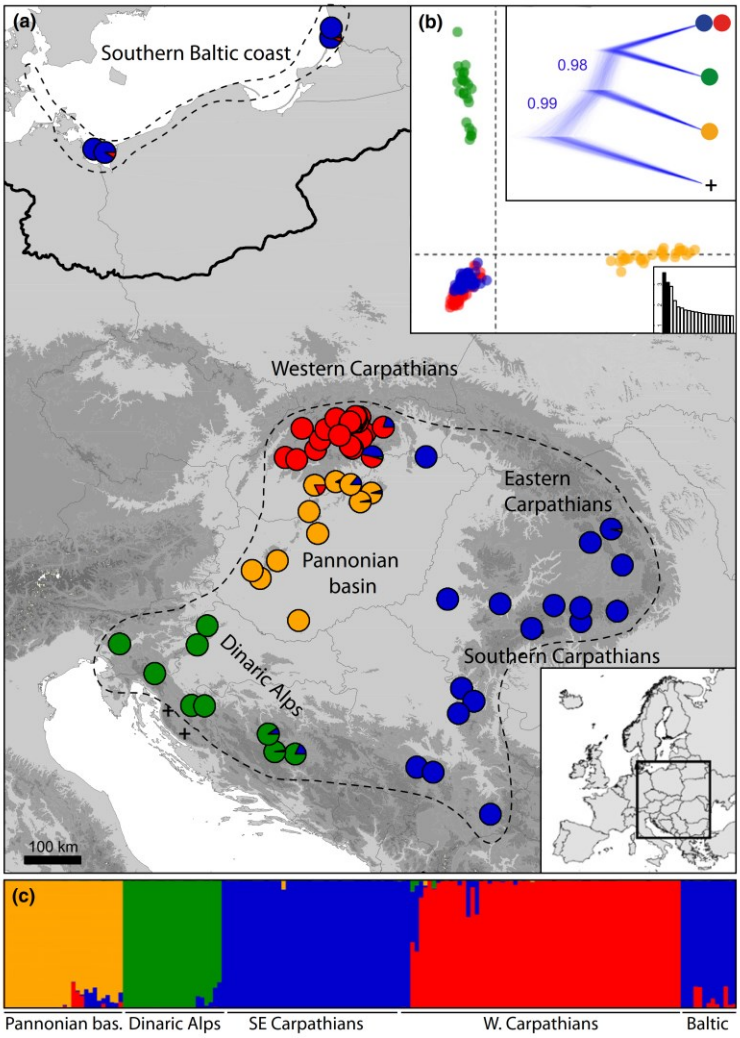


Fig. 1 Rangewide genetic differentiation of diploid *Arabidopsis arenosa*, (a) geographical distribution of sampled *A. arenosa* populations (pie charts reflecting proportional assignment to a particular STRUCTURE group) and *A. croatica* (black cross), bold line indicates maximal extent of the continental ice sheet during the last glaciation, dashed line denotes borders of the distribution range of the diploid cytotype; (b) principal coordinate analysis of the *A. arenosa* individuals (10 955 SNPs, first and second axis displayed, histogram shows proportional contribution in explaining variance by the first 20 axes) and species tree of the three most distinct *A. arenosa* groups inferred under multispecies coalescent analysis of 2313 SNPs (one random SNP per RAD locus), rooted with *A. croatica*; posterior probabilities are above the branches; (c) cluster assignment of the individuals revealed by STRUCTURE (2413 SNPs, one random SNP per RAD locus).

revealed optimal separation of the data into four groups, that is the partition exhibiting the highest similarity among runs and the highest delta *K* (see Fig. S1a, b, Supporting information). This grouping corresponded with geography (Fig. 1), clustering together populations from (i) Pannonian lowlands of Hungary and southern Slovakia (Pannonian group), (ii) the foothills of the Dinaric Alps and their surroundings in Slovenia, Croatia and Bosnia and Herzegovina (Dinaric group) and (iii) mid-altitudes to high altitudes of the western Carpathians in Slovakia (W Carpathian group) and (iv) mid-altitudes of southern and eastern Carpathians in Romania and the southern Dinarids in Serbia, as well as one population from southeastern Slovakia (SE Carpathian group). The spatially isolated populations from southern Baltic coast showed admixture between the W and SE Carpathian groups, however, with a higher probability of membership in the latter group (Fig. 1). Principal coordinate analysis (PCoA) of the SNP data confirmed the STRUCTURE–based grouping and revealed the Pannonian and Dinaric groups to be the most distinct, separated on the first and second PCoA axis, respectively (Fig. 1). For the follow-up analyses, we thus defined five genetically (pairwise F_{ST} ranging from 0.04 to 0.14, Table S2, Supporting information) and geographically distinct major lineages hereafter called the Pannonian, Dinaric, W Carpathian, SE Carpathian and Baltic groups. This grouping does not correspond with the current taxonomic treatment of *A. arenosa* (Fig. S2, Supporting information).

Considerable admixture among the major groups was indicated by STRUCTURE for populations in eastern

Slovakia (W Carpathian + SE Carpathian group; up to 0.45 admixture, i.e. probability of individual membership in the minor group), the northernmost Pannonian basin (Pannonian + W Carpathian and/or SE Carpathian groups; up to 0.21 admixture) and in the central Dinaric Alps (Dinaric + SE Carpathian group, up to 0.20 admixture, Fig. 1a). Finally, separate STRUCTURE analyses of each of the four main clusters (disregarding Baltic populations) resulted in an optimal $K = 2$ partition in each case (Fig. S1c, d, Supporting information), separating populations from distant regions within both Pannonian and Dinaric groups, part of the southern Carpathian populations from the rest of the SE Carpathian group and high-altitude populations (but partly with high admixture) from mid-altitude ones in the W Carpathian group (Fig. S3, Supporting information).

Phylogenetic relationships among the groups

We further inferred phylogenetic relationships among the major genetic groups by multispecies coalescent analysis (SNAPP). We did not find large differences in topologies and branch lengths either between different subsets analysed (different individuals from the same populations) or among different scales of the group delimitation (three vs. five *A. arenosa* groups, Fig. S4, Supporting information). The analyses jointly revealed the monophyly of all sampled *A. arenosa* diploids (rooted by *A. croatica*), the sister position of Pannonian populations to the remaining *A. arenosa* diploids (Fig. 1b) and, under a finer group delimitation, they also supported the monophyly of Carpathian + Baltic populations (Fig. S4, Supporting information). The estimated times of divergence of the major genetic groups (including the 95% HPD intervals) in all cases safely preceded the Holocene by approximately an order of magnitude (Fig. S4, Supporting information).

Testing the admixed origin of the Baltic group

Populations of *A. arenosa* from the southern Baltic coast appeared to be admixed between the W and SE Carpathian groups as indicated by their intermediate position in the PCoA ordination and admixed assignment in STRUCTURE analyses of the Carpathian + Baltic populations (Fig. 2) ($K = 2$ was the optimal partition with the highest similarity among runs and the highest delta K, Fig. S1, Supporting information). Ten STRUCTURE replicates run under $K = 3$ did not provide consistent outcomes, indicating either Baltic populations as a separate group or the same pattern as $K = 2$ plus an empty cluster (not shown). In the Treemix population graphs, the Baltic group was sister to the SE Carpathian group (with high bootstrap support) but was also linked to the W Carpathian group by a migration edge, suggesting admixture (Fig. 2b).

Consequently, we investigated the origin of Baltic populations by comparing three competing evolutionary scenarios (divergence from either W or SE Carpathian groups or admixture among these two groups, Fig. 2d) using the coalescent-based approximate Bayesian computation (ABC). This analysis also supported admixed origin of Baltic populations. This scenario (Scenario 1, Fig. 2d) exhibited the highest posterior probabilities, that is the numbers of simulated data sets with summary statistics similar to observed values, which were estimated by both the direct approach ($P > 0.87$; 95% confidence intervals 0.70–1) and the logistic regression ($P > 0.99$; 95% confidence intervals 0.99–1). In addition, scenario 1 exhibited a low probability of being erroneously selected even if it was not the true scenario (0.042 and 0.034 following the direct and logistic approaches, respectively), as evidenced by the comparison with 1000 pseudo-observed data sets. Finally, the allele frequency spectrum (AFS) of the Baltic group markedly differed from the spectra of the remaining groups (chi-squared tests with 2000 simulations, $P < 0.001$ in all pairwise comparisons). This difference was mainly caused by a shortage of rare alleles and a slight excess of intermediate frequency alleles, as shown by comparing to AFS under the expectation of a demographic equilibrium (Fig. S5c, Supporting information; chi-squared tests, 2000 simulations, $P = 0.002$).

Population-level variation

We inferred levels of intrapopulation diversity and interpopulation differentiation through genotyping an average of nine individuals per population using fourteen nuclear microsatellite loci. The high values of the global fixation index (F_{ST}, 0.25) and its standardized measure (G'_{ST}, 0.57) as well as high pairwise among-population F_{ST} values (mean 0.245, ranging from 0.012 to 0.568, Table 1) indicated strong genetic differentiation among diploid *A. arenosa* populations. Both diversity and differentiation of populations varied among the five major genetic groups (Table 1). The highest genetic diversity, measured as both expected heterozygosity (He) and Nei's estimator of genetic diversity (Hs), was detected in populations from the western Carpathians and the Baltic coast (Fig. 3a). Allelic richness was also highest in W Carpathian populations (Table 1). On the contrary, the Baltic and W Carpathian groups exhibited the lowest proportion of among-population genetic variance, as identified by AMOVA (20.9% and 25.4%, respectively), the lowest global and pairwise F_{ST} (Table 1), and they were the only groups lacking significant correlation among geographical and genetic distances

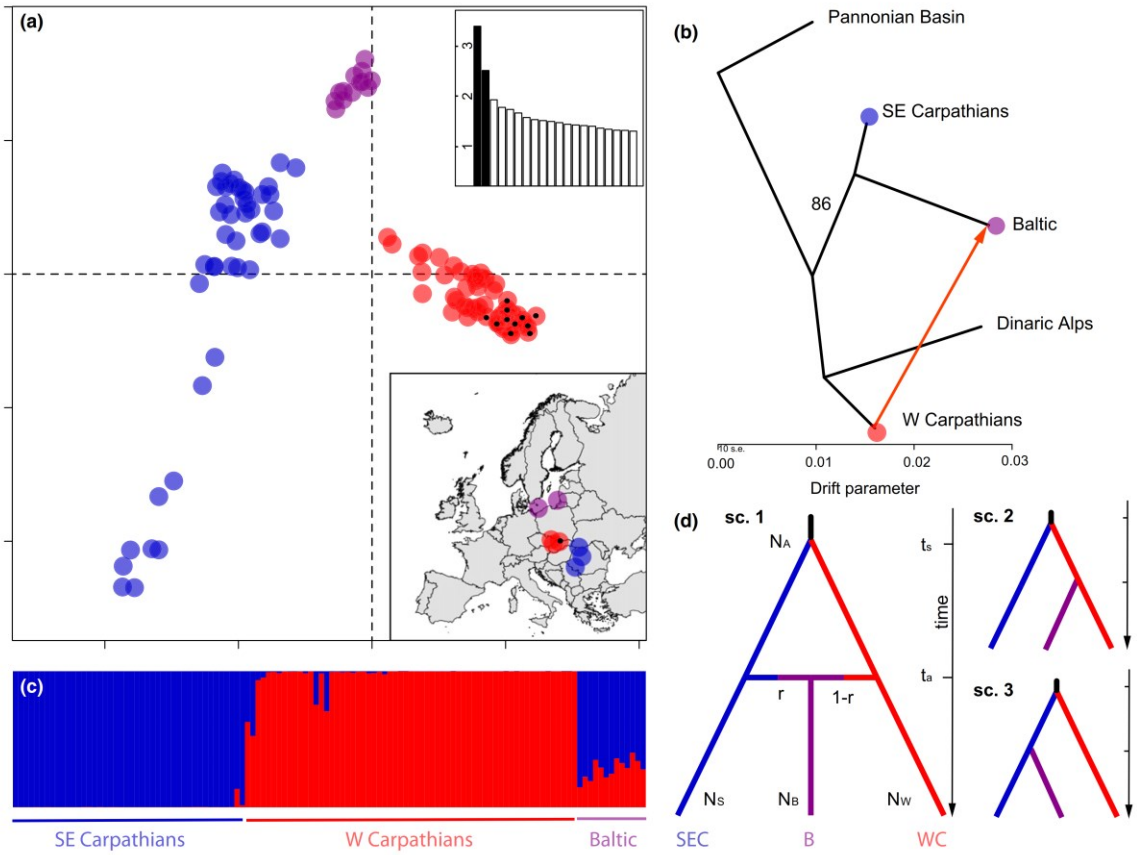


Fig. 2 Reconstruction of the relationships among the Baltic and Carpathian populations of diploid *Arabidopsis arenosa*. (a) Principal coordinate ordination (8949 SNPs, first and second axis displayed, histogram shows proportional contribution in explaining variance by the first 20 axes) of Carpathian + Baltic individuals; (b) Treemix maximum-likelihood graph (2323 SNPs) showing relationships among main lineages of *A. arenosa* with one migration edge (the single bootstrap support of >50% is above the corresponding branch); (c) STRUCTURE clustering (2323 SNPs); (d) three competing scenarios differing by the mode of origin of the Baltic populations simulated and tested in ABC framework with varying effective population sizes (*N*) and migration rate (*r*); scenario 1 was the most likely. Individuals from high-altitude populations of the W Carpathian group are marked by black dot in the principal coordinate diagram.

(isolation by distance, Table 1). Baltic populations also exhibited the low incidence of rare alleles (DW index, Table 1).

Dinaric populations, on the other hand, exhibited the lowest values of genetic diversity and the highest proportion of rare alleles (Table 1). Finally, the Pannonian and SE Carpathian populations varied considerably in both genetic diversity and rarity indices (Fig. 3), the highest diversity being expressed in admixed populations in northern Hungary in the Pannonian group and in the southeastern corner of the Carpathians in the SE Carpathian group. The three last-mentioned groups exhibited high overall population differentiation (0.21–0.22 and 0.46–0.48 for F_{ST} and G'_{ST}, respectively), a high among-population variance component of AMOVA (35.7–36.2%) and significant isolation-by-distance relationships (Table 1). Differentiation among the five major groups accounted for 15.3% of overall genetic variation in a hierarchical AMOVA, while this component was markedly reduced to 7.4% when only the Baltic, W and SE Carpathian groups were compared. There were six populations with significant heterozygote deficiency (i.e. nonzero lower 95% HPD for F_{IS}; average F_{IS} estimates ranged from 0.24 to 0.44 to in these populations), belonging to the W Carpathian, SE Carpathian and Pannonian groups (Table S1, Supporting information).

Finally, we focused on the W Carpathian group only and compared its ecologically highly divergent

Table 1 Genetic diversity and differentiation of diploid *A. arenosa* populations inferred from 14 microsatellite loci. The populations are grouped into five major genetic groups identified by SNP data; the W Carpathian group was further subdivided into two ecologically contrasting groups of populations from foothill and high-altitude habitats

Group	N indivs/ pops	N private alleles	Total N alleles	Expected heterozygosity†	Gene diversity (Hs)†	Rarity (DW)†	Allelic richness†	% of among-pop. variation‡	IBD§	F_{ST}	G'_{ST}	Among–population pairwise F_{ST}¶
Baltic	33/4	1	70	0.49 ± 0.065	0.58 ± 0.074	0.29 ± 0.47	2.77 ± 0.16	20.9	n.s.	0.12	0.40	(0.076-) 0.124 (–0.193)
Dinaric	76/9	11	123	0.44 ± 0.046	0.50 ± 0.032	0.49 ± 0.27	2.65 ± 0.26	35.7	0.78**	0.21	0.46	(0.031-) 0.214 (–0.354)
Pannonian	103/11	6	126	0.47 ± 0.055	0.54 ± 0.048	0.37 ± 0.18	2.71 ± 0.29	37.3	0.54***	0.22	0.48	(0.081-) 0.236 (–0.367)
SE Carpathian	152/17	6	121	0.45 ± 0.061	0.53 ± 0.060	0.26 ± 0.09	2.70 ± 0.40	36.2	0.29**	0.20	0.46	(0.045-) 0.203 (–0.422)
W Carpathian	206/23	24	162	0.51 ± 0.066	0.58 ± 0.058	0.38 ± 0.12	3.30 ± 0.28	25.4	n.s.	0.11	0.36	(0.012-) 0.113 (–0.310)
Significance of differences among groups				$P = 0.011$	$P = 0.004$	$P = 0.009$	$P < 0.001$					
W Carpathian – high-altitude pops.	85/10	7	119	0.47 ± 0.059	0.54 ± 0.056	0.35 ± 0.17	2.8 ± 0.36	22.6	n.s.	0.1	0.30	(0.012-) 0.097 (–0.239)
W Carpathian – foothill pops.	121/13	11	144	0.51 ± 0.079	0.60 ± 0.060	0.40 ± 0.14	3.39 ± 0.30	22.9	0.34*	0.1	0.38	(0.017-) 0.100 (–0.256)
Significance of differences between groups				$P = 0.79$	$P = 0.14$	$P = 0.37$	$P = 0.10$					
A. arenosa – all populations	570/64	—	200	0.48 ± 0.065	0.55 ± 0.061	0.36 ± 0.164	2.92 ± 0.423	42	0.33***	0.25	0.57	(0.012-) 0.245 (–0.568)

P-values were estimated by 500 permutations (*$P < 0.05$, **$P < 0.01$, ***$P = 0.002$).
†Mean ± standard deviation calculated from values of populations belonging to a particular genetic group (see Table S1, Supporting information).
‡As inferred by AMOVA performed on populations belonging to a particular genetic group.
§Isolation by distance tested by the Mantel test.
¶(min-) mean (-max) values calculated for all population pairwise comparisons within a particular major genetic group.

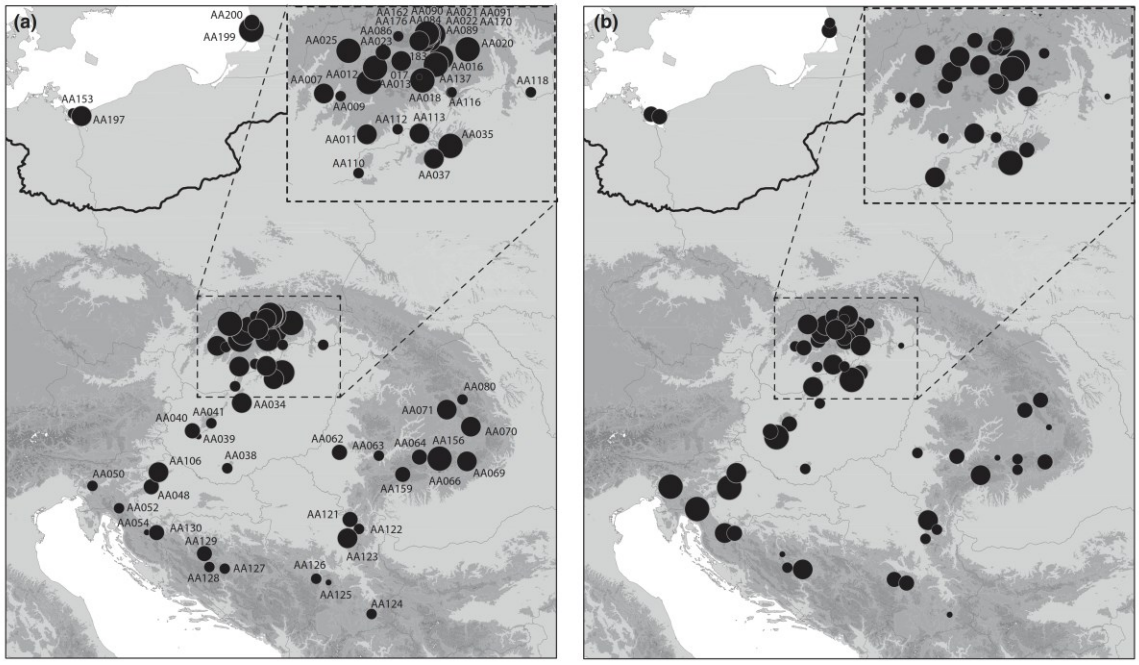


Fig. 3 Patterns of population-level diversity and rarity in diploid *A. arenosa*. (a) Genetic diversity of populations expressed by expected heterozygosity, H_e (range 0.30–0.61); (b) proportion of rare alleles in populations measured by frequency-downweighted index, DW (range 0.12–1.02). The bold line indicates maximal extent of the continental ice sheet during the last glaciation.

high-altitude and foothill populations. The groups were very weakly differentiated, as shown by both low F_{ST} (0.02) and a low among-population variation component of hierarchical AMOVA (3.4% of total variation). The two groups did not differ significantly either in indicators of population diversity or rarity (Table 1).

Correlation between niche differentiation and genetic structure

To identify the major ecological changes in the evolutionary history of diploid *A. arenosa*, we first tested whether the previously identified major genetic groups differ in the sampled ecological factors. Using linear discriminant analysis based on 117 ecologically screened populations (Fig. 4a), we confirmed the groups significantly differ in their climatic niche ($F_{8,108}$ = 22.4, P = 0.001) and identified the most important climatic factors. The first axis separated the Baltic and partly also the Pannonian groups (mostly representing areas with lower precipitation and lower isothermality, i.e. diurnal temperature oscillations relatively to annual oscillations, see Table S3, Supporting information), while the second axis separated Dinaric populations (occupying warmer and less seasonal areas with higher precipitation). By contrast, ecological niches of W and SE Carpathian populations were less distinct, as confirmed by their repeated misclassifications in a classificatory discriminant analysis (Table S4, Supporting information). The five major genetic groups also differed in the frequency with which their populations occupied neutral/siliceous vs. calcareous sites (χ^2 = 21.2, d.f. = 4, P < 0.001), but all groups contained at least a few populations from both calcareous and siliceous stands. Finally, the groups grew in different altitudes and in topographically different landscapes (the multinomial logistic regression model was significantly improved when including also altitude and slope inclination, likelihood ratio test, P < 000.1 in each case); the two Carpathian lineages grew in areas with generally highest slope inclination whereas the Baltic populations occupied flatlands. The effect of both variables remained significant also when the Baltic populations were removed (likelihood ratio test, P < 0.001 and P = 0.008 for altitude and slope, respectively).

We next addressed the general correspondence among the ecological and genetic data by testing for potential associations between SNP-based genetic distances and ecological requirements of the 64 genotyped populations. Although the studied ecological variables contributed to explaining genetic differences among all genotyped *A. arenosa* populations in a direct ordination

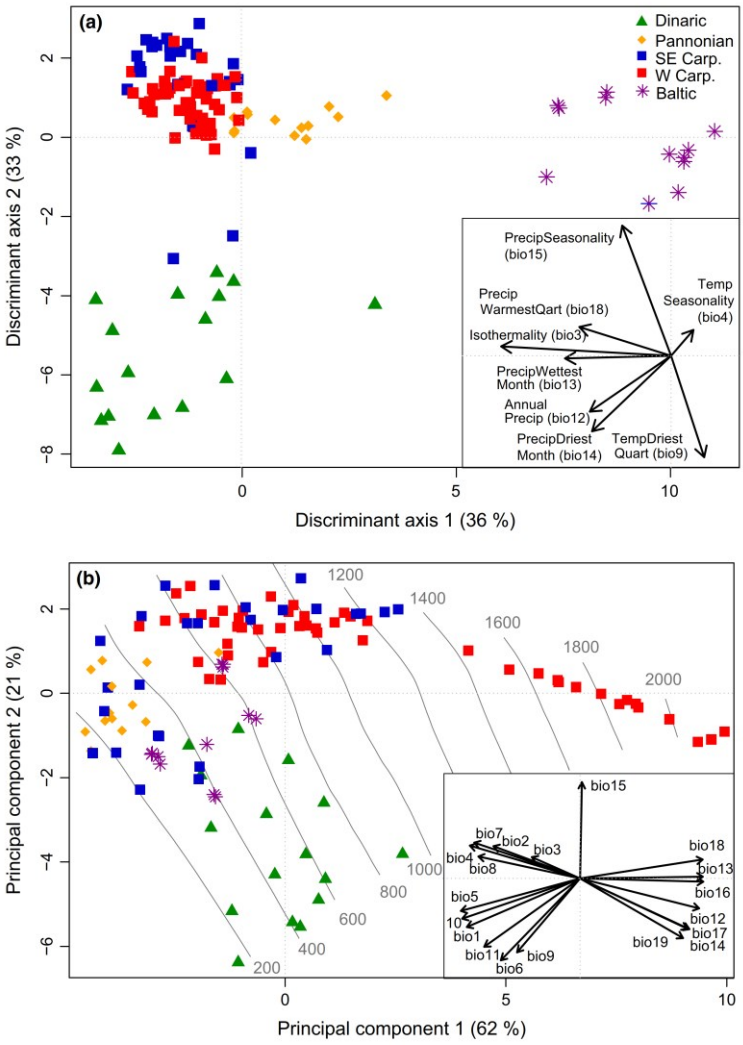


Fig. 4 Rangewide climatic niche variation in 117 ecologically screened populations of diploid *A. arenosa* denoted according to the five main genetic groups. (a) Linear discriminant analysis based on eight variables significantly contributing to the model (chosen using forward selection). (b) Principal component analysis based on all 19 Worldclim-derived bioclimatic variables. Insets display the contributions of individual variables (coded by their Bioclim numbers) to the first two ordination axes. Altitude (in m a.s.l.) was fitted onto the PCA diagram using loess smoother.

(constrained analysis of principal coordinates, CAP, Table 2), the overall correlation among ecological and genetic distances was low and nonsignificant (Mantel test, rM = 0.1). As shown by indirect ordination of the bioclimatic data, this was mainly due to a group of high-altitude populations from western Carpathians, which segregated from all remaining samples, including their nearby foothill counterparts, along the first PCA axis (explaining 62% of total variation, Fig. 4b). Importantly, the correlation among genetic and ecological distances became highly significant when the ecologically divergent high-altitude populations were excluded (rM = 0.27, P = 0.001, Fig. S6, Supporting information). Precipitation and temperature variables exhibited the strongest effects on the genetic structure of the populations (Table 2).

Finally, we examined ecological structuring of the genetic variation *within* each of the major genetic groups through separate CAP analyses. Environmental predictors had significant effect on the distribution of genetic variation within the two Carpathian lineages but not within the two groups restricted to low to mid-elevations, that is Dinaric and Pannonian ones (Table 2). Altitude and the composite temperature bioclimatic variable (strongly correlated, not shown) had the largest explanatory power in both W and SE Carpathian groups, being further complemented by the composite precipitation variable and soil reaction in W Carpathian populations (both were again correlated with the marked altitudinal gradient within this area) and by the composite precipitation variable and topography in the SE Carpathian group (Table 2, see also Fig. S7,

Table 2 Association of main environmental gradients with genetic distances among all 64 SNP-genotyped diploid *A. arenosa* populations and within the major genetic groups as inferred from a direct ordination (constrained analysis of principal coordinates). Marginal effects and unique contributions of corresponding environmental factors (in % of total variation) are shown before and after the slash, respectively

Group	*N*	All environmental variables	Altitude	Substrate	Topography[‡]	Precipitation[§]	Temperature[§]	rM ecology[¶]	rM geography[¶]
A. arenosa – all populations	64	**23.1*****	**6.1*****/**3.6****	**3.1***/2.0	**3.2***/1.3	**8.1*****/**7.0*****	**8.0*****/**5.6*****	0.10	**0.35*****
A. arenosa – high-altitude excluded	54	**22.1*****	**3.2*****/**3.3****	**3.0***/2.1	**2.6***/1.7	**6.5*****/**5.8*****	**3.9*****/**4.3*****	**0.27*****	**0.3*****
Dinaric	9	57.6	17.2/16.6	Not tested	10.3/8.6	11.4/15.1	9.4/13.4	0.09	**0.70*****
Pannonian	11	56.2	10.8/9.2	**16.5***/8.2	7.8/9.2	7.0/9.9	**18.5***/9.4	0.02	**0.41****
SE Carpathian	17	**42.4****	**15.1*****/4.7	5.1/5.8	**10***/4.8	**11.2*/10.9****	**15.7*****/5.3	**0.37****	**0.43*****
W Carpathian[†]	22	**34.4*****	**12*****/3.1	**12.3*****/**7.4***	5.5/3.9	**12.4*****/5.8	**12.3*****/3.3	**0.22****	**0.24***

P-values were estimated by 500 permutations (*$P < 0.05$, **$P < 0.01$, ***$P = 0.002$); significant values are shown in bold.
[†]Admixed population AA116 was excluded from the data set.
[‡]Slope inclination in the surrounding area as inferred from Worldclim data.
[§]Composite environmental variable effects of eleven temperature- and eight precipitation-related bioclimatic variables reflected as the scores of the first axis in a separate principal component analyses.
[¶]Mantel correlations among genetic and ecological (rM ecology) or geographical (rM geography) distances of the genotyped populations.

Supporting information for the ordination diagrams fitted by the environmental characteristics).

Discussion

Survival in northern glacial refugia

While glacial survival in the three southern European peninsulas (Iberian, Appennine and Balkan) is a paradigm (Taberlet *et al.* 1998; Hewitt 2004), continuous persistence of populations in more northerly areas of central Europe ('cryptic' northern refugia, Stewart & Lister 2001) through the last glacial maximum (LGM) is a subject of debate. Glacial survival in European 'cryptic' northern refugia is documented for some cold-tolerant arctic and boreal species (e.g. Tollefsrud *et al.* 2008; Tzedakis *et al.* 2013; Douda *et al.* 2014; Mandák *et al.* 2016), but it remains particularly controversial for their temperate counterparts (Willis & van Andel 2004; Birks & Willis 2008; Tzedakis *et al.* 2013; Robin *et al.* 2016). Evolutionary history of diploid *A. arenosa*, a temperate herb, significantly contributes to this debate as several lines of evidence suggest that at least two genetically distinct lineages currently occurring in eastern central Europe (Pannonian Basin and the western Carpathians) probably survived the LGM locally in northern refugia. First, at least two highly divergent groups of *A. arenosa* occur exclusively in central Europe, while the Balkan Peninsula, that is the 'classical' southern glacial refugium and source of postglacial recolonizers of temperate European species (Bennett *et al.* 1991; Taberlet *et al.* 1998; Tzedakis 2004), is occupied by one divergent endemic lineage and one lineage shared with southeastern Carpathians (Dinaric and SE Carpathian lineages, respectively, Fig. 1). Second, all Balkan populations exhibit markedly reduced levels of genetic diversity (Fig. 3a) and populations from the central Balkans (Serbia) also show low rarity and cluster with SE Carpathian populations. This pattern suggests (re)colonization of Serbia from the southern Carpathians (connections between the two regions have been documented also for other plants, e.g. Frajman & Oxelman 2007; Ronikier 2011) and isolated survival in small yet distinct populations in the northern Dinarids (genetically distinct, Dinaric, lineage comprising populations with high DW values; Figs 1 and 3b). Third, the high genetic differentiation of the Carpathian and Pannonian lineages, probably dating back to the Pleistocene (Fig. S4, Supporting information; see also Arnold *et al.* 2015), and topology of the species tree (Pannonian lineage in the basal position, Fig. 1b), rules out their recent postglacial segregation from Balkan populations. Although our time estimates, directly depending on a mutation rate used for rescaling (3.7×10^{-8}

substitutions/site/generation previously inferred for *A. arenosa*, Arnold *et al.* 2015), should be taken with caution, a very recent Holocene origin of the main *A. arenosa* differentiation is unlikely as (i) confidence limits of our divergences estimates are still an order of magnitude older (Fig. S4, Supporting information) and (ii) the other available estimate of *Arabidopsis* genome-wide mutation rates would imply even older divergences (7×10^{-9} substitutions/site/generation in *A. thaliana*, Ossowski *et al.* 2010). Finally, long-term in situ persistence of large populations in eastern and central Europe is supported by elevated genetic diversity of their present populations indicated jointly by microsatellites (Fig. 3), SNPs (no obvious deficit of low-frequency alleles in the allele frequency spectra, Fig. S5, Supporting information) and previous AFLP and plastid DNA survey (Schmickl *et al.* 2012). The opposite scenario, that is recent postglacial expansion from the Balkans, would imply weakly differentiated and genetically depauperate lineages in central Europe.

We hypothesize that the northern refugia of diploid *A. arenosa* might have been located in adjacent topographically diverse areas such as the Carpathian foothills. The Carpathians are already considered a strong candidate for a 'cryptic' northern refugium (e.g. Provan & Bennett 2008; but see Tzedakis *et al.* 2013) based on both fossil data indicating the persistence of patches of favourable temperate habitats throughout the LGM (e.g. Willis & van Andel 2004; Birks & Willis 2008) and the genetic structure of several temperate plants and animals (e.g. Babik *et al.* 2004; Kotlík *et al.* 2006; Magri *et al.* 2006; Wielstra *et al.* 2015). The western Carpathians represent a particularly good candidate region, as this area hosts the genetically most diverse populations of diploid *A. arenosa* (both in terms of population diversity and rarity, Table 1), it was continuously forested throughout the LGM (Jankovská & Pokorný 2008), and directly dated land snail fossils document here a whole truly temperate species assemblage from the LGM period (Juřičková *et al.* 2014). Although backward gene flow from co-occurring tetraploids (suggested by the coalescent simulations, Arnold *et al.* 2015) might have inflated the genetic variation of diploids in the western Carpathians, we do not expect it to have substantially altered the overall diversity patterns, as the tetraploids are direct and recent descendants of W Carpathian diploids (~11 000–30 000 generations ago; Arnold *et al.* 2015), and both lineages thus presumably share the same alleles due to common descent. In addition, there is virtually no chance of ongoing gene flow, as indicated by the nearly complete lack of naturally occurring triploid individuals, that is potential mediators of gene flow from tetraploids to diploids (Kolář *et al.* 2015a).

Allopatric differentiation was likely the major force behind the observed genetic differentiation of diploid *A. arenosa* into the four major groups in central and southeastern Europe. The role of spatiotemporal isolation is underlined by very good correspondence of the major genetic breaks with prominent barriers in species distributions such as the border of the Pannonian and Carpathian regions (e.g. Futák *et al.* 1966), the border between the western and eastern Carpathians (Wołoszczak 1896; Pawłowski 1970) and the split in the mid-Dinaric Alps (Kutnjak *et al.* 2014); all these barriers also structure the intraspecific genetic differentiation of other plants and animals (e.g. Kryštufek *et al.* 2007; Mráz *et al.* 2007; Ronikier *et al.* 2008; Těšitel *et al.* 2009; Ronikier 2011; Surina *et al.* 2011; Winkler *et al.* 2012; Caković *et al.* 2015). After a spread in open postglacial landscapes (e.g. the early Holocene, Ložek 1973), which might have provided favourable conditions for migration of weak competitors such as *A. arenosa*, the major lineages met in several areas and hybridized (admixture suggested by the STRUCTURE analyses, Fig. 1a, c). Past gene flow among these lineages, for example during previous interglacial(s), could not be ruled out at this stage of investigations but a larger haplotype-based data set should be used for precise inference of such complex demographic events (Schraiber & Akey 2015).

Origin of the northern postglacial colonizers

The northernmost disjunct outposts of the range of diploid *A. arenosa* occupy the southern Baltic coast, where diploids grow in ecologically distinct stands such as chalk cliffs and grey sand dunes. Baltic populations are genetically very close to both W and SE Carpathian groups (among-group F_{ST} ranging from 0.03 to 0.05, Table S2, Supporting information) and the exploratory techniques (Fig. 2b, c) as well as coalescent-based tests (Fig. 2d) jointly showed their likely origin from admixture between these two Carpathian lineages. The Baltic sea coastline is a novel environment that, after melting of the continental ice sheet, underwent dramatic changes driven by a dynamic equilibrium between rising global ocean levels and gradual uplift of the deglaciated land mass (Björck 1995). In its current place, it began to develop only ca 5700 years ago (Janke *et al.* 1993; Wohlfarth *et al.* 2008). It is thus plausible that the admixture event took part in the areas where both putative parents still co-occur and hybridize (i.e. the border between the western and eastern Carpathians, Fig. 1) and only after that the admixed individuals migrated northwards to the novel postglacial environment.

Such migration might have taken place in earlier periods of the Holocene, either gradually through less competitive nonforest habitats (dominating the landscape in

those times, Ralska-Jasiewiczowa 2004) or via long distance, for example, along rivers connecting the Carpathians and the Baltic Sea. There are no extant diploid *A. arenosa* populations known from the areas between Baltic coast and Carpathians, which corresponds with the lack of suitable habitats in these flat landscapes (Kolář *et al.* 2015a) and similar Carpathian – Baltic disjunctions documented for other plants preferring low-competitive environments (Zając & Zając 2011). All current occurrences of *A. arenosa* in this area that are known to us come from non-native stands and represent a different tetraploid lineage preferring man-made habitats such as roadsides and railway tracks (Arnold *et al.* 2015; Kolář *et al.* 2015a and our observations). We also do not expect very recent spread caused by humans, as this would result in much lower indicators of rare genetic diversity (DW indices are still well within the range of most other lineages, Table 1) and an absence of genetic differentiation from source populations (Baltic populations still have a somewhat distinct position, e.g. in the PCoA plots, Fig. 2a). Finally, that Baltic diploids prefer natural habitats and the total range spanning over 1000 km of coast (E Denmark – Latvia, Fig. S8, Supporting information) also speak against recent human introduction.

Ecological gradients as potential drivers of A. arenosa differentiation

Preferring a wide variety of substrates, climatic niches and habitats spanning over 2500 altitudinal metres, *Arabidopsis arenosa* represents a suitable model for testing hypotheses concerning niche conservatism vs. shifts throughout its evolutionary history (Schmickl *et al.* 2012; Hohmann *et al.* 2014; Kolář *et al.* 2015a). The major genetic lineages of diploid *A. arenosa* indeed occupy distinct climatic niches differentiated along the altitudinal gradient, although remarkable deviations from this general pattern exist. The two basal lineages, Pannonian and Dinaric (as well as the outgroup *A. croatica*), occupy relatively warmer lowland to foothill areas of the Pannonian steppe region and the sub-Mediterranean northwestern Balkans (Fig. 5). We thus infer that *A. arenosa* probably originated in warmer, low-elevation habitats, later colonized the higher and topographically more structured Carpathian mountain arch, and only from there, it finally formed the northernmost postglacial outposts in Baltic Sea coastal landscapes. Such a scenario fits well with niche evolution patterns in the *Arabidopsis* genus as a whole, whose ancestral niche was reconstructed to lie in relatively warmer areas, and several independent expansions towards cooler temperate/arctic climates were suggested (Hoffmann 2005). Climatic and topographic gradients also significantly correlated with genetic distances within the two Carpathian groups (Table 2). Similar trends observed at both rangewide and within-lineage spatial scales suggest that the altitudinal gradient and its bioclimatic correlates may play an essential role in the genetic differentiation of *A. arenosa*, its potential adaptive value, however, remaining to be determined by reciprocal transplants and/or physiological experiments.

Notably, we identified two exceptional cases of strongly ecologically yet only weakly genetically differentiated groups of populations (Table 2, Fig. S6, Supporting information), indicating probably recent dramatic niche expansions of *A. arenosa*. These populations occupied sites covered by mountain (high elevations of western Carpathians) or continental ice sheets (Baltic Sea coastal habitats) during the last glaciations (Svendsen *et al.* 2004; Lindner *et al.* 2010). In particular, high-altitude populations from the western Carpathians, although genetically close to their foothill counterparts (e.g. Fig. 2a), represent the climatically most divergent group of diploid *A. arenosa* (Fig. 4b). They are not only

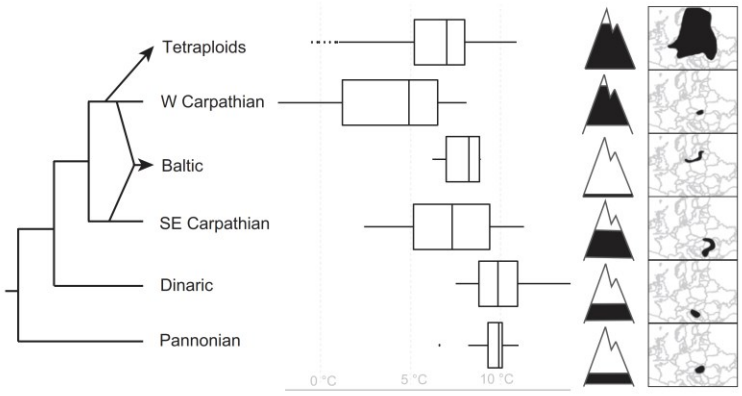


Fig. 5 Summary hypothesis on the history and niche evolution of diploid (based on this study) and autotetraploid (following Arnold *et al.* 2015 and Kolář *et al.* 2015a) lineages of *Arabidopsis arenosa*. Phylogenetic reconstruction, mean annual temperature, altitudinal range (proportionally in range 0–2700 m a.s.l.) and total distribution range are displayed for each lineage.

ecologically but also morphologically so distinct they were considered a separate species for last 200 years (*Arabis/Arabidopsis neglecta*; (Schultes 1814; Měsíček & Goliašová 2002; Schmickl *et al.* 2012). Second, the Baltic diploid populations occupy highly distinct habitats from those of both its genetically close Carpathian ancestors (Fig. 4a). Interestingly, such niche shifts were seemingly not linked with dramatic losses of the original genetic diversity, as in both cases, the colonizers and their putative ancestors shared similar diversity levels (and in W Carpathians also rare alleles, Table 1). Colonization by relatively large populations, gene flow (in the case of spatially close W Carpathian populations) and/or admixed origin (Baltic populations) may explain such pattern. This contrasts with the generally acknowledged trend of genetic depauperation linked with postglacial recolonization (Hewitt 2000; Widmer & Lexer 2001; Keppel *et al.* 2012), which has recently been challenged by several case studies, however (Prentice *et al.* 2011; Stone *et al.* 2012; Kolář *et al.* 2015b).

Taken together, diploid *A. arenosa* (or at least its Carpathian lineages) shows considerable potential for colonizing novel postglacial environments, and the colonizers still retain a large portion of the original genetic diversity and exhibit only incipient genetic differentiation. Moreover, such a process was paralleled in populations, which have undergone genome duplication. The widespread autotetraploid *A. arenosa* lineage, which originated from W Carpathian diploids (Arnold *et al.* 2015), was even more successful in expanding its niche into both high-altitude areas and deglaciated northern Europe (Kolář *et al.* 2015a; Fig. 5), but still retaining high levels of genetic diversity (Schmickl *et al.* 2012). Similarly, nearly 3000-m altitudinal variation in allotetraploid *Arabidopsis kamchatica* in Japan is attributed to a climatic niche shift (Shimizu-Inatsugi *et al.* 2009; Kenta *et al.* 2011). The suggestion of higher rates of colonization of novel environments after genome duplications is still a controversial topic (Brochmann *et al.* 2004; Husband *et al.* 2013), but here, we show that such potential can sometimes already be manifested in diploid progenitors of the polyploid lineage. Finally, local patterns of adaptation linked with altitude have been identified within other diploid species, namely *A. halleri* (Fischer *et al.* 2013; Kubota *et al.* 2015) and *A. thaliana* (Méndez-Vigo *et al.* 2011; Quèbre 2012), suggesting that altitude may be an important driver of genetic differentiation and adaptation in the entire model genus regardless of the presence of polyploidy (see also Hoffmann 2005).

Finally, in contrast to climatic and topographic variables, substrate preferences seem to have played a rather minor role in the broadest, rangewide genetic differentiation of the group. Although significant differences in frequencies were observed between the groups, there is no purely calcicolous or silicicolous lineage within diploid *A. arenosa*, and the change in frequencies of the soil types does not consistently decrease or increase throughout the its recent evolution. This stands in contrast to general patterns, known particularly from European mountains, where ecological sorting according to soil reaction (calcicolous vs. silicicolous species) is a major determinant of overall genetic structure within multiple species (e.g. Schönswetter *et al.* 2005; Alvarez *et al.* 2009) and represents a diversification trigger within some species groups (e.g. Dillenberger & Kadereit 2013; Moore & Kadereit 2013). On the other hand, our results are in line with the generally broad range of substrate preferences known also for *A. arenosa* tetraploids (Kolář *et al.* 2015a) and other *Arabidopsis* species: *Arabidopsis lyrata* (Černý *et al.* 2006; Turner *et al.* 2010) and *A. halleri* (Clauss & Koch 2006). This suggests a flexibility of these species to colonize varied substrates across their range.

Taxonomic implications

The documented segregation into several spatially and genetically well-delimited groups stands in strong contrast with the taxonomic concepts applied to *A. arenosa* so far. Three of the six provisionally recognized diploid taxa (*A. carpatica*, *A. neglecta*, *A. nitida*; Schmickl *et al.* 2012) fall within the single (W Carpathian) major genetic lineage, while the distinct Dinaric and SE Carpathian groups are completely neglected in the current taxonomic treatments. *Arabidopsis petrogena* (A. Kern) V. I. Dorof is the only taxon, which fits into one of our major lineages (the Pannonian lineage, see Fig. S2, Supporting information). Whether the observed genetic differentiation already lead to accumulation of reproductive incompatibilities (such as in *A. lyrata*; Leppälä *et al.* 2013) and speciation remains to be experimentally tested, the observed admixture in several areas suggest that the major lineages did not evolve impermeable reproductive barriers. A proper taxonomic assessment of the discovered genetic structure requires further morphological and experimental investigations, involving also the tetraploid cytotype. Our current results highlight that studies using *A. arenosa* as a model should not rely on the current highly inaccurate taxonomy but rather refer to the major genetically and spatially distinct lineages, as described here (Fig. 1), within a widely treated *Arabidopsis arenosa 'sensu lato'*.

Conclusions and further prospects

Here, we provide the first rangewide assessment of genetic structure and evolutionary history of diploid

Arabidopsis arenosa, an important emerging model in evolutionary biology. Genome-wide SNPs and microsatellites show that diploid *A. arenosa* splits into several genetically and ecologically differentiated lineages, which probably independently survived the last glacial maximum in distinct areas, including northern 'cryptic' refugia situated in eastern central Europe. Besides spatiotemporal segregation, altitude and its bioclimatic correlates seem to be important drivers of genetic differentiation, as we found a similar correspondence among genetic and ecological data both range-wide (expansion from warmer flatlands to the mountains) and within lineages occupying topographically variable landscapes (Carpathian Mts.). Further investigations are necessary to test whether such correspondence between genetic structure and ecology resulted from random population genetic processes (e.g. reflecting rapid colonization of higher altitudes and/or long-term isolation in allopatry) or local adaptation to different environments. The two putatively recent colonizers of ecologically distinct postglacial landscapes promise to be particularly useful for investigating the evolutionary drivers of local adaptation including the role of individual genes. Given the expected multiple switches among distinct substrates in different species, *A. arenosa* may also serve as a good model for addressing local substrate adaptation, including the role of standing variation and genome duplication. Finally, the intriguing question whether such ecological and geographical shifts might be linked with incipient homoploid speciation should also be tested by assessing the strength of reproductive isolation among the spatially, genetically and ecologically divergent diploid lineages. Importantly, as the molecular evolutionary studies published so far focused only on one or two major lineages of diploid *A. arenosa*, we highlight the so far neglected natural diversity of the *A. arenosa* diploids that can be used for addressing diverse evolutionary and molecular questions.

Acknowledgements

We thank Michael Nowak, Magnus Popp, Brian Arnold, Levi Yant and Tanja Slotte for useful suggestions during the preparation of this manuscript and Frederick Rooks, Zigmantas Gudžinskas, Stanislav Španiel, Eva Kolářová, Martin Kolník, Jaromír Kučera, Klára Kabátová, Dimitar Lakušić and Nevena Kuzmanović for their help with field collections. Frederick Rooks kindly improved our English. Computing was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR), http://www.hpc.uio.no/. This work was supported by the Czech Science Foundation (P506/12/0668 to K.M.), the Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (624421/WILDARABIDOPSIS to F.K.) and by the Grant Agency of the Charles University project (GAUK/391315 to G.F.).

References

Alvarez N, Thiel-Egenter C, Tribsch A *et al.* (2009) History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters*, **12**, 632–640.

Anderson MJ, Willis TJ (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology*, **84**, 511–525.

Arnold B, Kim S-T, Bomblies K (2015) Single geographic origin of a widespread autotetraploid *Arabidopsis arenosa* lineage followed by interploidy admixture. *Molecular Biology and Evolution*, **32**, 1382–1395.

Babik W, Branicki W, Sandera M *et al.* (2004) Mitochondrial phylogeography of the moor frog, *Rana arvalis*. *Molecular Ecology*, **13**, 1469–1480.

Baduel P, Arnold B, Weisman CM, Hunter B, Bomblies K (2016) Habitat-associated life history and stress-tolerance variation in *Arabidopsis arenosa*. *Plant Physiology*. doi: 10.1104/pp.15.01875.

Beaumont MA (2010) Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.

Bennett KD, Tzedakis PC, Willis KJ (1991) Quaternary refugia of north European trees. *Journal of Biogeography*, **18**, 103–115.

Birks HJB, Willis KJ (2008) Alpines, trees, and refugia in Europe. *Plant Ecology & Diversity*, **1**, 147–160.

Björck S (1995) A review of the history of the Baltic Sea, 13.0-8.0 ka BP. *Quaternary International*, **27**, 19–40.

Brochmann C, Brysting AK, Alsos IG *et al.* (2004) Polyploidy in arctic plants. *Biological Journal of the Linnean Society*, **82**, 521–536.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**, 1917–1932.

Caković D, Stešević D, Schönswetter P, Frajman B (2015) How many taxa? Spatiotemporal evolution and taxonomy of *Amphoricarpos* (Asteraceae, Carduoideae) on the Balkan Peninsula. *Organisms Diversity & Evolution*, **15**, 429–445.

Černý T, Petřík P, Boublík K, Kolbek J (2006) Habitat requirements of *Cardaminopsis petraea* — Rare and relict species of the Czech Republic. *Biologia*, **61**, 51–61.

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity*, **100**, 106–113.

Clauss MJ, Koch MA (2006) Poorly known relatives of *Arabidopsis thaliana*. *Trends in Plant Science*, **11**, 449–459.

Clauss MJ, Cobban H, Mitchell-Olds T (2002) Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaeae). *Molecular Ecology*, **11**, 591–601.

Comai L, Tyagi AP, Winter K *et al.* (2000) Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *The Plant Cell*, **12**, 1551–1567.

Cornuet J-M, Pudlo P, Veyssier J *et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide

polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187–1189.
Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.
Dillenberger MS, Kadereit JW (2013) The phylogeny of the European high mountain genus *Adenostyles* (Asteraceae-Senecioneae) reveals that edaphic shifts coincide with dispersal events. *American Journal of Botany*, **100**, 1171–1183.
Douda J, Doudová J, Drašnarová A *et al.* (2014) Migration patterns of subgenus *Alnus* in Europe since the last glacial maximum: a systematic review. *PLoS One*, **9**, e88709.
Dray S, Dufour A-B (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.
Ehrich D (2006) aflpdat: a collection of r functions for convenient handling of AFLP data. *Molecular Ecology Notes*, **6**, 603–604.
Eidesen PB, Ehrich D, Bakkestuen V *et al.* (2013) Genetic roadmap of the Arctic: plant dispersal highways, traffic barriers and capitals of diversity. *New Phytologist*, **200**, 898–910.
Estoup A, Lombaert E, Marin J-M *et al.* (2012) Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources*, **12**, 846–855.
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
Fagundes NJR, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences*, **104**, 17614–17619.
Fijarczyk A, Nadachowska K, Hofman S *et al.* (2011) Nuclear and mitochondrial phylogeography of the European fire-bellied toads *Bombina bombina* and *Bombina variegata* supports their independent histories. *Molecular Ecology*, **20**, 3381–3398.
Fischer MC, Rellstab C, Tedder A *et al.* (2013) Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology*, **22**, 5594–5607.
Frajman B, Oxelman B (2007) Reticulate phylogenetics and phytogeographical structure of *Heliosperma* (Sileneae, Caryophyllaceae) inferred from chloroplast and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, **43**, 140–155.
Futák J, Dostál J, Novák F (1966) *Flóra Slovenska [Flora of Slovakia]*. Vydav. Slov. Akad. Vied, Bratislava.
Griffiths HI, Kryštufek B, Reed JM (eds.) (2004) *Balkan Biodiversity*. Springer Netherlands, Dordrecht.
Havrdová A, Douda J, Krak K *et al.* (2015) Higher genetic diversity in recolonized areas than in refugia of *Alnus glutinosa* triggered by continent-wide lineage admixture. *Molecular Ecology*, **24**, 4759–4777.
Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.
Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.
Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **359**, 183–195.
Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
Hoffmann MH (2005) Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution*, **59**, 1425–1436.
Hohmann N, Schmickl R, Chiang T-Y *et al.* (2014) Taming the wild: resolving the gene pools of non-model *Arabidopsis* lineages. *BMC Evolutionary Biology*, **14**, 224.
Hollister JD, Arnold BJ, Svedin E *et al.* (2012) Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genetics*, **8**, e1003093.
Hu TT, Pattyn P, Bakker EG *et al.* (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics*, **43**, 476–481.
Husband BC, Baldwin SJ, Suda J (2013) The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In: *Plant Genome Diversity*(eds Greilhuber J, Doležel J, Wendel JF), pp. 255–276. Springer, Vienna.
Janke W, Kliewe H, Sterr H (1993) Holozäne Genese der Boddenküste Mecklenburg-Vorpommerns und deren künftige klimabedingte Entwicklung. In: *Klimaänderung und Küste* (eds Schellnhuber PDH-J, Sterr DH), pp. 137–152. Springer, Berlin.
Jankovská V, Pokorný P (2008) Forest vegetation of the last full-glacial period in the Western Carpathians (Slovakia and Czech Republic). *Preslia*, **80**, 307–324.
Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
Juřičková L, Horáčková J, Ložek V (2014) Direct evidence of central European forest refugia during the last glacial period based on mollusc fossils. *Quaternary Research*, **82**, 222–228.
Kamm A, Galasso I, Schmidt T, Heslop-Harrison JS (1995) Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationships between *Arabidopsis* species. *Plant Molecular Biology*, **27**, 853–862.
Kenta T, Yamada A, Onda Y (2011) Clinal variation in flowering time and vernalisation requirement across a 3000-m altitudinal range in perennial *Arabidopsis kamchatica* ssp. *kamchatica* and annual lowland subspecies *kawasakiana*. *Journal of Ecosystems and Ecography*, S6:001. doi: 10.4172/2157-7625.S6-001.
Keppel G, Van Niel KP, Wardell-Johnson GW *et al.* (2012) Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography*, **21**, 393–404.
Koch MA, Matschinger M (2007) Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, **104**, 6272–6277.
Kolář F, Lučanová M, Záveská E *et al.* (2015a) Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae). *Biological Journal of the Linnean Society*. doi: 10.1111/bij.12479.
Kolář F, Píšová S, Záveská E *et al.* (2015b) The origin of unique diversity in deglaciated areas: traces of Pleistocene processes in north-European endemics from the *Galium pusillum* polyploid complex (Rubiaceae). *Molecular Ecology*, **24**, 1311–1334.

