Annals of Botany **XX**: 1–13, 2018
doi: 10.1093/aob/mcy149, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Founded 1887

Unmasking cryptic biodiversity in polyploids: origin and diversification of *Aster amellus* aggregate

Mario Mairal[1,2,3,*], Mária Šurinová[1,2], Sílvia Castro[4] and Zuzana Münzbergová[1,2]

[1]Department of Botany, Faculty of Science, Charles University, 128 01 Prague, Czech Republic, [2]Department of Population Ecology, Czech Academy of Science, Zámek 1, 252 43 Průhonice, Czech Republic, [3]Department of Botany and Zoology, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa and [4]Centre for Functional Ecology, Department of Life Sciences of the University of Coimbra and Botanic Garden of the University of Coimbra, Calçada Martim de Freitas s/n, 3000–456 Coimbra, Portugal
**For correspondence. E-mail mariomairal@gmail.com*

Received: 7 March 2018 Returned for revision: 16 April 2018 Editorial decision: 15 July 2018 Accepted: 18 July 2018

• **Background and Aims** The origin of different cytotypes by autopolyploidy may be an important mechanism in plant diversification. Although cryptic autopolyploids probably comprise the largest fraction of overlooked plant diversity, our knowledge of their origin and evolution is still rather limited. Here we study the presumed autopolyploid aggregate of *Aster amellus*, which encompasses diploid and hexaploid cytotypes. Although the cytotypes of *A. amellus* are not morphologically distinguishable, previous studies showed spatial segregation and limited gene flow between them, which could result in different evolutionary trajectories for each cytotype.
• **Methods** We combine macroevolutionary, microevolutionary and niche modelling tools to disentangle the origin and the demographic history of the cytotypes, using chloroplast and nuclear markers in a dense population sampling in central Europe.
• **Key Results** Our results revealed a segregation between diploid and hexaploid cytotypes in the nuclear genome, where each cytotype represents a monophyletic lineage probably homogenized by concerted evolution. In contrast, the chloroplast genome showed intermixed connections between the cytotypes, which may correspond to shared ancestral relationships. Phylogeny, demographic analyses and ecological niche modelling supported an ongoing differentiation of the cytotypes, where the hexaploid cytotype is experiencing a demographic expansion and niche differentiation with respect to its diploid relative.
• **Conclusions** The two cytotypes may be considered as two different lineages at the onset of their evolutionary diversification. Polyploidization led to the occurrence of hexaploids, which expanded and changed their ecological niche.

Key words: Cryptic diversity, autopolyploidy, cytotypes, diversification, ecological niche, reproductive isolation, *Asteraceae*.

INTRODUCTION

Polyploidy or whole-genome duplication (WGD) is known as a major mechanism of adaptation and speciation in evolutionary history. Numerous WGD events have been detected in the last 500 million years in many eukaryotic taxa (Wendel, 2000; Van de Peer, 2017), being especially widespread in plants (Stebbins, 1970; Soltis and Soltis, 2000; Wendel, 2015) in contrast to most groups of animals (Alix *et al.*, 2017). It has in fact been established that all flowering plants have experienced episodes of polyploidization (Masterson, 1994; Wood *et al.*, 2009; Jiao *et al.*, 2011), with WGD driving plant evolution (Alix *et al.*, 2017).

Polyploids are traditionally classified as either autopolyploids, which arise within a single taxonomic species, or allopolyploids, which are the product of interspecific hybridization. While allopolyploidy has been extensively studied (Müntzing, 1932; Feldman and Levy, 2005; Catalán *et al.*, 2012; Barker *et al.*, 2016), autopolyploidy has received little attention in the past, as it was expected to be very rare and maladaptive in natural populations (Stebbins, 1970; Grant, 1981; Arrigo and Barker, 2012). Additionally, many autopolyploids have escaped recognition because they are morphologically similar to their diploid progenitors (Stebbins, 1947; Soltis *et al.*, 2007, 2010; Parisod *et al.*, 2010; Husband *et al.*, 2013). The sum of these factors has suggested that the diversity of autopolyploids has been underestimated (Soltis *et al.*, 2007; Parisod *et al.*, 2010). Recently it has been shown that cryptic polyploids may comprise the largest fraction of overlooked plant diversity (approx. 51 000–61 000 cryptic polyploid species; Barker *et al.*, 2016). Most of these polyploids would correspond to autopolyploids, which fulfil most of the species concepts (biological, evolutionary, phylogenetic and apomorphic; Soltis *et al.*, 2007; Barker *et al.*, 2016). Although available data show that autopolyploids are more numerous than previously thought, our knowledge of their origin and evolution is still very scarce (Barker *et al.*, 2016; Kolář *et al.*, 2017; Van de Peer, 2017).

Autopolyploidy could be an important driver of plant evolution and genetic differentiation (Ramsey *et al.*, 2008; Parisod *et al.*, 2010; Soltis *et al.*, 2016). However, the effects of autopolyploidy

on evolutionary divergence are barely known and detailed studies exploring cytotype origins and diversification are still necessary (Kellogg, 2016; Van de Peer, 2017; Kolář *et al.*, 2017). Polyploid origin may be explained by two alternative models: the single-origin model and the multiple-origin model. Under the single-origin model, an *n*-ploid cytotype is expected to arise only once from the diploid cytotype. This model also proposes that the polyploid cytotype originated a long time ago to allow its spread through its modern range (Halverson *et al.*, 2008). In contrast, under the multiple-origin model, it is expected that the cytotypes arose independently several times in different populations. The relative importance of these two models for the origin of autopolyploids is, however, not clear.

Here, we reconstruct the phylogenetic relationships, demographic history and ecological niche of diploid and presumed autohexaploid cytotypes of *Aster amellus*, a plant species widely distributed in central and eastern Europe (Münzbergová *et al.*, 2011). The genus *Aster sensu stricto* contains approx. 180 species, typified by *A. amellus* and restricted to the Northern Hemisphere of the Old World (Nesom, 1994). *Aster amellus* has been defined as an aggregate grouping diploid and hexaploid cytotypes, with a few non-fertile minority cytotypes found on very rare occasions (see Mandáková and Münzbergová, 2006; Castro *et al.*, 2012). Although both cytotypes grow in close proximity, each natural population is composed of reproductive plants with only one ploidy level (based on flow cytometric analyses of >7000 individuals in 327 populations) and only a mixed population has been found so far (see Strebersdorf population in Fig. 1; Mandáková and Münzbergová, 2006; Castro *et al.*, 2012); a distribution known as 'mosaic parapatry'. Central European populations show a cytogeographical structure: while diploid populations are distributed throughout most of the European area, hexaploid populations are longitudinally restricted, appearing exclusively east of Germany (Fig. 1; Castro *et al.*, 2012). In addition, a large contact zone including populations of both cytotypes exists across Poland, the Czech Republic, Slovakia, Austria and Romania. Despite this distribution, fertile intermediate ploidy levels (tetraploids) have not been detected so far. This may be attributed to their reproductive isolation due to spatial segregation, strong post-pollination barriers and, to a lesser extent, temporal and behavioural segregation (Castro *et al.*, 2011, 2012). An additional study using seven microsatellite loci (Münzbergová *et al.*, 2013) showed no or very limited gene flow between the cytotypes. Although this suggested that both cytotypes could be evolving independently, the evolutionary history of the group is not yet clear.

The extensive contact zone of *A. amellus* together with the particular spatial segregation of the cytotypes and all the previous knowledge already accumulated on this system makes *A. amellus* an ideal candidate to explore polyploid origin and its subsequent evolutionary history. Moreover, if the cytotypes within this aggregate show divergent histories (as suggested in Münzbergová *et al.*, 2013), it would offer us a unique opportunity to investigate both the origin and the demographic trajectories of the different cytotypes. To our knowledge, no study has tested, at the same time, the origin, divergence times, ecological niche and demographic imprints of an autopolyploid complex. This knowledge is important to gain insights into the connection between autopolyploidy and diversification (Kellogg, 2016).

In this study, we reconstructed the intraspecific-level phylogeny and phylogeography of the *A. amellus* aggregate using evidence from the nuclear ribosomal (nrDNA) internal transcribed spacer (ITS) region and non-coding chloroplast (pDNA) markers utilizing a large sample of populations – covering most of the species range. We combined Bayesian methods, population genetic analyses, statistical parsimony and ecological niche modelling tools to disentangle the phylogeographic distribution patterns of the cytotypes of the *A. amellus* aggregate. Our aims were to: (1) reconstruct the relationships and origin(s) of the diploid and hexaploid cytotypes and (2) search for ongoing differentiation in demography and ecogeographical niche between the cytotypes.

MATERIALS AND METHODS

Cytotype distribution and evidence for autopolyploidy

Previous literature surveys and massive flow cytometric screening of *A. amellus* populations across Europe revealed that most populations present only one cytotype, either diploid ($2n = 2x = 18$ chromosomes) or hexaploid ($2n = 6x = 54$), rarely accompanied by minority cytotypes (e.g. triploids, tetraploids, heptaploids and nonaploids) of which no breeding adults have been found (Mandáková and Münzbergová, 2008; Münzbergová *et al.*, 2011; Castro *et al.*, 2012).

Several lines of evidence, including data from cytology, isozymes and morphology, suggest that the hexaploid cytotype of *A. amellus* is of autopolyploid origin (Mandáková and Münzbergová, 2008; Castro *et al.*, 2012; Münzbergová *et al.*, 2013). Specifically, autopolyploidy is suggested by close morphological resemblance between the diploids and hexaploids (Mandáková and Münzbergová, 2008). In addition, allozyme analyses showed that the two cytotypes possess similar arrays of allozymes at all polymorphic loci and there was no evidence for fixed heterozygosity in the hexaploids [fixed heterozygosity is expected in allopolyploids, although it may be absent in diploidized allopolyploids (Mandáková and Münzbergová, 2008)]. Besides, no signal of allopolyploidy has been detected in the analyses of the karyotype of the species (Jarolímová V. *et al.*, unpubl. res.). Similarly, Münzbergová *et al.* (2013) also supported autopolyploidy by very similar microsatellite profiles in the two cytotypes, though they strongly differed in the frequencies of these microsatellites. Although there are many methods allowing confirmation of the allopolyploid origin of a species (e.g. Rosato *et al.*, 2008; Cuadrado *et al.*, 2017), there are no clear-cut methods allowing such a definitive confirmation for autopolyploids. This makes proving autopolyploidy extremely difficult, and such definite support is largely impossible in most systems (for an interesting approach, see Catalán *et al.* 2006). Although, separately none of our previous data provide unequivocal evidence of autopolyploidy in the system, all the evidence together indicate that this is the most likely explanation of the patterns observed.

