Molecular Phylogenetics and Evolution 139 (2019) 106572

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Multiple auto- and allopolyploidisations marked the Pleistocene history of the widespread Eurasian steppe plant *Astragalus onobrychis* (Fabaceae)

Eliška Záveská[a], Clemens Maylandt[a], Ovidiu Paun[b], Clara Bertel[a], Božo Frajman[a], The STEPPE Consortium[a,b,c,d,e,f,g,1], Peter Schönswetter[a,*]

[a] Department of Botany, University of Innsbruck, Sternwartestraße 15, 6020 Innsbruck, Austria
[b] Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, 1030 Vienna, Austria
[c] Department of Biosciences, University of Salzburg, Hellbrunnerstrasse 34, 5020 Salzburg, Austria
[d] Department of Ecology, University of Innsbruck, Technikerstraße 25, 6020 Innsbruck, Austria
[e] Real Jardín Botánico, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain
[f] Department of Life Sciences and Biotechnology, University of Ferrara, Via L. Borsari 46, 44121 Ferrara, Italy
[g] Institute for Alpine Environment, Eurac Research, Drususallee 1/Viale Druso 1, 39100 Bozen/Bolzano, Italy

ARTICLE INFO

Keywords:
Autopolyploidy
Allopolyploidy
Genotype-free inference
Eurasian steppes
Plastid DNA
RADseq
Reticulate evolution

ABSTRACT

The Eurasian steppes occupy a significant portion of the worldwide land surface and their biota have been affected by specific past range dynamics driven by ice ages-related climatic fluctuations. The dynamic alterations in conditions during the Pleistocene often triggered reticulate evolution and whole genome duplication events. Employing genomic, genetic and cytogenetic tools as well as morphometry we investigate the intricate evolution of *Astragalus onobrychis*, a widespread Eurasian steppe plant with diploid, tetraploid and octoploid cytotypes. To analyse the heteroploid RADseq dataset we employ both genotype-based and genotype-free methods that result in highly consistent results, and complement our inference with information from the plastid ycf1 region. We uncover a complex and reticulate evolutionary history, including at least one auto-tetraploidization event and two allo-octoploidization events; one of them involved also genetic contributions from other species, most likely *A. goktschaicus*. The present genetic structure points to the existence of four main clades within *A. onobrychis*, which only partly correspond to different ploidies. Time-calibrated diffusion models suggest that diversification within *A. onobrychis* was associated with ice age-related climatic fluctuations during the last million years. We finally argue for the usefulness of uniparentally inherited plastid markers, even in the genomic era, especially when investigating heteroploid systems.

1. Introduction

The Eurasian steppes are among the largest continuous biomes on Earth, covering ca 7% of the total land surface. Bordered by boreal forests to the North and (semi)deserts to the South, they range from the Pontic plains north of the Black Sea over southern Siberia and Central Asia to northern China (Wesche et al., 2016). The extent of the Eurasian steppe belt is determined by precipitation, which does not suffice to permit forest growth. Beyond this continuous area (the "zonal" steppe), isolated occurrences of steppe vegetation in central, western and southern Europe (the "extrazonal" steppe) are restricted to well-drained, shallow soils in southern exposition in dry and continental, but potentially forested areas (Kajtoch et al., 2016). Reflecting the biome's vast extent, many steppe species inhabit much of Eurasia. Their distributions are often continuous from the Pontic plains to China and become increasingly fragmented towards their western limits in central and southern Europe (Meusel et al., 1965, 1978; Meusel and Jäger, 1992).

The past range dynamics of the Eurasian steppes are characterised by the expansion of continental steppes at the expense of nemoral and boreal forests during cold and dry stages (glacials), and the disintegration of the cold-stage steppe belt leading to the fragmentation of steppe outliers, during warm and humid stages (interglacials; Schmitt and Varga, 2012; Varga, 2010). This implies that – as opposed to the temperate tree flora (Petit et al., 2003; Svenning and Skov, 2007) – cold and drought tolerant steppe species have been much more widely distributed during cold stages. Hence, their distribution ranges retracted

* Corresponding author.
E-mail address: peter.schoenswetter@uibk.ac.at (P. Schönswetter).

[1] The members of the STEPPE Consortium are listed in Frajman et al., 2019. Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Eurasian *Euphorbia seguieriana* s. l. (Euphorbiaceae). Mol. Phylogenet. Evol. 134, 250.

https://doi.org/10.1016/j.ympev.2019.106572
Received 28 March 2019; Received in revised form 12 July 2019; Accepted 23 July 2019
Available online 24 July 2019

during interglacials and the Holocene in response to the contraction of suitable niches (Kajtoch et al., 2016; but see Frajman et al., 2019). Phases of allopatry in disjunct and relatively isolated steppe refugia during warm stages have likely triggered evolutionary processes such as divergent evolution governed by genetic drift, whereas phases of range expansion during cold and dry glacial phases favoured secondary contact and gene flow among previously diverged lineages resulting in reticulate evolution (Comes and Kadereit, 1998).

In contrast to the similarly widespread and disjunctly distributed arctic-alpine species, which have been the focus of molecular biogeographic and taxonomic research for two decades (Gabrielsen et al., 1997; Winkler et al., 2012), our knowledge of the spatiotemporal evolution and intraspecific diversification of steppe species is still limited. The Pontic plains and the Pannonian basin have been suggested as major warm-stage refugia for steppe species in Europe (Stewart et al., 2010; Varga, 2010), whereas the disjoint steppe islands to the west of the zonal Eurasian steppe belt have been assumed to be the result of recolonization events, when this biome expanded during cold and dry periods (Gauckler, 1938; Jännicke, 1892; Kerner, 1863).

Recent molecular studies, however, have seriously challenged this view by providing genetic evidence for the existence of hitherto disregarded refugia outside of the zonal steppe belt in Central Europe, the Balkan and Apennines Peninsulas and the Alps, for instance in *Cheilotoma musciformis* (Kajtoch et al., 2013), *Melitaea cinxia* (Wahlberg and Saccheri, 2007), *Coronella austriaca* (Sztencel-Jabłonka et al., 2015), *Cricetus cricetus* (Neumann et al., 2004; Neumann et al., 2005), *Linum flavum* (Cieślak, 2014; Plenk et al., 2017), *Scorzonera purpurea* (Meindl, 2012) and *Euphorbia seguieriana* (Frajman et al., 2019). Interestingly, many of the examined species show highly differentiated populations not only in the Pontic plains and the Pannonian basin, but also further to the north (Kajtoch et al., 2016). For instance, in *Stipa capillata*, one of the most typical elements of Eurasian steppes, the divergence among Central European populations has been shown to be significantly greater than among Central Asian ones (Wagner et al., 2011), whereas for *E. seguieriana* westward dispersal from the zonal steppes was invoked (Frajman et al., 2019). In any event, due to the scarcity of studies based on both a comprehensive, range-wide sampling and a complementary set of genetic or genomic markers, no phylogeographic synthesis has yet emerged for the Eurasian steppe biome, not even for the extrazonal European steppes (Kajtoch et al., 2016).

Range oscillations linked to glacial cycles may have played a pivotal role in the evolution of steppe plants by triggering whole genome duplication (WGD). On the one hand, secondary contacts resulting from range expansions may have triggered reticulation events, potentially coupled with WGD (Kadereit et al., 2004). On the other hand, there is accumulating evidence that phases of climatic turmoil, such as changes in temperature, directly stimulate WGD through an increase in the production of unreduced gametes within populations (Bomblies and Madlung, 2014; Kreiner et al., 2017; Ramsey and Schemske, 1998). Both processes may lead to WGD within species (autopolyploidy) or involving hybridisation between different parental species (allopolyploidy). Once established, young polyploids may have a higher stress tolerance than diploids, as shown by their occurrence in recently formed, disturbed or harsh environments (Diallo et al., 2016; Madlung, 2013; Ramsey, 2011; Rice et al., 2019). Overall, through increased heterozygosity and gene redundancy, whole genome duplications can reduce the effect of environmental stress – such as that imposed by the rapid climatic changes during the Pleistocene (Van de Peer et al., 2017).

Elucidating phylogenetic relationships and evolutionary processes in polyploids remains a challenge to molecular systematics and evolutionary research due to mixed inheritance (Stift et al., 2008), difficulties in clearly identifying homoeologs and allele dosage, and a slow advancement in evolutionary theory for polyploids (Glover et al., 2016; Gregg et al., 2017; Blischak et al., 2018; Meirmans et al., 2018). Restriction-site Associated DNA (RAD) tags generated by next-generation sequencing (NGS) provide sensitive markers from many loci sampled over the genome of any organism (Baird et al., 2008; Griffin et al., 2011; Peterson et al., 2012) and, given the use of an appropriate sequencing depth, may overcome problems associated with cloning and Sanger sequencing of nuclear genes in polyploid organisms (Qi et al., 2015). RADseq has been recently shown to be a valuable method in genotyping, evolutionary and ecological genomics, phylogeography, and phylogenetic studies of various groups including polyploid species (Brandrud et al., 2017, 2019a,b; Buggs et al., 2012; Lai et al., 2012; Lu et al., 2013; Frajman et al., 2019). The strategy to analyse RADseq data on non-model polyploid species often depends on the resources already available for the study system. When a reference genome is missing, but several extant diploid relatives are present, a catalog of RADseq tags can be built *de novo* based on those. Such a catalog can then be used for variant calling across all samples, taking into account their respective ploidy (Wang et al., 2014; Brandrud et al., 2019a). However, in cases where parents of polyploids may not be diploids, are extinct, or belong to groups that comprise only a few diploid accessions, a (qualitative) analysis of individuals *a priori* ignoring ploidy might be useful, which should be based only on biallelic SNPs rather than entire haplotypes (e.g. Qi et al., 2015). This may lead to underestimates of genetic diversity and some genealogical biases (e.g., Arnold et al., 2013), even if these are expected to have a negligible effect, given that a considerable number of loci is being analysed. Recently developed genotype-free methods of estimation of allele frequencies (Nielsen et al., 2011) maximize the information content of low to intermediate coverage data sets such as those generated by RADseq towards overcoming such biases in diploid samples (Korneliussen et al., 2014; Warmuth and Ellegren, 2019). Whereas efficiently accounting for uncertainty within NGS datasets, genotype-free methods can reduce bias in population genetics estimates at biallelic single nucleotide variants in polyploids (Blischak et al., 2018; Brandrud et al., 2019a). Finally, plastid DNA markers, which are maternally inherited in most angiosperms (Corriveau and Coleman, 1988), remain ideal for disentangling reticulate relationships in non-model systems independent of ploidy by complementing RAD tags with uniparental information (Vargas et al., 2017).

