Biol Invasions
DOI 10.1007/s10530-016-1133-2

ORIGINAL PAPER

The population genetics of the fundamental cytotype-shift in invasive *Centaurea stoebe* s.l.: genetic diversity, genetic differentiation and small-scale genetic structure differ between cytotypes but not between ranges

Christoph Rosche · Walter Durka · Isabell Hensen · Patrik Mráz · Matthias Hartmann · Heinz Müller-Schärer · Susanne Lachmuth

Received: 3 June 2015 / Accepted: 25 March 2016

Abstract Polyploids are overrepresented in invasive species. Yet, the role of genetic diversity and drift in colonization success of polyploids remains unclear. Here, we investigate genetic diversity, genetic differentiation and small-scale genetic structure in our model system, the three geo-cytotypes of *Centaurea stoebe*: monocarpic diploids and polycarpic (allo)tetraploids coexist in the native range (Eurasia), but only tetraploids are reported from the invasive range (North America). For each geo-cytotype, we investigated 18–20 populations varying in size and habitat type (natural vs. ruderal). Population genetic analyses were conducted at eight microsatellite loci. Compared to diploids, tetraploids revealed higher genetic diversity and lower genetic differentiation, whereas both were comparable in tetraploids between both ranges. Within spatial distances of a few meters, diploid individuals were more strongly related to one another than tetraploids. In addition, expected heterozygosity in diploids increased with population size and was higher in natural than in ruderal habitats. However, neither relationship was found for tetraploids. The higher genetic diversity of tetraploid *C. stoebe* may have enhanced its colonization abilities, if genetic diversity is correlated with fitness and adaptive capabilities. Furthermore, the inheritance of a duplicated chromosome set as well as longevity and frequent gene flow reduces drift in tetraploids. This counteracts genetic depletion during initial introductions and in

Electronic supplementary material The online version of this article (doi:10.1007/s10530-016-1133-2) contains supplementary material, which is available to authorized users.

C. Rosche (✉) · I. Hensen · S. Lachmuth
Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108 Halle, Germany
e-mail: christoph.rosche@botanik.uni-halle.de

C. Rosche · W. Durka
Department Community Ecology (BZF), Helmholtz-Centre for Environmental Research – UFZ, Theodor-Lieser-Straße 4, 06120 Halle, Germany

C. Rosche · P. Mráz · M. Hartmann
Herbarium and Department of Botany, Faculty of Science, Charles University in Prague, Benátská 2, 128 01 Prague, Czech Republic

W. Durka · I. Hensen · S. Lachmuth
German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

H. Müller-Schärer
Department of Biology, Unit of Ecology and Evolution, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland

Published online: 30 March 2016

subsequent phases of small or fluctuating population sizes in ruderal habitats. Our findings advocate the importance of studying colonization genetic processes to gain a more mechanistic understanding of the role of polyploidy in invasion dynamics.

Keywords Biological invasion · *Centaurea stoebe* · Colonization · Genetic diversity · Geo-cytotype · Polyploidy

Introduction

Biological invasions are ecological enigmas: while some non-native species attain astoundingly high abundances in the introduced range (e.g. Shah et al. 2014); the vast majority of exotics fail to establish (Sax and Brown 2000). Understanding the mechanisms that determine the success of invaders has consequently attracted major interest in ecological research (Simberloff et al. 2013). More recently, polyploidy has received increasing attention as it may promote founder and invasion success (Pandit et al. 2014; Bock et al. 2015). Polyploids frequently possess broader ecological amplitudes and higher plasticity than diploids (Soltis et al. 2014). Both may be related to the high genetic diversity in polyploids, which often considerably exceeds that of their diploid ancestors (reviewed in Soltis and Soltis 2000).

Genetic diversity is considered to be a key determinant of invasion dynamics, largely because it inherently affects population growth rates and the adaptive potential of exotics (Forsman 2014). However, dispersal limitation may result in a random loss of overall allele diversity in the introduced range, and colonization usually involves founder effects, including genetic drift and inbreeding (Dlugosch and Parker 2008; Hufbauer et al. 2013; Szűcs et al. 2014). More specifically, colonization and initial range expansion mostly take place at ruderal sites, where higher disturbance frequencies ensure great resource supply (Dietz and Edwards 2006). However, such environmental stochasticity may cause more frequent fluctuations in population size and consequent genetic bottlenecks than in natural habitats. Nevertheless, invasion dynamics are, at least in the long term, not inevitably restricted by massive reductions in genetic diversity (reviewed in Uller and Leimu 2011). Since multiple introductions are rather the rule than the exception (Dlugosch and Parker 2008), invasions may involve both, the fission and fusion of native source gene pools (Keller and Taylor 2010). The resultant admixture of previously isolated gene pools may boost spread by counteracting genetic depletion (Verhoeven et al. 2011). Thus, traits that decelerate the loss of genetic diversity may ultimately facilitate invasion success as they may help founder populations persist though colonization bottlenecks until the gene pool can be restored (Theoharides and Dukes 2007).

Polyploidy antagonizes genetic depletion (Soltis et al. 2014), because genetic drift is known to affect polyploid genomes less strongly than diploid ones due to inheritance of a duplicated set of chromosomes per gamete (Ronfort et al. 1998). In addition, polyploidy often involves a switch from annual to perennial life history (te Beest et al. 2011), which can influence the small-scale genetic structure. Under fluctuating population sizes, perennials may exhibit less frequent (biparental) inbreeding and reduced drift due to their longevity and overlapping generations (Nybom 2004). However, population genetic consequences of ploidy level are poorly understood and can best be investigated within polyploid complexes that comprise diploid and polyploid subspecies (Hardy and Vekemans 2001).

