Flow cytometry, microsatellites and niche models reveal the origins and geographical structure of *Alnus glutinosa* populations in Europe

Bohumil Mandák1,2,*, Petr Vit1,2, Karol Krak1,2, Pavel Trávníček1,3, Alena Havrdová1,2, Věroslava Hadincová1, Petr Zákrauský1, Vlasta Jarolímová1, Cecile Fanny Emilie Bacles4 and Jan Douda1,2

1Institute of Botany, The Czech Academy of Sciences, Zámek 1, CZ-252 43, Průhonice, Czech Republic; 2Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamýcká 129, Praha 6–Suchdol, CZ-165 21, Czech Republic; 3Faculty of Agriculture, University of South Bohemia, Studentská 13, CZ-370 05, České Budějovice, Czech Republic and 4University of Pau and Pays Adour, UFR Sciences et Techniques, Département de Biologie, F-64100 Pau, France

* For correspondence. E-mail bohumil.mandak@ibot.cas.cz

Received: 8 January 2015  Returned for revision: 11 May 2015  Accepted: 1 September 2015  Published electronically: 14 October 2015

- **Background and Aims** Polyplody in plants has been studied extensively. In many groups, two or more cytotypes represent separate biological entities with distinct distributions, histories and ecology. This study examines the distribution and origins of cytotypes of *Alnus glutinosa* in Europe, North Africa and western Asia.
- **Methods** A combined approach was used involving flow cytometry and microsatellite analysis of 12 loci in 2200 plants from 209 populations combined with species distribution modelling using MIROC and CCSM climatic models, in order to analyse (1) ploidy and genetic variation, (2) the origin of tetraploid *A. glutinosa*, considering *A. incana* as a putative parent, and (3) past distributions of the species.
- **Key Results** The occurrence of tetraploid populations of *A. glutinosa* in Europe is determined for the first time. The distribution of tetraploids is far from random, forming two geographically well-delimited clusters located in the Iberian Peninsula and the Dinaric Alps. Based on microsatellite analysis, both tetraploid clusters are probably of autoployploid origin, with no indication that *A. incana* was involved in their evolutionary history. A projection of the MIROC distribution model into the Last Glacial Maximum (LGM) showed that (1) populations occurring in the Iberian Peninsula and North Africa were probably interconnected during the LGM and (2) populations occurring in the Dinaric Alps did not exist throughout the last glacial periods, having retreated southwards into lowland areas of the Balkan Peninsula.
- **Conclusions** Newly discovered tetraploid populations are situated in the putative main glacial refugia, and neither of them was likely to have been involved in the colonization of central and northern Europe after glacial withdrawal. This could mean that neither the Iberian Peninsula nor the western part of the Balkan Peninsula served as effective refugia for northern post-glacial expansion of *A. glutinosa*.

**Key words:** *Alnus glutinosa*, autoployploidy, cytotype, glacial refugia, microsatellites.

**INTRODUCTION**

Polyplody, the possession of more than two sets of chromosomes, has been an important factor in euakaryote evolution (Otto, 2007), with as many as 80 % of all taxa estimated to have had a polyploid origin (Stebbins, 1950, 1971; Grant, 1981; Masterson, 1994; Otto and Whitton, 2000). Recently it has been shown that almost all vascular plants have undergone at least one round of polyploidy in their evolutionary history. Genome doubling has been an important process in plant evolution, quickly producing novel cytotypes by two major processes. Hybridization between different cross-sterile taxa, which can escape from sterility by chromosome doubling, produces allopolyploid individuals (Kochert et al., 1996; Lowe and Abbott, 1996; Cook et al., 1998; Segraves et al., 1999; Kolář et al., 2009; Mandák et al., 2012), whereas hybridization between fully cross-fertile progenitors, accompanied by doubling of structurally similar, homologous genomes, produces auto-ployploid individuals (Thompson and Lumaret, 1992; Parisod et al., 2010). Allopolyploidy has been recognized as the most common process in polyploid formation for a long time (Stebbins, 1950, 1971; Wendel, 2000), but only recently have different authors shown that autoployploidy has probably been overlooked and that the number of autoployploids is underestimated (Mahy et al., 2000; Soltis et al., 2007; Parisod et al., 2010).

Both allo- and autoployploids potentially harbour more genetic variation than their diploid progenitors because they combine more than two gene copies, but differ in their mode of inheritance (disomy in allopolyploids vs. polysomy in autopolyploids) (Catalán et al., 2006; Parisod et al., 2010). Moreover, recent studies have demonstrated that polyploid genomes can be highly dynamic and undergo rapid structural and functional alterations (Doyle et al., 2008; Leitch and Leitch, 2008).

It has long been known that the frequency of polyploidy increases with latitude in the Northern Hemisphere (Hagerup, 1931; Stebbins, 1950; Löve and Löve, 1957; Johnsson and ...
Packer, 1965). Stebbins (1984, 1985) suggested a correlation between the frequency of polyploidy and the degree of glaciation, rather than purely between polyploidy and latitude. This led Stebbins to describe polyploidy as a process stabilizing new gene combinations derived from hybridization between genetically distinct parental gene pools adapted to different ecological conditions (‘secondary contact hypothesis’). The view of Stebbins (1984) was recently revisited by Brochmann et al. (2004), who used Arctic plants as a model system. They did not find any association between polyploidy and the degree of glaciation for the Arctic flora as a whole, but for Arctic specialist taxa with restricted distributions the frequency of diploids was higher in largely unglaciated areas during the last Ice Age than in heavily glaciated areas. Even differences in current distribution among cytotypes might be significantly affected by history; the pattern of polyploidy increasing with latitude might not be absolutely clear in many taxa (Mandáková and Münzbergová, 2006; Kolář et al., 2009, 2012; Duchošal et al., 2010). Therefore, the position of past cytotype refugia relative to sites that became available for colonization by a single cytotype is probably also important and may significantly complicate the pattern of increasing ploidy with latitude. Kolář et al. (2009, 2012) revealed a unique evolutionary pattern in the *Knautia arvensis* aggregate. They determined a wide variety of processes and mechanisms which probably took part in the rapid evolution of this complex, including isolation in Holocene refugia, repeated colonization by distinct lineages, hybridization and recurrent polyploidization. All these processes generate diffuse patterns of cytotype distribution not related to the general trend of increasing ploidy from south to north.

