The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

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

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

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

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

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

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

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

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

INTRODUCTION

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

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

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

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

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

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

MATERIAL AND METHODS

Field sampling

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

Flow cytometry

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

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

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

Chromosome preparations

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

Data analyses

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

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

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

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

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

RESULTS

Ploidy level variation and cytogeography

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

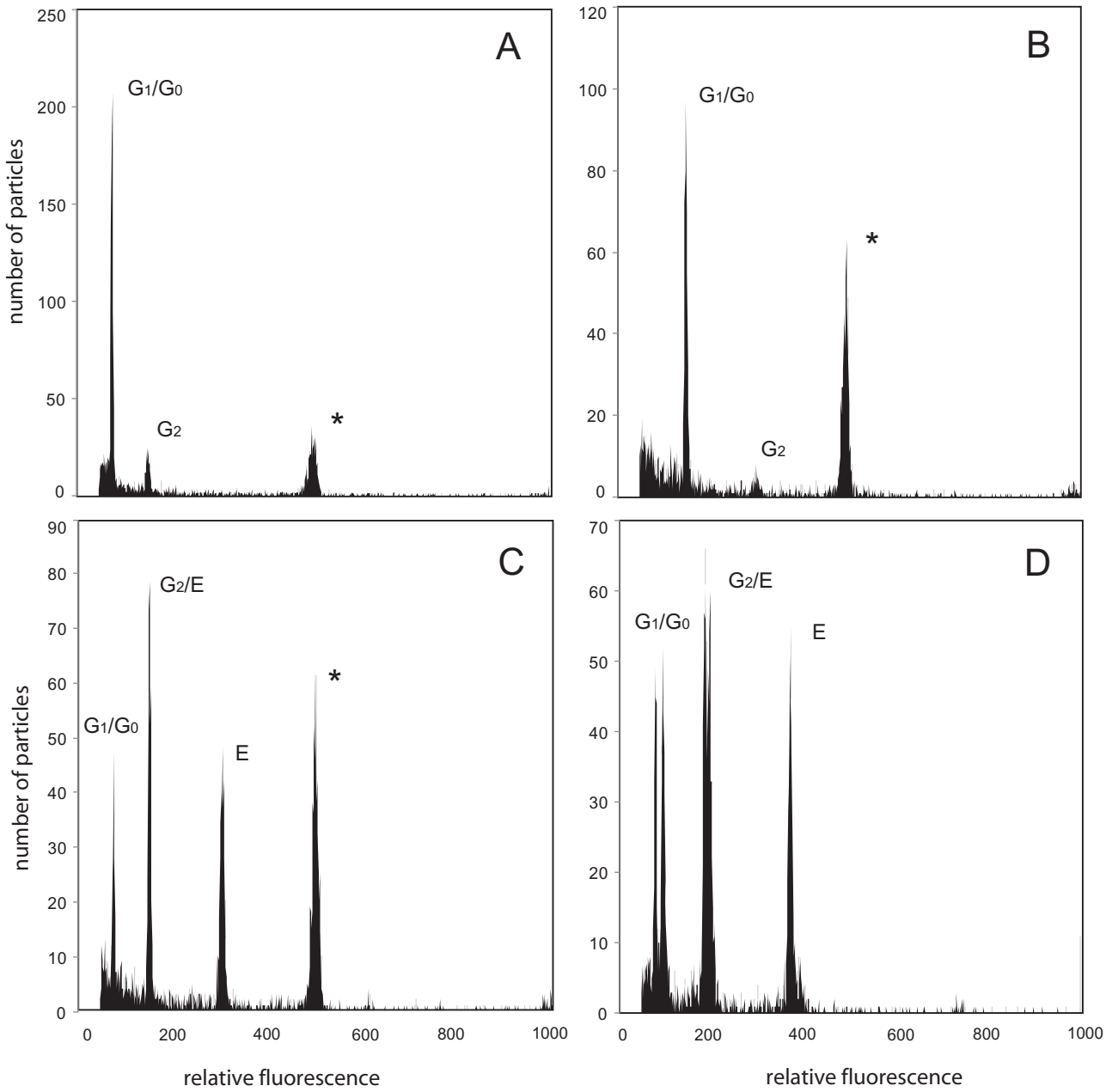

Figure 1. Flow cytometric histograms of suspensions of DAPI-stained nuclei isolated from diploid (A, C, D) and tetraploid (B) accessions of the *Arabidopsis arenosa* group. A + C, Analysis of nuclei of identical diploid individuals (pop. AA084) isolated from either fresh petal (A) or stem leaf (C). B,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

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

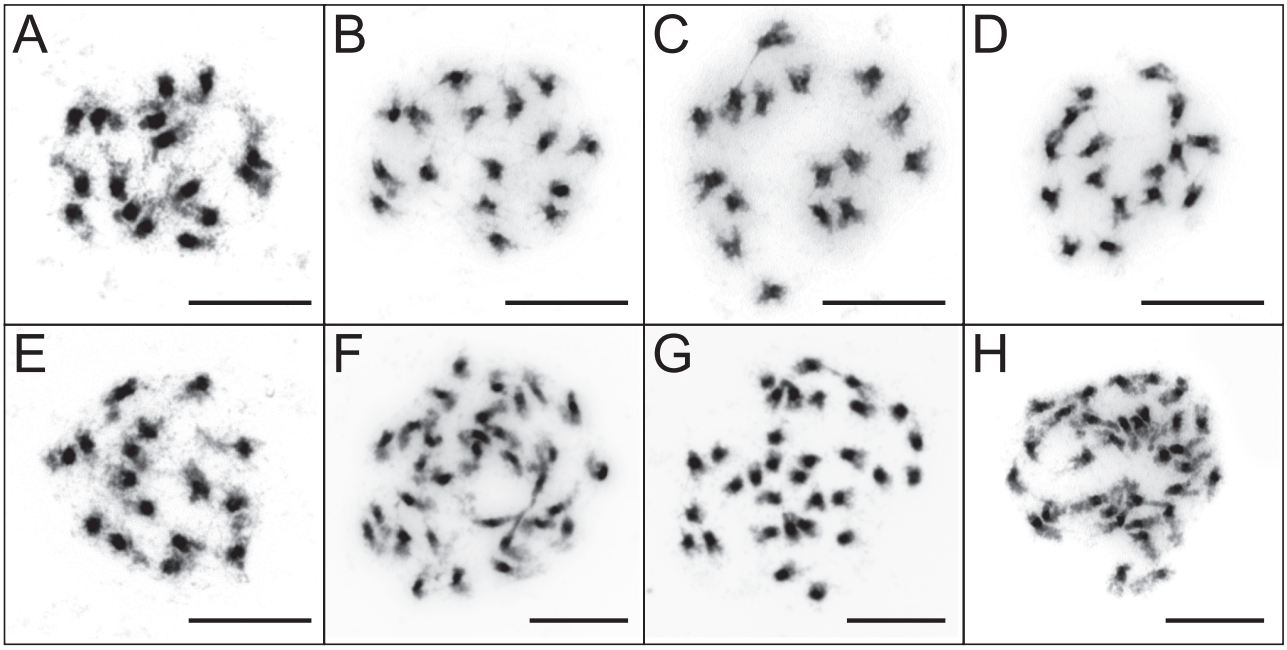

Figure 2. DAPI-stained mitotic chromosome spreads from flower bud tissue of the *Arabidopsis arenosa* group. A, *A. arenosa* s.l. AA200 (Lithuania, coastal sands; $2n = 2x = 16$). B, *A. arenosa* s.l. AA124 (Serbia, dry rocks; $2n = 2x = 16$). C, *A. arenosa* s.l. AA070 (Romania, dry rocks; $2n = 2x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