Here, we focus on the effect of range fragmentation and secondary contact on the heteroploid widespread Eurasian steppe plant *Astragalus onobrychis* (Fabaceae). This herbaceous perennial inhabits steppic grasslands from the foothills of the southwestern Alps in France to the Altai Mountains in southern central Siberia. The species is relatively abundant throughout its range; chromosome numbers range from $2n = 2x = 16$ over $2n = 4x = 32$ to $2n = 8x = 64$ (Magulaev, 1989; Rice et al., 2015). All three ploidy levels have been reported from the Caucasus area, while octoploids seem to predominate in Central Europe and a single diploid count has been published from the western Balkan Peninsula (Pustahija et al., 2013). Several taxonomic entities have been described; they were later either synonymised with *A. onobrychis* or retained as intraspecific taxa (Podlech, 2011), which underlines the species' high morphological variability.

We investigate the genetic, cytological and morphological variation of *A. onobrychis* across its distributional range in order to shed light on the factors and mechanisms driving the diversification of a widespread Eurasian steppe species, in particular the role played by hybridization and polyploidization, and to elucidate the effect of strongly alternating range size and connectivity during its evolutionary history. To this end we employed genome-wide (RADseq) as well as organellar (sequencing of the plastid *ycf1* region) markers, and analysed them in phylogenetic and phylogeographic frameworks. Further, we used relative genome size measurements and chromosome counts to infer ploidy, and multivariate morphometric analyses to quantify the amount of morphological divergence among the genetic and cytological groups. We ask the following specific questions: (1) How is the genetic and cytological diversity distributed across the species' range in *A. onobrychis* and what

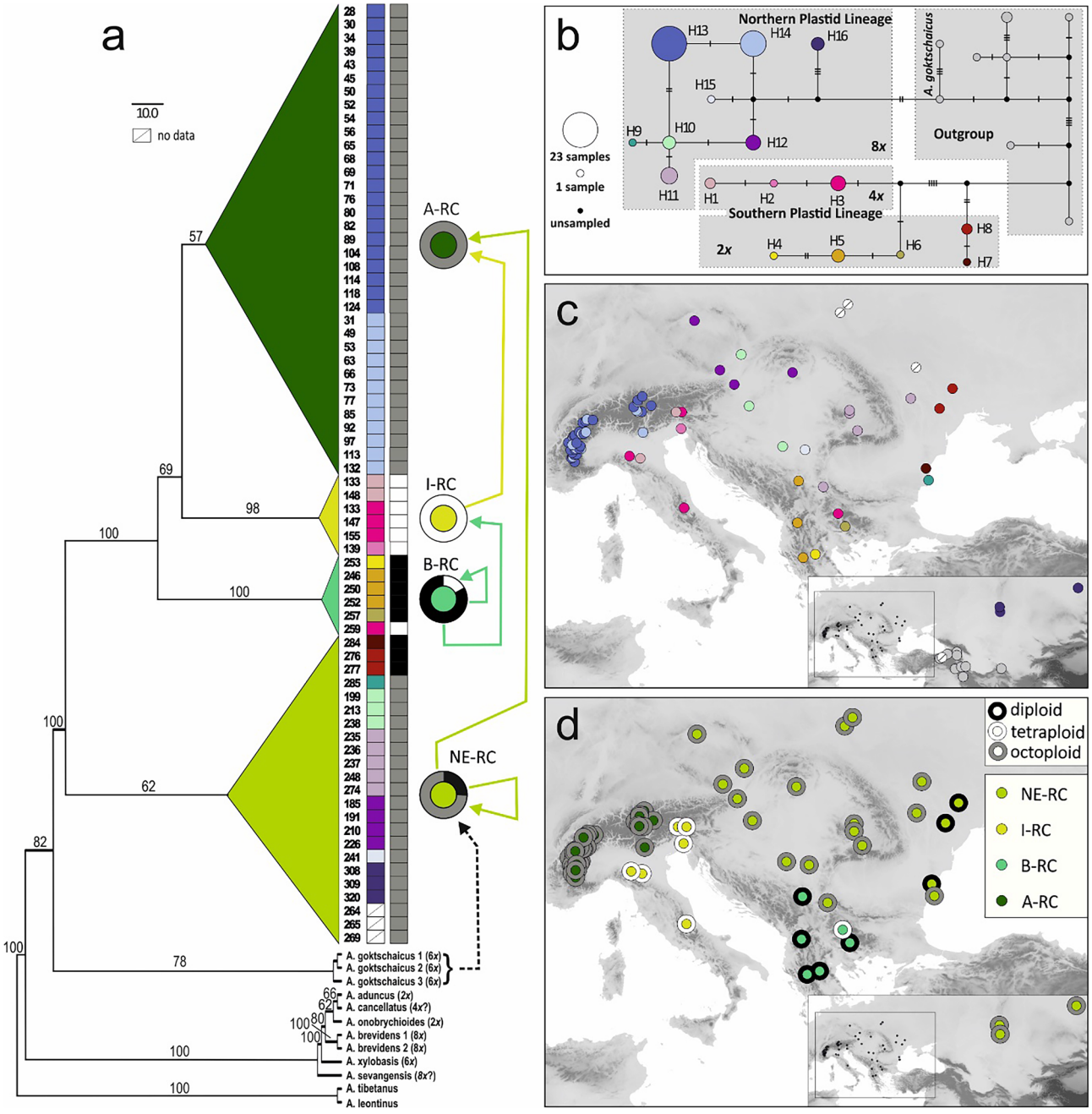

Fig. 1. Phylogenetic relationships in *Astragalus onobrychis* and other members of *A.* sect. *Onobrychoidei*. (a) Maximum likelihood tree based on 8730 SNPs derived from RADseq data. The tree was rooted with *A. leontinus* and *A. tibetanus*; labels at tips of branches are population identifiers corresponding to Supporting Table S1 and Supporting Fig. S1; numbers above branches are bootstrap support values. Nodes within the four major branches of *A. onobrychis* were collapsed (A-RC, Alpine RAD Clade; B-RC, Balkan RAD Clade; I-RC, Italian RAD Clade; NE-RC, Northeastern RAD Clade). The left vertical bar illustrates the plastid haplotypes, whose relationships are shown in (b), and the right vertical bar reflects ploidy (black, diploid; white tetraploid; grey, octoploid). Arrows represent reticulations inferred based on relatedness analyses (Fig. 2, Supporting Fig. S5a) and plastid DNA information. (b) Relationships among plastid DNA haplotypes illustrated as statistical parsimony network; unsampled haplotypes are shown as black dots. (c) Geographic distribution of plastid DNA haplotypes; populations for which we failed to obtain readable sequences are indicated. (d) Geographic distribution of ploidies (rings in greyscales) and of the four main clades (coloured dots) shown in (a). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is the spatio-temporal context of lineage diversification? (2) Are polyploid cytotypes auto- or allopolyploids? Did polyploidy drive range expansion? (3) On the methodological side we explored whether genotype-based and genotype-free analyses give highly comparable results. Finally, stimulated by the large number of taxonomic entities described within *A. onobrychis*, we asked, (4) if cytotypes and/or genetic groups in *A. onobrychis* differ consistently in morphology.

2. Material and methods

2.1. Plant material

In the years 2014 to 2016 we sampled 70 populations of *Astragalus onobrychis* throughout its range (Fig. 1 and Supporting Fig. S1). In addition, we sampled twelve outgroup populations, out of which ten from other members of *A.* sect. *Onobrychoidei* D.C. from Turkey, Armenia and

Iran and one population each of the more distantly related *A. leontinus* and *A. tibetanus* (Supporting Table S1). Altogether 82 populations have been analysed. Herbarium vouchers are stored at the herbarium IB. Outgroups were determined using national floras (Turkey: Chamberlain and Matthews, 1969; Ekici et al., 2015; former USSR: Komarov, 1948). For genetic analyses and relative genome size estimation, leaf material of at least four individuals per population was stored in silica gel.

In some populations only non-flowering individuals were encountered. Therefore, we complemented the 34 available flowering herbarium vouchers from 29 populations with additional herbarium vouchers and inflorescences (fixed in 70% ethanol) of 78 flowering individuals from 34 populations (herein termed additional populations A1–A34). The additional populations were selected to represent the three major RADseq clades present in Central Europe (see Results). From a subset of populations, seeds and/or living plants were collected and cultivated in the Botanical Garden of the University of Innsbruck for chromosome counts.

2.2. Ploidy estimation and chromosome counts

Relative genome size (RGS) was inferred from fluorescence intensities of 4′,6-diamidino-2-phenylindole (DAPI)-stained nuclei isolated from silica-gel dried leaf material following Suda and Trávnícek (2006) as described in detail in Supporting Methods S1. For all populations of *A. onobrychis* analysed with RADseq and the ten populations of other members of *A.* sect. *Onobrychoidei* the RGS of four individuals (average 3.96) was typically measured (Supporting Table S1). In addition, RGS was determined for one individual each from the additional populations A1–A34 as well as from three additional individuals, which were used for the determination of chromosome numbers. Chromosome numbers were determined for one individual from each RGS class (see Results), i.e., one individual each from populations 276, A17 and A25 (see Results section). A detailed protocol is given in Supporting Methods S1.

2.3. DNA extraction and RADseq

Total genomic DNA was extracted from dried leaf tissue (ca. 10 mg) using the modified CTAB protocol of Tel-Zur et al. (1999) and purified with the NucleoSpin gDNA clean-Up kit (Macherey-Nagel). Single-digest RADseq libraries were prepared using the restriction enzyme *Pst*I (New England Biolabs) and a protocol adapted from Paun et al. (2016). Briefly, we started with 110 ng DNA per individual and ligated 100 mM P1 adapters to the restricted samples. Shearing by sonication was performed with a M220 Focused-ultrasonicator (Covaris) with settings targeting a size range of 200–800 bp and mode at 400 bp (peak in power: 50, duty factor 10%, 200 cycles per burst and treatment time 90 s at 20 °C). One individual each of 82 populations (70 populations of *A. onobrychis* and twelve outgroup populations) were sequenced on Illumina HiSeq at VBCF NGS Unit (http://www.vbcf.ac.at/ngs/) as 100 bp single-end reads.

2.4. Variant calling and filtering

The raw reads were demultiplexed according to index barcodes using BamIndexDecoder v. 1.03 (from Illumina2bam package; available from https://github.com/wtsi-npg/illumina2bam) and based on inline barcodes with process_radtags.pl implemented in STACKS v. 1.35 (Catchen et al., 2011, 2013) with default settings.