Putative migration routes after glaciers retreated have mostly been described based on studies of post-glacial tree migrations (Bennet et al., 1991; Hewitt, 1996, 2000; Taberlet et al., 1998; Cruzan and Templeton, 2000; Tzedakis et al., 2013). These studies have demonstrated that the current distributions of different tree species in Europe are a result of multiple migration scenarios (Petit et al., 2002; Magri et al., 2006; Tollefsrud et al., 2008; Parducci et al., 2012; Wachowiak et al., 2013). King and Ferris (1998) suggested, based on an analysis of plastid DNA, that *Alnus glutinosa*, a foundation tree species of alluvial forests, may have expanded from a Balkan refugium to northern and western Europe after the Ice Age. Even though the species was present in Italian and Iberian refugia, these populations do not seem to have participated in post-glacial recolonization of northern Europe including the far north. However, the sampling of King and Ferris (1998) completely missed the western Russian plains, where at least one important refugium is thought to have been located. Furthermore, Douda et al. (2014) showed, on the basis of radiocarbon-dated pollen and macrofossil sites, that *Alnus* trees probably withstood the Last Glacial Maximum (LGM) in western Europe, the northern foothills of the Alps, the Romanian Carpathians and a large area of north-eastern Europe. It follows that, after withdrawal of glaciers, *Alnus* rapidly colonized southern Sweden and gradually expanded northward, most probably predominantly from a refugium located in Belarus and western Russia. The increase in *Alnus* occurrence in more southerly ice-free areas of Europe seems to reflect local expansions originating from regional refugia.

*Alnus glutinosa* and *A. incana* (Betulaceae) have been reported to be diploid (2n = 2x = 28) by many authors (Fedorov, 1969; Goldblatt, 1981, 1984, 1985, 1988; Goldblatt and Johnson, 1990, 1991, 1994, 1996, 1998, 2000, 2003; Ivanova et al., 2006). However, there is one tetraploid record (2n = 4x = 56) for *A. glutinosa* (Fedorov, 1969) that was not based on chromosome counting of European samples, but American material collected by Woodworth (1929, 1931). Recently, Lepais et al. (2013) described putative tetraploid populations from North Africa (Morocco) based on nuclear microsatellite genotyping.

*Alnus glutinosa* and *A. incana* are wind-pollinated, self-incompatible trees of riparian and water-logged habitats (McVean, 1953; Tallantire, 1974; Douda et al., 2012). The distribution of the two species differs significantly (Jalas and Suominen, 1976). *Alnus glutinosa* grows in lowlands and midlands throughout Europe, except the extreme north, extending as far as Siberia and the mountains of Turkey and North Africa (McVean, 1953). It is common in southern Fennoscandia, but northwards it is associated with a coastal oceanic climate (Tallantire, 1974). The range of *A. incana* is divided into a northern and a southern part, similar to that of Norway spruce (*Picea abies*) (Jalas and Suominen, 1976). However, the distribution range of the species during the LGM was probably significantly wider, also covering western Europe and part of the Iberian Peninsula, as shown by Svenning et al. (2008) using species distribution modelling. Both species were therefore probably in contact elsewhere across Europe throughout the last Ice Age.

Some individuals of *A. glutinosa* collected in 2011 in the Iberian Peninsula and analysed using microsatellites turned out not to be diploids, contrary to reports of numerous authors (but see Fedorov, 1969; Lepais et al., 2013), but probably tetraploids. The occurrence of a previously nearly unknown ploidy prompted the study of cytogeographical patterns in Europe using a method other than microsatellites which do not provide us with precise information concerning the ploidy of all collected individuals. We therefore conducted extensive screening of *A. glutinosa* populations across its distribution range to determine ploidy distribution and also included *A. incana* to test whether tetraploids are of an auto- or allopolyploid origin.

One of the hypotheses tested was whether the distribution of different ploidy levels fits the classical concept of increasing ploidy frequency from southern to northern Europe, i.e. whether polyploids are more likely to persist under extreme climatic conditions. The following questions were asked. (1) What are the frequencies and distribution patterns of plants of different ploidy levels? (2) Are there mixed-cytotype populations? (3) Are particular cytotypes restricted to certain geographical areas? (4) What is the origin of the tetraploid cytotype? Specifically, have tetraploids evolved by autoploidy or are they a result of hybridization with *A. incana* followed by a polyploidization event? (5) What is the Late Pleistocene history of the tetraploid cytotype?

**MATERIALS AND METHODS**

**Study area and sampling procedure**

The research area covers most of the European distribution range of *Alnus glutinosa*, although populations from Siberia were not included and the area of the Turkish mountains was
represented by only a few samples. Samples of *A. glutinosa* were collected across this distribution range in summer months between 2011 and 2013 exclusively from natural unmanaged forest stands. Several sampling strategies were applied to include as many samples as possible in the analysis.

1. **Detailed sampling (94 localities)** – a stratified random sampling procedure was used to sample populations and individuals within populations. Populations were at least 100 km apart. Whenever possible, 20 individuals per population were collected. In each population, samples were collected along a linear transect from individuals at least 50 m apart, i.e. each population represents an area of at least 1 km (Fig. 1; Supplementary Data Table S1). Areas with the detected presence of tetraploids were visited several times to ascertain ploidy homogeneity of populations and delimit better the boundaries between diploid and tetraploid populations. Some populations are therefore listed under ‘detailed sampling’ although they comprised <20 collected individuals and were <100 km apart.