In this study, we examined the genetic structure of spotted knapweed, *Centaurea stoebe* s.l. L. (Asteraceae, syn. *C. maculosa* Lam.), which constitutes a polyploid complex including a diploid (predominantly) monocarpic cytotype and a polycarpic tetraploid cytotype (Mráz et al. 2011). The complex is native to Europe and Asia Minor with diploids representing the majority of native populations (Broennimann et al. 2014). Both cytotypes occupy relatively similar habitats (i.e. dry natural and ruderal sites), but tetraploids are more frequent at ruderal sites (Treier et al. 2009; Otisková et al. 2014). Remarkably, so far, only tetraploids have been reported from the non-native range (Mráz et al. 2011). As such, we distinguish three geo-cytotypes (GCTs, defined by ploidy level and range) as follows: native diploid (EU2x), native tetraploid (EU4x), and invasive tetraploid (NA4x). Te Beest et al. (2011) highlighted this cytotype shift as "an excellent model system for evaluating the role of polyploidy in plant invasions".

A previous microsatellite study from Marrs et al. (2008) particularly focused on the introduction history

of tetraploid *C. stoebe*. Although it included two diploid populations, it did not allow for and did not aim at drawing conclusions about population genetic differences between the two cytotypes. In contrast, we examine the interplay of polyploidy, longevity and demographic history for overcoming founder effects and use the *C. stoebe* complex as a model system to highlight the relevance of polyploidy for the population genetics of colonizing species. Despite increasing awareness of the significance of polyploidy in invasions (e.g. Pandit et al. 2011, 2014), population genetic studies on GCTs are surprisingly scarce (but see Schlaepfer et al. 2008; Ferrero et al. 2015). Our investigations were directed by the following hypotheses:

1. Tetraploids of *C. stoebe* reveal higher genetic diversity than diploids. Current genetic diversity is not reduced in NA4x compared to EU4x.
2. Among population differentiation is stronger in diploids than in tetraploids. Tetraploids are more strongly differentiated in the native range.
3. Within populations, diploid individuals are more closely related on a small spatial scale than tetraploid individuals.
4. Within GCTs, genetic diversity increases with population size. Natural populations reveal higher genetic diversity than ruderal populations.

Our analyses will contribute to better understanding of the cytotype shift in *C. stoebe* and may provide important implications for polyploid versus diploid range dynamics in general.

Materials and methods

The model system *Centaurea stoebe* s.l.

Diploid *C. stoebe* L. subsp. *stoebe* and tetraploid *C. stoebe* L. subsp. *micranthos* (Gugler) Hayek exhibit a strong reproductive barrier (Mráz et al. 2012b). Based on cloned internal transcribed spacers, Mráz et al. (2012a) showed that the tetraploid cytotype originated from allopolyploidization events, which occurred within the last 2 mya, but, the second closely related parental taxon has not yet been identified. Despite the allopolyploid origin of tetraploids, Mráz et al. (unpublished data) found clear evidence for tetrasomic inheritance when they screened the inheritance of four microsatellite loci in controlled crosses of tetraploid plants. Due to its complexity, the nomenclature of both taxa remains unresolved, and the cytotypes are mainly treated at the subspecies level (Mráz et al. 2011). Both cytotypes are strictly self-incompatible; and they have similar gene dispersal capabilities: small Hymenoptera are considered as main pollinators (Mráz et al. 2012b), and achenes are dispersed by barochory with no differences in falling velocity between GCTs (Hahn et al. 2013). Although tetraploids are polycarpic, neither cytotype shows vegetative propagation.

In the native range, diploids are more common in central Europe, while tetraploids prevail in southeastern Europe (Broennimann et al. 2014). However, their native distributions overlap widely and include several mixed-ploidy populations (Mráz et al. 2012b). Moreover, EU4x recently expanded towards central Europe (Ochsmann 2000). In the invasive range, the first recorded introductions of NA4x were in the late nineteenth century, followed by a lag-phase of 50 years. Subsequently, the species spread rapidly along ruderal transport corridors of two separate invasion routes: one expanding from the east coast and one from the west coast (Broennimann et al. 2014). Nowadays, *C. stoebe* is a widespread, notorious weed that causes tremendous economic damage in the North American grasslands (Corn et al. 2006).

Sampling

Extensive field sampling was undertaken across large parts of the native distribution and of the western invasion route in North America, where *C. stoebe* is regarded as one of the most noxious invaders (Maron et al. 2013). Between 2012 and 2014, we sampled at least 18 populations of each GCT. We estimated population size as the number of flowering individuals. Therefore, we counted every flowering individual in populations with up to 500 individuals and rounded the counts. In larger populations, we counted 500 individuals in an area of representative density and extrapolated the population size to the entire population area. All sites were classified according to the European classification system of habitats (EUNIS 2008). Following the protocol of Broennimann et al. (2014), natural and semi-natural grasslands (EUNIS-category E), natural rocky outcrops (H), and diluvial sediments (C) were considered as (semi-)natural (for reason of simplicity referred to as "natural"

throughout this manuscript). Agricultural (I), artificial and industrial habitats (J) were considered as ruderal. We collected leaf samples for genetic analysis and, if available, seeds for flow cytometry, from 19 to 31 haphazardly selected adult individuals per population equally distributed across the population. If populations consisted of fewer than 20 adults, we added samples from rosettes.