As no reference genome for any *Astragalus* species was available, we started by applying the commonly used pipeline STACKS to build *de novo* a catalog of RADseq loci while optimizing the recovery of loci across the genomic heterogeneity present in the data. We first followed the methodology detailed by Heckenhauer et al. (2018) with minor modifications. For this, first, a catalog of RADseq loci was built using denovo_map.pl in STACKS using a minimum coverage to identify a stack of 10× (-m 10), a maximum number of differences between two stacks in a locus in each sample (-M) of three, and a maximum number of differences among loci to be considered as orthologous (-n) of three. We then used export_sql.pl from STACKS to extract from the catalog consensus sequences of the polymorphic loci with a maximum of ten SNPs per locus and with data present in at least 20% of individuals. These consensus sequences were retained to form a 'pseudo-reference' for mapping back all raw reads with the mem algorithm of bwa v. 0.7.5a-r405 (Li and Durbin, 2009). The mapping files were then sorted by reference coordinates and read groups were added with picard tools v. 2.0.1 (available from http://broadinstitute.github.io/picard). Realignments around indels have been performed for each bam file using the Genome Analysis Toolkit vers. 3.6-0-g89b7209 (McKenna et al., 2010). Finally, ref_map.pl from STACKS was applied to the mapped bam files with the requirement of a minimum coverage to identify an allele (-m) of 3×. For the final data filtering we used the software Populations from STACKS retaining the loci with maximum 10% missing samples, minimum minor allele frequency (–min_maf) 0.025 (i.e., corresponding to two individuals) and maximum observed heterozygosity (–max_obs_het) 0.65. The filtering on the number of heterozygous genotypes per locus was done to avoid combining paralogs within the same RAD locus. Datasets of SNPs from the filtered loci were exported in Phylip format to be used in RAxML (Stamatakis, 2014) using Populations flags "–phylip –phylip_var –phylip_var_all". To perform analysis of relatedness and PCA, data was exported in vcf format using the "–vcf" flag.

As an alternative approach we used the genotype-free method of ANGSD v. 0.910 (Korneliussen et al., 2014) on the indel-realigned mapping files to call variants implementing a GATK-derived algorithm (McKenna et al., 2010). This approach is expected to more accurately retain allele dosage information for polyploid samples within genotype likelihoods (Brandrud et al., 2019a). Here, only reads with base and mapping qualities of at least 20 were accepted (-minQ 20 -minMapQ 20). Major and minor alleles were inferred from genotype likelihoods (-doMajorMinor 1) and three possible genotypes were written in a beagle genotype likelihood format (-doGlf 2). Allele frequencies were estimated (-doMaf 2), filtering the variants by using a likelihood ratio test (LRT) to retain sites that are confidently polymorphic (-SNP_pval 1e-6). Finally, similar filters as used for STACKS-derived data were applied here, retaining sites with maximally 10% missing samples (-minInd 72) and if their minor allele was present in more than two individuals (-minMaf 0,025). A recent simulation study (Blischak et al., 2018) has shown that for high coverage datasets variant calling algorithms that assume either equal genotype probabilities (GATK-like) or Hardy-Weinberg equilibrium inherit negligible error rates for genotype estimates for polyploid datasets. The same study concludes that the use of genotype likelihoods can increase confidence for data affected by non-standard inheritance patterns and allelic dosage uncertainty, such as those including polyploids. Using the datasets obtained with these two alternative approaches (STACKS, ANGSD), we performed comparable analyses (relatedness, PCA/PCoA) in order to cross-evaluate the robustness of our inferences. A graphic representation of the workflow is illustrated in the Supporting Information (Supporting Methods S1).

2.5. Analyses of variant datasets

The SNPs called with STACKS were further processed as follows. To infer phylogenetic relationships among all 82 individuals we computed a maximum likelihood (ML) phylogeny using RAxML v. 8.2.8 (Stamatakis, 2014). Invariant sites were removed from the original phylip format using the script "deleteAlignColumn.pl" (available from https://www.biostars.org/p/55555/) and Felsenstein's ascertainment bias correction was further used to account for missing invariant sites as recommended (Leaché et al., 2015). Tree searches were done under a General Time Reversible model with categorical optimization of substitution rates (ASC_GTRCAT; Stamatakis, 2014). The best-scoring ML

tree was bootstrapped using 1000 replicates and the frequency-based stopping criterion (Pattengale et al., 2010).

For the 70 individuals of *A. onobrychis* and the ten individuals of outgroup species from *A.* sect. *Onobrychoidei* we used SplitsTree v.4.12.6 (Huson and Bryant, 2005) to create a NeighborNet based on Hamming distances (Hamming, 1950) to account for the potential presence of hybridization within the data set. Any ambiguous sites present in the SNP matrix were treated as an 'average' site, which means that the contribution at a site is averaged over all possible resolutions of the ambiguous codes (Huson and Bryant, 2018).

For the ingroup samples (70 individuals), a principal coordinate analysis (PCoA) was calculated with the R package ADEGENET v. 2.1.0 (Jombart, 2008; Jombart and Ahmed, 2011) based on Euclidean distances. For the same data set (but excluding the single tetraploid individual of the Balkan RAD Clade, see Results), we used the Populations program in STACKS to estimate the number of private alleles per group and the nucleotide diversity (π) per group. The private allele estimates were further corrected for small sample sizes by multiplying them by the factor $(N/(N-1))$, where N is the sample size of the respective group.

Starting with a vcf file we further calculated with VCFtools v. 0.1.15 (Danecek et al., 2011) a per-individual inbreeding coefficient *F* estimated using a method of moments (option –het). *F* can take values between −1 (lowest homozygosity) and 1 (highest homozygosity), but as it is calculated for individual rather than population samples and it is based only on a subset of polymorphic sites in the data, it should be considered as a relative measure. With VCFtools we further estimated the relatedness between each pair of individuals as the unadjusted Ajk statistic (Yang et al., 2010). Relatedness coefficients have been shown to be unbiased with respect to ploidy, even when allele dosage is missing (Meirmans et al., 2018). Ajk is expected to equal one when comparing a genotype to itself, to be close to zero when comparing individuals within the same genetic pool, and to take negative values when comparing accessions from different populations or species. The Ajk estimates were drawn as violin plots for four groups of polyploids, each from a different RAD Clade (i.e., as defined based on the ML tree) using Python v. 2.7.13 (van Rossum and Drake, 1995) and its libraries Pandas (McKinney, 2010), Matplotlib (Hunter, 2007) and Seaborn (https://doi.org/10.5281/zenodo.883859). The significance of pairwise differences in distributions between groups was estimated with Mann-Whitney tests using wilcox.test() in R v. 3.4.2 (R Development Core Team, 2017).

For the variants called with ANGSD, downstream analyses were performed directly on genotype likelihoods, as this approach has been shown to be more sensitive to weak structure than classical SNP calling analysis (Fumagalli et al., 2013); it also enables the recovery of SNPs with lower frequency of minor genotype that may be filtered out in genotype calling approaches when assuming diploid samples. We computed a covariance matrix using PCAngsd v. 0.95 (Meisner and Albrechtsen, 2018) that uses an iterative procedure based on the genotype likelihoods from ANGSD. Covariances were drawn as violin plots for each of the polyploid groups defined based on ML tree, similarly as (and comparable with) the Ajk relatedness values of the STACKS-based dataset. In order to allow for an additional comparison of the two variant calling approaches we performed a principal component analysis (PCA) on the covariance matrix of the ingroup samples with the R package ADEGENET v. 2.1.0 (Jombart, 2008; Jombart and Ahmed, 2011).

2.6. Amplification, sequencing and analyses of the plastid marker ycf1

Based on recommendations of Bartha et al. (2012) for the genus *Astragalus* we sequenced a ca. 1 kb portion of the plastid gene ycf1. Typically, one individual each (average: 1.3 individuals) of the 82 molecularly analysed populations (including 12 outgroup populations) was amplified and sequenced as described in Supporting Methods S1. Sequences were aligned using MAFFT (http://mafft.cbrc.jp/alignment/server/) and the alignment was improved manually in BioEdit v. 7.0.0. (Hall, 2004). A haplotype network was constructed using statistical parsimony as implemented in TCS 1.21 (Clement et al., 2000) using PopART (Leigh and Bryant, 2015). The analyses were run with the default parsimony connection limit of 95%; indels longer than 1 bp were reduced to single-base pair changes.

We conducted a Bayesian molecular dating analysis of the *A. onobrychis* populations and several distantly related *Astragalus* species with BEAST v. 1.7.5. (Baele et al., 2012; Drummond et al., 2012) using a (mean) clock rate of 4.5×10^{-3} substitutions per site per million years (Smith et al., 2007; Yamane et al., 2003) and a uniform prior to constrain the root age to a maximum age of 30 mya. Details about the analysis are given in Supporting Methods S1.

Finally, a continuous phylogeographic analysis using relaxed random walks (Lemey et al., 2010) was performed for the plastid data set using BEAST v. 1.8.1 (Drummond et al., 2012) as described in Supporting Methods S1. The method was originally used for epidemic data, but it is "a flexible statistical framework for biogeographical reconstructions that is easily expanded upon to accommodate various landscape genetic features" (Lemey et al., 2010, p. 1877) and was later applied in different biogeographic studies (e.g., Nylinder et al., 2014; Collevatti et al., 2015; Rešetnik et al., 2016). The maximum clade credibility tree with annotated diffusion estimates was visualized in the program SPREAD v. 1.0.6 (Bielejec et al., 2011) and projected together with polygons representing ancestral areas using ArcGIS 10.3. (ESRI, Redlands, California, USA).

2.7. Morphometric analyses

Analysis of morphological differentiation was performed across 112 individuals from 63 populations. Details about the 11 studied characters and the employed methods are given in Supporting Methods S1. Regrettably, sufficient material for morphometric analysis was only available for three of the four RADseq clades (Italian RAD Clade, Alpine RAD Clade and Northeastern RAD Clade; see Results), and not for any diploid samples. For each character, differences were compared by a one-factorial analysis of variance (ANOVA), which was followed by a Tukey's test at a significance level of $\alpha = 0.05$. Principal component analysis (PCA) of standardised variables and discriminant function analysis (DFA) implemented in the packages ade4 (Dray and Dufour, 2007) and MASS (Venables and Ripley, 2002), respectively, were conducted to display the overall variation pattern along the first two components. The grouping for the DFA followed the three above-mentioned RADseq clades. The probability of assigning an individual to a particular group based on its characters was calculated using the function predict.lda implemented in the MASS package. We conducted a multivariate analysis of variance (MANOVA) using Pillai trace statistics to test for differentiation between groups. Statistical analyses were performed using the software R v. 3.4.2 (R Development Core Team, 2017).