Taxon sampling and DNA sequencing

We used cytotype screening for 327 populations (inset in Fig. 1) performed in our previous studies (Mandáková and Münzbergová, 2006; Castro *et al.*, 2012) to identify the populations for this study and to perform ecological niche modelling. Based on this knowledge, we collected leaf material for 102 individuals in several field expeditions throughout central

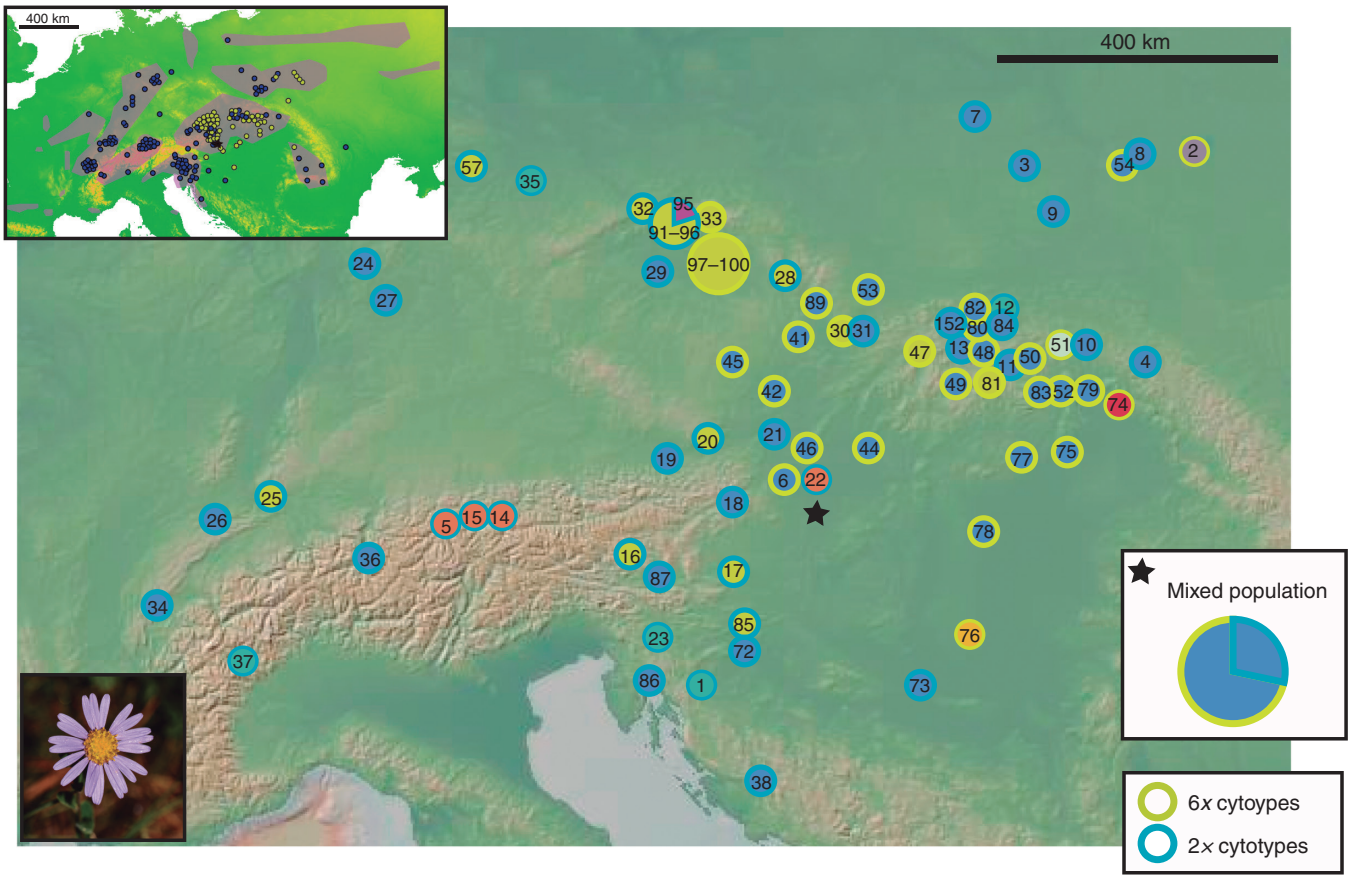

FIG. 1. Geographical location of the 72 populations used in the phylogenetic analyses. Numbers correspond to the sampled populations, with codes given in Supplementary Data Table S1. Pie charts are surrounded in blue for the diploids and in green for the hexaploids, while its circle size is proportional to population sampling. The colour filling each pie chart shows the chloroplast haplotypes, corresponding to the haplotypes shown in Fig. 4A. The star and the inserted box represent the only mixed-ploidy population of Strebersdorf. The inset 327 populations with known cytotype used in this study, coloured in blue for the diploids and in green for the hexaploids. The approximate distribution of *A. amellus* is shaded in purple in the inset. Maps have been modified from GeoMapApp (Ryan *et al.*, 2009; www.geoMapApp.org).

Europe between 2008 and 2011, representing 72 populations within the *Aster amellus* aggregate (Fig. 1; Supplementary Data Table S1) and covering the entire contact area of the cytotypes and adjacent areas throughout central Europe. Note that while the distribution range is wider than central Europe, samples from the eastern part of the distribution range are not available for this study. As we attempted to cover the contact zone completely, this does not limit the quality of our data set. Our previous knowledge, based on microsatellites (Münzbergová *et al.*, 2013 and an ongoing study) showed very low variation within populations, and we thus used mainly one individual per population. While we acknowledge that one individual per population is a low number, the basic patterns detected using nuclear markers correspond to those previously obtained with a large sample using microsatellites (Münzbergová *et al.*, 2013). This suggest that this sampling did not lead to a great loss of information. A special sampling effort was made in the Strebersdorf population (n = 22), the only mixed-ploidy population that has been detected so far (Castro *et al.*, 2012). Additional samples of a few other populations were also included (see Supplementary Data Table S1). In all, 596 sequences (ITS + pDNA) for 102 individuals were generated for this study. Species information and GenBank accession numbers for all the sequences are provided in Supplementary Data Table S1. DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN Inc., CA, USA) following the manufacturer's instructions from silica gel-dried leaves obtained from the specimens.

Several species representing hierarchical levels of phylogenetic relationships in the order Asterales were selected as alternative outgroups in the phylogenetic and dating analyses: Fam. Goodeniaceae (*Goodenia*, *Scaveola* and *Verreauxia*), Fam. Calyceraceae (*Acicarpa*, *Boopis*, *Calycera* and *Nastanthus*) and Fam. Asteraceae (*Aster*, *Bellis*, *Calendula*, *Conyza*, *Crinitina*, *Erigeron*, *Galatella*, *Grangea*, *Myriactis*, *Solidago* and *Tripolium*). For this, we downloaded 37 accessions from GenBank (Supplementary Data Table S1).

A comprehensive pilot study was carried out to select variable regions within the chloroplast genome. We first tested 17 plastid markers (reported as highly variable in Dong *et al.*, 2012), of which we selected five non-coding plastid regions: *atp*I–*atp*H, *rps*16 3′ exon–*trn*K (UUU) 5′ exon, *rpl*32–*trn*L and *Pet*N–*psb*M, which exhibited high levels of genetic variation, and *Psb*E–*pet*L with moderate variation (see Table 1). Amplification of these regions was as follows: 10 µL PCRs

TABLE 1. *Summary statistics of the chloroplast and nuclear regions analysed here for the* Aster amellus *markers (no outgroups)*

Aster amellus markers	*rps*16–*trn*K	*atp*I–*atp*H	*pet*N–*psb*M	*psb*E–*pet*L	*rpl*32–*trn*L	ITS
Unaligned length (bp)	874–919	995–1042	617–618	400–623	775–792	540–641
Aligned length (bp)	919	1042	618	623	814	642
Constant sites	902	1021	610	622	800	626
Variable sites	17	21	8	1	14	16
Haplotype diversity	0.48	0.483	0.497	0.503	0.533	0.74
Nucleotide diversity	0.0034	0.0035	0.0038	0.0013	0.0039	0.00705

Fragment length is given in bp; alignment length includes the indels.

contained 5 μL of QIAGEN Multiplex PCR Master Mix, 0.5 μL of each primer (10 mmol L^{-1} each in initial volume), 3 μL of double-distilled H_2O (ddH_2O) and 20 ng of DNA dissolved in 1 μL of ddH_2O. Pre-denaturation for 15 min at 95°C was followed by 37 cycles of 95 °C/30 s, 52 °C/30 s and 72 °C/2 min, and a final extension step at 72 °C for 10 min.

Additionally, we sequenced the nuclear marker, ITS. For amplification, we used ITS4 primer and modified ITS5 primer as described in Noyes and Rieseberg (1999). Amplification of ITS regions was as follows: 30 μL PCRs contained 15 μL of QIAGEN Multiplex PCR Master Mix, 0.9 μL of each primer (10 mmol L^{-1} each in initial volume), 12.2 μL of ddH_2O and 20 ng of DNA dissolved in 1 μL of ddH_2O. Pre-denaturation for 15 min at 95 °C was followed by 27 cycles of 95 °C/60 s, 60 °C/60 s and 72 °C/2 min, and a final extension step at 72 °C for 10 min. For each sample, both strands were directly sequenced.

Forty-eight individuals of *A. amellus* revealed double bands and unreadable electrophoretograms. For these, we followed the guidelines to obtain reliable ITS sequences in plants (Feliner and Rosselló, 2007) and used a cloning strategy when necessary to end up with consensual sequences. Cloning procedure included excision of the PCR products from 1 % agarose gels and purification with the Zymoclean Gel DNA Recovery kit (Zymoresearch, Orange, CA, USA). These fragments were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions, but downscaled to half-reactions. Twenty colonies per sample were transferred into 20 μL of ddH_2O, denaturated at 95 °C for 10 min and used as templates for subsequent PCR amplifications for sequencing. Sequence variability within clones of one sample ranged between 0.4 and 1.6 %. All the 20 sequences from the cloning procedure were merged into one consensus sequence per individual, which was used for all the phylogenetic analyses.