Flow cytometry and microsatellite amplification

The ploidy level of all populations was assessed by applying the identical protocol as in Mráz et al. (2011). For population genetic analyses of mixed-ploidy populations, we only made use of leaf samples from individuals for which we determined the majority cytotype. Thus, the microsatellite analyses concerned only one cytotype per population, as sample size of the minority cytotype was commonly too low (Table 1), and gene flow between the cytotypes was shown to be almost absent (Mráz et al. 2012b).

Microsatellite amplification and genotyping

We extracted DNA from 10 to 15 mg of lyophilized leaf tissue with the DNeasy 96 Plant Kit (Qiagen) following the manufacturer's protocol. We tested ten already established microsatellites, eight of which appeared to be highly polymorphic, and showed clear single bands for each allele: CM-730, CM-8337, CM-1922, CM-10060 described in Mráz et al. (2012b), and CM17, 42 CM27, CM26, CD9 from Marrs et al. (2008). Amplification was accomplished with M13R- or CAG-tailed primers in three multiplex PCR reactions. The final volume was 5 µL containing 3 µL QIAGEN Multiplex PCR kit, ~20 ng genomic DNA and 1 µL Mastermix. Mastermix contained 0.25 µM of forward primer (either CAG- or M13R-tailed), 0.25 µM reverse primer and 0.25 µM of the fluorescent-labeled CAG or M13R primer. We applied a touchdown PCR with following conditions: 95 °C for 15 min; 20 cycles of 94 °C for 30 s, 60 °C for 60 s (with an increment of −0.5 °C per cycle), 72 °C for 90 s; 20 cycles of 94 °C for 30 s, 50 °C for 60 s, 72 °C for 90 s; and finally, an elongation step of 10 min at 72 °C. Electropherograms were obtained by migration of amplification products on an ABI 3130 genetic analyzer (Applied Biosystems) with LIZ-500 (internal size standard). To bin the allele sizes, we used GeneMapper 5.0 (Applied Biosystems). 67 samples were deleted from the final data set, because more than one locus failed to amplify. The remaining proportion of missing loci was 2.5 %. In total, we genotyped 1321 individuals. We confirmed reliable and stable patterns at each SSR locus by repeating the amplification of 48 identical samples. Without segregation analysis, peak intensities are not reliable to estimate the quantum of null alleles or allelic doses (Blanchet et al. 2015; Dufresne et al. 2014). Exact genotypes of polyploids can be assigned only when marker phenotypes show a single allele or the number of alleles equals the ploidy level. We therefore choose programs that are robust for dealing with genotype uncertainty and occurrence of null alleles.

Note that allopolyploidy may result in disomic inheritance, which leads to biased population genetic parameter estimates when calculated under the assumption of tetrasomy (for details see Meirmans and van Tienderen 2013). However, in accordance with the above mentioned inheritance screening of microsatellites by Mráz et al. (unpublished data), we also found strong evidence for tetrasomic inheritance, as we did not find fixed heterozygosity at any of our eight loci. Thus, disomic inheritance seems rather unlikely, at least for large parts of the genome of tetraploid *C. stoebe*.

Genetic diversity within populations

Allelic richness (A_R, i.e. number of alleles rarefied to the minimum sample size of 19 individuals per population), was calculated with SPAGeDi 1.4 (Hardy and Vekemans 2002). Expected heterozygosity (H_e) was estimated using the unbiased estimator of Nei (1978) correcting for sample size in SPAGeDi for diploids. For tetraploids, we estimated H_e in ATetra 1.3 (van Puyvelde et al. 2010) by 10,000 Markov Chain Monte Carlo (MCMC) iterations (accounting for different probabilities of allele copy number combinations in partial heterozygotes). We further determined the numbers of private alleles per population, range and cytotype in R 3.12 (R development core team 2014). To test whether rare alleles were more frequent in EU4x than in NA4x, we calculated frequency down-weighted marker values (*DW*) per population according to Schönswetter and Tribsch (2005), and compared $\log_e$ (*DW*) between both tetraploid ranges in a linear model.

Table 1 Investigated populations: characteristics, ploidy levels, indices for genetic diversity and small-scale genetic structure