Kotlík P, Deffontaine V, Mascheretti S *et al.* (2006) A northern glacial refugium for bank voles (*Clethrionomys glareolus*). *Proceedings of the National Academy of Sciences*, **103**, 14860–14864.

Koutecký P (2014) MorphoTools: a set of R functions for morphometric analysis. *Plant Systematics and Evolution*, **301**, 1115–1121.

Kryštufek B, Buzan EV, Hutchinson WF, Hänfling B (2007) Phylogeography of the rare Balkan endemic Martino's vole, *Dinaromys bogdanovi*, reveals strong differentiation within the western Balkan Peninsula. *Molecular Ecology*, **16**, 1221–1232.

Kubota S, Iwasaki T, Hanada K *et al.* (2015) A genome scan for genes underlying microgeographic-scale local adaptation in a wild *Arabidopsis* species. *PLOS Genetics*, **11**, e1005361.

Kutnjak D, Kuttner M, Niketić M *et al.* (2014) Escaping to the summits: phylogeography and predicted range dynamics of *Cerastium dinaricum*, an endangered high mountain plant endemic to the western Balkan Peninsula. *Molecular Phylogenetics and Evolution*, **78**, 365–374.

Leppälä J, Bokma F, Savolainen O (2013) Investigating incipient speciation in *Arabidopsis lyrata* from patterns of transmission ratio distortion. *Genetics*, **194**, 697–708.

Lindner L, Dzierżek J, Marciniak B, Nitychoruk J (2010) Outline of Quaternary glaciations in the Tatra Mts.: their development, age and limits. *Geological Quarterly*, **47**, 269–280.

Ložek V (1973) *Příroda ve Čtvrtohorách. [Nature in the Quaternary]*. Academia, Praha.

Lunter G, Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, **21**, 936–939.

Madlung A, Masuelli RW, Watson B *et al.* (2002) Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant Physiology*, **129**, 733–746.

Magri D, Vendramin GG, Comps B *et al.* (2006) A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist*, **171**, 199–221.

Mandák B, Havrdová A, Krak K *et al.* (2016) Recent similarity in distribution ranges does not mean a similar postglacial history: a phylogeographical study of the boreal tree species *Alnus incana* based on microsatellite and chloroplast DNA variation. *New Phytologist*, **210**, 1395–1407.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

Méndez-Vigo B, Picó FX, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C (2011) Altitudinal and climatic adaptation is mediated by flowering traits and FRI, FLC, and PHYC genes in Arabidopsis. *Plant Physiology*, **157**, 1942–1955.

Měsíček J, Goliašová K (2002) *Cardaminopsis* (C. A. Mey.) Hayek. In: *Flóra Slovenska*(eds Goliašová K, Šípošová H), pp. 388–415. Veda, Bratislava.

Miller N, Estoup A, Toepfer S *et al.* (2005) Multiple transatlantic introductions of the western corn rootworm. *Science*, **310**, 992–992.

Moore AJ, Kadereit JW (2013) The evolution of substrate differentiation in *Minuartia* series *Laricifoliae* (Caryophyllaceae) in the European Alps: In situ origin or repeated colonization? *American Journal of Botany*, **100**, 2412–2425.

Mráz P, Gaudeul M, Rioux D *et al.* (2007) Genetic structure of *Hypochaeris uniflora* (Asteraceae) suggests vicariance in the Carpathians and rapid post-glacial colonization of the Alps from an eastern Alpine refugium. *Journal of Biogeography*, **34**, 2100–2114.

Nägele K-L, Hausdorf B (2015) Comparative phylogeography of land snail species in mountain refugia in the European Southern Alps. *Journal of Biogeography*, **42**, 821–832.

Nei M (1972) Genetic distance between populations. *The American Naturalist*, **106**, 283–292.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, **12**, 443–451.

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

Oksanen J, Blanchet FG, Kindt R *et al.* (2013) *vegan: Community Ecology Package*.

Ossowski S, Schneeberger K, Lucas-Lledó JI *et al.* (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, **327**, 92–94.

Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, **26**, 419–420.

Paun O, Schönswetter P, Winkler M, Tribsch A; IntraBioDiv Consortium (2008) Historical divergence vs. contemporary gene flow: evolutionary history of the calcicole *Ranunculus alpestris* group (Ranunculaceae) in the European Alps and the Carpathians. *Molecular Ecology*, **17**, 4263–4275.

Pawłowski B (1970) Remarques sur l'endémisme dans la flore des Alpes et des Carpates. *Vegetatio*, **21**, 181–243.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, **7**, e37135.

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, **8**, e1002967.

Prentice HC, Andersson S, Månsby E (2011) Mosaic variation in allozyme and plastid DNA markers in the European ranges of *Silene vulgaris* and its partially sympatric relative *S. uniflora* (Caryophyllaceae). *Botanical Journal of the Linnean Society*, **166**, 127–148.

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

Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, **23**, 564–571.

Qiu Y-X, Fu C-X, Comes HP (2011) Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Molecular Phylogenetics and Evolution*, **59**, 225–244.

Quèbre N (2012) *Altitudinal Adaptation in Arabidopsis Thaliana*. Diss. Eidgenössische Technische Hochschule ETH, Zürich.

Ralska-Jasiewiczowa M (2004) *Late Glacial and Holocene History of Vegetation in Poland Based on Isopollen Maps*. W. Szafer Institute of Botany, Polish Academy of Sciences, Krakow.

Robin V, Nadeau M-J, Grootes PM, Bork H-R, Nelle O (2016) Too early and too northerly: evidence of temperate trees in northern Central Europe during the Younger Dryas. *New Phytologist*. doi: 10.1111/nph.13844.

Ronikier M (2011) Biogeography of high-mountain plants in the Carpathians: an emerging phylogeographical perspective. *Taxon*, **60**, 373–389.

Ronikier M, Cieślak E, Korbecka G (2008) High genetic differentiation in the alpine plant *Campanula alpina* Jacq. (Campanulaceae): evidence for glacial survival in several Carpathian regions and long-term isolation between the Carpathians and the Alps. *Molecular Ecology*, **17**, 1763–1775.

Schmickl R, Koch MA (2011) *Arabidopsis* hybrid speciation processes. *Proceedings of the National Academy of Sciences*, **108**, 14192–14197.

Schmickl R, Paule J, Klein J, Marhold K, Koch MA (2012) The evolutionary history of the *Arabidopsis arenosa* complex: diverse tetraploids mask the western Carpathian center of species and genetic diversity. *PLoS One*, **7**, e42691.

Schmitt T, Varga Z (2012) Extra-Mediterranean refugia: the rule and not the exception? *Frontiers in Zoology*, **9**, 22.

Schönswetter P, Tribsch A (2005) Vicariance and dispersal in the Alpine perennial *Bupleurum stellatum* L. (Apiaceae). *Taxon*, **54**, 725–732.

Schönswetter P, Stehlik I, Holderegger R, Tribsch A (2005) Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology*, **14**, 3547–3555.

Schraiber JG, Akey JM (2015) Methods and models for unravelling human evolutionary history. *Nature Reviews Genetics*, **16**, 727–740.

Schultes JA (1814) *Österreichische Flora. Ein Handbuch auf Botanischen Excursionen Enthaltend Eine Kurze Beschreibung der in den Erbstaaten des Österreichischen Kaiserthumes Wildwachsenden Pflanzen*. C. Schaumburg und Comp., Wien.

Shimizu-Inatsugi R, Lihova J, Iwanaga H *et al.* (2009) The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. *Molecular Ecology*, **18**, 4024–4048.

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, **16**, 608–613.

Stone JL, Crystal PA, Devlin EE, Downer RHL, Cameron DS (2012) Highest genetic diversity at the northern range limit of the rare orchid *Isotria medeoloides*. *Heredity*, **109**, 215–221.

Surina B, Schönswetter P, Schneeweiss GM (2011) Quaternary range dynamics of ecologically divergent species (*Edraianthus serpyllifolius* and *E. tenuifolius*, Campanulaceae) within the Balkan refugium. *Journal of Biogeography*, **38**, 1381–1393.

Svendsen JI, Alexanderson H, Astakhov VI *et al.* (2004) Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews*, **23**, 1229–1271.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Těšitel J, Malinová T, Štech M, Herbstová M (2009) Variation in the *Melampyrum sylvaticum* group in the Carpathian and Hercynian region: two lineages with different evolutionary histories. *Preslia*, **81**, 1–22.

Thiel-Egenter C, Alvarez N, Holderegger R *et al.* (2011) Break zones in the distributions of alleles and species in alpine plants. *Journal of Biogeography*, **38**, 772–782.

Tollefsrud MM, Kissling R, Gugerli F *et al.* (2008) Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen. *Molecular Ecology*, **17**, 4134–4150.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics*, **42**, 260–263.

Tzedakis PC (2004) The Balkans as prime glacial refugial territory of European temperate trees. In: *Balkan Biodiversity*(eds Griffiths HI, Kryštufek B, Reed JM), pp. 49–68. Springer, Netherlands.

Tzedakis PC, Emerson BC, Hewitt GM (2013) Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, **28**, 696–704.

Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*. Springer, Berlin.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Weiss S, Ferrand N (2007) *Phylogeography of Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity*. Springer, Berlin.

Widmer A, Lexer C (2001) Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology & Evolution*, **16**, 267–269.

Wielstra B, Babik W, Arntzen JW (2015) The crested newt *Triturus cristatus* recolonized temperate Eurasia from an extra-Mediterranean glacial refugium. *Biological Journal of the Linnean Society*, **114**, 574–587.

Willis KJ, van Andel TH (2004) Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews*, **23**, 2369–2387.

Winkler M, Tribsch A, Schneeweiss GM *et al.* (2012) Tales of the unexpected: phylogeography of the arctic-alpine model plant *Saxifraga oppositifolia* (Saxifragaceae) revisited. *Molecular Ecology*, **21**, 4618–4630.

Wohlfarth B, Bjorck S, Funder S *et al.* (2008) Quaternary of Norden. *Episodes*, **31**, 73–81.

Wołoszczak E (1896) Z granicy flory zachodnio- i wschodnio-karpackiej. *Spraw Kom Fizjogr Ak Um*, **31**, 119–159.

Wright KM, Arnold B, Xue K *et al.* (2015) Selection on meiosis genes in diploid and tetraploid *Arabidopsis arenosa*. *Molecular Biology and Evolution*, **32**, 944–955.

Yant L, Hollister JD, Wright KM *et al.* (2013) Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Current Biology*, **23**, 2151–2156.

Zając M, Zając A (2011) Directional northern element in the flora of vascular plants of Poland. *Biodiversity: Research and Conservation*, **21**, 19–26.

Zieliński P, Nadachowska-Brzyska K, Wielstra B *et al.* (2013) No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (*Lissotriton montandoni*). *Molecular Ecology*, **22**, 1884–1903.

F.K., K.M. and E.Z. designed the study and, together with G.F. and M.L., collected the samples. G.F., E.Z., A.N. and L.H. performed the laboratory work. F.K. performed the data analyses. F.K., K.M. and H.K. drafted the manuscript with contributions of all authors. All

authors have revised and approved the final manuscript.

Data accessibility

The matrices of SNP data formatted as input files for particular programs are included as Appendix S1–S4 of the Supporting Information; microsatellite matrix is in Appendix S5 of the Supporting Information. Raw reads and are available at GenBank (project PRJNA301691).

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Matrix of SNP data formatted as input files for STRUCTURE.

Appendix S2 Matrix of SNP data formatted as input files for SNAPP - subset 1.

Appendix S3 Matrix of SNP data formatted as input files for SNAPP - subset 2.

Appendix S4 Matrix of SNP data formatted as input files for DIYABC.

Appendix S5 Microsatellite data matrix.

Table S1 Details on localities and genetic diversity measures (inferred from microsatellite data) of the 64 populations of *Arabidopsis arenosa* and two of *A. croatica* included in the study.

Table S2 Pairwise FST among major genetic groups of diploid *Arabidopsis arenosa* inferred from 10,955 SNPs and 14 microsatellite loci above and below the diagonal, respectively.

Table S3 Details on the individual bioclimatic factors used in the linear discriminant analysis of the 117 ecologically-screened diploid *A. arenosa* populations divided into the five major genetic groups.

Table S4 Results of the classificatory discriminant analysis evaluating the success of classification of the 117 ecologically-screened diploid *A. arenosa* populations into five major genetic groups based on eight forward-selected bioclimatic variables (bio 3, bio4, bio9, bio12–bio15).

Fig. S1 Summary of STRUCTURE analyses of SNP datasets of diploid *Arabidopsis arenosa*.

Fig. S2 The traditional taxonomy of diploid *Arabidopsis arenosa* group does not correspond with major trends in genetic differentiation.

Fig. S3 Finer sub-structuring within each of the four major clusters of diploid *Arabidopsis arenosa* marked by different colour shades.

Fig. S4 Phylogenetic relationships of major *A. arenosa* groups (rooted with *A. croatica*) inferred from multispecies coalescent analysis of 2,313 (a, c) and 2,213 (b, d) unlinked SNPs using SNAPP.

Fig. S5 Folded allele frequency spectra (AFS) inferred from SNP data.

Fig. S6 Relationship among Nei's (1972) population genetic distances and ecological distances, i.e. Euclidean distances of populations characterized by all investigated bioclimatic, substrate and topographical variables.

Fig. S7 Relationships among genetic variation and ecological preferences of the SNP-genotyped populations of diploid *Arabidopsis arenosa*.

Fig. S8 Location of the 171 ecologically-screened populations used in tests of ecological differentiation among major genetic groups (shown in corresponding colours).

Appendix S6 Methods.

Paper III

Šrámková-Fuxová G, Záveská E, Kolář F, Lučanová M, Španiel S, Marhold K (2017) Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction, *Botanical Journal of the Linnean Society*, Volume 185, Issue 3, Pages 321–342.

Botanical Journal of the Linnean Society, 2017, **185**, 321–342. With 5 figures.

Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction

GABRIELA ŠRÁMKOVÁ-FUXOVÁ[1†], ELIŠKA ZÁVESKÁ[2†], FILIP KOLÁŘ[1,3†], MAGDALENA LUČANOVÁ[1,3], STANISLAV ŠPANIEL[1,4] and KAROL MARHOLD[1,4*] FLS

[1]*Department of Botany, Faculty of Science, Charles University, CZ-128 01 Prague, Czech Republic*
[2]*Institute of Botany, University of Innsbruck, AT-6020 Innsbruck, Austria*
[3]*Institute of Botany, The Czech Academy of Sciences, CZ-252 43 Průhonice, Czech Republic*
[4]*Institute of Botany, Plant Science and Biodiversity Centre, Slovak Academy of Sciences, SK-845 23 Bratislava, Slovak Republic*

Received 18 March 2017; revised 4 July 2017; accepted for publication 18 August 2017

Wild relatives of *Arabidopsis thaliana* are moving into the spotlight of plant evolutionary biologists and molecular geneticists, but patterns of genetic diversity and phenotypic variation in natural populations are often overlooked. This gap in knowledge may in turn hamper generalizations of results from experimental studies using populations of unclear evolutionary background. Here we present a comprehensive assessment of the genetic structure and morphological variation of *Arabidopsis halleri*, a model species for studying heavy metal tolerance and phytoremediation. Based on extensive sampling of 768 individuals from 82 populations across the entire distribution of the species, genotyping using multiple molecular markers (AFLP, nuclear microsatellites and sequences of single-copy nuclear regions and plastid DNA) and phenotyping by multivariate morphometrics, we aimed to reconstruct the range-wide phylogeography and morphological trait evolution in *A. halleri* populations. In addition, we address general biogeographical questions related to the origin of the striking Northern Hemisphere disjunction (Europe–East Asia) and glacial survival in extra-Mediterranean refugia in Europe. East Asian (Japanese) populations were genetically distinct and slightly depauperate, but their divergence was at levels comparable to major splits within Europe, rejecting both an ancient (old vicariance) and recent (human-mediated spread) origin of the Northern Hemisphere disjunction. In Europe we detected three major genetic lineages of *A. halleri*, corresponding well with geography (Western–Central Europe, the Alps and the south-eastern Carpathians + the Balkans). Sequence-based divergence estimates indicated a probable Pleistocene origin of these three lineages. This, together with elevated diversity and rarity within each group, suggests *in situ* glacial persistence of *A. halleri* in multiple northern refugia of eastern Central Europe. The extensive morphological variation of European *A. halleri* populations only partly correlated with genetic structure. Rather, it was driven by local environmental characteristics. This suggests a remarkably plastic response of the species to major environmental gradients, manifested by the parallel origin of a distinct alpine phenotype.

ADDITIONAL KEYWORDS: Cruciferae – extra-Mediterranean refugia – DNA sequencing – parallel origin of alpine phenotype – phylogeography.

INTRODUCTION

Knowledge of patterns of natural variation across the range of a species is a necessary prerequisite for addressing general questions on how intraspecific diversity evolves in space and time. The recent burst of new sequencing techniques has allowed various evolutionary questions using natural populations of model species and their close relatives to be addressed. Knowledge of the demographic history of such species across their native ranges, however, still remains fragmentary or has become a study topic only in recent years (e.g. Arnold, Kim & Bomblies, 2015;

*Corresponding author. E-mail: karol.marhold@savba.sk
†These authors contributed equally to the study.

The 1001 Genomes Consortium, 2016), although such understanding may be crucial for interpreting experimental results in a broader evolutionary framework. *Arabidopsis* Heynh. has been the leading model genus in plant biology for many decades (Koenig & Weigel, 2015), but researchers have only recently started revealing the natural variation and evolutionary relationships of *Arabidopsis* spp. across their distribution ranges (Pauwels *et al.*, 2012; Hohmann *et al.*, 2014; Arnold *et al.*, 2015; The 1001 Genomes Consortium, 2016; Kolář *et al.*, 2016a).

Arabidopsis halleri (L.) O'Kane & Al-Shehbaz, one of the closest relatives of *A. thaliana* (L.) Heynh. (Clauss & Koch, 2006; Novikova *et al.*, 2016), is an important model for the study of hyperaccumulation of heavy metals and phytoremediation (Bert *et al.*, 2002; Willems *et al.*, 2007; Verbruggen, Hermans & Schat, 2009; Stolpe, Krämer & Müller, 2017). This outcrossing perennial herb prefers moist grassland and woodland habitats in the montane zone of temperate areas of Eurasia. Its native area splits into two areas (mountain zones of Central and south-eastern Europe and mountains of East Asia, mainly Japan, but reported also from some localities in Korea, the Russian Far East and neighbouring provinces of China) that are separated by a large distribution gap spanning the whole of Western and Central Asia (Berkutenko, 1988; Jalas & Suominen, 1994; Zhou *et al.*, 2001; Al-Shehbaz & O'Kane, 2002; Hoffmann, 2005; Pak, 2005; Al-Shehbaz, Arai & Ohba, 2006). The origin of this disjunction is an interesting issue on its own. Much attention has been paid to East Asian–North American disjunctions (Donoghue, Bell & Li, 2001, and references therein), but little is known about the origins of the European–Eastern Asian disjunction (but see, e.g., Carlson, Linder & Donoghue, 2012). The species is well known to spread into man-made sites contaminated by heavy metals and into river alluvia. Thus, *A. halleri* has a wide ecological amplitude spanning a range of substrates from the lowlands to the subalpine zone (Jones & Akeroyd, 1993; Kubota *et al.*, 2015). Besides genus-wide studies documenting the monophyly of the species across its disjunct area (Hohmann *et al.*, 2014; Novikova *et al.*, 2016), there has been no comprehensive evaluation of the intraspecific diversity in a species over its entire range. In Japan, Shimizu-Inatsugi *et al.* (2009) addressed its role in the origin of a new allopolyploid species and several other studies have addressed patterns of genetic variation in the European part of the *A. halleri* range. Variation in plastid DNA (large-scale genotyping of a panel of known variants) indicated the presence of two major groups in Europe (Pauwels *et al.*, 2005, 2008), a pattern that was later supported by nuclear microsatellites (Pauwels *et al.*, 2012). A recent study screening plastid variation detected an additional split between the Western and Eastern Carpathians, but the distribution of the eastern lineage is unknown due to missing sampling from most of the south-eastern Carpathians (Wasowicz *et al.*, 2016). Despite these significant contributions, we still cannot draw a complete picture of the natural variation in *A. halleri* across its range because sampling efforts have been skewed towards metal-contaminated sites, the almost complete lack of sampling from eastern and south-eastern Europe, i.e. areas well known for their elevated species and genetic diversity (Kryštufek, 2004; Mráz & Ronikier, 2016) and the absence of a range-wide study revealing the relationships between European and East Asian populations. In addition, apart from local taxonomic assessments (Měsíček & Goliašová, 2002; Kolník & Marhold, 2006) we lack information on the extent of range-wide phenotypic morphological variation of natural populations and its major ecogeographical drivers. Consequently, there is still controversy in the taxonomic classification of the species (Kolník & Marhold, 2006; Hohmann *et al.*, 2014). This may bring practical problems when studying this species across the broad *Arabidopsis* research community, such as inconsistent identification of accessions in experimental studies.

To fill in these gaps, we described the genetic structure of *A. halleri* across its entire distribution, based on extensive range-wide sampling of natural populations. Using genome-wide AFLP data, nuclear microsatellites and sequences from one plastid and one nuclear region, we reconstructed phylogenetic relationships, revealed range-wide patterns of genetic diversity and inferred the evolutionary history of *A. halleri*. In addition, we addressed specific biogeographical hypotheses regarding (1) the origin of the phytogeographical disjunction between Europe and East Asia and (2) glacial survival in extra-Mediterranean refugia such as montane regions of Central Europe. Finally, we asked whether the described genetic structure corresponds with morphological variation (suggesting morphological separation of the major groups) and/or ecological conditions (suggesting niche differentiation of the lineages) of natural populations or, alternatively, whether *A. halleri* phenotypes reflect local environmental conditions.

MATERIAL AND METHODS

Plant material

Plant material was collected between 2011 and 2013, aiming to achieve homogeneous sampling across the entire distribution area of *A. halleri* in Europe, and was complemented by seven populations spanning most of the distribution of the species in Japan. Sampled

populations thus included multiple representatives of all five hitherto recognized subspecies [*A. halleri* subsp. *halleri*, *A. halleri* subsp. *tatrica* (Pawł.) Kolník, *A. halleri* subsp. *dacica* (Heuff.) Kolník, *A. halleri* subsp. *ovirensis* (Wulfen) O'Kane & Al-Shehbaz and *A. halleri* subsp. *gemmifera* (Matsum.) O'Kane & Al-Shehbaz] (Kolník & Marhold, 2006; Hohmann *et al.*, 2014). Our dataset comprised 768 individuals from 82 natural populations (see Supporting Information, Table S1 for locality details). Individual plants were sampled throughout their sites at maximum distances, but no less than 3 m apart, to minimize the chance of collecting clones (Van Rossum *et al.*, 2004). Leaf tissue from five to 15 (ten on average) plants per population was collected and quickly desiccated in silica gel. Fresh flowering stems from the same individuals were also collected for flow cytometry. Finally, from a subset of 41 European populations, we collected specimens for morphometric analyses. Vouchers are deposited in the Herbarium of Charles University in Prague (PRC).

Analytical approach

To clarify relationships among intraspecific entities of *A. halleri*, we employed a phylogeographical approach instead of classical phylogenetic methods, which, being designed for historical relationships at the species level, often assume that taxonomic units represent non-reticulate lineages (e.g. Schaal *et al.*, 1998). We thus used a combination of several independent genetic markers that may show different levels of genetic divergence: AFLP and microsatellite data (reflecting a more recent population genetic structure) and plastid and nuclear DNA sequences (that might still retain traces of the common ancestor of currently diversified lineages; e.g. Eidesen *et al.*, 2007; Kolář, Dušková & Sklenář, 2016b). Firstly, we identified major genetic groupings in the data (ancestral lineages from which current populations are derived) using AFLP markers and estimated levels of genetic diversity within and among groups and populations, using both AFLPs and microsatellites. Then, we reconstructed gene tree genealogies from DNA sequences and, using a multi-species coalescent model, a population tree in order to reconstruct hierarchical relationships among those major groups and to estimate their times of divergence.

Flow cytometry

Individuals from all 82 populations were analysed using flow cytometry (FCM) to obtain DNA ploidy (72 populations) and absolute nuclear genome size (53 populations) data. For details about analysed individuals, see Supporting Information, Table S1. For DNA ploidy estimations, individuals from 72 populations (910 individuals, in total, one to 23 per population, 13 on average) were analysed using a simplified two-step protocol (Doležel, Greilhuber & Suda, 2007) with Otto buffers optimized for *Arabidopsis* as described by Kolář *et al.* (2016a). Briefly, *c.* 10 mm^2 sample tissue (preferably petals in order to avoid endopolyploidy) was chopped together with the same amount of tissue of *Solanum pseudocapsicum* L., which was used as the internal standard [2C = 2.59 pg, Temsch, Greilhuber & Krisai (2010); only in a few cases we used *Glycine max* (L.) Merr. 'Polanka', 2C = 2.50 pg, Doležel, Doleželová & Novák (1994)] in 0.5 mL Otto I buffer (0.1 M citric acid, 0.5% Tween 20). After filtration we added 1 mL Otto II buffer (0.4 M $Na_2HPO_4.12H_2O$) enriched with 4 μg mL^{-1} of the fluorochrome 4,6′-diamino-2-phenylindole (DAPI) and 2 μg mL^{-1} β-mercaptoethanol. After 10 min of staining, we measured the relative fluorescence of 3000 particles in each sample, using a Cyflow ML flow cytometer (Partec) equipped with a UV-led lamp.

In addition, a set of 85 individuals from 53 populations (partly new and partly overlapping with those analysed using DAPI, see Supporting Information, Table S1) was subjected to absolute genome size estimation using a similar protocol but with propidium iodide (PI) staining with the following modifications. (1) Instead of DAPI we added PI and RNase IIA (both at final concentrations of 50 μg mL^{-1}) and β-mercaptoethanol (at 2 μg mL^{-1}) to the Otto II buffer. (2) Analyses were done using a CyFlow SL flow cytometer (Partec) equipped with a green (532 nm) solid-state laser. (3) All samples were run for 5000 particles, preferably three times on different days, to minimize between-day variation, and the resulting values per individual were averaged. The resulting cytometric histograms were analysed in FloMax (Partec).

AFLP genotyping

In total, 768 individuals from 82 populations in Europe and Japan (two to 15 plants per population, nine on average) were genotyped by AFLP analysis (Vos *et al.*, 1995). We followed the protocol provided by Applied Biosystems (2010) with some modifications described here. Double-digestion of DNA (*c.* 300 ng) was performed using the *Eco*RI and *Mse*I enzymes at 37 °C for 3 h. The reaction mix (6 μL volume) contained 3 U *Eco*RI (Thermo Scientific), 1.2 U *Mse*I (New England Biolabs) and 1.2 μL 10× Tango buffer. Adaptors were ligated to the digested fragments by adding 3 μL ligation mix containing 0.6 U T4 DNA ligase (Thermo Scientific), 0.9 μL T4 DNA ligase buffer (including ATP) and 0.6 μL each adaptor pair (Sigma), and the reaction (total volume 9 μL) was incubated for 12 h at 16 °C. Ligated DNA fragments were diluted 1:4 with ddH_2O. Preselective amplification reactions (total volume 6.25 μL) contained 1.25 μL restricted/ligated

DNA, 4 μL Pre-Amp Primer Mix I (Invitrogen), 0.6 μL 10× buffer for AmpliTaq Gold (Applied Biosystems), 0.35 μL 25 mM magnesium chloride and 0.25 U AmpliTaq Gold DNA Polymerase (5 U μL^{-1}; Applied Biosystems). The PCR cycle profile was 95 °C (7 min), 25 cycles with 94 °C (30 s), 56 °C (30 s), 72 °C (2 min), followed by 60 °C for 30 min. Selective amplification was performed using 5 μL 10× diluted pre-amplification product as a template, 1 μL 10× buffer for AmpliTaq Gold, 2 mM dNTP, 1 μL 25 mM magnesium chloride, 0.5 pmol *Eco*RI-selective fluorescence-labelled primer, 2.5 pmol *Mse*I-selective primer and 0.5 U AmpliTaq Gold DNA Polymerase (5 U μL^{-1}) in a total volume of 10 μL. The following four pairs of selective primers were used: *Eco*RI-ATC-(6-FAM)/*Mse*I-CTT, *Eco*RI-AAG-(VIC)/*Mse*ICAC, *Eco*Ri-ACC-(NED)/*Mse*I-CAA and *Eco*RI-AAC-(PET)/*Mse*I-CTG. Fragments were resolved on an ABI 3130x1 Avant Genetic Analyzer (Applied Biosystems) using the GeneScan-LIZ-600 size standard (Applied Biosystems). For 120 samples (15%), the whole AFLP protocol from isolation of DNA onwards was repeated to test the reproducibility of the method (Bonin *et al.*, 2004). Unambiguous fragments in the range of 50–500 bp were scored regardless of their intensity (Tribsch, Schönswetter & Stuessy, 2002) using GeneMarker 1.8 (Softgenetics LLC). The resulting data matrix is available as Supporting Information, Data S1.

Microsatellite amplification

We genotyped five unlinked microsatellite loci (ICE13, NGA, ATT, ICE14 and ATH), previously used in a population genetic study of *A. halleri* (Clauss, Cobban & Mitchell-Olds, 2002), in the same 82 populations as in those used for AFLP genotyping, and 765 individuals were analysed (two to 15 plants per population, nine on average). Amplifications were performed in a multiplex PCR assay using the QIAGEN Multiplex PCR Kit. Forward primers were fluorescently labelled with 6-FAM, NED, PET or VIC fluorescent dyes. Primer concentrations ranged from 0.075 to 0.1 μM (see Supporting Information, Table S2 for details regarding individual loci). Thermocycling conditions were 95 °C for 15 min; 35 cycles of 94 °C for 30 s, 60 °C for 90 s and 72 °C for 60 s; and a final extension of 60 °C for 30 min. Amplification products were separated using a 3130xl Genetic Analyzer (DNA Sequencing Laboratory, Faculty of Science, Charles University) using a GeneScan 600 LIZ (Life Technologies) as the internal size standard. Fragment sizes were scored automatically by GeneMarker 1.8.0 (SoftGenetics) using respective panels for each locus, and each sample was then checked manually. The resulting data matrix is available in the Supporting Information, Data S2.

Plastid and nuclear low-copy gene DNA sequencing

Primers for three variable plastid DNA regions (further named as *cp3*, *cp4*, *cp6*) have been designed based on the set of plastomes of five geographically different accessions of *A. halleri* (Novikova *et al.*, 2016) using Prime3Plus (Untergasser *et al.*, 2007; for primer sequences see Supporting Information, Table S2). Initially, we sequenced a test dataset of 17 individuals spanning the variation range of *A. halleri* (see Supporting Information, Table S1 for samples included in the test dataset) and built an alignment in MAFFT v.6 (Katoh & Standley, 2013). Then we selected a single region (*cp3*) that exhibited the greatest number of variable sites in the test dataset. Only *cp3* was further used for the haplotyping of 99 individuals of *A. halleri* and ten individuals of its closest relatives [*A. cebennensis* (DC.) O'Kane & Al-Shehbaz, *A. arenosa* (L.) Lawalrée, *A. croatica* (Schott) O'Kane & Al-Shehbaz and *A. thaliana* (L.) Heynh.]. As an additional marker, we used a nuclear gene which encodes the enzyme chalcone synthase (*CHS*; EC 2.3.1.74) and has been successfully used to infer the phylogeny of tribe Arabideae, including multiple *Arabidopsis* spp. (Koch, Haubold & Mitchell-Olds, 2000). The primers CHS-FOR1 (5′-CTTCATCTGCCCGTCCATCTAACC-3′) (promoter specific) and CHS-REV (5′-GGAACGCTGTGCAAGAC-3′) in exon 2 from the same study (Koch *et al.*, 2000) were used to amplify the locus. PCRs of all loci were done with MyTaq polymerase (Bioline) following the manufacturer's instructions but downscaled to one-third volume reactions and with an annealing temperature of 60 °C for all loci. PCR products were checked on 1% agarose gels and sequenced by Macrogen from both DNA strands, using the original PCR primers. The *CHS* marker was sequenced from 38 *A. halleri* individuals and nine individuals of its closest relatives (*A. cebennensis*, *A. arenosa*, *A. croatica* and *A. thaliana*). Details of all individuals from which the *cp3* and *CHS* regions were sequenced are given in Supporting Information, Table S1.

Analysis of AFLP data

To detect clustering of AFLP genotypes, we used Structure 2.3.2 (Pritchard, Stephens & Donnelly, 2000; Falush, Stephens & Pritchard, 2007). The admixture model with uncorrelated allele frequencies and recessive alleles (i.e. 0) was run ten times for each K (number of groups) ranging from 1 to 10, using a burn-in of 100 000 iterations followed by 1 000 000 additional Markov chain Monte Carlo iterations. The analyses were performed separately for (1) the entire dataset of European and Asian *A. halleri* and (2) only European

accessions in order to identify structuring among the intensively sampled European accessions. The estimation of the optimal number of groups as the value of K was calculated using the R script Structure-sum (Ehrich, 2007), which compares the posterior probabilities (PP) of the runs (Rosenberg *et al.*, 2002), the similarity coefficient between the runs (Nordborg *et al.*, 2005) and delta K (Evanno, Regnaut & Goudet, 2005). The optimal value of K was defined as the partition where the curve of likelihood started to flatten out, where the result of replicate runs was similar and where the cluster was not empty. We plotted genetic distances among individuals using principal coordinate analysis based on Jaccard distance (PCoA) calculated in adegenet v1.4-2 (Jombart, 2008).

We quantified the partitioning of genetic variation within and among populations in the whole dataset as well as within each of the major genetic groups identified by Structure using analysis of molecular variance (AMOVA) calculated in Arlequin 3.5.1.2 (Excoffier & Lischer, 2010). Hierarchical AMOVA was used to estimate the levels of genetic differentiation among major groups. We calculated Nei's gene diversity (Nei, 1987) and the frequency-down-weighted marker index (DW; Schönswetter & Tribsch, 2005) from a presence–absence matrix of alleles using the R script *AFLPdat* (Ehrich, 2006) for each population and for each of the major geographical groups identified with Structure. The DW value is expected to be greater in populations that harbour a high number of rare alleles, i.e. alleles with low frequency in the total dataset (e.g. Paun *et al.*, 2008). Differences in diversity and the DW index between the three main groups were tested by one-way ANOVA (R package *stats*). Finally, we tested for a significant correlation among matrices of genetic [chord distances derived from AFLP fragment frequencies (Zhivotovsky, 1999), as implemented in FAMD 1.31 (Schlüter & Harris, 2006)] and Euclidean geographical distances between populations (isolation by distance) using the Mantel test as implemented in the *adegenet* package.

Analysis of microsatellite data

As a complementary approach to the dominant AFLP markers, we used five microsatellite loci to characterize the patterns of within-population diversity and among-population differentiation. For each population, we calculated observed (H_o) and expected heterozygosity (H_e; both averaged over loci variable in that population) and average number of alleles (allelic richness computed through rarefaction on the small sample size of a minimum of six individuals, 1000 permutations) in MSA v4.05 (Dieringer & Schlötterer, 2003). Differences in heterozygosity indices, allelic richness and the DW index between the three main groups were tested by one-way ANOVA (R package *stats*).

We inferred the differentiation among populations in the whole dataset and within each of the three major groups separately using the fixation index F_{ST} (Weir & Cockerham, 1984) and its standardized extension for multi-allelic states (G'_{ST}, Hedrick, 2005) calculated in MSA. In addition, we quantified the partitioning of genetic variation within and among populations in the whole dataset and within each of the major genetic groups by AMOVA calculated in Arlequin 3.5.1.2 (Excoffier & Lischer, 2010).

Analysis of plastid DNA

Sequences were proofread and assembled into contigs using Geneious v.7.0.6 (Kearse *et al.*, 2012) and aligned in MAFFT v.6 (Katoh & Standley, 2013) and the alignments were manually improved in BioEdit v.7.0.0. A plastid DNA haplotype network was constructed using TCS 1.21 (Clement, Posada & Crandall, 2000) based on substitution variation and indel variation of the dataset including only *A. halleri* individuals. Indels were coded in a simple coding manner (Simmons & Ochoterena, 2000) using SeqState (Müller, 2005) with subsequent transformation of the resulting 0–1 indel matrix to an A–T matrix that was added to the indel-free nucleotide alignment, in order to include the presence/absence indel information in the TCS analysis. The TCS analysis was then run with default settings.

To identify hierarchical relationships among plastid haplotypes, we performed phylogenetic analyses using BEAST v.1.7.5 (Drummond & Rambaut, 2007; Drummond *et al.*, 2012), together with additional individuals of the closely related species *A. cebennensis*, *A. arenosa*, *A. croatica* and *A. thaliana* as outgroups. Additionally, we constructed a maximum likelihood (ML) tree in MEGA 5.2.1 (Tamura *et al.*, 2011) with 1000 bootstrap replicates to support estimations based on the Bayesian approach. The fit of nucleotide substitution models was assessed using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) as implemented in JMODELTEST 0.1.1 (Posada, 2008).

Divergence dating analysis of the *CHS* gene tree and the multi-species coalescent tree

To infer approximate divergence dates of *A. halleri* and its subgroups, we first inferred a scaled *CHS* gene tree in BEAST v.1.7.5, including our own sampling of *A. halleri* and its closest relatives (*A. cebennensis*, *A. arenosa*, *A. croatica* and *A. thaliana*) and several samples from Koch *et al.* (2000), one accession of *A. halleri* (GenBank number gi_6684388), three accessions of *A. lyrata* (GenBank numbers gi_6684393, gi_6684396, gi_6684397) and one accession of *A. thaliana* (GenBank number gi_6684379). As a prior for the

gene tree calibration, we used a secondary calibration point of divergence of *Arabidopsis* as inferred by Koch *et al.* (2000) based on the same *CHS* region. In particular, we set the divergence of the *Arabidopsis* crown group to obtain a lognormal distribution with a mean of 5.2 Mya, with a standard deviation of 0.4 and an offset of 0.5. To infer the topology by an alternative approach, we inferred an ML tree also for this dataset, using MEGA 5.2.1 as described above.

Secondly, we employed plastid DNA and *CHS* markers to construct a time-calibrated species tree under a multispecies-coalescent model using *BEAST (Heled & Drummond, 2010). This analysis also enabled us to directly compare divergence times of plastid DNA and *CHS* gene trees and to evaluate the presence/extension of ancestral polymorphisms in both datasets. Ingroup *A. halleri* populations were *a priori* assigned to four 'species' (significantly distinguishable genetic groups as identified by Structure analysis of AFLP data), namely Japanese populations (JP, *A. halleri* subsp. *gemmifera*), the north-western group (NW), the south-eastern group (SE) and the Alpine group of *A. halleri*. In addition, we included three outgroup species: *A. cebennensis*, *A. croatica* and *A. thaliana*.

For the species tree calibration, we used the following priors. First, the prior for the substitution rate of plastid DNA was applied with wide margins to cover most biologically realistic values by covering previously published plastid DNA rates (Wolfe, Li & Sharp, 1987; Winkler *et al.*, 2012). We therefore used a lognormal distribution with a mean of 0.007 and a standard deviation of 1.0, which ensured the median value to be around 4×10^{-3} substitutions per site per My. Secondly, for the diversification of the genus *Arabidopsis*, we set the prior to be a uniform distribution from 0 to 9 Mya. This was in accordance with the results of Koch *et al.* (2000), based on divergence times computed from *CHS* and *ADH* synonymous substitution rates. In *BEAST the rate of one marker has to be fixed, while the rates of the other marker(s) are estimated and scaled according to the fixed one. In our settings, the rate of the *CHS* marker was fixed whereas the rate of plastid DNA was estimated, because the fixed marker is suggested to be the one with the stronger phylogenetic signal (A. J. Drummond, pers. comm.). To get reasonable values of the plastid DNA rate while the *CHS* rate was fixed, we ran three alternative runs of the *BEAST analysis with a different fixed rate of the *CHS* marker: (1) 1.5×10^{-2} substitutions per site per My (Koch *et al.*, 2000); (2) 9×10^{-3} substitutions per site per My; and (3) 7×10^{-3} substitutions per site per My to cover the range of *CHS* mutation rates suggested by Wolfe, Sharp & Li (1989) for maize versus barley and Durbin *et al.* (1995) for *Ipomoea* L. The resulting log files of three runs were compared to evaluate the co-estimated plastid DNA rate, and the analysis where the distribution of plastid DNA rate was closest to the prior settings (i.e. lognormal distribution with a mean of 0.007 and a standard deviation of 1.0 as defined above) was selected for interpretation of the results. The rate of evolution was modelled in a strict clock framework, as the null hypothesis of an equal evolutionary rate throughout the tree was not rejected at the 5% significance level ($P < 0.99$; computation performed with MEGA 5.05, with a test of molecular clocks using ML, Tamura *et al.*, 2011).

Similarly for all dating analyses, four independent chains were run for a total of 100 000 000 generations. Log files were analysed using TRACER v1.5 to assess convergence and ensure that the effective sample size (ESS) for all parameters was > 200. The resulting trees were combined using LogCombiner v.1.7.2 with a burn-in of 25%. Subsequently, a maximum credibility tree was constructed using TreeAnnotator v.1.7.2.

Analysis of ecological and morphological data

We searched for concordance between genetic structure of the 75 European AFLP-genotyped populations and the environmental conditions of their original localities using constrained analysis of principal coordinates (CAP), which is a multivariate extension of multiple regression based on distances between objects (Anderson & Willis, 2003). We calculated CAP models from a matrix of genetic distances (among-population genetic chord distances of AFLP fragment frequencies inferred in FAMD 1.31; Schlüter & Harris, 2006) by the *capscale* procedure in vegan 2.3-0 (Oksanen *et al.*, 2015) and tested (1000 permutations) for effects of elevation and two composite climatic parameters. These latter two parameters were represented by scores on the first axis of a separate unconstrained ordination (principal component analysis) of the 11 temperature-linked and eight precipitation-linked bioclimatic variables (derived from Worldclim data with *c.* 1-km resolution; 19 bioclimatic variables, Hijmans *et al.*, 2005), respectively. We addressed two CAP models, the first based on environmental variables only and the second estimating the explanatory power of environmental variables conditioned by geographical distances among the populations. We transformed the Euclidean matrix of among-population geographical distances into a rectangular matrix using principal coordinates of a neighbourhood matrix (PCNM; Borcard & Legendre, 2002), using the *pcnm* function in the vegan package with the default truncation threshold and supplied the first two PCNM vectors as conditional variables for the CAP model.

In addition, we examined the relative contributions of the environment, genetic structure and geographical distances to the European range-wide morphological variation of 41 *A. halleri* populations. We used a

matrix of 33 morphological characters (see Supporting Information, Table S5 for details) that were scored on vouchers of field-collected individuals (10–49 individuals per population, 20 on average, 825 in total) and averaged over a population. Both environmental (19 bioclimatic variables) and morphological Euclidean distances were calculated from variables that were scaled to zero mean and unit variance; genetic distances were again represented by among-population genetic chord distances (see above). We used two analytical approaches to characterize the drivers of morphological variation. Firstly, using the Mantel test with 1000 permutations, we tested for correlations of either environmental or genetic distances on the one hand and morphological distances on the other (with and without conditioning for geographical distances using the functions *mantel* and *mantel.partial* in vegan, respectively). Secondly, we tested for the marginal effects of the environment, genetic structure and geographical distances on morphology and the unique contributions of these factors, using direct ordination [redundancy analysis (RDA) function *rda* in vegan, tested by 1000 Monte Carlo permutations] and quantified their relative contribution through variation partitioning (function *varpart* in vegan). The scaled matrix of morphological characters was constrained by the following predictors: (1) elevation and scores on the first axis of separate unconstrained ordinations (principal component analyses) of the 11 temperature-linked and eight precipitation-linked bioclimatic variables (environmental data); (2) scores on the first two PCNM vectors derived from a matrix of genetic; and (3) geographical distances. We performed this set of analyses for the entire dataset and separately for populations belonging to each of the three major genetic groups in Europe.

RESULTS

Flow-cytometric analysis documented nuclear DNA content values corresponding to diploid chromosome numbers (average 2C = 0.44 pg DNA) in multiple individuals from all populations except for one triploid individual from population AH086 with *c*. 1.5-fold greater nuclear DNA content (2C = 0.71 pg). Diploid individuals showed little variation in nuclear DNA content (ranging from 0.41 to 0.47 pg among the subset of 87 individuals analysed for absolute DNA content, see Supporting Information, Table S1 for details).

Major genetic groupings inferred from AFLP markers

AFLP fingerprinting of 768 individuals from 75 European and seven Japanese populations (Eurasian dataset) yielded a total of 585 reliable fragments (96% overall reproducibility), of which 328 (56%) fragments were polymorphic; only these were used in the subsequent analysis.

Bayesian clustering of the Eurasian dataset using Structure converged into similar results among replicate runs under $K = 2$ and $K = 6$ (Supporting Information, Fig. S1A); the former partition showed also the greatest delta K. The $K = 2$ partition distinguished Alpine populations (except for the Austrian Tyrol) from the rest; $K = 6$ separated populations from Japan, the Alps (except for the Austrian Tyrol), the Hercynian massif (+ Western Europe and the Austrian Tyrol) and the Western Carpathians, and finally distinguished three groups from the Eastern Carpathians, the Southern Carpathians and the Balkans with varying levels of admixture (Fig. 1A). Japanese populations were clearly separated from their European counterparts by the combination of the first and second PCoA axes (Fig. 1C); this separation accounted for 21% of the total AFLP variation in AMOVA. When focusing on European accessions only (European dataset, based on 325 polymorphic fragments), populations clustered into three (both highest delta K and similarity among runs, Supporting Information, Fig. S1B) geographically well-delimited groups, hereafter called the north-western group (Hercynian massif, Western Europe and Western Carpathians, and two populations from the Austrian Tyrol), the south-eastern group (Southern and Eastern Carpathians and Balkan Peninsula) and the Alpine group (Southern and Eastern Alps) (Figs 1B, 2A). This grouping accounted for 13% of the total variation in AFLP phenotypes (hierarchical AMOVA, Supporting Information, Table S3) and was well reflected by separation into three clear clusters in the PCoA (Fig. 1D). For the following analyses we thus assigned the populations to four genetically and geographically delimited major lineages: both the spatially and genetically distinct Japanese group and the three European groups that were clearly delimited in the Structure analysis of the European dataset.

Patterns of genetic diversity inferred from AFLP and microsatellite loci

Using both dominant and co-dominant markers, we assessed levels of intrapopulation variation and interpopulation differentiation of *A. halleri* populations assigned to the four major lineages. Japanese plants exhibited the lowest diversity as inferred from AFLPs (Nei's gene diversity, 0.05), but the north-western group exhibited the lowest proportion of rare AFLP fragments (DW index, 0.32). The south-eastern group exhibited both the highest diversity (0.11) and proportion of rare fragments (0.58). Correspondingly, populations from the major groups differed with respect to

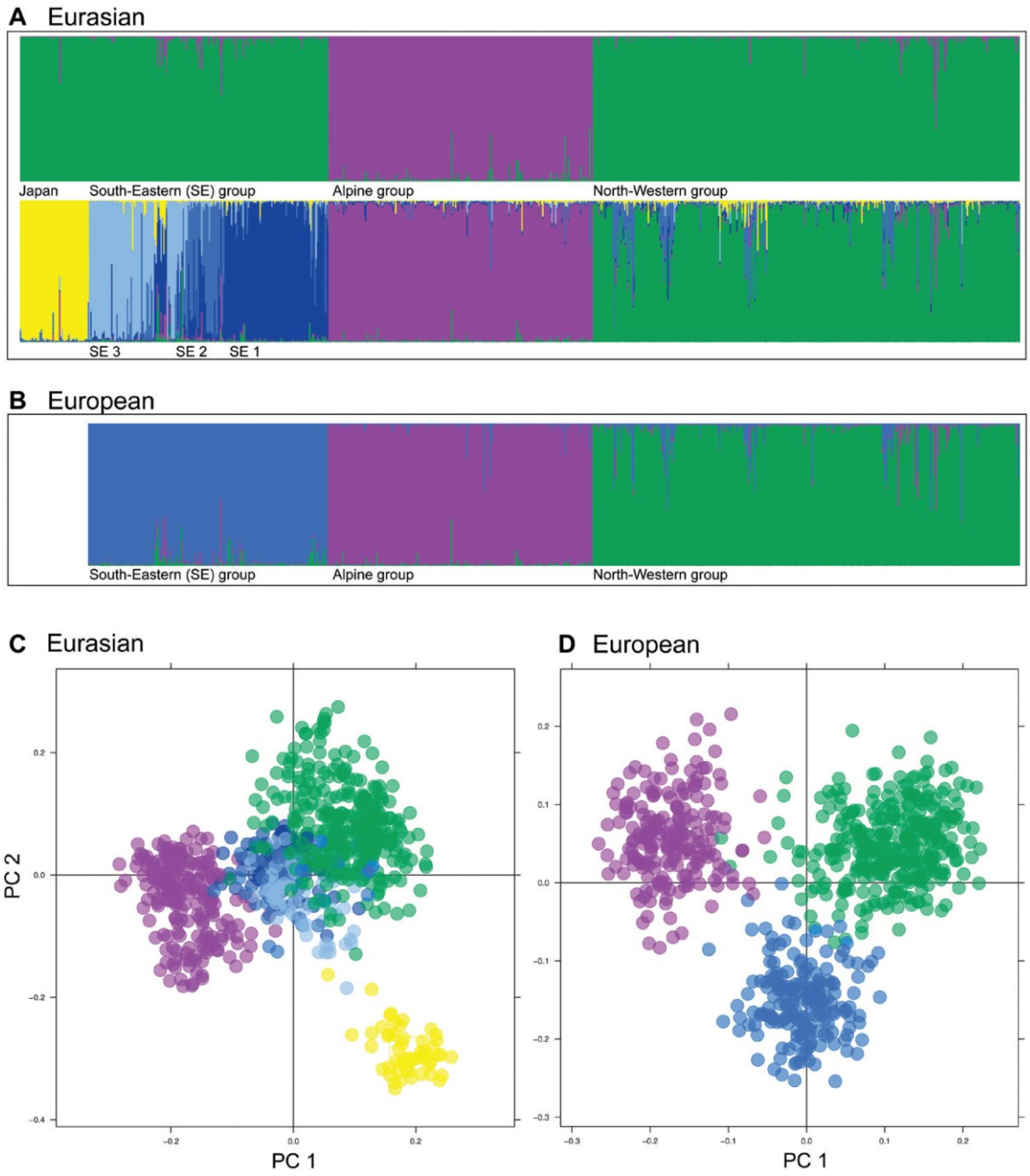


Figure 1. Genetic structure of 768 Eurasian (A, C) and 715 European (B, D) samples of *Arabidopsis halleri* inferred from AFLP markers. Ancestry proportions inferred using Structure under $K = 2$ and $K = 6$ (A) and $K = 3$ (B) for Eurasian and European datasets, respectively. Principal coordinate analysis based on Jaccard distances of all individuals in the Eurasian (C) and European (D) dataset, colour coded according to their major ancestry inferred by the corresponding Structure analysis. Subgroups of the south-eastern (SE) group resolved by $K = 6$ depicted in (A) correspond to Eastern Carpathians (SE 1), Southern Carpathians (SE 2) and Balkans (SE 3) that are described in the text.

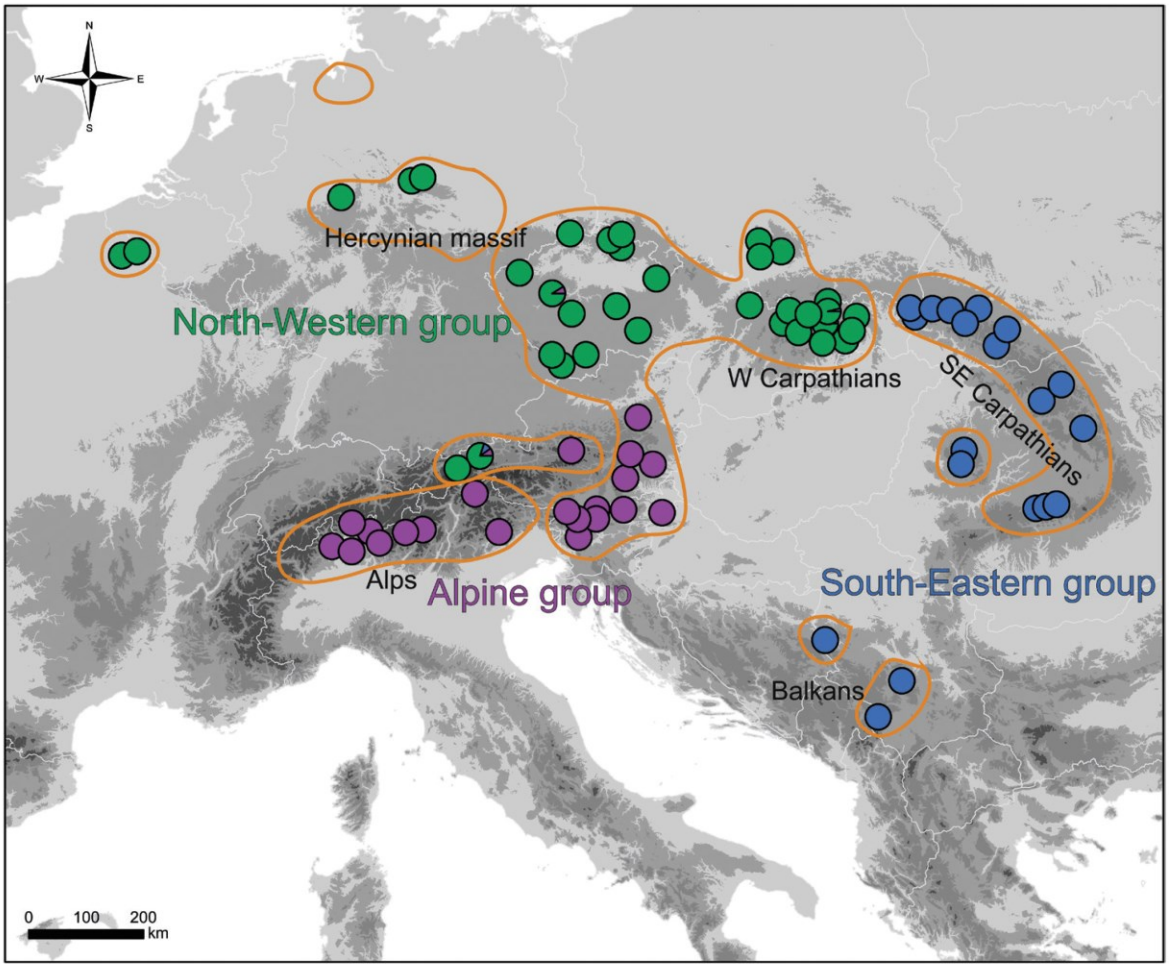


Figure 2. Geographical distribution of major genetic groups of *Arabidopsis halleri* in Europe as inferred by Structure analysis of AFLP data (the colour scheme corresponds to Fig. 1B). Pie charts reflect the proportion of individuals belonging to particular Structure groups (based on their major inferred ancestry). The orange line denotes the borders of the *A. halleri* distribution range following Hoffmann (2005).

diversity ($F_{3,78} = 11.7, P < 0.001$) and rarity ($F_{3,78} = 12.0$, $P < 0.001$), although the overlap was high (Table 1).