All reactions were run on Eppendorf Mastercycler Pro S (Eppendorf, Hamburg, Germany). PCR products were purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany) and sequenced (Seqme, Dobříš, Czech Republic). The five plastid DNA loci selected, together with the nuclear ribosomal ITS, were successfully amplified using the primers listed in Supplementary Data Table S2. Sequences were edited in Geneious 10.1.3, and aligned using the 'global alignment' option with free end gaps, and manually adjusted when necessary following alignment rules described by Kelchner (2000).

To address different objectives, we constructed three data sets: (1) the 'pDNA outgroups data set' ($n = 21$), which included accessions of the concatenated pDNA sequences for the Asterales outgroups plus several accessions of *Aster* (*A. alpinus*, *A. amellus*, *A. foliaceus*, *A. lavandulifolius* and *A. tongolensis*); (2) the 'amellus ITS data set' ($n = 102$), which included at least one ITS accession of each population within the *A. amellus* aggregate plus one sequence of *A. alpinus* as outgroup [identified as the sister species of *A. amellus* (Li *et al.*, 2012)]; and (3) the 'amellus pDNA data set' ($n = 100$), which included at least one concatenated pDNA accession of each population within the *A. amellus* aggregate, plus one sequence of *A. alpinus*.

Phylogenetic inference and divergence time estimation

Phylogenetic relationships were estimated for each marker separately using Bayesian inference, implemented in MrBayes 3.2.2 (Ronquist *et al.*, 2012). Choice of substitution models was based on the Akaike information criterion implemented in JModelTest 2.2 (Posada, 2008): the models selected for each region are specified in Supplementary Data Table S3. Two independent analyses of four chains each were run for a minimum of 10 million generations, sampling every 1000th chain. Convergence was assessed by monitoring cumulative split frequencies. After discarding the first 25 % of samples as burn-in, we pooled the remaining trees to construct a 50 % majority rule consensus tree. Additionally, maximum likelihood analyses were run in the software RAxML (Stamatakis *et al.*, 2008) using the online tool (http://embnet.vital-it.ch/raxml-bb/). The 'amellus ITS data set' and the 'amellus pDNA data set' were rooted using *A. alpinus* as outgroup. Before concatenating the plastid genes into a combined data set, we checked for topological congruence in the inferred relationships by examining the Bayesian consensus trees and searching for well-supported clades [posterior probability (PP) >0.95]. For the final analyses, we concatenated the five plastid regions into a combined pDNA matrix: the 'amellus pDNA data set'. As the 'amellus ITS data set' did not support the same signal, we analysed it separately. The concatenated data matrix was analysed under the GTR + G model, partitioned by gene and allowing the overall mutation rate to differ among partitions. To support further the results based on our single nuclear marker, we constructed a Neighbor–Joining dendogram using the nuclear microsatellite data set developed in our previous study (Münzbergová *et al.*, 2013). While the data have already been published, they were not previously used for dendrogram construction. Here we wanted to clarify whether the results based on the microsatellites correspond to those based on the ITS.

Simultaneous dating of plastid and nuclear genomes, which might have very different evolutionary rates, may lead to potential artefacts (Wolfe *et al.*, 1987). To avoid this and because the ITS marker may be subjected to concerted evolution, we performed the dating analysis only for the plastid data set. To

provide a temporal framework, lineage divergence times were estimated using the Bayesian relaxed-clock models implemented in BEAST v.1.7 (Drummond *et al.*, 2012). We were not interested in absolute time estimates but rather in obtaining molecular evidence for testing our hypothesis. Two Markov Chain Monte Carlos (MCMCs) were run for a minimum of 20 million generations. We used Tracer v.1.6 (Rambaut *et al.*, 2013) to monitor convergence and EES values (>200) for all parameters, and TreeAnnotator v. 1.7 (Rambaut and Drummond, 2013) to construct a maximum clade credibility tree from the posterior distribution after discarding 20 % of samples as burn-in.

There are no suitable fossils for *Aster*, so we relied on two approaches to estimate lineage divergence times. While the two approaches have their limitations (explained in detail in Mairal *et al.*, 2015*a*), they are the most suitable for dating in our system. First, we used a standard 'secondary calibration approach' in which the more inclusive higher level data set ('pDNA outgroups data set') was used to estimate divergence times within the ingroup. We used a GTR + G model and a uniform prior for the ucld.mean within values commonly observed in plant plastid markers (10^{-6}–10^{-1} substitutions per site Ma^{-1}; Wolfe *et al.*, 1987) and a default exponential prior for the standard deviation (s.d.). As calibration points, we used secondary age constraints drawn from the most comprehensive fossil-rich, meta-calibrated angiosperm phylogenetic tree reconstruction of Magallón *et al.* (2015). Two nodes were calibrated using a normal prior: the split between Goodeniaceae and the rest of the Asterales {mean = 57.05 Ma [high posterior density (HPD) 50.87–65.69], s.d. = 3}, and the split between Calyceraceae and Asteraceae [mean = 47.34 Ma (HPD 47.69–53.83), s.d.= 1.5]. The divergences between *A. alpinus* and *A. amellus* estimated in the previous analysis (1.87 Ma, HPD 0.09–6.2) were used to calibrate the *A. alpinus*–*A. amellus* node in the 'amellus pDNA data set'. Secondly, because the root and stem nodes of the *A. amellus* data set are both constrained with deep-time calibration events, we used a nested dating approach, similar to that adopted in previous studies (Pokorny *et al.*, 2011; Mairal *et al.*, 2015*a*), in which the higher level data sets calibrated with external evidence are used to constrain the molecular clock rate of the data set containing population-level data. This approach allowed us to avoid using 'all-encompassing' priors for the mean clock rate, so that we could assign a branching tree prior for the outgroup data set and a coalescent constant-size prior for the intraspecific amellus' data sets. Choice of model priors was based on Bayes factor comparisons using the path sampling (PS) and stepping stone (SS) sampling methods in BEAST (Baele *et al.*, 2012).

Haplotype analyses and demographic history

The relationships among lineages were investigated through haplotype network analyses, examining separately the plastid and nuclear genomes. Genealogical relationships among haplotypes were inferred via the statistical parsimony algorithm (Templeton *et al.*, 1992) implemented in TCS 1.21 (Clement *et al.*, 2000). The number of mutational steps resulting from single substitutions among haplotypes was calculated with 95 % confidence limits, and gaps were represented as missing data.

As we detected strong assortative mating between the different cytotypes of the *A. amellus* aggregate (see results here and in Münzbergová *et al.*, 2013), we performed demographic analyses for each cytotype separately. We used different approaches to infer the demographic processes within each cytotype. First of all, to test for evidence of population expansion, we carried out a neutrality test – Fu and Li's tests (Fu and Li, 1993; Fu, 1996) and Tajima's *D* test (Tajima, 1989) – for each cytotype. We used the DNAsp program, version 5.0 (Librado and Rozas, 2009), and assessed the significance level of each test [Fu's F_S (Fu, 1996) and Tajima's *D* (Tajima, 1989) and raggedness]) by generating 10 000 random samples, using coalescent simulations (Hudson, 1990; Nordborg *et al.*, 2003) under the infinite-site model. Secondly, we plotted the mismatch distribution for each cytotype using the observed number of differences between all pairs of sequences with DNAsp. The goodness of fit of the observed mismatch distribution to the theoretical distribution under a constant population size model was tested by generating 10 000 samples by coalescent simulations between observed and expected mismatch distributions and raggedness index (*r*) as test statistics (Harpending, 1994).

Ecological niche modelling

To understand whether the two cytotypes differ in their realized niche, we modelled the present distribution of each cytotype. Our extensive sampling with flow cytometry allowed us to use 327 records: 167 for the diploid cytotype and 160 for the hexaploid cytotype (inset in Fig. 1; Supplementary Data Table S4), covering the entire distribution range of the *A. amellus* aggregate in central Europe (see inset in Fig. 1), including adjacent areas. This includes entire species range to the north, south and west, while, to the east, our sampling only extends to Moldavia. We combined the available occurrences for the species within each lineage with a set of bioclimatic variables available from the WorldClim database (www.worldclim.org; Hijmans *et al.*, 2005). To avoid the high level of correlation usually found when using many of these variables, we chose bioclimatic variables based on prior ecological knowledge of our species (e.g. Münzbergová *et al.*, 2011) and that were not correlated (Pearson correlation coefficient $r < 0.75$). We chose: BIO1 = annual mean temperature, BIO5 = maximum temperature of the warmest month, BIO6 = minimum temperature of the coldest month, BIO12 = annual precipitation, BIO13 = precipitation of the wettest month and BIO14 = precipitation of the driest month. Pseudoabsences were generated by selecting 5000 random points across the modelled region. We used ensemble modelling (a procedure integrating the results from multiple modelling techniques) to generate our models. This procedure integrates the results from multiple modelling techniques within an ensemble framework to achieve more robust reconstructions (Araújo and New, 2007). We used four modelling algorithms that estimate species distribution using environmental variables and species occurrences: generalized linear models (GLMs), generalized additive models (GAMs), general boosting method (GBM) and random forests (RFs). These models were run in the R package 'biomod2' (Thuiller *et al.*, 2013) and summarized using additional R packages (R Core Team, 2014): 'foreign', 'raster' (Hijmans and van Etten, 2016), 'SDMTools' (VanDerWal *et al.*, 2011), 'rms' (Harrell, 2016), 'gbm' (Ridgeway, 2015), 'gam' (Hastie, 2016), 'rJava' (Urbanek, 2010), 'dismo' (Hijmans *et al.*, 2016) and 'randomForest' (Liaw and Wiener, 2002) with default

settings. We used repeated split sampling to evaluate the performance of the models, successively splitting the data set into 70 % for calibration and 30 % for evaluation by measuring the area under the curve (AUC). We quantified the performance of the model using the true skill statistics (TSS; Allouche *et al*., 2006). The final ensemble model was obtained considering models with AUC >0.8 and TSS >0.6.