3. Results

3.1. Ploidy estimation and chromosome counts

Flow cytometry yielded high-resolution histograms and revealed three different ploidies among the populations of *A. onobrychis* (Fig. 1, Supporting Table S1, Supporting Fig. S2). Ploidy within populations was uniform and no odd-ploidy cytotypes were detected. Diploids (verified with a chromosome count from population 276: $2n = 16$; Supporting Fig. S3a) are restricted to the central and southern Balkan Peninsula (Albania, Serbia, Greece, Bulgaria; RGS 0.49 to 0.50) as well as the western Pontic plains (Romania, Moldova, Ukraine; RGS 0.58–0.59; Supporting Table S1). Tetraploids (RGS 0.96–0.99; count from population A17: $2n = 32$, Supporting Fig. S3b) extend from the

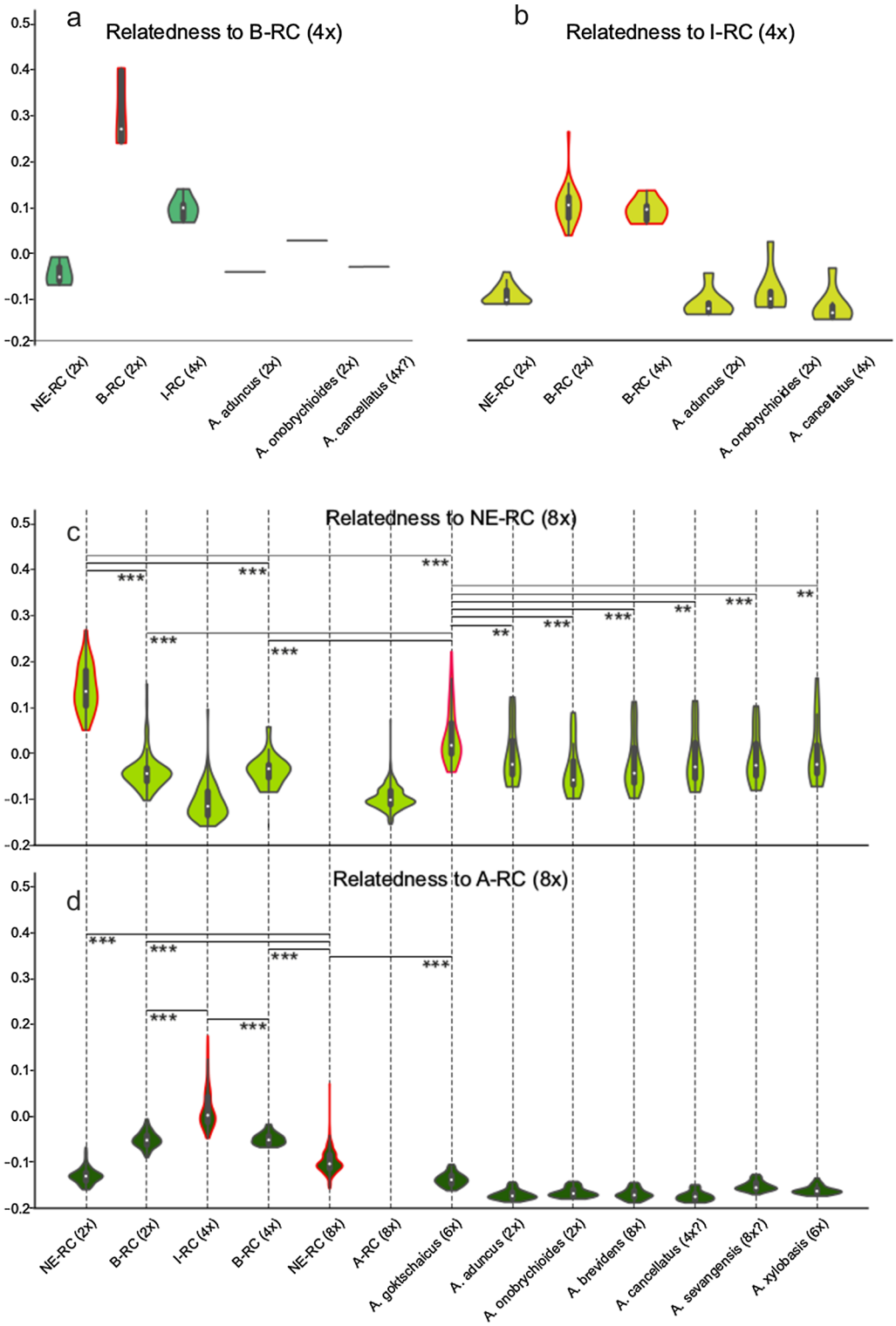

Fig. 2. Violin plots of pairwise relatedness among different *Astragalus* groups, calculated based on a covariance matrix from genotype likelihoods at 74,352 variant sites. Relatedness of all diploid and tetraploid individuals to (a) the tetraploid from the Balkan RAD Clade, B-RC, and (b) to the tetraploids from the Italian RAD Clade, I-RC. Relatedness of diploid, tetraploid, hexaploid and octoploid individuals to (c) octoploid individuals from the Northeastern RAD Clade, NE-RC, and (d) the Alpine RAD Clade, A-RC. Colours of violin plots correspond to the colour of the RAD Clade in Fig. 1 to which relatedness is computed. Hypothesized ancestors of particular polyploids are highlighted with a red outline. Tests of significance of pairwise differences in distributions between groups were not applicable in (a) and (b) due to the low number of observations. In (c) and (d) significances (***, p-value < 0.001; **, p < 0.05) are given for comparisons of hypothesized ancestors of a particular octoploid group to other groups. Note that in (d) the high relatedness of A-RC to the two B-RC groups is interpreted as result of their involvement in the ancestry of I-RC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Apennines to the southeastern Alps. Octoploids (RGS 1.84–2.01; count from population A25: $2n = 64$, Supporting Fig. S3c) are most widespread and span the extremes of the species' sampled distribution area (French Alps to Altai Mountains). Establishing the ploidy level of outgroup species of *A.* sect. *Onobrychoidei* was not straightforward in the absence of chromosome counts; ploidies ranged from diploid (*A. aduncus*, *A. onobrychioides*) over likely tetraploid (*A. cancellatus*) and hexaploid (*A. goktschaicus*, *A. xylobasis*) to likely octoploid (*A. sevangensis*) and octoploid (*A. brevidens*; Fig. 1a, Supporting Table S1).

3.2. RADseq

The average number of high-quality reads per sample retained after demultiplexing and quality filtering was 0.85 (SD = 0.33) million. The data have been deposited in the NCBI Short Archive (BioProject ID PRJNA510656, SRA accessions SAMN10612725–SAMN10612806).

After *de novo* assembly of RADseq loci and filtering, we retained 16,709 polymorphic loci in the pseudo-reference that was used for mapping. On average, the mapping success to this reference was 23.1%

(SD 2%). After the ref_map.pl catalog building, STACKS-based SNP calling and filtering, we selected up to 5145 RADseq loci present in at least 90% of the analysed samples and up to 11,158 high-quality SNPs, depending on the analysis (see details in Supporting Table S2). The average coverage per retained SNP was 33×. After variant calling using ANGSD, up to 74,352 variable sites were retained that were present in at least 90% of analysed samples (the exact percentage depended on the analysis; see details in Supporting Table S2).

The RAxML tree (Fig. 1a) rooted with *A. leontinus* and *A. tibetanus* resolved *A. onobrychis* as monophyletic with maximum bootstrap support; it was sister to *A. goktschaicus*. Within *A. onobrychis* four main clades were identified with bootstrap support higher than 50%. The Northeastern RAD Clade (BS 62%) constituted mostly by octoploids with a few additional diploids from the Pontic plains and the Pannonian basin (termed NE-RC in the following) was sister to the three other clades, which formed a maximally supported monophylum. Within the latter, the Balkan RAD Clade (B-RC; BS 100%) comprised diploid and a single tetraploid population from the Balkans. It was sister to a clade constituted by the exclusively tetraploid Italian RAD Clade (I-RC; BS 98%) and the exclusively octoploid Alpine RAD Clade (A-RC; BS 57%). In addition to the four *A. onobrychis* clades recovered with the RAxML tree, the NeighbourNet analysis shows a split (highlighted in Supporting Fig. S4) supporting a relationship between the NE-RC and most individuals of the A-RC.

The PCoA of the ingroup accessions, based on 11,158 SNP obtained by the STACKS-based pipeline allowed separating five genetic clusters along the first and second axes explaining together 22% of the variation in the data (Supporting Fig. S5). Particular clusters reflect the four groups resolved in the RAxML tree plus diploids of the NE-RC, which represent a fifth cluster. Similarly, the genotype-free PCA based on 67,596 variants recovered by ANGSD separated five clusters along the first two components explaining 12.6% of the variation (Supporting Fig. S5). Distances between clusters corresponding to the B-RC and the I-RC and between diploids and octoploids of the NE-RC are, however, smaller than in the PCoA.

The number of private alleles (Supporting Table S3) with reference to other ingroup samples ranged between 123 for tetraploids of the I-RC, 152–165 for the diploid groups, and 1306–1479 for the octoploid groups, indicating that a significant portion of the genomes of the octoploids is different from all other ingroup samples. Nucleotide diversity π increased in general with ploidy, except for tetraploids of the I-RC, which had π estimates within the range of the diploids as expected for autopolyploids. These patterns corroborate estimates of the per-individual inbreeding coefficient, *F*, which took values in a similar range for tetraploids of the I-RC and for the two diploid groups (Supporting Fig. S6). In contrast, both octoploid groups showed significantly increased heterozygosity, as indicative of an allopolyploid nature.

Estimates of Ajk relatedness based on the STACKS-derived data and the relatedness based on the covariance matrix resulting from the ANGSD data were directly compared using split violin plots (Supporting Fig. S5) and gave highly similar results. Therefore, in the following we only interpret the results of covariance-based relatedness (Fig. 2), because this approach fits better our heteroploid dataset. The single tetraploid individual from the B-RC showed the highest relatedness to diploids from the same clade. Tetraploids of the I-RC exhibited high relatedness to both diploids and tetraploids of the B-RC, suggesting that tetraploids of the I-RC are likely derived from tetraploids of the B-RC. Octoploids of the NE-RC exhibit a close relationship with diploids of the NE-RC and with the hexaploid *A. goktschaicus* suggesting an allopolyploid origin. For the exclusively octoploid A-RC the situation is less clear. The values of relatedness are generally lower, the highest and partly positive ones are exhibited by tetraploids of the I-RC – and thus also by their inferred ancestors – and by octoploids of the NE-RC and their inferred ancestors.

