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

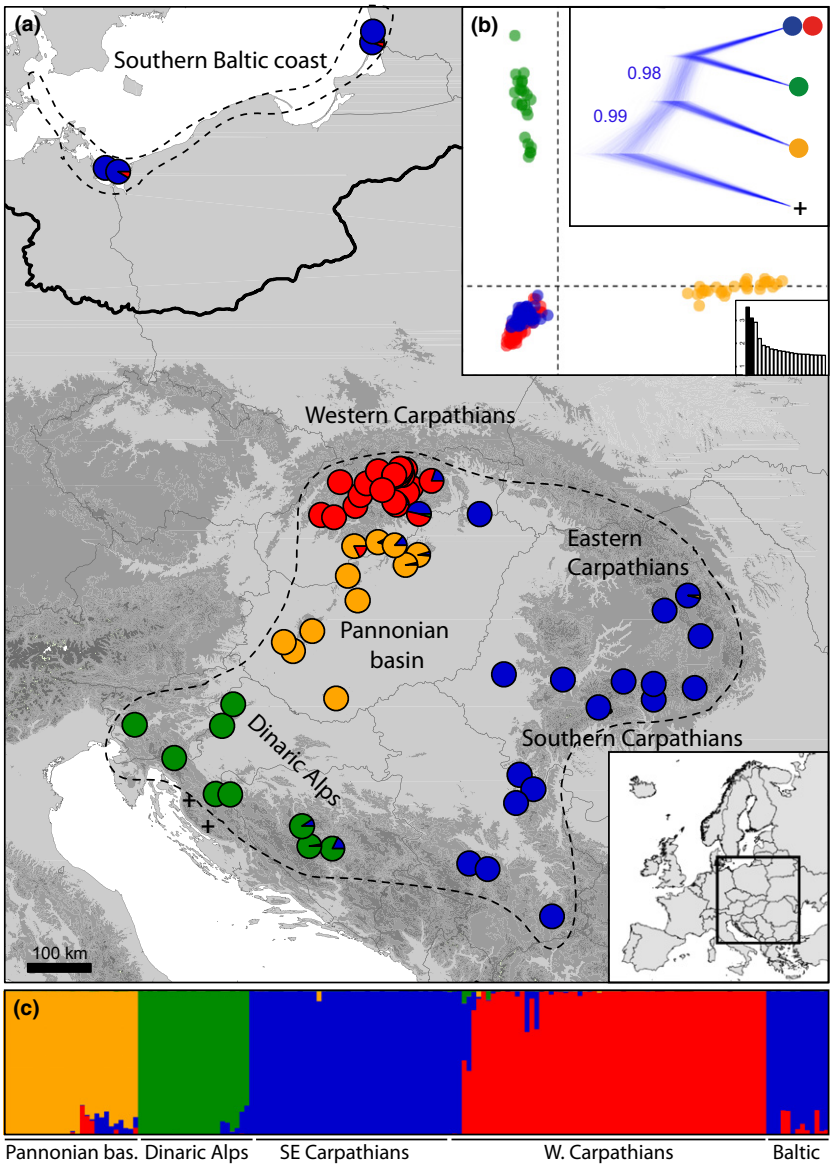

Fig. 1 Rangewide genetic differentiation of diploid *Arabidopsis arenosa*, (a) geographical distribution of sampled *A. arenosa* populations (pie charts reflecting proportional assignment to a particular STRUCTURE group) and *A. croatica* (black cross), bold line indicates maximal extent of the continental 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