ID	Country, locality, GPS (°N,°E)	Habitat type	Population size	4x	2x	Sample size	H_e	A_R	*DW*	H_A	b_{log}
Native range, diploid (EU2x)											
1	DE, Federow, 53.48°, 12.76°	Ruderal	40	0	22	20	0.32	2.58	n.e.	0.05	−0.004
2	DE, Feldberg, 53.32°, 13.43°	Ruderal	250	0	27	27	0.55	3.57	n.e.	0.14	n.e.
3	DE, Hillersleben, 52.29°, 11.48°	Ruderal	10	0	15	24	0.56	3.34	n.e.	0.15	n.e.
4	DE, Steinthaleben, 51.39°, 11.04°	Natural	500	0	22	27	0.64	5.09	n.e.	0.86	n.e.
5	DE, Lieskau, 51.5°, 11.86°	Natural	200	0	24	29	0.61	4.43	n.e.	0.93	n.e.
6	DE, Amselgrund, 51.5°, 11.94°	Natural	500	0	25	23	0.63	4.51	n.e.	0.54	−0.041*
7	DE, Neue Göhle, 51.23°, 11.78°	Ruderal	250	0	20	21	0.47	3.25	n.e.	0.07	n.e.
8	DE, Bautzen, 51.18°, 14.42°	Ruderal	9	0	9	26	0.41	3.23	n.e.	0.18	n.e.
9	DE, Isteiner Klotz, 47.66°, 7.53°	Natural	200	0	20	20	0.58	3.72	n.e.	0.82	n.e.
10	CH, Ramosch, 46.83°, 10.4°	Natural	250	0	33	23	0.69	5.53	n.e.	0.31	−0.028
11	IT, Castelle Penede, 45.88°, 10.89°	Natural	140	0	32	23	0.67	4.39	n.e.	0.28	−0.037*
12	IT, Rafenstein, 46.53°, 11.36°	Natural	250	0	30	29	0.73	6.04	n.e.	0.54	−0.006
13	CZ, Rájov, 48.84°, 14.37°	Natural	10	0	12	31	0.49	3.11	n.e.	0.06	n.e.
14	AT, Völkermarkt, 46.65°, 14.91°	Ruderal	150	0	30	24	0.67	5.24	n.e.	0.69	−0.071***
15	SI, Murska Sobota, 46.63°, 16.21°	Ruderal	50	0	9	20	0.54	3.66	n.e.	0.95	n.e.
16	SK, Sandberg, 48.2°, 16.97°	Natural	1000	2	31	20	0.71	5.29	n.e.	0.93	−0.002
17	HU, Balatongyörök, 46.76°, 17.34°	Ruderal	600	0	31	25	0.67	5.4	n.e.	0.83	−0.029*
18	HU, Csepel Island, 47.33°, 18.95°	Ruderal	400	6	27	27	0.70	5.63	n.e.	0.85	−0.019*
19	SK, Gelnica, 48.85°, 20.93°	Ruderal	50	0	15	24	0.54	4.16	n.e.	0.58	−0.085**
20	RO, Valea lui David, 47.2°, 27.47°	Natural	2000	0	24	21	0.75	5.86	n.e.	0.35	n.e.
Native range, tetraploid (EU4x)											
21	DE, Lichterfelde Süd, 52.4°, 13.31°	Ruderal	750	24	0	30	0.77	5.96	5.12	0.78	−0.027*
22	DE, Dieskau, 51.44°, 12.04°	Ruderal	300	30	0	23	0.62	4.75	2.73	0.33	−0.007
23	DE, Pratzschwitz, 50.97°, 13.9°	Ruderal	70	25	5	21	0.74	5.5	4.84	0.57	n.e.
24	DE, Nürnberg, 49.39°, 11.08°	Ruderal	200	33	0	25	0.74	6.05	3.01	0.68	−0.029*
25	AT, Starkenbach, 47.19°, 10.64°	Ruderal	1500	32	0	28	0.73	5.41	3.22	0.62	−0.03*
26	AT, Krems, 48.43°, 15.65°	Ruderal	400	33	0	26	0.78	6.26	7.2	0.41	−0.005
27	SK, Závod, 48.52°, 17.02°	Ruderal	200	33	0	23	0.77	5.6	2.96	0.73	−0.052***
28	SK, Trnava, 48.38°, 17.6°	Ruderal	200	23	0	25	0.75	5.94	8.11	0.61	−0.044*
29	SK, Nové Mesto, 48.76°, 17.84°	Ruderal	50	25	0	21	0.66	4.74	3.56	0.62	−0.022
30	HU, Tapolca, 46.88°, 17.43°	Ruderal	25	22	10	19	0.78	6.01	2.29	0.85	n.e.
31	HU, Gellérthegy, 47.49°, 19.05°	Natural	1000	30	0	26	0.79	6.2	11.03	0.82	n.e.
32	RO, Urziceni Pădure, 47.7°, 22.44°	Natural	2000	28	0	20	0.77	6.39	4.24	0.79	−0.007
33	RO, Cheile Turzii, 46.56°, 23.7°	Natural	500	31	2	23	0.77	6.08	3.89	0.82	n.e.
34	RO, Poşaga de Jos, 46.43°, 23.45°	Natural	400	28	0	25	0.77	5.68	4.92	0.86	−0.02*
35	RO, Oprişeni, 47.48°, 26.27°	Natural	750	26	0	24	0.70	5.24	3.69	0.47	n.e.
36	RO, Lepşa, 45.94°, 26.59°	Natural	800	15	0	20	0.71	4.88	1.96	0.12	n.e.
37	RO, Paraul Cacaina, 47.19°, 27.59°	Natural	750	25	0	20	0.68	5.23	3.92	0.3	n.e.
38	MD, Tiraspol, 46.87°, 29.58°	Natural	600	24	0	19	0.76	5.63	5.16	0.6	n.e.
Invasive range, tetraploid (NA4x)											
39	CA, Tsawwassen, 49.02°, −123.11°	Ruderal	1500	22	0	28	0.78	6.44	4.66	0.71	−0.033**
40	US, Seattle, 47.48°, −122.24°	Ruderal	155	18	0	22	0.72	5.26	2.11	0.59	n.e.
41	US, Stevens Pass, 47.79°, −120.89°	Ruderal	300	28	0	26	0.76	6.28	3.23	0.81	−0.018