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

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

Homoploid differentiation in DNA content

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

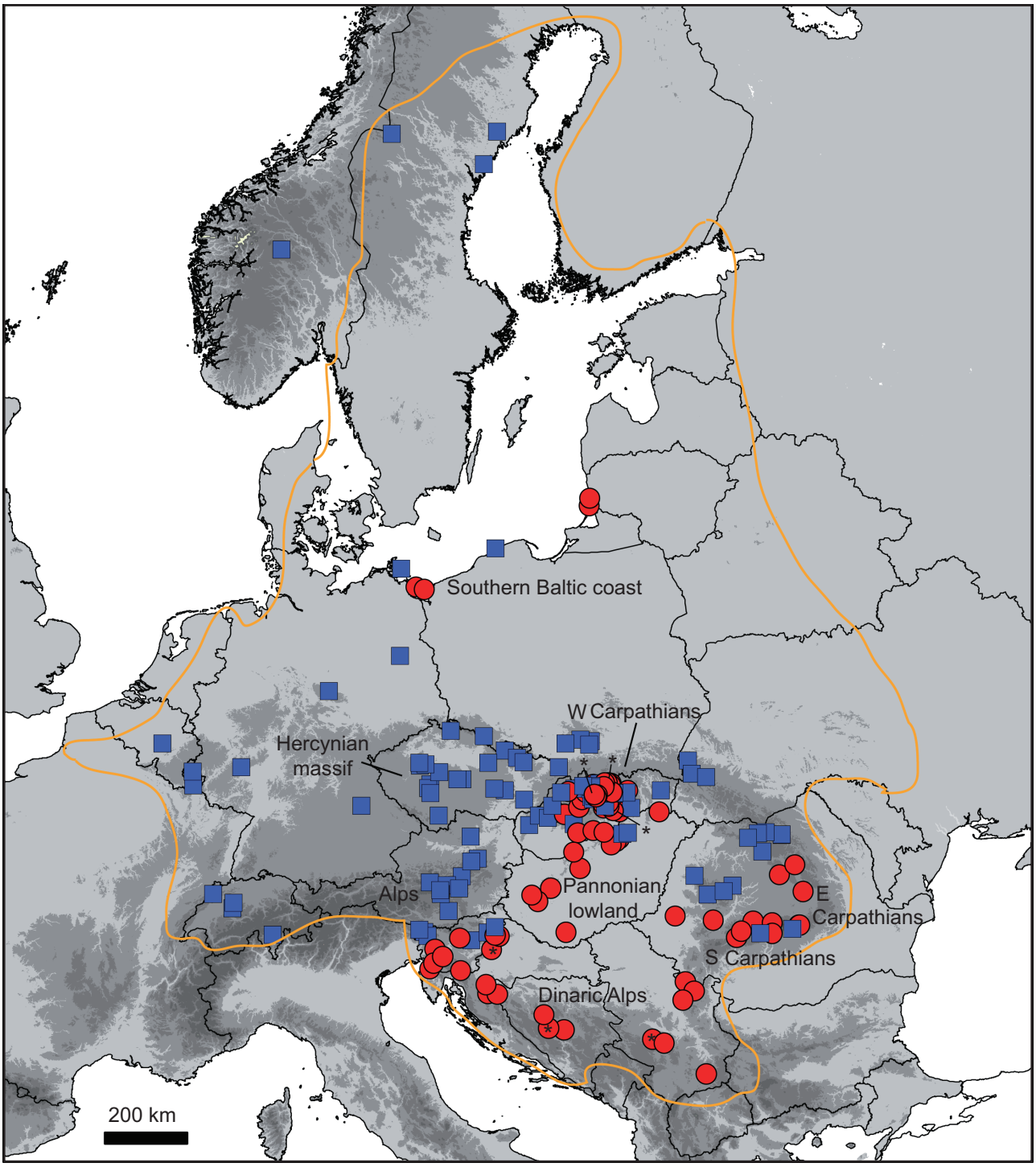

Figure 3. Distribution and ploidy level of the 194 studied populations of the *Arabidopsis arenosa* group in Europe (red – diploid, blue – tetraploid, asterisk – triploid, 2963 individuals investigated in total). The continuous distribution range of the whole species complex is marked by the orange outline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