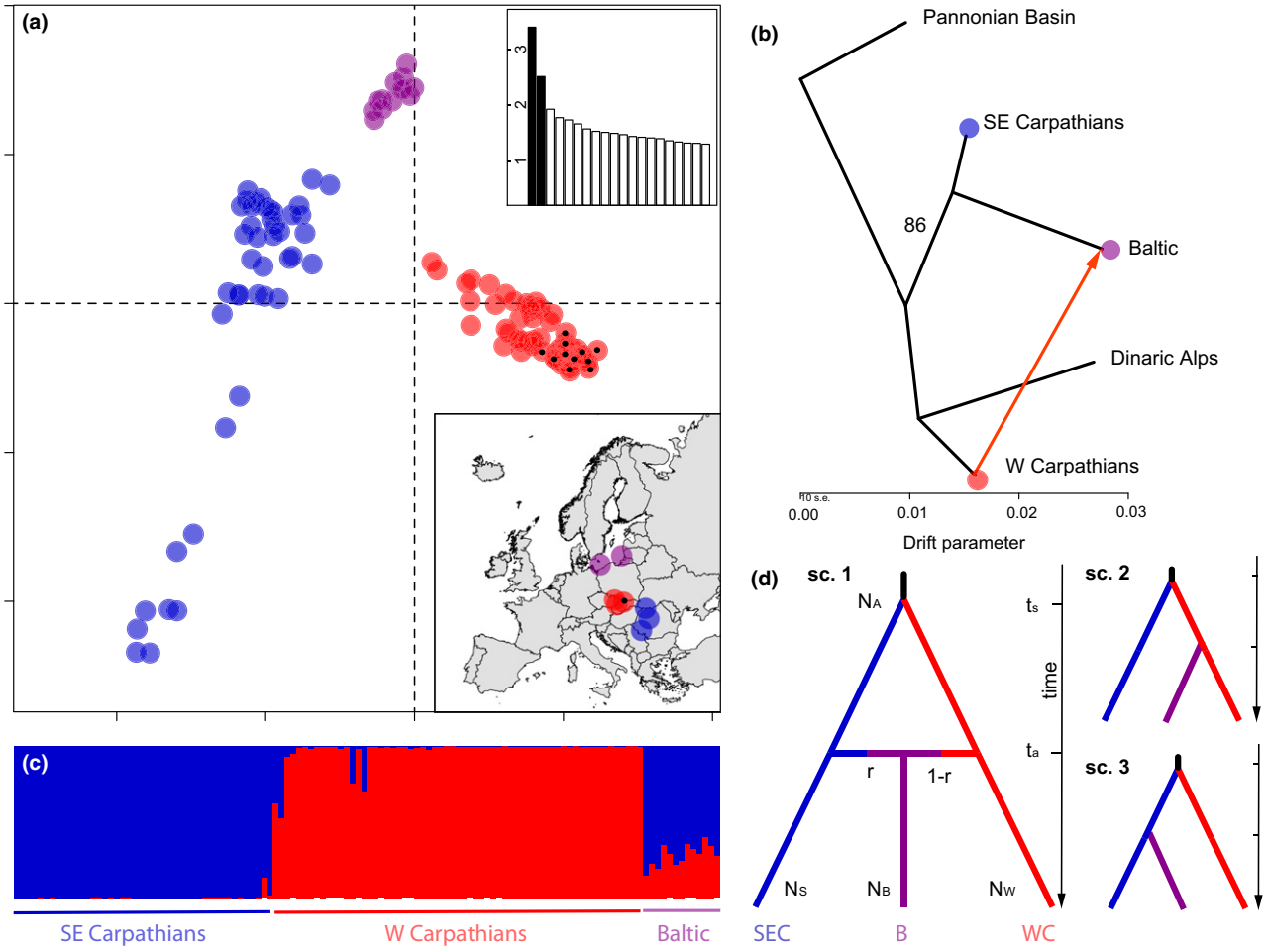

Fig. 2 Reconstruction of the relationships among the Baltic and Carpathian populations of diploid *Arabidopsis arenosa*. (a) Principal coordinate ordination (8949 SNPs, first and second axis displayed, histogram shows proportional contribution in explaining variance by the first 20 axes) of Carpathian + 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.