2. **Coarse sampling (70 localities)** – at least three individuals per population were collected in some part of the range by our collaborators. In these cases, individuals were at least 500 m apart (Fig. 1; Supplementary Data Table S1).

3. **Samples from seed collections (45 localities)** – to extend the number of examined populations, we also included populations from across Europe maintained by the International Alder Seed Bank at the Research Institute for Nature and Forest in Belgium (Fig. 1; Supplementary Data Table S1). Plants were cultivated in the experimental garden of the Institute of Botany, The Czech Academy of Sciences, Průhonice, Czech Republic (49°59′30″N, 14°34′00″E, approx. 320 m above sea level). Seeds were germinated in 5 × 5 cm bedding cells with homogeneous garden compost and later moved from the bedding cells to 19 × 19 × 19 cm (6.9 L) pots filled with a common garden substrate.

For detailed sampling, all samples were stored in silica gel, and, whenever possible, fresh leaves were transported to the

![Fig. 1. Geographical locations of populations of *Alnus glutinosa* and *A. incana* across the entire study area. *A. incana*, yellow circles; *A. glutinosa*, others symbols and colours. Different sampling strategies are represented by different symbols: circles, detailed sampling; squares, coarse sampling; triangles, samples from seed collections (see the Materials and Methods for details). Cytotypes in *A. glutinosa* populations are represented by different colours: green, diploid populations; red, tetraploid populations; violet, mixed populations of diploids and tetraploids; light blue, mixed populations of diploids and triploids; white, mixed populations of triploids and tetraploids; dark blue, mixed populations of diploids, triploids and tetraploids.](image-url)
laboratory and analysed immediately. This, however, was not the case for all samples because some expeditions were too long to keep samples fresh enough for flow cytometry analysis or DNA extraction. All coarse samples were stored in silica gel and analysed later, whereas all samples from collections planted in the experimental garden were analysed fresh.

To test the hypothesis that tetraploid *A. glutinosa* has an allopolyploid origin with *A. incana* serving as one parental species, we also collected samples of *A. incana* in the same way as in the detailed sampling of *A. glutinosa*. Only populations growing in southern Europe, i.e. close to tetraploid populations, were included in the microsatellite analysis (see Fig. 1; Supplementary Data Table S1).

**Estimation of DNA ploidy and genome size**

DNA ploidy (Suda et al., 2006) and absolute genome sizes (C-values; Greilhuber et al., 2005) of *Alnus* species were estimated using flow cytometry. For flow cytometry analyses, 209 populations and 2200 individuals of *A. glutinosa* were collected in total. Thirty-three plants from 13 populations were subjected to analysis of absolute genome size, i.e. 17 plants estimated by flow cytometry as diploids from populations 1, 85, 86, 91, 98 and 100, and 16 plants estimated as tetraploids from populations 12, 34, 48, 63, 66, 102 and 106 (see Supplementary Data Table S1 for exact locations). Both fresh leaves and leaves stored in silica gel were used for analyses. Young, intact leaf tissue of the analysed plants and an appropriate amount of leaf tissue of the internal reference standard [Bellis perennis; 2C-value set to 3.38 pg following Schönswetter et al. (2007)] were co-chopped using a sharp razor blade in a plastic Petri dish containing 0.5 mL of ice-cold Otto I buffer (0.1 M citric acid, 0.5 % Tween-20) (Otto, 1990; Dolezˇel et al., 2007). The crude suspension was filtered through a 0.2-μm nylon mesh to remove tissue debris and then incubated for at least 30 min at room temperature. Isolated nuclei were stained with 1 mL of Otto II buffer (0.4 M Na2HPO4·12H2O) supplemented with the AT-selective fluorochrome 4′,6-diamidino-2-phenylindole (DAPI) and β-mercaptoethanol at final concentrations of 4 and 2 μg mL−1, respectively. Immediately after staining, the relative fluorescence intensity of at least 3000 particles was recorded on a CyFlow Space flow cytometer (Partec GmbH, Münster, Germany) equipped with a diode UV chip as an excitation light source.

A different staining procedure was used for absolute genome size estimation. The suspension of isolated nuclei was stained with a solution containing 1 mL of Otto II buffer (0.4 M Na2HPO4·12H2O), β-mercaptoethanol (final concentration of 2 μL mL−1), propidium iodide and RNase IIA (both at final concentrations of 50 μg mL−1). Samples were stained for 5 min at room temperature before being run through a CyFlow SL flow cytometer (Partec GmbH). Isolated stained nuclei were excited with a laser beam of 532 nm (solid-state laser Samba, Cobolt AB, Solna, Sweden), and the fluorescence intensity of 5000 particles was recorded. Resulting histograms were evaluated using the application FloMax (Partec GmbH); DNA ploidy and absolute genome sizes were determined on the basis of the sample/standard ratio. Each plant was analysed separately. Our previous pilot study confirmed the lack of variation in the sample/standard ratio between fresh and silica-dried samples analysed in the same way. The reliability of flow cytometry measurements (i.e. between-plant differences) was repeatedly confirmed in simultaneous runs of *Alnus* accessions yielding distinct fluorescence intensities (i.e. resulting in furcate double peaks in flow cytometry histograms (Greilhuber, 2005).

**Chromosome counts**

To confirm the reliability of the ploidy estimates, flow cytometry results were supplemented by conventional chromosome counts. Two diploids (locality no. 1) and two tetraploids (locality no. 48) were analysed. Chromosome counts were obtained from somatic mitotic cells in root tips of pot-cultivated plants. The root tips were pre-treated in a saturated water solution of p-dichlorobenzene for approx. 2 h, then fixed in a 3:1 mixture of 96 % ethanol and acetic acid, macerated in 1:1 mixture of ethanol and hydrochloric acid for 15 s, washed in water and stained with lacto-propionic orcein. The number of chromosomes was determined under an NU Zeiss microscope with an Olympus E 510 camera attached.