3.3. Plastid ycf1 sequences

The ycf1 sequences were 747–810 bp long (ingroup only: 753–810 bp); the alignment was 840 bp long (ingroup only: 816 bp). GenBank accession numbers are given in Supporting Table S1. In the ingroup dataset 17 variable characters and 15 indels were found, resulting in 4.3% variability. In total, 24 haplotypes were found, which formed two lineages (Southern Plastid Lineage, S-PL, H1–H8; Northern Plastid Lineage, N-PL, H9–H16) separated by at least 21 mutational steps (Fig. 1b). The two lineages did not form a monophylum but rather connect to eight distantly related haplotypes, recovered exclusively from the outgroup. The S-PL comprised all diploid and tetraploid populations from the I-RC and the B-RC as well as the three diploid populations from the NE-RC, whereas the N-PL comprised exclusively octoploid populations falling into the A-RC and the NE-RC. The next-related haplotypes to the N-PL were from *A. goktschaicus*. For populations 264, 265 and 269 we failed to obtain readable sequences most likely due to a mutation in a priming site.

The ycf1 Bayesian dated phylogeny with expanded outgroup sampling (Supporting Fig. S7) reflected the structure revealed by the parsimony network. Haplotypes of *A. onobrychis* constituting the N-PL (posterior probability, PP, 1, mean age 0.77 mya, highest posterior density interval, HDP, 0.29–1.46) formed a clade (PP 0.98, mean age 1.42 mya, HPD 0.52–2.75) with *A. aduncus*, *A. brevidens*, *A. cancellatus*, *A. goktschaicus* and *A. sevangensis*. This clade, which corresponds to *A.* sect. *Onobrychoidei*, was sister (PP 0.99, mean age 2.71 mya, HPD 1.05–5.25) to an unsupported group including *A. onobrychioides*, *A. xylobasis* and a clade constituted by *A. onobrychis* haplotypes of the S-PL (PP 1, mean age 1.08 mya; HPD 0.32–2.19).

The continuous phylogeographic analysis (Fig. 3) using relaxed random walks estimated the onset of geographical diversification of *A. onobrychis* and other members of *A.* sect. *Onobrychoidei* to be in the Caucasus (Fig. 3b; geographical toponymes are specified in Fig. 3a), where two lineages diversified; the first lineage is ancestral to the S-PL (onset of diversification 1.8 mya; Supporting Fig. S7) and the second lineage is ancestral to the N-PL (1.4 mya). The S-PL later diversified in the Balkans and spread to the western Pontic plains (1.0 mya; Fig. 3c) and to the southeastern Alps (0.5 mya; Fig. 3d), whereas the N-PL spread from the Caucasus to Central Asia and the Pannonian basin (0.8 mya; Fig. 3c) and from the latter further to the Alps (0.6 mya; Fig. 3d). Present distributions of S-PL and N-PL based on all sampled populations are shown in Fig. 3e.

3.4. Morphometry

One-factorial ANOVAs followed by a Tukey's test revealed significant ($\alpha = 0.05$) differences among the three RADseq clades in all eight characters (illustrated as box plot diagrams in Supporting Fig. S8). The PCA of the morphometric dataset (112 individuals, eight characters including three ratios; Fig. 4a) suggested a clear structure along the first two axes. The first axis (explaining 41.6% of the overall variation) was highly correlated with characters describing the length of flower bracts (BL), calyx (CL) and leaflets (LL) as well as the shape of the standard (SL/SW). The second axis (23.5%) was highly correlated with characters derived from pistil length (PiL), calyx tube width (CW) and wing shape (WL/WW). The discriminant function analysis yielded three non-overlapping groups (Fig. 4b), supporting morphological differentiation among tetraploids of the I-RC from the Alps, octoploids from the A-RC and octoploids from the NE-RC. Characters BL, SL/SW, CW, and LL contributed most to the separation along the first axis; the separation along the second axis was mostly governed by BL and CW. Assignment of individuals to a RADseq clade was on average correct in 96.4% of all cases (tetraploids from the Alps: 100%; octoploids from the Alps: 96.4%; octoploids from the Pontic-Pannonian area: 95.2%). The MANOVA indicated significant differentiation among the three groups (Pillai trace = 1.53, $p < 0.001$) and yielded significant differences for

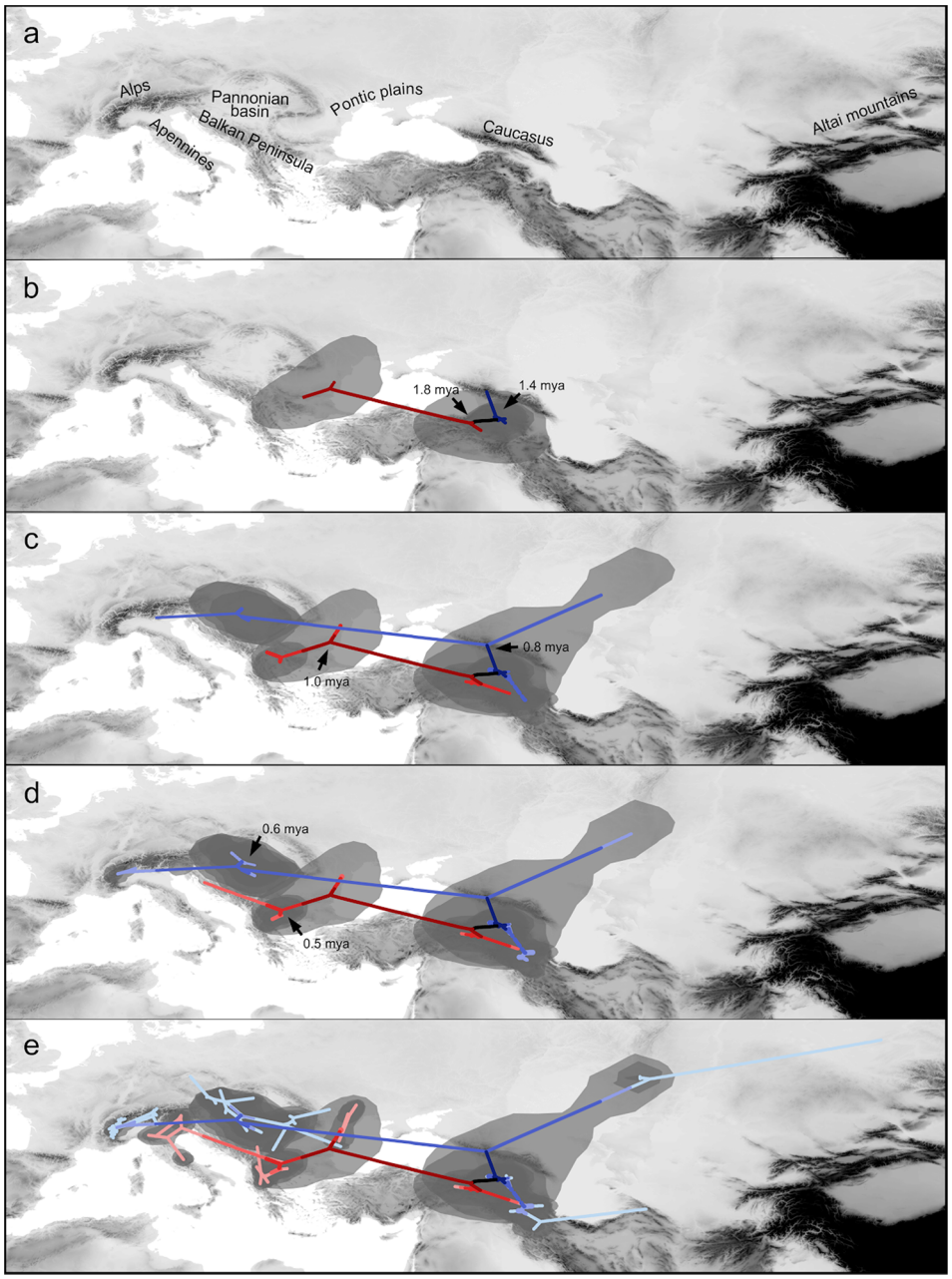

Fig. 3. (a) Toponyms used in this study. (b–e) Spatial diffusion of *Astragalus onobrychis* and other members of *A.* sect. *Onobrychoidei* illustrated as snapshots of estimated ancestral node areas in the maximum clade credibility tree of plastid DNA data at different time horizons as visualized using the software SPREAD. Lines illustrate migration events (blue: Northern Plastid Lineage; red: Southern Plastid Lineage). More intensive colours indicate older events; the colour coding thus does not reflect the plastid haplotypes but rather the sequence of events. Polygons represent 80% highest posterior density areas for nodes. Timing of diversification events is indicated at the respective nodes. Due to the lack of independent calibration points, the absolute dates should be viewed with appropriate caution. (e) Present distribution of plastid lineages (terminals represent all sampled populations). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

all eight characters (including three ratios). All characters had high discriminatory power at least between two of the three groups (Supporting Fig. S8).

4. Discussion

4.1. The evolution of Astragalus onobrychis was complex and reticulate

In *A. onobrychis* phylogenetic reconstructions exhibit striking incongruities between nuclear and plastid genomes; furthermore, the three ploidies and the genetic groups detected within the species are not fully overlapping. Phylogenetic analysis of mostly nuclear-derived RADseq data (Fig. 1a) resolved the species as monophyletic and revealed four moderately to well supported allopatric clades termed Balkan RAD Clade (B-RC), Italian RAD Clade (I-RC), Alpine RAD Clade (A-RC) and Northeastern RAD Clade (NE-RC). The three ploidies are largely allopatric and overlap with these clades only partially: diploids occur in the B-RC and the NE-RC within a relatively small area (southern Balkan Peninsula, western Pontic area), tetraploids – which have not been previously reported from Europe outside the Caucasus (Rice et al., 2015) – were found in the Apennines, the southeastern Alps and the southern Balkan Peninsula, and belong to the B-RC and the I-RC. Octoploids, which dominate the NE-RC and exclusively constitute the A-RC, are most widespread and span the extremes of the species' distribution area from southeastern France to the Altai Mountains in Central Asia. Plastid DNA, which is maternally inherited in closely related *A. cicer* (Harris and Ingram, 1991), identified two haplotype lineages within *A. onobrychis*, which are connected via haplotypes retrieved from outgroup species of *A.* sect. *Onobrychoidei* (Fig. 1b, Supporting Fig. S7). The Southern Plastid Lineage (S-PL) is characteristic for all diploids and tetraploids and likely comprises the non-introgressed set of plastid haplotypes of *A. onobrychis*, whereas the Northern Plastid Lineage (N-PL) found in all octoploid accessions was probably acquired by hybridisation of *A. onobrychis* as paternal parent with another species of *A.* sect. *Onobrychoidei* as maternal parent.