Niche overlap between both cytotypes was quantified by using the proportional similarity of the distribution using the metric D (Schoener, 1970; Warren *et al*., 2010), an index used in conjunction with SDMs, using ENMtools (Warren *et al*., 2010), appropriate for intraspecific lineages that differ in their geographical distribution (Broennimann *et al*., 2012). This metric ranges from zero (no overlap) to 1 (complete overlap).

RESULTS

Phylogenetic relationships and molecular dating

The 'amellus ITS data set' consisted of 103 sequences of 642 nucleotides (102 individuals + *A. alpinus*), while the 'amellus pDNA data set' consisted of 101 sequences of 3887 nucleotides (100 individuals + *A. alpinus*) (Table 1). Nuclear and plastid Bayesian phylogenetic reconstructions showed different topologies (Fig. 2). Analyses of each plastid marker separately showed polytomies and some structured subclades with varying levels of support (Supplementary Data Fig. S1). The concatenated pDNA showed a topology where both cytotypes were intermingled within different clades (Fig. 2A). The nuclear phylogeny clearly separated the hexaploid individuals into one monophyletic clade and the diploid individuals into another monophyletic clade, except for one hexaploid individual (no. 33 in Supplementary Data Table S1) from the Vrutice Sad, Czech Republic population, which was grouped within the diploid clade. To avoid mistakes, we checked the unusual position of this individual by resequencing it (confirming its position) and by sequencing additional individuals in this populations, which fitted within the clade of hexaploids from other locations. Both cytotypes of the mixed-ploidy population of Strebersdorf were also separated into these two clades. The cytotype segregation was supported with a moderate clade support in the MrBayes analysis (PP = 88) and a strong support in the BEAST analysis (PP = 1, not included). ML analyses also separated the two cytotypes, though with lower resolution (Fig. 2B). The separation of the

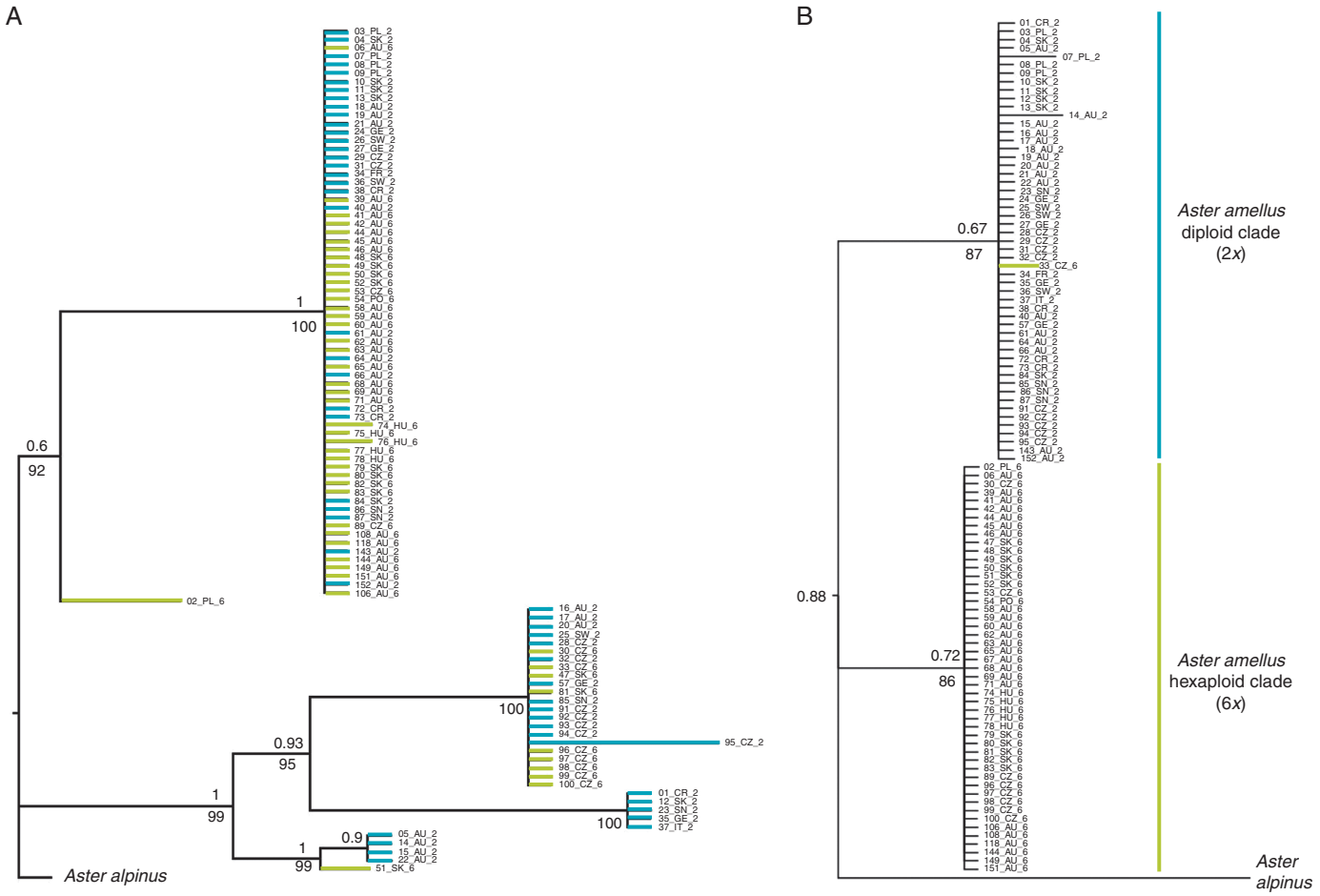

FIG. 2. Bayesian majority-rule consensus trees obtained by MrBayes from: (A) the concatenated *A. amellus* chloroplast data set (*atp*I–*atp*H, *rps*16–*trn*K, *rpl*32–*trn*L, *psb*E–*pet*L and *pet*N–*psb*M); (B) the *A. amellus* nuclear ribosomal (ITS) data set. Numbers above branches indicate Bayesian credibility values (PP); numbers below branches indicate maximum-likelihood bootstrap support values. Depending on their cytotype, each individual is marked in blue (diploids) or green (hexaploids). Codes for *A. amellus* populations correspond to those shown in Supplementary Data Table S1.

cytotypes based on ITS was partly supported by the Neighbor–Joining tree constructed using previously obtained microsatellite data. It showed lower genetic distances within diploids and within hexaploids, although the clades did not precisely sort into diploids and hexaploids (Supplementary Data Fig. S2).

The 'standard' (Supplementary Data Fig. S3) and 'nested' (Fig. 3) plastid dating approaches were congruent with each other and with the MrBayes analyses. Additionally, they showed overlapping confidence intervals: the divergence of Goodeniaceae and Calyceraceae was dated in the Paleocene [55.62 Ma in the nested approach (Fig. 3) vs. 56.14 Ma in the standard approach (Supplementary Data Fig. S3)], while the divergence of Calyceraceae and Asteraceae was dated in the Eocene (49.61 Ma in nested vs. 51.62 Ma in standard). *Aster alpinus* and *A. amellus* diverged in the Late Miocene (1.94 Ma in nested vs. 1.87 Ma in standard) and the first divergence within the *A. amellus* cytotypes was dated in the Pliocene (3.16 Ma, PP = 1 in nested vs. 3.24 Ma, PP = 1 in standard). Results for all the dating analyses implemented in BEAST are provided in Supplementary Data Table S5. PS and SS selected the Yule model and the uncorrelated log-normal distribution as the tree and clock model priors for all the analyses (Supplementary Data Table S5).

Haplotype network distribution and demographic analyses

Among the 102 individuals sampled from the 72 populations, we observed nine plastid DNA haplotypes (Hp1–Hp9 in Fig. 4A; geographically represented in Fig. 1) and seven different nuclear haplotypes (Hn1–Hn7 in Fig. 4B; geographically represented in Supplementary Data Fig. S4). The plastid network (Fig. 4A) showed that the two dominant haplotypes grouped populations including both cytotypes (see Hp1 and Hp8 in Fig. 4A), and the haplotypes did not show closer relationships among them, being separated by a large number of nucleotide changes. In the nuclear network (Fig. 4B), each haplotype was composed exclusively of either diploids or hexaploids. In addition, the haplotypes of each cytotype showed closer relationships (fewer changes) among them than with haplotypes of the other cytotype. Cytotypes occurring sympatrically in the mixed-ploidy population were also pooled into different haplotypes in the network. Four haplotypes dominated and were widespread geographically (Supplementary Data Fig. S4): two of the diploid cytotype (Hn1 and Hn2) and two of the hexaploid cytotype (Hn6 and Hn7).

Demographic analysis showed positive values of Fu's and Tajima's tests (Table 2), with significant values for the diploid