**Microsatellite analysis**

**DNA extraction.** Fresh leaves were collected and stored in silica gel. DNA from *A. glutinosa* and *A. incana* samples was isolated using the DNeasy 96 Plant Kit (Qiagen, Germany). The quality and yield of isolated DNA were checked on 1 % agarose gels, and the fluorescence intensity of at least 3000 particles was recorded on a CyFlow Space flow cytometer (Partec GmbH, Münster, Germany) equipped with a diode UV chip as an excitation light source.

**PCR and fragment analyses.** For microsatellite analysis, 31 populations and 619 individuals of *A. glutinosa*, and 10 populations and 194 individuals of *A. incana* were used (see Fig. 1 and Supplementary Data Table S1 for the geographical distribution of populations). We analysed genetic variation at 12 nuclear microsatellite loci. These loci have been cross-amplified from closely related species by Drašnarová et al. (2014) (multiplex PCR 1: A2, A22, A35, A37, A38) or developed specifically for both *Alnus* species by Lepais and Bacles (2011) (multiplex PCR 2: Ag1, Ag5, Ag9, Ag10, Ag13, Ag20, Ag30). Concerning multiplex PCR 1, we used only five microsatellite loci out of ten published in Drašnarová et al. (2014) due to an amplification failure in one of the species studied (A6, A7, A18) and mutations in the sequence region flanking microsatellite loci (A2, A10, A26). Such mutations can lead to erroneous estimates of allele homology because alleles with identical lengths will not have the same number of repeat units (Drašnarová et al., 2014). In the case of multiplex PCR 2, only seven microsatellite loci out of 12 published by Lepais and Bacles (2011) were used. The reasons for rejection of some loci were the same as in the case of multiplex PCR 1, i.e. amplification failure in one of the species studied (Ag14, Ag23, Ag25, Ag35) and different fragment length in homologous loci between species (Ag9).
DNA was amplified using the Qiagen Multiplex PCR kit in a total reaction volume of 5 μL of PCR mix plus 5 μL of mineral oil to keep the PCR mix from evaporating. The mix contained 20–25 ng, i.e. 1 μL of DNA, 0.1–0.5 μm of each primer and 2.5 μL of Master Mix (Qiagen). PCR amplifications were conducted in a Mastercycler (Eppendorf, Germany) under the following conditions for multiplex PCR 1: 15 min of denaturation at 95 °C, followed by 40 cycles at 94 °C for 30 s, 30s at 58 °C, 60s at 72 °C and a final extension of 10 min at 72 °C; and for multiplex PCR 2: 5 min denaturation at 95 °C, 30 cycles at 95 °C for 30 s, 58 °C for 3 min, 72 °C for 30 s and extension of 30 min at 60 °C. PCR products were separated by electrophoresis in an ABI PRISM 3130 sequencer (Applied Biosystems, USA). A 1 μL aliquot of PCR product was mixed with 0.2 μL of GeneScan-500 LIZ (Applied Biosystems) and 12 μL of Hi-Di formamide (Applied Biosystems). Allele sizes were determined using GeneMarker version 2.4.0 (SoftGenetics, USA). A microsatellite locus was treated as missing data after two or more amplification failures.

Data analysis

Genetic diversity. To examine the genetic diversity of diploid and tetraploid populations of A. glutinosa and diploid A. incana, we computed Nei’s (1978) gene diversity (Ht) corrected for sample size and the average number of alleles (A) in each population using SPAGeDi version 1.2 (Hardy and Vekemans, 2002), a program that computes statistics and permutation tests of relatedness and differentiation among populations for organisms of any ploidy (Hardy and Vekemans, 2002). We used analysis of variance (ANOVA) to compare gene diversity (Ht) and the average number of alleles (A) between A. glutinosa and A. incana, and among cytotypes within A. glutinosa (StatSoft, Inc., www.statsoft.com).

Population structure and the origin of tetraploid populations. To resolve the origin of tetraploid populations situated in two distinct geographical areas (Fig. 1), we used two approaches. (1) To examine genetic similarities and relationships among individuals of the two species and among different cytotypes in A. glutinosa, we performed a principal co-ordinates analysis (PCoA) in R (R Core Team, 2014) using the POLYSAT package (Clark and Jasienski, 2011). A pairwise distance matrix among all samples was calculated using Bruvo distances (Bruvo et al., 2004) as implemented in POLYSAT 1.3-0. This measure of genetic distance has been developed specifically for polyploids and takes distances between microsatellite alleles into account without knowledge of the allele copy number (Clark and Jasienski, 2011). (2) Bayesian model-based clustering of microsatellite data was employed using the procedure implemented in STRUCTURE 2.3.3 (Pritchard et al., 2000), accounting for different ploidy and allelic copy ambiguity in the co-dominant data set (Falush et al., 2007). This analysis was run to infer homogeneous genetic clusters and detect genetic admixture. POLYSAT 1.3-0 was used to generate the input data file for STRUCTURE computations. Ten replicates for each K = 1–10 (the user-defined number of clusters) with the burn-in length of 100 000 generations and the data collection of an additional 1 000 000 generations were run, using the admixture model and correlated allele frequencies. We analysed three data sets to determine the origin of polyploid individuals of A. glutinosa.