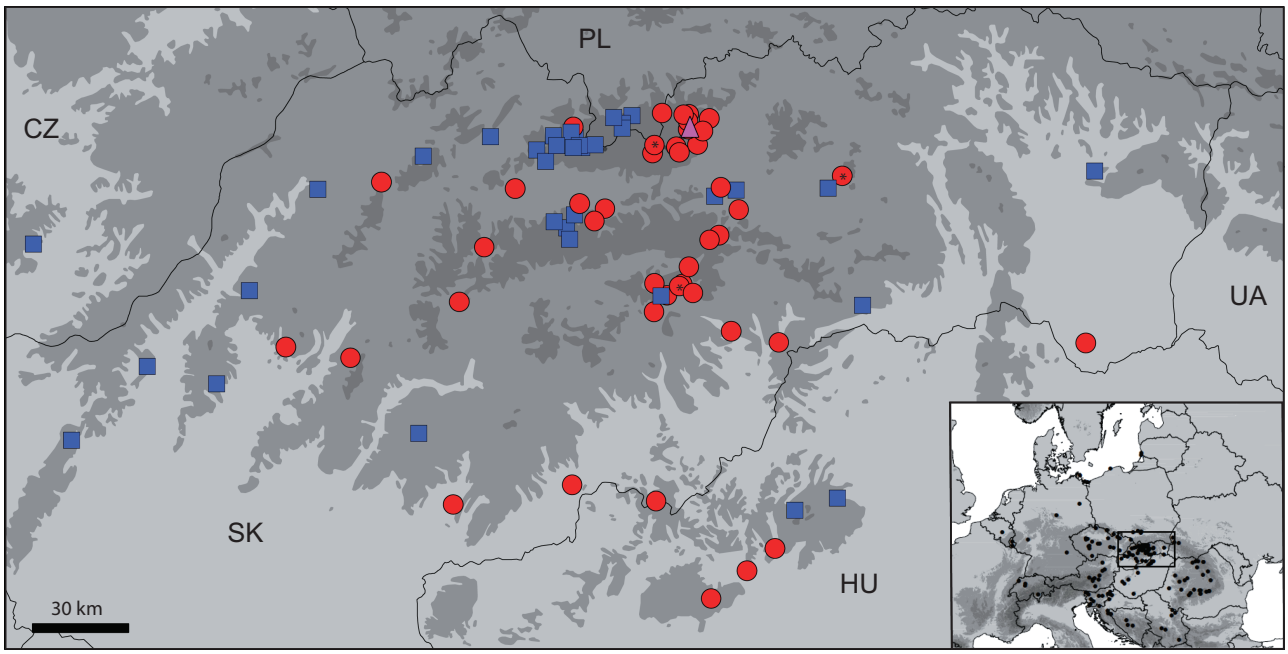

Figure 4. Detail of the contact zone of cytotypes of the *Arabidopsis arenosa* group in the Western Carpathians (red circle – diploid, blue square – tetraploid, pink triangle – mixed di-tetraploid population, asterisk – triploid; based on 1374 individuals from 79 populations).

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

Niche differentiation

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

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

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

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

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

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

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

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

DISCUSSION

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

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

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

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

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

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

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

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

High cytogeographical complexity in the Carpathians contrasts with intrapopulation cytotype uniformity

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

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

colonization history and genetic drift should be taken into account.

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

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

Large niche overlap among cytotypes

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

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

ACKNOWLEDGEMENTS

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

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

SUPPORTING INFORMATION

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

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