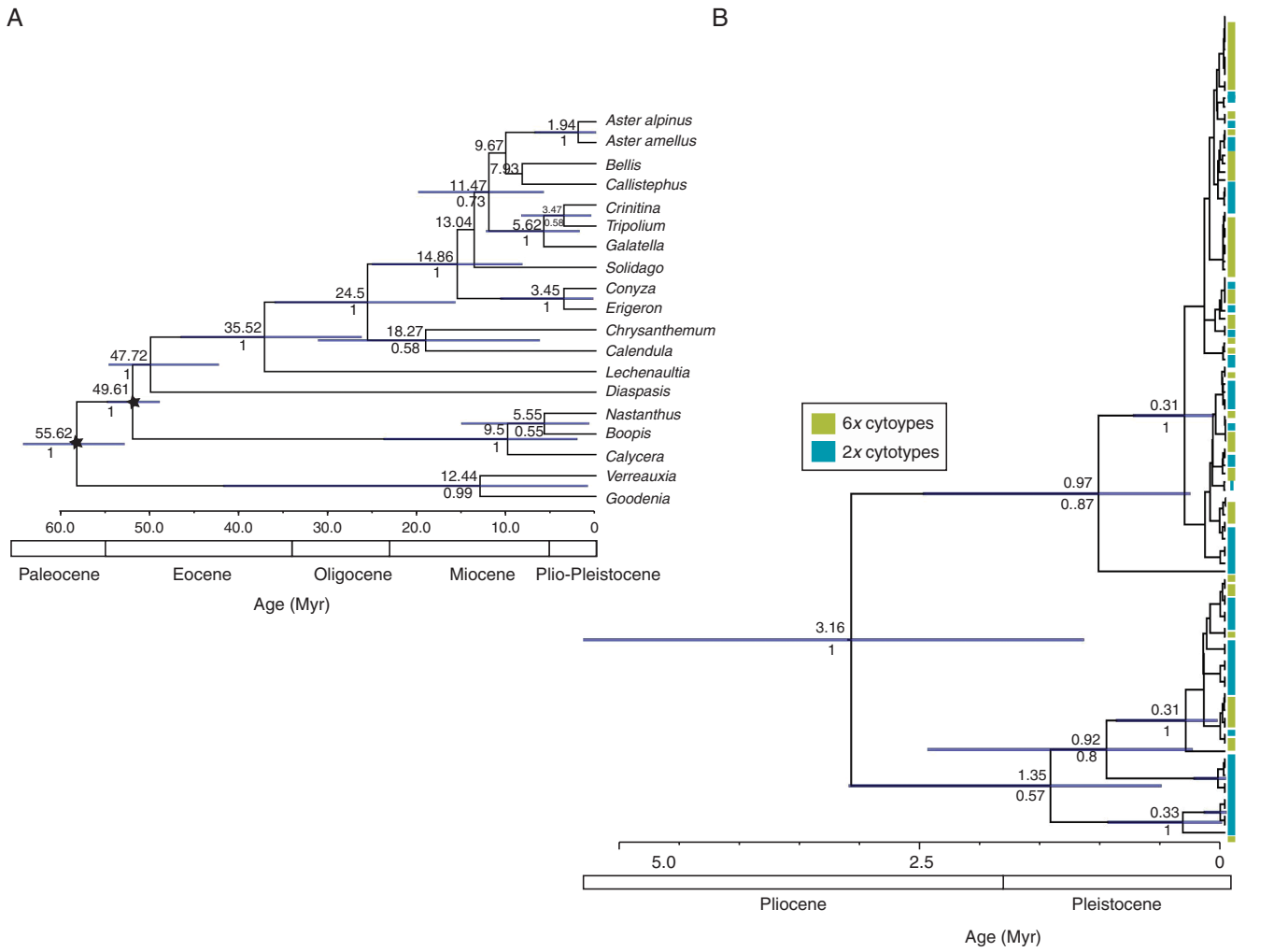

FIG. 3. Bayesian majority-rule consensus trees obtained by the nested analyses of the linked data sets: (A) 'pDNA outgroups data set' and (B) 'amellus pDNA data set'. Stars indicate constrained nodes (see text for more details). Horizontal bars show the 95 % HPD confidence intervals for the supported nodes. Numbers above branches indicate mean ages, and numbers below branches indicate Bayesian PP.

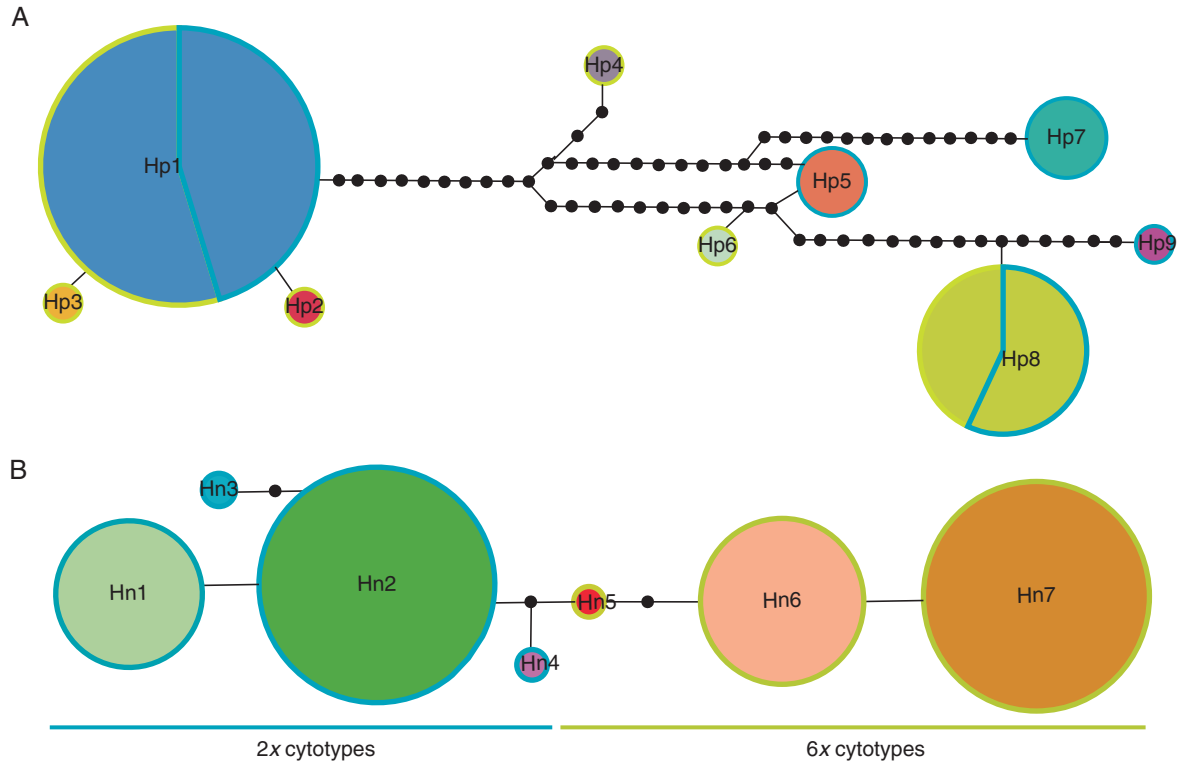

FIG. 4. Statistical parsimony networks inferred using the (A) chloroplast and (B) nuclear sequences by TCS. Black points on connecting lines indicate nucleotide changes. The circle size of the pie charts is proportional to the frequency of haplotypes. Pie charts are surrounded in blue for the diploid individuals and in green for the hexaploids. Each haplotype is shown in a different colour, where codes for the (A) chloroplast correspond to populations shown in Fig. 1, while codes for the (B) nucleus correspond to the populations shown in Supplementary Data Fig. S4.

individuals. The frequencies of pairwise differences in the mismatch distribution analyses resulted in bimodal distributions for the diploid individuals, and unimodal distributions with a second small peak for the hexaploid individuals (Fig. 5). The bimodal distribution in diploids does not conform to distributions expected for a population under the sudden expansion model and it might be interpreted as constant population size (e.g. Schneider and Excoffier, 1999; McMillen-Jackson and Bert, 2003). The unimodal distribution in the hexaploids was consistent with a recent demographic expansion, while the second small peak showed signs of sub-structure (Rogers and Harpending, 1992). The raggedness statistics derived from the mismatch distribution were significant (Table 2).

Ecological niche modelling

Species distribution models indicated that both cytotypes had high suitability values in the current contact zones (Fig. 6), where they grow together. However, overall, there were differences in ranges of the cytotype, with the niche of the diploids distributed to the west of the contact zone and the niche of the hexaploid to the east (Fig. 6). Additionally, the niches of the two cytotypes were not equivalent (Shoener's D total overlap = 0.52, P = 0.020). However, the relatively high value of D might indicate some biological overlap of resource use (Wallace, 1981). It is noteworthy that hexaploids showed a greater potential to occupy new areas to the east, and partly also to the west of their current distribution (see Fig. 6B). The AUC values were generally high (with values ranging between 0.80 and 0.96), suggesting that the models are consistent.

DISCUSSION

Can diploids and hexaploids be considered as distinct evolutionary lineages?

The nuclear and plastid trees showed different phylogenetic signals, which may indicate distinct evolutionary histories (Fig. 2). Our resolved ITS phylogeny supported an explicit phylogenetic hypothesis, separating diploid and hexaploid individuals into two different monophyletic clades (Fig. 2B). However, the inferences obtained from the ITS region could be limited due to additional difficulties such as paralogy, concerted evolution or directional bias in the homogenization process (Buckler *et al*., 1997; Eidesen *et al*., 2017). We can discard the possibility of paralogy, because the polymorphism detected when sequencing the ITS did not show shared sequences between cytotypes (Feliner and Roselló, 2007). The ITS may be subjected to concerted evolution leading to clear ITS differentiation of the two clades (Fig. 2B). Additionally, the phylogram method based on genetic distances using simple sequence repeats (SSRs) showed low support values, with diploids and hexaploids not clearly sorting into distinct clades (Supplementary Data Fig. S2). This is likely to be because microsatellites are widely distributed in

TABLE 2. *Results from the DNA polymorphism of plastid and nuclear haplotypes, neutrality test and mismatch raggedness for the* Aster amellus *aggregate*

	DNA polymorphism				Fu's F_S test	Tajima's D test	Mismatch distribution	
	π	θ	H (d)	n	F_S	D	Distribution	Raggedness
2*x* Cytotype	0.00377	0.00225	0.773	52	2.55**	2.12*	Multimodal	0.016*
6*x* Cytotype	0.00223	0.00196	0.471	49	1.533	0.421	Unimodal	0.091*

The mismatch distribution analyses are shown in Fig. 5.

Asterisks denote significant differences: $*P < 0.05$, $**P < 0.001$.