Table 1 continued

ID	Country, locality, GPS (°N,°E)	Habitat type	Population size	4x	2x	Sample size	H_e	A_R	DW	H_A	b_{log}
42	CA, Kamloops, 50.71°, −120.37°	Ruderal	2500	20	0	26	0.77	6.19	3.43	0.87	−0.041***
43	CA, Revelstoke, 51.02°, −118.21°	Natural	80	27	0	25	0.77	6.21	3.74	0.77	−0.052**
44	US, Coeur dÁlene, 47.65°, −116.72°	Natural	20,000	22	0	26	0.77	6.24	4.49	0.87	0.003
45	US, Emerald Creek, 47.07°, −116.33°	Ruderal	5000	25	0	31	0.78	6.47	4.49	0.82	−0.034**
46	CA, Okotoks, 50.72°, −113.95°	Ruderal	6	11	0	22	0.70	4.61	1.68	0.39	n.e.
47	US, Logan Pass, 48.7°, −113.69°	Natural	1500	25	0	26	0.66	4.29	2.46	0.57	n.e.
48	US, Missoula, 46.87°, −113.99°	Ruderal	12,000	21	0	22	0.77	6.28	5.57	0.83	−0.024*
49	US, Salmon Lake, 47.12°, −113.43°	Natural	25,000	22	0	21	0.82	6.84	3.64	0.82	−0.002
50	US, Butte, 46°, −112.61°	Ruderal	120	21	0	20	0.79	6.6	2.91	0.74	n.e.
51	CA, Manyberries, 49.43°, −110.72°	Ruderal	22	22	0	23	0.79	6.15	3.09	0.66	n.e.
52	US, Big Sandy, 48.19°, −110.11°	Natural	14	19	0	20	0.76	5.38	2.29	0.69	n.e.
53	US, West Yellowstone, 44.7°, −111.1°	Natural	10	10	0	19	0.62	3.95	2.36	0.23	n.e.
54	US, Kendallś, 45.53°, −111.18°	Natural	300	22	0	23	0.72	5.05	2.84	0.2	−0.005
55	US, Burke Park, 45.67°, −111.03°	Ruderal	160	20	0	19	0.72	5.21	1.94	0.49	n.e.
56	US, Mammoth, 44.97°, −110.69°	Natural	250	13	0	20	0.74	5.08	1.89	0.65	−0.073**

ID: population ID; 4x: tetraploid individuals identified by flow cytometry; 2x: diploid individuals identified by flow cytometry; H_e: expected heterozygosity (corrected for sample size); A_R: allelic richness (rarefied to minimum sample size of 19 samples); DW: down-weighted marker value; H_A: degree of admixture; b_{log}: regression slope of spatial genetic autocorrelation; AT: Austria; CA: Canada; CH: Switzerland; CZ: Czech Republic; DE: Germany; HU: Hungary; IT: Italy; MD: Moldova; RO: Romania; SI: Slovenia; SK: Slovakia; US: USA; n.e.: not estimated

Significance level: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$

Geographic distribution of genetic clusters

We studied the among-population genetic structure with Structure 2.3.4 (Pritchard et al. 2000) employing a Bayesian assignment analysis. We performed two admixture models with correlated allele frequencies, one for each cytotype. Data were coded as co-dominant allele matrix. To handle genotypes with ambiguous allele copy numbers, we analyzed the tetraploid subset with the recessive allele option.

We ran both models with 20 replicate chains of 100,000 MCMC iterations after discarding 100,000 burn-in iterations for each K (i.e. number of genetically distinct partitions). The most likely partitioning was determined according to Evanno et al. (2005) using Structure Harvester (Earl and vonHoldt 2012). Tested K values ranged from $K = 1$ to $K = 18$ for the tetraploid subset, and to $K = 20$ for the diploid subset. Individual as well as population mean posterior assignment probabilities were inferred with CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007). Barplots of individual assignments were illustrated in Distruct 1.1 (Rosenberg 2004). We visualized mean population cluster memberships as barplots and plotted them on a geographic map in ArcMap 10.1 (ESRI). We attributed a population or an individual to a distinct cluster when its membership probability (qK) was higher than an arbitrary threshold of 80 % (Zorić et al. 2012). Mean assignment probabilities of populations were regressed against longitude and latitude in linear models. We further estimated an admixture index (H_A) according to Keller and Taylor (2010), and tested for differences in degree of admixture within individuals in NA4x and EU4x in a linear mixed effect model with population set as a random effect (package lme4 in R; Bates et al. 2014).

Since Structure may be less reliable when dealing with different ploidy levels (Dufresne et al. 2014), we assessed genetic similarity between the GCTs with a principal component analysis (PCA) to illustrate gene flow between cytotypes. We used Bruvo distances (Bruvo et al. 2004) calculated with the R-package polysat (Clark and Jaseniuk 2011), which is

particularly recommended for analyzing mixed-ploidy data (Dufresne et al. 2014).

Differentiation among populations

To quantify genetic differentiation of populations, we calculated the most frequently used estimator of F_{ST} (Nei 1978), and ρ_{ST} (Ronfort et al. 1998) in SPAGeDi. ρ_{ST}-statistics exhibit identical expectations for population differentiation in different ploidy levels under identical gene flow conditions (Hardy and Vekemans 2001). Moreover, while F_{ST} may be underestimated under disomic inheritance (Dufresne et al. 2014), ρ_{ST} was shown to be least sensitive to ploidy level and double reduction rate and, consequently, mode of inheritance (Meirmans and van Tienderen 2013).

Small-scale genetic structure within populations

In 30 populations (ten populations of each GCT, see Table 1), we geo-referenced all sampled individuals. To compare small-scale genetic structure of diploids and tetraploids, we used the coefficient of relationship (ρ), as it is not affected by reduced drift in polyploids (Hardy and Vekemans 2001). We computed pair-wise ρ_{ij} by applying Moran's I statistics in SPAGeDi. Spatial distance was divided into distinct distance intervals that ensured a high spatial resolution and a sufficient number of individual pairs per distance class (i.e. 5, 10, 20, 40, 80, 160, and 320 m). To illustrate whether the within population structure differed between GCTs, an averaged ρ was computed for each given distance interval over all pair-wise comparisons within GCTs and plotted against spatial distance in correlograms. Within each GCT, significance of each mean ρ per class was tested with 1000 permutations of multilocus genotypes. We correlated matrices of pair-wise ρ_{ij} and $\log_e$ spatial distances for each population and for each GCT, and tested b_{log} (i.e. slope of the regression) with Mantel tests (1000 randomizations).