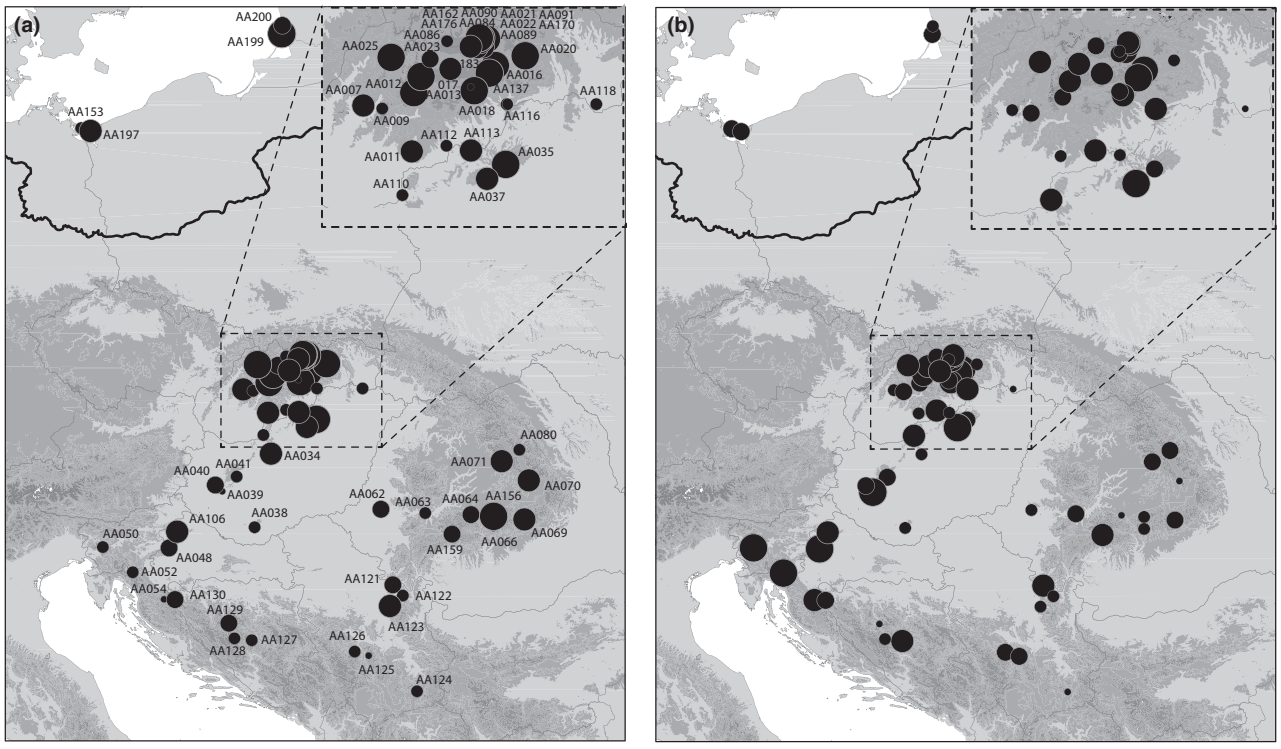

Fig. 3 Patterns of population-level diversity and rarity in diploid *A. arenosa*. (a) Genetic diversity of populations expressed by expected heterozygosity, H_e (range 0.30–0.61); (b) proportion of rare alleles in populations measured by frequency-downweighted index, DW (range 0.12–1.02). The bold line 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 ($\chi^2 = 21.2$, 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