Using relatedness analysis (Fig. 2 and Supporting Fig. S5), one of the

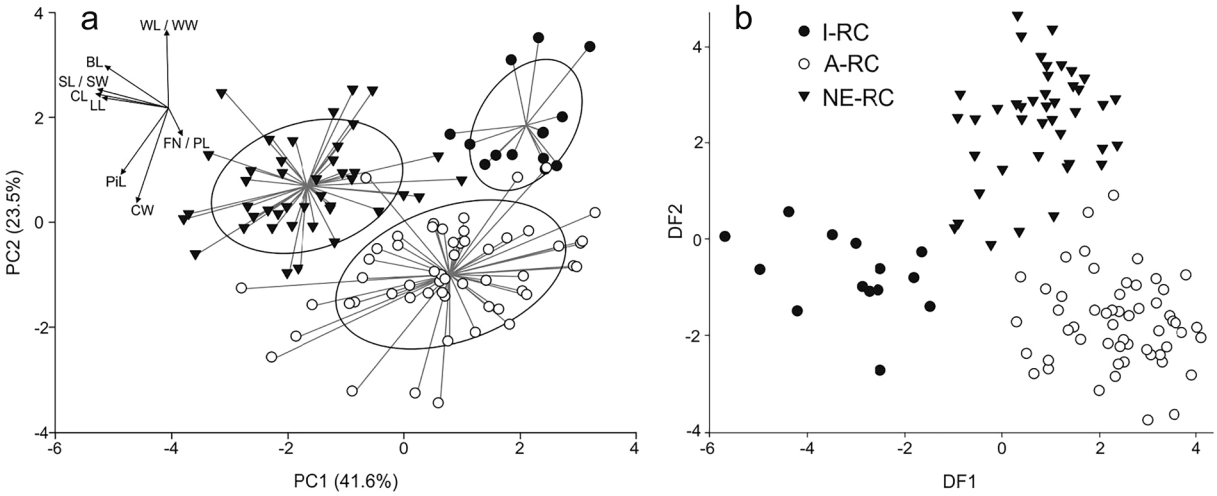

Fig. 4. Ordination diagrams illustrating the morphological variation among three of the four major RADseq clades delineated within *Astragalus onobrychis*. Black dots, I-RC, Italian RAD Clade; white dots, A-RC, Alpine RAD Clade; black triangles, NE-RC, Northeastern RAD Clade. Analyses are based on a dataset of 112 individuals and eight characters including three ratios. (a) Principal component analysis (PCA). Arrows in the inset represent the contribution of the characters to the overall explained variation. (b) Discriminant function analysis (DFA). The three a priori defined groups are those mentioned above.

few approaches that does not show ploidy-bias and is therefore applicable to heteroploid SNP data sets (Meirmans et al., 2018; Brandrud et al., 2019b), both tetraploid groups are of autopolyploid origin. However, according to the phylogenetic analysis (Fig. 1a) they likely have distinct histories. While the single tetraploid population of the B-RC is related only to diploids of the B-RC (Fig. 2a), the closest relatives of the exclusively tetraploid I-RC are either diploids or tetraploids from the B-RC. Additional arguments for an autotetraploid origin of the I-RC (for the single available tetraploid population of the B-RC no further statistics were calculated due to the small sample size) come from (i) their low nucleotide diversity, within the range of the diploid groups (Supporting Table S3), suggesting that only one gene pool participated in their origin; (ii) the per-individual inbreeding coefficients obtained (Supporting Fig. S6), which is not significantly different from diploid values, indicating a lack of widespread fixed heterozygosity otherwise expected for an allopolyploid; and (iii) the estimates of private alleles (Supporting Table S3), which are the lowest of all groups.

In contrast, both octoploid lineages likely result from allopolyploidization events. They show an increased nucleotide diversity (Supporting Table S3) and the lowest per-individual inbreeding coefficient *F* (Supporting Fig. S6), consistent with fixed heterozygosity. The octoploid groups also feature roughly ten times higher numbers of private alleles than related diploids and tetraploids from the ingroup, indicating a significant contribution to their genetic make-up from outside *A. onobrychis*. An important role of reticulate evolution in the history of both the NE-RC and the A-RC is also supported by plastid DNA. Specifically, diploid populations of the NE-RC carry haplotypes of the S-PL, whereas octoploids bear haplotypes falling into the N-PL crown group. It is likely, therefore, that diploids or (auto)tetraploids carrying haplotypes of the S-PL came into contact with bearers of the N-PL, probably in the Pontic plains, and gave rise to octoploids. It is unclear if the donor of the N-PL belonged to *A. onobrychis* or rather to another species of *A.* sect. *Onobrychoidei*, a likely candidate supported by both plastid DNA (Fig. 1) and RADseq (Fig. 2) being hexaploid *A. goktschaikus*. However, we stress that the lack of a range-wide sampling of all members of *A.* sect. *Onobrychoidei* may question the role of that particular species in the evolution of *A. onobrychis*.

Reconstructing the evolutionary history of the exclusively octoploid A-RC is not straightforward; it probably evolved via contact of tetraploids of the I-RC with octoploids present in the Alps or their southern and eastern periphery. Whereas relatedness analysis (Fig. 2d) and NeighbourNet diagram (Supporting Fig. S4) weakly indicate that these octoploids belonged to (ancestors of) the NE-RC, the plastid data clearly support such a relationship (Fig. 1b). Hybridisation with these octoploids, which acted as maternal parent, gave rise to allo-octoploid populations that later colonised the Alps. In any event, significant morphological differentiation between octoploids of the A-RC and the NE-RC (Fig. 4, Supporting Dataset S1, Supporting Fig. S8) supports the hypothesis of their independent origins.

Octoploid plants of *A. onobrychis* are taller-growing and have more and larger flowers per inflorescence than tetraploids (Supporting Fig. S8), which reflects patterns of morphological differentiation in other polyploid complexes (Flatscher et al., 2015). These properties may have increased the octoploids' ability to colonize and/or survive in different environments. This is supported by two lines of evidence. First, the A-RC has a much wider distribution in the Alps than the I-RC but both have colonised their present habitats postglacially, since these were covered by ice during the Last Glacial Maximum (van Husen, 1987; Ehlers and Gibbard, 2004). Second, octoploids of the NE-RC span a broad range throughout much of Eurasia, which was colonised late in the species' evolutionary history (Fig. 3). Altogether, the distribution of both octoploid lineages is in accordance with the hypothesis that polyploids tend to cover larger distribution areas and occupy more extreme and northern habitats than their diploid relatives (Comai, 2005; Van de Peer et al., 2017). This is generally explained by a breakdown of self-incompatibility (supported for *A. onobrychis* by preliminary crossing experiments by C. Maylandt), supplemented by increased heterozygosity and genetic buffering (Comai, 2005).

4.2. Plastid data provide a temporally explicit hypothesis of range dynamics

Time calibrated diffusion models applied to the plastid data suggest that bearers of ancestral haplotypes found in *A.* sect. *Onobrychoidei* were distributed in the southern Balkans, the Pontic plains and the Caucasus, and diverged into two lineages ca. 2.5 mya (Supporting Fig. S7). We stress that due to the lack of independent calibration points and the consequently wide HPD intervals the exact dates should be viewed with caution. This shortcoming should, however, not compromise the mere sequence of events. This coincides with the climatic cooling at the onset of the Pleistocene, thus placing this initial evolutionary divergence of *A. onobrychis* firmly in that period. Within the first lineage, the haplotypes retrieved from diploid *A. onobrychis* populations currently found in the Balkans and the western Pontic plains diverged ca. 1.8 mya from Caucasian ancestors belonging to *A.* sect. *Onobrychoidei*, giving

rise to the S-PL (Fig. 3b). Within the second lineage, the N-PL diverged – as well from Caucasian ancestors – ca. 1.1 mya (Supporting Fig. S7) and colonised areas north of the Caucasus. Roughly 1 mya, Balkan and western Pontic populations diverged within the S-PL (Fig. 3c). In the same line, diploids from the Pontic plains (populations 276, 277, 284) have plastid haplotypes that are closest to the outgroup in the parsimony network, and are sister to all other haplotypes of the S-PL in the Bayesian phylogeny (Supporting Fig. S7). Subsequently, diversification within the S-PL took place on the Balkans.

The N-PL crown group – which comprises exclusively octoploid populations – diversified ca. 0.8 mya (Fig. 3c). This time, therefore, marks the maximum age of the hybridisation event giving rise to the octoploids of the NE-RC, which was followed by a range expansion eastwards to Central Asia and westwards to the Pannonian basin. The two haplotypes present in the A-RC diverged from Pannonian haplotypes ca. 0.6 mya, probably during isolation of the Alpine populations in warm stage refugia during the Pleistocene (Fig. 3d). Shortly afterwards, around 0.5 mya, tetraploids carrying S-PL haplotypes diverged from Balkan diploids and migrated northwards to the southeastern Alps, and west- and southwards to and along the Apennines (Fig. 3d, e). As outlined above, at some point these tetraploids probably came into contact with octoploids carrying N-PL haplotypes, giving rise to the allo-octoploid populations distributed in the western and the central eastern Alps.

4.3. Methodological aspects

We have presented one of the first phylogeographic studies of a convolute heteroploid species complex using a RADseq approach. In such a complex system the application of currently established SNP calling approaches is not straightforward. Therefore, we have used two types of data analyses (genotype-based and genotype-free) to evaluate the consistency of the obtained results. The approach based on genotype likelihoods was able to retain six to seven times more variable sites while using similar filters (Supporting Table S2). From a practical point of view we emphasise that – at least in our study group – the general signal remained highly comparable across all analyses, despite different expected biases between the two analytical approaches.

Our study strongly suggests that even in species such as *A. onobrychis*, for which uniformity in ploidy was suggested for large parts of the distribution area (with a single exception only octoploids were reported from extra-Caucasian Europe; Pustahija et al., 2013; Rice et al., 2015), information on the ploidy of each investigated population is crucial for the interpretation of genetic or genomic data. Furthermore, it is obvious that plastid gene trees are highly complementary to RADseq data, and enable tracking spatio-temporal dynamics of maternal lineages as well as reconstructing past reticulations between lineages. Hence, we argue that plastid data should remain an integral part of evolutionary phylogenetic research in plants, as they may help uncover hidden reticulate histories (see also Garrick et al., 2015; Morris and Shaw, 2018).