The five microsatellite loci comprised in total 74 alleles (seven to 26 alleles observed per locus). Populations from the south-eastern group exhibited the highest average values of genetic diversity, in terms of expected heterozygosity (0.56) and allelic richness (3.39); on the other hand, Alpine populations showed the strongest differentiation ($F_{ST} = 0.29$, $G'_{ST} = 0.65$, Table 1). Japanese populations exhibited similar levels of both observed and expected heterozygosity to their European counterparts (non-significant differences) but significantly lower allelic richness (1.64 on average; $F_{3,76} = 10.99$, $P < 0.001$ and Tukey *post-hoc* comparisons at alpha = 0.05). In summary, both AFLP and microsatellite data show that Japanese populations are slightly genetically depauperate, whereas the three European groups exhibited rather similar levels of genetic diversity, and each of the three groups comprised multiple populations with elevated levels of genetic diversity and proportion of rare markers (Fig. 3).

Environmental and geographical correlates of genetic diversity

The genetic structure of European populations of *A. halleri* (AFLP-based chord distances) strongly correlated with the environmental conditions of their sites (Table 1), suggesting remarkable niche differentiation of the European *A. halleri* lineages. This correlation was linked to geographical structure in the data

Table 1. Genetic diversity and differentiation of *Arabidopsis halleri* populations inferred from multiple genetic markers [AFLPs, microsatellites (= SSR) and plastid and nuclear chalcone synthase (= *CHS*) DNA sequences]. The populations are grouped into four major genetic groups identified by AFLP data (see Results for details on their delimitation): three in Europe and one in East Asia (Japan). A summary for all European populations is also provided

	Alpine group	South-eastern group	North-western group	All European *A. halleri*	Japanese group	All *A. halleri*
Number of populations/individuals for AFLP	21/203	19/184	35/328	75/815	7/53	82/868
Number of individuals for plastid DNA/*CHS* sequencing	27/9	25/7	32/14	84/30	15/8	99/38
Elevation range (m a.s.l.)	210–2100	300–2250	50–1880	50–2250	20–500	20–2250
Total number of polymorphic fragments (AFLP)	244	277	269	325	131	328
Private AFLP fragments[†]	**15/17**	14/**17**	07/09	197/–	3/–	–
Private SSR alleles[†]	03/**06**	05/05	**06/06**	41/–	1/–	–
DW index (AFLP)[†‡]	0.46 (0.23–0.81)/0.48 (0.26–0.82)	**0.58** (0.37–1.03)/**0.61** (0.38–1.10)	0.33 (0.14–0.82)/0.35 (0.14–0.89)	0.43 (0.14–1.03)/0.45 (0.14–1.10)	0.39 (0.23–0.55)/–	–
Gene diversity (AFLP)[‡]	0.07 (0.04–0.10)	**0.08** (0.05–0.12)	0.07 (0.03–0.12)	0.07 (0.03–0.12)	0.04 (0.02–0.06)	0.07 (0.02–0.06)
Expected heterozygosity (SSR)[‡]	0.49 (0.33–0.66)	**0.56** (0.40–0.69)	0.50 (0.50–0.26)	0.51 (0.33–0.69)	0.50 (0.31–0.59)	0.51 (0.31–0.69)
F_{ST} (SSR)	**0.29*****	0.19***	0.23***	0.31***	0.32***	0.32***
G'_{ST} (SSR)	**0.65*****	0.54***	0.53***	0.63***	0.61***	0.65***
Allelic richness (SSR)	2.77 (1.77–4.53)	**3.39** (1.64–4.69)	2.77 (1.57–3.78)	2.93 (1.57–4.69)	1.64 (1.25–2.11)	2.82 (1.25–4.69)
Isolation by distance (r_M)	**0.55*****	0.40**	0.35***	0.43***	–	–
Plastid haplotypes unique/shared[§]	Ad, Abb/**Aa**, **Ba**, Ab, Bf	Bb, Be, C/**Aa**, **Ba**, Ac, Bd, Bf	Bc/Ab, Ac	all	–/**Ba**, Ab	all
r_M environment[¶]	0.15*/0.07	0.57***/0.45**	0.20*/0.04	0.13*/0.01	–	–
Genetics-environment (CAP)[††]	0.24*/0.19***	0.36**/0.26***	0.16***/0.09	0.14***/0.06***	–	–

*$P < 0.05$, **$P < 0.01$, ***$P = 0.001$

[†]Values for the Eurasian and the European dataset are given before and after the slash, respectively; the highest value is in bold type.

[‡]Mean (min–max) per-population values.

[§]Haplotypes in bold are interpreted as ancestral haplotypes.

[¶]Mantel correlations among genetic and environmental distances of AFLP-genotyped populations. Results of a simple Mantel test and a partial Mantel test conditioned by the matrix of geographical distances are shown before and after the slash, respectively.

[††]Association of main environmental gradients with genetic distances among all genotyped populations as inferred from direct ordination (constrained analysis of principal coordinates, CAP). The relative explanatory power of environmental parameters (bioclimatic variables and elevation), with and without the inclusion of geographical distance as a covariate, is shown before and after the slash, respectively.

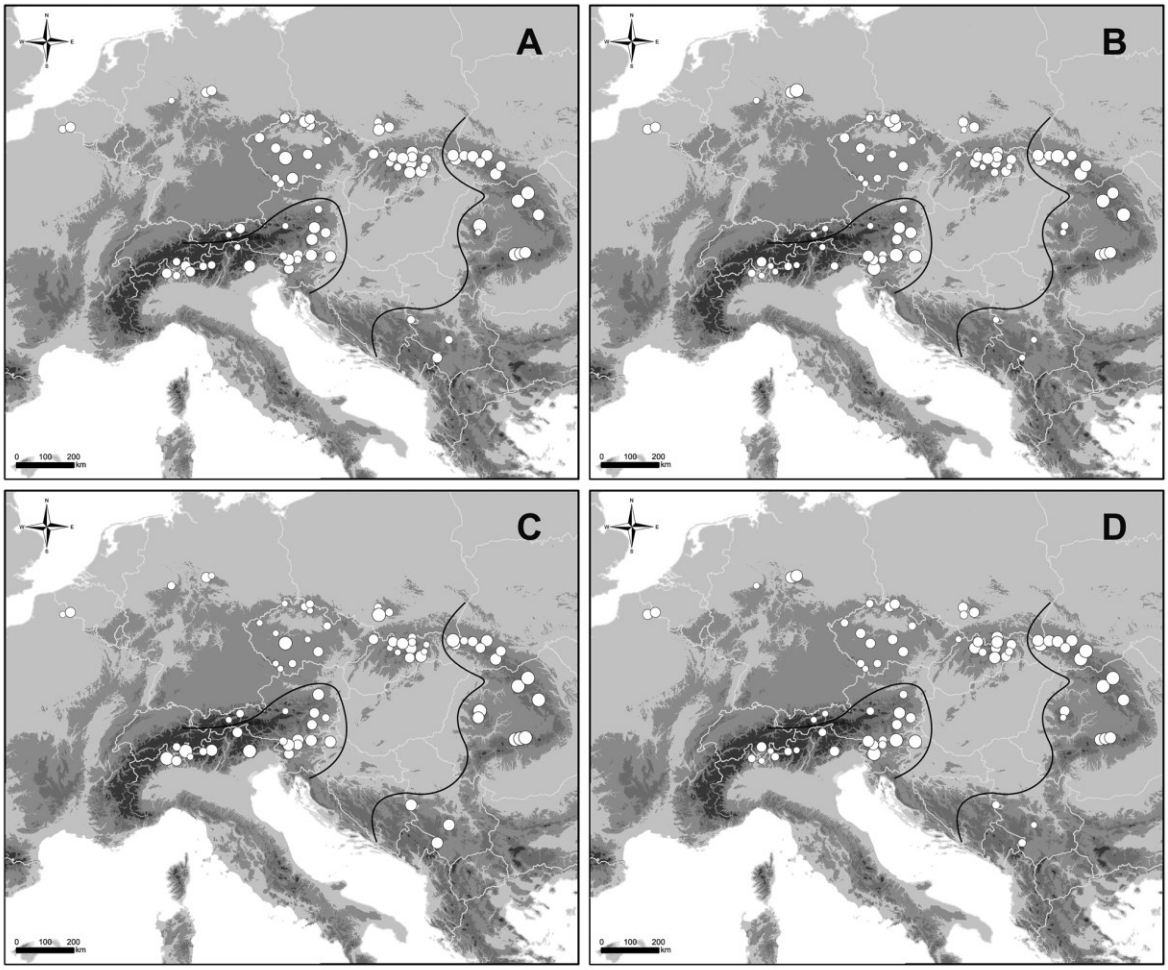


Figure 3. Patterns of population-level diversity and proportion of rare markers in European *Arabidopsis halleri* inferred from AFLP (A, C) and microsatellite data (B, D) from the European dataset. A, gene diversity of AFLP fragments; B, expected heterozygosity inferred from microsatellite data; C, proportion of rare AFLP fragments (DW index); and D, microsatellite allelic richness. The black line indicates borders of the three major European lineages.

(isolation by distance relationship, $r_M = 0.43$), and the environmental–genetic relationship got weaker when conditioned by geography (Table 1). The strength of this correlation differed between the three major genetic groups in Europe, south-eastern populations showing the strongest (and significant) relationship, whereas this effect was weakest and non-significant among north-western populations, particularly in models conditioning for the effect of geography (Table 1).

Phylogenetic relationships and geographical distribution of plastid haplotypes

The lengths of the *cp3* non-coding region in the full *A. halleri* dataset (including all 100 specimens) was 629 bp and comprised 11 variable and potentially informative characters, including three indels. Twelve haplotypes were identified in 99 *A. halleri* individuals and three more haplotypes were observed when datasets including outgroup taxa were analysed. The original alignment and GenBank accession numbers are provided in Supporting Information, Data S3.

Two internal haplotypes were interpreted as putative ancestral haplotypes based on the rationale of Schaal *et al.* (1998), i.e. based on their central position in the statistical parsimony (TCS) network, early branching position in the phylogenetic tree and wide geographical distribution (Fig. 4A–C, Table 2). These putative ancestral haplotypes are: (1) haplotype *Aa* present in several populations distributed in the Eastern

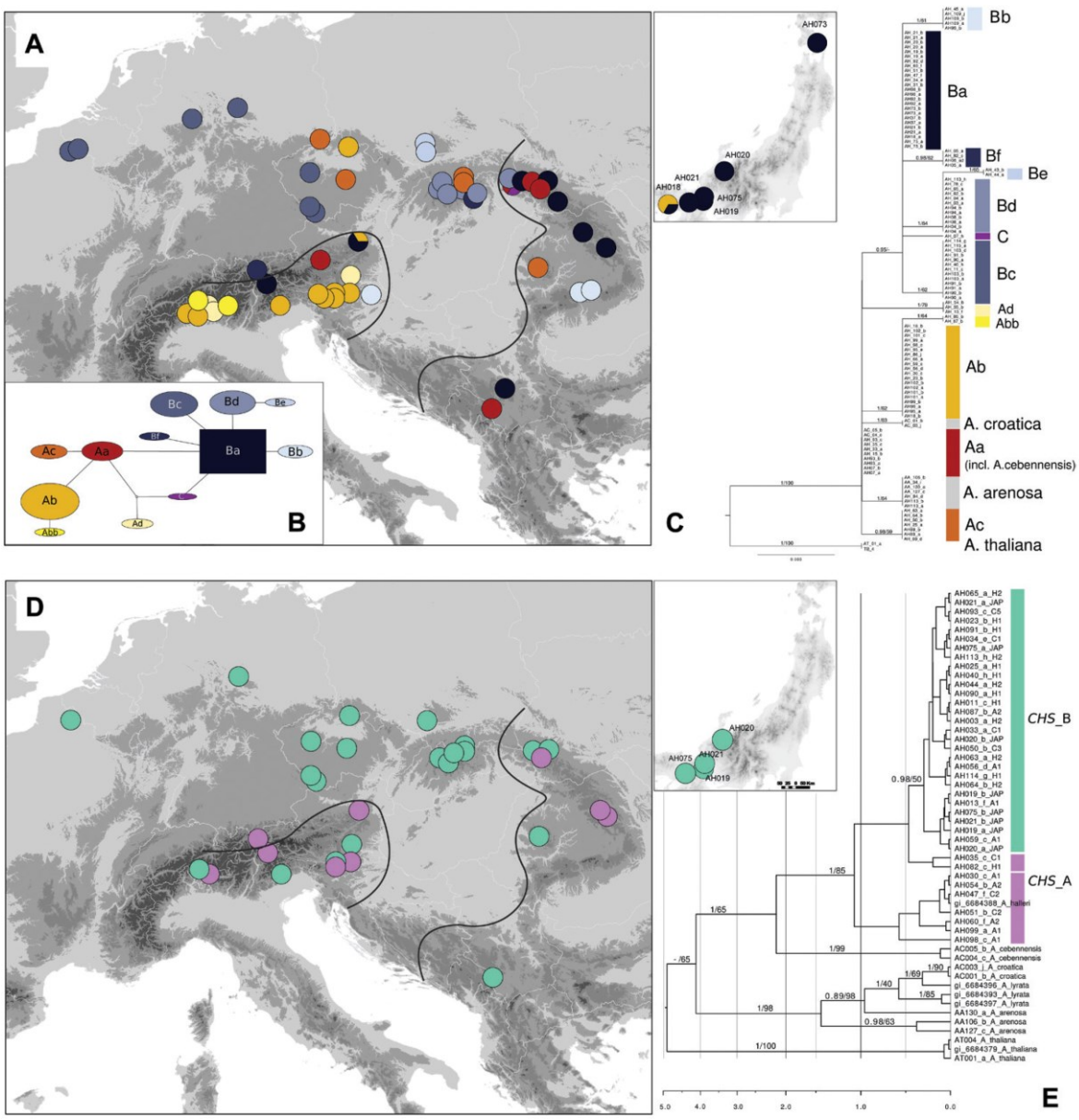


Figure 4. Distribution of genetic diversity of *Arabidopsis halleri* populations in Europe and Japan (inset) inferred from plastid and nuclear DNA sequences. A, geographical distribution of plastid haplotypes. B, relationship of the plastid haplotypes visualized using statistical parsimony [the colour shading reflects divergence from the two putative ancestral internal haplotypes, *Aa* for the brown group and *Ba* for the grey group, respectively (see details in the text)]. C, hierarchical relationships among the plastid haplotypes shown by maximum likelihood 50% bootstrap majority rule tree (scale below the tree indicates the mean number of nucleotide substitutions per site on the respective branch). D, geographical distribution of two main lineages inferred from a nuclear low-copy gene (chalcone synthase). E, relationships among the lineages indicated by a dated Bayesian consensus tree (scale below the tree in Mya). The black line in the maps denotes borders among the three major European lineages inferred from AFLP data. Values of posterior probabilities (based on Bayesian analysis) and bootstrap support (maximum likelihood analysis) are indicated above the branches of each tree before and after the slash, respectively.

Table 2. Distribution of genetic diversity in *Arabidopsis halleri* inferred from AFLPs and plastid and nuclear (chalcone synthase, *CHS*) DNA sequences across different biogeographical regions; putative ancestral haplotypes are denoted in type

Geographical areas	AFLP groups	Plastid haplotypes	*CHS* groups
Unique			
Southern and Eastern Carpathians (SE-Carp.)	South-eastern group	Bb, C	–
Western Carpathians (W-Carp.)	North-western group	Be	–
Hercynian massif & Western Europe (Herc.)*	North-western group	Bc	–
Alps	Alpine group	Ad, Abb	–
Balkans	South-eastern group	–	–
Japan	Japanese group	–	–
Shared among areas			
SE-Carp., Alps, Balkans, Japan	–	**Ba**	–
SE-Carp., Alps, Balkans	–	**Aa**	–
SE-Carp., Alps	–	–	A basal grade
SE-Carp., W-Carp., Herc.	–	Ac	–
SE-Carp., W-Carp.	–	Bd	–
W-Carp., Alps	–	Bf	–
Alps, Herc., Japan	–	Ab	–
all areas	–	–	B clade

*Including the two northern Alpine (Tyrolian) localities in anthropogenic habitats.

Carpathians, Alps and Balkans and in two individuals of *A. cebennensis* (the closest relative of *A. halleri*) and (2) haplotype *Ba*, which is the most abundant, similarly distributed in the Eastern Carpathians, Alps, Balkans and Japan, although prevailing in the Carpathians (Fig. 4A–C, Table 2). Most of the remaining haplotypes are directly derived from one of the putative ancestral ones (Fig. 4B). With the exception of two haplotypes (*Ad*, *Abb*), none of the derived haplotypes differs from the ancestral one by more than one mutation step. Haplotypes derived from ancestral haplotype *Aa* are distributed mainly (*Ab*) or exclusively (*Abb*, *Ad*) in the Alps with the exception of haplotype *Ac*, which is distributed outside of the Alps, in the Carpathians and the Hercynian area. Haplotypes derived from *Ba* are distributed either exclusively in the Carpathians (*Bb*, *Bd*, *Be*, *C*), exclusively in the Hercynian area (*Bc*) or they occur in both regions (*Bf*; Table 2). None of the derived haplotypes has a broader distribution than their putative ancestral haplotypes (Fig. 4A).

Hierarchical (phylogenetic) relationships of the plastid haplotypes were only poorly resolved in the Bayesian tree due to a low number of informative sites in the dataset. *Arabidopsis thaliana* accessions were resolved as a true outgroup, whereas *A. cebennensis*, *A. arenosa* and *A. croatica* accessions were included in the ingroup with all accessions of *A. halleri* (PP = 1). Within the ingroup, several accessions (including two accession of *A. cebennensis*) were placed in a basal grade, all representing the putatively ancestral haplotype *Aa*. Furthermore, within the ingroup, six clades were detected in the basal polytomy, each supported by PP > 0.95. One of those clades includes all accessions of *A. arenosa* (plus one *A. halleri* accession, AH94_d), one clade includes accessions of *A. croatica* and one clade includes the putative ancestral haplotype *Ba* and all its derivatives. The other clades represent derivatives of the putative ancestral haplotype *Aa* (Fig. 4C).

Variation in the *CHS* nuclear marker

The *CHS* region in the full dataset of 39 *A. halleri* accessions and 13 outgroup specimens (including five GenBank accessions from Koch *et al.*, 2000) was 606 bp long and comprised 57 variable and 51 potentially parsimony-informative characters (but only 13 and ten variable and parsimony-informative characters, respectively, within the *A. halleri* ingroup). No indels were detected in this dataset. As the individuals were not cloned, several polymorphic sites discovered were coded with IUPAC ambiguity codes. Twenty-six genotypes were identified among the *A. halleri* individuals. The original alignment and GenBank accession numbers are provided in Supporting Information, Data S3.

The Bayesian phylogenetic tree (Fig. 4E) was well resolved at the species level, i.e. all accessions belonging to a particular species grouped together with maximum support (PP = 1). Within the *A. halleri*

ingroup, however, there was only one supported clade (PP > 0.95) that included most of the sampled accessions from the entire distribution range of *A. halleri* in Eurasia (*CHS*_B in Fig. 4B). The remaining genotypes from the Alps and Eastern Carpathians were placed in the basal polytomy (*CHS*_A grade in Fig. 4D, E). On the time scale, the *CHS* genotypes of *A. halleri* diversified *c.* 1 Mya [with a 95% highest posterior density interval (HPD) of 0.3–2.3 Mya]. The diversification of the genus *Arabidopsis* was then estimated at 4.9 Mya (with a 95% HPD of 1.9–9.9 Mya), i.e. in a similar interval as previously suggested by Koch *et al.* (2000).

Divergence dating analysis based on plastid and nuclear DNA sequences

Three independent species tree reconstructions were performed with different prior settings of the mutation rate for the *CHS* marker in order to obtain a range of possible divergence times of nodes of interest in the species tree but also in both gene trees, plastid DNA and *CHS* (Supporting Information, Fig. S2 and Table S4). All resulting 'species' trees (with the major *A. halleri* groups coded as separate lineages) provided high support for the monophyly of *A. halleri*, whereas neither the relationships among the four major groups within *A. halleri* nor the relationships of *A. halleri* to its three outgroup species (*A. cebennensis*, *A. croatica* and *A. thaliana*) were resolved. The low resolution of the species trees allows us to describe only the two well-supported nodes, i.e. tree root corresponding to the divergence of the genus *Arabidopsis* and node corresponding to the diversification of *A. halleri*. Table S4 summarizes in detail the estimates of divergence times of *Arabidopsis* and *A. halleri* from three different species tree analyses and also presents estimated ranges of cpDNA mutation rates that have been used to evaluate the reliability of particular analyses. The diversification of the genus *Arabidopsis* according to the species tree (based on all three independent estimates) took place 0.06–4.85 Mya whereas the diversification of this group based on particular gene trees is suggested to be much older, 0.68–5.41 Mya based on plastid DNA and 1.36–6.18 Mya based on *CHS*. The diversification of *A. halleri* into the four major lineages according to the species tree (again, based on all three independent estimates) occurred 0.01–0.16 Mya, i.e. falling into the late Pleistocene, whereas the diversification estimates based on particular gene trees are suggested to be much older, 0.07–0.82 Mya based on plastid DNA and 0.22–1.43 Mya based on *CHS*, corresponding to the presence of deep coalescence at these loci.

Correlates of morphological variation

The morphological variation of the 41 European *A. halleri* populations was strongly determined by the environmental conditions of the sites occupied, partly as a single effect of the environment itself and partly in correlation with their genetic background, as shown by variation partitioning (Fig. 5; see Supporting Information Fig. S3 for the loadings of individual morphological characters). Environmental variables explained the largest proportion of morphological variation when used as the only predictors (34%) and showed the strongest Mantel correlations (r_M = 0.46), and their unique contribution, after conditioning for the effect of geography (27%) and/or geography+genetics (18 %), was also the strongest and statistically significant (Table 3). The elevational gradient and its bioclimatic correlates seem to be the major driver of the morphological variation [the largest portion (11%) of the variation was explained by the combined effect of elevational + bioclimatic variables only, Fig. S4, see also Fig. 5A]. In particular, populations from ecologically divergent subalpine stands (> 2000 m a.s.l.) exhibited the morphologically most distinct *A. halleri* phenotype. This phenotype was observed in representatives of the two major genetic lineages (the south-eastern group and the Alpine group, Fig. 5A), indicating its parallel origin. Finally, the strength of the drivers of morphological variation differed between populations belonging to the three major lineages. Alpine and south-eastern populations exhibited a stronger effect of the environment and/or genetic structure, but a weak effect was detected within populations of the north-western group (Table 3).

DISCUSSION

Origin of the Northern Hemisphere disjunction

By including populations from across the European and the East Asian ranges of *A. halleri* (the latter being represented by populations sampled over the entire range of the species in Japan), we tried to address the origin of the striking disjunction spanning the entire Palaearctic region. Although there are several cases of European–East Asian disjunctions, e.g. in the related genus *Cardamine* L. (European *C. resedifolia* L. vs. Japanese *C. nipponica* Franch. & Sav.; Ikeda, Fujii & Setoguchi, 2011), relatively little attention has been paid to their origin (e.g. Carlson *et al.*, 2012). Two extreme scenarios may explain this pattern: either recent human-mediated dispersal or ancient vicariance reflecting past contacts of the European and East Asian temperate biota, which were frequent since the early Tertiary (Sanmartín, Enghoff

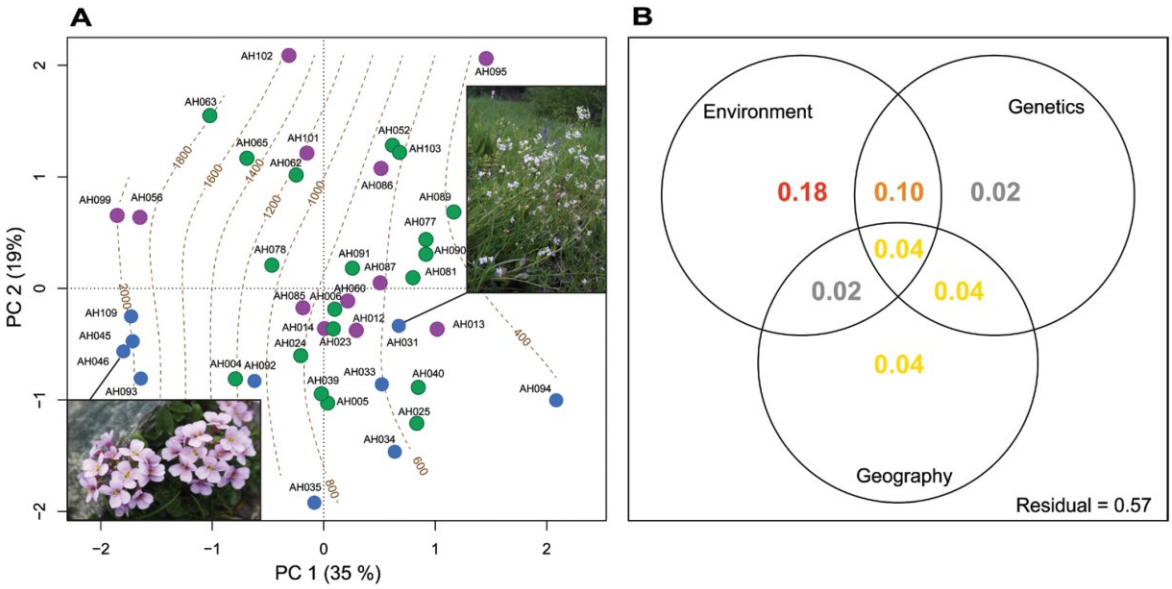


Figure 5. Morphological differentiation of a subset of 41 European *Arabidopsis halleri* populations related to genetic structure, geographical location and environmental preferences. A, principal component analysis of morphologically screened populations (based on population averages) coloured according to their assignment to the three major AFLP lineages; brown contour lines denote elevation fitted onto the ordination plot by a thin plate spline generalized additive model; B, relative contributions of environmental parameters (bioclimatic variables and elevation), genetic structure (AFLP-based genetic distances) and geographical distances to the morphological variation of populations, inferred by variation partitioning through a set of direct ordinations (redundancy analyses).

& Ronquist, 2001). Instead of favouring either of these extremes, our data suggest an 'intermediate scenario' of long-distance dispersal from Europe to East Asia, probably during the Pleistocene. Ancient vicariance seems unlikely because the Japanese populations in fact possess a subset of European haplotypes of both nuclear (*CHS*) and plastid DNA markers (Fig. 4, Table 2) and AFLP fragments (Table 1). Although this pattern could be due to incomplete lineage sorting in both areas, the virtual absence of private genetic diversity (no private haplotypes, only three private AFLP fragments) and comparable levels of genetic differentiation among the three European groups and between European and Japanese populations (21% vs. 13% in AMOVA, Supporting Information, Table S3) rule out long-term isolation linked with parallel divergence in each area. Based on the prior used for the plastid DNA mutation rate (4×10^{-3} substitutions per site per My), we could expect diversification of plastid alleles about 120–640 kya and no more than two mutations from the ancestral haplotype since that time. Although we can observe the haplotype variation caused by those mutations within Europe, populations in Japan contain only ancestral haplotypes (Fig. 4). Therefore, we presume that the ancestral haplotypes were introduced there not more than 250 kya, as there are no private haplotypes that differ by even a single mutation. On the other hand, recent (e.g. human-mediated) dispersal could be rejected based on the relatively high levels of genetic diversity in Japanese populations, which are either similar to or only slightly lower than the average among their European counterparts (Table 1). If *A. halleri* was a recently introduced alien, we would expect markedly reduced levels of genetic diversity, as is usually detected in species recently introduced by humans (e.g. Molina-Freaner & Jain, 1992; Novak & Mack, 1993; Bartlett, Novak & Mack, 2002). In addition, Japanese *A. halleri* occupies a large area spanning the entire northern half of the country, preferring natural habitats (Sato & Kudoh, 2014), unlike a typical alien species. In summary, although Japanese populations form a genetically distinct group, this group is divergent from European populations at levels comparable to that among the three European groups.

The increased levels of diversity in all markers studied (Fig. 4, Tables 1, 2), the presence of private AFLP fragments and plastid and nuclear DNA haplotypes in Europe, as well as the presence of all other *Arabidopsis* spp. on the continent (Hoffmann, 2005; Hohmann *et al.*, 2014) indicate Europe as the potential ancestral area. Taking into account the clear delimitation of the three European groups and the

Table 3. Contribution of genetic structure, geography and environmental preferences to morphological differentiation of the 41 morphologically screened *Arabidopsis halleri* populations from Europe

	N	r_M environment†	r_M genetics†	RDA complete model‡	RDA environment‡	RDA genetics‡	RDA geography†	RDA environment (geography)§	RDA genetics (geography)§
Alpine group	12	0.55***/0.56***	0.34*/0.41*	0.64**	0.47**/0.05	0.45***/0.05	0.26/0	0.44**	0.43***
South-eastern group	10	0.75***/0.65***	0.54**/0.38*	0.53	0.44***/0.18	0.34*/0	0.13/0	0.48*	0.15
North-western group	19	0.38**/0.27*	0.34**/0.14	0.24*	0.14**/0	0.19**/0	0.22***/0	0.01	0.01
European *A. halleri*	41	0.46***/0.43***	0.31***/0.25***	0.43***	0.34***/0.18***	0.20***/0.02	0.14***/0.04*	0.27***	0.12**

*$P < 0.05$, **$P < 0.01$, ***$P = 0.001$, inferred by 1000 permutations.
†Mantel correlations among morphological and either environmental parameters (bioclimatic variables and elevation, 'r_M environment') or AFLP-based among-population genetic distances ('r_M genetics'). Results of a simple Mantel test and a partial Mantel test conditioned by the matrix of geographical distances are shown before and after the slash, respectively.
‡Association of main environmental gradients, genetic and geographical distances with morphology of *A. halleri* populations tested by direct ordination (redundancy analysis, RDA). Relative explanatory power is shown for all explanatory variables ('RDA complete model') and separately as the contributions of environmental parameters (bioclimatic variables and elevation, 'RDA environment'), AFLP-based genetic distances ('RDA genetics') and geographical distances ('RDA geography') to the complete RDA model. Marginal effects (effect of the set of factors when used as the only explanatory variables) and unique contributions (after conditioning for the effect of all the other tested factors) are shown before and after the slash, respectively.
§Separate RDAs with environmental or genetic parameters as the only explanatory variables and geography as the only conditional factor.

Japanese group together with the fact that the relationships among them are unresolved, we hypothesize that the four major groups identified in our study might have originated during a single radiation event, which probably dates back to the Pleistocene (Supporting Information, Fig. S2). Any precise dating of this event and evaluation of demographic consequences of the Northern Hemisphere dispersion, however, will require further investigations using multiple unlinked DNA sequence markers.

Major genetic groupings in Europe

Although European populations have already been subjected to population-genetic and biosystematic investigations, previous studies were focused on specific geographical regions (Kolník & Marhold, 2006; Wasowicz *et al.*, 2016) or metal-polluted sites and their surroundings (Van Rossum *et al.*, 2004; Pauwels *et al.*, 2005, 2006, 2008, 2012). The present study is the first based on homogeneous sampling across the whole distribution in Europe, including previously undersampled (Western, Eastern and Southern Carpathians) and completely unsampled (the Balkans) areas. Our AFLP data suggest the existence of three distinct gene pools in Europe, the distribution of which is strongly correlated with major biogeographical boundaries in the Central European mountain system, i.e. barriers separating the Alps and Hercynian mountains, and the border between the Western and Eastern Carpathians (Wołoszczak, 1896; Pawłowski, 1970; Fig. 2). Although the differentiation between Hercynian and Alpine populations was previously detected by examining microsatellite and plastid DNA polymorphism variation (Pauwels *et al.*, 2012), we newly describe a third group which spans the Southern and Eastern Carpathians and reaches the mountains of the central Balkan Peninsula. This south-eastern group is not only distinct in having different frequencies of AFLP bands (Figs 1, 2) and plastid haplotypes (Fig. 4A–C), but it also possesses its own set of private AFLP fragments, microsatellite alleles and plastid DNA haplotypes (Fig. 4A–C, Table 1). Although the south-eastern Carpathians are poorly known from a plant phylogeographical perspective, the few studies available so far suggest the separation of this region from the Western Carpathians and a high frequency of unique genetic diversity there, which is in line with our findings (e.g. Těšitel *et al.*, 2009; Ronikier, 2011; Kolář *et al.*, 2016c; Mráz & Ronikier, 2016). Finally, the three main genetic lineages, despite being congruent with geography, do not match the current taxonomical concept of five subspecies in *A. halleri* (e.g. Hoffmann, 2005; Kolník &

Marhold, 2006; Hohmann *et al.*, 2014), implying a need for a further taxonomic revision of the species.

Although few south-eastern Carpathian populations were included in previous investigations, they were usually grouped together with the Southern Alpine populations, based on patterns in microsatellite and plastid DNA variation (Pauwels *et al.*, 2005, 2012). We, however, assume that this assignment might reflect unequal sampling (*c.* 10% of the samples coming from the area; Pauwels *et al.*, 2012) and/or lower discriminative power of microsatellites in detecting old divergences and range-wide patterns in genetic variation than AFLP (e.g. Skrede, Borgen & Brochmann, 2009). In addition, the previously observed extensive haplotype sharing among south-eastern Carpathian and Alpine populations (Pauwels *et al.*, 2012) might reflect an ascertainment bias because the plastid genotyping technique was optimized to screen variants known in the western part of the range of *A. halleri* (France, Germany, Czech Republic, Slovakia and Poland; Pauwels *et al.*, 2005), thus hampering the discovery of private variants in either Alpine or south-eastern Carpathian populations. In contrast, our direct sequencing approach identified two shared but presumably ancient (ancestral) haplotypes and several haplotypes that were private to each region (Fig. 4A–C, Table 2). In fact, the south-eastern Carpathians appear to be the most genetically diverse area hosting the largest proportion of private/rare diversity, consistently over all markers analysed in this study (Table 1, Fig. 3).

On the other hand, both plastid and nuclear DNA sequences show some levels of haplotype sharing among major lineages and/or geographically distant areas (Fig. 4), raising the question of whether this is mere incomplete lineage sorting or a result of past hybridization among the major lineages. The admixture model applied to the clustering of AFLP data allows potential recent contact between three European groups to be uncovered (Falush *et al.*, 2007). The low levels of admixture among the three major genetic groups, observed in the Structure analysis (Fig. 1B), however, suggest only minor admixture between north-western and Alpine populations in the Austrian Tyrol (all those populations occupy man-made grassland habitats such as ornamental parks, and thus they probably reflect recent human-mediated spread followed by admixture) and between north-western and south-eastern populations in central Slovakia (Fig. 2). We therefore assume that the ancestral haplotype sharing may reflect old vicariance linked to incomplete lineage sorting and/or ancient gene flow between the areas rather than recent hybridization events.

Glacial persistence in multiple areas of Central–Eastern Europe

The typical historical scenario for temperate lowland European biodiversity involves glacial survival in southern European refugia (the Iberian, Apennine and Balkan Peninsulas) and post-glacial recolonization of presumably hostile peri-glacial (low to mid-elevations of Central Europe) or glaciated (high mountains of Central and most of Northern Europe) areas (Taberlet *et al.*, 1998; Hewitt, 2004; Tzedakis, Emerson & Hewitt, 2013). An alternative scenario, however, suggests glacial survival of temperate elements in more northern areas such as the Carpathian mountains, which were only scarcely glaciated (Ronikier, 2011) and hosted forest communities during the last glacial maximum (LGM) (Jankovská & Pokorný, 2008; Juřičková, Horáčková & Ložek, 2014). Among other factors (reviewed by Tzedakis *et al.*, 2013), this controversy persists also due to a shortage of phylogeographical data from south-eastern Europe, although several recent studies indicated survival of the LGM based on patterns of elevated genetic diversity and distinctness in various parts of the Carpathians (Magri *et al.*, 2006; Těšitel *et al.*, 2009; Kolář *et al.*, 2016c).

European *A. halleri* may bring important clues to this debate, as it is a species with a temperate distribution but preferring mid-elevations (montane habitats) across various European mountain ranges, i.e. a so far little-explored ecological group (Ronikier, Cieślak & Korbecka, 2008; Mráz & Ronikier, 2016). The three European groups detected by our study probably represent separate vicariant lineages with origins probably pre-dating the LGM, as indicated by divergence dating (Supporting Information, Fig. S2) and high AFLP divergence (13% of among-group differentiation in AMOVA). In addition, the *in situ* persistence of each group during the LGM in at least part of the current range of the species is suggested by the presence of genetically variable populations with high proportions of variable markers in each lineage (high DW index, allelic richness, Table 1, Fig. 3; private plastid haplotypes, Tables 1, 2). Populations hosting elevated genetic diversity and proportion of rare fragments (probably corresponding to areas of long-term persistence, e.g. Tribsch *et al.*, 2002; Paun *et al.*, 2008) are located in the southern and eastern foothills of the Alps (Alpine group), the Western Carpathians and partly the Hercynian mountains of the Czech Republic (north-western group), and multiple regions of the Southern and Eastern Carpathians (south-eastern group; Fig. 3). Finally, the relatively high and statistically significant differentiation of populations in each of the groups (F_{ST} of 0.19–0.29, Table 1) speaks against their recent spread from a single gene pool

that would be expected under a simple recent recolonization scenario.

Previous studies have suggested areas of Western Europe as glacial refugia for *A. halleri* (Pauwels *et al.*, 2012; Wasowicz *et al.*, 2016), but ours is the first to indicate persistence in the Carpathians, probably reflecting the denser and more homogeneous sampling in the area. The putative climatic unsuitability of this area for LGM survival indicated by Wasowicz *et al.* (2016) may also reflect the limited number of occurrence points from Eastern Europe available in distribution databases and thus possibly underestimating the climatic niche of *A. halleri* in this region.

Environmentally driven morphological variation

Preferring a wide variety of substrates, climatic niches and habitats spanning an elevational range of > 2300 m, *A. halleri* represents a suitable model for testing hypotheses concerning niche conservatism vs. shifts throughout its evolutionary history (Pauwels *et al.*, 2012; Hohmann *et al.*, 2014) and the influence of ecological parameters on its phenotypic variation. The genetic structure in European *A. halleri* indeed correlates with local conditions (characterized by current bioclimatic conditions), suggesting ecological differentiation among the major genetic lineages of *A. halleri*, although this correlation is largely driven by covariance with geography (Table 1).

Genetic divergence is not the main cause of the observed morphological variation among populations. Instead, the ecological conditions of the original sampling locations explain the largest proportion of observed phenotypic variation, as suggested by variation partitioning (Fig. 5B). This is best illustrated by the presence of similar morphotypes in high-alpine stands in both the Eastern Alps and the Southern Carpathians, i.e. areas harbouring major distinct genetic groups of *A. halleri* (Fig. 2). Alpine populations from both regions are morphologically similar to each other, yet distinct from their lower-elevation counterparts from the same genetic group (Fig. 5A), suggesting parallel evolution of morphologically distinct alpine phenotypes in *A. halleri*. A similar phenotypic effect of the high-alpine environment was recently observed in diploid *A. arenosa* (Kolář *et al.*, 2016c). Whether this pattern represents phenotypic plasticity or genetically fixed differences, probably reflecting genetic drift in high-alpine populations and/or local adaptation, requires further study. For example, the analogous case of the Alpine *Heliosperma pusillum* (Waldst. & Kit.) Rchb. species complex reflects multiple origins of genetically fixed elevation-differentiated ecotypes (Bertel *et al.*, 2016; Trucchi *et al.*, 2016).

ACKNOWLEDGEMENTS

We are indebted to Hiroshi Kudoh, Martin Kolník, Pavol Mereďa, Eva Kolářová, Petr Koutecký and Milan Štech for help with sampling. The research was supported by the Czech Science Foundation (grant number P506/12/0668) and by the Slovak Research and Development Agency (APVV; grant number APVV-0139-12).

REFERENCES

Al-Shehbaz IA, Arai K, Ohba H. 2006. *Arabidopsis*. In: Iwatsuki K, Boufford DE, Ohba H, eds. *Flora of Japan, vol. IIa, Angiospermae, Dicotyledoneae, Archichlamydeae*. Tokyo: Kodansha, 482–490.

Al-Shehbaz IA, O'Kane SL. 2002. Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). *The Arabidopsis Book* **6:** 1–22.

Anderson JM, Willis JT. 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* **84:** 511–525.

Arnold B, Kim ST, Bomblies K. 2015. Single geographic origin of a widespread autotetraploid *Arabidopsis arenosa* lineage followed by interploidy admixture. *Molecular Biology and Evolution* **32:** 1382–1395.

Bartlett EA, Novak SJ, Mack RN. 2002. Genetic variation in *Bromus tectorum* (Poaceae): differentiation in eastern United States. *American Journal of Botany* **89:** 626–636

Berkutenko AN. 1988. Rod 31. Serdechnikovidnik – *Cardaminopsis* (C.A. Mey.) Hayek. In: Kharkevich SS, ed. *Sosudistye rasteniya Sovetskogo Dal'nego Vostoka*, *vol*. **3**. Leningrad: Izdatel'stvo Nauka, 101–103.

Bert V, Bonnin I, Saumitou-Laprade P, de Laguérie P, Petit D. 2002. Do *Arabidopsis halleri* from nonmetallicolous populations accumulate zinc and cadmium more effectively than those from metallicolous populations? *New Phytologist* **155:** 47–57.

Bertel C, Buchner O, Schönswetter P, Frajman B, Neuner G. 2016. Environmentally induced and (epi-) genetically based physiological trait differentiation between *Heliosperma pusillum* and its polytopically evolved ecologically divergent descendent, *H. veselskyi* (Caryophyllaceae: Sileneae). *Botanical Journal of the Linnean Society* **182:** 658–669.

Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13:** 3261–3273.

Borcard D, Legendre P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* **153:** 51–68.

Carlson SE, Linder HP, Donoghue MJ. 2012. The historical biogeography of *Scabiosa* (Dipsacaceae): implications for Old World plant disjunctions. *Journal of Biogeography* **39:** 1086–1100.

Clauss MJ, Cobban H, Mitchell-Olds T. 2002. Cross-species microsatellite markers for elucidating population genetic

structure in *Arabidopsis* and *Arabis* (Brassicaeae). *Molecular Ecology* **11:** 591–601.

Clauss MJ, Koch MA. 2006. Poorly known relatives of *Arabidopsis thaliana*. *Trends in Plant Science* **11:** 449–459.

Clement M, Posada D, Crandall KA. 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology* **9:** 1657–1659.

Dieringer D, Schlötterer C. 2003. microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3:** 167–169.

Doležel J, Doleželová M, Novák FJ. 1994. Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). *Biologia Plantarum* **36:** 351–357.

Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* **2:** 2233–2244.

Donoghue MJ, Bell CD, Li J. 2001. Phylogenetic patterns in Northern Hemisphere plant geography. *International Journal of Plant Sciences* **162:** S41–S52.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7:** 214.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29:** 1969–1973.

Durbin ML, Learn GH, Huttley GA, Clegg MT. 1995. Evolution of the chalcone synthase gene family in the genus *Ipomoea*. *Proceedings of the National Academy of Sciences of the United States of America* **92:** 3338–3342.

Ehrich D. 2006. AFLPDAT: a collection of R functions for convenient handling of AFLP data. *Molecular Ecology Notes* **6:** 603–604.

Ehrich D. 2007. Structure-sum v. 2007: a series of R functions for summarizing the outputs of the program Structure ver. 2.2. Unpublished, available from the author upon request.

Eidesen PB, Alsos IG, Popp M, Stensrud O, Suda J, Brochmann C. 2007. Nuclear vs. plastid data: complex Pleistocene history of a circumpolar key species. *Molecular Ecology* **16:** 3902–3925.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611–2620.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7:** 574–578.

Hedrick PW. 2005. A standardized genetic differentiation measure. *Evolution* **59:** 1633–1638.

Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27:** 570–580.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **359:** 183–195.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25:** 1965–1978.

Hoffmann MH. 2005. Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution* **59:** 1425–1436.

Hohmann N, Schmickl R, Chiang T-Y, Lučanová M, Kolář F, Marhold K, Koch MA. 2014. Taming the wild: resolving the gene pools of non-model *Arabidopsis* lineages. *BMC Evolutionary Biology* **14:** 1–21.

Ikeda H, Fujii N, Setoguchi H. 2011. Molecular evolution of cryptochrome genes and the evolutionary manner of photoreceptor genes in *Cardamine nipponica* (Brassicaceae). *Journal of Plant Research* **124:** 85–92.

Jalas J, Suominen J, eds. 1994. *Atlas florae Europaeae. Distribution of vascular plants in Europe. 10. Cruciferae (*Sisymbrium *to* Aubrieta*)*. Helsinki: The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo.

Jankovská V, Pokorný P. 2008. Forest vegetation of the last full-glacial period in the western Carpathians (Slovakia and Czech Republic). *Preslia* **80:** 307–324.

Jombart T. 2008. adegenet: an R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403–1405.

Jones BMG, Akeroyd JR. 1993. 41. *Cardaminopsis* L. In: Tutin TG, Burges NA, Chater AO, Edmondson JR, Heywood VH, Moore DM, Valentine DH, Walters SM, Webb DA, eds. *Flora Europaea 1, Psilotaceae to Platanaceae, 2nd edn*. Cambridge: Cambridge University Press, 351–352.

Juřičková L, Horáčková J, Ložek V. 2014. Direct evidence of central European forest refugia during the last glacial period based on mollusc fossils. *Quaternary Research* **82:** 222–228.

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

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28:** 1647–1649.

Koenig D, Weigel D. 2015. Beyond the thale: comparative genomics and genetics of *Arabidopsis* relatives. *Nature Reviews, Genetics* **16:** 285–298.

Koch MA, Haubold B, Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Molecular Biology and Evolution* **17:** 1483–1498.

Kolář F, Lučanová M, Záveská E, Fuxová G, Mandáková T, Španiel S, Senko D, Svitok M, Kolník M, Gudžinskas Z, Marhold K. 2016a. Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae). *Biological Journal of the Linnean Society* **119:** 673–688.

Kolář F, Dušková E, Sklenář P. 2016b. Niche shifts and range expansions along cordilleras drove diversification in

a high-elevation endemic plant genus in the tropical Andes. *Molecular Ecology* **25:** 4593–4610.

Kolář F, Fuxová G, Záveská E, Nagano AJ, Hyklová L, Lučanová M, Kudoh H, Marhold K. 2016c. Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. *Molecular Ecology* **25:** 3929–3949.

Kolník M, Marhold K. 2006. Distribution, chromosome numbers and nomenclature conspect of *Arabidopsis halleri* (Brassicaceae) in the Carpathians. *Biologia* **61:** 41–50.

Kryštufek B. 2004. A quantitative assessment of Balkan mammal diversity. In: Griffiths HI, Kryštufek B, Reed JM, eds. *Balkan biodiversity: patterns and process in the European hotspot*. Dordrecht: Kluwer Academic Publishers, 79–108.

Kubota S, Iwasaki T, Hanada K, Nagano AJ, Asao F, Atsushi T, Sumio S, Yutaka S, Kouki H, Motomi I, Schin-Ichi M. 2015. A genome scan for genes underlying microgeographic-scale local adaptation in a wild *Arabidopsis* species. *PLoS Genetics* **11:** 1–26.

Magri D, Vendramin GG, Comps B, Dupanloup I, Geburek T, Gömöry D, Latałowa M, Litt T, Paule L, Roure JM, Tantau I, Van Der Knaap WO, Petit RJ, De Beaulieu JL. 2006. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist* **171:** 199–221.

Měsíček J, Goliašová K. 2002. *Cardaminopsis* (C. A. Mey.) Hayek. In: Goliašová K, Šípošová H, eds. *Flóra Slovenska V/4*. Bratislava: Veda, 388–415.

Molina-Freaner F, Jain SK. 1992. Isozyme variation in Californian and Turkish populations of the colonizing species *Trifolium hirtum*. *Journal of Heredity* **83:** 423–430.

Mráz P, Ronikier M. 2016. Biogeography of the Carpathians: evolutionary and spatial facets of biodiversity. *Biological Journal of the Linnean Society* **119:** 528–559.

Müller K. 2005. SeqState – primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics* **4:** 65–69.

Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology* **3:** 1289–1299.

Novak SJ, Mack RN. 1993. Genetic variation in *Bromus tectorum* (Poaceae): comparison between native and introduced populations. *Heredity* **71:** 167–176.

Novikova PY, Hohmann N, Nizhynska V, Tsuchimatsu T, Ali J, Muir G, Guggisberg A, Paape T, Schmid K, Fedorenko OM, Holm S, Säll T, Schlötterer C, Marhold K, Widmer A, Sese J, Shimizu KK, Weigel D, Krämer U, Koch MA, Nordborg M. 2016. Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics* **48:** 1077–1082.

Oksanen FJ, Blanchet G, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos PM, Stevens HH, Wagner H. 2015. *Vegan: community ecology package. R package Version 2.3-0*. https://CRAN.R-project.org/package=vegan

Pak J-H. 2005. *The coloured flora for practical use in Korea, herbaceous plants*. Taegu: Tonga Munhwasa [in Hangul].

Paun O, Schönswetter P, Winkler M, Tribsch A; IntraBioDiv Consortium. 2008. Historical divergence vs. contemporary gene flow: evolutionary history of the calcicole *Ranunculus alpestris* group (Ranunculaceae) in the European Alps and the Carpathians. *Molecular Ecology* **17:** 4263–4275.

Pauwels M, Saumitou-Laprade P, Holl AC, Petit D, Bonnin I. 2005. Multiple origin of metallicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in central Europe: the cpDNA testimony. *Molecular Ecology* **14:** 4403–4414.

Pauwels M, Frérot H, Bonnin I, Saumitou-Laprade P. 2006. A broad-scale analysis of population differentiation for Zn tolerance in an emerging model species for tolerance study: *Arabidopsis halleri* (Brassicaceae). *Journal of Evolutionary Biology* **19:** 1838–1850.

Pauwels M, Willems G, Roosens N, Frérot H, Saumitou-Laprade P. 2008. Merging methods in molecular and ecological genetics to study the adaptation of plants to anthropogenic metal-polluted sites: implications for phytoremediation. *Molecular Ecology* **17:** 108–119.

Pauwels M, Vekemans X, Godé C, Frérot H, Castric V, Saumitou-Laprade P. 2012. Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). *New Phytologist* **193:** 916–928.

Pawłowski B. 1970. Remarques sur l'endémisme dans la flore des Alpes et des Carpates. *Vegetatio* **21:** 181–243.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25:** 1253–1256.

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

Ronikier M. 2011. Biogeography of high-mountain plants in the Carpathians: an emerging phylogeographical perspective. *Taxon* **60:** 373–389.

Ronikier M, Cieślak E, Korbecka G. 2008. High genetic differentiation in the alpine plant *Campanula alpina* Jacq. (Campanulaceae): evidence for glacial survival in several Carpathian regions and long-term isolation between the Carpathians and the Alps. *Molecular Ecology* **17:** 1763–1775.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. *Science* **298:** 2381–2385.

Sanmartín I, Enghoff H, Ronquist F. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society* **73:** 345–390.

Sato Y, Kudoh H. 2014. Fine-scale genetic differentiation of a temperate herb: relevance of local environments and demographic change. *AoB Plants* **6:** plu070.

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA. 1998. Phylogeographic studies in plants: problems and prospects. *Molecular Ecology* **7:** 465–474.

Schlüter PM, Harris SA. 2006. Analysis of multilocus fingerprinting data sets containing missing data. *Molecular Ecology Notes* **6:** 569–572.

Schönswetter P, Tribsch A. 2005. Vicariance and dispersal in the Alpine perennial *Bupleurum stellatum* L. (Apiaceae). *Taxon* **54:** 725–732.

Shimizu-Inatsugi R, Lihová J, Iwanaga H, Kudoh H, Marhold K, Savolainen O, Watanabe K, Yakubov VV, Shimizu KK. 2009. The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. *Molecular Ecology* **18:** 4024–4048.

Simmons MP, Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49:** 369–381.

Skrede I, Borgen L, Brochmann C. 2009. Genetic structuring in three closely related circumpolar plant species: AFLP versus microsatellite markers and High-Arctic versus Arctic-Alpine distributions. *Heredity* **102:** 293–302.

Stolpe C, Krämer U, Müller C. 2017. Heavy metal (hyper) accumulation in leaves of *Arabidopsis halleri* is accompanied by a reduced performance of herbivores and shifts in leaf glucosinolate and element concentrations. *Environmental and Experimental Botany* **133:** 78–86.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7:** 453–464.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28:** 2731–2739.

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

Těšitel J, Malinová T, Štech M, Herbstová M. 2009. Variation in the *Melampyrum sylvaticum* group in the Carpathian and Hercynian region: two lineages with different evolutionary histories. *Preslia* **81:** 1–22.

The 1001 Genomes Consortium. 2016. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166:** 481–491.

Tribsch A, Schönswetter P, Stuessy TF. 2002. *Saponaria pumila* (Caryophyllaceae) and the Ice Age in the European Alps. *American Journal of Botany* **89:** 2024–2033.

Trucchi E, Frajman B, Haverkamp THA, Schönswetter P, Paun O. 2016. Genomic and metagenomic analyses reveal parallel ecological divergence in *Heliosperma pusillum* (Caryophyllaceae). *bioRxiv 044354*; doi: https://doi.org/10.1101/044354

Tzedakis PC, Emerson BC, Hewitt GM. 2013. Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology and Evolution* **28:** 696–704.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* **35:** 71–74.

Van Rossum F, Bonnin I, Fénart S, Pauwels M, Petit D, Saumitou-Laprade P. 2004. Spatial genetic structure within a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy-metal-tolerant species. *Molecular Ecology* **13:** 2959–2967.

Verbruggen N, Hermans C, Schat H. 2009. Molecular mechanisms of metal hyperaccumulation in plants. *New Phytologist* **181:** 759–776.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23:** 4407–4414.

Wasowicz P, Pauwels M, Pasierbinski A, Przedpelska-Wasowicz EM, Babst-Kostecka AA, Saumitou-Laprade P, Rostanski A. 2016. Phylogeography of *Arabidopsis halleri* (Brassicaceae) in mountain regions of Central Europe inferred from cpDNA variation and ecological niche modelling. *PeerJ PrePrints* **3:** 1–29.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38:** 1358–1370.

Willems G, Dräger DB, Courbot M, Godé C, Verbruggen N, Saumitou-Laprade P. 2007. The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): an analysis of quantitative trait loci. *Genetics* **176:** 659–674.