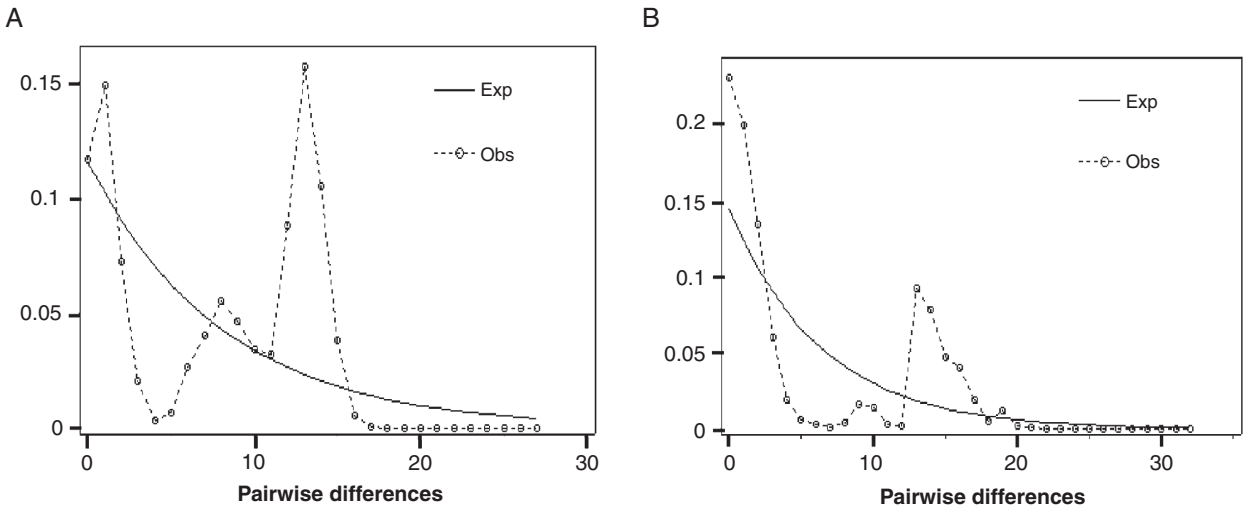

FIG. 5. Demographic analyses for the diploid (A) and hexaploid (B) cytotypes of *Aster amellus* showing the mismatch distribution. The observed (Obs) and expected (Exp) mismatch distributions (left) show the frequencies of pairwise differences.

the euchromatin (Schlötterer and Harr, 2001), and it is thus very unlikely that a concerted evolution process would have homogenized all the alleles. Still, we can, however detect separate genetic clusters in the microsatellite data due to assortative mating within cytotypes (Münzbergová *et al.*, 2013). Altogether, the results of the ITS and SSRs seem to indicate that the current mating patterns are strongly assortative within cytotypes.

Conversely, the different topology shown by the plastid data could arise due to processes such as hybridization, chloroplast capture or deep coalescence. We can discard hybridization since breeding barriers were detected in *A. amellus* and the intermediate forms found previously are inviable (Mandáková and Münzbergová, 2006; Castro *et al.*, 2011, 2012). In terms of chloroplast capture, we detected some identical sequences in cytotypes growing sympatrically in the only mixed-ploidy population detected in nature. Because individuals with intermediate ploidy levels have rarely been found in the populations (but were never fertile), one cannot completely dismiss the possibility of punctual hybridization and backcrossing in sympatric locations, though it seems unlikely. Thus, the discrepancy between the ITS and chloroplast trees seems to be due to shared ancestral relationships (deep coalescence events) or incomplete lineage sorting (ILS) among the chlorotypes (Maddison and Knowles, 2006).

Chloroplast DNA is smaller and more conserved compared with the nuclear genome, usually reflecting deep evolutionary events (Zurawski, 1987; Patwardhan *et al.*, 2014). In this case, the deep pDNA evolutionary events are further supported by the deep divergence found among the plastid cytotypes (Fig. 3; Supplementary Data Fig S3). While ILS may occur during divergences, gene flow decreases over time and eventually disappears (Rogers and Gibbs, 2014). These theoretical expectations agree with our findings: while pDNA patterns seem to reflect the ancient history of colonization of the seeds, the nuclear genome does not show admixture, confirming the absence of gene flow (pollen transport) between the cytotypes (Münzbergová *et al.*, 2013).

Morphological and cytological differences between the cytotypes do not all appear at once, leading to conflicts between the different species concepts in the early stages of autopolyploid evolution (De Queiroz, 2007). For example, in the case of *A. amellus*, the phylogenetic species concept is debatable because of phenomena such as concerted evolution and ILS. However, the cytotypes seem to behave as separate species according to the biological species concept as the cytotypes are reproductively isolated (Castro *et al.*, 2011). Regarding the ecological species concept (different habitat), our evidence demonstrates that habitats of diploid and hexaploid populations differ, with cytotypes showing some signs of local adaptation (Mandáková and Münzbergová, 2006; Raabová *et al.*, 2008; Münzbergová *et al.*, 2011)). However, the magnitude of this differentiation was low, indicating that the two cytotypes may occupy the same habitats in the contact zone, with broader niche differentiation across all the geographical range (Fig. 6). In contrast, *A. amellus* does not satisfy the morphological species concept, since both cytotypes are morphologically indistinguishable

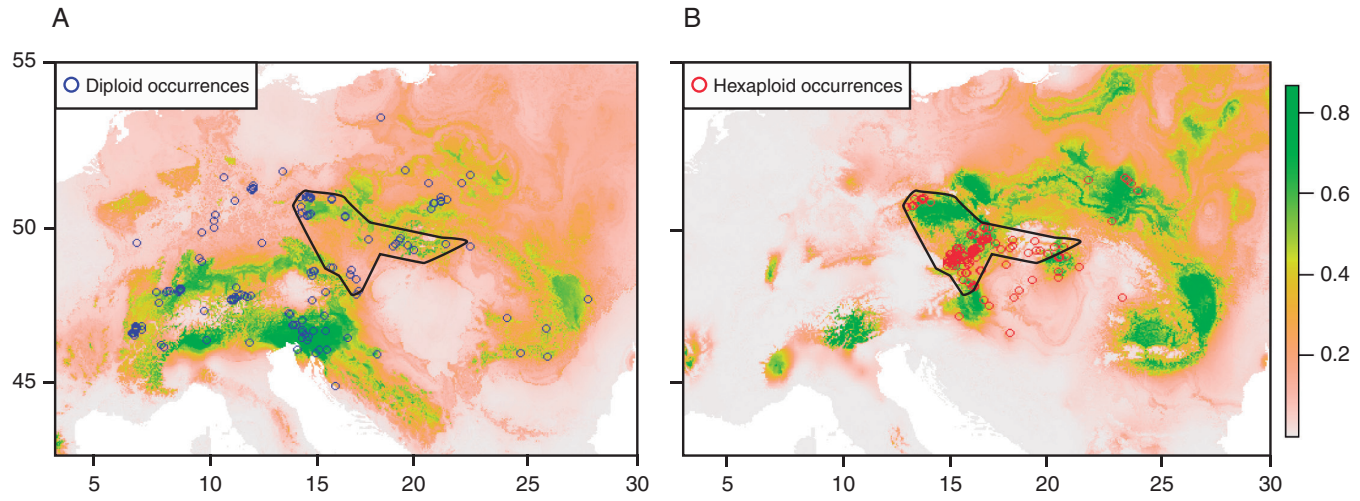

FIG. 6. Geographic projections of the climatic niche model of the (A) diploid and (B) hexaploid cytotypes of *Aster amellus* for current climate. The color scale indicates habitat suitability. The area delimited geometrically in black corresponds to the contact area of the cytotypes. The *x*-axis shows the longitude coordinates, while the *y*-axis shows the latitude.

(Mandáková and Münzbergová, 2008). However, this is an intrinsic feature of autopolyploids, which tend to be morphologically similar to their diploid progenitors (Soltis *et al.*, 2007). This may be because species divergence and phenotypic diversification are often highly temporally detached from the WGD event (Robertson *et al.*, 2017). Overall, the reproductive isolation between the *A. amellus* cytotypes seems to indicate an ongoing separation of the two lineages. These lineages can be viewed as two cryptic species (Soltis *et al.*, 2007) yet carrying the signatures of ancient relationships.

Multiple-origin vs. single-origin model

A previous study of *A. amellus* using microsatellites (Münzbergová *et al.*, 2013) postulated that *A. amellus* cytotypes probably had a single origin. However, our new evidence shows the need to include chloroplast markers to unravel the origin of cytotypes. The ancient divergences and topological relationships reflected by the pDNA suggest that the hexaploid cytotype arose and was established several times from the diploid cytotype (Figs 2A and 3B). On the other hand, the parapatric mosaic of *A. amellus* cytotypes makes it difficult to interpret the geographical patterns clearly, and a single-origin model could also be expected if the secondary contact occurred after Pleistocence range expansion (Mandáková and Münzbergová, 2006). Differentiation between these alternatives requires an accurate temporal framework, which is not a straightforward task due to the difficulties that exist when dating polyploids (Doyle and Egan, 2010). However, it is remarkable that our different dating approaches were congruent and resulted in very similar age estimates (Supplementary Data Table S5). We need, however, to emphasize that our dating identified the point at which gene trees coalescent, which does not necessarily coincide with the polyploidization event (Doyle and Egan, 2010). Overall, dating estimates agree well with an ancient origin of the genetic diversification in *A. amellus*, suggesting a multiple-origin model for the hexaploid cytotypes. This is further supported by the large number of undetected mutation events separating the chlorotypes (gaps in Fig. 4A), which probably corresponds to extinct ancestral haplotypes (Mairal *et al.*, 2015*b*), showing older relationships in this marker.

After the emergence of polyploid lineages, their subsequent success requires reproductive isolation (Soltis *et al.*, 2007). The establishment of a new cytotype will only be possible when intracytotype mating increases (Rieseberg and Willis, 2007; Paun *et al.*, 2009), or the new entity has an advantage (such as increased asexual reproduction or selfing) or disperses to other localities to avoid minority cytotype exclusion. In *A. amellus*, the hexaploid establishment has probably been supported by reproductive isolation caused by strong intercytotype gametic barriers together with some level of selfing that enabled offspring production in the hexaploids (Castro *et al.*, 2011) and resulted in functional isolation between the ploidy levels (Münzbergová *et al.*, 2013).

Changes in population size and niche differentiation of the cytotypes

The evolutionary advantages intrinsic to autopolyploidy may provide higher reproductive success over the diploid entities (Lewis, 1980; Levin, 1983). If this hypothesis is correct, the polyploid cytotype should be more persistent, increasing the probability of detecting its demographic expansion. In this study, we found significant differences in the trajectories of the two *A. amellus* cytotypes. On the one hand Fu's and Tajima's tests (Table 2) for the diploid individuals suggest either balancing selection or a recent population decrease. On the other hand, our results for hexaploids allowed us to accept the null hypothesis of recent population expansion (Rogers and Harpending, 1992) (Table 2; Fig. 5), supporting the idea that, after their origin, the hexaploids settled in a growing number of habitats. This is further suggested by the greater potential of the hexaploids to colonize new areas, especially to the east [compare distribution (inset in Fig.1) with Fig. 6]. To the east, hexaploids clearly found open niches beyond the limits of its diploid progenitor where they could easily establish (Levin, 1975; Mandáková and Münzbergová, 2006).