Palaeo-distribution modelling

We used the maximum entropy machine-learning approach implemented in Maxent 3.3.3 k (Elith et al., 2006; Phillips et al., 2006) to infer the potential present-day distribution of tetraploid populations and their potential range during the LGM (approx. 21 000 years ago) and the Last Interglacial (LIG) periods (approx. 115 000 years ago). The aim was to explore whether the potential distribution of tetraploids during the LGM corresponds to their recent range or whether they have experienced population withdrawal or expansion during the LGM in some areas. For distributional modelling, 33 co-ordinates of sites where tetraploids were detected using flow cytometry analysis were used. Climatic layers for 19 climatic variables were obtained from the WorldClim database (Hijmans et al., 2005; http://www.worldclim.org). The LIG climatic model simulation with 30 s (approx. 1 km) resolution followed Otto-Bliensier et al. (2006). For the LGM, two climatic models with 2.5 min (approx. 4 km) resolution were used: MIROC (Model for Interdisciplinary Research on Climate; Hasumi and Emori, 2004) and CCSM (Community Climate System Model; Collins et al., 2006). First, we clipped all climatic layers to span from 35.90°N to 75.71°N and from 12.07°W to 48.05°E, including all populations of A. glutinosa with determined ploidy. Then we selected the nine most biologically relevant and relatively uncorrelated (r < 0.7) bioclimatic variables, i.e. annual mean temperature (BIO1), temperature seasonality (BIO4), minimum temperature of the coldest month (BIO6), mean temperature of the wettest quarter (BIO9), mean temperature of the warmest quarter (BIO8), mean temperature of the coldest quarter (BIO10), mean temperature of the coldest quarter (BIO11), annual precipitation (BIO12), precipitation seasonality (BIO15) and precipitation of the warmest quarter (BIO18).

In MAXENT, 5-fold cross-validation and default settings with the ‘auto features’ option, 10−5 convergence threshold, 1000 maximum iteration, regularization parameter β = 1 and 100 000 background points were selected. In each of five replicates, 80 % of sites were used as the training data set, and 20 % of the remaining sites formed the testing data set. The quality of the final model was evaluated using the area under the curve (AUC) statistic (Peterson et al., 2008). The resulting
model projections were the medians from five replicates. To delimit the potential area with suitable conditions for the occurrence of tetraploids at present, the LGM and the LIG, we used the ‘equal training sensitivity and specificity’ logistic threshold (Liu et al., 2005). The potential area of long-term stable occurrence of tetraploid populations resulted from an overlap of suitable areas predicted for the present, LGM and LIG.

RESULTS

Cytotype composition

Chromosome numbers were obtained for two putatively diploid and tetraploid individuals, confirming the existence of diploid (2n = 2x = 28) and tetraploid (2n = 4x = 56) cytotypes of A. glutinosa (Fig. 2). Two thousand two hundred A. glutinosa samples from 209 populations were analysed to assess their ploidy, and a sub-set was used to estimate absolute nuclear DNA content (genome size). The three sampling strategies (detailed sampling, coarse sampling and sampling of seed collections) yielded 94, 70 and 45 localities, respectively. Three different DNA ploidy levels were recorded. Of the 209 populations sampled, diploids occurred in 85.6 %, triploids in 1.9 % and tetraploids in 15.8 %. Populations consisted of one, two or three ploidy levels (Fig. 1; Supplementary Data Table S1). Most of the populations sampled (97.6 %) consisted exclusively of one ploidy. Populations comprising two or three ploidy levels were extremely rare (1.4 % and 1.0 %, respectively). Among the populations consisting of a single ploidy, 83.7 % consisted of diploids and 13.9 % consisted of tetraploids.

The distribution of ploidy levels in Europe, North Africa and western Asia (Georgia, Turkey) departed significantly from a random pattern (Fig. 1). Diploid populations prevailed in south-western Asia (Georgia, Turkey) and western Asia (Fig. 1). Tetraploid populations occupied two geographically distinct areas, one in the Iberian Peninsula (Portugal, Spain) and northern Europe (Morocco) and one in the Dinaric Alps and adjacent areas of Greece, Albania, Montenegro, Serbia, and Bosnia and Herzegovina (Fig. 1). We found two mixed populations consisting of all three ploidies in the contact zone between the area dominated by tetraploids in the Dinaric Alps and diploids (Fig. 1). No mixed populations were found in the Iberian Peninsula or North Africa (Fig. 1).

Mean 2C-values ± s.d. for diploids, Iberian tetraploids and Dinaric tetraploids of A. glutinosa cytotypes were estimated to be 1.010 ± 0.010, 2.091 ± 0.016 and 2.070 ± 0.007 pg, respectively. Monoploid genome sizes (1C-values ± s.d.) were 0.505 ± 0.005, 0.522 ± 0.004 and 0.518 ± 0.002 pg for diploids, Iberian tetraploids and Dinaric tetraploids, respectively (significantly different according to one-way ANOVA, F2, 33 = 61.4, P < 10^-6). Diploids differed significantly from tetraploid populations in both areas, but tetraploid populations from the Iberian Peninsula and the Dinaric Alps did not differ significantly at P < 0.05 according to a multiple-range Tukey’s test.

Gene diversity

The overall mean number of alleles per locus (A) and gene diversity (Hs) for A. glutinosa and A. incana (see Supplementary Data Table S1 for individual values) differed significantly (F1, 492 = 91.98, P < 10^-2 and F1, 492 = 54.69, P < 10^-6, respectively); A. glutinosa reached higher values (Table 1). A comparison of the number of alleles per locus (A) and gene diversity (Hs) in A. glutinosa for different regions characterized by the occurrence of different ploidy levels also yielded a significant result (F3, 372 = 23.29, P < 10^-6, F3, 372 = 9.11, P < 10^-3). The highest genetic diversity was found in tetraploid populations of A. glutinosa from the Balkan Peninsula and the lowest was found in diploid populations from North Africa (Table 1).

Origin of polyploids and gene exchange between diploids and polyploids

The PCoA clearly separated A. glutinosa and A. incana as two distinct species irrespective of ploidy in A. glutinosa (Fig. 3). The same was true for the STRUCTURE analyses of the whole data set. Similarity coefficients indicated that two and three clusters best explained the genetic structuring of Alnus populations (Supplementary Data Fig. S1). Two clusters of individuals clearly separated the two species irrespective of ploidy (Fig. 4A), and three clusters corresponded to (1) A. incana; (2) diploid populations of A. glutinosa from Europe and North Africa and tetraploid populations from the Dinaric Alps; and (3) tetraploid A. glutinosa. Frequencies of the different species and gene diversity for the whole data set were estimated with a weighted AMOVA (Table 2).