Winkler M, Tribsch A, Schneeweiss GM, Brodbeck S, Gugerli F, Holderegger R, Abbott RJ, Schönswetter P. 2012. Tales of the unexpected: phylogeography of the Arctic-Alpine model plant *Saxifraga oppositifolia* (Saxifragaceae) revisited. *Molecular Ecology* **21:** 4618–4630.

Wolfe KH, Li W-H, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America* **84:** 9054–9058.

Wolfe KH, Sharp PM, Li W-H. 1989. Mutation rates differ among regions of the mammalian genome. *Nature* **337:** 283–285.

Wołoszczak E. 1896. Z granicy flory zachodnio- i wschodnio-karpackiej. *Akademija Umiejętności w Krakowie, Sprawozdanie Komisji Fizjograficznej* **31:** 119–159.

Zhivotovsky LA. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology* **8:** 907–913.

Zhou TY, Lu LL, Yang G, Al-Shehbaz IA. 2001. Brassicaceae. In: Wu ZY, Raven PH, eds. *Flora of China*, *vol*. **8**. Beijing & St. Louis: Science Press & Missouri Botanical Garden Press, 1–193.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Details of the 82 populations of *A. halleri* sampled in the present study. Samples selected for testing of variation in plastid DNA regions are highlighted in bold in the column 'cpDNA_ind'.
Table S2. List of primers used for the amplification of five microsatellite loci and the amplification and sequencing of three plastid DNA regions.
Table S3. AMOVA – analysis of molecular variance based on AFLP data
Table S4. Divergence times (Mya) estimated in three independent species tree (population tree) reconstructions with different *a priori* settings for the mutation rate of the *CHS* marker. A visualization of three alternative divergence time estimates presented in this table is presented in Fig. S2.
Table S5. Morphometric characters scored in morphologically screened populations.
Data S1. Matrix of AFLP fragments used in the analyses.
Data S2. Matrix of microsatellite alleles used in the analyses.
Data S3. Plastid DNA and *CHS* sequence alignments and list of GenBank accessions used in the analyses.
Figure S1. Summary of results of Structure analyses of the Eurasian (A) and European (B) AFLP dataset of *Arabidopsis halleri*. Values of ln probability of the data for each number of groups (*K*) are plotted against *K* values in the first panel; similarity coefficients among runs are plotted against *K* values in the second panel and the delta *K* value in the third panel.
Figure S2. Three alternative dated species tree reconstructions of the genus *Arabidopsis* with a focus on infraspecific relationships in *Arabidopsis halleri* based on plastid DNA and *CHS* data. Alternative reconstructions differ by prior settings of the *CHS* mutation rate: at the top *CHS* rate was set up to be 9.0×10^{-3} substitutions per site per My (Wolfe, Sharp & Li, 1989), in the middle it was 1.5×10^{-2} substitutions per site per My (Koch *et al.*, 2000) and at the bottom 7.0×10^{-3} substitutions per site per My (Durbin *et al.*, 1995). Although the topology of the alternative species trees is the same (depicted only once with the first alternative) diversification times estimated for the plastid DNA gene tree (light grey and light red), *CHS* gene tree (darker grey and darker red) and species tree (dark grey and dark red) differ as depicted by 95% HPD intervals and mean values (in kya). See legend below the picture for colour scale explanations.
Figure S3. Loadings of 33 morphological characters on the first two axes of a principal component analysis of the 41 morphologically screened *A. halleri* populations.
Figure S4. Relative contributions of elevation, bioclimatic variables, genetic structure (AFLP-based genetic distances) and geographical distances to the morphological variation of populations based on variation partitioning from direct ordinations (redundancy analyses).

Paper IV

Šrámková G, Kolář F, Záveská E, Lučanová M, Španiel S, Kolník M, Marhold K (2019) Phylogeography and taxonomic reassessment of *Arabidopsis halleri* – a montane species from Central Europe. *Plant Syst Evol* 305, 885–898,
and (2020) Correction to: Phylogeography and taxonomic reassessment of *Arabidopsis halleri* – a montane species from Central Europe. *Plant Systematics and Evolution* 306, 60.

Plant Systematics and Evolution (2019) 305:885–898
https://doi.org/10.1007/s00606-019-01625-y

ORIGINAL ARTICLE

Phylogeography and taxonomic reassessment of *Arabidopsis halleri* – a montane species from Central Europe

Gabriela Šrámková[1] · Filip Kolář[1,2,3] · Eliška Záveská[2,3] · Magdalena Lučanová[1,3,5] · Stanislav Španiel[1,4] · Martin Kolník[4] · Karol Marhold[1,4]

Received: 21 August 2019 / Accepted: 29 October 2019 / Published online: 27 November 2019

Abstract
Evolutionary histories of plants from the mid-elevation (montane) zone of European mountain ranges have only rarely been documented, standing in contrast to those of well-researched inhabitants of (sub-)alpine and foothill zones. To fill this gap, we have reconstructed the phylogeography of *Arabidopsis halleri*, a species preferring coniferous woodlands and corresponding secondary habitats in the montane zone of the Alps, Carpathians, Hercynian massif and Dinaric Alps. Based on range-wide sampling and finer-scale analyses of multiple multilocus DNA markers, we have addressed phylogeographic patterns among the Carpathian populations and inferred their relationships to *A. halleri* from neighbouring mountain ranges. We also present a taxonomic re-evaluation of the species in Europe, based on the revealed genetic structure complemented by morphological data. Besides two distinct Alpine groups, we identified a major phylogeographic split between the Western and South-Eastern Carpathians. Interestingly, Western and South-Eastern Carpathian populations were genetically closer to populations from neighbouring mountain ranges (the Hercynian massif and the Dinaric Alps for the Western and South-Eastern Carpathians, respectively) than they were to each other, likely reflecting long-term isolation in different parts of the Carpathians or different (re)colonization pathways during the Holocene. In spite of the considerable environmentally correlated variation, the five major European genetic groups exhibited distinctive morphological characters, and we therefore propose treating them as separate subspecies: *A. halleri* subsp. *halleri* (Western Europe, Hercynian massif), *A. halleri* subsp. *tatrica* (Western Carpathians), *A. halleri* subsp. *ovirensis* (Eastern Alps), *A. halleri* subsp. *occidentalis* (Western Alps) and *A. halleri* subsp. *dacica* (Eastern and Southern Carpathians and Dinaric Alps).

Keywords *Arabidopsis* · AFLPs · Microsatellites · Multivariate morphometrics · Taxonomy

Introduction

Over the last two decades, considerable attention has been paid to the phylogeographic patterns of numerous plant species in Central Europe, in particular those occurring in subalpine and alpine belts of the Alps and Carpathians (Ronikier 2011; Puşcaş and Choler 2012; Pachschwöll et al. 2015; Mráz and Ronikier 2016). Less attention has been paid to

Handling editor: Michal Ronikier.

Contribution to "Biogeography of the Carpathians".

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00606-019-01625-y) contains supplementary material, which is available to authorized users.

✉ Karol Marhold
karol.marhold@savba.sk

[1] Department of Botany, Faculty of Science, Charles University, Benátská 2, 128 01 Prague, Czechia

[2] Department of Botany, University of Innsbruck, Sternwartestraße 15, 6020 Innsbruck, Austria

[3] Institute of Botany, The Czech Academy of Sciences, 252 43 Průhonice, Czechia

[4] Plant Science and Biodiversity Centre, Institute of Botany, Slovak Academy of Sciences, Dúbravská cesta 9, 845 23 Bratislava, Slovakia

[5] Department of Botany, Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czechia

species that occur at middle elevations or span from middle to (sub-)alpine elevations. Although the distribution patterns of genetic lineages in the Alps and Carpathians often considerably differ from species to species, major phylogeographic trends are emerging, for example strong the biogeographic border between the Western and Eastern Alps or between the Western and South-Eastern Carpathians, and the existence of multiple glacial refugia along the entire Carpathian arch (Thiel-Egenter et al. 2011; Alvarez et al. 2012; Ronikier et al. 2012). Moreover, contrary to early phylogeographic hypotheses that saw the only European glacial refugia of the temperate biota in the three major southern peninsulas (Iberian, Apennine and Balkan ones; Taberlet et al. 1998; Hewitt 2000), extended sampling and detailed population-level analyses of genetic variation as well as palaeopalynological data provided evidence for additional, extra-Mediterranean refugia in Central and Eastern Europe (Stewart and Lister 2001; Bhagwat and Willis 2008; Juřičková et al. 2014). In sum, recent botanical and zoological research in the area has pointed out the significant role of the Carpathians both in glacial survival of temperate species and as a hotspot of (still poorly known) diversity and endemism in Europe (Juřičková et al. 2014; Mráz and Ronikier 2016).

Arabidopsis halleri (L.) O´Kane and Al-Shehbaz is a diploid, clonal, self-incompatible, outcrossing perennial species. The substantial part of its genetic variation is concentrated in the Central European mountain ranges, particularly in the Alps, the Hercynian massif and the Carpathians (Pauwels et al. 2012; Šrámková-Fuxová et al. 2017; Wasowicz et al. 2016). It is mainly a species of the montane belt, which makes it an interesting model for addressing phylogeographies of European mountain ranges, complementing those of more intensively researched foothill and (sub-)alpine areas. *Arabidopsis halleri* also represents a key model for studying the genetic basis of phytoremediation and heavy metal hyperaccumulation (Willems et al. 2007; Pauwels et al. 2008a, b, 2012; Verbruggen et al. 2009; Krämer 2010; Stein et al. 2017; Stolpe et al. 2017; Preite et al. 2019). Although the related model species *A. thaliana* (L.) Heynh. has been under intensive study for more than 40 years, the circumscription, taxonomy and evolutionary relationships within and among its wild relatives, *A. halleri* and *A. arenosa* in particular, are only poorly known (Hohmann et al. 2014; Koch 2019). Consequently, experimental studies using *Arabidopsis* species may get confounded by a misleading taxonomy or an inaccurate phylogenetic framework. This hampers the comparison of results between studies and also poses a risk of inaccurate conclusions drawn within a misleading evolutionary context (Koch et al. 2008).

According to recent taxonomic treatments (Jones and Akeyrod 1993; Jalas and Suominen 1994; Kolník and Marhold 2006), two to four subspecies are recognized within European *A. halleri*, namely *A. halleri* subsp. *halleri*, *A. halleri* subsp. *tatrica* (Pawł.) Kolník, *A. halleri* subsp. *ovirensis* (Wulfen) O'Kane & Al-Shehbaz, *A. halleri* subsp. *dacica* (Heuff.) Kolník. One additional subspecies, *A. halleri* subsp. *gemmifera* (Matsum.) O'Kane & Al-Shehbaz (sometimes treated as a separate species *A. gemmifera* (Matsum.) Kadota), occurs in East Asia, is genetically distinct and separated from European populations by a large distribution gap (Šrámková-Fuxová et al. 2017) and is therefore beyond the scope of the present study.

The genetic structure of European populations of *Arabidopsis halleri* has been the subject of several studies using a wide spectrum of molecular markers (Pauwels et al. 2008a, b, 2012; Wasowicz et al. 2016; Šrámková-Fuxová et al. 2017). In a recent study based on the most comprehensive range-wide sampling and multiple markers (AFLPs, nuclear microsatellites and sequences of single-copy nuclear regions and plastid DNA), Šrámková-Fuxová et al. (2017) identified three major genetic lineages within the European range of this species that correspond well with geography: (1) Western and Central Europe (including the Western Carpathians), (2) the Alps and (3) the Eastern and Southern Carpathians, and the Dinaric Alps. Molecular data have also indicated the Pleistocene origin of these lineages and their in situ glacial persistence in multiple northern refugia of eastern Central Europe.

However, apart from the clearly separate position of the Japanese taxon *A. halleri* subsp. *gemmifera*, the genetic structure revealed in neither of these studies fully fits the taxonomic concept of European subspecies of *A. halleri* suggested by Kolník and Marhold (2006), which was, in turn, based solely on morphological data. Furthermore, the overall morphological variation within this species is strongly influenced by plastic responses to major environmental gradients (Šrámková-Fuxová et al. 2017), which is reflected particularly in the parallel origin of its distinct alpine phenotype. Therefore, a number of authors have called for the revision of the taxonomic treatment of *Arabidopsis halleri*, in particular by integrating molecular and morphological data (Koch and German 2013; Wasowicz et al. 2016; Šrámková-Fuxová et al. 2017; see also Table 1).

To provide a detailed phylogeographic perspective and revised taxonomic classification of European populations of *A. halleri*, we re-evaluated the molecular (amplified fragment length polymorphism, AFLP, and nuclear microsatellite) and morphometric data published by Šrámková-Fuxová et al. (2017), complemented with morphological data from additional populations. The main aims of the current article were (1) to explore the finer-scale genetic structure of *A. halleri* within its European distribution area with special focus on Carpathian biogeography; (2) to check whether there are any morphological characters defining the major genetic lineages within this species; and (3) to provide a taxonomic treatment of the infraspecific variation of *A. halleri* enabling

Table 1 Comparison of genetic lineages revealed in this paper with taxonomic classifications of European *Arabidopsis halleri*

Genetic lineages recognized here	Hercynian	Western Alpine	Eastern Alpine	Western Carpathian	South-Eastern
Classification proposed here	halleri	occidentalis	ovirensis	tatrica	dacica
Phylogeographic units and subunits by Pauwels et al. (2012)	North-Western 1–6, hybrid zone 4	South-Eastern 1–3	Hybrid Zone 3	Hybrid zones 1,2	South-Eastern 4,5
Jones and Akeroyd (1993)	halleri	halleri	halleri and ovirensis	halleri and intermediate populations between halleri and ovirensis	halleri and ovirensis
Jalas and Suominen (1994)	halleri	halleri	halleri and ovirensis	halleri and ovirensis	halleri and ovirensis
Kolník and Marhold (2006)	halleri	halleri	halleri and ovirensis	tatrica	halleri and dacica

dacica–*A. halleri* subsp. *dacica*, halleri–*A. halleri* subsp. *halleri*, occidentalis–*A. halleri* subsp. *occidentalis*, ovirensis–*A. halleri* subsp. *ovirensis*, tatrica–*A. halleri* subsp. *tatrica*

correct reference to experimental material of this important model species.

Materials and methods

Plant material

Samples from 88 populations were collected during 2002–2013 with the aim to include multiple representatives of all four hitherto recognized European subspecies (*A. halleri* subsp. *halleri*, *A. halleri* subsp. *tatrica*, *A. halleri* subsp. *dacica* and *A. halleri* subsp. *ovirensis* (Kolník and Marhold 2006; Hohmann et al. 2014) and genetic lineages (Šrámková-Fuxová et al. 2017) across their entire distribution areas (see Online Resource 1 for locality details). Individual plants were sampled throughout their sites at maximum distances, but no less than 3 m apart, to minimize the chance of collecting clones (Van Rossum et al. 2004). All population samples were checked for variation in ploidy level using flow cytometry as described by Šrámková-Fuxová et al. (2017). Four of these samples are documented by chromosome number counts by Kolník and Marhold (2006).

The morphometric dataset comprises 55 population samples (8–49 individuals per population, 19 on average, 1047 plants altogether). Of these, 42 population samples were already explored by Šrámková-Fuxová et al. (2017) although in a different context of a study focused on major environmental correlates of morphological variation. Thirteen additional population samples (altogether 135 plants, scored using the same methodology, see Online Resource 1) were included in this study to provide representative sampling for taxonomic evaluation. Voucher specimens of the samples are deposited in the herbarium of Charles University in Prague, Czechia (PRC; samples collected in 2011–2013) and in the herbarium of the Plant Science and Biodiversity Centre of the Slovak Academy of Sciences, Bratislava, Slovakia (SAV; samples collected in 2002–2003).

The genetic dataset comprised 75 populations genotyped by means of AFLPs and microsatellites that were taken from our previous study focused on broad genetic relationships across the entire species (Šrámková-Fuxová et al. 2017) and re-analysed here in order to infer finer genetic structure among European populations only. Specifically, we analysed (i) 715 individuals from 75 populations in Europe (5–15 plants per population, 10 on average) genotyped by amplified fragment length polymorphism (AFLP) analysis (Vos et al. 1995) and (ii) 711 individuals from the same 75 populations genotyped by five unlinked microsatellite loci (ICE13, NGA, ATT, ICE14 and ATH; Clauss et al. 2002; 215 plants per population, 10 on average). The detailed protocol is specified in Šrámková-Fuxová et al. (2017). The resulting data matrices are available in Online Resources 6 and 7.

In total, 42 populations were included in both the morphometric and genetic datasets; the assignment of the remaining 13 populations, evaluated for morphometry only, to genetic groups was based on their geographical location. Such an assignment was made possible by the clear spatial differentiation of all genetic clusters and very low admixture among them evidenced by genetic data (see Results).

Analysis of genetic data

To detect structuring among the AFLP genotypes, we used STRUCTURE 2.3.2 (Pritchard et al. 2000; Falush et al. 2007). We applied the following hierarchical approach in order to detect finer genetic substructuring. First, we analysed the total dataset of all European populations and inferred major genetic grouping. Secondly, we carried out separate analyses of each of the major groups. Thirdly, to test the stability of substructuring within each major group given the sampling available, we created (for each major group) five subsampled data matrices: Three new

data matrices were created by random exclusion of 20% of populations and two by excluding the same number of furthermost populations (one matrix without westernmost and one without easternmost populations, northernmost and southernmost for the South-Eastern group, respectively). For all STRUCTURE analyses in each step, we used the admixture model with uncorrelated allele frequencies and recessive alleles (i.e. 0, suitable for dominant data) that was run ten times for each K (number of groups) ranging from 1 to 10, using a burn-in of 100,000 iterations followed by 1000,000 additional MCMC iterations. To infer the optimal number of groups for each STRUCTURE analysis, we calculated the following parameters using the R script Structure-sum (Ehrich 2006): (i) the posterior probabilities of the runs (Rosenberg et al. 2002), (ii) the similarity coefficient between the runs, and (iii) delta K as defined by (Evanno et al. 2005). The optimal value of K was defined as the partition where values of likelihood (natural logarithm of probability) started to flatten out, the results of replicate runs were close to identity (i.e. with a high similarity coefficient with no scatter between runs), and mean delta K was the greatest (Evanno et al. 2005). To verify the STRUCTURE partitioning, we also displayed genetic distances among the individuals assigned to particular groups using principal coordinate analysis (PCoA) based on Jaccard distance calculated in *adegenet v1.4-2* (Jombart 2008). To capture the pairwise genetic separation of genetic lineages, we also visualized genetic differentiation between each pair using this method (PCoA).

We calculated Nei's (1987) gene diversity and the frequency-down-weighted marker index (DW; Schönswetter and Tribsch 2005) from the presence-absence matrix of alleles using the R script AFLPdat (Ehrich 2006) for each population and for each of the major geographical groups identified by STRUCTURE. The DW value is expected to be greater in populations that harbour a high number of rare alleles, meaning alleles with low frequency in the total dataset (e.g. Paun et al. 2008). Differences in diversity and the DW index between the five lineages were tested by one-way ANOVA (R package stats).

Five microsatellite loci were used as a complementary approach to the dominant AFLP markers to characterize the patterns of within-population diversity and among-population differentiation. We estimated observed (H_o) and expected heterozygosity (H_e; both averaged over loci variable in that population) and average number of alleles (allelic richness computed through rarefaction on the small sample size of a minimum of six individuals, 1000 permutations) in MSA v4.05 (Dieringer and Schlötterer 2003). Inter-population differentiation was characterized by means of the fixation coefficient (Fst) calculated also in MSA v4.05 (Dieringer and Schlötterer 2003). Differences in heterozygosity indices, allelic richness and the DW index between the five lineages were tested by one-way ANOVA (R package stats).

Morphological measurements

We used a matrix of 33 morphological characters (19 quantitative, 9 binary and 5 ratios, see Online Resource 3 for details) that were scored on voucher specimens of 1047 field-collected individuals. Morphological characters were chosen with an emphasis on key morphological characters used in the *Brassicaceae* taxonomy (e.g. trichome morphology and occurrences). The resulting data matrices are available in Online Resource 8.

Analysis of morphological data

The point of the morphometric analyses carried out in the present study was to test whether genetic lineages identified within European populations of *A. halleri* are sufficiently morphologically different to merit their taxonomic recognition and to find the best morphological characters for their identification. To this end, we created matrices of morphological characters based on values measured on individual plants as well as on mean values of characters per given population. To reveal the correlation structure of all measured/scored characters and to ensure that characters were not very highly correlated (values > 0.95), which would potentially distort some of the discriminant analyses, the Pearson and the nonparametric Spearman rank correlation coefficients were computed from the matrix for individual plants. Then, to explore the extent of morphological differentiation among genetic lineages and to find which characters or groups of characters best differentiate these lineages, we employed canonical and nonparametric k-nearest neighbours classificatory discriminant analyses (Klecka 1980; Krzanowski 1990; Marhold 2011). Canonical discriminant analyses visualize relationships among objects in the space defined by canonical axes while maximizing differences between predefined groups of objects. Characters that differ between groups and have low within-group variation have the greatest influence on canonical axes. On the other hand, classificatory discriminant analyses derive one or more functions aiming at the identification of objects. Whereas canonical discriminant functions maximize the separation of predefined groups, classificatory functions minimize the number of misidentified objects (Marhold 2011).

Initially, we performed canonical discriminant analysis of populations, characterized by the mean values of measured/scored characters using genetic lineages identified by molecular analyses (AFLP analyses in particular) as groups (as already stated above, populations that were not included in molecular analyses were assigned to groups according to

geographical criteria and populations with admixture were assigned according to the majority rule). In the second step, we performed separate canonical and classificatory discriminant analyses of each pair of genetic lineages as groups, using individual plants. To find out whether genetic lineages can be treated as morphologically defined taxa, the percentage of plants correctly classified into the predefined groups was calculated and the separation of individuals along the discriminant axes was visualized. Total canonical structure (correlations of morphological characters with canonical axes) was used to identify characters that best discriminate each pair of genetic lineages.

Linear discriminant functions for identification purposes, based on characters with the greatest utility for identifying genetic lineages/taxa, were calculated for each pair of them (Klecka 1980; Krzanowski 1990; Marhold 2011). Characters for discriminant functions were selected by means of stepwise discriminant analysis (Klecka 1980).

Results

Ploidy level

Based on flow-cytometric analyses of multiple plants from all populations studied, all of the material of *A. halleri* examined was diploid (having DNA content values on average 2C = 0.44 pg DNA) except for one triploid individual from population AH086 (with approx. 1.5-fold greater nuclear DNA content, 2C = 0.71 pg, for more details see Šrámková-Fuxová et al. 2017).

Genetic data

AFLP fingerprinting of 715 individuals from 75 European populations yielded a total of 325 polymorphic fragments (96% overall reproducibility). Bayesian clustering of all sampled populations resulted in clustering into three geographically well-delimited groups (this number had both the greatest delta K and similarity among runs, Online Resource 2) occupying (i) the Hercynian massif and Western Carpathians (North-Western group), (ii) the South-Eastern Carpathians and Dinaric Alps (South-Eastern group) and (iii) the Alps (Alpine group). To examine the potential substructuring of the data, which might correspond with one of the previously proposed classifications, we analysed each major group separately (see Online Resource 2). Bayesian clustering suggested the presence of two subgroups within each North-Western and Alpine group and five subgroups of South-Eastern populations. To test the stability of these substructures, we applied the same analyses to subsampled data matrices for each major group. This clustering revealed a stable genetic structure within the North-Western and Alpine groups, but not in the South-Eastern one, where each subsampled matrix led to different genetic subgroups (data not shown). For the interpretation below, we thus define the following five stable genetically defined lineages: Western Alpine, Eastern Alpine, Hercynian, Western Carpathian and South-Eastern Carpathian (see Fig. 1). Principal coordinate analyses of AFLP data confirmed this clustering (see Fig. 1). Pairwise genetic separation of five genetic lineages based on AFLP markers visualized by principal coordinate analyses revealed good separation of almost all pairs, the only exception being the Hercynian and Western Carpathian lineages (however, their overlap was only marginal, see Fig. 4).

We further analysed the patterns of population genetic diversity and differentiation using AFLP as well as codominant microsatellite data (Table 2). Populations belonging to the Western Carpathian lineage exhibited negligible differentiation from the Hercynian lineage (Fst based on microsatellites = 0.05) while being markedly distinct from the South-Eastern lineage (Fst = 0.23, Table 3). Both AFLP (DW index) and microsatellite data (allelic richness) congruently showed that both Western Carpathian and Eastern Alpine populations of *A. halleri* harboured higher levels of rare genetic markers compared to populations from neighbouring regions (Hercynian massif and Western Alps, Table 2). Populations from the South-Eastern lineage exhibited on average greater values of indicators of both genetic diversity and rarity than their Western Carpathian counterparts. To compare populations from the South-Eastern Carpathians with those from neighbouring regions and to discuss possible phylogeographic pattern, we also calculated genetic diversities separately for geographically defined group from the Dinaric mountains (populations AH92, AH93; DW index = 0.52 (0.50–0.55), allelic richness = 1.91 (1.64–2.17), $Fst_{(vs\ SEgroup)} = 0.11$).

Multivariate morphometrics

Because the distribution of characters deviated from a normal one, we employed nonparametric methods wherever possible. Therefore, we used not only Pearson, but also nonparametric Spearman correlation coefficients, to check for highly correlated characters. As a result, we found only one character pair, namely Length_Leafy_Stem and Dist_Shortest_Leaf, with a correlation greater than 0.9 (0.948 of Pearson and 0.930 of Spearman correlation coefficients). As none of character pairs exhibited very high correlations (> 0.95), all characters were used in further analyses.

Ordination diagrams of canonical discriminant analysis of populations characterized by mean values of characters measured on individual plants clearly show the morphological separation of the five genetic lineages. Four lineages are separable along the first two canonical axes whereas the last two (Eastern Alpine and Western Carpathian) lineages are

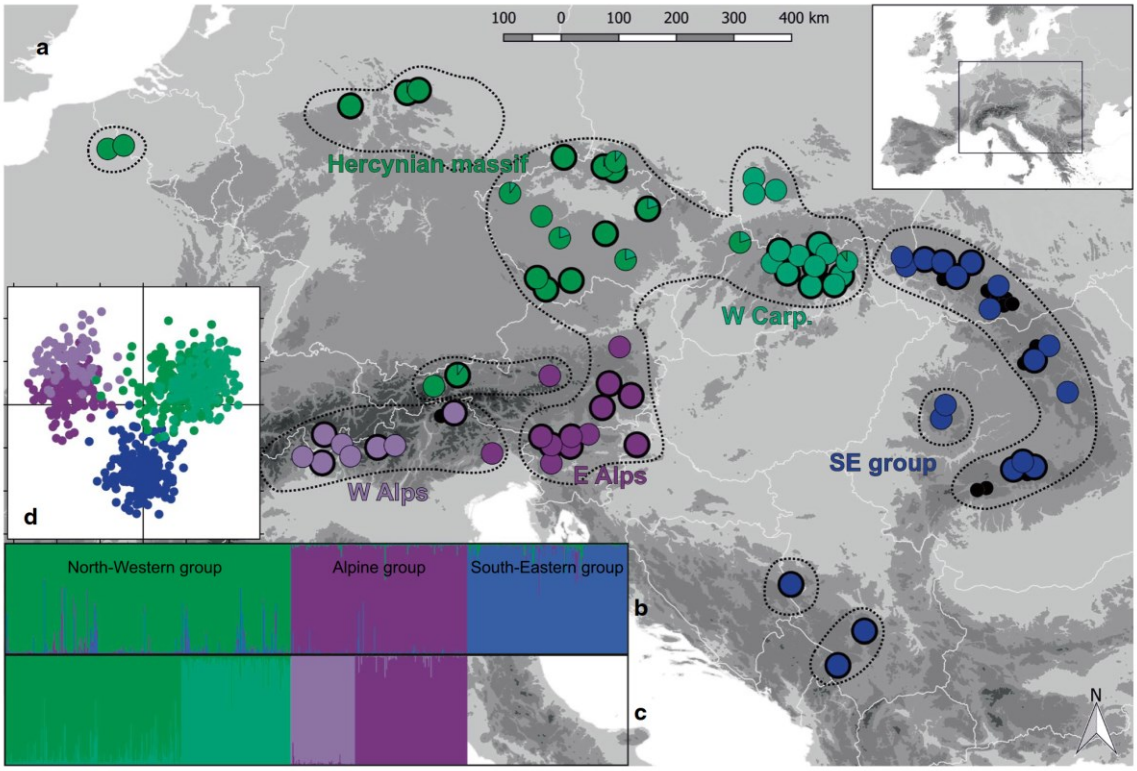


Fig. 1 Range-wide genetic differentiation of *Arabidopsis halleri* across its European range. **a** Geographic distribution of populations sampled (colour pie charts reflecting the proportional assignment to particular STRUCTURE subgroups within each group, thick borders represent populations used also for morphological measurements, and black dots represent populations sampled for morphological measurements only); the dotted line denotes the borders of the distribution range of *A. halleri*. **b** Cluster assignment of the 715 individual AFLP phenotypes revealed by STRUCTURE for the complete dataset. **c** Cluster assignment of individuals revealed by separate STRUCTURE analyses of the North-Western and Alpine groups (analysis of the South-Eastern group did not result in consistent results, see the text). **d** Principal coordinate analysis based on Jaccard distances between all AFLP phenotypes; the colour coding matches the assignment to the respective lineages as inferred by STRUCTURE

Table 2 Genetic diversity and differentiation of *Arabidopsis halleri* lineages identified by Bayesian clustering inferred from AFLP and five microsatellite markers

	Western Alpine	Eastern Alpine	Hercynian	Western Carpathian	South-Eastern
N# pops/indivs for AFLP	8/74	13/129	21/202	14/126	19/184
Elevation range (m a.s.l.)	330–1510	210–2100	50–1050	300–1880	300–2250
Total# of polymorph. fragments (AFLP)	212	210	223	222	277
Private AFLP fragments	4	10	4	1	17
Private SSR alleles	0	4	2	3	5
DW index (AFLP)[a]	0.45 (0.27–0.82)	0.50 (0.26–0.75)	0.32 (0.14–0.89)	0.40 (0.20–0.75)	0.62 (0.38–1.10)
Gene diversity (AFLP)[a]	0.06 (0.05–0.07)	0.07 (0.04–0.10)	0.06 (0.03–0.12)	0.07 (0.06–0.09)	0.08 (0.05–0.12)
Expected heterozygosity (SSR)[a]	0.36 (0.27–0.51)	0.51 (0.30–0.66)	0.45 (0.21–0.67)	0.47 (0.33–0.58)	0.51 (0.23–0.65)
Fst (SSR)	0.35***	0.16***	0.25***	0.14***	0.19***
G'st (SSR)	0.54***	0.41***	0.51***	0.33***	0.54***
Allelic richness (SSR)	2.19 (1.77–2.83)	3.12 (2.15–4.52)	2.61 (1.57–3.63)	3.03 (2.25–3.78)	3.39 (1.64–4.70)

[a]Mean (min–max) per-population values

*$p<0.05$, **$p<0.01$, ***$p=0.001$

Table 3 Pairwise genetic differentiation (Fst) among geographically and genetically delimited regional groups/lineages of *Arabidopsis halleri* populations inferred from the microsatellite data

	W Alps	E Alps	Herc.	W Carp.
E Alps	0.166			
Herc.	0.162	0.07		
W Carp.	0.214	0.067	0.049	
SE lineage	0.2612	0.235	0.161	0.226

All values were significant ($P = 0.001$)

W Alps Western Alpine lineage (*A. halleri* subsp. *occidentalis*), *E Alps* Eastern Alpine lineage (*A. halleri* subsp. *ovirens*), *Herc.* Hercynian lineage (*A. halleri* subsp. *halleri*), *W Carp.* Western Carpathian lineage (*A. halleri* subsp. *tatrica*), *SE lineage* South-Eastern lineage (*A. halleri* subsp. *dacica*)

separated along the third canonical axis(Fig. 2). Specifically, populations from the Western Alps, Hercynian massif and Western Carpathians + Eastern Alps are separated along the first canonical axis (highly correlated with the following characters: number of pairs of teeth of the longest basal rosette leaf, width of a narrower sepal, presence/absence of forked trichomes on the second lowermost stem leaf or simple trichomes on flower buds) whereas South-Eastern populations are separated from the others along the second canonical axis (highly correlated with the following characters: ratio of the length of the leafy part of the stem and number of stem leaves on the main stem, presence or absence of trichomes on the leafy part of the stem, distance from the base of the stem to the longest leaf) and characters correlated with the third canonical axis (presence or absence of 3-forked trichomes on flower buds, distance from the base of the stem to the longest stem leaf, ratio of the distance from the base of the stem to the longest stem leaf and length of the leafy part of the stem) separate Western Carpathian and Eastern Alpine populations. Boxplots displaying the diversity per group of quantitative characters are presented in Fig. 3; trichome variation is summarized in Table 4. (10 and 90 percentiles and median values of all characters are compiled in Online Resource 4.)

Canonical and classificatory discriminant analyses based on individual plants that were grouped according to genetic clustering (five lineages) and compared in a pairwise manner showed at least reasonable morphological separation of all lineages. In all classificatory discriminant analyses, at least 90% of plants were classified to the correct group (Fig. 4). The best-separated groups were Western Alpine versus Western Carpathian populations (100% classification success), whereas the worst-separated, with 97% and 90% of correctly classified plants, were Hercynian and Western Carpathian populations, respectively, and with 92% and 91% of correctly classified plants, Eastern Alpine and Hercynian populations, respectively. Canonical discriminant analyses yielded similar results to those depicted in histograms in Fig. 4. Correlations of morphological characters with particular axes are presented in Online Resource 5.

Discriminant functions enabling the identification of infraspecific taxa, together with their full morphological descriptions, are provided in Online Resource 9.

Discussion

Overall genetic differentiation of European Arabidopsis halleri

Over the last 15 years, several studies have dealt with phylogeographic patterns of European populations of *A. halleri* based on various organellar and nuclear markers, mostly in the context of addressing metal tolerance and occurrence of populations on metalliferous and non-metalliferous soils (Pauwels et al. 2005, 2008a, b, 2012; Wasowicz et al. 2016; Šrámková-Fuxová et al. 2017). For this reason, the sampling of the majority of these studies was focused on regions rich in metalliferous localities. Contrary to most of the previous

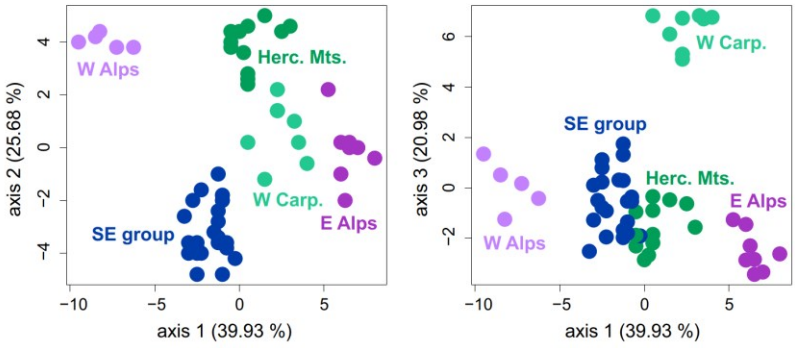


Fig. 2 Morphological separation of the five subgroups of European *Arabidopsis halleri* revealed by means of canonical discriminant analysis based on population means of the complete dataset (33 characters, 55 populations)

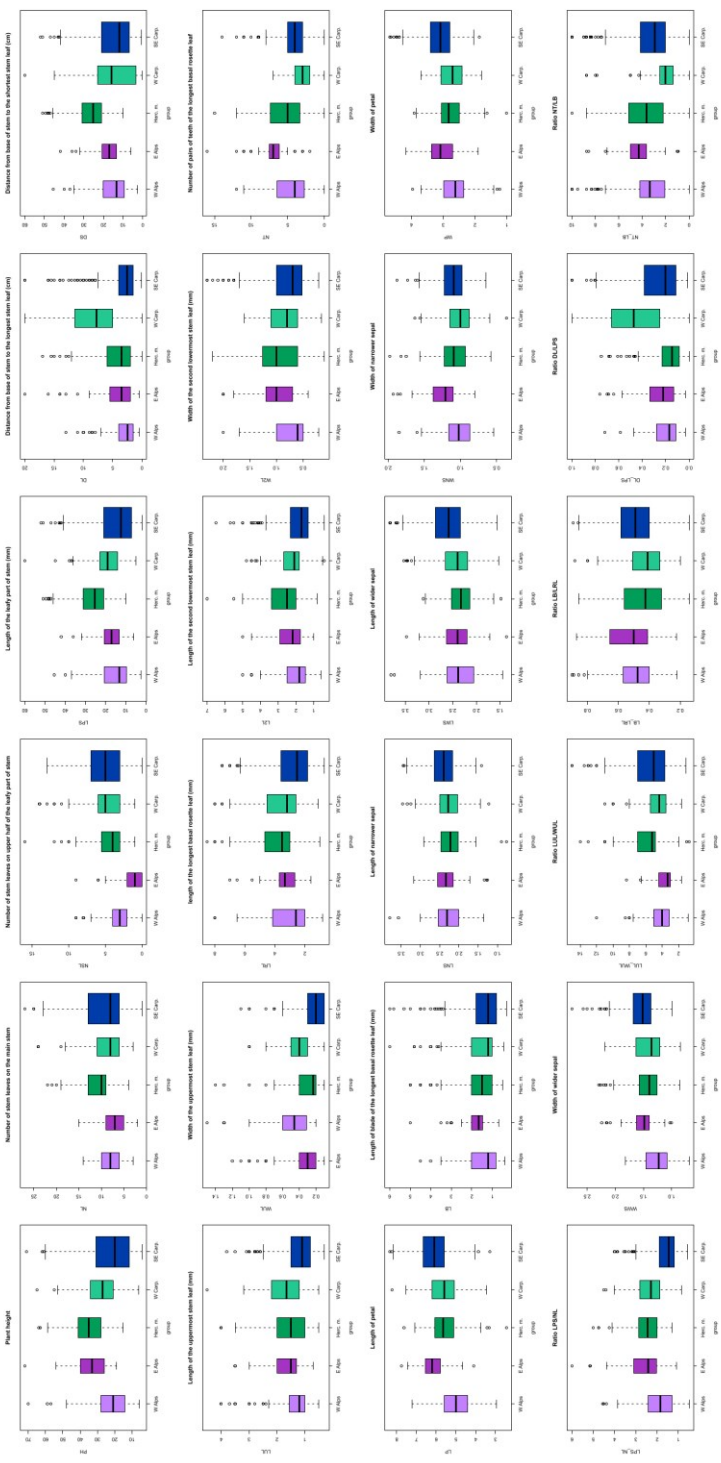


Fig. 3 Range of values of 24 continuous morphological characters (19 quantitative and 5 ratios) for five genetic lineages of European *Arabidopsis halleri*. The box represents values between the first and third quartile of the data, the thick horizontal line represents the second quartile (median), whiskers show values greater/smaller than 1.5× the interquartile range from the given quartile, and points represent outliers

Table 4 Summary of nine binary morphological characters among five lineages, numbers of morphologically scored individuals from lineage are in the brackets

	W Alps ($n = 99$)		E Alps ($n = 132$)		Herc. ($n = 198$)		W Carp. ($n = 165$)		SE lineage ($n = 233$)	
	Absence	Presence	Absence	Presence	Absence	Presence	Absence	Presence	Absence	Presence
Colour of flower white	2	97	34	98	2	196	31	134	13	220
Trichomes on leafy part of stem	0	99	40	92	95	103	97	68	1	232
Trichomes on inflorescence	25	74	59	73	142	57	145	20	91	142
Simple trichomes on the second lowermost stem leaf	1	98	38	94	7	191	42	123	9	224
Forked trichomes on the second lowermost stem leaf	3	96	42	90	35	163	93	72	45	188
3-forked trichomes on the second lowermost stem leaf	29	70	59	73	79	119	146	19	129	104
Simple trichomes on flower buds	12	87	40	92	7	191	106	59	122	111
Forked trichomes on flower buds	27	72	42	90	20	178	135	30	132	101
3-forked trichomes on flower buds	90	9	92	40	170	28	163	2	220	13

W Alps Western Alpine lineage (*A. halleri* subsp. *occidentalis*), *E Alps* Eastern Alpine lineage (*A. halleri* subsp. *ovirens*), *Herc.* Hercynian lineage (*A. halleri* subsp. *halleri*), *W Carp.* Western Carpathian lineage (*A. halleri* subsp. *tatrica*), *SE lineage* South-Eastern lineage (*A. halleri* subsp. *dacica*)

studies, we homogeneously cover here the entire range of the species and include all hitherto described intraspecific units from Europe. The results of studies by Pauwels et al. (2005, 2008a,b) are generally congruent with our structure of genetic lineages (see Table 1). In particular, they separate Western Carpathian and South-Eastern lineages from each other, Western and Eastern Alpine ones, as well as both Alpine lineages from the Hercynian one. Nevertheless, their study did not separate Western Alpine and South-Eastern lineages, perhaps as a consequence of a different sampling scheme and markers used (see Šrámková-Fuxová et al. (2017) for a further discussion of this). Plastid DNA data of Wasowicz et al. (2016) clearly support the separation of Western and Eastern Carpathian populations of *A. halleri*; nevertheless, they, again, show a more detailed structure within our Hercynian lineage.

Carpathian populations of Arabidopsis halleri – *biogeographic implications*

The Carpathian mountains are among the major harbours of biodiversity in extra-Mediterranean Europe. However, their role in preserving intraspecific plant variation had, until recently, been unknown (reviewed by Mráz and Ronikier 2016). Considerable attention was paid during the last two decades particularly to the phylogeography of Carpathian vascular plant taxa distributed above the timberline, frequently in studies under the IntraBioDiv consortium (Gugerli et al. 2008) or inspired by the activities of this consortium. Still, however, species occurring predominantly at lower elevations remained neglected.

Arabidopsis halleri, preferring mid-elevations in the montane belt, represents an ecological group that has so far been overlooked by phylogeographic studies. Interestingly, the major genetic split dividing European populations of this species follows the depression between the Eastern and Western Carpathians (see also Wasowicz et al. 2016; Šrámková-Fuxová et al. 2017) despite a continuous presence of suitable habitats for *A. halleri* (montane forests and meadows) in the border zone. This borderline has long been known from the distributions of plant species having various ecological preferences (since Woloszczak 1896), and recently, it was shown to be the main barrier inphylogeographies of alpine plants (Mráz and Ronikier 2016). This barrier was studied using floristic (Pax 1898; Jasiewicz 1965), cytological (Mráz and Szeląg 2004) as well as genetic data (Mráz et al. 2007; Ronikier et al. 2008; Těšitel et al. 2009; Kolář et al. 2016). The fact that mountain groups on both sides of this barrier host populations with great genetic variation and a high proportion of rare markers (Table 2) suggests that this pattern resulted from historical factors, namely recolonization of the area from at least two different glacial refugia,

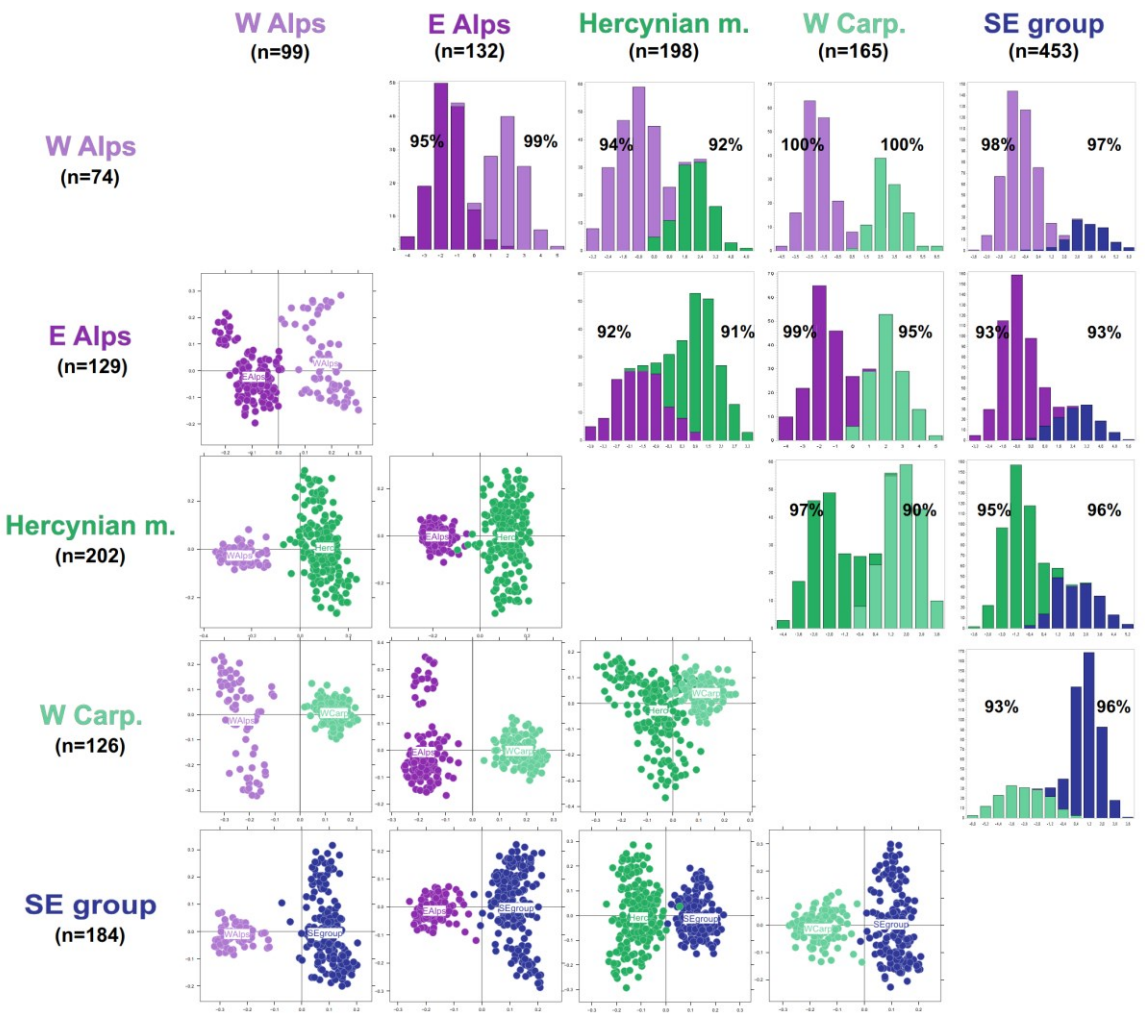


Fig. 4 Pairwise morphological and genetic separation of the five genetic lineages of European *Arabidopsis halleri*. Below the diagonal: genetic differentiation based on AFLP markers visualized by principal coordinate analyses (numbers of individuals per group are in brackets in the left column). Above the diagonal: morphological differentiation based on 33 morphological characters visualized as the positions of individuals on the first (discriminant) axis of the canonical discriminant analysis (numbers of individuals per group are in brackets in the first row; percentages of individuals correctly assigned to each group by classificatory discriminant analysis are denoted above corresponding parts of each histogram). The colour coding indicates the respective lineages as inferred by STRUCTURE: *W Alps* Western Alpine lineage (*A. halleri* subsp. *occidentalis*), *E Alps* Eastern Alpine lineage (*A. halleri* subsp. *ovirens*), *Hercynian m.* Hercynian lineage (*A. halleri* subsp. *halleri*), *W Carp.* Western Carpathian lineage (*A. halleri* subsp. *tatrica*), *SE group* South-Eastern lineage (*A. halleri* subsp. *dacica*)

rather than by an orographic barrier between the Western and Eastern Carpathians.

Overall, Carpathian populations of *A. halleri* from both the Western Carpathian and the South-Eastern groups exhibited elevated levels of genetic diversity and proportion of rare genetic markers, reaching levels that exceeded those in their neighbouring regions (the Hercynian massif and the Dinaric Alps). In addition, populations from these adjacent areas were genetically markedly closer (Hercynian massif) or exhibited similar divergence (Dinaric Alps) to their spatially closest Carpathian groups than was the divergence within the Carpathians themselves (Table 3). This pattern of genetically closer but genetically poorer adjacent areas fits well the scenario of the Carpathian Mts as a refugium of multiple distinct lineages that later contributed to the (re)colonization of the neighbouring

areas. To our knowledge, only a single species occupying a similar ecological niche as *A. halleri* (i.e. the montane zone) has been investigated so far—the annual hemiparasite *Melampyrum sylvaticum* (Těšitel et al. 2009). Interestingly, despite its very different life history, this species shows strikingly similar patterns of (i) genetic differentiation within the Carpathians (a major split between the Eastern and Western Carpathians), (ii) relationships to adjacent regions (close Western Carpathian and Hercynian populations) and (ii) distribution of genetic diversity (greatest diversity and proportion of rare haplotypes in the Eastern Carpathians). The case of *A. halleri* thus supports an emerging picture of at least two major harbours of intraspecific diversity in the Carpathians, situated in their Western and South-Eastern part, which correspond with probable glacial refugia for these montane species. These areas might have then served as sources for the later recolonization of the adjacent regions both towards the (north) west (Hercynian massif) and the south(east) (Dinaric Alps) (Mráz and Ronikier 2016; Kolář et al. 2016).

Phylogeographic division of populations in the Alps

Similarly to the genetic division within the Carpathians, we found considerable genetic (and also morphological) differences between Eastern and Western Alpine populations of *A. halleri*. The genetic break zone identified in *A. halleri* almost completely matches one of the two lines delimiting phylogeographically the Eastern and Western Alps referred to by Schönswetter et al. (2005: Fig. 2) and Thiel-Egenter et al. (2011) and called the "Brenner line". This line is located in the area west of the Dolomites, connecting Innsbruck (Austria) and Lago di Garda (Italy), and this delimitation of the Eastern and Western Alps goes back already to the work by Kerner (1870). There are a number of examples both from plant and animal taxa where genetic lineages are divided along the break lines of the Eastern and Western Alps (e.g. Albach et al. 2006; Kuss et al. 2011; Schmitt 2017 and references therein), although the exact placement of the borderline is not consistent across all cases. Schönswetter et al. (2005: Fig. 2) and Thiel-Egenter et al. (2011) report, apart from the "Brenner line", the "Aosta zone" (or the Pennine-Savoy break (in the sense of Merxmüller 1952), which is particularly important for silicicolous alpine plant taxa.

Taxonomic implications

The authors of the account of the genus *Cardaminopsis* (in which *A. halleri* had been placed) in the Flora Europaea (Jones and Akeyrod 1993) and the Atlas Florae Europaeae (Jalas and Suominen 1994) recognized only two subspecies of *A. halleri* in Europe: *A. halleri* subsp. *halleri*, occurring throughout most of the range of the species and *A. halleri* subsp. *ovirensis*, representing alpine morphotypes with pinkish petals, spatially restricted to higher elevations in certain areas of the South-Eastern Alps, Carpathians and Dinaric Alps. Jones and Akeroyd (1993) mentioned that populations from the Tatry mountains (part of the Western Carpathians) are intermediate between these two subspecies, whereas Jalas and Suominen (1994) classified them within *A. halleri* subsp. *ovirensis*. Kolník and Marhold (2006) split *A. halleri* subsp. *ovirensis* into three taxa and classified Western Carpathian populations as *A. halleri* subsp. *tatrica* and high-mountain populations from the Eastern and Southern Carpathians and the Dinaric Alps as *A. halleri* subsp. *dacica*, restricting *A. halleri* subsp. *ovirensis* to the area around Mt Obir in the Eastern Alps. Nevertheless, Šrámková-Fuxová et al. (2017) demonstrated that alpine morphotypes with pinkish petals represent mere ecotypes that have been formed repeatedly in multiple geographical regions and genetic lineages. Neither *A. halleri* subsp. *ovirensis* nor *A. halleri* subsp. *dacica* in the circumscription covering only high-mountain populations are supported by molecular data, as shown by Šrámková-Fuxová et al. (2017) and results presented herein.

Consequently, the results of molecular and morphometric analyses presented here require some adaptation of the above taxonomic treatments of infraspecific variation of *A. halleri*, including somewhat different circumscriptions of the taxa occurring in Europe (for a comparison of the classification proposed here with those by previous authors, see Table 1). We therefore present here a new infraspecific classification of *A. halleri* based primarily on genetic and geographical patterns, which, as our morphometric analyses show, are correlated also with morphology. Such a classification, compared with the previous taxonomic treatments, in our opinion, better reflects the overall structure of the species.

Arabidopsis halleri was described from around the town of Clausthal (currently Clausthal-Zellerfeld), which is located on the Upper Harz Plateau in central Germany (our population AH039). In this respect, the type locality of the nominal subspecies *A. halleri* subsp. *halleri* fits well into the Hercynian lineage recognized in this paper. Contrary to the concepts by Jones and Akeroyd (1993), Jalas and Suominen (1994) and Kolník and Marhold (2006), we propose here to restrict the application of this name to the Hercynian lineage.

The subspecies *A. halleri* subsp. *tatrica* was originally described as *Arabis halleri* var. *tatrica* Pawł. (Pawłowski 1930–1931) from the Polish part of the Tatry Mts between the mountains of Krzesanica and Kopa Kondracka (very close to our population AH113). Again, the type locality of *A. halleri* subsp. *tatrica* fits into the Western Carpathian lineage and we consider it its proper name. The circumscription of the subspecies proposedhere is consistent with the one presented by Kolník and Marhold (2006).

Arabidopsis halleri subsp. *ovirensis* was described from Mt Hochobir in North Karawanken, Carinthia, Austria (our population AH056) as *Arabis ovirensis* by Wulfen in Jacquin (1786; see also Kolník and Marhold 2006; Hohmann et al. 2014). The different circumscriptions of this taxon spanned from considering it a local endemic of Mt. Hochobir (Koch and German 2013), with a suggestion to treat it as a separate species, to the wide concept presented in the Flora Europaea or the Atlas Florae Europaeae (Jones and Akeroyd 1993; Jalas and Suominen 1994). None of these concepts fits the actual genetic grouping, so we propose here to apply this name at the subspecies level to the whole Eastern Alpine lineage. Although the resulting taxon is somewhat morphologically variable, comprising populations of middle elevations, formerly classified as *A. halleri* subsp. *halleri*, as well as alpine morphotypes corresponding to *A. halleri* subsp. *ovirensis* in the narrow sense, it forms a genetically compact group of populations.

Heuffel (1858) described *Arabis ovirensis* var. *dacica* from the alpine belt of the mountains of Banat of Temeshwar (historical province of Habsburg Austria, now divided among Hungary, Serbia and Romania, its higher mountains being now part of Romania). Although the sampling location of the type specimen of this name is not known, it is clear from the detailed description that Heuffel (1858) had in mind the South-Carpathian high-mountain morphotype with lilac to purple petals. It was interpreted in this sense also by Kolník and Marhold (2006). Similarly as in the case of *A. halleri* subsp. *ovirensis*, we propose to apply this name not only to the alpine morphotype, but also to all other populations of the South-Eastern genetic lineage, as it is the oldest name, which should be used according to the rules of nomenclature (Turland et al. 2018).