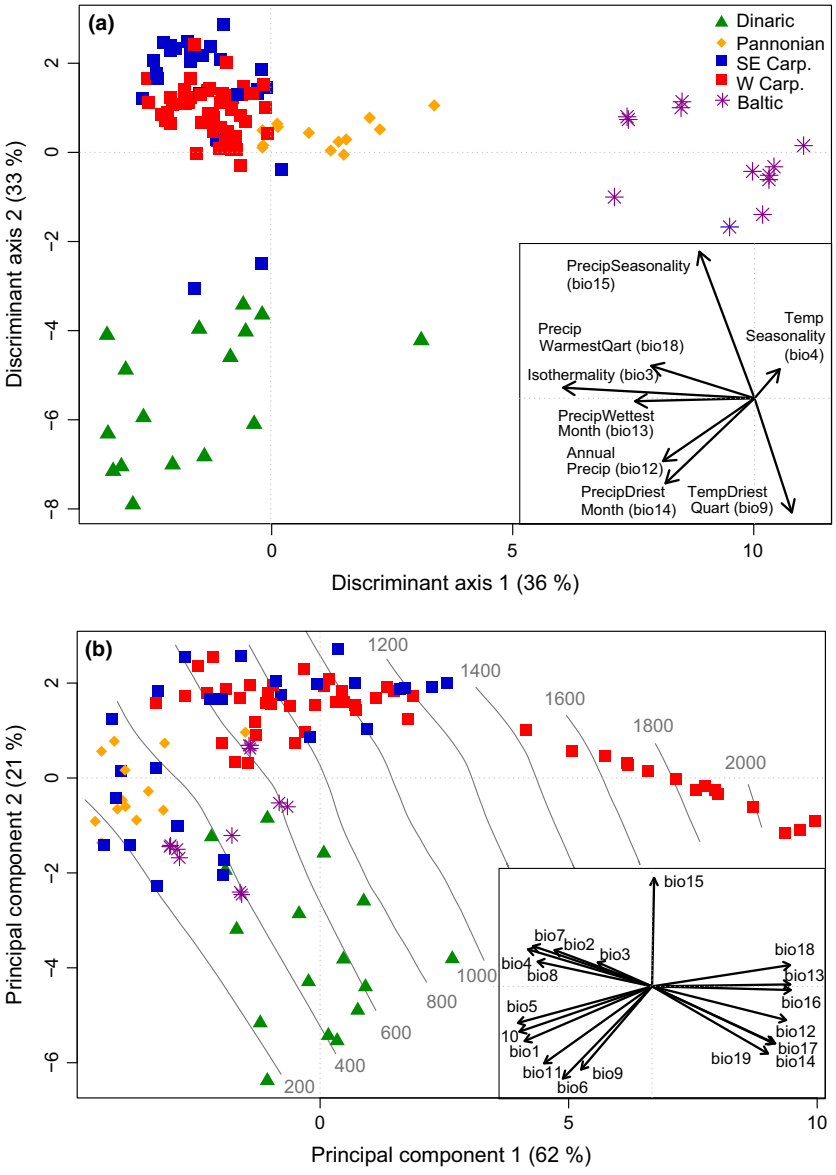

Fig. 4 Rangewide climatic niche variation in 117 ecologically screened populations of diploid *A. arenosa* denoted according to the five main genetic groups. (a) Linear discriminant analysis based on eight variables significantly contributing to the model (chosen using forward selection). (b) Principal 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

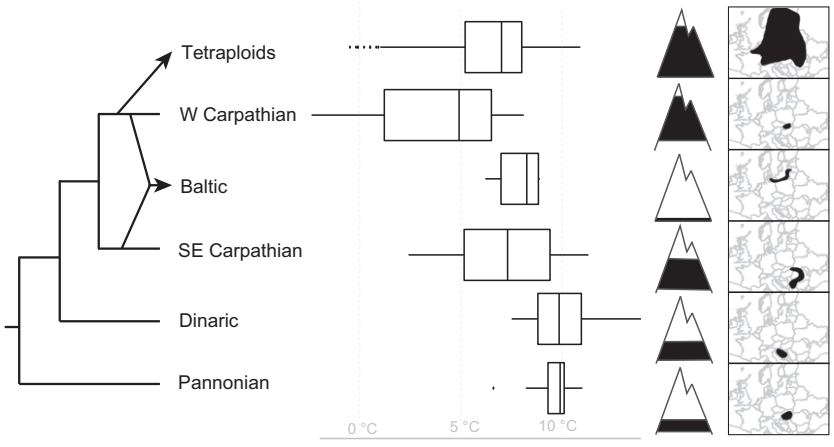

Fig. 5 Summary hypothesis on the history and niche evolution of diploid (based on this study) and autotetraploid (following Arnold *et al.* 2015 and Kolář *et al.* 2015a) lineages of *Arabidopsis arenosa*. Phylogenetic reconstruction, mean annual temperature, altitudinal range (proportionally in range 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 wschodniokarpackiej. *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.