### Table 1. Summary of the average number of alleles (A) and gene diversities (Hs) for Alnus glutinosa and A. incana, individual ploidy and regions

<table>
<thead>
<tr>
<th>Species</th>
<th>Ploidy</th>
<th>Region</th>
<th>A</th>
<th>Hs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. incana</td>
<td>2x</td>
<td>Europe</td>
<td>3.77</td>
<td>0.437</td>
</tr>
<tr>
<td>A. glutinosa</td>
<td>2x + 4x</td>
<td>Europe + North Africa</td>
<td>6.70</td>
<td>0.643</td>
</tr>
<tr>
<td>A. glutinosa</td>
<td>2x</td>
<td>Europe</td>
<td>6.09a</td>
<td>0.610a</td>
</tr>
<tr>
<td></td>
<td>2x</td>
<td>North Africa</td>
<td>4.67b</td>
<td>0.522b</td>
</tr>
<tr>
<td></td>
<td>4x</td>
<td>Iberia</td>
<td>7.97a</td>
<td>0.721a</td>
</tr>
<tr>
<td></td>
<td>4x</td>
<td>Balkan</td>
<td>8.50c</td>
<td>0.730c</td>
</tr>
</tbody>
</table>

Means within a column followed by a different superscript letter in A. glutinosa were significantly different at P < 0.05 in multiple range Tukey’s test.
Alnus glutinosa populations from the Iberian Peninsula and North Africa (Fig. 4A). The hypothesis that A. glutinosa is of allopolyploid origin and that A. incana is one of the parental species is therefore not supported, suggesting that tetraploid cytotypes of A. glutinosa are probably of autoploid origin.

Separate analyses of all A. glutinosa populations yielded a similar pattern. The similarity coefficients indicated that two clusters best explained the genetic structuring of A. glutinosa populations (Supplementary Data Fig. S2). Two clusters of individuals clearly separated Iberian and North African tetraploid populations from European diploid and Dinaric tetraploid populations (Fig. 4B).

The analysis of Balkan populations combining exclusively diploid and tetraploid populations together with mixed populations comprising di, tri- and tetraploid individuals and one Austrian population (13) comprising diploids and a triploid showed that two clusters best explained the genetic structuring of these populations (Supplementary Data Fig. S3). The analysis clearly separated diploid and tetraploid populations and individuals of different ploidy from cytotypically mixed populations (Fig. 4C). All triploids in Balkan populations had a high proportion of the tetraploid genetic cluster (Fig. 4C), indicating that they have not arisen by fusion of reduced and unreduced gametes of diploids, but by hybridization of diploid and tetraploid individuals. In this case, tetraploids provided two-thirds of the genome. This was reflected in the STRUCTURE analysis, which assigned triploids to the tetraploid genetic cluster. A different origin was detected in the case of the triploid individual from the Austrian population assigned to the diploid genetic cluster (Fig. 4C). This means that the triploid evolved by fusion of reduced and unreduced gametes of diploid individuals with no participation of tetraploids.

Palaeo-distribution modelling

A high value of the operating characteristic curve (AUC = 0.967) indicated a better than random prediction of tetraploid distribution based on nine BIOCLIM variables. A prediction of the recent range showed that the area with suitable conditions for tetraploid populations is larger than their current occurrence, also including large areas of temperate Europe (Figs. 5 and 6). In contrast, LGM models suggested that the distribution of tetraploids was mainly restricted to the Mediterranean area (Figs. 5 and 6). Balkan and Dinaric populations retreated to the western coastal area of the Balkan Peninsula based on both the MIROC and the CCSM climatic models (Figs. 5 and 6). Both LGM climatic models supported a continuous distribution of tetraploids in lowlands of southern Portugal and Spain connecting African and Iberian populations (Figs. 5 and 6). The occurrence of tetraploids in northern Portugal, north-eastern Spain and southern France was predicted only by the MIROC climatic model (Fig. 5). An overlap among recent, LGM and LIG predictions based on the MIROC model indicated suitable conditions for survival of tetraploid populations in the Balkan, Iberian and Apennine Peninsulas and North Africa, at least since the LIG (Fig. 5).

DISCUSSION

Distribution and origin of tetraploid cytotypes

Analysis of ploidy variation in populations of Alnus glutinosa all across Europe combined with microsatellite analysis and palaeodistribution modelling allowed us to infer the distribution, origins and survival of different cytotypes of A. glutinosa. Alnus glutinosa has been considered a diploid species in a large number of cytological studies from many European countries (e.g. Fedorov, 1969; Goldblatt, 1981, 1985, 1988; Goldblatt and Johnson, 1990, 1991, 1994, 2000, 2003; Ivanova et al., 2006). The first indication that this might not be entirely true was presented by Woodworth (1929, 1931), who was later cited by Fedorov (1969). Woodworth discovered tetraploid individuals of A. glutinosa for the first time, but he did not work with European samples directly, collecting the material for his studies at the Arnold Arboretum of Harvard University, Boston.
FIG. 4. Percentage assignment of *Alnus* individuals (represented by vertical bars) to each of the determined clusters (represented by different colours) inferred by STRUCTURE. Site codes (see Supplementary Data Table S1) indicate the geographical location of individuals along the x-axis. Ploidy and the species to which each population belongs are indicated for each population at the top of the figure. (A) Percentage assignment of diploid, tetraploid and cytotypically mixed populations of *Alnus glutinosa* and *A. incana*. (B) Percentage assignment of diploid, tetraploid and cytotypically mixed populations of *A. glutinosa* only. (C) Percentage assignment of pure diploid, pure tetraploid and cytotypically mixed Balkan populations of *Alnus glutinosa* and one cytotypically mixed Austrian population (13). Diploids and triploids from cytotypically mixed populations (13, 32, 35) are marked by arrows.
Alnus glutinosa is not native to North America, and therefore these individuals must have been introduced to the Arnold Arboretum from somewhere. It is impossible to identify the location of the source population, because the source trees that were growing at the Arnold Arboretum around the period when Woodworth worked there are no longer alive and there is no chance of determining the source population by genetic means (Michael S. Dosmann, Curator of Living Collections, The Arnold Arboretum of Harvard University, pers. comm.). The possibility that A. glutinosa has more than one cytotype in Europe has been completely overlooked, probably because there are so many diploid chromosome counts and because Woodworth’s counts seem doubtful. A second study pointing out the existence of tetraploid A. glutinosa has recently been published by Lepais et al. (2013), who reported tetraploid populations from Morocco based on microsatellite genotyping. Lepais et al. (2013) postulated that these relictual populations could have evolved by hybridization between differentiated A. glutinosa populations and subsequent polyploidization at the southern edge of the distribution range. In this study, we for the first time report the occurrence of tetraploid A. glutinosa in Europe. The distribution of tetraploids is far from random, as they form two geographically well delimited populations. The first is situated in the Iberian Peninsula, extending to North Africa, where it has previously been reported by Lepais et al. (2013). The second one is in the Dinaric Alps, extending to south-western Greece. Bayesian clustering analysis revealed a clear pattern of genetic structure spanning A. incana and diploid and tetraploid populations of A. glutinosa. The PCoA yielded the same results. Both tetraploid populations are therefore probably of autoploid origin, with no indication that A. incana has been involved in their evolutionary history.