4.4. Conclusions

The intersection of genomic, genetic, flow-cytometric and morphometric data clearly suggests that the latest taxonomic treatment of *A. onobrychis* (Podlech, 2011) underestimated its actual diversity. Here, we refrain from taking taxonomic conclusions; these should be based on deeper sampling across the Balkan Peninsula, including the type localities of all described taxa, a more complete morphometric dataset covering all genetic lineages, and, ideally, all described taxa. Finally, although not specifically addressed here with a modelling approach, it appears likely that the distribution of *A. onobrychis* in Europe during interglacials was similar to the current distribution. Albeit weakened by a low sampling density in the zonal steppe belt, our plastid DNA data show that the western Pontic plains, the Balkan Peninsula, the Apennines, the Alps and the Pannonian basin harbour distinct sets of haplotypes suggesting divergent evolution during successive phases of allopatry. Overall, genetic and phenotypic variation is unexpectedly high within the fragmented steppe areas to the west of the zonal steppe belt, suggesting existence of several warm-stage refugia. A forthcoming comparative phylogeographic paper based on several plant and animal taxa of the Eurasian steppes will show if the unravelled pattern is specific to the study species or a pervasive feature of the enigmatic extrazonal European steppes.

Funding

The present study was financed by the Austrian Science Fund (FWF, project P25955 “Origin of steppe flora and fauna in inner-Alpine dry valleys” to PS).

Acknowledgements

We thank the collectors of plant material listed in Supporting Table S1. Filippo Prosser kindly provided locality data from northern Italy. Various institutions issued collecting permits, which are cited in Supporting Table S1. Our thanks go to Juliane Baar for her introduction into the laboratory aspects of RADseq as well as to Daniela Pirkebner and Marianne Magauer for preparing the libraries. Barbara Pernfuss introduced CM to chromosome counting and Andreas Holzinger gave us access to his Axio Imager M2 microscope. Dirk Meyer allow us to use his M220 Focused-ultrasonicator (Covaris) and Agilent 2100 Bioanalyzer. We are most grateful to P. Daniel Schlorhaufer and his colleagues from the Botanical Gardens of the University of Innsbruck for successfully cultivating our living collection of *A. onobrychis*. Filip Kolář provided helpful suggestions on analyses and presentation of the results.

Declaration of Competing Interest

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

Author contributions

EZ and PS planned and designed the research, EZ, CM, OP, CB and BF performed the research, EZ, OP, CM and CB analyzed the data, and EZ, PS and OP wrote a first manuscript draft. Other members of The STEPPE Consortium collected samples, advised on RADseq analyses and improved drafts of the manuscript, which was finally approved by all authors.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.106572.

References

Arnold, B., Corbett-Detig, R.B., Hartl, D., Bomblies, K., 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol. Ecol. 22, 3179–3190.

Baele, G., Li, W.L.S., Drummond, A.J., Suchard, M.A., Lemey, P., 2012. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. Mol. Biol. Evol. 30, 239–243.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., et al., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS One 3, e3376.

Bartha, L., Sramkó, G., Dragoş, N., 2012. New PCR primers for partial ycf1 amplification in *Astragalus* (Fabaceae): promising source for genus wide phylogenies. Stud. Univ. Babes-Bolyai, Ser. Biol. Studia Biologia 57, 33–45.

Bielejec, F., Rambaut, A., Suchard, M.A., Lemey, P., 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27, 2910–2912.

Blischak, P.D., Kubatko, L.S., Wolfe, A.D., 2018. SNP genotyping and parameter

estimation in polyploids using low-coverage sequencing data. Bioinformatics 34, 407–415.
Bomblies, K., Madlung, A., 2014. Polyploidy in the *Arabidopsis* genus. Chromosome Res. 22, 117–134.
Brandrud, M.K., Baar, J., Lorenzo, M.T., Athanasiadis, A., Bateman, R.M., Chase, M.W., Hedrén, M., Paun, O., 2019a. Phylogenomic relationships of diploids and the origins of allotetraploids in *Dactylorhiza* (Orchidaceae). Syst. Biol. https://doi.org/10.1093/sysbio/syz035. (in press).
Brandrud, M.K., Paun, O., Lorenz, R., Baar, J., Hedren, M., 2019b. Restriction-site associated DNA sequencing supports a sister group relationship of *Nigritella* and *Gymnadenia* (Orchidaceae). Mol. Phylogenet. Evol. 136, 21–28.
Brandrud, M.K., Paun, O., Lorenzo, M.T., Nordal, I., Brysting, A.K., 2017. RADseq provides evidence for parallel ecotypic divergence in the autotetraploid *Cochlearia officinalis* in Northern Norway. Sci. Rep. 7, 5573.
Buggs, R.J., Renny-Byfield, S., Chester, M., Jordon-Thaden, I.E., Viccini, L.F., Chamala, S., et al., 2012. Next-generation sequencing and genome evolution in allopolyploids. Amer. J. Bot. 99, 372–382.
Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J.H., 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes Genom. Genet. 1, 171–182.
Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140.
Chamberlain, D.F., Matthews, V.A., 1969. *Astragalus* L. In: Davis, P.H. (Ed.), Flora of Turkey and the East Aegean Islands. vol. 3. Edinburgh University Press, Edinburgh, UK, pp. 49–254.
Cieślak, E., 2014. Phylogeography of Pontic-Pannonian species in Central Europe. Pol. Bot. Stud. 30, 1–53.
Clement, M., Posada, D.C.K.A., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1659.
Collevatti, R.G., Terribile, L.C., Rabello, S.G., Lima-Ribeiro, M.S., 2015. Relaxed random walk model coupled with ecological niche modeling unravel the dispersal dynamics of a Neotropical savanna tree species in the deeper Quaternary. Front. Plant Sci. 6, 653.
Comai, L., 2005. The advantages and disadvantages of being polyploid. Nat. Rev. Genet. 6, 836–846.
Comes, H.P., Kadereit, J.W., 1998. The effect of Quaternary climatic changes on plant distribution and evolution. Trends Plant Sci. 3, 432–438.
Corriveau, J.L., Coleman, A.W., 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. Am. J. Bot. 75, 1443–1458.
Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., et al., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.
Diallo, A.M., Nielsen, L.R., Kjær, E.D., Petersen, K.K., Ræbild, A., 2016. Polyploidy can confer superiority to West African *Acacia senegal* (L.) Willd. trees. Front. Plant Sci. 7, 821.
Dray, S., Dufour, A.B., 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22, 1–20.
Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.
Ehlers, J., Gibbard, P.L., 2004. Quaternary glaciations-extent and chronology: Part I: Europe, vol. 2, Elsevier, Amsterdam.
Ekici, M., Akan, H., Aytaç, Z., 2015. Taxonomic revision of *Astragalus* L. section *Onobrychoidei* DC. (Fabaceae) in Turkey. Turk. J. Bot. 39, 708–745.
Flatscher, R., Escobar García, P., Hülber, K., Sonnleitner, M., Winkler, M., Saukel, J., Schneeweiss, G.M., Schönswetter, P., 2015. Underestimated diversity in one of the world's best studied mountain ranges: the polyploid complex of *Senecio carniolicus* (Asteraceae) contains four species in the European Alps. Phytotaxa 213, 1–21.
Frajman, B., Záveská, E., Gamisch, A., Moser, T., The STEPPE Consortium, Schönswetter, P., 2019. Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Eurasian *Euphorbia seguieriana* s. l. (Euphorbiaceae). Mol. Phylogenet. Evol. 134, 238–252.
Fumagalli, M., Vieira, F.G., Korneliussen, T.S., Linderoth, T., Huerta-Sánchez, E., Albrechtsen, A., Nielsen, R., 2013. Quantifying population genetic differentiation from next-generation sequencing data. Genetics 195, 979–992.
Gabrielsen, T.M., Bachmann, K., Jakobsen, K.S., Brochmann, C., 1997. Glacial survival does not matter: RAPD phylogeography of Nordic *Saxifraga oppositifolia*. Mol. Ecol. 6, 831–842.
Garrick, R.C., Bonatelli, I.A., Hyseni, C., Morales, A., Pelletier, T.A., Perez, M.F., et al., 2015. The evolution of phylogeographic data sets. Mol. Ecol. 24, 1164–1171.
Gauckler, K., 1938. Steppenheide und Steppenheidewald der Fränkischen Alb in pflanzensoziologischer, ökologischer und geographischer Betrachtung. Botanisches Institut, Erlangen.
Glover, N.M., Redestig, H., Dessimoz, C., 2016. Homoeologs: what are they and how do we infer them? Trends Plant Sci. 21, 609–621.
Gregg, W.T., Ather, S.H., Hahn, M.W., 2017. Gene-tree reconciliation with MUL-trees to resolve polyploidy events. Syst. Biol. 66, 1007–1018.
Griffin, P.C., Robin, C., Hoffmann, A.A., 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. BMC Biol. 9, 19.
Hall, T., 2004. Bioedit 7.0. 0. North Carolina State University, Raleigh.
Hamming, R.W., 1950. Error-detecting and error-correcting codes. Bell Syst. Tech. J. 29, 147–160.
Harris, S.A., Ingram, R., 1991. Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. Taxon 393–412.
Heckenhauer, J., Samuel, R., Ashton, P.S., Salim, K.A., Paun, O., 2018. Phylogenomics resolves evolutionary relationships and provides insights into floral evolution in the tribe *Shoreeae* (Dipterocarpaceae). Mol. Phylogenet. Evol. 127, 1–13.
Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9, 90–95.
Huson, D.H., Bryant, D., 2005. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.
Huson, D.H., Bryant, D., 2018. User manual for SplitsTree4 V4.14.8. Available from: < http://ab.inf.uni-tuebingen.de/data/software/splitstree4/download/manual.pdf > .
Jännicke, W., 1892. Die Sandflora von Mainz, ein Relict aus der Steppenzeit. Gebr. Knauer, Frankfurt am Main.
Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.
Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071.
Kadereit, J.W., Griebeler, E.M., Comes, H.P., 2004. Quaternary diversification in European alpine plants: pattern and process. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359, 265–274.
Kajtoch, Ł., Cieślak, E., Varga, Z., Paul, W., Mazur, M.A., Sramkó, G., Kubisz, D., 2016. Phylogeographic patterns of steppe species in Eastern Central Europe: a review and the implications for conservation. Biodiv. Cons. 25, 2309–2339.
Kajtoch, Ł., Kubisz, D., Lachowska-Cierlik, D., Mazur, M.A., 2013. Conservation genetics of endangered leaf-beetle *Cheilotoma musciformis* populations in Poland. J. Insect Cons. 17, 67–77.
Kerner, A., 1863. Pflanzenleben der Donauländer. Universitäts Verlag Wagner, Innsbruck.
Komarov, V.L., 1948. Flora of the U.S.S.R., vol. 12, Israel Program for Scientific Translations, Jerusalem.
Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinform. 15, 356.
Kreiner, J.M., Kron, P., Husband, B.C., 2017. Frequency and maintenance of unreduced gametes in natural plant populations: associations with reproductive mode, life history and genome size. New Phytol. 214, 879–889.
Lai, K., Duran, C., Berkman, P.J., Lorenc, M.T., Stiller, J., Manoli, S., et al., 2012. Single nucleotide polymorphism discovery from wheat next-generation sequence data. Plant Biotechnol. J. 10, 743–749.
Leaché, A.D., Banbury, B.L., Felsenstein, J., De Oca, A.N.M., Stamatakis, A., 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. Syst. Biol. 64, 1032–1047.
Leigh, J.W., Bryant, D., 2015. PopART: full-feature software for haplotype network construction. Methods Ecol. Evol. 6, 1110–1116.
Lemey, P., Rambaut, A., Welch, J.J., Suchard, M.A., 2010. Phylogeography takes a relaxed random walk in continuous space and time. Mol. Biol. Evol. 27, 1877–1885.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Lu, F., Lipka, A.E., Glaubitz, J., Elshire, R., Cherney, J.H., Casler, M.D., et al., 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet. 9, e1003215.
Magulaev, A.J., 1989. Chromosome numbers in some *Astragalus* (Fabaceae) species of the Caucasus flora. Bot. Z. 74, 1519–1521.
Madlung, A., 2013. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110, 99–104.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.
McKinney, W., 2010. Data structures for statistical computing in Python. In: Millman, J., van der Walt, S. (Eds.), Proceedings of the 9th Python in Science Conference, SciPy, pp. 51–56. Available from < https://conference.scipy.org/proceedings/scipy2010/ > .
Meindl, C., 2012. New aspects in plant conservation – phylogeography, population dynamics, genetics and management of steppe plants in Bavaria. University of Regensburg (PhD thesis). Available from: < https://epub.uni-regensburg.de/23047/ > .
Meirmans, P.G., Liu, S., van Tienderen, P.H., 2018. The analysis of polyploid genetic data. J. Hered. 109, 283–296.
Meisner, J., Albrechtsen, A., 2018. Inferring population structure and admixture proportions in low depth NGS data. bioRxiv, 302463.
Meusel, H., Jäger, E., 1992. Vergleichende Chorologie der zentraleuropäischen Flora (vol. 3). Gustav Fischer Verlag, Jena.
Meusel, H., Jäger, E., Weinert, E., 1965. Vergleichende Chorologie der zentraleuropäischen Flora (vol. 1). Gustav Fischer Verlag, Jena.
Meusel, H., Jäger, E., Rauschert, S., Weinert, E., 1978. Vergleichende Chorologie der zentraleuropäischen Flora. Gustav Fischer Verlag, Jena.
Morris, A.B., Shaw, J., 2018. Markers in time and space: a review of the last decade of plant phylogeographic approaches. Mol. Ecol. 27, 2317–2333.
Neumann, K., Jansman, H., Kayser, A., Maak, S., Gattermann, R., 2004. Multiple bottlenecks in threatened western European populations of the common hamster *Cricetus cricetus* (L.). Conserv. Genet. 5, 181–193.
Neumann, K., Michaux, J.R., Maak, S., Jansman, H.A.H., Kayser, A., Mundt, G., Gattermann, R., 2005. Genetic spatial structure of European common hamsters (*Cricetus cricetus*) – a result of repeated range expansion and demographic bottlenecks. Mol. Ecol. 14, 1473–1483.
Nielsen, R., Paul, J.S., Albrechtsen, A., Song, Y.S., 2011. Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. 12, 443–451.
Nylinder, S., Lemey, P., De Bruyn, M., Suchard, M.A., Pfeil, B.E., Walsh, N., Anderberg, A.A., 2014. On the biogeography of Centipeda: a species-tree diffusion approach. Syst. Biol. 63, 178–191.
Paun, O., Turner, B., Trucchi, E., Munzinger, J., Chase, M.W., Samuel, R., 2016. Processes driving the adaptive radiation of a tropical tree (*Diospyros*, Ebenaceae) in New Caledonia, a biodiversity hotspot. Syst. Biol. 65, 212–227.