Influence of population size and habitat on genetic diversity

To analyze differences in genetic diversity (H_e, A_R, both untransformed) between GCTs and habitat type, we used ANOVAs including GCT and habitat, as well as their interaction, as fixed effects in R. Moreover, we performed ANCOVAs to analyze the effects of centered $\log_e$ population size, GCT and their interaction on genetic diversity. Transformation decisions were based on graphical assessment of normality of errors and homogeneity of variance (i.e. model checking plots; Crawley 2014). Significance of all terms was tested with F tests (type III sums of squares; R-package car; Fox and Weisberg 2010). For significant terms of the ANOVA models, pair-wise comparisons among factor levels were performed with Tukey post hoc tests (R-package multcomp, Hothorn et al. 2008). When the interaction of GCT and $\log_e$ population size was significant in the ANCOVAs, we fitted single linear models to assess the significance of $\log_e$ population size on genetic diversity for the single GCTs. To assess potentially confounding relationships between population size and other explanatory variables, we tested whether $\log_e$ population size depended on GCT, habitat or their interaction. While population size differed between habitats within EU4x (ruderal < natural; $F_{1,18} = 5.73$, $P < 0.05$), all other tested potentially confounding relationships were non-significant.

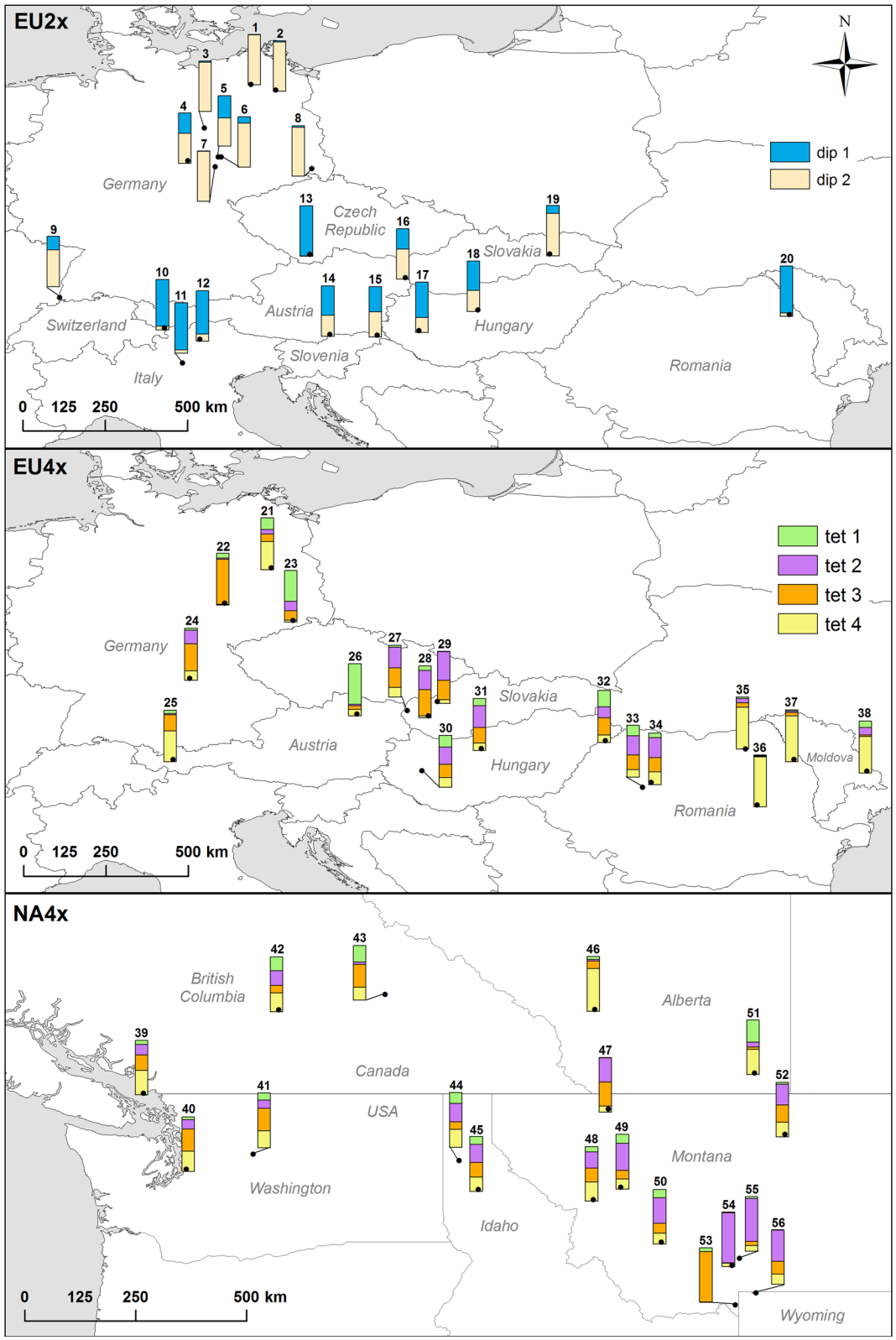
Results

Cytotype distribution

All North American samples were tetraploid (Table 1). In Europe, we confirmed 18 populations to consist exclusively of diploids and 15 populations to be tetraploid. Five populations consisted of both cytotypes, from which two were dominated by diploids and three by tetraploids. We did not observe any triploid individual.