Establishment and maintenance of polyploids

Two scenarios have been proposed to explain differences in patterns of cytotype distributions. First, the adaptive evolutionary scenario is based on the assumption that newly arising polyploids possess novel genetic combinations enabling them to thrive under a wider range of ecological conditions (Levin, 1983; Soltis and Soltis, 1993; Otto and Whitton, 2000; Soltis et al., 2003). It has been repeatedly shown that polyploid populations are able successfully to colonize different niches from those inhabited by their progenitors (Flégróvá and Krahulec, 1999; Ramsey, 2011; Mráz et al., 2012; Collins et al., 2013; Hahn et al., 2013). Due to ecological sorting along abiotic or biotic gradients, polyploids can occupy different ecological niches, resulting in spatial segregation in diploid-polyploid complexes. The other scenario is the non-adaptive scenario (the so-called ‘minority cytotype exclusion model’, Levin, 1975; Fowler and Levin, 1984; Ramsey and Schemske, 1998) used to explain spatial segregation by frequency-dependent mating success that results from low fitness of hybrids formed from between-cytotype mating gradually leading to the elimination of the minority cytotype from the population. In this case,

![Fig. 5. Distributions of Alnus glutinosa tetraploids since (A) the Last Glacial Maximum (LGM; approx. 21,000 years ago) and (B) the Last Interglacial (LIG; approx. 115,000 years ago) predicted based on the MIROC climatic model. The equal training sensitivity and specificity threshold of 0.143 were used for delimiting the area with suitable climatic conditions. ‘Recent + LGM’ and ‘Recent + LGM + LIG’ mark areas predicted as suitable in more than one time period. Grey in coastal zones indicates land areas above the sea level during the LGM.](https://academic.oup.com/aob/article-abstract/117/1/107/2195961)
cytologically uniform populations occur in different locations and the coexistence of multiple cytotypes is viewed as a transient phenomenon.

Although our knowledge concerning the distribution of *A. glutinosa* cytotypes in the field does not allow an accurate distinction between the adaptive and non-adaptive scenarios (Levin, 1975; Husband et al., 2013), many studies of other groups provide information about contact zones between or among cytotypes. There are two types of contact zones: narrow contact zones with only a few cytotypically mixed populations (Husband and Schemske, 1998; Baack, 2004) and contact zones extending over large areas (Burton and Husband, 1999; Halverson et al., 2008; Duchoslav et al., 2010, Fialová et al., 2014). Even in areas where the distribution of two cytotypes is diffuse, mixed-cytotype populations tend to be quite rare (Kolář et al., 2009; Trávníček et al., 2010; Castro et al., 2012). Contact zones may result from secondary contact between previously allopatric chromosomal races (secondary contact zones; Petit et al., 2002) or the expansion of newly formed polyploids from within diploid populations. It is, however, difficult to apply the concept of secondary contact zones to *A. glutinosa* cytotypes. In our view, the distribution of its cytotypes has been formed by past climatic changes during glacial and interglacial times.

In the case of *A. glutinosa*, the distribution areas of diploids and tetraploids overlap only to a small degree in the Dinaric Alps. Diploid and tetraploid populations are almost parapatric, suggesting differences in ecological tolerance. Pure tetraploid populations occur almost exclusively at the bottoms of deep valleys, whereas diploids are distributed at lower elevations around core tetraploid populations. The observed pattern might be explainable by variation in ecological tolerance, as polyploids have a wider spectrum of tolerance and are adapted to ecological conditions not suitable for diploids (Levin, 1983).

The small number of triploids found suggests restricted gene exchange and increasing reproductive isolation. However, the occurrence of triploids was detected in both (1) populations situated within the range of diploids, i.e. locality 13 in Austria, and (2) on the border between diploid and tetraploid populations ranges, i.e. localities 32 and 35. In the case of population 13, the triploid individual originated by fusion of reduced and unreduced gametes, as has been shown by the STRUCTURE analysis. On the other hand, triploids in Balkan populations originated by hybridization between diploid and tetraploid individuals. We have thus demonstrated restricted gene flow between diploids and tetraploids that could probably have occurred anywhere individuals of different ploidy are in contact.