Pattengale, N.D., Alipour, M., Bininda-Emonds, O.R., Moret, B.M., Stamatakis, A., 2010. How many bootstrap replicates are necessary? J. Comput. Biol. 17, 337–354.
Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS one 7, e37135.
Petit, R.J., Aguinagalde, I., de Beaulieu, J.L., Bittkau, C., Brewer, S., Cheddadi, R., et al., 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300, 1563–1565.
Plenk, K., Bardy, K., Höhn, M., Thiv, M., Kropf, M., 2017. No obvious genetic erosion, but evident relict status at the westernmost range edge of the Pontic-Pannonian steppe plant *Linum flavum* L. (Linaceae) in Central Europe. Ecol. Evol. 7, 6527–6539.
Podlech, D., 2011. Thesaurus Astragalorum, index of all taxa described within the genus *Astragalus* L. (Taxa of the Old World and related taxa of the New World). Ludwig-Maximilians-Universität, München.
Pustahija, F., Brown, S.C., Bogunić, F., Bašić, N., Muratović, E., Ollier, S., et al., 2013. Small genomes dominate in plants growing on serpentine soils in West Balkans, an exhaustive study of 8 habitats covering 308 taxa. Plant Soil 373, 427–453.
Qi, Z.C., Yu, Y., Liu, X., Pais, A., Ranney, T., Whetten, R., Xiang, Q.Y., 2015. Phylogenomics of polyploid *Fothergilla* (Hamamelidaceae) by RAD-tag based GBS – insights into species origin and effects of software pipelines. J. Syst. Evol. 53, 432–447.
R Development Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Ramsey, J., 2011. Polyploidy and ecological adaptation in wild yarrow. Proc. Nat. Acad. Sci. USA 108, 7096–7101.
Ramsey, J., Schemske, D.W., 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu. Rev. Ecol. Syst. 29, 467–501.
Rešetnik, I., Frajman, B., Schönswetter, P., 2016. Heteroploid *Knautia drymeia* (Caprifoliaceae) includes *K. gussonei* and cannot be separated into diagnosable subspecies. Am. J. Bot. 103, 1300–1313.
Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N.M., Salman-Minkov, A., et al., 2015. The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers. New Phytol. 206, 19–26.
Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabath, N., Meiri, S., Belmaker, J., Mayrose, I., 2019. The global biogeography of polyploid plants. Nat. Ecol. Evol. 3, 265–273.
Schmitt, T., Varga, Z., 2012. Extra-Mediterranean refugia: the rule and not the exception? Front. Zool. 9, 22.
Smith, C.I., Pellmyr, O., Althoff, D.M., Balcazar-Lara, M., Leebens-Mack, J., Segraves, K.A., 2007. Pattern and timing of diversification in *Yucca* (Agavaceae): specialized pollination does not escalate rates of diversification. Proc. R. Soc. Lond. B Biol. Sci. 275, 249–258.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Stewart, J.R., Lister, A.M., Barnes, I., Dalén, L., 2010. Refugia revisited: individualistic responses of species in space and time. Proc. R. Soc. Lond. B Biol. Sci. 277, 661–671.
Stift, M., Berenos, C., Kuperus, P., van Tienderen, P.H., 2008. Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to *Rorippa* (yellow cress) microsatellite data. Genetics 79, 2113–2123.
Suda, J., Trávnícek, P., 2006. Estimation of relative nuclear DNA content in dehydrated plant tissues by flow cytometry. In: Robinson, J.P., Darzynkiewicz, Z., Dobrucki, J., Hyun, W., Nolan, J., Orfao, A., Rabinovitch P. (Eds.), Current Protocols in Cytometry, Wiley, New York, pp. 7.30.1–7.30.14.
Svenning, J.C., Skov, F., 2007. Ice age legacies in the geographical distribution of tree species richness in Europe. Glob. Ecol. Biogeogr. 16, 234–245.
Sztencel-Jabłonka, A., Mazgajski, T.D., Bury, S., Najbar, B., Rybacki, M., Bogdanowicz, W., Mazgajska, J., 2015. Phylogeography of the smooth snake *Coronella austriaca* (Serpentes: Colubridae): evidence for a reduced gene pool and a genetic discontinuity in Central Europe. Biol. J. Linn. Soc. 115, 195–210.
Tel-Zur, N., Abbo, S., Myslabodski, D., Mizrahi, Y., 1999. Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera *Hylocereus* and *Selenicereus* (Cactaceae). Plant Mol. Biol. Rep. 17, 249–254.
Van de Peer, Y., Mizrachi, E., Marchal, K., 2017. The evolutionary significance of polyploidy. Nat. Rev. Genet. 18, 411–424.
Van Husen, D., 1987. Die Ostalpen in den Eiszeiten. Geologische Bundesanstalt, Wien.
Van Rossum, G., Drake Jr., F.L., 1995. Python tutorial. Centrum voor Wiskunde en Informatica, Amsterdam.
Varga, Z., 2010. Extra-Mediterranean refugia, post-glacial vegetation history and area dynamics in Eastern Central Europe. In: Assmann, T., Habel, J.C. (Eds.), Relict Species. Springer, Berlin, pp. 57–87.
Vargas, O.M., Ortiz, E.M., Simpson, B.B., 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae: *Diplostephium*). New Phytol. 214, 1736–1750.
Venables, W.N., Ripley, B.D., 2002. Tree-based methods. In: Venables, W.N., Ripley, B.D. (Eds.), Modern Applied Statistics with S. Springer, New York, pp. 251–269.
Wagner, V., Durka, W., Hensen, I., 2011. Increased genetic differentiation but no reduced genetic diversity in peripheral vs. central populations of a steppe grass. Am. J. Bot. 98, 1173–1179.
Wahlberg, N., Saccheri, I., 2007. The effects of Pleistocene glaciations on the phylogeography of *Melitaea cinxia* (Lepidoptera: Nymphalidae). Eur. J. Entomol. 104, 675–684.
Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., et al., 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol. J. 12, 787–796.
Warmuth, V.M., Ellegren, H., 2019. Genotype-free estimation of allele frequencies reduces bias and improves demographic inference from RADseq data. Mol. Ecol. Res. 19, 586–596. https://doi.org/10.1111/1755-0998.12990.
Wesche, K., Ambarlı, D., Kamp, J., Török, P., Treiber, J., Dengler, J., 2016. The Palaearctic steppe biome: a new synthesis. Biodiv. Cons. 25, 2197–2231.
Winkler, M., Tribsch, A., Schneeweiss, G.M., Brodbeck, S., Gugerli, F., Holderegger, R., et al., 2012. Tales of the unexpected: phylogeography of the arctic-alpine model plant *Saxifraga oppositifolia* (Saxifragaceae) revisited. Mol. Ecol. 21, 4618–4630.
Yamane, K., Yasui, Y., Ohnishi, O., 2003. Intraspecific cpDNA variations of diploid and tetraploid perennial buckwheat, *Fagopyrum cymosum* (Polygonaceae). Am. J. Bot. 90, 339–346.
Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., et al., 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42, 565–569.