There is no available name for the fifth European subspecies, covering the Western Alpine populations. Taking into account the genetic and clear morphological differentiation between Eastern and Western Alpine populations, we have decided to describe them here as a new subspecies bearing the name *Arabidopsis halleri* subsp. *occidentalis*.

The proposed subspecies are, indeed, not easy to recognize based on single characters; nevertheless, they can be relatively reliably identified using data on geographic origin or discriminant functions based on morphological characters. The weak morphological differentiation is also the reason why we keep the recognition of this geographic and genetic variation at the level of subspecies.

Taxonomic Treatment

Arabidopsis halleri* subsp. *occidentalis Šrámková & Marhold, **subsp. nova.**—HOLOTYPE: Switzerland, Campocologno, 800 m NW of the village, NW of Tirano, meadow along the road, 46.234°N, 10.131°E, 955 m a.s.l., 15 May 2013, *M. Lučanová s.n.* (PRC 455664; isotype: SAV).

Diagnosis: Arabidopsis halleri subsp. *occidentalis* differs from the closely related *A. halleri* subsp. *halleri* and *A. halleri* subsp. *ovirensis* by a complete absence of trichomes on the leafy part of the stem (vs their presence in most plants of *A. halleri* subsp. *ovirensis* and *A. halleri* subsp. *halleri*), by longer (5.26–6.92 mm) and wider (2.48–3.57 mm) petals (vs 4.7–6.4 × 2.2–3.35 mm *A. halleri* subsp. *halleri*; 3.91–6.14 × 1.84–3.17 mm *A. halleri* subsp. *ovirensis*), wider (1.02–1.47 mm) narrower sepals (vs 0.83–1.39 mm *A. halleri* subsp. *halleri*; 0.71–1.33 mm *A. halleri* subsp. *ovirensis*), smaller (2.5–5) ratio of length and width of the uppermost stem leaf (vs –3.33–10 *A. halleri* subsp. *halleri*; 2.5–6.67 *A. halleri* subsp. *ovirensis*), greater (2.96–4.29) ratio of distance from the base of the stem to the longest stem leaf and length of the leafy part of the stem (vs 1.27–3.65 *A. halleri* subsp. *halleri*; 1.18–3.36 *A. halleri* subsp. *ovirensis*); from *A. halleri* subsp. *halleri* it differs in having fewer (4–11) stem leaves on the main stem (vs 7–15 subsp. *halleri*), and from *A. halleri* subsp. *ovirensis* by shorter (23–45.4 cm) plant height (vs 11–38 cm *A. halleri* subsp. *ovirensis*).

Acknowledgements We are grateful to all colleagues that helped us with the field sampling (see the Appendix) or provided locality data. The research was supported by the Czech Science Foundation (Grant No. 16-10809S) and by the Grant Agency VEGA, Bratislava, Slovakia (Grant No. 2/0137/17). The STRUCTURE computations were performed using resources provided by the Slovak Infrastructure for High Performance Computing (SIVVP, www.sivvp.sk).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Information on electronic supplementary material

Online resource 1. Details on the 88 populations of *Arabidopsis halleri* sampled for the present study.
Online resource 2. Summary of results of STRUCTURE analyses based on AFLP data on *Arabidopsis halleri.*
Online resource 3. Morphometric characters scored in morphologically screened populations of *Arabidopsis halleri*.
Online resource 4. Summary of 19 quantitative characters and five ratios scored in the morphological dataset of five *Arabidopsis halleri* lineages.
Online resource 5. Canonical discriminant analysis based on morphological characters and individual plants of pairs of genetic lineages/subspecies of *Arabidopsis halleri*, correlations of morphological characters with canonical axis (total correlation structure).
Online resource 6. Matrix of AFLP fragments used in analyses of *Arabidopsis halleri*.

Online resource 7. Matrix of microsatellite alleles used in analyses of *Arabidopsis halleri*.
Online resource 8. Matrix of morphometric measurements used in analyses of *Arabidopsis halleri.*
Online resource 9. Discriminant functions enabling the determination of infraspecific taxa with morphological descriptions.

References

Albach DC, Schönswetter P, Tribsch A (2006) Comparative phylogeography of the *Veronica alpina* complex in Europe and North America. Molec Ecol 15:3269–3286. https://doi.org/10.1111/j.1365-294X.2006.02980.x

Alvarez N, Manel S, Schmitt T (2012) Contrasting diffusion of Quaternary gene pools across Europe: the case of the arctic-alpine *Gentiana nivalis* L. (Gentianaceae). Flora Morphol Distrib Funct Ecol Pl 207:408–413. https://doi.org/10.1016/j.flora.2012.03.006

Bhagwat SA, Willis KJ (2008) Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? J Biogeogr 35:464–482. https://doi.org/10.1111/j.1365-2699.2007.01861.x

Clauss MJ, Cobban H, Mitchell-Olds T (2002) Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaeae). Molec Ecol 11:591–601. https://doi.org/10.1046/j.0962-1083.2002.01465.x

Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molec Ecol Notes 3:167–169. https://doi.org/10.1046/j.1471-8286.2003.00351.x

Ehrich D (2006) AFLPDAT: a collection of R functions for convenient handling of AFLP data. Molec Ecol Notes 6:603–604. https://doi.org/10.1111/j.1471-8286.2006.01380.x

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

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molec Ecol Notes 7:574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Gugerli F, Englisch T, Niklfeld H, Tribsch A, Mirek Z, Ronikier M, Zimmermann NE, Holderegger R, Taberlet P (2008) Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation—a project synopsis. Perspect Pl Ecol Evol Syst 10:259–281. https://doi.org/10.1016/j.ppees.2008.07.001

Heufell J (1858) Enumeratio plantarum Banatus Temesiensis sponte crescentium et frequentius cultariim. Verh Zool-Bot Ges Wien 8:39–240

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. Nature 405:907–913. https://doi.org/10.1038/35016000

Hohmann N, Schmickl R, Chiang TY, Lučanová M, Kolář F, Marhold K, Koch MA (2014) Taming the wild: resolving the gene pools of non-model *Arabidopsis* lineages. BMC Evol Biol 14:1–21. https://doi.org/10.1186/s12862-014-0224-x

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

Jasiewicz A (1965) Rośliny naczyniowe Bieszczadów Zachodnich. Monogr Bot 22:1–340

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

Jones BMG, Akeyrod RJ (1993) *Cardaminopsis*. In: Tutin TG, Burges NA, Chater AO, Edmondson JR, Heywood VH, Moore DM, Valentine DH, Walters SM, Webb DA (eds) Flora europaea, 2nd edn. Cambridge University Press, Cambridge, pp 351–352

Juřičková L, Horáčková J, Ložek V (2014) Direct evidence of central European forest refugia during the last glacial period based on mollusc fossils. Quatern Res 82:222–228. https://doi.org/10.1016/j.yqres.2014.01.015

Kerner A (1870) Die natürlichen Floren im Gelände der deutschen Alpen. Fromann, Jena

Klecka WR (1980) Discriminant analysis. Sage Publications, Beverly Hills

Koch MA (2019) The plant model system *Arabidopsis* set in an evolutionary, systematic, and spatio-temporal context. J Exp Bot 70:55–67. https://doi.org/10.1093/jxb/ery340

Koch MA, German DA (2013) Taxonomy and systematics are key to biological information: *arabidopsis*, *Eutrema* (*Thellungiella*), *Noccaea* and *Schrenkiella* (Brassicaceae) as examples. Frontiers Pl Sci 4:1–14. https://doi.org/10.3389/fpls.2013.00267

Koch MA, Wernisch M, Schmickl R (2008) *Arabidopsis thaliana*'s wild relatives: an updated overview on systematics, taxonomy and evolution. Taxon 57:933–943

Kolář F, Fuxová G, Záveská E, Nagano AJ, Hyklová L, Lučanová M, Kudoh H, Marhold K (2016) Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. Molec Ecol 25:3929–3949. https://doi.org/10.1111/mec.13721

Kolník M, Marhold K (2006) Distribution, chromosome numbers and nomenclature conspect of A*rabidopsis halleri* (Brassicaceae) in the Carpathians. Biologia (Bratislava) 61(41–50):486

Krämer U (2010) Metal hyperaccumulation in plants. Annual Rev Pl Biol 61:517–534. https://doi.org/10.1146/annurev-arplant-042809-112156

Krzanowski WJ (1990) Principles of multivariate analysis. Clarendon press, Oxford

Kuss P, Armbruster GFJ, Ægisdóttir HH, Scheepens JF, Stöcklin J (2011) Spatial genetic structure of *Campanula thyrsoides* across the European Alps: indications for glaciation-driven allopatric subspeciation. Perspect Pl Ecol Evol Syst 13:101–110. https://doi.org/10.1016/j.ppees.2011.02.003

Marhold K (2011) Multivariate morphometrics and its application to monography at specific and infraspecific levels. In: Stuessy TF, Lack HW (eds) Monographic plant systematics: fundamental assessment of plant biodiversity. Gantner, Ruggel, pp 73–99

Merxmüller H (1952) Untersuchungen zur Sippengliederung und Arealbildung in den Alpen. Jahrb Vereins Schutze Alpenpfl Alpentiere 17–19:1–105

Mráz P, Ronikier M (2016) Biogeography of the Carpathians: evolutionary and spatial facets of biodiversity. Biol J Linn Soc 119:528–559

Mráz P, Szeląg Z (2004) Chromosome numbers and reproductive systems in selected species of *Hieracium* and *Pilosella* (Asteraceae) from Romania. Ann Bot Fenn 41:405–414

Mráz P, Gaudeul M, Rioux D, Gielly L, Choler P, Taberlet P (2007) Genetic structure of *Hypochaeris uniflora* (Asteraceae) suggests vicariance in the Carpathians and rapid post-glacial colonization of the Alps from an eastern Alpine refugium. J Biogeogr 34:2100–2114. https://doi.org/10.1111/j.1365-2699.2007.01765.x

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Pachschwöll C, García PE, Winkler M, Schneeweiss GM, Schönswetter P (2015) Polyploidisation and geographic differentiation drive diversification in a European high mountain plant group (*Doronicum clusii* aggregate, Asteraceae). PLoS ONE 10:1–30. https://doi.org/10.1371/journal.pone.0118197

Paun O, Schönswetter P, Winkler M, Tribsch A (2008) Historical divergence vs. contemporary gene flow: evolutionary history of the calcicole *Ranunculus alpestris* group (Ranunculaceae) in the European Alps and the Carpathians. Molec Ecol 17:4263–4275. https://doi.org/10.1111/j.1365-294X.2008.03908.x

Pauwels M, Saumitou-Laprade P, Holl AC, Petit D, Bonnin I (2005) Multiple origin of metallicolous populations of the pseudometallophyte *Arabidopsis halleri* (Brassicaceae) in central Europe: the cpDNA testimony. Molec Ecol 14:4403–4414. https://doi.org/10.1111/j.1365-294X.2005.02739.x

Pauwels M, Roosens N, Frérot H, Saumitou-Laprade P (2008a) When population genetics serves genomics: putting adaptation back in a spatial and historical context. Curr Opin Pl Biol 11:129–134. https://doi.org/10.1016/j.pbi.2008.01.005

Pauwels M, Willems G, Roosens N, Frérot H, Saumitou-Laprade P (2008b) Merging methods in molecular and ecological genetics to study the adaptation of plants to anthropogenic metal-polluted sites: implications for phytoremediation. Molec Ecol 17:109–119. https://doi.org/10.1111/j.1365-294X.2007.03486.x

Pauwels M, Vekemans X, Godé C, Frérot H, Castric V, Saumitou-Laprade P (2012) Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, *Arabidopsis halleri* (Brassicaceae). New Phytol 193:916–928. https://doi.org/10.1111/j.1469-8137.2011.04003.x

Pax F (1898) Grundzüge der Pflanzenverbreitung in der Karpathen. 1. Verlag von Wilhelm Engelmann, Leipzig

Preite V, Sailer C, Syllwasschy L, Bray S, Ahmadi H, Krämer U, Yant L (2019) Convergent evolution in *Arabidopsis halleri* and *Arabidopsis arenosa* on calamine metalliferous soils. Philos Trans Ser B 374:20180243. https://doi.org/10.1098/rstb.2018.0243

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

Puşcaş M, Choler P (2012) A biogeographic delineation of the European Alpine System based on a cluster analysis of *Carex curvula*-dominated grasslands. Flora Morphol Distrib Funct Ecol Pl 207:168–178. https://doi.org/10.1016/j.flora.2012.01.002

Ronikier M (2011) Biogeography of high-mountain plants in the Carpathians: an emerging phylogeographical perspective. Taxon 60:373–389

Ronikier M, Cieślak E, Korbecka G (2008) High genetic differentiation in the alpine plant *Campanula alpina* Jacq. (Campanulaceae): evidence for glacial survival in several Carpathian regions and long-term isolation between the Carpathians and the Alps. Molec Ecol 17:1763–1775. https://doi.org/10.1111/j.1365-294X.2008.03664.x

Ronikier M, Schneeweiss GM, Schönswetter P (2012) The extreme disjunction between Beringia and Europe in *Ranunculus glacialis* s. l. (Ranunculaceae) does not coincide with the deepest genetic split—a story of the importance of temperate mountain ranges in arctic-alpine phylogeography. Molec Ecol 21:5561–5578. https://doi.org/10.1111/mec.12030

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385. https://doi.org/10.1126/science.1078311

Schmitt T (2017) Molecular biogeography of the high mountain systems of Europe: an overview. In: Catalan J, Ninot JM, Aniz MM (eds) High mountain conservation in a changing world. Advances in global change research, vol 62. Springer, Berlin, pp 63–74. https://doi.org/10.1007/978-3-319-55982-7_3

Schönswetter P, Tribsch A (2005) Vicariance and dispersal in the alpine perennial *Bupleurum stellatum* L. (Apiaceae). Taxon 54:725–732. https://doi.org/10.2307/25065429

Schönswetter P, Stehlik I, Holderegger R, Tribsch A (2005) Molecular evidence for glacial refugia of mountain plants in the European Alps. Molec Ecol 14:3547–3555. https://doi.org/10.1111/j.1365-294X.2005.02683.x

Šrámková-Fuxová G, Záveská E, Kolář F, Lučanová M, Španiel S, Marhold K (2017) Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction. Bot J Linn Soc 185:321–342. https://doi.org/10.1093/botlinnean/box064

Stein RJ, Höreth S, de Melo JRF, Syllwasschy L, Lee G, Garbin ML, Clemens S, Krämer U (2017) Relationships between soil and leaf mineral composition are element-specific, environment-dependent and geographically structured in the emerging model *Arabidopsis halleri*. New Phytol 213:1274–1286. https://doi.org/10.1111/nph.14219

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. Trends Ecol Evol 16:608–613. https://doi.org/10.1111/mec.13585

Stolpe C, Krämer U, Müller C (2017) Heavy metal (hyper)accumulation in leaves of *Arabidopsis halleri* is accompanied by a reduced performance of herbivores and shifts in leaf glucosinolate and element concentrations. Environm Exp Bot 133:78–86. https://doi.org/10.1016/j.envexpbot.2016.10.003

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. Molec Ecol 7:453–464. https://doi.org/10.1046/j.1365-294x.1998.00289.x

Těšitel J, Malinová T, Štech M, Herbstová M (2009) Variation in the *Melampyrum sylvaticum* group in the Carpathian and Hercynian region: two lineages with different evolutionary histories. Preslia 81:1–22

Thiel-Egenter C, Alvarez N, Holderegger R, Tribsch A, Englisch T, Wohlgemuth T, Colli L, Gaudeul M, Gielly L, Jogan N, Linder HP, Negrini R, Niklfeld H, Pellecchia M, Rioux D, Schönswetter P, Taberlet P, van Loo M, Winkler M, Gugerli F (2011) Break zones in the distributions of alleles and species in alpine plants. J Biogeogr 38:772–782. https://doi.org/10.1111/j.1365-2699.2010.02441.x

Turland NJ, Wiersema JH, Barrie FR, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Kusber W-H, Li D-Z, Marhold K, May TW, McNeill J, Monro AM, Prado J, Price MJ, Smith GF (eds) (2018) International code of nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the nineteenth international botanical congress Shenzhen, China, July 2017, Regnum Vegetabile 159. Koeltz Botanical Books, Glashütten. Available at: https://doi.org/10.12705/Code.2018

Van Rossum F, Bonnin I, Fénart S, Pauwels M, Petit D, Saumitou-Laprade P (2004) Spatial genetic structure within a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy-metal-tolerant species. Molec Ecol 13:2959–2967. https://doi.org/10.1111/j.1365-294X.2004.02314.x

Verbruggen N, Hermans C, Schat H (2009) Molecular mechanisms of heavy metal hyperaccumulation in plants. Phytoremediat Environm Pollut 181:759–776. https://doi.org/10.4324/9781315161549

Vos P, Hogers R, Bleeker M, Reijans M, van De Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. Nucl Acids Res 23:4407–4414. https://doi.org/10.1093/nar/23.21.4407

Wasowicz P, Pauwels M, Pasierbinski A, Przedpelska-Wasowicz EM, Babst-Kostecka AA, Saumitou-Laprade P, Rostanski A (2016) Phylogeography of *Arabidopsis halleri* (Brassicaceae) in mountain regions of Central Europe inferred from cpDNA variation and ecological niche modelling. PeerJ 4:e1645. https://doi.org/10.7287/peerj.preprints.1259v1

Willems G, Dräger DB, Courbot M, Godé C, Verbruggen N, Saumitou-Laprade P (2007) The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): an analysis of quantitative trait loci. Genetics 176:659–674. https://doi.org/10.1534/genetics.106.064485

Woloszczak E (1896) Z granicy flory zachodnio- i wschodnio-karpackiej. Spraw Komis Fizjogr Akad Umiejetn 31:119–159

Plant Systematics and Evolution (2020) 306:60
https://doi.org/10.1007/s00606-020-01686-4

CORRECTION

Correction to: Phylogeography and taxonomic reassessment of *Arabidopsis halleri* – a montane species from Central Europe

Gabriela Šrámková[1] · Filip Kolář[1,2,3] · Eliška Záveská[2,3] · Magdalena Lučanová[1,3,5] · Stanislav Španiel[1,4] · Martin Kolník[4] · Karol Marhold[1,4]

Published online: 27 May 2020

Correction to:
Plant Systematics and Evolution (2019) 305:885–898
https://doi.org/10.1007/s00606-019-01625-y

List of errata:

In text:

Taxonomic treatment:

Diagnosis: *Arabidopsis halleri* subsp. *occidentalis* differs from the closely related *A. halleri* subsp. *halleri* and *A. halleri* subsp. *ovirensis* by a complete presence of trichomes on the leafy part of the stem (vs their potential absence in plants of *A. halleri* subsp. *ovirensis* and *A. halleri* subsp. *halleri*), by longer (5.26–6.92 mm) and wider (2.48–3.57 mm) petals (vs 4.7–6.4 × 2.2–3.35 mm *A. halleri* subsp. *halleri*; 3.91–6.14 × 1.84–3.17 mm *A. halleri* subsp. *ovirensis*), wider (1.02–1.47 mm) narrower sepals (vs 0.83–1.39 mm *A. halleri* subsp. *halleri*; 0.71–1.33 mm *A. halleri* subsp. *ovirensis*), smaller (2.5–5) ratio of length and width of the uppermost stem leaf (vs 3.33–10 *A. halleri* subsp. *halleri*; 2.5–6.67 *A. halleri* subsp. *ovirensis*), greater (2.96–4.29) ratio of distance from the base of the stem to the longest stem leaf and length of the leafy part of the stem (vs 1.27–3.65 *A. halleri* subsp. *halleri*; 1.18–3.36 *A. halleri* subsp. *ovirensis*); from *A. halleri* subsp. *halleri* it differs in having fewer (4–11) stem leaves on the main stem (vs 7–15 subsp. *halleri*), and from *A. halleri* subsp. *ovirensis* by longer (23–45.4 cm) plant height (vs 11–38 cm *A. halleri* subsp. *ovirensis*) (Fig. 3).

The original article can be found online at https://doi.org/10.1007/s00606-019-01625-y.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00606-020-01686-4) contains supplementary material, which is available to authorized users.

✉ Karol Marhold
karol.marhold@savba.sk

[1] Department of Botany, Faculty of Science, Charles University, Benátská 2, 128 01 Prague, Czechia

[2] Department of Botany, University of Innsbruck, Sternwartestraße 15, 6020 Innsbruck, Austria

[3] Institute of Botany, The Czech Academy of Sciences, 252 43 Průhonice, Czechia

[4] Plant Science and Biodiversity Centre, Institute of Botany, Slovak Academy of Sciences, Dúbravská cesta 9, 845 23 Bratislava, Slovakia

[5] Department of Botany, Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czechia

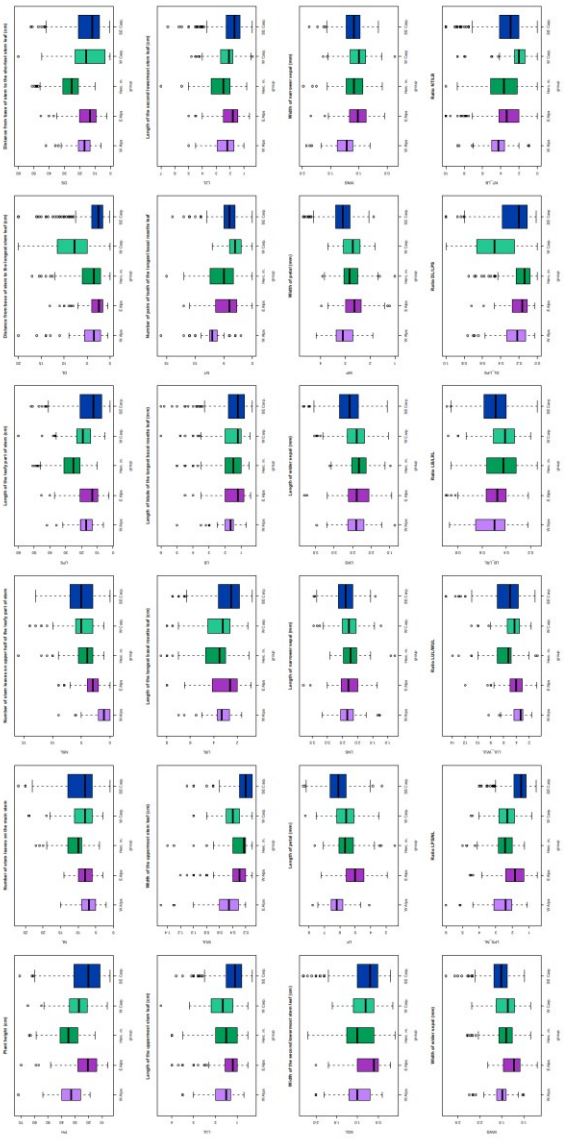


Fig. 3 Range of values of 24 continuous morphological characters (19 quantitative and 5 ratios) for five genetic lineages of European *Arabidopsis halleri*. The box represents values between the first and third quartiles of the data, the thick horizontal line represents the second quartile (median), whiskers show values greater/smaller than 1.5× the interquartile range from the given quartile, and points represent outliers

Paper V

Padilla-García N, **Šrámková G**, Kolář F, Šlenker M, Zeisek V, Záveská E, Clo J, Lučanová M, Rurane I, Marhold K (manuscript prepared for submisison) Niche differentiation following whole-genome duplication? The importance of considering the evolutionary history of genetic lineages when assessing climatic niche evolution.

Niche differentiation following whole-genome duplication? The importance of considering the evolutionary history of genetic lineages when assessing climatic niche evolution

Nélida Padilla-García[1]*, Gabriela Šrámková[1]*, Marek Šlenker[3], Eliška Záveská[2], Vojtěch Zeisek[1,2], Magdalena Lučanová[4], Ieva Rurane[5], Filip Kolář[1,2], Karol Marhold[1,3]
[1]Department of Botany, Faculty of Science, Charles University, Prague CZ-128 01, Czech Republic
[2]Institute of Botany, The Czech Academy of Sciences, Pruhonice CZ-252 43, Czech Republic
[3]Plant Science and Biodiversity Centre, Institute of Botany, Slovak Academy of Sciences, Bratislava SK-845 23, Slovak Republic
[4]Department of Botany, Faculty of Science, University of South Bohemia, České Budějovice CZ-370 05, Czech Republic
[5]Institute of Biology, University of Latvia, Miera 3, Salaspils LV-20169, Latvia
* These authors contributed equally to the study

ABSTRACT

Aim
Although whole genome duplication (WGD, polyploidization) is an important speciation force, we still lack a consensus on the role of niche differentiation in polyploid evolution. In addition, the role of genome doubling per se vs. later divergence on polyploid's niche evolution remains obscure. One reason for this might be that the often intricate intraspecific genetic structure of polyploid complexes and interploidy gene flow is frequently neglected in ecological studies. Here, we aim to investigate to which extent these evolutionary impacts our inference on niche differentiation of autopolyploids.
Location: Europe
Taxon: *Arabidopsis arenosa* (Brassicaceae)
Methods
Leveraging a total of 352 jointly genotyped and cytotyped populations of diploid-autotetraploid *A. arenosa*, we examined differences among climatic niches of diploid and tetraploid lineages globally and independently for each tetraploid lineage with respect to the niche of its evolutionary closest diploid and/or tetraploid relative. Then, we tested if there was an effect of additional interploidy introgression from other sympatric yet non-ancestral diploid lineages of *A. arenosa* on climatic niches of tetraploids by integrating the niche data of the respective diploid lineages to the ecological models.
Results
The ecological niche shift of tetraploids is detected only when each tetraploid lineage is compared with the evolutionary closest diploid ancestor lineage, not when examined globally. Different patterns of climatic niche evolution (i.e., niche conservatism, contraction or expansion) are found in each tetraploid lineage. Climatic niches of tetraploid lineages are significantly different among them. We observe an effect of interploidy gene flow in patterns of climatic niche evolution of tetraploid ruderal plants of *A. arenosa*.
Main conclusions
The niche shift of tetraploids in *A. arenosa* is not driven by WGD per se but rather reflects dynamic post-WGD evolution in the species, involving tetraploid migration out of their ancestral area and interploidy introgression with other diploid lineages. Our study supports that evolutionary processes following WGD, which usually remain undetected by studies neglecting intraspecific genetic structure, may play a key role in the adaptation of polyploids to challenging environments.

KEYWORDS
Arabidopsis arenosa, climatic niche evolution, ecological niche comparison, genetic structure, polyploidy

INTRODUCTION

Polyploidy is a leading evolutionary force driving speciation and diversification of all plant lineages (Stebbins, 1950; Otto and Whitton, 2000; Wendel, 2000; Soltis et al., 2015). All flowering plants have experienced one or more episodes of whole-genome duplication (WGD) during their evolutionary history (Jiao et al., 2011; Wendel, 2015) and it has been estimated that up to 15% of all speciation events in angiosperms are associated with a ploidy increase (Wood et al., 2009). Polyploid speciation has been considered a mechanism of sympatric speciation given the potential reproductive isolation that newly formed polyploids can experience due to strong postzygotic barriers among cytotypes (e.g. differences in the number of sets of chromosomes; Coyne and Orr, 2004). However, incomplete postzygotic isolation among cytotypes is common in plants, thus promoting interploidy reproduction (Sutherland and Galloway, 2017). Accordingly, polyploid speciation is not an "instantaneous" process and prezygotic reproductive barriers such as differences in ecological niches between polyploids and closest lower-ploidy progenitors can also play a key role in assortative mating and thus, polyploid speciation (Levin, 2004; Husband et al., 2016). According to the niche shift hypothesis, novel phenotypic, physiological and genetic combinations associated with polyploidy potentially could drive up polyploids to expand to new ecological niches that would have remained unavailable to their diploid progenitors (Madlung, 2013). According to this, niche ecological divergence with respect to progenitors may be more prone to occur in allopolyploids (i.e. formed by genome duplication after hybridization of two different parental genomes more or less divergent) than autopolyploids (i.e., formed within the same species or genetic lineage) because high heterozygosity could allow allopolyploids to colonize new habitats. Many empirical studies have addressed niche evolution after WGD, with some of them supporting niche divergence between diploids and polyploids (Theodoridis et al., 2013, Thompson et al., 2014, Visger et al., 2016, Muñoz-Pajares et al., 2018, Decanter et al., 2020), while others do not (Godsoe et al., 2013, Glennon et al., 2014, Čertner et al. 2015, Visser and Molofsky, 2015, Castro et al., 2019; 2020a). Yet, there is not a consistent or universal pattern concerning the role of niche differentiation in polyploid establishment and evolution. This incongruence between studies has been explained by methodological issues as an inappropriate resolution of environmental variables (Kirchheimer et al., 2016), because niche differentiation might be occurring at a different spatial scale (Čertner et al., 2019) and/or simply by different evolutionary histories of polyploid cytotype between studied species – an aspect that has been, however, often neglected in purely ecological studies. The presence or absence of niche evolution in polyploids strongly depends on species' history: e.g. polyploid's age, multiple origins or the number of ploidy levels, among others (see Duchoslav et al., 2020). There can also be a phylogenetic signal in environmental traits (Burns and Strauss, 2011), thus, ecological niches of polyploids can be potentially affected by ancestral niches of progenitors. Consequently, patterns of ecological differentiation could be misunderstood if niches of the polyploids are not compared to their closest lower-ploidy progenitors. Unfortunately, the intraspecific genetic structure of polyploid complexes is often unknown due to limitations associated with challenges surrounding population genetic data analysis of polyploids (Rothfels, 2021), and if known, it is rarely taken into account when evaluating polyploid niche evolution (see López-Jurado et al., 2019 as the exception). Niches of allopolyploids are predicted by the sum of the niche of their progenitors (Parisod and Broennimman, 2016). This is not the case with strict autopolyploids, which usually conserve the climatic niche of their unique progenitor species. Nevertheless, hybridization and interploidy admixture between geographically close diploid and tetraploid populations are common in many plant species (Aagaard et al., 2005; Koutecký et al., 2011; Monnahan et al., 2019; Šmíd et al., 2020). Thus, it might also occur that autopolyploids come into secondary contact with distance-related diploid ancestors that have diverged before the WGD event. This genetic exchange may result in adaptive introgression that can also influence patterns of climatic niche evolution (Schmickl and Yant, 2021). Nevertheless, few studies have assessed

whether introgression promotes ecological niche evolution of polyploids, most of them being focused on strict allopolyploids (Arrigo et al., 2016, Blaine Marchant et al., 2016). The effect of interploidy introgression in the ecological niche of autopolyploids remained unexplored.

Arabidopsis arenosa (Brassicaceae) is a diploid-autotetraploid species with a well-described evolutionary history of its lineages, that has recently become an interesting system not only to study polyploid evolution (Monnahan et al., 2019; Morgan et al., 2021a; 2021b; Bohutínská et al., 2021a)but also adaptation to extreme conditions (Konečná et al., 2021; Bohutínská et al., 2021b, Knotek et al., 2020, Wos et al., 2021). Multiple studies have addressed the role of niche differentiation in the autopolyploid evolution of this system, however, reaching strikingly inconsistent outcomes. Ecological niche modelling was used in a study conducted by Molina-Henao and Hopkins (2018), which concluded niche expansion but not the divergence of tetraploids of *A. arenosa*. These results contrast with other studies in which an absence of ecological niche differentiation was described at both the landscape (Kolář et al., 2016b) and intra-population scales (Wos et al., 2019), and in three contact zones independently (Morgan et al., 2020). Unfortunately, neither of these studies did cover the whole distribution range of the species (Wos et al., 2019; Morgan et al., 2020) and they had not integrated the intricate evolutionary history of this species involving the intraspecific genetic sub-structure of each cytotype and interploidy gene flow (Kolář et al., 2016b, Molina-Henao and Hopkins 2018). Autotetraploid cytotype of *A.arenosa* originated only once in the Western Carpathians 20,000 to 31,000 generations ago, where it still coexists with its diploid progenitor until now, but also from where it spread through most of Europe from Romania in the south to Belgium in the west and Scandinavia in the north (Arnold et al., 2015; Monnahan et al., 2019). During this expansion, tetraploids also encountered other earlier diverged diploid lineages of *A. arenosa* and got introgressed by them in at least two contact zones, in SE Carpathians and the Baltic coast (Monnahan et al. 2019). An intriguing question about to which extent such intricate evolutionary history has been translated to niche divergence, however, remained unanswered.

In the present study, we aim to examine the impact of the intraspecific genetic structure and interploidy introgression when testing for autopolyploid niche evolution. More specifically, we test the following hypothesis: (i) whether niche shift of polyploids is detectable only when polyploid lineages are compared with their corresponding diploid ancestor(s), not globally; and (ii) whether interploidy introgression events promote ecological divergence and/or expansion of polyploids due to additional genetic material from other source lineages. First, we used genome-wide single nucleotide polymorphism (SNP) genotyping to investigate the genetic structure within *A. arenosa* tetraploids across the landscape including for the first time samples from the whole geographic distribution area of the tetraploid cytotype. Second, we compared the environmental niche occupied by diploid and tetraploid lineages of *A. arenosa*, both globally, and independently for each tetraploid lineage with respect to the niche of the evolutionary closest diploid lineage, and among them. Third, we investigated the effect of interploidy gene flow in polyploid niche shift by comparing niches of tetraploid Ruderal and SE Carpathian lineages to the niches of its respective sympatric locally-adapted diploid-related lineages.

MATERIAL AND METHODS

Genomic data

Dataset and library preparation

The cytogeographical patterns in *A. arenosa* have been well documented (Kolář et al. 2016b) and the genetic structure of diploid populations across their entire distributional range is well known (Kolář et al. 2016a). However, in the case of tetraploids, the identification of genetic lineages remains unclear mainly because the whole distribution range of tetraploids of *A. arenosa* has not been fully covered in previous studies (Arnold et al. 2015), or the number of sampled populations was limited (Monnahan et al. 2019). In order to identify range-wide tetraploid genetic structure, a dataset of genome-wide single-

nucleotide polymorphisms (SNPs) was generated for a total of 275 tetraploid individuals from 125 populations (1-4 indiv/pop) using double-digest RADseq (as described in Wos et al. 2019). Available published WGS data (Monnahan et al. 2019) from 62 tetraploid populations were also integrated to generate a more robust dataset, consisting of a total of 428 individuals from 186 populations.

Raw data processing, variant calling and filtration

Raw reads were demultiplexed using FASTX toolkit 0.0.14 and quality trimmed (> 20 Phred quality score) in Trimmomatic 0.36. Mapping on *Arabidopsis lyrata* reference genome v. 1.0.25 (Hu et al. 2011) was performed using BWA v. 0.7.3a and the resulting BAM was processed with Picard Tools v. 2.22.1. The Genome Analysis Toolkit v. 3.8 (GATK, McKenna et al. 2010) was used following the best practice recommendations (www.boradinstitute.org/gatk). Variant calling was performed for each individual using the HaplotypeCaller module, setting the ploidy=4 option. Then, we aggregated variants and performed genotyping across all individuals using GenotypeGVCFs.

We used the coordinates of the identified RAD loci to retrieve the corresponding SNPs from a previous set of genome resequencing data (Monnahan et al. 2019) that was mapped to the same reference genome. In this way, we extracted the SNP data from the same sites for an additional 155 individuals using GATK. Both vcf files were merged and the final vcf contained 430 tetraploid individuals from a total of 187 populations. We only considered biallelic sites that passed the filter parameters indicated by GATK best practices (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants): 'QD < 2.0', 'FS > 60.0', 'MQ < 40.0', 'MQRankSum < -12.5', 'ReadPosRankSum < -8.0', 'SOR > 3.0'. Additionally, for subsequent analyses, we only keep variants that were present in at least 80% of individuals at a minimum sequencing depth of 8x, reaching a final dataset of 179,698 SNPs. Scripts used for processing the data are available at https://github.com/V-Z/RAD-Seq-scripts.

Identification of genetic lineages

We inferred population genetic structure using Bayesian clustering analyses in STRUCTURE v. 2.3.2 (Pritchard et al. 2000), which allows us to take into consideration autotetraploid genotypes. Before running STRUCTURE, we pruned the dataset to avoid linkage among SNPs. Taking into account the average length of RADseq fragments (350 bp), we randomly selected one SNP per each 1000-bp window to avoid linkage disequilibrium. We discarded those SNPs showing a lower minor allele frequency of 0.05 and a higher minor frequency of 0.95 to remove uninformative singletons and errors in the dataset. Python scripts used for pruning and formatting the input data for STRUCTURE are available at https://github.com/MarekSlenker/vcf_prune. In STRUCTURE, we run ten replicates per each value of *K* between 1 and 10 applying a burn-in of 10^4 iterations followed by 10^5 MCMC iterations. Convergence among different replicates per each *K* value was evaluated in R using a script that ran modified functions previously coded by Ehrich (2006) (https://github.com/MarekSlenker/structureSum). The results for every value of *K* were visualized using CLUMPAK (Kopelman et al. 2015). The optimal value of groups in our dataset was identified according to several criteria (i.e. $K = 5$ is the highest value of *K* showing a positive delta *K* value (Evanno et al. 2005) and convergence of the results among the ten replicates (Supplementary Figure 1). Each population was assigned to the cluster for which the highest proportion of membership was observed. We further investigated the population structure inferenced by STRUCTURE running principal component analysis (PCA). Based on putatively neutral four-fold degenerate (4dg) SNPs and using ADEGENET package in R (Jombart 2008) we summarized the neutral genetic variability among the identified lineages within our tetraploid samples. Additionally, we calculated pairwise Nei's genetic distances (Nei 1987) among the same lineages to quantify the genetic differentiation between them using StAMPP R package (Pembleton et al. 2013).

Niche comparison analyses

Occurrence data and climatic variables

For the climatic niche comparison analyses, we used both diploid and tetraploid occurrences of *A. arenosa* with known affiliations to the genetic lineages. We collected leaf material of 10-20 individuals from diploid and tetraploid populations of *A. arenosa* in 2011-2020 across their entire European distribution range (Kolář et al. 2016b) and we checked the ploidy level of each individual using flow cytometry (as described in Kolář et al. 2016b, Supplementary Table 1). Geographical coordinates were obtained from GPS during field surveys. The lineage assignment to tetraploid populations was done according to the Bayesian clustering of the SNP data obtained in this study (see details above). To avoid potential bias caused by admixture between different tetraploid lineages, we excluded equivocal populations showing less than 50% membership to one single cluster from the niche comparison analyses (in total 21 populations, see Supplementary Table 2). In the case of diploids, the assignment was based on previous Bayesian clustering of a set of populations covering the entire range of the diploid cytotype (Kolář et al. 2016) and the geographical location of additionally sampled populations. A total of 352 localities including 2*x* and 4*x* cytotypes of *A. arenosa* assigned to intraspecific genetic lineages were analyzed (see Supplementary Table 1). In order to avoid unequal representation caused by biased sampling in different areas, the obtained sampling points were filtered and those that were closer than 10-km distance were removed. The number of occurrences per cytotype and lineage is indicated in Table S3. Environmental data related to temperature (BIO1-BIO11 variables) and precipitation (BIO12-BIO19 variables) were extracted for all occurrence points from WorldClim at 30-s (ca. 1 km) resolution (Hijmans et al., 2005).

Niche quantification and comparison

Quantification and comparison of climatic niches were performed using a statistical framework developed by Broennimann et al. (2012). It applies a kernel density function to calculate the smoothed density of occurrences and environmental values along the first two axes of multivariate analysis (PCA-env). This method ensures that the niche overlap is independent of the resolution of the grid. We considered the first two axes of the PCA calibrated on the environmental space of the study area, which was divided into a grid of 100 x 100 cells with each cell corresponding to a unique set of environmental conditions. The environmental space was produced by extracting the same climatic values for 10,000 occurrence points randomly sampled from 100-km buffer zones around the occurrences of diploid and tetraploid *A. arenosa* records. This common environmental space, which theoretically corresponds to the potential habitat of the species, was used for all pairs of comparisons. Occurrence density grids had a resolution of 100 and a species density threshold of zero. Niche overlap, equivalence and similarity tests implemented in the R package "ecospat" (Di Cola et al., 2017) were performed to compare the divergence between lineages. Niche overlap calculation is based on Schoener's *D* metric (Schoener, 1968) that ranges from 0 (no overlap) to 1 (complete overlap). To evaluate the significance of niche overlap ($\alpha = 0.05$), we performed an equivalency test (Warren et al. 2008), which uses random bootstrap resampling of presence occurrence points of both lineages to calculate if a null distribution of *D* and the observed *D* are significantly different ($p < 0.05$; niches are not statistically equivalent) or not (niches are equivalent). The significance of niche overlap was also evaluated by similarity tests, which use bootstrap resampling to assess whether the niche of one lineage predicts the other better than would be expected by chance ($\alpha = 0.05$). If observed *D* is greater than the null distribution, niches are more similar than expected. Values lower than the null distribution indicate that niches are not similar and non-significant values mean a lack of power of the test to detect differences or similarities. However, simply testing if niches of diploid and tetraploid cytotypes are equivalent or different does neither fully account for dynamics of niche evolution, nor reflects the complex reticulated evolutionary history of the species. In order to understand alternative processes driving niche evolution in *A. arenosa* we have quantified niche dynamics per each genetic lineage, using niche unfilling (U), stability (S) and expansion (E) indices (Guisan et al., 2014) and calculating niche optimum and breadth along the axes

of temperature and precipitation variation. The niche optimum and breadth of each lineage was calculated following the procedure described in Theodoridis et al. (2013) and Kirchheimer et al. (2016). We randomly sampled 100 cells of the gridded space of each lineage with the probability of selection weighted by the density of the species occurrences. We calculated the niche optimum and breadth, calculating the median and the standard deviation of the scores along the two PCA axes, respectively. This re-sampling was repeated 1000 times and differences in the distribution of optimum and breadth values were compared using Welch's t-tests in R. The results were visualized using boxplots for each PCA axis. To test for the effect of the studied lineages on the niches' optimum and breadth for each PCA axis, we first performed an ANOVA test, and then we performed a Tukey HSD test to perform all the possible pairs-comparisons.

RESULTS

Genetic structure within *A. arenosa* tetraploids

We obtained a total of 179,698 filtered SNPs for the 428 tetraploid individuals of *A.arenosa* included in our study. The average depth is 11× and our dataset comprises 13.1% of missing data over populations and sites. Bayesian clustering analyses support the existence of five distinct genetic clusters among tetraploids of *A. arenosa*. These clusters are geographically separated (Fig. 1). Two of them (red and blue) are restricted to populations located in Western and Southeastern Carpathians, respectively. The majority of populations from the Western part of Central Europe (further referred to as C Europe) are included in a third cluster (light orange), only populations from the Eastern Alps form a separate, fourth, cluster (yellow). Finally, some populations sampled in human-made ruderal stands (mainly railways tracks and roadsides) in the Alps, Germany and the Czech Republic are clustered together with all populations located in Northern Europe and the Baltic Sea coast (purple).

Principal component analysis based on genetic data confirms the genetic differentiation of *A. arenosa* tetraploids into the five lineages and further indicates the level of differentiation among them. Ruderal populations (4x-RUD) are separated from the others along the first axis, while populations located in SE Carpathians (4x-SEC) are differentiated from the rest along the second axis (Supplementary Figure 2A). Alpine (4x-ALP), W Carpathians (4x-WCA) and C European (4x-CEU) populations cluster together in the PCA of the complete dataset but get clearly distinct in a separate analysis excluding the populations identified within the 4x-RUD and 4x-SEC lineages (Supplementary Figure 2B). Nei's genetic distances calculated among lineages reflect the fact that the genetic divergence among them is generally low (Supplementary Figure 2C).

Niche quantification and comparison among diploids and tetraploids globally

The variation explained by the two first axes of the PCA-env is 41.5% and 30.4% respectively, which means a total of 71.9% of the variance. Environmental variables related to precipitation (BIO12-BIO19) are highly correlated to PC1, while temperature-related variables (BIO1-BIO11) are mainly correlated to PC2 (Fig. 2A). The contribution of each variable to the two first axes of the PCA is summarized in Supplementary Table 4. Annual precipitation (BIO12), precipitation of the driest quarter and month (BIO17 and BIO14) together with precipitation of the coldest quarter (BIO19) show the highest percentage of correlation to PC1. The mean temperature of the coldest quarter (BIO11) and minimum temperature of the coldest month (BIO6) are the most correlated variables to PC2.

When the niche of diploids and tetraploids of *A. arenosa* is compared globally, i.e. ignoring the assignment to intraspecific genetic lineages, the niche equivalency test does not have enough power to detect niche differentiation, while the similarity test indicates that niches are more similar than expected by chance independent of the direction of the test (Table 1). These results indicate niche conservatism between diploids and tetraploids when they are globally compared. Accordingly, tetraploids show a relative high niche overlap with diploids (50.2 %), very high stability index value (S = 0.947) and almost zero expansion and niche unfilling (E = 0.053 and U = 0.079, respectively).

Niche quantification and comparison between related diploid and tetraploid intraspecific lineages

To integrate niche data with evolutionary history, we firstly reconstructed niche evolution reflecting the single evolutionary origin of all *A. arenosa* tetraploid lineages from the ancestral diploid lineage (2x-WCA). Niche conservatism is observed when comparing all *A. arenosa* tetraploid lineages to the 2x-WCA lineage, but excluding other diploid lineages from the analysis that did not contribute to the origin of autotetraploids (Fig. 2, Table 1). In this case, the proportion of niche overlap of tetraploids with the 2x-WCA diploid lineage is 62.8 %. The stability index value is high (S = 0.748) while the expansion and unfilling values are E = 0.252 and U = 0.006. Furthermore, when ancestral 2x-WCA lineage was compared independently to each tetraploid lineage, the resulting patterns differed in each comparison (Fig. 2). Specifically in the W Carpathians, niches of diploids and tetraploids are more similar than expected by chance, independent of the direction of the test (Table 1c). Furthermore, a high niche overlap between 2x and 4x W Carpathians lineages (64.9 %) with high stability and very low values for expansion and niche unfilling indices (S = 0.983, E = 0.017 and U = 0.127) is found. These results together suggest that in the W Carpathians, diploid and tetraploid lineages show similar ecological niches. In contrast, similarity tests were non-significant when comparing other tetraploid lineages to the ancestral diploid (2x-WCA). A high overlap was found between niches corresponding to 4x-ALP and 2x-WCA (52.1%), also showing a high stability value (S = 0.785) and intermediate values of niche expansion (E = 0.215) and unfilling (U = 0.350). In the case of 4x-SEC lineage, niche contraction compared to the 2x-WCA lineage is indicated by an intermediate value of the unfilling index (U = 0.400) and a high stability index (S = 0.903) but almost zero expansion (E = 0.097). In this case, the percentage of niche overlap is 24.6 % (Table 1). The 4x-RUD lineage has experienced niche expansion compared to the 2x-WCA lineage (S = 0.664 and E = 0.336), whose niche was partially filled by the Ruderal one (U = 0.207). These results indicate that niche expansion of 4x-RUD has occurred towards areas of more extreme temperatures while it has not filled the "high-alpine" niche portion occupied by its 2x-WCA ancestor lineage, which is characterized by high precipitation values. The niche occupied by 4x-CEU lineage has also significantly expanded (S = 0.521 and E = 0.479) but in this case, tetraploids almost did not fill the niche occupied by 2x-WCA lineage (U = 0.607).

Niche optimum (calculated as the median of the scores along each axis of PCA-env) reflects a higher optimal precipitation value (PC1) for 2x-WCA when compared to all tetraploid lineages, except for4x-ALP (Fig. 3). Optimal temperature values (PC2) are higher for tetraploids than 2x-WCA except for 4x-SEC. Niche breadth in terms of precipitation (PC1) is lower for all tetraploid lineages compared to 2x-WCA lineage (Fig. 3). In terms of temperature (PC2), niche breadth is not significantly different among 2x-WCA, 4x-WCA and 4x-CEU lineages. Temperature breadth is higher for 2x-WCA than for 4x-ALP and 4x-SEC lineages but lower than for the 4x-RUD one.

Niche quantification and comparison among tetraploid lineages

Furthermore, we inferred niche evolution during tetraploid expansion. To do so, we compared the niche of each tetraploid lineage to the niche of tetraploids from the W Carpathians (4x-WCA) which occupy the presumed area of origin of tetraploid *A. arenosa* cytotype and are thus closest to the "ancestral polyploid" niche. Similarity tests were not significant for any pair-wise comparison (Table 1). Percentages of niche overlap varied from 14.4 % (4x-CEU) to 46.76 % (4x-ALP). All tetraploid lineages showed some degree of niche expansion relative to the 4x-WCA lineage occupying the ancestral area (Fig. 3), with values of E indices ranging between 0.258 (4x-SEC) and 0.447 (4x-CEU). The 4x-RUD lineage almost entirely filled the niche of 4x-WCA (U = 0.098) while other lineages show values of unfilling indices that vary between 0.347 (4x-ALP) and 0.438 (4x-CEU). The optimal value of precipitation (PC1) is shown to be highest for 4x-ALP and lowest for 4x-RUD and 4x-CEU lineages, while in terms of temperature (PC2) is the 4x-SEC which showed the coldest optimal value. Specifically, 4x-RUD and 4-CEU lineages

show the highest optimal values of temperature (Fig. 3). Niche breadth analyses indicate that 4x-WCA occupies the broadest gradient of precipitation compared to all other tetraploid lineages, while the temperature breadth is much higher for the 4x-RUD lineage than from other tetraploid lineages (Fig. 3).

Niche quantification and comparison between tetraploid and locally sympatric diploid lineages

Last, we investigated niche evolution when taking into account the influence of strong inter-ploidy introgression from more divergent diploid lineages of *A. arenosa* which co-occur with particular tetraploid lineages within two contact zones: Southern Carpathians (4x-SEC lineage introgressed by 2x-SEC) and Baltic coast (4x-RUD lineage introgressed by 2x-BAL). When we compared the observed niche of 4x-SEC lineage to the combined niche of both diploid lineages that served as source gene pools (2x-WCA and 2x-SEC), we do not observe substantive changes in niche dynamics index values concerning the previous comparison including 4x-SEC lineage and only the ancestral diploid 2x-WCA lineage (Fig. 2, Fig. 4). It is shown that tetraploids have barely expanded their niche (S = 0.923 and E = 0.077) and that a considerable proportion of the niche of diploid lineages has not been filled by 4x-SEC (U = 0.417). Thus, niche contraction with respect to both diploid lineages has occurred in 4x-SEC lineage towards warmer and more humid areas.

Comparing the observed niche of 4x-RUD to those predicted by the combination of niches of both diploid lineages that contributed to its genetic make-up (2x-WCA and 2x-BAL), we observe differences concerning the previous comparison including just the ancestral diploid 2x-WCA lineage (Fig. 2). We observe a lower niche expansion of 4x-RUD lineage for both diploid lineages (E = 0.270). A lower unfilling index value (U = 0.133) and a higher stability (S = 0.730) are also shown (Fig. 4).

DISCUSSION

1. Varying extent of niche differentiation and expansion across tetraploid intraspecific lineages.

It is usually assumed that polyploids have a different ecological niche than diploids, reflecting a WGD-driven shift in important functional traits, which in turn strengthens prezygotic isolation between ploidies and contributes to polyploid speciation. However, the existence and potentially the extent of ploidy-related niche differentiation differ between individual species and sometimes among case studies (Glennon et al., 2014), even within the same species. One explanation for this controversy is that niche evolution of polyploids might be detectable only when polyploid lineages are compared with their corresponding diploid ancestors but, unfortunately, the intraspecific genetic structure of polyploid complexes is often ignored when testing climatic niche evolution of polyploids. To test for this effect, we compared niches of ploidy cytotypes of *A. arenosa* - a species surrounded with controversy regarding the level of niche differentiation of its diploid and autotetraploid cytotype (Molina Henao and Hopkins, 2018; Morgan et al., 2020) - using the largest population sampling to date coupled with genotyping of each population. We assigned each autotetraploid population to one of the five major genetic lineages which corresponded with previous genetic structuring results based on smaller sampling (Arnold et al., 2015; Monnahan et al., 2019) and put our results in the context of polyploid origin and interploidy gene flow by integrating with previous population genomic investigations involving both ploidies (Monnahan et al. 2019).

In our study, niche shift was not supported when we compared diploid and tetraploid cytotypes globally, i.e. ignoring the intraspecific genetic structure (Fig. 2, Table 1). However, we identified highly variable patterns of niche differentiation when the climatic niche of each tetraploid lineage was compared independently to the niche of the ancestral diploid lineage (2x-WCA; Arnold et al., 2015; Monnahan et al., 2019). Niche conservatism between diploids and tetraploids was only found within the zone of primary coexistence (W Carpathians), where tetraploids of *A. arenosa* originated. This result

corresponds with previous studies that found neither regional climatic nor local habitat differences between diploid and tetraploid populations of *A. arenosa* in this contact zone (Wos et al., 2019; Morgan et al., 2020). The fact that diploids and tetraploids do not show niche differentiation indicates that tetraploids can coexist in the same niche as their diploid progenitors. These results confirm a growing body of empirical works showing that polyploids and diploids living in similar ecological niches tend to be common in autopolyploids (Godsoe et al., 2013; Glennon et al., 2014, Kirchheimer et al., 2016; Castro et al., 2019). Some such autopolyploids can escape or reduce competition with diploid progenitors by other changes in phenology, pollinators or parasite interactions (Thompson et al., 1997; Segraves and Thompson, 1999), or post-pollination prezygotic barriers involving pollen competition and altered pollen-stigma interactions (Husband, 2016, Castro et al., 2020b). In the autotetraploid *A. arenosa*, other processes than niche differentiation might promote reproductive isolation between cytotypes, which in turn allowed tetraploids to get established in sympatry with their diploid progenitors.

After the presumed single origin ~ 20-30 thousands of generations ago (Monnahan et al., 2019), the tetraploids spread across Europe and diversified into four other lineages. When niches of these allopatric tetraploid lineages are compared to the ancestral WCA diploids, the results largely vary among lineages. Tetraploids from the Alps (4x-ALP) exhibited a high overlap with the niche of 2x-WCA lineage, but also a niche shift towards more humid climatic conditions (as compared to 2x-WCA lineage; Table 1, Fig. 2 and Fig. 3) is observed. Other studies have found similar patterns of a considerable degree of niche overlap but some niche shift of polyploids to more humid conditions with respect to progenitors (Muñoz-Pajares et al., 2018; Gaynor et al., 2018). Tetraploids from SE Carpathians have experienced a considerable niche contraction mainly related to precipitation gradient (Fig. 2). On the other hand, the Ruderal and C European tetraploid lineages have significantly expanded their niche to the ancestral diploids, 2x-WCA (Table 1, Fig. 2) either towards warmer climatic conditions (4x-CEU) or both colder and warmer climates (4x-RUD; Table 1, Fig. 2). Indeed, it has been shown in ruderal *A. arenosa* populations that these plants are heat and cold stress-tolerant and do not require vernalization (Baduel et al., 2016). All these phenotypic traits could have facilitated the colonization of human-associated habitats, especially in Northern Europe which in turn explains the observed expansion of their climatic niche to wider gradients of temperature. Furthermore, when we compare the climatic niches of these diverged tetraploid lineages to the tetraploid lineage that still occupies the ancestral area (4x-WCA) we also observe significant differentiation, suggesting notable post-WGD evolution of tetraploid niches (Table 1, Fig. 3). Most of them have experienced some niche expansion with respect to the ancestral 4x-WCA (Table 1) and they have colonized very different climatic conditions either towards drier and warmer (4x-RUD and 4x-CEU), colder (4x-SEC) and more humid environments (4x-ALP; Fig. 3).