Despite the niche differentiation detected across the whole distribution range, the niche of both cytotypes seems to be overlapping in the contact zone (Fig. 6). However, a previous study that performed a reciprocal transplant experiment among neighbouring populations in the field showed niche differentiation between the two cytotypes and local adaptation within each cytotype (Raabová *et al.*, 2008). This may contribute to the maintenance of single-ploidy populations of *A. amellus* along the contact zone.

The only mixed-ploidy population detected (Strebersdorf) could provide new insights into the trajectories of mixed ploidy populations. This population was detected at a contact zone of secondary origin, where the hexaploids show higher seed set and seedling emergence (Castro *et al.*, 2012). The hexaploids are thus expected to displace the diploids by means of minority cytotype exclusion, possibly also linked with direct competition. In line with this expectation, reduction in the proportion of diploids in this population has been confirmed in its resampling in 2017 (J. Raabová, pers. comm.). Additionally, recent field observations have detected a second mixed-ploidy population where, interestingly, cytotypes show clear niche differentiation: the hexaploids have been detected growing in grasslands, while the diploids are relegated to sub-optimal zones inside the forest (J. Raabová, pers. comm.). This further supports the expansion of the hexaploids, and subsequent displacement of diploid individuals.

Additionally, the accumulated evidence detected in *A. amellus* further supports an increasing colonization potential of hexaploids. For example: (1) hexaploids show wider ecological amplitude (Mandáková and Münzbergová, 2006); (2) hexaploids grow equally well with and without arbuscular mycorrhizal fungal (AMF) symbiosis, while diploids grow significantly better only with AMF (Sudova *et al.*, 2014); (3) hexaploids occur in both low and high productive habitats, while diploids are confined only to low productive habitats (Mandáková and Münzbergová, 2006; Münzbergová, 2007); and (4) the damage to seeds by herbivory decreases with habitat isolation in hexaploids, whereas no such trend can be found for diploids (Münzbergová, 2006).

The greater genomic flexibility acquired by polysomic inheritance provides polyploids with new evolutionary advantages for colonization over their diploid relatives (Ramsey *et al.*, 2008; Parisod *et al.*, 2010; Alix *et al.*, 2017). While evolutionary advantages of polyploids have been extensively documented for allopolyploids (Lewis, 1980; Soltis *et al.*, 2014), in autopolyploids the evolutionary consequences of these advantages remain largely unknown (Parisod *et al.*, 2010; Ramsey, 2011). Although our demographic analyses seem to point in this direction, our data set is not powerful enough to test this hypothesis, and further studies are necessary to test the possible evolutionary advantages of autopolyploidy.

Conclusions

Autopolyploid evolution has been overlooked, underestimating its evolutionary implications (Soltis *et al.*, 2007; Barker *et al.*, 2016). Most studies have focused on diploid–tetraploid allopolyploids, where the effects of genome duplication with interspecific hybridization may obscure the signature of ancient relationships. In addition, identifying autopolyploid lineages has been limited due to the need for a high sampling, an insufficient differentiation in the markers among cytotypes and a lack of taxonomic recognition (Kolář *et al.*, 2017). The analytical approach performed here provides clues about an ongoing diversification process in a unique diploid–autohexaploid aggregate, in which cytotypes are morphologically indistinguishable. Here we demonstrate that the two cytotypes have common evolutionary history and probably diverged due to multiple polyploidization events. This knowledge is essential to understand the cryptic diversity of morphologically identical autopolyploids, and shows the importance of performing additional studies on autopolyploid cytotypes, potentially recognizable as different lineages. We suggest that the great cryptic diversity masked by the autopolyploidy and its role as a source of evolutionary advantages should be further evaluated.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/aob and consist of the following. Table S1: voucher information and GenBank accession numbers for all samples included in this study. Table S2: primers used for PCR amplification and sequencing. Table S3: evolutionary model selected with Jmodeltest. Table S4: geographical co-ordinates for the diploid and hexaploid individuals used in the ecological niche modelling of *Aster amellus*. Table S5: results for the analyses implemented in BEAST to test accuracy of dating estimates. Figure S1: Bayesian majority-rule consensus trees for each chloroplast marker individually. Figure S2: Neighbor–Joining tree with microsatellites. Figure S3: Bayesian majority-rule consensus trees obtained by the standard dating analyses of the linked data sets. Figure S4: haplotype nuclear distribution for the 72 populations sampled for *Aster amellus* in central Europe.

ACKNOWLEDGEMENTS

We thank Andrea Jarošová for her help in the laboratory, and Judith Fehrer for the help provided in the laboratory in the initial stages of the project. We are grateful to the Population Ecology Seminar Group of Průhonice (Czech Academy of Science) for a critical review of the manuscript. We also thank Isabel Marques, Myriam Heuertz, Pilar Catalán, Tomáš Fér and two anonymous reviewers for comments to the manuscript, and J. Raabová for providing the unpublished information on the mixed-ploidy populations. This work was supported by Grant Agency Czech Republic [grant no. P505-13-32048S]; Institutional research projects Academy of Science, Czech Republic [grant no. 67985939]; Ministry of education, youth and sports and Programa Operacional Potencial Humano – Fundo Social Europeu/Portuguese Foundation for Science and Technology [grant no. IF/01267/2013 to S.C.].

LITERATURE CITED

Alix K, **Gérard PR**, **Schwarzacher T**, **Heslop-Harrison JS. 2017**. Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Annals of Botany* **120**: 183–194.

Allouche O, **Tsoar A**, **Kadmon R. 2006**. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**: 1223–1232.

Araújo MB, **New M. 2007**. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* **22**: 42–47.

Arrigo N, **Barker MS. 2012**. Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology* **15**: 140–146.

Baele G, **Lemey P**, **Bedford T**, **Rambaut A**, **Suchard MA**, **Alekseyenko AV. 2012**. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* **29**: 2157–2167.

Barker MS, **Arrigo N**, **Baniaga AE**, **Li Z**, **Levin DA. 2016**. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist* **210**: 391–398.

Broennimann O, **Fitzpatrick MC**, **Pearman PB**, ***et al.* 2012**. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* **21**: 481–497.

Buckler ES, Ippolito A, Holtsford TP. 1997. The evolution of ribosomal DNA divergent paralogues and phylogenetic implications. *Genetics* **145:** 821–832.

Castro S, **Münzbergová Z**, **Raabová J**, **Loureiro J. 2011**. Breeding barriers at a diploid–hexaploid contact zone in *Aster amellus*. *Evolutionary Ecology* **25**: 795–814.

Castro S, **Loureiro J**, **Procházka T**, **Münzbergová Z. 2012**. Cytotype distribution at a diploid–hexaploid contact zone in *Aster amellus* (Asteraceae). *Annals of Botany* **110**: 1047–1055.

Catalán P, **Segarra-Moragues JG**, **Palop-Esteban M**, **Moreno C**, **González-Candelas F. 2006**. A Bayesian approach for discriminating among alternative inheritance hypotheses in plant polyploids: the allotetraploid origin of genus *Borderea* (Dioscoreaceae). *Genetics* **172**: 1939–1953.

Catalán P, **Müller J**, **Hasterok R**, ***et al.* 2012**. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Annals of Botany* **109**: 385–405.

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

Cuadrado Á, **de Bustos A**, **Jouve N. 2017**. On the allopolyploid origin and genome structure of the closely related species *Hordeum secalinum* and *Hordeum capense* inferred by molecular karyotyping. *Annals of Botany* **120**: 245–255.

De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* **56**: 879–886.

Dong W, **Liu J**, **Yu J**, **Wang L**, **Zhou S. 2012**. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS One* **7**: e35071.

Doyle JJ, **Egan AN. 2010**. Dating the origins of polyploidy events. *New Phytologist* **186**: 73–85.

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

Eidesen PB, **Alsos IG**, **Popp M**, **Stensrud Ø**, **Suda J**, **Brochmann C. 2017**. Nuclear vs. plastid data: complex Pleistocene history of a circumpolar key species. *Molecular Ecology* **16**: 3902–3925.

Feldman M, **Levy AA. 2005**. Allopolyploidy – a shaping force in the evolution of wheat genomes. *Cytogenetic and Genome Research* **109**: 250–258.

Feliner GN, **Rosselló JA. 2007**. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* **44**: 911–919.

Fu YX. 1996. New statistical tests of neutrality for DNA samples from a population. *Genetics* **143**: 557–570.

Fu YX, **Li WH. 1993**. Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.

Grant V. 1981. *Plant speciation*. New York: Columbia University Press.

Halverson K, **Heard SB**, **Nason JD**, **Stireman JO. 2008**. Origins, distribution, and local co-occurrence of polyploid cytotypes in *Solidago altissima* (Asteraceae). *American Journal of Botany* **95**: 50–58.

Harrell JFE. 2016. *rms: regression modeling strategies. R package version 3.4-0*. https://cran.r-project.org/web/packages/ rms/rms.pdf

Harpending H. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66**: 591–600.

Hastie T. 2016. *gam: generalized additive models. R package version, 1*. https://cran.r-project.org/web/packages/gam/ gam.pdf.

Hijmans RJ, **van Etten J. 2016**. *raster: geographic data analysis and modeling. R package version, 2–15*. https://cran.r-project.org/web/packages/ raster/ras- ter.pdf

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.

Hijmans RJ, **Phillips S**, **Leathwick J**, **Elith J. 2016**. *dismo: species distribution modeling. R package version 0.8–17*. https://cran.r-project.org/web/ packages/dismo/dismo.pdf

Hudson RR. 1990. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**: 44.

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

Jiao Y, **Wickett NJ**, **Ayyampalayam S**, ***et al.* 2011**. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97–100.

Kelchner SA. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* **87**: 482–498.

Kellogg EA. 2016. Has the connection between polyploidy and diversification actually been tested? *Current Opinion in Plant Biology* **30**: 25–32.