Genetic diversity within populations

We recorded a total of 115 alleles, 16 of which were found exclusively in tetraploids, and two in diploids only. Within diploids, we determined 17 private alleles (i.e. found only within a single sample location) in ten populations. For tetraploids, 14 private alleles were found in ten populations. At the continent scale, more alleles were unique to EU4x (16) than to NA4x (4), and DW was significantly higher ($F_{1,34} = 6.03$; $P < 0.05$) in EU4x (mean 4.55; range 1.96–11.03; Table 1) than in NA4x (mean 3.16; range 1.68–5.57). A_R ranged from 2.58 to 6.04 in EU2x (mean 4.39),


EU2x
N
dip 1
dip 2
Germany
Czech Republic
Slovakia
Switzerland
Austria
Hungary
Slovenia
Italy
Romania
0 125 250 500 km
EU4x
tet 1
tet 2
tet 3
tet 4
Germany
Slovakia
Austria
Hungary
Moldova
Romania
0 125 250 500 km
NA4x
British Columbia
Alberta
Canada
USA
Washington
Idaho
Montana
Wyoming
0 125 250 500 km

◀**Fig. 1** Maps of the sampled *Centaurea stoebe* populations including *barplots* of the structure results, subdivided by the geo-cytotypes (EU2x = native range, diploid; EU4x = native range, tetraploid; NA4x = invasive range, tetraploid). Two separate structure analyses revealed two clusters for the diploid and four clusters for the tetraploid data set. The *stacked barplots* show proportions of populations' posterior assignment probabilities to the different genetic clusters. Population IDs are given in Table 1. Note that we only analyzed samples from the majority cytotype in mixed-ploidy populations (i.e. 16, 18, 23, 30 and 33)

from 4.74 to 6.39 in EU4x (mean 5.64), and from 3.95 to 6.84 in NA4x (mean 5.69). We estimated an average H_e of 0.6 in EU2x (0.32–0.75), 0.74 in EU4x (0.62–0.79) and 0.75 in NA4x (0.62–0.82). Both A_R and H_e differed highly significantly between GCTs ($P < 0.001$). Posthoc-tests for the main effect of GCT consistently revealed significantly lower values in EU2x compared to EU4x and NA4x, whereas EU4x did not differ from NA4x in both cases.

Geographic distribution of genetic clusters

The Bayesian inferences of genetic structure revealed an optimal number of two clusters for the diploid data set, and four clusters for the tetraploid subset (Electronic Supplementary Material Fig. S1). Within the diploid data set, the first cluster (dip1, Fig. 1) was more abundant in southern populations (decreasing with latitude, $F_{1,18} = 27.04$, $P < 0.001$), while dip2 mainly occurred in northern populations. 12 out of 20 EU2x populations were concerned as distinct, because they showed an assignment to one of both clusters that was higher than the arbitrary threshold of 80 % (Zoric et al. 2012). At the individual level, 79.2 % of all diploid individuals were assigned to a distinct cluster (Fig. 2). For tetraploids, the genetic structure was considerably more ambiguous, as the majority of populations (30 out of 36) comprised mixtures of different clusters (Fig. 1). We found significant relationships of cluster membership coefficients with longitude in EU4x (tet 3: $F_{1,16} = 8.33$, $P < 0.05$; tet 4: $F_{1,16} = 7.53$, $P < 0.05$) and with latitude in NA4x (tet 1: $F_{1,16} = 5.9$, $P < 0.05$; tet 2: $F_{1,16} = 7.34$, $P < 0.05$; tet 4: $F_{1,16} = 12.79$, $P < 0.01$). The proportion of individuals belonging to a distinct cluster was 28.4 % for EU4x and 21.2 % for NA4x (Fig. 2). There was no significant difference ($\chi^2_{(1)} = 2.91$, $P = 0.09$) in degree of admixture between diploids (mean $H_A = 0.51$, Table 1) and tetraploids (mean $H_A = 0.64$). The level of admixture did not differ