Overall, our study confirms that niche evolution of polyploids is detectable only when polyploid lineages are compared with their corresponding diploid ancestor lineage, not globally. While niche similarity is observed when comparing the niches of cytotypes coexisting in the ancestral area of the tetraploid cytotype, niche expansion is mainly driven by post-WGD diversification into several lineages, primarily those that colonized warmer and drier anthropogenic habitats. These results demonstrate that niche shift is likely not driven by WGD per se in *A. arenosa* but rather reflects dynamic post-WGD evolution in the species, involving tetraploid migration and potential further interactions of tetraploids with other diploid lineages as is discussed in the next section.

2. Contribution of interploidy introgression to tetraploid niche expansion

Previous analyses demonstrated that genetic distinctness of particular lineages (4x-SCA, 4x-RUD) reflects strong interploidy gene flow after secondary contact with distantly related diploid lineages that have diverged before the WGD event (Monnahan et al., 2019). Such interploidy gene flow could have increased phenotypic and genetic variation which

may allow the tetraploid to colonize new ecological niches, in a way that is somewhat analogous to expectations stemming from the hybrid origin of allopolyploids (e.g. Parisod and Broenniman, 2016). Accordingly, the effect of introgression can be tested using an ecological niche comparison approach by looking at changes in terms of niche dynamic components such as unfilling and expansion. Here we leveraged the exceptionally well-described evolutionary history of *A. arenosa* involving localized post-WGD introgression from divergent diploids into two distinct tetraploid lineages (SCA and RUD, Monnahan et al. 2019) to assess if considering a niche of such additional diploid "donor" in addition to the 2x-WCA ancestor help to explain inter-ploidy niche divergence. Although notable effects have been observed in both cases, the results, once again, differed for each lineage. In the case of the 4x-SCA lineage, adding niche requirements of the sympatric 2x-SCA lineage had not explained tetraploid expansion but rather indicated tetraploid's contraction in terms of temperature. These results show that interploidy gene flow between the 4x-SCA and the sympatric but evolutionary distant 2x-SCA has not led to a broader niche of the 4x-SCA lineage (Fig. 4). On the other hand introgression of the Baltic diploid lineage partly explains the massive expansion of the tetraploid Ruderal lineage towards more northerly habitats that are characteristic of the Baltic lineage. Indeed, previous studies have demonstrated that introgression of alleles involved in flowering time regulation from 2x-BAL to 4x-RUD has played a role in the early flowering of tetraploid ruderal *A. arenosa* (Baduel et al., 2018). These results suggest a potential effect of interploidy gene flow in patterns of climatic niche evolution of tetraploid ruderal plants of *A. arenosa*. Thus, our study supports that interploidy introgression may be an important process for the adaptation of plants to challenging environments.

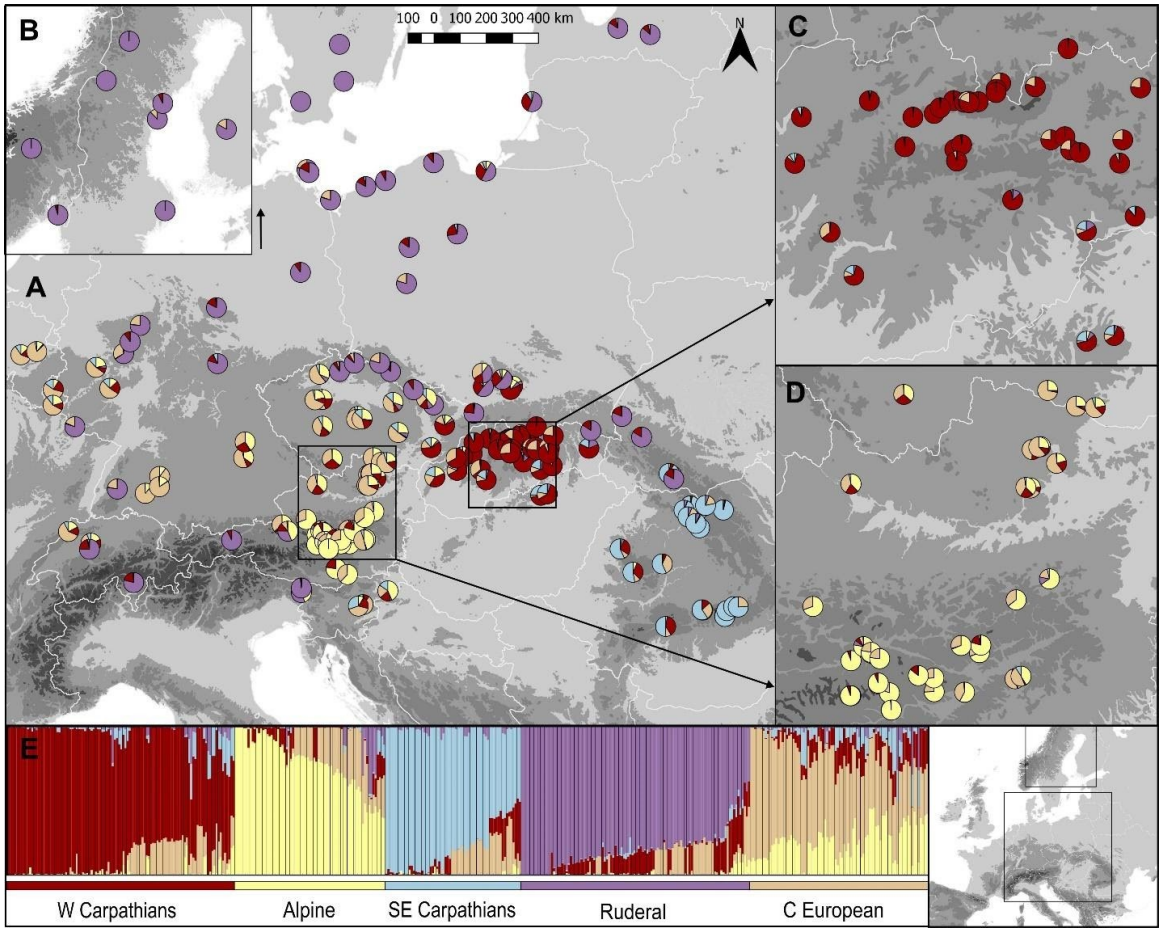


Figure 1. A Distribution of tetraploid *A. arenosa* populations used in the genomic analyses (colour pie charts reflecting the proportional assignment to particular clusters identified by STRUCTURE); **B** Detail of Scandinavia; **C** Detail of the Western Carpathian region; **D** Detail of the Austrian region; **E** Individual assignment into different clusters representing tetraploid lineages within *A. arenosa*: W Carpathian (dark red), Alpine (light yellow), SE Carpathian (light blue), Ruderal (purple), C European (light orange) lineages.

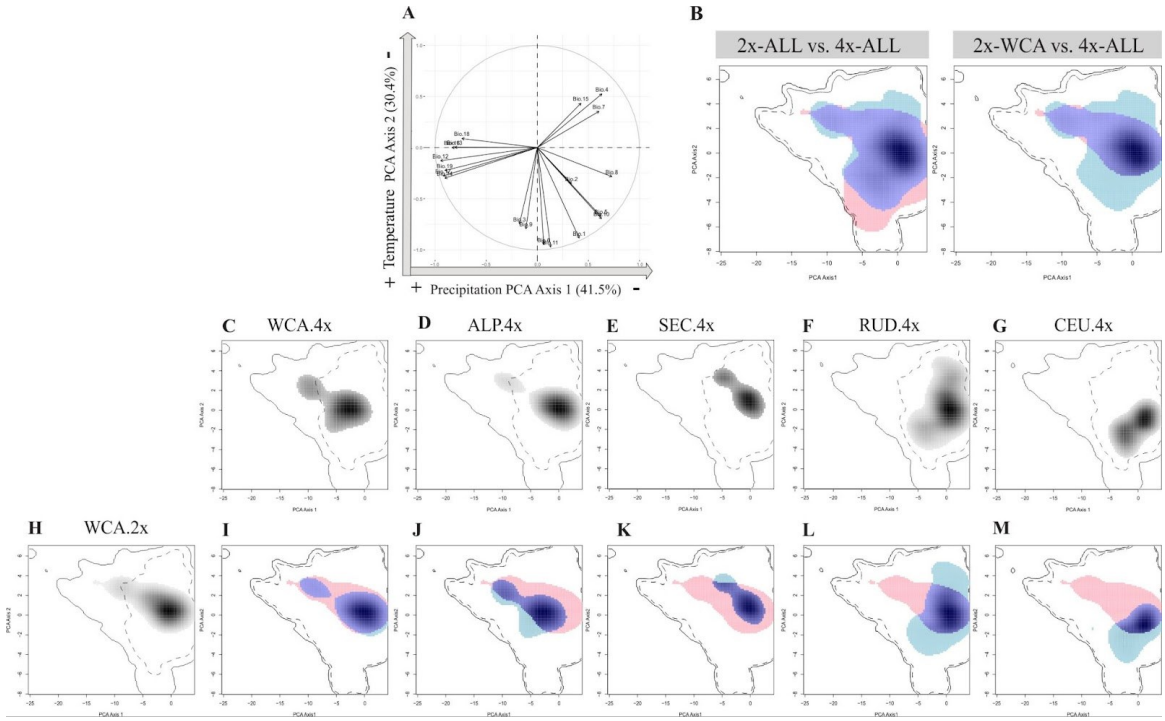


Figure 2. Climatic niche dynamics of diploid and tetraploid *A. arenosa* lineages. The fraction of diploid niche that remains unfilled by tetraploids represents tetraploid niche contraction (pink), whereas climatic space occupied by the tetraploid and not by the diploids indicates niche expansion (blue). The fraction of the tetraploid climatic niche that overlaps with diploids indicates niche stability (dark blue). **A** Correlation circle of 19 environmental variables along the first two axes of PCA-env. The percentage of variation and the main environmental gradient explained by each axis is indicated; **B** Niche of all *A. arenosa* diploids compared to the niche of all *A. arenosa* tetraploids without considering their assignment to different lineages and the niche of the ancestral diploid lineage (2x-WCA) compared to the niche of all A. arenosa tetraploids; **C-H** Climatic niche reconstruction of each *A. arenosa* tetraploid lineage and the ancestral diploid W Carpathian lineage. Occurrence density grids are represented by a black-to-white downward gradient along the two first axes of PCA-env; **I-M** Comparisons of climatic niches of each tetraploid lineage to their ancestral diploid lineage.

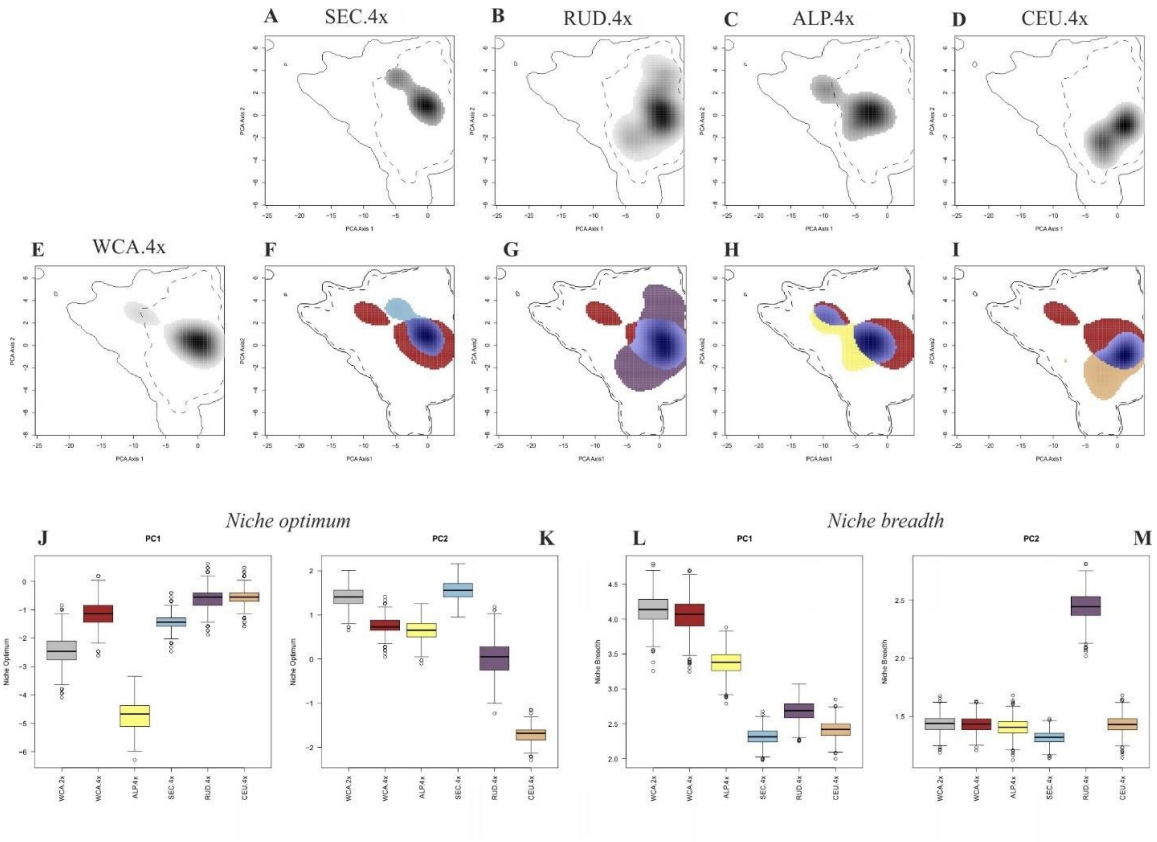


Figure 3. A-E Climatic niche reconstruction for each tetraploid lineage of *A. arenosa*. Occurrence density grids are represented by a black-to-white downward gradient along the two first axes of PCA-env. **F-I** Niche comparisons between each tetraploid lineage and the ancestral W Carpathian tetraploid. The fraction of the W Carpathian tetraploid niche that remains unfilled is represented in dark red, whereas the niche expansion of each tetraploid lineage coloured according to Fig. 2. The niche overlap is indicated in dark blue. **J-K** Niche optimum and breadth for W Carpathian diploid and each of the tetraploid *A. arenosa* lineages along the PC1 and PC2 axes.

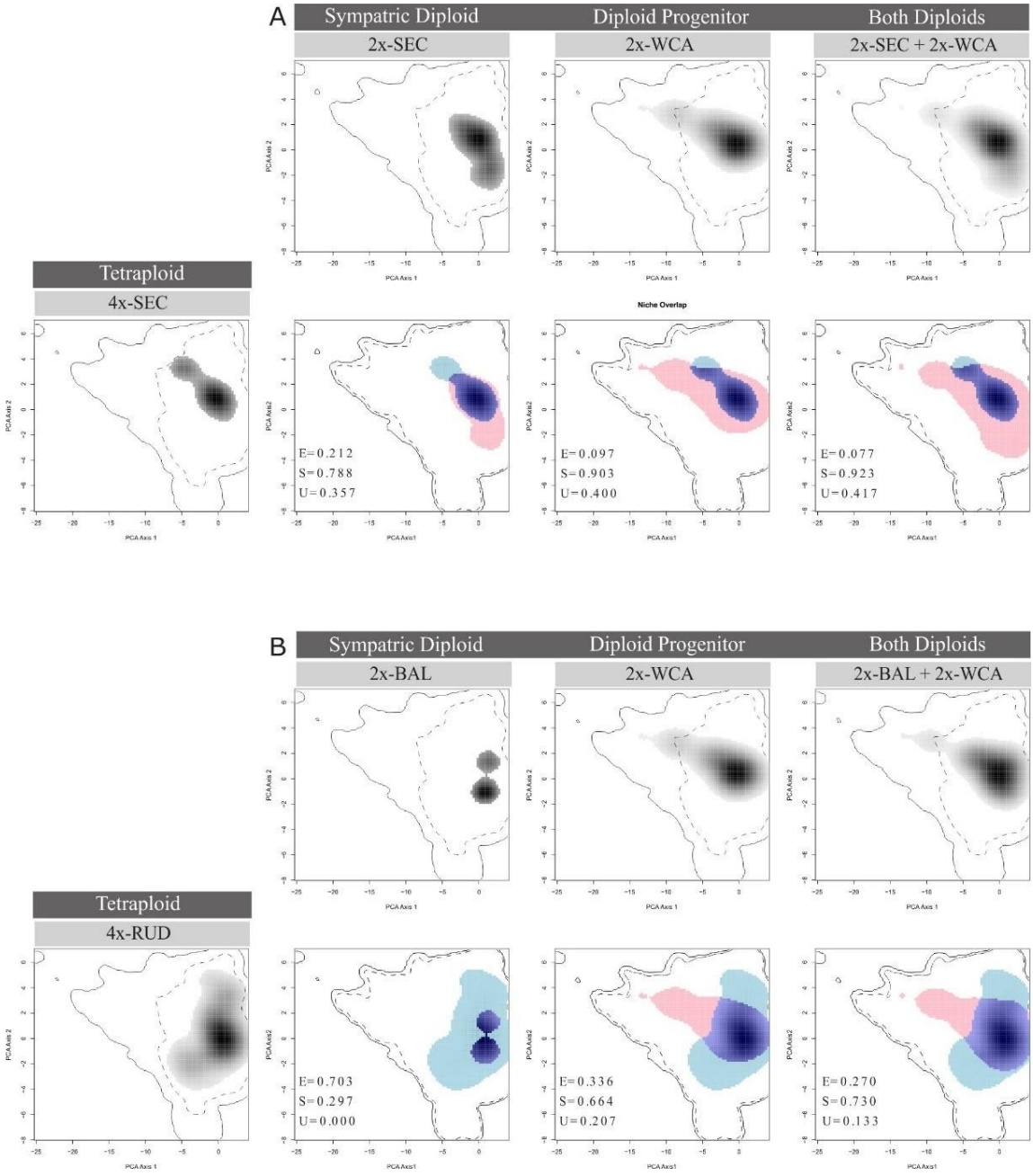
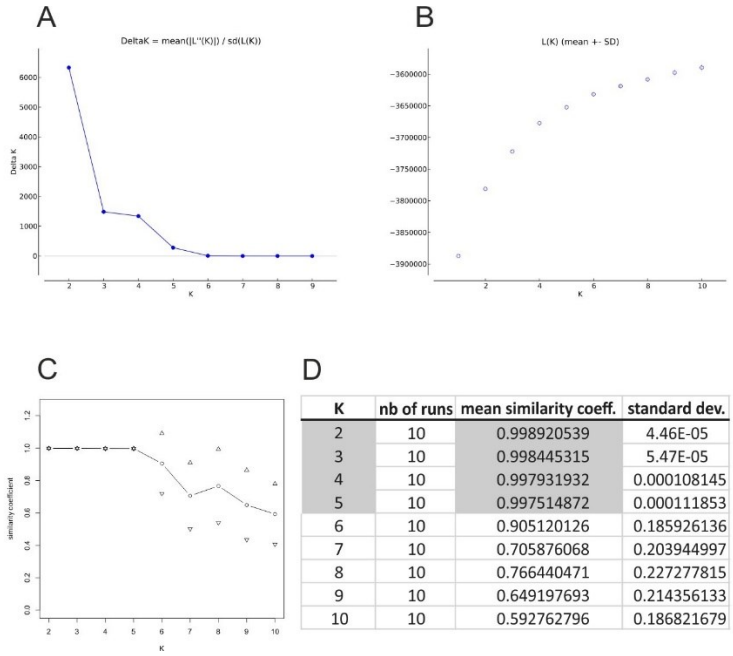
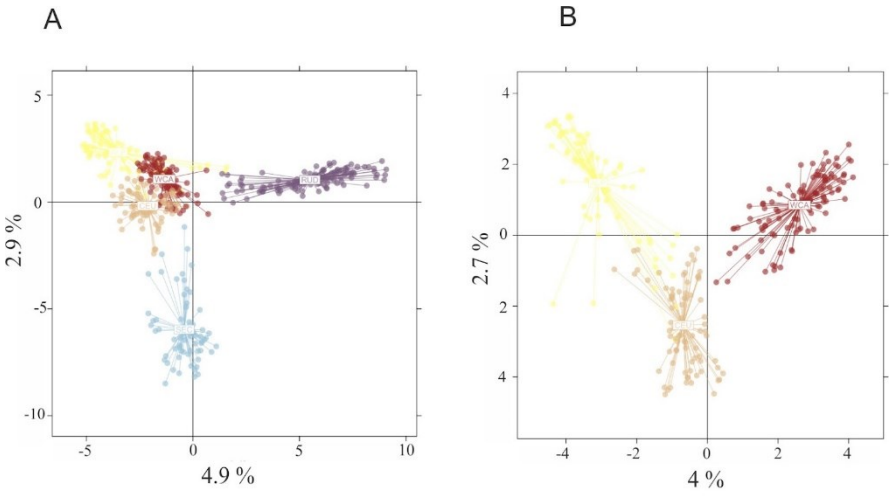


Figure 4. Climatic niche comparisons of observed and expected niches for SE Carpathian tetraploid (**A**) and Ruderal tetraploid (**B**) lineages. The occurrence density grid for each diploid and tetraploid lineage is represented in the climatic space along with the first two principal components (black and white plots). The observed niche of each tetraploid lineage is compared to the observed niche of their respective putative diploid progenitor and sympatric diploid and to their expected niches, which are calculated as their combined niche spaces. The fraction of the tetraploid climatic niche that overlaps with the diploid niche space indicates niche stability (S, dark blue), whereas the fraction that remains unfilled represents tetraploid niche contraction (U, pink). Climatic space occupied by the tetraploid lineages while not predicted by the combination of diploid progenitors indicates niche expansion (E, blue).



K	nb of runs	mean similarity coeff.	standard dev.
2	10	0.998920539	4.46E-05
3	10	0.998445315	5.47E-05
4	10	0.997931932	0.000108145
5	10	0.997514872	0.000111853
6	10	0.905120126	0.185926136
7	10	0.705876068	0.203944997
8	10	0.766440471	0.227277815
9	10	0.649197693	0.214356133
10	10	0.592762796	0.186821679

Figure S1. **A, B** Mean delta *K* and likelihood values respectively, both calculated based on the Evanno method (Evanno et al. 2005), **C, D** Plot and values of mean similarity coefficients and standard deviation considering 10 replicates for each value of *K*.



C

	CEU	RUD	ALP	SEC	WCA
CEU	0				
RUD	0.00396	0			
ALP	0.00181	0.00473	0		
SEC	0.00282	0.00443	0.00403	0	
WCA	0.00194	0.00361	0.00248	0.00327	0

Figure S2. **A** Principal Components Analysis (PCA) of all tetraploid *A. arenosa* lineages identified in this study. Axis 1 and axis 2 explain 4.9% and 2.9% of the variation, respectively, **B** PCA including W Carpathians, C European and Alpine tetraploid lineages of *A. arenosa*. Two first axes explain 4% and 2.7% of the total variance, respectively. **C** Nei's genetic distances among the identified lineages.

Table 1. Niche overlap values (D), results for equivalency and similarity tests and niche dynamic indices (expansion, stability and unfilling) are shown for pairwise comparisons (lineage 1 vs. lineage 2)

lineages		% Niche overlap (*D*)	Niche equivalency	Niche similarity		Indices of niche change		
1	2			1 --> 2	2 --> 1	Expansion (E)	Stability (S)	Unfilling (U)
2x-ALL	4x-ALL	50.2	ns (EQUIVALENT)	More similar *	More similar *	0.053	0.947	0.079
2x-WCA	4x-ALL	62.8	ns (EQUIVALENT)	More similar *	More similar *	0.252	0.748	0.006
2x-WCA	4x-WCA	64.9	ns (EQUIVALENT)	More similar *	More similar *	0.017	0.983	0.127
2x-WCA	4x-ALP	52.1	ns (EQUIVALENT)	ns	ns	0.215	0.785	0.350
2x-WCA	4x-SEC	24.6	ns (EQUIVALENT)	ns	ns	0.097	0.903	0.400
2x-WCA	4x-RUD	28.1	ns (EQUIVALENT)	ns	ns	0.336	0.664	0.207
2x-WCA	4x-CEU	7.7	ns (EQUIVALENT)	ns	ns	0.479	0.521	0.607
4x-WCA	4x-ALP	46.8	ns (EQUIVALENT)	ns	ns	0.329	0.672	0.347
4x-WCA	4x-CEU	14.4	ns (EQUIVALENT)	ns	ns	0.447	0.553	0.438
4x-WCA	4x-RUD	32.4	ns (EQUIVALENT)	ns	ns	0.364	0.636	0.098
4x-WCA	4x-SEC	21.6	ns (EQUIVALENT)	ns	ns	0.258	0.742	0.390

Table S1. Details of the *Arabidopsis arenosa* populations included in this study. N_{ma} = number of individuals used for molecular analyses, N_{nc} = population used for niche comparison method.

Ploidy	Lineage	Pop code	N_{ma}	N_{nc}	Locality details
2x	BA	**AA153**	2	yes	D, Mecklenburg-Vorpommern, Usedom, Bansin, sandy slopes along the coast of Baltic sea (primary habitat), 22 m a.s.l., sand dunes, 53.9857N, 14.1251E, E. Záveská
2x	BA	**AA197**	3	yes	PL, Województwo Zachodniopomorskie, Miedzyzdróje, at the top of the sabulous traverse along seashore cca 1km western from the town Miedzyzdróje and around the pathway in the hut colonies in pinewood at the western border the town (distance between forest and seashore ca 100m), 19 m a.s.l., shady forest pathway, sand dunes, 53.9211N, 14.4216E, E. Záveská
2x	BA	**AA199**	3	yes	LT, Klaipėda, Curonian Spit, environs of Preila village, coast of the Baltic Sea, eastern slope of foredune, 1 m a.s.l., grey dune habitat, 55.3782N, 21.0323E, Z. Gudžinskas
2x	BA	**AA217**		yes	LV, Riga, Vecdangara (Riga), 11 m a.s.l., open pine forest along railway, 57.0423N, 24.1004E, J. Kalůsková, P. Vít
2x	BA	**AA258**	2	yes	LT, Riga, Between Pludmales iela and Pakalnes iela, 1 m a.s.l., secondary sand dunes, 57.0794N, 24.1037E, I. Rurane
2x	BA	**AA259**		yes	LT, Riga, At the end of Ilmena iela at the sea, 1 m a.s.l., secondary sand dunes, 57.0190N, 23.9640E, I. Rurane
2x	BA	**AA268**	2	yes	DK, Sjælland, Møns Klint, slopes 30 m SE of GeoCenter, 93 m a.s.l., open sandy slope between chalk cliffs, 54.9633N, 12.5506E, F. Kolář, E. Záveská
2x	BA	**AA269**		yes	DK, Sjælland, Stevns Klint, Holtug Kritbud old chalk pit, 22 m a.s.l., chalk rocks and scree, 55.3405N, 12.4441E, F. Kolář, E. Záveská
2x	BA	**AA275**	3	yes	PL, Województwo Zachodniopomorskie, Świnoujście, forest 2 km ESE of the port, 4 m a.s.l., pine forest and road bank, 53.8947N, 14.2924E, F. Kolář, E. Záveská
2x	BA	**AA276**		yes	PL, Województwo Zachodniopomorskie, Świnoujście, railway station 5 km E of the port, 3 m a.s.l., railway embankment, sandy soils, 53.8979N, 14.3336E, F. Kolář, E. Záveská
2x	BA	**AA278**		yes	PL, Województwo Zachodniopomorskie, Mielno, along the road in Unieście E of the town, 8 m a.s.l., sandy sites along the road, 54.2736N, 16.1041E, F. Kolář, E. Záveská
2x	BA	**AA313**	2	yes	UA, Tarnopilska oblasť, Horodnytsia, rocks E of the village, 376 m a.s.l., shady crevices in dry limestone rocks, 49.40776N, 26.065374E, E. Záveská
2x	BA	**AA394**	3	yes	UA, Kremenets, calcareous rocks above the NE margin of the town, rock margin semi-shaded by deciduous forest, 50.118939N, 25.739772E, F. Kolář, D. Požárová, M. Holcovár
2x	BA	**AA395**	1	yes	UA, Bilokrynytsia, old peat bog N of the village, dry places in excavated peatbog and mire with scattered sandy outcrops, 50.16107N, 25.726966E, F. Kolář, D. Požárová
2x	BA	**AA454**		yes	D, Mecklenburg-Vorpommern, Landkreis, Vorpommern-Rügen (VR), Amt Bergen, Lietzow, Bahnhof, direkt am Zugang zu den Bahnsteigen, 1 m a.s.l., 54.48115N, 13.5105E, C.N.Schröder, R. Schmickl
2x	BA	**AA455**		yes	D, Mecklenburg-Vorpommern, Landkreis, Vorpommern-Rügen (VR), Amt Bergen, Lietzow, Gehölzrand ca. 400 m E des Bahnhofs, 6 m a.s.l., 54.48146N, 13.51544E, C.N.Schröder, R. Schmickl
2x	BA	**AA456**		yes	D, Mecklenburg-Vorpommern, Landkreis, Vorpommern-Rügen (VR), Amt Mönchgut-Granitz, Göhren, Nordperd, Steilhang am Nordstrand (die Population erstreckt sich bis zum Punkt 1844), 1 m a.s.l., 54.37808N, 13.70537E, C.N.Schröder, R. Schmickl
2x	BA	**AA457**		yes	D, Mecklenburg-Vorpommern, Landkreis, Vorpommern-Rügen (VR), Amt Mönchgut-Granitz, Sellin, Steilabbruch ca. 800 m südlich der Seebrücke (die Population erstreckt sich bis zum Punkt 1844), 3 m a.s.l., 54.33922N, 13.75609E, C.N.Schröder, R. Schmickl
2x	BA	**AA458**	2	yes	D, Mecklenburg-Vorpommern, Landkreis, Vorpommern-Rügen (VR), Amt Mönchgut-Granitz, Göhren, Nordperd, Sandhang am Südstrand (die Population erstreckt sich bis zum Punkt 1837), 1 m a.s.l., 54.34256N, 13.75531E, C.N.Schröder, R. Schmickl
2x	BA	**AA459**		yes	D, Mecklenburg-Vorpommern, Landkreis, Vorpommern-Rügen (VR), Amt Mönchgut-Granitz, Göhren, Nordperd, Steilhang am Nordstrand (die Population erstreckt sich bis zum Punkt 1846), 1 m a.s.l., 54.34228N, 13.7568E, C.N.Schröder, R. Schmickl
2x	BA	**AA460**		yes	D, Mecklenburg-Vorpommern, Landkreis, Vorpommern-Rügen (VR), Amt Mönchgut-Granitz, Göhren, Nordperd, Steilhang am Nordstrand (die Population erstreckt sich bis zum Punkt 1848), 1 m a.s.l., 54.34146N, 13.76391E, C.N.Schröder, R. Schmickl
2x	BA	**AA461**	2	yes	D, Mecklenburg-Vorpommern, Landkreis, Vorpommern-Rügen (VR), amtsfreie Gemeinde Binz, Granitzer Ort (abruptes Ende der Population bei Punkt 1863), 1 m a.s.l., 54.40117N, 13.66298E, C.N.Schröder, R. Schmickl
2x	DI	**AA048**		yes	HR, Zagrebačka županija, Medvedica, slope above a forest road and shady rocks, 326 m a.s.l., slope above forest road + dolomitic rocks (P-Y), 45.8691N, 15.8506E, G. Fuxová
2x	DI	**AA050**		yes	SLO, Idrija, Idrija, hillside along the road from Idrija to Godovič, 500 m a.s.l., roadside, 45.9693N, 14.0621E, G. Fuxová

2x	DI	**AA052**		yes	HR, Primorsko-goranska županija, Gusti Laz, river bank near the village, 276 m a.s.l., alluvial gravel, river bank, 45.4606N, 14.8255E, G. Fuxová
2x	DI	**AA054**		yes	HR, Ličko-Senjska županija, Plitvička jezera lakes, along tourist paths, steep rocky valley, 561 m a.s.l., rocky valley, 44.9043N, 15.6111E, G. Fuxová
2x	DI	**AA106**	3	yes	HR, Krapina-Zagorje, Belecgrad (the ruin of the Belec castle), N of the village of Juranščina, N of the town of Zlatar, in the Ivanščica Mountain, 550 m a.s.l., along the hiking path, open sites in the forest and walls of the castle ruin, 46.1617N, 16.115E, S. Španiel, K. Marhold, J. Zozomová
2x	DI	**AA108**		yes	SLO, Koper, Koper: 11.5 km E of the town, 260 m WSW of Predloka village, 97 m a.s.l., rocks and eroded slope above road, 45.54N, 13.8747E, M. Lučanová
2x	DI	**AA109**		yes	SLO, Divača, Škocjan: slope along the road below the church at N end of the village, 422 m a.s.l., rocks above road, 45.6658N, 13.9939E, M. Lučanová
2x	DI	**AA127**	3	yes	BIH, Federacija Bosna i Hercegovina, Fojnica, slope above road from Fojnica to Turkovići, 6 km W of Fojnica, 754 m a.s.l., shady rocky slope above a road, 43.975N, 17.8245E, F. Kolář, G. Fuxová
2x	DI	**AA128**		yes	BIH, Federacija Bosna i Hercegovina, Bugojno, rocks above the road Bugojno-Kupres, 5 km W of the town, 772 m a.s.l., rocks and alluvial gravel, 44.0426N, 17.3741E, F. Kolář, G. Fuxová
2x	DI	**AA129**		yes	BIH, Federacija Bosna i Hercegovina, Jajce, rocks below parking place next to the waterfall in the city centre, 384 m a.s.l., rocks and eroded slope, 44.3376N, 17.27E, F. Kolář, G. Fuxová
2x	DI	**AA130**	3	yes	BIH, Federacija Bosna i Hercegovina, Bihać, rocks above road to Bosanska Krupa, 8 km N of the town, 217 m a.s.l., rocks, 44.8818N, 15.8988E, F. Kolář, G. Fuxová
2x	DI	**AA131**		yes	HR, Karlovac, Slunj, rocks below main road bridge across the Korana river, 221 m a.s.l., rocky slope, 45.1216N, 15.5889E, F. Kolář, G. Fuxová
2x	DI	**AA132**		yes	HR, Krapina-Zagorje, Golubovac, slopes above the road Golubovac-Lepoglava and close limestone quarry, 255 m a.s.l., disturbed slopes above road, limestone rocks, 46.1918N, 15.9889E, F. Kolář, G. Fuxová
2x	DI	**AA155**		yes	SLO, Lukovica, Podmilj: beech forest slope 1.3 km NE of the village, 520 m a.s.l., eroded slope above road, 46.1812N, 14.8552E, M. Lučanová
2x	DI	**AA161**		yes	SLO, Cerknica, Rakov Škocjan: Veliky naravni most bridge in a canyon N of Rakov Škocjan village, 543 m a.s.l., rocks, 45.7958N, 14.2879E, M. Lučanová
2x	DI	**AA266**		yes	SLO, Kočevje, Bezgarska planina, Bezgovica: around the route in the mountain saddle ca 1.3 km ENE of the village, 970 m a.s.l., disturbed slopes and forest rocks above the road, natural site, 45.5533N, 14.7249E, M. Štech, J. Košnar, J. Laburdová
2x	DI	**AA342**		yes	SLO, Logatec, Rocky slopes along the path to Vrania jama cave, 500 m a.s.l., rocks, 45.8739414N, 14.2464336E, F. Kolář, M. Holcová, D. Požárová
2x	DI	**AA343**	2	yes	SLO, Planina, Rocks and gravely patches along the path to the entrance to Planinska jama cave, 520 m a.s.l., rocks, gravel, disturbed slopes, 45.8213028N, 14.2467769E, F. Kolář, M. Holcová, D. Požárová
2x	DI	**AA379**	2	yes	SLO, Tolmin, N of Tolmin, rocks along roadside, 46.187126N, 13.723346E, F. Kolář, E. Morgan
2x	DI	**AA380**		yes	SLO, Goriška, rocky slope above road, 46.160365N, 13.815243E, F. Kolář, E. Morgan
2x	DI	**AA381**		yes	SLO, Gorenjska, rocky slope above road, 46.192323N, 14.025872E, F. Kolář, E. Morgan
2x	DI	**AA382**		yes	SLO, Goriška, gorge, 46.1522367N, 14.0275394E, F. Kolář, E. Morgan
2x	DI	**AA385**		yes	SLO, Jugovzhodna Slovenija, Dolenjska, Dolenjske Teplice - semič: 1.2 km NNW of Črmošnjice, 385 m a.s.l., rocks in forest, 45.682778N, 15.098611E, B. Frajman
2x	DI	**AA389**		yes	SLO, Savinjska, 46.083122N, 15.173869E, F. Kolář, E. Morgan
2x	DI	**AA393**		yes	SLO, Osrednjeslovenska, 46.166988N, 14.569014E, F. Kolář, E. Morgan
2x	DI	**AA408**	2	yes	SLO, Savinjska, Zidani Most, alongside the small road N of the village, rocky slopes and gravel, 46.09921N, 15.20572E, E. Morgan, M. Holcová, D. Bohutínský
2x	DI	**AA424**		yes	SLO, Sevnica, rocks above the road , 183.7 m a.s.l., rocky slope above the road, 45.998683N, 15.272037E, F. Kolář
2x	DI	**AA426**		yes	SLO, Celje, Gračnica, rocks left of the road to Jurklošter, 228.5 m a.s.l., base of limestone rocks, 46.106603N, 15.24263E, F. Kolář
2x	DI	**AA427**	2	yes	SLO, Celje, Lipni Dol, rocks left of the road to Jurklošter, 303.4 m a.s.l., rocky scree at the base of limestone rocks, 46.10676N, 15.30565E, F. Kolář
2x	DI	**AA429**		yes	SLO, Celje, Along rock wall, west of Pod Kojzico,rock wall, 46.1321N, 15.1854E, E. Morgan, C. Lafon Placette
2x	DI	**AA508**		yes	BIH, Olovo, Krivaja river gorge downstream of the town, 600 m a.s.l., rocks and road bank, 44.1249357N, 18.5713039E, F. Kolář, D. Požárová
2x	DI	**AA509**		yes	BIH, Ribnica, serpentine rocks above the road at the SE end of the village, 330 m a.s.l., rocks, 44.3418063N, 18.4052057E, F. Kolář, D. Požárová
2x	PA	**AA011**		yes	SK, Banskobystrický kraj, Medovarce, bank of the road Medovarce-Domaníky, 1 km NW of the village, 247 m a.s.l., road ditch and open forest, 48.237N, 18.9957E, E. Záveská, J. Kučera, F. Kolář

2x	PA	**AA034**	3	yes	HU, Pest megye, Budaörs, Törökugrató rock at the N border of the town, 252 m a.s.l., rocky steppe, 47.458N, 18.9248E, F. Kolář
2x	PA	**AA035**		yes	HU, Heves megye, Szarvaskő, rocks along the road N of the town, 231 m a.s.l., rocks, 47.9902N, 20.3274E, F. Kolář
2x	PA	**AA036**		yes	HU, Heves megye, Sirok, castle ruins, 314 m a.s.l., rocks, castle ruins, 47.9391N, 20.1963E, F. Kolář
2x	PA	**AA037**	1	yes	HU, Heves megye, Kékes, Sas-kő rock E of the summit of Kékes, 898 m a.s.l., rocks, 47.8737N, 20.0304E, F. Kolář
2x	PA	**AA038**		yes	HU, Baranya megye, Pécs, SE slope of Misina-tető, close to Dömörkapu, 408 m a.s.l., rocky steppe, 46.0992N, 18.2339E, F. Kolář
2x	PA	**AA039**		yes	HU, Veszprém megye, Veszprém, N exposed rocks in the town, 318 m a.s.l., rocks, 47.0972N, 17.8988E, F. Kolář
2x	PA	**AA040**		yes	HU, Veszprém megye, Kisapáti, W and NW exposed slopes of Szent György-hegy, 402 m a.s.l., rocky steppe, 46.8439N, 17.4514E, F. Kolář
2x	PA	**AA041**		yes	HU, Veszprém megye, Sümeg, castle hill and ruins, 318 m a.s.l., rocks, castle ruins, 46.9823N, 17.2815E, F. Kolář
2x	PA	**AA110**		yes	SK, Nitrianský kraj, Kováčovské kopce hills: SE hillside, W edge of the Kováčov village, 161 m a.s.l., rocky hillside, 47.8239N, 18.7782E, M. Lučanová, E. Záveská
2x	PA	**AA112**		yes	SK, Banskobystrický kraj, Dolná Strehová: forest 1.6 km E from the village, 203 m a.s.l., oak and hornbeam forest, sidehill with stones, 48.2494N, 19.5107E, M. Lučanová, E. Záveská
2x	PA	**AA113**		yes	SK, Banskobystrický kraj, Lučenec district, Šiatorská Bukovinka village, Šomoška castle, 496 m a.s.l., grasslands, screes and rocks round the whole castle, 48.1715N, 19.8564E, M. Lučanová, E. Záveská
2x	PA	**AA211**		yes	HU, Pest, Pomáz, forest NW of the village, 450 m a.s.l., open woodland, 47.6727N, 18.9880E, C. Pachschwöll
2x	PA	**AA347**	3	yes	SK, for locality details see Arnold et al. 2015, 280 m a.s.l., 48.26694N, 19E, B. Arnold et al.
2x	PA	**AA348**	3	yes	HU, for locality details see Arnold et al. 2015, 330 m a.s.l., 47.72444N, 18.77917E, B. Arnold et al.
2x	PA	**AA349**	3	yes	HU, for locality details see Arnold et al. 2015, 130 m a.s.l., 46.80667N, 17.43444E, B. Arnold et al.
2x	PA	**AA489**	1	yes	SK, Banskobystrický kraj; Šášovské Podhradie, for locality details see Novikova et al. 2016, 48.5817328N, 18.8930924E, R. Schmickl, G. Muir
2x	SEC	**AA062**		yes	RO, Arad, Şoimoş, castle ruin above the village, 261 m a.s.l., shady rocks, walls of castle ruin, 46.1089N, 21.7232E, F. Kolář, G. Fuxová
2x	SEC	**AA063**		yes	RO, Hunedoara, Deva, walls along the old road to the castle (citadela), 327 m a.s.l., ruined walls, 45.8897N, 22.8967E, F. Kolář, G. Fuxová
2x	SEC	**AA064**		yes	RO, Sibiu, Cisnădioara, SE slopes of the hill with old church S of the town, 570 m a.s.l., open oak forest with rocks, 45.7022N, 24.1115E, F. Kolář, G. Fuxová
2x	SEC	**AA066**	3	yes	RO, Argeş, Vidraru, slope above the Transfăgăraş road, 400 m ESE of the dam, 900 m a.s.l., eroded slope above road and road margins, + stone wall, 45.3639N, 24.6376E, F. Kolář, G. Fuxová
2x	SEC	**AA068**		yes	RO, Prahova, Buşteni, Bucegi Mts, upper part of Valea Jepilor, 1511 m a.s.l., eroded patches above path, 45.4065N, 25.4965E, F. Kolář, G. Fuxová
2x	SEC	**AA069**		yes	RO, Prahova, Buşteni, Bucegi Mts, middle part of Valea Jepilor, 1368 m a.s.l., shady rocks, 45.4056N, 25.5008E, F. Kolář, G. Fuxová
2x	SEC	**AA070**		yes	RO, Covasna, Bixad, forest margin, limestone rocks and old quarry 1.5 km NW village, 609 m a.s.l., shady rocks, quarry, 46.1091N, 25.8464E, F. Kolář, G. Fuxová
2x	SEC	**AA071**		yes	RO, Harghita, Praid, stream valley along the road from Praid to Gheorgheni, 8 km ENE of the village, 572 m a.s.l., shady rocks, stones along a brook, 46.5794N, 25.2321E, F. Kolář, G. Fuxová
2x	SEC	**AA080**	2	yes	RO, Harghita, Bălan, W slopes of Haşmaş Mare, 8 km NNW of the town, 1693 m a.s.l., shady rocks, eroded slope, 46.7159N, 25.7768E, F. Kolář, G. Fuxová
2x	SEC	**AA118**	2	yes	SK, Košický kraj, Ladmovce: 2.4 km NNW from the village, 267 m a.s.l., oak forest, disturbed from wild boars, 48.4331N, 21.77E, M. Lučanová, E. Záveská, J. Smatanová
2x	SEC	**AA121**		yes	SRB, Veliko Gradište, Golubac, slopes above the road Golubac-Dobra, 6 km E of Golubac, 81 m a.s.l., slopes above the road and shady rocks, 44.6557N, 21.7071E, F. Kolář, G. Fuxová
2x	SEC	**AA122**		yes	SRB, Petrovac, Gornjak, rocks behind the monastery, 184 m a.s.l., shady limestone rocks, 44.2653N, 21.5427E, F. Kolář, G. Fuxová
2x	SEC	**AA123**	3	yes	SRB, Raška, Jošanička Banja, Drenjska klisura gorge along the road Jošanička Banja-Grčak, 3.5 km ENE of Još. Banja, 704 m a.s.l., rocks and gravel below, 43.3995N, 20.7909E, F. Kolář, G. Fuxová
2x	SEC	**AA124**		yes	SRB, Vranje, Vranje, rocks above the road to Golemo Selo N of the town, 888 m a.s.l., shady rocks above road, 42.5919N, 21.8802E, F. Kolář, G. Fuxová
2x	SEC	**AA125**		yes	SRB, Uščje, Studenica, rocks above the road Studenica-Međurečje, in a river gorge NW of the village, 583 m a.s.l., rocks and gravel below, 43.5161N, 20.4526E, F. Kolář, G. Fuxová

2x	SEC	**AA126**		yes	SRB, Temška, Temštica river gorge along the road towards Topli Do, before the junction to power station, 470 m a.s.l., rocks, 43.2814282N, 22.5770235E, F. Kolář, D. Požárová
2x	SEC	**AA156**		yes	RO, Argeş, Fagaraš - Mircii stream valley, 1518 m a.s.l., roadside in deciduous forest, 45.582N, 24.7025E, G. Fuxová
2x	SEC	**AA157**		yes	RO, Argeş, Fagaraš - Mircii stream valley, 1330 m a.s.l., roadside in deciduous forest, 45.5749N, 24.7033E, G. Fuxová
2x	SEC	**AA158**		yes	RO, Argeş, Cetatea Poenari - along the road below lacul Vidraru dam, 714 m a.s.l., rocks along the road, 45.3523N, 24.6377E, G. Fuxová
2x	SEC	**AA159**		yes	RO, Hunedoara, Parang - Cabana Mija - stream valley, 985 m a.s.l., roadside, 45.4068N, 23.5057E, G. Fuxová
2x	SEC	**AA160**		yes	RO, Sibiu, Cindrel - Potoci stream valley, 1402 m a.s.l., roadside, gravel, in a rocky valley, 45.5294N, 23.6991E, G. Fuxová
2x	SEC	**AA214**		yes	RO, Caraș-Severin, Banát, Sv, Helena, rocks in forest 3,5 km NE of the village, 450 m a,s,l,, calcareous rocks in beech forest, 44.7065N, 21.7419E, M. Hroneš
2x	SEC	**AA220**	3	yes	RO, Bistrița-Năsăud, Parva, slopes above the road in a deep valley of Rebra brook at N margin of the village, 545 m a.s.l., open sites above road in deciduous forest, 47.4023N, 24.5461E, F. Kolář, M. Holcová
2x	SEC	**AA221**	2	yes	RO, Sibiu, Făgăraș Mts., Cârțișoara, rocky slopes along Transfăgărășan road, 10 km S of the village, 1203 m a.s.l., rocks and scree, 45.6431N, 24.6055E, F. Kolář, M. Holcová
2x	SEC	**AA224**		yes	RO, Vâlcea, Călimănești, rockas above the main road opposite of Mănăstirea Turnu monastery (left bank of Olt), 6 km NNW of the settlement , 315 m a.s.l., rocks, 45.2898N, 24.2965E, F. Kolář, M. Holcová
2x	SEC	**AA250**	2	yes	RO, Argeş, Pisicii, river canyon , 873 m a.s.l., shady rocks in a canyon, dry drain of the stream, looks primary, 45.5299N, 25.2874E, E. Záveská
2x	SEC	**AA257**		yes	RO, Braşov, Piatra Craiului Mica, 1205 m a.s.l., rocks along path in the forest, 45.5348N, 25.2816E, S. Španiel
2x	SEC	**AA290**		yes	UA, Zakarpatska oblasť, Dilove - along H09 road in valley, 373 m a.s.l., gravel/calcareous rock along road, 47.95309N, 24.18688E, M. Holcová, J. Hojka, D. Bohutínský, F. Rooks
2x	SEC	**AA292**	2	yes	UA, Zakarpatska oblasť, In the canyon of the river, right bank downstream, on the rock, 423 m a.s.l., rock, 48.25683N, 23.62612E, M. Holcová, J. Hojka, D. Bohutínský, F. Rooks
2x	SEC	**AA317**	2	yes	RO, Suceava, Broșteni, along the road Bicaz - Vatra Dornei, 650 m a.s.l., rocks, slope above road and road ditch, 47.266199N, 25.656536E, F. Kolář
2x	SEC	**AA328**	3	yes	RO, Hunedoara, Rau de Mori, rocky slope above the road to Gura Apelor damm, 828 m a.s.l., eroded rocky slope and gravely road bank, 45.37778N, 22.75833E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
2x	SEC	**AA329**	2	yes	RO, Brașov , Satu Lung, small rock next to the road to Cheia, 832.7 m a.s.l., hornbeam forest, rocks, 45.543118N, 25.821988E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
2x	SEC	**AA331**		yes	RO, Brașov, Fundata, open rocky slope above road Bran-Rucăr, W of the village, 1252.7 m a.s.l., rocks, 45.437866N, 25.258971E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
2x	SEC	**AA332**	3	yes	RO, Argeș, Dâmbovicioara, cabana Brusturet, rocks in gorges and hills below and above the cabin, 1013.8 m a.s.l., rocks, mostly in gorges, 45.467007N, 25.227617E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
2x	SEC	**AA333**		yes	RO, Brașov, Bran, rock with a cross opposite of the Dracula castle, 792,7 m a.s.l., rocks, 45.514675N, 25.364979E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
2x	SEC	**AA334**		yes	RO, Brașov, Moieciu de Sus, rocky slopes in gorge along road to Moieciu de Jos, NW of the village, 956.2 m a.s.l., rocks and screes, in a gorge, 45,46641N, 25.305526E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
2x	SEC	**AA336**		yes	RO, Brașov, Zărneşti, rocks along a tourist path from Poiana Zănoaga to the town, upper part of the gorge, 1223.9 m a.s.l., rocks in a Fagus forest, 45.534954N, 25.281389E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
2x	SEC	**AA337**		yes	RO, Alba, Căpâlna, rocks along the road to Șugag, S of the village, 424.8 m a.s.l., open rocks and screes above the road, 45.813945N, 23.616676E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
2x	SEC	**AA420**		yes	RO, Bistrița-Năsăud, Along side road leading to restaurant, 622 m a.s.l., Rocks above road and gravel to side of road, 47.43722N, 24.66341E, E. Morgan, M. Lučanová
2x	SEC	**AA421**		yes	RO, Suceava, Along roadside, 913 m a.s.l., Rocky slope above road, 47.56396N, 25.15036E, E. Morgan, M. Lučanová
2x	SEC	**AA422**		yes	RO, Suceava, Along roadside, 789 m a.s.l., Grassy scree slopes next to road, 47.39135N, 25.48212E, E. Morgan, M. Lučanová
2x	SEC	**AA423**		yes	RO, Suceava, Along roadside, 672 m a.s.l., Rocky/mossy slopes above road, 47.31408N, 25.62566E, E. Morgan, M. Lučanová
2x	SEC	**AA441**		yes	RO, Brašov, Piatra Craiului, Fundatica, 1089.2 m a.s.l., 45.41626N, 25.28812E, F. Kolář, D. Požárová, M. Bohutínská, D. Bohutínský
2x	SEC	**AA442**		yes	RO, Brašov, Piatra Craiului, near la Uluce cave, 1019.6 m a.s.l., 45.40828N, 25.26371E, F. Kolář, D. Požárová, M. Bohutínská, D. Bohutínský