Kolář F, **Čertner M**, **Suda J**, **Schönswetter P**, **Husband BC. 2017**. Mixed-ploidy species: progress and opportunities in polyploid research. *Trends in Plant Science* **22**: 1041–1055.

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

Levin DA. 1983. Polyploidy and novelty in flowering plants. *American Naturalist* **122**: 1–25.

Lewis WH. 1980. Polyploidy in species populations. In: Lewis WH, ed. *Polyploidy: biological relevance*. New York: Plenum Press, 103–144.

Lexer C, **van Loo M. 2006**. Contact zones: natural labs for studying evolutionary transitions. *Current Biology* **16**: 407–409.

Li WP, **Yang FS**, **Jivkova T**, **Yin GS. 2012**. Phylogenetic relationships and generic delimitation of Eurasian *Aster* (*Asteraceae: Astereae*) inferred from ITS, ETS and *trn*L-F sequence data. *Annals of Botany* **109**: 1341–1357.

Liaw A, **Wiener M. 2002**. Classification and regression by randomForest. *R News* **2**: 18–22.

Librado P, **Rozas J. 2009**. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.

Maddison WP, **Knowles LL. 2006**. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* **55**: 21–30.

Magallón S, **Gómez–Acevedo S**, **Sánchez–Reyes LL**, **Hernández–Hernández T. 2015**. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**: 437–453.

Mairal M, **Pokorny L**, **Aldasoro JJ**, **Alarcón M**, **Sanmartín I. 2015*a***. Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: the case of the Rand Flora genus *Canarina* (Campanulaceae). *Molecular Ecology* **24**: 1335–1354.

Mairal M, **Sanmartín I**, **Aldasoro JJ**, **Culshaw V**, **Manolopoulou I**, **Alarcón M. 2015*b***. Palaeo-islands as refugia and sources of genetic diversity within volcanic archipelagos: the case of the widespread endemic *Canarina canariensis (*Campanulaceae). *Molecular Ecology* **24**: 3944–3963.

Mandáková T, **Münzbergová Z. 2006**. Distribution and ecology of cytotypes of the *Aster amellus* aggregates in the Czech Republic. *Annals of Botany* **98**: 845–856.

Mandáková T, **Münzbergová Z. 2008**. Morphometric and genetic differentiation of diploid and hexaploid populations of *Aster amellus* agg. in a contact zone. *Plant Systematics and Evolution* **274**: 155–170.

Masterson J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264**: 421–423.

McMillen-Jackson, AL, **Bert TM. 2003**. Disparate patterns of population genetic structure and population history in two sympatric penaeid shrimp species (*Farfantepenaeus aztecus* and *Litopenaeus setiferus*) in the eastern United States. *Molecular Ecology* **12**: 2895–2905.

Müntzing A. 1932. Cytogenetic investigations on synthetic *Galeopsis tetrahit*. *Hereditas* **16**: 105–154.

Münzbergová Z. 2006. Ploidy level interacts with population size and habitat conditions to determine the degree of herbivory damage in plant populations. *Oikos* **115**: 443–452.

Münzbergová Z. 2007. Population dynamics of diploid and hexaploid populations of a perennial herb. *Annals of Botany* **100**: 1259–1270.

Münzbergová Z, **Raabová J**, **Castro S**, **Pánková H. 2011**. Biological flora of central Europe: *Aster amellus* L. (Asteraceae). *Perspectives in Plant Ecology, Evolution and Systematics* **13**: 151–162.

Münzbergová Z, **Šurinová M**, **Castro S. 2013**. Absence of gene flow between diploids and hexaploids of *Aster amellus* at multiple spatial scales. *Heredity* **110**: 123–130.

Nesom GL. 1994. Subtribal classification of the Astereae (Asteraceae). *Phytologia* **76**: 193–274.

Nordborg M. 2003. Coalescent theory. In: Balding D, Bishop M, Cannings C, eds. *Handbook of statistical genetics*, 2nd edn. New York: John Wiley and Sons, 602–635.

Noyes RD, **Rieseberg LH. 1999**. ITS sequence data support a single origin for North American Astereae (Asteraceae) and reflect deep geographic divisions in *Aster* sl. *American Journal of Botany* **86**: 398–412.

Parisod C, **Holderegger R**, **Brochmann C. 2010**. Evolutionary consequences of autopolyploidy. *New Phytologist* **186**: 5–17.

Patwardhan A, **Ray S**, **Roy A. 2014**. Molecular markers in phylogenetic studies – a review. *Journal of Phylogenetics & Evolutionary Biology* **2**: 131. doi: 10.4172/2329-9002-2-131

Paun O, **Forest F**, **Fay MF**, **Chase MW. 2009**. Hybrid speciation in angiosperms: parental divergence drives ploidy. *New Phytologist* **182**: 507–518.

Pokorny L, **Oliván G**, **Shaw A. 2011**. Phylogeographic patterns in two southern hemisphere species of *Calyptrochaeta* (Daltoniaceae, Bryophyta). *Systematic Botany* **36**: 542–553.

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

Raabová J, **Fischer M**, **Münzbergová Z. 2008**. Niche differentiation between diploid and hexaploid *Aster amellus*. *Oecologia* **158**: 463–472.

Rambaut A, **Drummond AJ. 2013**. *TreeAnnotator v1. 7.0*.

Rambaut A, **Drummond AJ**, **Suchard M. 2013**. *Tracer v1. 6*.

Ramsey J. 2011. Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences, USA* **108**: 7096–7101.

Ramsey J, **Robertson A**, **Husband B. 2008**. Rapid adaptive divergence in New World *Achillea*, an autopolyploid complex of ecological races. *Evolution* **62**: 639–653.

R Core Team. **2014**. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rieseberg LH, **Willis JH. 2007**. Plant Dpeciation. *Science* **317**: 910–914.

Ridgeway G. 2015. *Generalized boosted regression models. Documentation on the R Package 'gbm', version 1? 5, 7*. https://cran.r-project.org/web/packages/gbm/gbm.pdf

Robertson FM, **Gundappa MK**, **Grammes F**, ***et al.* 2017**. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biology* **18**: 111. doi: 10.1186/s13059-017-1241-z.

Rogers AR, **Harpending H. 1992**. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**: 552–569.

Rogers J, Gibbs RA. 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nature Reviews Genetics* **15:** 347.

Ronquist F, **Teslenko M**, **van der Mark P**, ***et al.* 2012**. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

Rosato M, **Castro M**, **Rosselló**. **2008**. Relationships of the woody *Medicago* species (section Dendrotelis) assessed by molecular cytogenetic analyses. *Annals of Botany* **102**: 15–22.

Ryan WB, **Carbotte SM**, **Coplan JO**, ***et al.* 2009**. Global multi-resolution topography synthesis. *Geochemistry, Geophysics, Geosystems* **10**: Q03014. doi.org/10.1029/2008GC002332.

Schlötterer C, **Harr B. 2001**. Microsatellite instability. *eLS*. doi.org/10.1038/npg.els.0000840.

Schneider S, **Excoffier L. 1999**. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152**: 1079–1089.

Schoener TW. 1970. Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology* **51**: 408–418.

Soltis DE, **Soltis PS**, **Schemske DW**, ***et al.* 2007**. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* **56**: 13–30.

Soltis DE, **Buggs RJ**, **Doyle JJ**, **Soltis PS. 2010**. What we still don't know about polyploidy. *Taxon* **59**: 1387–1403.

Soltis DE, **Visger CJ**, **Marchant DB**, **Soltis PS. 2016**. Polyploidy: pitfalls and paths to a paradigm. *American Journal of Botany* **103**: 1146–1166.

Soltis PS, **Soltis DE. 2000**. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences, USA* **97**: 7051–7057.

Soltis PS, **Liu X**, **Marchant DB**, **Visger CJ**, **Soltis DE. 2014**. Polyploidy and novelty: Gottlieb's legacy. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**: 20130351. doi:10.1098/rstb.2013.0351.

Stamatakis A, **Hoover P**, **Rougemont J. 2008**. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* **57**: 758–771.

Stebbins GL. 1947. Types of polyploids: their classification and significance. *Advances in Genetics* **1**: 403–429.

Stebbins GL. 1970. Variation and evolution in plants: progress during the past twenty years. In: Hecht MK, Steere WC, eds. *Essays in evolution and genetics in honor of Theodosius Dobzhansky: a supplement to Evolutionary Biology*. New York: Appleton-Century-Crofts, 173–208.

Sudová R, **Pánková H**, **Rydlová J**, **Münzbergová Z**, **Suda J. 2014**. Intraspecific ploidy variation: a hidden, minor player in plant–soil–mycorrhizal fungi interactions. *American Journal of Botany* **101**: 26–33.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis pDNA polymorphism. *Genetics* **123**: 585–595.

Templeton AR, **Crandall KA**, **Sing CF. 1992**. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619–633.

Thuiller W, **Georges D**, **Engler R. 2013**. *biomod2: ensemble platform for species distribution modeling. R package version, 2, r560*. http://www2.uaem.mx/r-mirror/web/ packages/biomod2/biomod2.pdf

Urbanek S. 2010. *rJava: Low-level R to Java interface*. https://cran.r-project.org/web/packages/rJava/rJava.pdf

Van de Peer Y, **Mizrachi E**, **Marchal K. 2017**. The evolutionary significance of polyploidy. *Nature Reviews. Genetics* **18**: 411–424.

Wallace RK. 1981. An assessment of diet-overlap indexes. *Transactions of the American Fisheries Society* **110**: 72–76.

VanDerWal J, **Falconi L**, **Januchowski S**, **Shoo L**, **Storlie C. 2011**. *SDMTools: species distribution modelling tools: tools for processing data associated with species distribution modelling exercises*. R package version, 1.

Warren DL, **Glor RE**, **Turelli M. 2010**. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* **33**: 607–611.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* **42**: 22–249.

Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *American Journal of Botany* **102**: 1753–1756.

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

Wood TE, **Takebayashi N**, **Barker MS**, **Mayrose I**, **Greenspoon PB**, **Rieseberg LH. 2009**. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA* **106**: 13875–13879.

Zurawski G, **Clegg MT. 1987**. Evolution of higher plant chloroplast DNA-encoded genes: Implications for structure–function and phylogenetic studies. *Annual Review of Plant Physiology and Plant Molecular Biology* **38**: 391–418.