**Glacial refugia for Alnus glutinosa and ploidy distribution**

Two previous studies have considered the post-glacial migration of *A. glutinosa* in Europe (King and Ferris, 1998; Douda et al., 2014). The first used variation in plastid DNA to reveal the main migration routes of the species after the retreat of glaciers and located LGM refugia in the Carpathians. Douda et al.
(2014), in contrast, showed, based on a review of radiocarbon-dated pollen and macrofossil sites for species of Alnus subgenus Alnus from the last 50,000 years in Europe, that an expansion of Alnus is supported by the presence of alders during the LGM in southern peninsulas and northern areas of western Europe, the foothills of the Alps, the Carpathians and north-eastern Europe. However, post-glacial recolonization routes have been interpreted without information about cytotype variation of A. glutinosa in Europe. Tetraploids were found in two separate populations located in two important glacial refugia, one in the Iberian Peninsula and one in the western Balkan Peninsula. The results of this study corroborate previous studies indicating that northward post-glacial expansion from the Iberian Peninsula is unlikely and also rules out the western part of the Balkan Peninsula as a putative source refugium. Therefore, diploid populations located in southern European peninsulas that might be taken into consideration for northward expansion are located only along the border between the Iberian Peninsula and Europe, i.e. in the Pyrenees, in the eastern part of the Balkan Peninsula, the Apennine Peninsula and more northern refugial areas suggested by Douda et al. (2014). To sum up, because some putative glacial refugia harbour the tetraploid cytotype, these areas could not have served as effective refugia for A. glutinosa diploids growing in the rest of Europe.

**Ecological niche models and tetraploid glacial refugia**

Areas predicted as currently suitable for tetraploids of A. glutinosa by ecological niche model projections onto current climate layers cover a larger area than the actual distribution of the species. Iberian tetraploid populations might occur in western France, southern Ireland, western Great Britain and the foothills of the Alps and certain areas of Germany. The same is true for tetraploids of the Dinaric Alps, the predicted area of which covers the whole of the former Yugoslavia, extending to south-eastern Greece. The wider distribution predicted by the model compared with the actual situation might not be simply explainable by sparse sampling in the predicted areas because the whole of Europe was sampled quite densely and evenly. One reason for this mismatch might reside in the fact that species distribution models do not consider information on species history, dispersal ability, water availability, soil types, competing vegetation or genetic diversity, and assume a species–climate equilibrium, namely that species occupy all environmentally suitable areas (Nogués-Bravo, 2009; Angert et al., 2011; Abeli et al., 2013).

The projection of the distribution model into the LGM situated the distribution range of the tetraploid cytotype exclusively in the Mediterranean area. It has been shown that populations found in the Iberian Peninsula and North Africa are likely to have been interconnected during the LGM, which is also in agreement with the genetic structure: tetraploid populations from the Iberian Peninsula fall in the same genetic cluster as populations from North Africa. Moreover, some Iberian populations were probably present at localities of continuous occurrence during the last glacial period. This should support the view that formerly recognized refugia did not constitute single refugia throughout the Quaternary, but instead several independent or interconnected ‘refugia within refugia’ (Gómez and Lunt, 2006; Nieto-Feliner, 2011; Fernández-Mazuecos and Vargas, 2013). Hence, as Gutiérrez-Larena et al. (2002) and Martín-Bravo et al. (2010) argued in the case of the Quaternary history of Mediterranean species, elevational migrations were more likely than large-scale range shifts. Long-term isolation within each refugium led to the evolution of genetically distinct tetraploid populations in Europe and North Africa and diploid populations in North Africa.

However, a different picture might be drawn in the case of tetraploid populations occurring in the Dinaric Alps. Most of them probably did not exist throughout the last glacial period and found refuge further south in lowland and coastal areas of the Balkan Peninsula. By combining data from ecological niche models and Bayesian analyses of microsatellite data, it is therefore possible to postulate two contrasting scenarios for each area. Some populations in the Iberian Peninsula were probably stable over a long period of time, whereas tetraploid populations in the Dinaric Alps withdrew during glacial times and expanded to new suitable localities in interglacial times. These populations now survive at the bottom of deep river valleys with specific microclimatic conditions, e.g. in the canyons of the Tara and Mrtvica rivers in Montenegro.

**Conclusions**

Our cytological and molecular data combined with species distribution modelling provide new insights into the cytological variation, cytotype origin and late Quaternary history of A. glutinosa. Newly discovered tetraploid populations in the Iberian Peninsula and the Dinaric Alps probably persisted there for a long time. Both tetraploid clusters are situated in the putative main glacial refugia, and neither was probably involved in the colonization of central and northern Europe after glacial retreat. This means that neither the Iberian Peninsula nor the western part of the Balkan Peninsula served as effective refugial areas for northward post-glacial expansion because the ploidy is different. The diploid populations currently growing in most of Europe are therefore likely to have descended from populations located in western Europe and the eastern Balkan Peninsula or the Apennine Peninsula. This result partly supports the outcomes of King and Ferris (1998), who placed the main refugium of A. glutinosa in the Carpathians, and Douda et al. (2014), who emphasized the importance of eastern refugia (i.e. Belarus and western Russia) and also individual regional refugia serving as important centres of local spread.

**SUPPLEMENTARY DATA**

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Table S1: summary of geographic locations and genetic diversities of Alnus glutinosa and A. incana populations. Figures S1–S3: average similarity coefficients for each K with standard deviations.

**ACKNOWLEDGEMENTS**

We would like to thank Martina Hadincová and Eva Ničová for their help in the laboratory, and Fred Rooks for helping us out with the language in the manuscript. Michael S. Dosmann

**REFERENCES**

Abeli et al. — Distribution and origin of cytotypes in Alnus glutinosa

Mandák et al. — Distribution and origin of cytotypes in Alnus glutinosa
kindly informed us about the history of *Alnus* accessions growing at the Arnold Arboretum of Harvard University. We also wish to thank all the people who collected samples of *Alnus* for us, especially Filip Kolář, Vít Dvořák, Milan Stech and Boudewijn Michiels. Serge Müller deserves thanks for sharing samples from North Africa. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme ‘Projects of Large Infrastructure for Research, Development, and Innovations’ (LM2010005), is greatly appreciated. This study was supported by the Grant Agency of the Czech Republic (P504/11/0402) and as part of long-term research development project (RVO 67985939) and project Postdoc USB (CZ.1.07/2.3.00/30.0006).

**LITERATURE CITED**


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