2x	SEC	**AA445**		yes	RO,Brašov, Piatra Craiului, at the end of Drumul Fruntes road, 1283.3 m a.s.l., 45.49815N, 25.22847E, F. Kolář, D. Požárová, M. Bohutínská, D. Bohutínský
2x	SEC	**AA446**		yes	RO, Brašov, Piatra Craiului, near the turistic path, 1271.4 m a.s.l., 45.52626N, 25.2402E, F. Kolář, D. Požárová, M. Bohutínská, D. Bohutínský
2x	SEC	**AA448**		yes	RO, Brašov, Piatra Craiului, Solomon, at the resting place, 754.3 m a.s.l., 45.61705N, 25.5588E, F. Kolář, D. Požárová, M. Bohutínská, D. Bohutínský
2x	SEC	**AA507**		yes	SRB, Donji Milanovac, Majdanpek, mine tailings at the SW border of the town, 382 m a.s.l., mine tailings, 44.4193N, 21.9299E, F. Kolář, G. Fuxová
2x	WC	**AA007**	1	yes	SK, Trenčiansky kraj, Uhrovec, walls of the monument at Jankov Vŕšok hill, 541 m a.s.l., crevices in walls, 48.739N, 18.3664E, E. Záveská, J. Kučera, F. Kolář
2x	WC	**AA009**		yes	SK, Trenčiansky kraj, Podhradie, rocks below Sivý hrad castle, 598 m a.s.l., rocky steppe, 48.6863N, 18.6381E, E. Záveská, J. Kučera, F. Kolář
2x	WC	**AA012**	2	yes	SK, Banskobystrický kraj, Špania Dolina, along the road at the end of the village, 683 m a.s.l., eroded slope above road, 48.8063N, 19.1315E, E. Záveská, J. Kučera, F. Kolář
2x	WC	**AA013**		yes	SK, Žilinský kraj, Liptovská Osada, bottom of Žiar hill, at the NE end of the village, 637 m a.s.l., limestone rocks and slope in open forest, 48.9528N, 19.2687E, E. Záveská, J. Kučera, F. Kolář
2x	WC	**AA016**	6	yes	SK, Košický kraj, Podlesok, rocks in the entrance to Suchá Belá gorge, 600 m a.s.l., limestone rocks, 48.9603N, 20.3833E, E. Záveská, J. Kučera, F. Kolář
2x	WC	**AA017**		yes	SK, Prešovský kraj, Vernár, Kopanecké lúky 2.7 km SSE of the village, 929 m a.s.l., river bank and sediments, 48.8957N, 20.2821E, E. Záveská, J. Kučera, F. Kolář
2x	WC	**AA018**		yes	SK, Banskobystrický kraj, Muráň, old quarry at the entrance to Hrdzavá dolina valley, 467 m a.s.l., rocks in old quarry, 48.7462N, 20.0228E, E. Záveská, J. Kučera, F. Kolář
2x	WC	**AA020**		yes	SK, Prešovský kraj, Korytné, road bend 0.5 km E of the church, 604 m a.s.l., slopes above the road in deciduous forest, 49.0115N, 20.8475E, E. Záveská, F. Kolář
2x	WC	**AA021**	3	yes	SK, Prešovský kraj, Tatranská kotlina, rocks next to the entrance to Belianska jaskyňa cave, 900 m a.s.l., slopes and limestone rocks in mixed forest, 49.2289N, 20.3117E, E. Záveská, F. Kolář
2x	WC	**AA022**		yes	SK, Prešovský kraj, Tatranská Lesná, banks of the Studený potok brook, 1.5 km NW of the village, 1077 m a.s.l., brook banks, 49.1597N, 20.2456E, E. Záveská, F. Kolář
2x	WC	**AA023**	2	yes	SK, Žilinský kraj, Bešeňová, travertine rock 1 km NNE from the church in the village, 574 m a.s.l., travertine rock, 49.1073N, 19.4347E, E. Záveská, F. Kolář
2x	WC	**AA025**	3	yes	SK, Žilinský kraj, Strečno, limestone rocks at the entrance to the castle, 425 m a.s.l., limestone rocks, 49.1741N, 18.8617E, E. Záveská, F. Kolář
2x	WC	**AA084**	3	yes	SK, Prešovský kraj, Vysoké Tatry, wet rocks and gravel in Velická Dolina, between Sliezsky dom and Dlhé pleso, 1823 m a.s.l., alluvial gravel, wet rocks, 49.162N, 20.1542E, F. Kolář, E. Záveská, S. Španiel, M. Kolník, K. Marhold
2x	WC	**AA085**		yes	SK, Prešovský kraj, Vysoké Tatry, limestone gravel along road to Sliezsky dom in Velická Dolina, 1714 m a.s.l., roadside gravel, secondary, 49.1456N, 20.1627E, F. Kolář, E. Záveská, S. Španiel, M. Kolník, K. Marhold
2x	WC	**AA086**	2	yes	SK, Žilinský kraj, Zuberec, Zverovka, limestone rocks at the top of Osobitá mountain, 1552 m a.s.l., exposed rocks, gravel, 49.2593N, 19.7218E, F. Kolář, E. Záveská, S. Španiel, M. Kolník, K. Marhold
2x	WC	**AA089**		yes	SK, Prešovský kraj, Ždiar, Belianské Tatry, along tourist path in Monkova dolina valley, 1207 m a.s.l., shady rocks, alluvial gravel, 49.2484N, 20.2276E, F. Kolář, S. Španiel
2x	WC	**AA090**	3	yes	SK, Prešovský kraj, Vysoké Tatry, screes and rocks above the Zelené pleso lake up to Medený vodopád waterfall and Dlhý vodopád waterfall, 1625 m a.s.l., scree, wet rocks, 49.2065N, 20.2151E, F. Kolář, E. Záveská, S. Španiel, M. Kolník, K. Marhold
2x	WC	**AA091**	3	yes	SK, Prešovský kraj, Ždiar, Belianské Tatry, along blue tourist path in Kopské sedlo pass, 1751 m a.s.l., exposed rocks, eroded slope, 49.2299N, 20.219E, F. Kolář, E. Záveská, S. Španiel, M. Kolník, K. Marhold
2x	WC	**AA092**		yes	SK, Prešovský kraj, Vysoké Tatry, Kežmarská Biela voda valley, 1056 m a.s.l., roadside gravel, secondary, 49.1964N, 20.2767E, E. Záveská, K. Marhold
2x	WC	**AA116**		yes	SK, Košický kraj, Slovenský kras karst: Silická planina plateau, Gombasek, 1.8 km SSE from the village of Slavec, 400 m a.s.l., rocks beside the road, 48.569N, 20.4713E, M. Lučanová, E. Záveská, J. Smatanová
2x	WC	**AA120**		yes	SK, Prešovský kraj, Poprad: 20 km S from the town, Vernár: crossroads 4.2 km SW from the village, 901 m a.s.l., gravel along the road, 48.8853N, 20.2396E, M. Lučanová, E. Záveská, J. Smatanová
2x	WC	**AA134**		yes	SK, Banskobystrický kraj, Muráň, along tourist path from Predná Hora saddle to Muráň castle, 714 m a.s.l., eroded slope above path, 48.7725N, 20.0965E, F. Kolář
2x	WC	**AA135**		yes	SK, Banskobystrický kraj, Muráň, along tourist path from Predná Hora saddle to Muráň castle, 667 m a.s.l., shady rocks, 48.768N, 20.0804E, F. Kolář
2x	WC	**AA136**		yes	SK, Banskobystrický kraj, Tisovec, rocks at Voniaca cottage, 1017 m a.s.l., half-shaded rocks and disturbed sites in open forest, 48.7053N, 19.9592E, F. Kolář
2x	WC	**AA137**		yes	SK, Banskobystrický kraj, Muráň, eastern part of Veľká Stožka rock system, 1339 m a.s.l., half-shaded rocks, 48.7855N, 19.9771E, F. Kolář

2x	WC	**AA138**		yes	SK, Banskobystrický kraj, Jelšava, rocks at Slovenská skala (622 m) hill, 609 m a.s.l., shady rocks, 48.62N, 20.2748E, F. Kolář
2x	WC	**AA139**		yes	SK, Banskobystrický kraj, Muránska Zdychava, rocks in the village, 1034 m a.s.l., shady rocks, 48.743N, 20.1364E, F. Kolář
2x	WC	**AA141**		yes	SK, Banskobystrický kraj, Červená Skala, rocks above Hron river next to the bridge, 785 m a.s.l., half-shaded rocks, 48.8186N, 20.134E, F. Kolář
2x	WC	**AA162**	2	yes	SK, Prešovský kraj, Vysoké Tatry, Mlynická dolina Skok waterfall, 1750 m a.s.l., wet rocks, alluvial gravel, 49.1534N, 20.0458E, F. Kolář, M. Lučanová, K. Marhold, J. Smatanová, J. Bayerová
2x	WC	**AA170**	3	yes	SK, Prešovský kraj, Žiar, Tristárska dolina valley, 1380 m a.s.l., open gravely sites, screes, 49.2502N, 20.2053E, F. Kolář, M. Lučanová, K. Marhold, J. Smatanová, J. Bayerová
2x	WC	**AA173**	3	yes	SK, Prešovský kraj, Gánovce, travertine outcrop in the village, 664 m a.s.l., exposed rocks, 49.0299N, 20.3207E, F. Kolář, M. Lučanová, J. Bayerová
2x	WC	**AA175**		yes	SK, Prešovský kraj, Lysá Poľana, Biela Voda river bed, 950 m a.s.l., gravel on the river bed, 49.2633N, 20.1136E, F. Kolář
2x	WC	**AA176**	3	yes	SK, Prešovský kraj, Vysoké Tatry, Mengusovská dolina, Veľké Hincovo pleso lake, 1950 m a.s.l., open moist gravely patches, lake shore, 49.1756N, 20.0605E, F. Kolář
2x	WC	**AA182**	3	yes	SK, Žilinský kraj, Kráľova Lehota, rock above Čierny Váh river, 670 m a.s.l., exposed dry rock, 49.016N, 19.8093E, K. Marhold, J. Smatanová
2x	WC	**AA183**	4	yes	SK, Žilinský kraj, Malužiná, slope above road to the quarry, 735 m a.s.l., eroded slope, 48.9845N, 19.7573E, K. Marhold, J. Smatanová
2x	WC	**AA184**	3	yes	SK, Žilinský kraj, Liptovský Hrádok, slopes above Čierny Váh, Borová Sihoť, 624 m a.s.l., eroded slope, 49.0386N, 19.7005E, K. Marhold, J. Smatanová
2x	WC	**AA208**	3	yes	SK, Prešovský kraj, Svit, Baba hill, SE slopes, 844 m a.s.l., forest clearing in pine forest, 49.0435N, 20.1807E, J. Bayerová, J. Smatanová
2x	WC	**AA225**	3	yes	SK, Žilinský kraj, Ružomberok - Jazierce, rocks 1 km SSW of the settlement, 576 m a.s.l., rocks and sparse vegetation, 49.017836N, 19.283829E, J. Bayerová
2x	WC	**AA227**	3	yes	SK, Prešovský kraj, Vysoké Tatry, Velká Studená dolina - Prielom, along tourist path, 2070 m a.s.l., alpine scree, 49.1761N, 20.1499E, J. Bayerová
2x	WC	**AA228**	3	yes	SK, Prešovský kraj, Vysoké Tatry, north facing slope below Pod Polskym hrebeňom saddle, along tourist path, 2147 m a.s.l., rocks, 49.1738N, 20.1390E, J. Bayerová
2x	WC	**AA236**	3	yes	SK, Prešovský kraj, Lipovce, at the lowest waterfall in Kamenná Baba gorge, 1.5 km W of the village, 659 m a.s.l., shady limestone rocks, 49.0589N, 20.9306E, F. Kolář, G. Fuxová, J. Smatanová
2x	WC	**AA237**		yes	SK, Prešovský kraj, Ždiar, N-NW facing slopes of Monkova dolina, N of Hlúpy hill, 1888 m a.s.l., alpine scree, small ravines in alpine grassland, 49.2384N, 20.2192E, F. Kolář, J. Bayerová
2x	WC	**AA238**		yes	SK, Prešovský kraj, Vysoké Tatry, Kôprová dolina, waterfall 150 m N of N shore of Temnosmrečianske pleso lake, 1675 m a.s.l., rock with waterfall, 49.1960N, 20.0285E, K. Marhold, S. Španiel, J. Smatanová
2x	WC	**AA239**		yes	SK, Prešovský kraj, Vysoké Tatry, Tatranská Lomnica, top of Lomnický Štít peak, 2630 m a.s.l., paved path at the top, 49.1952N, 20.2129E, F. Kolář, J. Bayerová, G. Fuxová
2x	WC	**AA241**		yes	SK, Prešovský kraj, Branisko, Rajtopíky, slopes ca 1.5 km S of the Branisko saddle of the road no. 18, 990 m a.s.l., disturbed sites and rocks in mixed forest, 49.0032N, 20.8612E, J. Bayerová, K. Marhold, S. Španiel, J. Smatanová
2x	WC	**AA243**	4	yes	PL, Województwo Małopolskie, Babia Gora, rocky ridge along the path from saddle, ca 600 m ENE from the top, 1660 m a.s.l., rock crevices, 49.5748N, 19.5381E, F. Kolář, A. Knotek
2x	WC	**AA244**	3	yes	SK, Žilinský kraj, Šútovo, small limestone hill at E margin of the village, 480 m a.s.l., rocky and gravely sunny slope, 49.1519N, 19.0851E, F. Kolář, A. Knotek, G. Fuxová
2x	WC	**AA247**		yes	SK, Prešovský kraj, Rysy - just under the Rysy chalet, 2229 m a.s.l., rock along turistic path, 49.17465N, 20.08639E, M. Holcová
2x	WC	**AA248**	4	yes	SK, Prešovský kraj, Rysy- just under the top of the mountain, 2488 m a.s.l., top of the Rysy mountain - rock, 49.1795N, 20.0880E, M. Holcová
2x	WC	**AA249**		yes	SK, Žilinský kraj, Thermal pool in the Liptovský Ján village, 650 m a.s.l., just above the water level of the thermal swimming pool, 49.0425N, 19.6780E, M. Holcová
2x	WC	**AA308**		yes	SK, Nízké Tatry, road above Vrbicke lake, 1152 m a.s.l., ditch along the asphalt road, 48.96848N, 19.57413E, J. Bayerová
2x	WC	**AA309**		yes	SK, Západné Tatry - Teply zlab, tourist pathway through the spruce forest in the direction of Osobita, along the stream, 1149 m a.s.l., forrest, 49.258917N, 19.706397E, J. Bayerová
2x	WC	**AA311**		yes	SK, Chočské vrchy, Velký Choč along the road in the direction Lúčky, 619 m a.s.l., along the road, 49.138626N, 19.393525E, M. Holcová
2x	WC	**AA323**		yes	SK, Západné Tatry - Smutna dolina, NE slope, 1596 m a.s.l., recently accumulated scree (ca 2 years ago), 49.202365N, 19.750004E, J. Bayerová
2x	WC	**AA365**		yes	SK, 910 m a.s.l., open S-facing limestone rocks, shady N-facing limestone rocks, shady slope above road, silicate, 48.774769N, 20.338676E, F. Kolář, M. Holcová, D. Požárová

2x	WC	**AA369**		yes	SK, 689 m a.s.l., semi-shaded S facing slopes, 48.618337N, 20.780072E, F. Kolář, M Holcová, D. Požárová
2x	WC	**AA372**	5	yes	SK, Košický kraj, Prielom Hornádu river gorge from Čingov to Podlesok, 550 m a.s.l., rocks, eroded slopes in forest, 48.95523N, 20.4161E, F. Kolář, D. Požárová
2x	WC	**AA373**		yes	SK, Vernár, bottom of the rocks at the N entrance to Vernárska tiesňava gorge, 750 m a.s.l., rocks and scree, 48.93288N, 20.28592E, F. Kolář, D. Požárová
2x	WC	**AA374**	1	yes	SK, Královská Lehota, rocky slope above the road to the damm, 730 m a.s.l., rocky slope, 49.0071N, 19.91276E, F. Kolář, D. Požárová
2x	WC	**AA375**		yes	SK, Ružomberok, travertine outcrop Bukovinka S of Jazierce settlement, 600 m a.s.l., travertine rocks and boulders, 49.01242N, 19.29029E, F. Kolář, D Požárová
2x	WC	**AA407**		yes	SK, Prešovský kraj, Vernár, at sides of road near stream and wall by monument, 728 m a.s.l., rocks and scree, 48.9327489N, 20.285812989E, E. Morgan, M. Lučanová
2x	WC	**AA411**		yes	SK, Banskobystrický kraj, along the path from Hronec to Mt. Chvatimech, 710 m a.s.l., open woodland, gravel, 48.7948661N, 19.5600703E, G. Šrámková
2x	WC	**AA412**		yes	SK, Banskobystrický kraj, Jaskyňa mrtvych netopierov, 1460 m a.s.l., , 48.9258822N, 19.6376167E, G. Šrámková
2x	WC	**AA413**		yes	SK, Banskobystrický kraj, , 520 m a.s.l., , 48.774422N, 19.5845428E, G. Šrámková
2x	WC	**AA432**		yes	SK, Lúčky, travertines in the village, 612.2 m a.s.l., rocks, 49.129828N, 19.398517E, E. Morgan, K.Kubíková, F. Kolář
2x	WC	**AA433**	4	yes	SK, Harmanec, scree slope above road bend, 675.4 m a.s.l., scree, eroded slope, 48.82446N, 19.022602E, E. Morgan, K. Kubíková, F. Kolář
2x	WC	**AA437**		yes	SK, Stratená, rock above river in Stratenska tiesňava gorge, 841.5 m a.s.l., N-facing wet rock with Carex firma, 48.874552N, 20.329507E, E. Morgan, K.Kubíková, F. Kolář
2x	WC	**AA474**	1	yes	SK, Banskobystrický kraj; Vel'ká Fatra-Nízké Tatry, Horehronie, in forest above Donovaly, along red-marked trail leading to summit of Zvolen, 48.8832602N, 19.226398E, A. Guggisberg, G. Mansion
2x	WC	**AA477**	2	yes	SK, Prešovský kraj; Vysoke Tatry, Hlinska dolina, along Hlinsky potok, along blue-marked trail leading to Vysne Kôprovské sedlo, 9.17N, 44640E, R. Schmickl, A. Guggisberg, G. Mansion, Z. Kyselova
4x	ALP	AA029	2	yes	A, Steiermark, Tragöss, forest road towards Grüner see and rocks around the lake, 792 m a.s.l., forest road + small rocks, 47.5405N, 15.06044444E, F. Kolář
4x	ALP	AA042	3	yes	A, Steiermark, Öblarn, slopes of valley of Walchenbach S of the village, 713 m a.s.l., rocky valley, 47.45083333N, 14.00644444E, F. Kolář
4x	ALP	AA083	2	no	CZ, Moravskoslezský kraj, Malá Morávka, Velká Kotlina glacial cirque, 1250 m a.s.l., shady rocks, 50.05571667N, 17.23646111E, F. Kolář
4x	ALP	AA133	2	no	SLO, Ptuj, Muretinci, rock at right bank of Drava river, S of the village, S of the bridge across Drava, 255 m a.s.l., shady rocks, 46.37016667N, 15.99561111E, F. Kolář, G. Fuxová
4x	ALP	AA144	2	yes	A, Niederösterreich, Hohenberg, castle ruin, 607 m a.s.l., castle ruin, 47.90415N, 15.62319E, E. Záveská, S. Španiel
4x	ALP	AA145	3	yes	A, Steiermark, Niedere Tauern, valley and rocks along the trekking path towards Gamskögel, 1752 m a.s.l., moist valley (with Petasites sp.) and rocks, 47.37326N, 14.55397E, E. Záveská, S. Španiel
4x	ALP	AA146	2	yes	A, Kärnten, Eberstein, dolomitic rocks above the village, 585 m a.s.l., shady rocks, eroded slopes, 46.802787N, 14.553129E, F. Kolář
4x	ALP	AA147	2	yes	A, Steiermark, Schönberg-Lachtal, along a road L 514, 856 m a.s.l., gravel and stones along the road, 47.182594N, 14.337868E, F. Kolář, M. Hanzl, E. Záveská, S. Španiel
4x	ALP	AA148	3	yes	A, Steiermark, Niedere Tauern, Schießeck, N exposed rocky slopes, quartzite rocks, 2240 m a.s.l., screes and rocks, 47.277662N, 14.321904E, F. Kolář, M. Hanzl, S. Španiel
4x	ALP	AA149	3	yes	A, Steiermark, Kraubath, rocks in open forest S of the village, 628 m a.s.l., shady rocks, 47.281679N, 14.927647E, F. Kolář, M. Hanzl, E. Záveská, S. Španiel
4x	ALP	AA252	4	yes	A, Steiermark, Wölzer Tauern: Schönberg bei Niederwölz, 820 m a.s.l., marble rocks, 47.18194N, 14.33694E, P. Schönswetter
4x	ALP	AA254	1	yes	A, Steiermark, Seckauer Alpen: Hochreichart, northwestern crest, 2360 m a.s.l., stabilized amphibolite screes, 47.36444N, 14.68083E, P. Schönswetter
4x	ALP	AA255	3	yes	A, Steiermark, Seckauer Alpen: lower-most Ingeringgraben, 970 m a.s.l., siliceous rocks, 47.28417N, 14.68194E, P. Schönswetter
4x	ALP	AA265	2	no	SLO, Tolmin, Soča river valley, Kal-Koritnica, 430 m a.s.l., gravel road margin and riverine beds, natural canyon, 46.33281N, 13.59064E, P Koutecký
4x	ALP	AA300	3	yes	AT, Steiermark, Pusterwald, shores of Wildsee below Eiskarspitz, 2117 m a.s.l., snowbed in a glacial cirque, 47.3256N, 14.23038E, F. Kolář, A. Knotek, S. Španiel, P. Schönswetter, K. Hülber
4x	ALP	AA301	3	yes	AT, Steiermark, Pusterwald, ridge in the saddle between Hohenwart and Eiskarspitz, 2296 m a.s.l., N-facing stony ridge and scree, 47.3293N, 14.23066E, F. Kolář, A. Knotek, S. Španiel, P. Schönswetter, K. Hülber
4x	ALP	AA302	3	yes	AT, Steiermark, Krakaudorf, screes in Sauoffensee glacial cirque, W of the lake, 2184 m a.s.l., wet rocks and scree slope below, 47.25861N, 14.01033E, F. Kolář, P. Schönswetter, K. Hülber

4x	ALP	AA303	3	yes	AT, Steiermark, Krakaudorf, slope above Sauoffensee lake, 2030 m a.s.l., stony snowbed slope, 47.2588N, 14.00491E, A. Knotek, S. Španiel
4x	ALP	AA304	3	yes	AT, Steiermark, Aigen, rocks in Gullingtal valley, next to the bridge W of a quarry, 800 m a.s.l., rocks , 47.49355N, 14.17224E, F. Kolář, A. Knotek, S. Španiel
4x	ALP	AA315	2	no	AT, Salzburg, Distr. Sankt Johann im Pongau, steep slope of valley of Salzach River E of motorway A10, 400 m NE of Blientau, 3.2 km N–NNW of Werfen, 590 m a.s.l., scree in open forest, 47.510116N, 13.174991E, Z. Kaplan
4x	ALP	AA339	3	yes	A, Castle ruin Rabenstein cca 2 km S of Sankt Paul im Lavanttal, 620 m a.s.l., rocks and walls of castle ruins, 46.68833N, 14.87167E, A. Knotek; D. Požárová
4x	ALP	AA340	3	yes	A, Calcareous riverbank near the road from Bad Ischl to Ebensee, 460 m a.s.l., river bank, 47.74694N, 13.68972E, A. Knotek; D. Požárová
4x	ALP	AA350	3	no	D, for locality details see Arnold et al. 2015, 640m a.s.l., 48.13972N, 8.23667E, B. Arnold et al.
4x	ALP	AA397	3	yes	AT, Steiermark, Panzriedl, south edge of the serpentine ridge, 1750 m a.s.l., open forest, rocks, 47.46403N, 14.23989E, V. Konečná, G. Wos
4x	ALP	AA398	3	yes	AT, Steiermark, Vorberg, rocks next to the road, 2 km of Ritzmannsdorf, 1010 m a.s.l., rocks, 47.49876N, 14.16964E, V. Konečná, G. Wos
4x	ALP	AA414	2	yes	AT, Steiermark, Steinach-Irdning railway station, 650 m a.s.l., railway bank gravel, 47.529083N, 14.107585E, D. Požárová, V. Zeisek, G. Wos
4x	ALP	AA475	1	yes	AT, Steiermark, Bezirk Leoben, Stadt Eisenerz, for locality details see Novikova et al. 2016, 47.54N, 14.9E, N. Hohmann
4x	ALP	AA484	1	yes	AT, Styria; Nordöstliche Kalkalpen, 4 km NE Mariazell, Rechengraben, near bridge to "Wuchtelwirtin" Inn, rock on the road, 47.7876124N, 15.3516894E, R. Schmickl, G. Muir
4x	ALP	AA491	1	yes	AT, Steiermark, Bezirk Bruck an der Mur, Gemeinde Tragoss, road from Eisenerz to Tragoss, 47.5N, 15.05E, N. Hohmann
4x	CEU	AA001	2	no	CZ, Jihočeský kraj, Boršov, rocks at the left bank of Vltava river, 1 km WSW of the church in the village, 420 m a.s.l., rocks in open forest in river canyon, 48.91797222N, 14.41841667E, M. Lučanová
4x	CEU	AA032	3	yes	CZ, Středočeský kraj, Křivoklát, SE facing slope "Brdatka" above Berounka river, 2.5 km NE of the castle, 350 m a.s.l., open forest in river canyon, 50.04966667N, 13.89080556E, E. Záveská, M. Lučanová, G. Fuxová, F. Kolář
4x	CEU	AA033	2	no	A, Niederösterreich, Wachau: surroundings of homestead “Seiber”, along the road LH78 called “Seibererstraße”, district (Bezirk): Krems-Land, 620 m a.s.l., along the roadside, on rocks and rock crevices, at the forest edge, 48.40877778N, 15.43375E, C. Pachschwöll, H. P. Grohmann
4x	CEU	AA043	2	no	CZ, Středočeský kraj, Pukňov, exposed rocks along the forest track, above Vltava river, 482 m a.s.l., rocky valley, forest road margin, 49.54276111N, 14.128275E, E. Záveská
4x	CEU	AA045	2	no	CZ, Pardubický kraj, Choceň, S exposed rocks above Orlice river, 332 m a.s.l., rocks, 50.00322222N, 16.23080556E, F. Kolář, E. Záveská, J.Malinská
4x	CEU	AA046	2	no	CZ, kraj Vysočina, Ostrov u Ledče nad Sázavou, confluence of Nezdínský potok brook and Sázava river, 365 m a.s.l., railway track close to rocks, 49.68711111N, 15.30325E, F. Kolář, E. Záveská, J.Malinská
4x	CEU	AA103	2	yes	CH, Bern, Burgdorf: grassland below the castle, close to the road, 545 m a.s.l., grassland, 47.05508889N, 7.630075E, M. Lučanová
4x	CEU	AA104	3	yes	CH, Jura, Muriaux: rock under the viewpoint Les Sommetrēs 1.4 km SW of the village, 1072 m a.s.l., limestone rock, 47.23685278N, 6.967061111E, M. Lučanová
4x	CEU	AA187	2	no	D, Bayern, Hohenfels, rocky slope below the castle ruins, 400 m a.s.l., rocks and grassy terraces, 49.203502N, 11.850417E, J. Chrtek, K kabátová
4x	CEU	AA188	2	yes	D, Rheinland-Pfalz, Bacharach, forested slate slopes (oaks), along the forest pathway; rocks along the roadside, 246 m a.s.l., forested rocky slopes, 50.04973N, 7.74165E, E. Záveská
4x	CEU	AA189	2	no	L, Luxembourg, Luxembourg - city walls, in the centre of the Luxembourg city, 189 m a.s.l., stone walls, 49.60911N, 6.12819E, E. Záveská
4x	CEU	AA190	2	yes	L, Luxembourg, Lipperscheid, rocky slopes along the roadside, 177 m a.s.l., exposed rocky slopes, 49.91278N, 6.08039E, E. Záveská
4x	CEU	AA191	2	yes	B, Wallonie, Namêche, Bois de la Sarte, forested slopes, on the margins of the slate-quarry, 125 m a.s.l., forested rocky slopes, 50.49126N, 4.98465E, E. Záveská
4x	CEU	AA195	2	yes	CZ, Jihomoravský kraj, Borač - Prudká, 364 m a.s.l., rocky outcrop in river valley, 49.4206N, 16.366E, T. Urfus, L. Musilová
4x	CEU	AA203	3	yes	B, Wallonie, Chokier, rocks in old quarry above the village, 103 m a.s.l., rocky slopes, 50.592977N, 5.443828E, F. Kolář, E. Záveská, G. Fuxová, K. Marhold, R. Schmickl
4x	CEU	AA205	2	yes	D, Rheinland-Pfalz, Maria Laach, next to the road along Lacher see, 360 m a.s.l., rocky slopes above road, 50.425635N, 7.272982E, F. Kolář, E. Záveská, G. Fuxová, K. Marhold, R. Schmickl
4x	CEU	AA246	2	no	A, Oberösterreich, Mühlviertel: Kettenturm near Untermühl, Naturschutzgebiet Neuhaus; Bezirk (county): Rohrbach., steep, SW-exposed forest c. 50 m above the Danube dominated by Quercus petraea and Carpinus betulus, 335 m a.s.l., forest, 48.422717N, 13.988168E, C. Pachschwöll and E. Schandl

4x	CEU	AA285	2	yes	CZ, Jihomoravský kraj, Znojmo, left side of river Dyje canyon, cca 2 km above Znojmo dam, 275 m a.s.l., rocky slopes in loose Quercus pubescens and Pinus sylvestris forest, 48.8433611N, 16.0171667E, A. Knotek
4x	CEU	AA286	3	yes	AT, Lower Austria, Wachau, Weißenkirchen, 359 m a.s.l., oak-pine forest and former vineyard, 48.405022N, 15.472906E, R.Schmickl
4x	CEU	AA338	3	no	A, Calcareous rocks and screes near railroad; near the road cca 4 km NE from Mautstatt, calcareous rocks and bank between railroad and small river, 560 m a.s.l., calcareous rocks and bank, 47.37N, 15.38667E, A. Knotek; D. Požárová
4x	CEU	AA341	2	yes	CZ, Opárno, Old railway, m a.s.l., , 50.542778N, 14.011944E, M, Holcová
4x	CEU	AA358	3	yes	CZ, Středočeský kraj, Bernartice, east-exposed slopes above Želivka water damm, NNE of the village, open pine forest, rocks, 49.6838164N, 15.1332558E, V. Konečná, M. Holcová, F. Kolář, L. Yant
4x	CEU	AA359	3	yes	CZ, Středočeský kraj, Vlastějovice, steep slope above the road to Pertoltice, on the N border of the village,eroded slope, rocks, 49.7349631N, 15.1748464E, V. Konečná, M. Holcová, F. Kolář, L. Yant
4x	CEU	AA360	3	yes	AT, Niederösterreich, Fuglau, slopes above the road to Steinegg, above the first bend of the serpentine road, open oak forest, N, 15.5572367E, V. Konečná, M. Holcová, F. Kolář, L. Yant
4x	CEU	AA361	3	yes	AT, Niederösterreich, Steinegg, steep w-exposed slope above Kamp river, 1,2 km W of the village, 394 m a.s.l., rocks, open pine forest, 48.62993N, 15.542567E, V. Konečná, M. Holcová, F. Kolář, L. Yant
4x	CEU	AA362	2	no	AT, Niederösterreich, Wegschied a.d. Kamp, open forest N of the village,open pine forest, 48.61599N, 15.48787E, V. Konečná, M. Holcová, F. Kolář
4x	CEU	AA383	1	yes	CZ, Středočeský kraj, Unhošť, rocks around hill 750 m, N of Markův mlýn mill, 3 km WSW of the village, 369 m a.s.l., open rocks, scree in small quarries, 50.05857N, 14.10875E, F. Kolář, D. Požárová
4x	CEU	AA390	2	no	SLO, Savinjska, 46.110795N, 15.2257089E, F. Kolář, E. Morgan
4x	CEU	AA430	3	no	AT, Steiermark, serpentine rock and path, 530 m a.s.l., 47.35557N, 15.33653E, V. Konečná, M. Bohutínská, D. Bohutínský
4x	CEU	AA464	2	yes	D, Bayern, Weltenburg, rocks around the monastery and ca 500 m E of the monastery, 440 m a.s.l., limestone rocks, N and W facing, 48.8968367N, 11.8274881E, F. Kolar
4x	CEU	AA476	1	yes	AT, Hardegg, silicate, 400 m a.s.l., , 48.85166N, 15.85833E
4x	CEU	AA481	1	yes	AT, Lower Austria; Waldviertel, Kamptal, E Altenhof, at sign "Der Muhlsteinbruch von Altenhof - Kulturpark Kamptal, 48.5448994N, 15.6881331E, R. Schmickl, G. Muir
4x	CEU	AA483	1	yes	CZ, Jihomoravský kraj; Znojmo; Lubnice, E Zeletavka river, Forest slope facing river Zeletavka on east side (2021), 48.9410337N, 15.6171967E, R. Schmickl
4x	CEU	AA487	3	yes	D, Reiftal, partly sunny, S-facing rock in valley above Neidingen, calcareous, 790 m a.s.l., rock, 48.10104N, 9.049581E
4x	CEU	AA496	1	yes	D, Baden-Wurttemberg, Schwabische Alb, Wimsener Hohle, for locality details see Novikova et al. 2016, 48.25N, 9.44E, R. Schmickl
4x	CEU	AA498	3	yes	D, Grindel Steige + Upfinger Steige, shady rocks in beech forest, NE-facing, calcareous, 700 m a.s.l., shady rocks in beech forest, NE-facing, calcareous, 48.44784N, 9.422422E, F. Kolář
4x	RUD	AA028	2	yes	PL, Województvo Dolnośląskie, Kletno, limestone rock above the valley, N of jaskynia Niedzwiedzia, 890 m a.s.l., rocks and scree, 50.23861111N, 16.84255556E, F. Kolář
4x	RUD	AA030	2	yes	S, Jämtland, Rotviken, roadside of the road Ostersund-Laksjoen, 316 m a.s.l., roadside, 63.95894444N, 14.18986111E, F. Kolář
4x	RUD	AA051	2	yes	SLO, Radovljica, Via Gorizia, 1119 m a.s.l., roadside, 46.42091667N, 13.58669444E, G. Fuxová
4x	RUD	AA056	1	yes	UA, L'vivs'ka oblasť, Staryj Sambir, road bank, 358 m a.s.l., road bank, gravels, 49.46091667N, 22.98736111E, S. Španiel, P. Mereďa
4x	RUD	AA058	2	yes	UA, L'vivs'ka oblasť, Skole (NE of Kozeva, SW of Stryj), along railway, 455 m a.s.l., railway bank, gravels, 49.02633333N, 23.49883333E, S. Španiel, P. Mereďa
4x	RUD	AA061	2	no	PL, Województwo Śląskie, Katowice, railway tracks surrounded by moist forest, suburb of Katowice, common locality with A. halleri, 307 m a.s.l., railway track, 50.24313889N, 18.94513889E, E. Záveská, Z. Khodlová, P. Trávníček
4x	RUD	AA097	2	yes	CZ, Královéhradecký kraj, Teplice nad Metují, railway track bank in the town, 480 m a.s.l., railway track, 50.59457N, 16.16786E, J. Suda
4x	RUD	AA098	2	yes	CZ, Moravskoslezský kraj, Český Těšín, railway station, 300 m a.s.l., railway track, 49.742075N, 18.62208889E, F. Kolář
4x	RUD	AA099	2	yes	CZ, Moravskoslezský kraj, Valšov, railway station, 550 m a.s.l., railway track, 49.93383333N, 17.43708333E, P. Koutecký
4x	RUD	AA101	2	yes	CH, Ticino, Biasca: 5 km NW from the village, 800 m NW from Personico village , 325 m a.s.l., dry edge of the meadow, 46.37548333N, 8.909727778E, M. Lučanová
4x	RUD	AA102	2	yes	CH, Bern, Enggistein: 760 m NE from the village, 685 m a.s.l., edge of the meadow along the small stream, 46.93521111N, 1518497E, M. Lučanová

4x	RUD	AA150	2	yes	D, Sachsen-Anhalt, Elbigenrode, border of a large quarry SW of the town, 505 m a.s.l., gravely road at the border of limestone quarry, 51.753164N, 10.785386E, J. Chrtek, K. Kabátová
4x	RUD	AA152	3	yes	D, Mecklenburg-Vorpommern, Rugen, Prora, open grass, medial strip between road and cycle path (secondary habitat), close to natural sand dunes, 18 m a.s.l., grass on sandy soil , 54.42222222N, 13.58305556E, E. Záveská
4x	RUD	AA154	2	yes	D, Berlin, Berlin, Yorckstrasse, ruins of the former railway station (S-bahn) and railway tracks, 40 m a.s.l., railway track, 52.498803N, 13.373365E, E. Záveská
4x	RUD	AA180	2	yes	S, Vasternorrland, Skuleskogen NP, parksite at the N border of the park, 40 m a.s.l., gravely road margin, 63.134726N, 18.51725E, F. Kolář, M. Lučanová
4x	RUD	AA181	2	yes	N, Vågå, Jotunheimen, Gjendesheim, campsite E of the main road, 970 m a.s.l., gravel in campsite, 61.488668N, 2386344E, F. Kolář, M. Lučanová
4x	RUD	AA206	2	yes	D, Nordrhein-Westfalen, Erndtebrück, railway tracks near the village, 500 m a.s.l., gravel in the railway track, 50.977096N, 8.225071E, F. Kolář, E. Záveská, K. Marhold, R. Schmickl
4x	RUD	AA207	2	yes	D, Thüringen, Ilmenau, railway station in the town, 483 m a.s.l., gravel in the railway track, 50.683383N, 10.922354E, F. Kolář, E. Záveská, K. Marhold
4x	RUD	AA215	2	yes	RUS, Leningradskaya oblast', Kirovskii Raion, vicinity of the village of Maluksa, sandy place, close to dirt road in the forest, 65 m a.s.l., open sandy forest, 59.68867N, 31.36073E, K. Marhold, P. Efimov, A. Sennikov
4x	RUD	AA216	3	yes	RUS, Leningradskaya oblast', Gatchinskii Raion, Vyritsa, railway station, along the railway tracks, 68 m a.s.l., railway tracks, 59.41427N, 30.34568E, K. Marhold, P. Efimov, A. Sennikov
4x	RUD	AA231	2	yes	UA, Zakarpatska oblasť , Kvasy, railway station, 547 m a.s.l., railway track, 48.15797N, 24.28025E, J. Chrtek, K. Kabátová
4x	RUD	AA261	2	yes	LT, Salaspils, Railway Riga - Ogre, 300 m from station Salapsis direction Ogre, 20 m a.s.l., railway embankment, 56.8602702N, 24.3556685E, I. Rurane
4x	RUD	AA270	2	yes	S, Skåne, Örtofta, railway station, 20 m a.s.l., railway embankment, 55.781835N, 13.250972E, F. Kolář, E. Záveská
4x	RUD	AA272	2	yes	S, Kronoberg, Alvesta, railway station, 147 m a.s.l., gravel between rails, 56.899507N, 14.557429E, F. Kolář, E. Záveská
4x	RUD	AA274	1	yes	S, Blekinge, Mörrum, railway E of Kråketorp station, 23 m a.s.l., railway embankment, 56.190655N, 14.717061E, F. Kolář, E. Záveská
4x	RUD	AA277	2	yes	PL, Województwo Zachodniopomorskie, Kołobrzeg, parking place E of Grzybovo village, 17 m a.s.l., sandy sites at parking place, 54.162953N, 15.451701E, F. Kolář, E. Záveská
4x	RUD	AA278	2	yes	PL, Województwo Zachodniopomorskie, Mielno, along the road in Unieście E of the town, 8 m a.s.l., sandy sites along the road, 54.273669N, 16.104121E, F. Kolář, E. Záveská
4x	RUD	AA281	2	yes	PL, Województwo Pomorskie, Nowa Wieś Lęborska, old railway track N of the village at the crossing with road to Łeba, 62 m a.s.l., railway embankment, 54.59279N, 17.704579E, F. Kolář, E. Záveská
4x	RUD	AA282	2	yes	PL, Województwo Kujawsko-Pomorskie, Unisław, railway station, 83 m a.s.l., sandy and gravely sites around railway embankment, 53.20466N, 18.377266E, F. Kolář, E. Záveská
4x	RUD	AA283	2	yes	PL, Województwo Wielkopolskie, Chodziez, along the side roads close to the road Chodziedz - Czarnkow, 7 km WSW of the town, 100 m a.s.l., road ditch in pine forest, 52.978428N, 16.818735E, F. Kolář, E. Záveská
4x	RUD	AA287	2	yes	S, Åkersberga, along Hästängsuddsvägen road, 19 m a.s.l., gravel along a road, 59.441573N, 18.270786E, F. Kolář
4x	RUD	AA288	2	yes	PL, Województwo Malopolskie, Pustynia Błędowska near village Chechlo, view point, 340 m a.s.l., sandy dune around concrete thing, sandy places in pine forest, 50.360296N, 19.519871E, J. Bayerová
4x	RUD	AA320	2	yes	S, Västerbottens län, Dikanäs, disturbed bank of a road to Umnäs, along a bridge over a small stream N of the village, 500 m a.s.l., gravel, road bank, 65.3169N, 16.0367E, F. Kolář
4x	RUD	AA326	2	yes	LV, Koknese, 100 m from "1905. gada iela" street eastwards, 88 m a.s.l., railway embankment, 56.6500444N, 25.4440991E, I. Rurane
4x	RUD	AA345	2	yes	SLO, Tirol, Fritzens, railway station, 550 m a.s.l., railway track gravel and sand, 47.3020922N, 11.5879617E, F. Kolář
4x	RUD	AA353	3	yes	D, for locality details see Novikova et al. 2016 , 50.7414898N, 8.05339983333333E, Christian Sailer
4x	RUD	AA356	3	yes	PL, for locality details see Arnold et al. 2015, 80m a.s.l., 52.28028N, 16.70944E, B. Arnold et al.
4x	RUD	AA363	2	yes	CZ, Liberecký kraj, along the walk path and on the slopes above the pond, 375 m a.s.l., , 50.7627061N, 15.074645E, G. Šrámková
4x	RUD	AA440	2	yes	CZ, Jestřebí, railway bank between Jestřebí and Doksy, below Konvalinkový vrch hill, 270 m a.s.l., railway ballast, 50.6042969N, 14.6194675E, F. Kolář
4x	RUD	AA449	2	yes	SK, Lupkovský priesmyk, along railway close to the tunnel, railway, 49.2508467N, 22.0384328E, K. Šemberová, M. Folbrová

4x	RUD	AA452	1	no	PL, Vistula Spit, Krynica Morska: along the road and turistic paths heading towards sea coast, 10 m a.s.l., sandy pine forest near Baltic sea, 54.37969N, 19.422601E, G. Šrámková
4x	RUD	AA453	2	yes	LT, Curonian Spit, Neringa, Juodkarte: sand dunes ca 11 km N of the village, 30 m a.s.l., , 55.64261N, 21.12501E, M. Štech
4x	RUD	AA472	2	yes	N, Viken, Greaker fort, at the fortress, 50 m a.s.l., rocky scree and fortress walls, 59.271923N, 11.032923E, F. and M. and K. Kolář
4x	RUD	AA473	1	yes	D, for locality details see Novikova et al. 2016, 290 m a.s.l., , 51.308N, 8.487028E, R. Schmickl
4x	RUD	AA478	3	yes	S, Harnosand, for locality details see Novikova et al. 2016, 62.6N, 44638E, M. Nordborg
4x	RUD	AA479	3	yes	PL, for locality details see Preite et al. 2019, 50.2440833N, 16.848417E, Christian Sailer
4x	RUD	AA480	3	yes	PL, Kowary, secondary gravel, for locality details see Preite et al. 2019, 670 m a.s.l., , 50.763153N, 15.8439E
4x	RUD	AA482	1	yes	D, Mecklenburg-Vorpommern, Landkreis Rugen, Lietzow, next to the train station, for locality details see Novikova et al. 2016, 54.48N, 13.51E, C. N. Schroder
4x	RUD	AA485	3	no	PL, for locality details see Preite et l. 2019, 50.5030833N, 18.93816E, Christian Sailer
4x	RUD	AA486	1	yes	F, for locality details see Novikova et al. 2016, 62.05209N, 23.077866E, Johanna Leppala, P. Paajanen
4x	RUD	AA488	1	yes	D, Saarland, Volklingen, area of the world cultural heritage "Volklinger Hutte, for locality details see Novikova et al. 2016, 49.25N, 6.833E, R. Schmickl
4x	RUD	AA490	2	yes	PL, for locality details see Novikova et al. 2016, 53.897702N, 14.298695E, P. Baduel
4x	RUD	AA497	3	yes	D, for locality details see Arnold et al. 2015, 570m a.s.l., 47.62806N, 13.00167E, B. Arnold et al.
4x	SCA	AA065	3	yes	RO, Argeş, Făgăraş Mts, slopes above lake Balea, close to Transfăgăraş road, 2269 m a.s.l., alpine scree, 45.602N, 24.62263889E, F. Kolář, G. Fuxová
4x	SCA	AA067	3	yes	RO, Argeş, Dâmbovicioara, river canyon W of the village, 858 m a.s.l., shady rocks in a canyon, 45.44163889N, 25.22394444E, F. Kolář, G. Fuxová
4x	SCA	AA072	2	yes	RO, Bistriţa-Năsăud, Bistriţa Bârgăului, rocky slope above road from the village to the dam, 653 m a.s.l., rocks, eroded slope, 47.16969444N, 24.83372222E, F. Kolář, G. Fuxová
4x	SCA	AA074	2	yes	RO, Suceava, Rarău Mts, summit rocks, 1573 m a.s.l., exposed rocks, scree, 47.447N, 25.56175E, F. Kolář, G. Fuxová
4x	SCA	AA075	3	yes	RO, Suceava, Cârlibaba, rocky slope above the road to Borşa, W of the village, 981 m a.s.l., shady rocks, eroded slope, 47.57594444N, 25.07711111E, F. Kolář, G. Fuxová
4x	SCA	AA077	2	yes	RO, Bistriţa-Năsăud, Romuli, valley of Stramba stream E of the village, 665 m a.s.l., shady rocks, alluvial deposits along brook, 47.53655556N, 24.46319444E, F. Kolář, G. Fuxová
4x	SCA	AA078	2	yes	RO, Alba, Gârda de Sus, bottom of the canyon NE of the village, 808 m a.s.l., shady rocks in a canyon, 46.46741667N, 22.84219444E, F. Kolář, G. Fuxová
4x	SCA	AA079	2	yes	RO, Bihor, Şuncuiuş, rocky slopes above Crişul Repede river, SE of the village, 332 m a.s.l., rocks in a steppe, 46.933327N, 22.547718E, F. Kolář, G. Fuxová
4x	SCA	AA081	3	yes	RO, Cluj, Turda, Cheia, bottom of Cheile Turzii gorge, 500 m a.s.l., shady rocks in a canyon, 46.56472222N, 23.6775E, F. Kolář, G. Fuxová
4x	SCA	AA220	3	yes	RO, Bistriţa-Năsăud, Parva, slopes above the road in a deep valley of Rebra brook at N margin of the village, 545 m a.s.l., open sites above road in deciduous forest, 47.402365N, 24.546105E, F. Kolář, M. Holcová
4x	SCA	AA222	3	yes	RO, Sibiu, Făgăraș Mts., Cârțișoara, rocky slopes along Transfăgărășan road, 10 km S of the village, 1203 m a.s.l., rocks and scree, 45.643126N, 24.605534E, F. Kolář, M. Holcová
4x	SCA	AA250	2	yes	RO, Argeş, Pisicii, river canyon, 873 m a.s.l., shady rocks in a canyon, dry drain of the stream, looks primary, 45.5299N, 25.28747E, E. Záveská
4x	SCA	AA251	3	yes	RO, Argeş, Dâmbovicioara, river canyon S of the village, 915 m a.s.l., gravel along the road, seems like secondary habitat, not on the rocks only below them in the gravel, 45.42667N, 25.21327E, E. Záveská
4x	SCA	AA293	3	yes	UA, Zakarpatska oblasť, S slope, bellow the top of Bliznica, in glacial cirque, 70 - 100 vertical meters above lakes, 1614 m a.s.l., limestone in glacial cirque, 48.22862N, 24.2323E, M. Holcová, J. Hojka, D. Bohutínský, F. Rooks
4x	SCA	AA294	3	yes	RO, Maramureş, Borșa, around the entrance to glacial cirque at E slopes of Pietrosul Rodnei, along the dirt road, 5.7 km SSW of the town, 1775 m a.s.l., rocky slope with sparse vegetation, 47.60311N, 24.648798E, M. Holcová, J. Hojka, D. Bohutínský, F. Rooks
4x	SCA	AA295	3	yes	UA, Zakarpatska oblasť, in glacial cirque above lake, 1596 m a.s.l., glacial cirque, 48.27113N, 24.16347E, M. Holcová, J. Hojka, D. Bohutínský, F. Rooks
4x	SCA	AA296	3	yes	RO, Maramureş, Borșa, in the left side of glacial cirque at pietrosul mountain, 5.7 km SSW of the town, 1775 m a.s.l., glacial cirque, 47.603123N, 24.648798E, M. Holcová, J. Hojka, D. Bohutínský, F. Rooks
4x	SCA	AA299	3	yes	RO, Maramureş, NW slope of the NW glacial cirque below Ineu peak, 2017 m a.s.l., glacial cirque, 47.52734N, 24.88062E, M. Holcová, J. Hojka, D. Bohutínský, F. Rooks

4x	SCA	AA330	3	yes	RO, Braşov, Timişu de Sus, rocks above railway, ca 5 km NNE of the village, 796.6 m a.s.l., shady rocks, travertine spring, 45.569999N, 25.608265E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
4x	SCA	AA332	3	yes	RO, Argeş, Dâmbovicioara, cabana Brusturet, rocks in gorges and hills below and above the cabin, 1013.8 m a.s.l., rocks, mostly in gorges, 45.467007N, 25.227617E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
4x	SCA	AA335	4	yes	RO, Braşov, Zărneşti, narrowest part of a gorge Prăpăstiile Zărneştilor, 5.5 km SW of the railway station in the town, 978.4 m a.s.l., rocks in the gorge, N, 25.281208E, F. Kolář, M. Holcová, D. Požárová, F. Rooks
4x	SCA	AA419	2	yes	RO, Bistriţa-Năsăud, Both sides of road in valley, and continues further along into quarry-area, 760 m a.s.l., Scree slopes down to stream, rocky slopes and old quarry, 47.2581N, 24.74585E, E. Morgan, M. Lučanová
4x	SCA	AA447	2	yes	RO, 706.8 m a.s.l., , 45.58573N, 25.48632E, F. Kolář, D. Požárová, M. Bohutínská, D. Bohutínský
4x	SCA	AA463	2	no	RO, Parang Mts., Parang Mts., northern slope of peak Găuri, about two hours of hiking away from the road, 2092 m a.s.l., , 45.363007N, 23.588303E, Dana Suteu
4x	WCA	AA002	2	yes	CZ, Zlínský kraj, Břestek, rocks 700 m E of Barborka hill, E of Buchlov castle, 350 m a.s.l., along sedimentary rocks, 49.10938889N, 17.32858333E, F. Kolář
4x	WCA	AA005	2	no	SK, Trnavský kraj, Buková, slope above the road at the E end of the village, 287 m a.s.l., rocks in open pine forest, 48.54255556N, 17.40652778E, E. Záveská, J. Kučera, F. Kolář
4x	WCA	AA008	2	yes	SK, Nitriansky kraj, Podhradie, rocks along the road to Topoľčiansky hrad castle, 465 m a.s.l., crevices, screes in limestone rock, 48.65808333N, 18.05122222E, E. Záveská, J. Kučera, F. Kolář
4x	WCA	AA010	2	yes	SK, Banskobystrický kraj, Banská Štiavnica, bend of the road B. Štiavnica - Počúvadlianske jazero, next to Banské múzeum, 682 m a.s.l., slopes in deciduous forest, 48.45030556N, 18.88813889E, E. Záveská, J. Kučera, F. Kolář
4x	WCA	AA019	2	yes	SK, Prešovský kraj, Humenné, limestone rocks at the E slopes of Sokol hill, 3 km SE of the railway station, 251 m a.s.l., crevices in limestone rock, 48.91175N, 21.92844444E, E. Záveská, F. Kolář
4x	WCA	AA024	2	yes	SK, Žilinský kraj, Terchová, bank of the road 0.5 km W of Štefanová, 601 m a.s.l., limestone road bank, 49.23261111N, 19.05541667E, E. Záveská, F. Kolář
4x	WCA	AA026	2	yes	SK, Žilinský kraj, Omšenie, calcareous rocks at the top of Omšenská Baba (668 m), N of the village, 668 m a.s.l., rocks, 48.91152778N, 18.23677778E, F. Kolář
4x	WCA	AA027	2	yes	SK, Žilinský kraj, Súľov, slopes below the castle ruins along the path to Lúka pod Hradom, 2 km J of Súľov, 600 m a.s.l., forest on rocky slope, 49.17551944N, 18.58361944E, F. Kolář
4x	WCA	AA059	2	yes	PL, Województwo Małopolskie, Ojcow, National park near the village Ojcow, steep rocky slopes, 348 m a.s.l., rocks, 50.22447222N, 19.82922222E, E. Záveská, Z. Khodlová, P. Trávníček
4x	WCA	AA087	1	yes	SK, Žilinský kraj, Zuberec, Zverovka, cliffs on the ridge between Plačlivý Roháč and Smutné sedlo saddle, 2031 m a.s.l., exposed rocks, 49.19702778N, 19.74477778E, F. Kolář, E. Záveská, S. Španiel
4x	WCA	AA095	1	yes	SK, Žilinský kraj, Nízke Tatry, W facing slope of Ludárova hoľa mountain, N of the peak of Ďumbier mountain , 1700 m a.s.l., old mine tailings and scree, 48.943475N, 19.63951389E, M. Kolník, K. Marhold
4x	WCA	AA096	4	yes	SK, Žilinský kraj, Nízke Tatry, Liptovský Ján, Jánska dolina valley, S of the village, 744 m a.s.l., alluvial gravel, eroded slope above road, 49.01075N, 19.67327778E, F. Kolář, S. Španiel, M. Kolník, K. Marhold
4x	WCA	AA114	2	yes	HU, Heves megye, Bükk Mts, 3.8 km SE Szilvásvárad, below Gerennavár, 675 m a.s.l., sidehill in oak forest, 48.08972806N, 20.43356889E, M. Lučanová, E. Záveská, J. Smatanová
4x	WCA	AA115	2	yes	HU, Borsod-Abaúj-Zemplén, Bükk Mts, rocks beside the road at the exit of Lillafüred spa for the village of Újmassa, 309 m a.s.l., scree below the former quarry, 48.10605N, 20.6203E, M. Lučanová, E. Záveská, J. Smatanová
4x	WCA	AA119	3	yes	SK, Košický kraj, Dreveník hill, 1 km NW from the village of Žehra, 552 m a.s.l., rock on the edge of pine forest and dry grassland, 48.98296667N, 20.77745E, M. Lučanová, E. Záveská, J. Smatanová
4x	WCA	AA142	2	yes	SK, Banskobystrický kraj, Muráň, Hrdzavá Dolina, bottom of a scree with azonal mire, 772 m a.s.l., bottom of a scree with dealpine species, 48.748136N, 19.996945E, F. Kolář
4x	WCA	AA164	2	yes	PL, Województwo Małopolskie, Zakopane, rocky S slopes of Mt. Giewont along blue and red tourist paths, 1838 m a.s.l., exposed rocks, screes and eroded slopes, 49.250383N, 19.934022E, F. Kolář, J. Smatanová, J. Bayerová
4x	WCA	AA165	1	yes	PL, Województwo Małopolskie, Zakopane, snowbed on NE slope of Kopa Kondraczka mountain, 1953 m a.s.l., screes in a snowbed, 49.236847N, 19.931016E, F. Kolář, J. Smatanová, J. Bayerová
4x	WCA	AA167	3	yes	SK, Žilinský kraj, Pribylina, Račkova dolina valley and Račkove plesá lakes, 1690 m a.s.l., open places in grassland disturbed by marmots, 49.200048N, 19.804658E, M. Lučanová, K. Marhold

4x	WCA	AA168	3	yes	SK, Žilinský kraj, Zuberec, Zverovka, rocky slopes above Horné Roháčske pleso and on N slopes of Tri Kopy mountain, 1783 m a.s.l., wet rocks, screes, 49.204509N, 19.735202E, F. Kolář, K. Marhold
4x	WCA	AA169	3	yes	SK, Žilinský kraj, Vysoké Tatry, Baníkovské sedlo saddle: northern rocky slope under the Baníkovské sedlo saddle, 1864 m a.s.l., wet rocks, 49.201166N, 19.708372E, M. Lučanová, J. Smatanová, J. Bayerová
4x	WCA	AA170	3	yes	SK, Prešovský kraj, Žiar, Tristárska dolina valley, 1380 m a.s.l., open gravely sites, screes, 49.250216N, 20.205255E, F. Kolář, M. Lučanová, K. Marhold, J. Smatanová, J. Bayerová
4x	WCA	AA171	3	yes	SK, Prešovský kraj, Hranovnica, slopes above the road to Poprad, N of the village, 720 m a.s.l., rocky eroded slope above the road and open oak forest, 49.00716N, 20.286407E, F. Kolář, M. Lučanová, J. Bayerová
4x	WCA	AA172	3	yes	SK, Prešovský kraj, Primovce, Primovské skaly rocks in the village, 605 m a.s.l., shady rocks, 49.015874N, 20.382482E, F. Kolář, M. Lučanová, J. Bayerová
4x	WCA	AA178	3	yes	SK, Žilinský kraj, Zuberec, rocks at the top ridge between Biela Skala and Sivý Vrch, 1700 m a.s.l., exposed rocks and gravel, 49.2124N, 19.63682E, F. Kolář
4x	WCA	AA192	2	yes	PL, Małopolska, Kobylany, rocks in the gorge N of the village, 330 m a.s.l., rocks and gravely slope, 50.156219N, 19.755231E, F. Kolář, E. Záveská, G. Fuxová
4x	WCA	AA225	3	yes	RO, Argeș, Făgăraș Mts., slopes above path from second bend of the Transfăgărășan road S of the tunnel to Lacul Capra, 2092 m a.s.l., calcareous rocks and scree, 45.59535N, 24.63458E, F. Kolář, M. Holcová
4x	WCA	AA226	3	yes	SK, Žilinský kraj, Prosiek, transect through Prosiecka dolina N of the village, 656 m a.s.l., semi-shady rocks in a gorge, 49.160034N, 19.496156E, J. Bayerová
4x	WCA	AA229	4	yes	SK, Žilinský kraj, Kvačany, rocks close to southern end of Kvačianska dolina, N of the village, 673 m a.s.l., shady rocks, disturbed sites along forest road, 49.183171N, 19.541024E, J. Bayerová
4x	WCA	AA234	4	yes	SK, Prešovský kraj, Lesnica, N facing slopes of the Dunajec river canyon 1.5 km NW of the village, 437 m a.s.l., shady rocks in beech forest, 49.411733N, 20.448837E, F. Kolář, G. Fuxová, J. Smatanová
4x	WCA	AA235	3	yes	SK, Prešovský kraj, Kamenica, rocky outcrops ("bradla") 3 km NW of the village, 633 m a.s.l., rocks and scree, 49.210747N, 20.928184E, F. Kolář, G. Fuxová, J. Smatanová
4x	WCA	AA242	3	yes	SK, Košiský kraj, Spišská Nová Ves - Čingov, N facing slopes above Hornád river, 500 m WSW of the settlement, 530 m a.s.l., shady north-facing rock, 48.940175N, 20.477409E, J. Bayerová, K. Marhold, S. Španiel, J. Smatanová
4x	WCA	AA310	3	yes	SK, Trenčiansky kraj, Trenčín, Trenčiansky hrad castle, 226 m a.s.l., rocks in castle, 48.894439N, 18.04673E, F. Kolář
4x	WCA	AA321	4	yes	SK, Nízké Tatry - below Ďumbier peak, N slope, 1893 m a.s.l., narrow north-facing gorge, 48.93825N, 19.632066E, J. Bayerová
4x	WCA	AA322	3	yes	SK, Nízké Tatry - between Krakova hoľa and Pustý hill, NE slope, 1472 m a.s.l., stony slope at forest clearing, 48.991868N, 19.623027E, J. Bayerová
4x	WCA	AA324	1	yes	SK, Dolný Liptov, Chočské vrchy - Veĺký Choč along the E and SE path, 1259 m a.s.l., forest and dwarf pine below the peak, along path, 49.148497N, 19.348828E, M. Holcová
4x	WCA	AA357	2	yes	CZ, Olomoucký kraj, ruins of Svrčov castle, next Hranická propast doline, rocks above the river, 295 m a.s.l., výslunné skály nad řekou, jihozápadně orientovaný svah, 49.5372064N, 17.7473575E, M. Holcová
4x	WCA	AA366	2	yes	SK, 375m a.s.l., shady and semi-open limestone rocks, 48.585645N, 20.484106E, F. Kolář, D. Požárová
4x	WCA	AA370	2	yes	SK, 500m a.s.l., shady limestone rocks, 48.63584N, 20.822238E, F. Kolář, D. Požárová
4x	WCA	AA418	3	yes	PL, Zakopane, 915 m a.s.l., , 49.278343N, 19.96706E, C. Sailer
4x	WCA	AA435	2	yes	SK, Čičmany, along a side forest road, 626.7m a.s.l., small rocks and slope above forest road, 48.96722N, 18.523201E, E. Morgan, K.Kubíková, F. Kolář
4x	WCA	AA438	2	yes	SK, Poráč, screes above newly built relax-centre, 610.7 m a.s.l., screes, 48.879229N, 20.743836E, E. Morgan, K.Kubíková, F. Kolář
4x	WCA	AA495	1	yes	SK, Vtacnik – Dererov Mlyn, for locality details see Novikova et al. 2016, 48.65N, 18.74E, R. Schmickl, G. Muir

Table S2. Population code, number of genotyped samples per population (sample size), ploidy level, methodological approach used to generate the data, lineage assignment and admixture proportions obtained by STRUCTURE are given by each tetraploid population included in our genomic analyses. Populations included in the niche analyses are also indicated.

					STRUCTURE ADMIXTURE PROPORTIONS					
Pop code	Sample size	Ploidy	Dataset	Lineage Assignment	K1-SEC	K2-CEU	K3-WCA	K4-RUD	K5-ALP	Niche analyses
AA001	**2**	**4x**	RADseq	**CEU**	**0.007**	**0.345**	**0.283**	**0.001**	**0.363**	**no**
AA002	2	4x	RADseq	WCA	0.112	0.145	0.595	0.038	0.111	yes
AA005	**2**	**4x**	**RADseq**	**WCA**	**0.205**	**0.204**	**0.394**	**0.007**	**0.190**	**no**
AA008	2	4x	RADseq	WCA	0.193	0.131	0.638	0.002	0.036	yes
AA010	2	4x	RADseq	WCA	0.173	0.103	0.685	0.034	0.005	yes
AA019	2	4x	RADseq	WCA	0.176	0.001	0.638	0.175	0.010	yes
AA024	2	4x	RADseq	WCA	0.014	0.002	0.946	0.017	0.020	yes
AA026	2	4x	RADseq	WCA	0.054	0.081	0.801	0.004	0.059	yes
AA027	2	4x	RADseq	WCA	0.074	0.003	0.898	0.005	0.020	yes
AA028	2	4x	RADseq	RUD	0.002	0.001	0.092	0.904	0.001	yes
AA029	2	4x	RADseq	ALP	0.021	0.025	0.158	0.001	0.795	yes
AA030	2	4x	RADseq	RUD	0.000	0.000	0.000	0.999	0.000	yes
AA032	3	4x	WGS	CEU	0.039	0.649	0.094	0.005	0.213	yes
AA033	**2**	**4x**	**RADseq**	**CEU**	**0.024**	**0.389**	**0.200**	**0.013**	**0.373**	**no**
AA042	3	4x	RADseq	ALP	0.003	0.005	0.049	0.002	0.941	yes
AA043	**2**	**4x**	**RADseq**	**CEU**	**0.077**	**0.384**	**0.157**	**0.016**	**0.366**	**no**
AA045	**2**	**4x**	**RADseq**	**CEU**	**0.127**	**0.398**	**0.077**	**0.095**	**0.302**	**no**
AA046	**2**	**4x**	**RADseq**	**CEU**	**0.132**	**0.411**	**0.174**	**0.017**	**0.267**	**no**
AA051	2	4x	RADseq	RUD	0.004	0.001	0.012	0.971	0.013	yes
AA056	1	4x	RADseq	RUD	0.011	0.004	0.172	0.809	0.005	yes
AA058	2	4x	RADseq	RUD	0.003	0.001	0.160	0.834	0.001	yes
AA059	2	4x	RADseq	WCA	0.012	0.076	0.697	0.055	0.160	yes
AA061	**2**	**4x**	**RADseq**	**RUD**	**0.047**	**0.076**	**0.276**	**0.485**	**0.115**	**no**
AA065	3	4x	WGS	SEC	0.594	0.278	0.126	0.000	0.002	yes
AA067	3	4x	WGS	SEC	0.787	0.211	0.001	0.001	0.000	yes
AA072	2	4x	RADseq	SEC	0.982	0.004	0.002	0.011	0.002	yes
AA074	2	4x	RADseq	SEC	0.962	0.001	0.015	0.021	0.001	yes
AA075	3	4x	WGS	SEC	0.783	0.161	0.000	0.055	0.000	yes
AA077	2	4x	RADseq	SEC	0.939	0.006	0.002	0.051	0.001	yes
AA078	2	4x	RADseq	SEC	0.529	0.097	0.308	0.001	0.065	yes
AA079	2	4x	RADseq	SEC	0.571	0.083	0.276	0.046	0.024	yes
AA081	3	4x	WGS	SEC	0.566	0.365	0.055	0.010	0.004	yes
AA083	**2**	**4x**	**RADseq**	**ALP**	**0.111**	**0.275**	**0.136**	**0.091**	**0.387**	**no**
AA087	1	4x	WGS	WCA	0.000	0.197	0.799	0.000	0.002	yes
AA095	1	4x	WGS	WCA	0.001	0.263	0.732	0.000	0.004	yes
AA096	4	4x	RADseq	WCA	0.014	0.004	0.976	0.002	0.005	yes
AA097	2	4x	RADseq	RUD	0.009	0.001	0.111	0.868	0.011	yes
AA098	2	4x	RADseq	RUD	0.015	0.001	0.216	0.752	0.016	yes

AA099	2	4x	RADseq	RUD	0.004	0.002	0.141	0.824	0.029	yes
AA101	2	4x	RADseq	RUD	0.009	0.009	0.192	0.784	0.006	yes
AA102	2	4x	RADseq	RUD	0.028	0.014	0.204	0.749	0.005	yes
AA103	2	4x	RADseq	CEU	0.071	0.557	0.185	0.001	0.187	yes
AA104	3	4x	RADseq	CEU	0.075	0.612	0.135	0.003	0.176	yes
AA114	2	4x	RADseq	WCA	0.246	0.036	0.581	0.083	0.054	yes
AA115	2	4x	RADseq	WCA	0.222	0.115	0.606	0.049	0.008	yes
AA119	3	4x	WGS	WCA	0.001	0.239	0.758	0.001	0.001	yes
AA133	**2**	**4x**	**RADseq**	**ALP**	**0.153**	**0.216**	**0.173**	**0.023**	**0.434**	**no**
AA142	2	4x	RADseq	WCA	0.029	0.003	0.833	0.130	0.006	yes
AA144	2	4x	RADseq	ALP	0.036	0.184	0.000	0.126	0.654	yes
AA145	3	4x	RADseq	ALP	0.007	0.004	0.134	0.001	0.854	yes
AA146	2	4x	RADseq	ALP	0.002	0.007	0.217	0.001	0.773	yes
AA147	2	4x	RADseq	ALP	0.001	0.000	0.017	0.000	0.981	yes
AA148	3	4x	WGS	ALP	0.001	0.192	0.001	0.000	0.806	yes
AA149	3	4x	WGS	ALP	0.014	0.406	0.001	0.001	0.578	yes
AA150	2	4x	RADseq	RUD	0.005	0.013	0.157	0.822	0.004	yes
AA152	3	4x	RADseq	RUD	0.003	0.001	0.168	0.826	0.002	yes
AA154	2	4x	RADseq	RUD	0.003	0.001	0.098	0.896	0.002	yes
AA164	2	4x	RADseq	WCA	0.001	0.001	0.995	0.001	0.003	yes
AA165	1	4x	RADseq	WCA	0.002	0.001	0.992	0.003	0.002	yes
AA167	3	4x	RADseq	WCA	0.002	0.001	0.993	0.001	0.003	yes
AA168	3	4x	WGS	WCA	0.000	0.212	0.787	0.000	0.001	yes
AA169	3	4x	RADseq	WCA	0.001	0.000	0.994	0.000	0.004	yes
AA170	3	4x	WGS	WCA	0.000	0.202	0.797	0.001	0.000	yes
AA171	3	4x	WGS	WCA	0.001	0.230	0.759	0.009	0.001	yes
AA172	3	4x	RADseq	WCA	0.000	0.000	0.998	0.001	0.000	yes
AA178	3	4x	RADseq	WCA	0.002	0.004	0.988	0.002	0.004	yes
AA180	2	4x	RADseq	RUD	0.002	0.000	0.075	0.922	0.001	yes
AA181	2	4x	RADseq	RUD	0.000	0.000	0.001	0.998	0.000	yes
AA187	**2**	**4x**	**RADseq**	**ALP**	**0.004**	**0.327**	**0.247**	**0.016**	**0.406**	**no**
AA188	2	4x	RADseq	CEU	0.122	0.536	0.211	0.001	0.130	yes
AA189	**2**	**4x**	**RADseq**	**CEU**	**0.058**	**0.285**	**0.158**	**0.440**	**0.060**	**no**
AA190	2	4x	RADseq	CEU	0.148	0.563	0.201	0.001	0.087	yes
AA191	2	4x	RADseq	CEU	0.095	0.596	0.172	0.003	0.134	yes
AA192	2	4x	RADseq	WCA	0.036	0.015	0.778	0.056	0.115	yes
AA195	2	4x	RADseq	CEU	0.130	0.521	0.018	0.069	0.262	yes
AA203	3	4x	WGS	CEU	0.029	0.842	0.013	0.006	0.110	yes
AA205	2	4x	RADseq	CEU	0.119	0.534	0.087	0.040	0.220	yes
AA206	2	4x	RADseq	RUD	0.002	0.002	0.098	0.898	0.002	yes
AA207	2	4x	RADseq	RUD	0.071	0.007	0.123	0.799	0.001	yes
AA215	2	4x	RADseq	RUD	0.003	0.001	0.053	0.942	0.001	yes

AA216	3	4x	WGS	RUD	0.001	0.210	0.009	0.78	0.000	yes
AA220	3	4x	WGS	SEC	0.753	0.180	0.000	0.066	0.000	yes
AA222	3	4x	WGS	SEC	0.616	0.275	0.108	0.000	0.001	yes
AA225	3	4x	RADseq	WCA	0.022	0.003	0.971	0.002	0.002	yes
AA226	3	4x	RADseq	WCA	0.002	0.001	0.993	0.001	0.003	yes
AA229	4	4x	RADseq	WCA	0.009	0.002	0.970	0.003	0.015	yes
AA231	2	4x	RADseq	RUD	0.011	0.001	0.159	0.824	0.004	yes
AA234	4	4x	RADseq	WCA	0.002	0.002	0.986	0.006	0.004	yes
AA235	3	4x	WGS	WCA	0.000	0.228	0.770	0.001	0.001	yes
AA242	3	4x	RADseq	WCA	0.014	0.001	0.970	0.003	0.012	yes
AA246	**2**	**4x**	**RADseq**	**CEU**	**0.035**	**0.377**	**0.205**	**0.001**	**0.381**	**no**
AA250	2	4x	RADseq	SEC	0.974	0.001	0.021	0.003	0.001	yes
AA251	3	4x	RADseq	SEC	0.879	0.001	0.117	0.001	0.002	yes
AA252	4	4x	RADseq	ALP	0.001	0.000	0.001	0.000	0.998	yes
AA254	1	4x	WGS	ALP	0.001	0.259	0.002	0.000	0.738	yes
AA255	3	4x	WGS	ALP	0.001	0.254	0.002	0.000	0.744	yes
AA261	2	4x	RADseq	RUD	0.032	0.002	0.124	0.839	0.003	yes
AA265	**2**	**4x**	**RADseq**	**ALP**	**0.02**	**0.078**	**0.007**	**0.452**	**0.443**	**no**
AA270	2	4x	RADseq	RUD	0.000	0.000	0.000	0.999	0.000	yes
AA272	2	4x	RADseq	RUD	0.000	0.000	0.000	0.999	0.000	yes
AA274	1	4x	RADseq	RUD	0.000	0.000	0.000	0.999	0.000	yes
AA277	2	4x	RADseq	RUD	0.032	0.004	0.132	0.830	0.001	yes
AA278	2	4x	RADseq	RUD	0.001	0.001	0.079	0.917	0.001	yes
AA281	2	4x	RADseq	RUD	0.010	0.002	0.102	0.883	0.004	yes
AA282	2	4x	RADseq	RUD	0.049	0.002	0.225	0.712	0.011	yes
AA283	2	4x	RADseq	RUD	0.018	0.002	0.135	0.842	0.003	yes
AA285	2	4x	RADseq	CEU	0.052	0.581	0.146	0.002	0.219	yes
AA286	3	4x	WGS	CEU	0.013	0.670	0.079	0.002	0.236	yes
AA287	2	4x	RADseq	RUD	0.000	0.000	0.001	0.998	0.000	yes
AA288	2	4x	RADseq	RUD	0.044	0.058	0.275	0.537	0.086	yes
AA293	3	4x	RADseq	SEC	0.948	0.002	0.035	0.001	0.013	yes
AA294	3	4x	RADseq	SEC	0.968	0.001	0.003	0.015	0.012	yes
AA295	3	4x	RADseq	SEC	0.924	0.002	0.052	0.001	0.021	yes
AA296	3	4x	RADseq	SEC	0.964	0.001	0.015	0.018	0.002	yes
AA299	3	4x	WGS	SEC	0.828	0.166	0.000	0.005	0.000	yes
AA300	3	4x	WGS	ALP	0.000	0.199	0.001	0.000	0.800	yes
AA301	3	4x	RADseq	ALP	0.001	0.001	0.063	0.000	0.935	yes
AA302	3	4x	RADseq	ALP	0.001	0.001	0.018	0.000	0.981	yes
AA303	3	4x	RADseq	ALP	0.006	0.003	0.045	0.001	0.946	yes
AA304	3	4x	RADseq	ALP	0.005	0.004	0.065	0.000	0.925	yes
AA310	3	4x	WGS	WCA	0.002	0.329	0.664	0.002	0.003	yes
AA315	**2**	**4x**	**RADseq**	**ALP**	**0.017**	**0.095**	**0.135**	**0.314**	**0.439**	**no**

AA320	2	4x	RADseq	RUD	0.000	0.000	0.001	0.998	0.001	yes
AA321	4	4x	RADseq	WCA	0.009	0.049	0.934	0.001	0.007	yes
AA322	3	4x	RADseq	WCA	0.010	0.004	0.969	0.001	0.016	yes
AA324	1	4x	RADseq	WCA	0.002	0.002	0.982	0.002	0.011	yes
AA326	2	4x	RADseq	RUD	0.034	0.003	0.101	0.858	0.004	yes
AA330	3	4x	WGS	SEC	0.760	0.237	0.002	0.000	0.001	yes
AA332	3	4x	WGS	SEC	0.815	0.183	0.001	0.000	0.000	yes
AA335	4	4x	RADseq	SEC	0.985	0.002	0.010	0.001	0.001	yes
AA338	**3**	**4x**	**WGS**	**CEU**	**0.091**	**0.473**	**0.022**	**0.002**	**0.411**	**no**
AA339	3	4x	WGS	ALP	0.009	0.367	0.003	0.002	0.620	yes
AA340	3	4x	WGS	ALP	0.001	0.302	0.002	0.001	0.695	yes
AA341	2	4x	RADseq	CEU	0.057	0.551	0.005	0.071	0.315	yes
AA345	2	4x	RADseq	RUD	0.001	0.001	0.072	0.925	0.001	yes
AA350	**3**	**4x**	**WGS**	**ALP**	**0.002**	**0.361**	**0.026**	**0.196**	**0.416**	**no**
AA353	3	4x	WGS	RUD	0.003	0.215	0.007	0.775	0.001	yes
AA356	3	4x	WGS	RUD	0.001	0.203	0.004	0.789	0.002	yes
AA357	2	4x	RADseq	WCA	0.042	0.147	0.633	0.052	0.126	yes
AA358	3	4x	WGS	CEU	0.030	0.697	0.032	0.001	0.240	yes
AA359	3	4x	WGS	CEU	0.058	0.700	0.012	0.011	0.218	yes
AA360	3	4x	WGS	CEU	0.010	0.621	0.120	0.002	0.247	yes
AA361	3	4x	WGS	CEU	0.003	0.614	0.104	0.004	0.275	yes
AA362	**2**	**4x**	**RADseq**	**CEU**	**0.018**	**0.478**	**0.153**	**0.021**	**0.331**	**no**
AA363	2	4x	RADseq	RUD	0.011	0.070	0.057	0.860	0.002	yes
AA366	2	4x	RADseq	WCA	0.203	0.104	0.529	0.163	0.001	yes
AA370	2	4x	RADseq	WCA	0.116	0.005	0.853	0.017	0.009	yes
AA372	3	4x	WGS	WCA	0.001	0.232	0.765	0.001	0.001	yes
AA383	1	4x	RADseq	CEU	0.055	0.503	0.206	0.004	0.232	yes
AA390	**2**	**4x**	**RADseq**	**CEU**	**0.301**	**0.411**	**0.194**	**0.007**	**0.087**	**no**
AA397	3	4x	WGS	ALP	0.000	0.233	0.001	0.000	0.766	yes
AA398	3	4x	WGS	ALP	0.000	0.219	0.001	0.000	0.780	yes
AA414	2	4x	RADseq	ALP	0.011	0.065	0.072	0.075	0.777	yes
AA418	3	4x	WGS	WCA	0.000	0.224	0.774	0.000	0.001	yes
AA419	2	4x	RADseq	SEC	0.912	0.018	0.002	0.066	0.002	yes
AA427	3	4x	WGS	CEU	0.249	0.647	0.080	0.017	0.007	yes
AA430	**3**	**4x**	**WGS**	**CEU**	**0.050**	**0.488**	**0.012**	**0.002**	**0.448**	**no**
AA435	2	4x	RADseq	WCA	0.088	0.051	0.819	0.008	0.033	yes
AA438	2	4x	RADseq	WCA	0.016	0.055	0.923	0.005	0.001	yes
AA440	2	4x	RADseq	RUD	0.004	0.044	0.059	0.884	0.009	yes
AA447	2	4x	RADseq	SEC	0.923	0.013	0.051	0.002	0.010	yes
AA449	2	4x	RADseq	RUD	0.014	0.002	0.143	0.836	0.006	yes
AA452	**1**	**4x**	**RADseq**	**RUD**	**0.058**	**0.054**	**0.311**	**0.493**	**0.083**	**no**
AA453	2	4x	RADseq	RUD	0.091	0.003	0.339	0.539	0.029	yes

AA463	**2**	**4x**	**RADseq**	**SEC**	**0.494**	**0.097**	**0.386**	**0.001**	**0.023**	**no**
AA464	2	4x	RADseq	CEU	0.041	0.516	0.092	0.003	0.348	yes
AA472	2	4x	RADseq	RUD	0.008	0.003	0.036	0.946	0.005	yes
AA473	1	4x	WGS	RUD	0.014	0.321	0.011	0.642	0.012	yes
AA475	1	4x	WGS	ALP	0.000	0.299	0.012	0.000	0.688	yes
AA476	1	4x	WGS	CEU	0.005	0.673	0.043	0.033	0.247	yes
AA478	3	4x	WGS	RUD	0.000	0.121	0.000	0.878	0.000	yes
AA479	3	4x	WGS	RUD	0.001	0.172	0.001	0.826	0.000	yes
AA480	3	4x	WGS	RUD	0.001	0.203	0.003	0.792	0.001	yes
AA481	1	4x	WGS	CEU	0.007	0.598	0.189	0.000	0.206	yes
AA482	1	4x	WGS	RUD	0.001	0.213	0.006	0.779	0.000	yes
AA483	1	4x	WGS	CEU	0.004	0.713	0.026	0.007	0.251	yes
AA484	1	4x	WGS	ALP	0.001	0.350	0.003	0.003	0.643	yes
AA485	**3**	**4x**	**WGS**	**RUD**	**0.002**	**0.342**	**0.168**	**0.391**	**0.098**	**no**
AA486	1	4x	WGS	RUD	0.001	0.167	0.001	0.830	0.001	yes
AA487	3	4x	WGS	CEU	0.009	0.882	0.002	0.000	0.107	yes
AA488	1	4x	WGS	RUD	0.001	0.186	0.013	0.798	0.002	yes
AA490	2	4x	WGS	RUD	0.002	0.193	0.010	0.795	0.001	yes
AA491	1	4x	WGS	ALP	0.001	0.308	0.004	0.000	0.688	yes
AA495	1	4x	WGS	WCA	0.018	0.337	0.631	0.002	0.012	yes
AA496	1	4x	WGS	CEU	0.011	0.830	0.005	0.001	0.153	yes
AA497	3	4x	WGS	RUD	0.002	0.227	0.008	0.762	0.001	yes
AA498	3	4x	WGS	CEU	0.002	0.895	0.003	0.000	0.100	yes

Table S3. Number of occurrences per cytotype and per lineage included in the niche quantification analyses before and after the filtering (closer than 10-km distance were removed).

Cytotype	No.occ.	Stratified no.occ
2x	188	113
4x	164	132
Total	352	245
Lineage	**No.occ.**	**Stratified no.occ**
BAL.2x	22	12
RUD.4x	50	48
SCA.2x	48	34
SCA.4x	23	16
WCA.2x	69	29
WCA.4x	42	29
ALP.4x	24	16
CEU.4x	24	22
DIN.2x	32	24
PAN.2x	17	15

Table S4. Description of environmental variables extracted from WorldClim and their contribution to the two first axis of the PCA-env.

WorldClim Variables	Description	PC1	PC2	%PC1	%PC2
BIO1	Annual Mean Temperature	0.41	-0.89	2.13	13.60
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	0.33	-0.35	1.40	2.14
BIO3	Isothermality (BIO2/BIO7) (* 100)	-0.18	-0.74	0.40	9.56
BIO4	Temperature Seasonality (Standard deviation *100)	0.63	0.52	5.07	4.75
BIO5	Max Temperature of Warmest Month	0.62	-0.67	4.90	7.81
BIO6	Min Temperature of Coldest Month	0.06	-0.95	0.05	15.60
BIO7	Temperature Annual Range (BIO5-BIO6)	0.60	0.35	4.60	2.16
BIO8	Mean Temperature of Wettest Quarter	0.73	-0.28	6.75	1.39
BIO9	Mean Temperature of Driest Quarter	-0.11	-0.79	0.15	10.90
BIO10	Mean Temperature of Warmest Quarter	0.63	-0.70	5.00	8.38
BIO11	Mean Temperature of Coldest Quarter	0.13	-0.97	0.21	16.40
BIO12	Annual Precipitation	-0.95	-0.13	11.38	0.29
BIO13	Precipitation of Wettest Month	-0.81	0.00	8.31	0.03
BIO14	Precipitation of Driest Month	-0.90	-0.30	10.38	1.54
BIO15	Precipitation Seasonality (Coefficient of Variation)	0.43	0.44	2.30	3.27
BIO16	Precipitation of Wettest Quarter	-0.84	0.00	8.86	0.00
BIO17	Precipitation of Driest Quarter	-0.92	-0.28	10.81	1.32
BIO18	Precipitation of Warmest Quarter	-0.74	0.09	6.93	0.14
BIO19	Precipitation of Coldest Quarter	-0.90	-0.22	10.38	0.88

REFERENCES

Aagaard, S. M. D., Såstad, S. M., Greilhuber, J., and Moen, A. (2005). A secondary hybrid zone between diploid *Dactylorhiza incarnata* ssp. *cruenta* and allotetraploid *D. lapponica* (Orchidaceae). *Heredity*, *94*(5), 488–496. https://doi.org/10.1038/sj.hdy.6800643

Arnold, B., Kim, S., and Bomblies, K. (2015). Single Geographic Origin of a Widespread Autotetraploid *Arabidopsis arenosa* Lineage Followed by Interploidy Admixture Article Fast Track. *Molecular Biology and Evolution*, *32*(6), 1382–1395. https://doi.org/10.1093/molbev/msv089

Arrigo, N., de La Harpe, M., Litsios, G., Zozomová-Lihová, J., Španiel, S., Marhold, K., … Alvarez, N. (2016). Is hybridization driving the evolution of climatic niche in *Alyssum montanum*. *American Journal of Botany*, *103*(7), 1348–1357. https://doi.org/10.3732/ajb.1500368

Baduel, P., Arnold, B., Weisman, C. M., Hunter, B., and Bomblies, K. (2016). Habitat-Associated Life History and Stress-Tolerance Variation in *Arabidopsis arenosa*, *171*(May), 437–451. https://doi.org/10.1104/pp.15.01875

Baduel, P., Hunter, B., Yeola, S., and Bomblies, K. (2018). Genetic basis and evolution of rapid cycling in railway populations of tetraploid *Arabidopsis arenosa*. *PLoS Genetics*, *14*(7), 1–26. https://doi.org/10.1371/journal.pgen.1007510

Blaine Marchant, D., Soltis, D. E., and Soltis, P. S. (2016). Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. *New Phytologist*, *212*(3), 708–718. https://doi.org/10.1111/nph.14069

Bohutínská, M., Handrick, V., Yant, L., Schmickl, R., Kolář, F., Bomblies, K., and Paajanen, P. (2021). De Novo Mutation and Rapid Protein (Co-)evolution during Meiotic Adaptation in *Arabidopsis arenosa*. *Molecular Biology and Evolution*, *38*(5), 1980–1994. https://doi.org/10.1093/molbev/msab001

Bohutínská, M., Vlček, J., Yair, S., Laenen, B., Konečná, V., Fracassetti, M., … Kolár, F. (2021). Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(21). https://doi.org/10.1073/pnas.2022713118

Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., … Guisan, A. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, *21*(4), 481–497. https://doi.org/10.1111/j.1466-8238.2011.00698.x

Burns, J. H., and Strauss, S. Y. (2011). More closely related species are more ecologically similar in an experimental test. *Proceedings of the National Academy of Sciences*, *108*(13), 5302–5307. https://doi.org/10.1073/pnas.1013003108

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. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.00315

Castro, M., Loureiro, J., Husband, B. C., and Castro, S. (2020). The role of multiple reproductive barriers: strong post-pollination interactions govern cytotype isolation in a tetraploid–octoploid contact zone. *Annals of Botany*, *126*(6), 991–1003. https://doi.org/10.1093/aob/mcaa084

Castro, M., Loureiro, J., Serrano, M., Tavares, D., Husband, B. C., Siopa, C., and Castro, S. (2019). Mosaic distribution of cytotypes in a mixed-ploidy plant species, *Jasione montana*: nested environmental niches but low geographical overlap. *Botanical Journal of the Linnean Society*, *190*(1), 51–66. https://doi.org/10.1093/botlinnean/boz007

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*? *Ecology and Evolution*, *5*(19), 4263–4276. https://doi.org/10.1002/ece3.1686

Čertner, M., Kúr, P., Kolář, F., and Suda, J. (2019). Climatic conditions and human activities shape diploid–tetraploid coexistence at different spatial scales in the common weed *Tripleurospermum inodorum* (Asteraceae). *Journal of Biogeography*, *00*, jbi.13629. https://doi.org/10.1111/jbi.13629

Coyne, J. A., and Orr, H. A. (2004). *Speciation*. Sinauer.

Decanter, L., Colling, G., Elvinger, N., Heiðmarsson, S., and Matthies, D. (2020). Ecological niche differences between two polyploid cytotypes of *Saxifraga rosacea*. *American Journal of Botany*, *107*(3), 423–435. https://doi.org/10.1002/ajb2.1431

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., … Guisan, A. (2017). ecospat: an R package to support spatial analyses and modelling of species niches and distributions. *Ecography*, *40*(6), 774–787. https://doi.org/10.1111/ecog.02671

Duchoslav, M., Jandová, M., Kobrlová, L., Šafářová, L., Brus, J., and Vojtěchová, K. (2020). Intricate Distribution Patterns of Six Cytotypes of *Allium oleraceum* at a Continental Scale: Niche Expansion and Innovation Followed by Niche Contraction With Increasing Ploidy Level. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.591137

Ehrich, D. (2006). AFLPDAT: A collection of R functions for convenient handling of AFLP data. *Molecular Ecology Notes*, *6*(3), 603–604. https://doi.org/10.1111/j.1471-8286.2006.01380.x

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

Gaynor, M. L., Marchant, D. B., Soltis, D. E., and Soltis, P. S. (2018). Climatic niche comparison among ploidal levels in the classic autopolyploid system, *Galax urceolata*. *American Journal of Botany*, *105*(10), 1631–1642. https://doi.org/10.1002/ajb2.1161

Glennon, K. L., Ritchie, M. E., and Segraves, K. A. (2014). Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecology Letters*, *17*(5), 574–582. https://doi.org/10.1111/ele.12259

Godsoe, W., Larson, M. A., Glennon, K. L., and Segraves, K. A. (2013). Polyploidization in *Heuchera cylindrica* (Saxifragaceae) did not result in a shift in climatic requirements. *American Journal of Botany*, *100*(3), 496–508. https://doi.org/10.3732/ajb.1200275

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., and Kueffer, C. (2014). Unifying niche shift studies: insights from biological invasions. *Trends in Ecology and Evolution*, *29*(5), 260–269. https://doi.org/10.1016/j.tree.2014.02.009

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., … Guo, Y.-L. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics*, *43*(5), 476–481. https://doi.org/10.1038/ng.807

Husband, B. C. (2016). Effect of inbreeding on pollen tube growth in diploid and tetraploid *Chamerion angustifolium*: Do polyploids mask mutational load in pollen? *American Journal of Botany*, *103*(3), 532–540. https://doi.org/10.3732/ajb.1500243

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., … dePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature*, *473*, 97–100. https://doi.org/10.1038/nature09916

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kirchheimer, B., Schinkel, C. C. F. F., Dellinger, A. S., Klatt, S., Moser, D., Winkler, M., … Dullinger, S. (2016). A matter of scale: apparent niche differentiation of diploid and tetraploid plants may depend on extent and grain of analysis. *Journal of Biogeography*, *43*(4), 716–726. https://doi.org/10.1111/jbi.12663

Knotek, A., Konečná, V., Wos, G., Požárová, D., Šrámková, G., Bohutínská, M., … Kolář, F. (2020). Parallel Alpine Differentiation in *Arabidopsis arenosa*. *Frontiers in Plant Science*, *11*(December), 1–12. https://doi.org/10.3389/fpls.2020.561526

Kolář, F., Fuxová, G., Záveská, E., Nagano, A. J., Hyklová, L., Lučanová, M., … Marhold, K. (2016). Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant

model *Arabidopsis arenosa*. *Molecular Ecology*, *25*(16), 3929–3949. https://doi.org/10.1111/mec.13721
Kolář, F., Lučanová, M., Záveská, E., Fuxová, G., Mandáková, T., Španiel, S., … Marhold, K. (2016). Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae). *Biological Journal of the Linnean Society*, *119*(3), 673–688. https://doi.org/10.1111/bij.12479
Kolář, F., Píšová, S., Záveská, E., Fér, T., Weiser, M., Ehrendorfer, F., and Suda, J. (2015). The origin of unique diversity in deglaciated areas: Traces of Pleistocene processes in north-European endemics from the *Galium pusillum* polyploid complex (Rubiaceae). *Molecular Ecology*, *24*(6), 1311–1334. https://doi.org/10.1111/mec.13110
Konečná, V., Bray, S., Vlček, J., Bohutínská, M., Požárová, D., Choudhury, R. R., … Kolář, F. (2021). Parallel adaptation in autopolyploid *Arabidopsis arenosa* is dominated by repeated recruitment of shared alleles. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-25256-5
Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387
Koutecký, P., Baďurová, T., Štech, M., Košnar, J., and Karásek, J. (2011). Hybridization between diploid *Centaurea pseudophrygia* and tetraploid *C. jacea* (Asteraceae): the role of mixed pollination, unreduced gametes, and mentor effects. *Biological Journal of the Linnean Society*, *104*(1), 93–106. https://doi.org/10.1111/j.1095-8312.2011.01707.x
Levin, D. A. (2004). The ecological transition in speciation. *New Phytologist*, *161*(1), 91–96. https://doi.org/10.1046/j.1469-8137.2003.00921.x
López-Jurado, J., Mateos-Naranjo, E., and Balao, F. (2019). Niche divergence and limits to expansion in the high polyploid *Dianthus broteri* complex. *New Phytologist*, *222*(2), 1076–1087. https://doi.org/10.1111/nph.15663
Madlung, A. (2013). Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity*, *110*(2), 99–104. https://doi.org/10.1038/hdy.2012.79
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110
Molina-Henao, Y. F., and Hopkins, R. (2019). Autopolyploid lineage shows climatic niche expansion but not divergence in *Arabidopsis arenosa*. *American Journal of Botany*, *106*(1), 61–70. https://doi.org/10.1002/ajb2.1212
Monnahan, P., Kolář, F., Baduel, P., Sailer, C., Koch, J., Horvath, R., … Yant, L. (2019). Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nature Ecology and Evolution*, *3*, 457–468. https://doi.org/10.1101/411041
Morgan, C., White, M. A., Franklin, F. C. H., Zickler, D., Kleckner, N., and Bomblies, K. (2021). Evolution of crossover interference enables stable autopolyploidy by ensuring pairwise partner connections in *Arabidopsis arenosa*. *Current Biology*, *31*(21), 4713-4726.e4. https://doi.org/10.1016/j.cub.2021.08.028
Morgan, E. J., Čertner, M., Lučanová, M., Deniz, U., Kubíková, K., Venon, A., … Kolář, F. (2021). Disentangling the components of triploid block and its fitness consequences in natural diploid–tetraploid contact zones of *Arabidopsis arenosa*. *New Phytologist*, *232*(3), 1449–1462. https://doi.org/10.1111/nph.17357
Morgan, E. J., Čertner, M., Lučanová, M., Kubíková, K., Marhold, K., and Kolář, F. (2020). Niche similarity in diploid-autotetraploid contact zones of *Arabidopsis arenosa* across spatial scales. *American Journal of Botany*, *107*(10), 1–14. https://doi.org/10.1002/ajb2.1534
Muñoz-Pajares, A. J., Perfectti, F., Loureiro, J., Abdelaziz, M., Biella, P., Castro, M., … Gómez, J. M. (2018). Niche differences may explain the geographic distribution of cytotypes in *Erysimum mediohispanicum*. *Plant Biology*, *20*(Stebbins 1971), 139–147. https://doi.org/10.1111/plb.12605
Nei, M. (1987). *Molecular evolutionary genetics*. New York: Columbia University Press, New York.
Otto, S. P., and Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review of Genetics*, *34*(1), 401–437. https://doi.org/10.1146/annurev.genet.34.1.401
Parisod, C., and Broennimann, O. (2016). Towards unified hypotheses of the impact of polyploidy on ecological niches. *The New Phytologist*, *212*(3), 540–542. Retrieved from https://www.jstor.org/stable/newphytologist.212.3.540
Pembleton, L. W., Cogan, N. O. I., and Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, *13*(5), 946–952. https://doi.org/10.1111/1755-0998.12129
Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945
Rothfels, C. J. (2021). Polyploid phylogenetics. *New Phytologist*, *230*(1), 66–72. https://doi.org/10.1111/nph.17105
Schmickl, R., and Yant, L. (2021). Adaptive introgression: how polyploidy reshapes gene flow landscapes. *New Phytologist*, *230*(2), 457–461. https://doi.org/10.1111/nph.17204
Schoener, T. W. (1968). Anolis lizards of Bimini: Resource partitioning in a com-plex fauna. *Ecology*, *49*, 704–726.

Segraves, K. A., and Thompson, J. N. (1999). Plant Polyploidy and Pollination: Floral Traits and Insect Visits to Diploid and Tetraploid *Heuchera grossulariifolia*. *Evolution*, *53*(4), 1114. https://doi.org/10.2307/2640816

Šmíd, J., Douda, J., Krak, K., and Mandák, B. (2020). Analyses of Hybrid Viability across a Hybrid Zone between Two *Alnus* Species Using Microsatellites and cpDNA Markers. *Genes*, *11*(7), 770. https://doi.org/10.3390/genes11070770

Soltis, D. E., Segovia-Salcedo, M. C., Jordon-Thaden, I., Majure, L., Miles, N. M., Mavrodiev, E. V., … Gitzendanner, M. A. (2014). Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al . (2011) . *New Phytologist*, *202*(4), 1105–1117. https://doi.org/10.1111/nph.12756

Soltis, P. S., Marchant, D. B., Van de Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Current Opinion in Genetics and Development*, *35*, 119–125. https://doi.org/10.1016/j.gde.2015.11.003

Sutherland, B. L., and Galloway, L. F. (2017). Postzygotic isolation varies by ploidy level within a polyploid complex. *New Phytologist*, *213*(1), 404–412. https://doi.org/10.1111/nph.14116

Theodoridis, S., Randin, C., Broennimann, O., Patsiou, T., and Conti, E. (2013). Divergent and narrower climatic niches characterize polyploid species of European primroses in *Primula* sect. *Aleuritia*. *Journal of Biogeography*, *40*(7), 1278–1289. https://doi.org/10.1111/jbi.12085

Thompson, J. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, *25*(24), 4876–4882. https://doi.org/10.1093/nar/25.24.4876

Thompson, K. A., Husband, B. C., and Maherali, H. (2014). Climatic niche differences between diploid and tetraploid cytotypes of *Chamerion angustifolium* (Onagraceae). *American Journal of Botany*, *101*(11), 1868–1875. https://doi.org/10.3732/ajb.1400184

Visger, C. J., Germain-Aubrey, C. C., Patel, M., Sessa, E. B., Soltis, P. S., and Soltis, D. E. (2016). Niche divergence between diploid and autotetraploid *Tolmiea*. *American Journal of Botany*, *103*(8), 1396–1406. https://doi.org/10.3732/ajb.1600130

Visser, V., and Molofsky, J. (2015). Ecological niche differentiation of polyploidization is not supported by environmental differences among species in a cosmopolitan grass genus. *American Journal of Botany*, *102*(1), 36–49. https://doi.org/10.3732/ajb.1400432

Warren, D. L., Glor, R. E., and Turelli, M. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, *62*(11), 2868–2883. https://doi.org/10.1111/j.1558-5646.2008.00482.x

Wendel, J F. (2000). Genome evolution in polyploids. *Plant Molecular Biology*, *42*(1), 225–249.

Wendel, Jonathan F. (2015). The wondrous cycles of polyploidy in plants. *American Journal of Botany*, *102*(11), 1753–1756. https://doi.org/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. *Proceedings of the National Academy of Sciences*, *106*(33), 13875–13879. https://doi.org/10.1073/pnas.0811575106

Wos, G., Bohutínská, M., Nosková, J., Mandáková, T., and Kolář, F. (2021). Parallelism in gene expression between foothill and alpine ecotypes in *Arabidopsis arenosa*. *Plant Journal*, *105*(5), 1211–1224. https://doi.org/10.1111/tpj.15105

Wos, G., Mořkovská, J., Bohutínská, M., Šrámková, G., Knotek, A., Lučanová, M., … Kolář, F. (2019). Role of ploidy in colonization of alpine habitats in natural populations of *Arabidopsis arenosa*. *Annals of Botany*. *124*(2), 255–268 https://doi.org/10.1093/aob/mcz070

Appendix I: Copyright permissions

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS
Jan 04, 2022

This Agreement between Mrs. Gabriela Šrámková ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	5221950870865
License date	Jan 04, 2022
Licensed content publisher	Oxford University Press
Licensed content publication	Biological Journal of the Linnean Society
Licensed content title	Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae)
Licensed content author	Kolář, Filip; Lučanová, Magdalena
Licensed content date	Oct 27, 2016
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Evolution of the genus Arabidopsis in its centre of diversity
Publisher of your work	Faculty of Science, Charles University
Expected publication date	Mar 2022
Permissions cost	0.00 EUR
Value added tax	0.00 EUR
Total	0.00 EUR
Title	Evolution of the genus Arabidopsis in its centre of diversity
Institution name	Faculty of Science, Charles University
Expected presentation date	Mar 2022
Portions	Page 1 to 16 including figures and tables
Requestor Location	Mrs. Gabriela Šrámková Benátská 2 Prague, 12800 Czechia Attn: Mrs. Gabriela Šrámková
Publisher Tax ID	GB125506730
Customer VAT ID	CZ00216208

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.
5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com
7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.
8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.
9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.
10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
12. Other Terms and Conditions: v1.4

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS
Jan 04, 2022

This Agreement between Mrs. Gabriela Šrámková ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	5222041127239
License date	Jan 04, 2022
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Molecular Ecology
Licensed Content Title	Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model Arabidopsis arenosa
Licensed Content Author	Karol Marhold, Hiroshi Kudoh, Magdalena Lučanová, et al
Licensed Content Date	Jul 29, 2016
Licensed Content Volume	25
Licensed Content Issue	16
Licensed Content Pages	21
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title	Evolution of the genus Arabidopsis in its centre of diversity
Institution name	Faculty of Science, Charles University
Expected presentation date	Mar 2022
Requestor Location	Mrs. Gabriela Šrámková Benátská 2 Prague, 12800 Czechia Attn: Mrs. Gabriela Šrámková
Publisher Tax ID	EU826007151
Customer VAT ID	CZ00216208
Total	0.00 EUR

TERMS AND CONDITIONS
This copyrighted material is owned by or exclusively licensed to John Wiley and Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.
- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley and Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is

granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.
- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS
Jan 04, 2022

This Agreement between Mrs. Gabriela Šrámková ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	5222041352152
License date	Jan 04, 2022
Licensed content publisher	Oxford University Press
Licensed content publication	Botanical Journal of the Linnean Sociey
Licensed content title	Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction
Licensed content author	Šrámková-Fuxová, Gabriela; Záveská, Eliška
Licensed content date	Oct 31, 2017
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Evolution of the genus Arabidopsis in its centre of diversity
Publisher of your work	Faculty of Science, Charles University
Expected publication date	Mar 2022
Permissions cost	0.00 EUR
Value added tax	0.00 EUR
Total	0.00 EUR
Title	Evolution of the genus Arabidopsis in its centre of diversity
Institution name	Faculty of Science, Charles University
Expected presentation date	Mar 2022
Portions	pages 1-22 including figures and tables
Requestor Location	Mrs. Gabriela Šrámková Benátská 2 Prague, 12800 Czechia Attn: Mrs. Gabriela Šrámková
Publisher Tax ID	GB125506730
Customer VAT ID	CZ00216208
Total	0.00 EUR

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.
5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com
7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.
8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.
9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.
10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
12. Other Terms and Conditions:

v1.4

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Jan 04, 2022

This Agreement between Mrs. Gabriela Šrámková ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5222050105276
License date	Jan 04, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Plant Systematics and Evolution
Licensed Content Title	Phylogeography and taxonomic reassessment of Arabidopsis halleri – a montane species from Central Europe
Licensed Content Author	Gabriela Šrámková et al
Licensed Content Date	Nov 19, 2019
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	yes
Title	Evolution of the genus Arabidopsis in its centre of diversity
Institution name	Faculty of Science, Charles University
Expected presentation date	Mar 2022
Requestor Location	Mrs. Gabriela Šrámková Benátská 2 Prague, 12800 Czechia Attn: Mrs. Gabriela Šrámková
Customer VAT ID	CZ00216208
Total	0.00 EUR

Springer Nature Customer Service Centre GmbH
Terms and Conditions
This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. **Grant of License**
    1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.
    2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).
    3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.
2. **Scope of Licence**
    1. You may only use the Licensed Content in the manner and to the extent permitted by these TsandCs and any applicable laws.
    2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.
    3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.
    4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.
    5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

☐ **Duration of Licence**

1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

☐ **Acknowledgement**

1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

☐ **Restrictions on use**

1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.
2. You must not use any Licensed Material as part of any design or trademark.
3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

☐ **Ownership of Rights**

1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

☐ **Warranty**

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

☐ **Limitations**

1. ***BOOKS ONLY:***Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).
2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

☐ **Termination and Cancellation**

1. Licences will expire after the period shown in Clause 3 (above).
2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)
For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)
For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)
Note: For any republication from the British Journal of Cancer, the following credit line style applies:
Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)
For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])
For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions: Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Jan 04, 2022

This Agreement between Mrs. Gabriela Šrámková ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5222050254587
License date	Jan 04, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Plant Systematics and Evolution
Licensed Content Title	Correction to: Phylogeography and taxonomic reassessment of Arabidopsis halleri – a montane species from Central Europe
Licensed Content Author	Gabriela Šrámková et al
Licensed Content Date	May 27, 2020
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	yes
Title	Evolution of the genus Arabidopsis in its centre of diversity
Institution name	Faculty of Science, Charles University
Expected presentation date	Mar 2022
Requestor Location	Mrs. Gabriela Šrámková Benátská 2 Prague, 12800 Czechia Attn: Mrs. Gabriela Šrámková
Customer VAT ID	CZ00216208
Total	0.00 EUR

Springer Nature Customer Service Centre GmbH
Terms and Conditions
This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and

completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. **Grant of License**
    1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.
    2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).
    3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.
2. **Scope of Licence**
    1. You may only use the Licensed Content in the manner and to the extent permitted by these TsandCs and any applicable laws.
    2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.
    3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.
    4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.
    5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

☐ **Duration of Licence**

1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

☐ **Acknowledgement**

1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

☐ **Restrictions on use**

1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.
2. You must not use any Licensed Material as part of any design or trademark.
3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

☐ **Ownership of Rights**

1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

☐ **Warranty**

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND

WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

☐ **Limitations**

1. ***BOOKS ONLY:***Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).
2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

☐ **Termination and Cancellation**

1. Licences will expire after the period shown in Clause 3 (above).
2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:
Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions: Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Appendix II: Curriculum vitae

Personal data

Gabriela Šrámková
* 6.2.1986
Maiden name: Fuxová
One child * 2017

Education

Since 2012: PhD study in Botany, Department of Botany, Faculty of Science, Charles University (supervisor: prof. RNDr. Karol Marhold, DrSc.)

2011–2015: Lifelong learning programme – Profession oriented, Department of Education and Didactics of Biology, Faculty of Science, Charles University (supervisor: Mgr. Zuzana Chumová)

2008–2011: MSc. in Botany, Department of Botany, Faculty of Science, Charles University (supervisor: Mgr. Tomáš Fér, Ph.D.)

2005–2008: BSc in Biology, Department of Botany, Faculty of Science, Charles University (supervisor: Mgr. Tomáš Fér, Ph.D.)

Appointments

since 2013: Department of Botany, Faculty of Science, Charles University in Prague – researcher

2011–2013: Department of Botany, Faculty of Science, Charles University in Prague – laboratory technician

2007–2011: Department of Botany, Faculty of Science, Charles University in Prague – microsatellite analyses related to MSc. thesis

2007–2010: Institute of Botany, Academy of Sciences of the Czech Republic – part-time lab work and database work

Fellowships

5. –11. 8. 2018 Earlham Institute, Norwich, United Kingdom – WGS method LITE training in the laboratory of Dr. Darren Heavens, labwork, library preparation, consultations, troubleshooting

15. –20. 12. 2015 Faculty Centre for Biodiversity, University of Vienna, Austria – RADseq method consultations with the team of Dr. Ovidiu Paun, lab work

18. 4. –11. 5. 2014 Centre of Ecological Research, University of Kyoto, Japan – joining the team of prof. Hiroshi Kudoh, lab work, ddRADseq library preparation, consultations (troubleshooting, use for phylogeographic studies)

Teaching

Lectures at Faculty of Science, Charles University in Prague:

Practices
- Molecular markers in systematics and plant population biology (master course)
- Plant morphology (bachelor course)
- Phylogeny and morphology of vascular plants (bachelor course)

Lectures
- Advanced methods in DNA sequence and multilocus data analyses (master and phd course)

Supervision of the thesis

Filip Holič (consultant): *Cardamine dentata,* its distribution in Central Europe and relationship to *C. pratensis* (master thesis, Department of Botany, Charles University)

Martina Gruntová (consultant): Invasiveness and hybridization in evolution of closely related species (bachelor student, Department of Botany, Charles University)

SCI publications

Melichárková A., Šlenker M., Zozomová-Lihová J., Skokanová K., Šingliarová B., Kačmárová T., Caboňová M., Kempa M., **Šrámková G.**, Mandáková T., Lysák M.A., Svitok M., Mártonfiová L., Marhold K. (2020): So Closely Related and Yet So Different: Strong Contrasts Between the Evolutionary Histories of Species of the *Cardamine pratensis* Polyploid Complex in Central Europe, Frontiers in Plant Science 11:588856.

Knotek A., Konečná V., Wos G., Požárová D., **Šrámková G.**, Bohutínská M., Zeisek V., Marhold K., Kolář F. (2019): Parallel alpine differentiation in *Arabidopsis arenosa*, Frontiers in Plant Science 11:561526.

Šrámková G., Kolář F., Záveská E., Lučanová M., Španiel S., Kolník M., Marhold K. (2019): Phylogeography and taxonomic reassessment of *Arabidopsis halleri* – a montane species from Central Europe, Plant Systematics and Evolution 205(10):885-898.

Wos G., Mořkovská J., Bohutínská M., **Šrámková G.**, Knotek A., Lučanová M., Španiel S., Marhold K., Kolář F. (2019): Role of ploidy in colonization of alpine habitats in natural populations of *Arabidopsis arenosa,* Annals of Botanny 124/2:255-268.

Monnahan P., Kolář F., Baduel P., Sailer Ch., Koch J., Horvath R., Laenen B., Schmickl R., Paajanen P., **Šrámková G.**, Bohutínská M., Arnold B., Weisman C.M., Marhold K., Slotte T., Bomblies K., Yant L. (2019): Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*, Nature Ecology and Evolution 3:457-468.

Šrámková-Fuxová G., Záveská E., Kolář F., Lučanová M., Španiel S., Marhold K. (2017): Range-wide genetic structure of *Arabidopsis halleri* (Brassicaceae): glacial persistence in multiple refugia and origin of the Northern Hemisphere disjunction, Botanical Journal of the Linnean Society 185/3:321–342.

Kolář F*, **Fuxová G***, Záveská E*, Nagano AJ, Hyklová L, Lučanová M, Kudoh H, Marhold K (2016): Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. – Molecular Ecology, 25:3929-3949. *equal contribution

Kolář F, Lučanová M, Záveská E, **Fuxová G**, Mandáková T, Španiel S, Senko D, Svitok M, Kolník M, Gudžinskas Z, Marhold K (2015): Ecological segregation does not drive the intricate parapatric distribution of diploid and tetraploid cytotypes of the *Arabidopsis arenosa* group (Brassicaceae). – Biological Journal of the Linnean Society 119(3):673-688.

Non-SCI publications

Šrámková G., Slovák M., Kolář F. (2019): Pátrání po dědictví ledových dob v karpatské flóře (Searching for the Ice Age Heritage in Carpathian Flora). Živa 5:236-239.

Reviewing experience: BMC Plant Biology