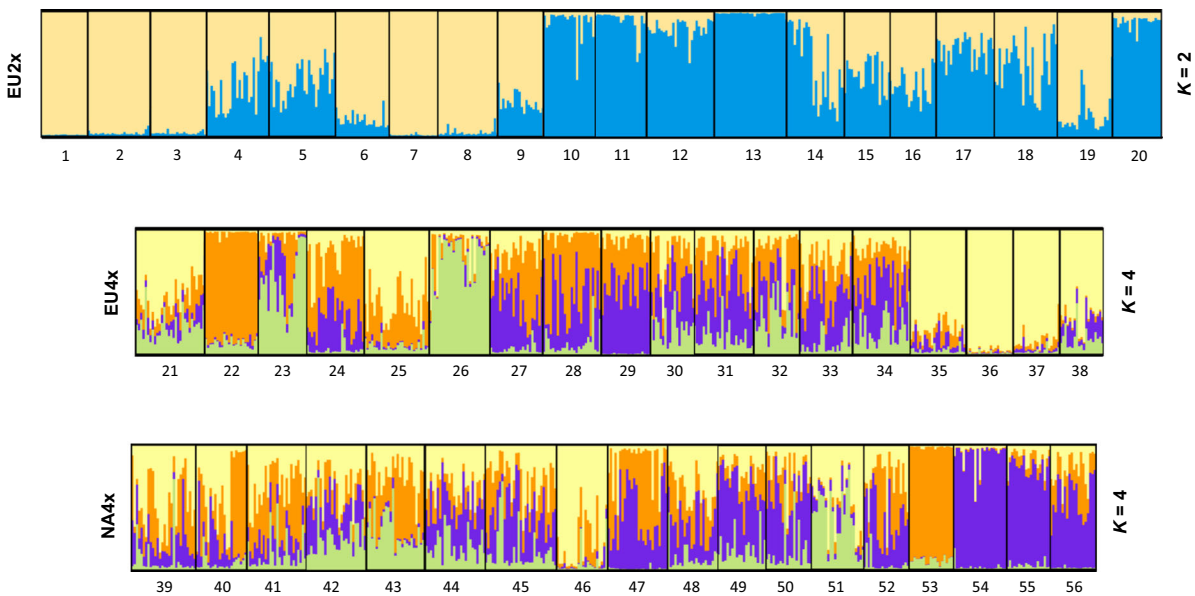


Fig. 2 Stacked barplots of individual posterior assignment probabilities to the clusters identified in the structure analyses, subdivided by the geo-cytotypes (EU2x = native range, diploid; EU4x = native range, tetraploid; NA4x = invasive range, tetraploid). Two separate structure analyses revealed two clusters for the diploid and four clusters for the tetraploid data set, respectively. Population IDs are given in Table 1. Note that we only analyzed samples from the majority cytotype in mixed-ploidy populations (i.e. 16, 18, 23, 30 and 33)

($\chi^2_{(1)} = 0.35$, $P = 0.55$) between EU4x (mean $H_A = 0.61$) and NA4x (mean $H_A = 0.66$).

The PCA revealed a strict separation of both cytotypes. We found two clusters, one including EU2x and one including both tetraploid GCTs (Electronic Supplementary Material Fig. S2). NA4x and EU4x did not show clear separation from one another.

Differentiation among populations

The overall population structure was significant in all GCTs ($P < 0.001$, Table 2). Global differentiation of diploids ($F_{ST} = 0.17$; $\rho_{ST} = 0.24$) was substantially higher than that of tetraploids ($F_{ST} = 0.07$; $\rho_{ST} = 0.13$). Differentiation was almost identical among EU4x populations ($F_{ST} = 0.07$; $\rho_{ST} = 0.13$) compared to NA4x ($F_{ST} = 0.07$; $\rho_{ST} = 0.12$).

Table 2 Genetic differentiation of the populations: F_{ST}-statistics and ρ_{ST}-statistics

Group	F_{ST}	ρ_{ST}
Total	0.114	0.204
4x total	0.072	0.127
EU4x	0.073	0.131
US4x	0.069	0.122
EU2x	0.168	0.238

Total: all samples; 4x total: all tetraploid samples; EU4x: native tetraploid; US4x: invasive tetraploid; EU2x: native diploid

Small-scale genetic structure within populations

Diploid individuals showed higher spatial autocorrelation within the first two distance classes than tetraploids (Fig. 3). Overall, we observed similar patterns of small-scale genetic structure across all GCTs with the highest relationship coefficients occurring in the first distance intervals, and decreasing continuously thereafter with increasing spatial distance. All observed mean coefficients of relationship per distance class were highly significant in all GCTs ($P < 0.001$). The correlation of pair-wise ρ_{ij} and $\log_e$ spatial distances was highly significant in all GCTs (EU2x: $b_{log} = -0.042$, $P < 0.001$; EU4x: $b_{log} = -0.028$, $P < 0.001$; NA4x: $b_{log} = -0.32$, $P < 0.001$) and significant in the majority of populations (Table 1).

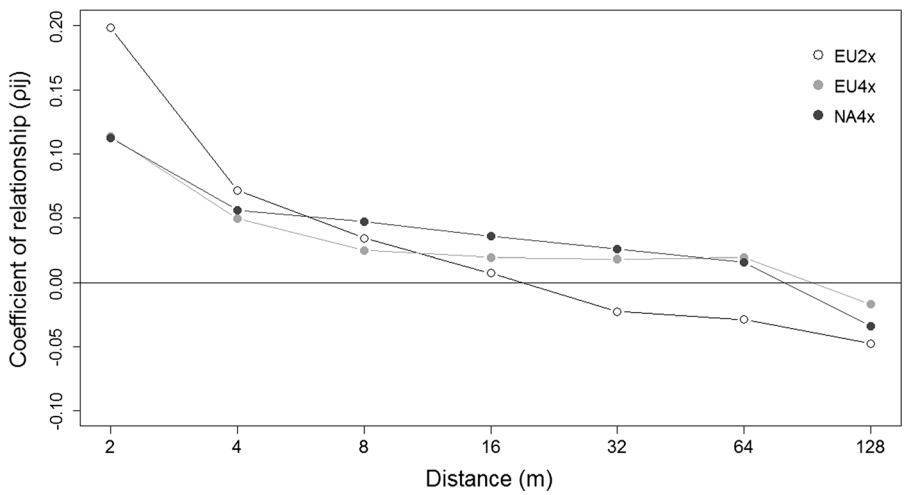


Fig. 3 Correlogram of averaged coefficients of relationship between individuals per distance class (log scale). Coefficients were computed for pairs of individuals within 30 populations in SPAGeDi [10 per geo-cytotype (GCT)]. *Colors of dots* correspond to the GCTs [*white* = EU2x (native range, diploid); *light grey* = EU4x (native range, tetraploid); *dark grey* = NA4x (invasive range, tetraploid); see legend]. The *solid horizontal line* (y = 0) represents the average relationship between individuals of the overall gene pool within GCTs under Hardy–Weinberg equilibrium. All observed coefficients of relationship per distance class were highly significant higher than y = 0 ($P < 0.001$, 1000 permutations). Sample size for each distance within each GCT was $N > 80$. Mantel Tests revealed highly significant slopes (b_{log}) for the regression between relationship coefficients and spatial distance in all GCTs (see legend)

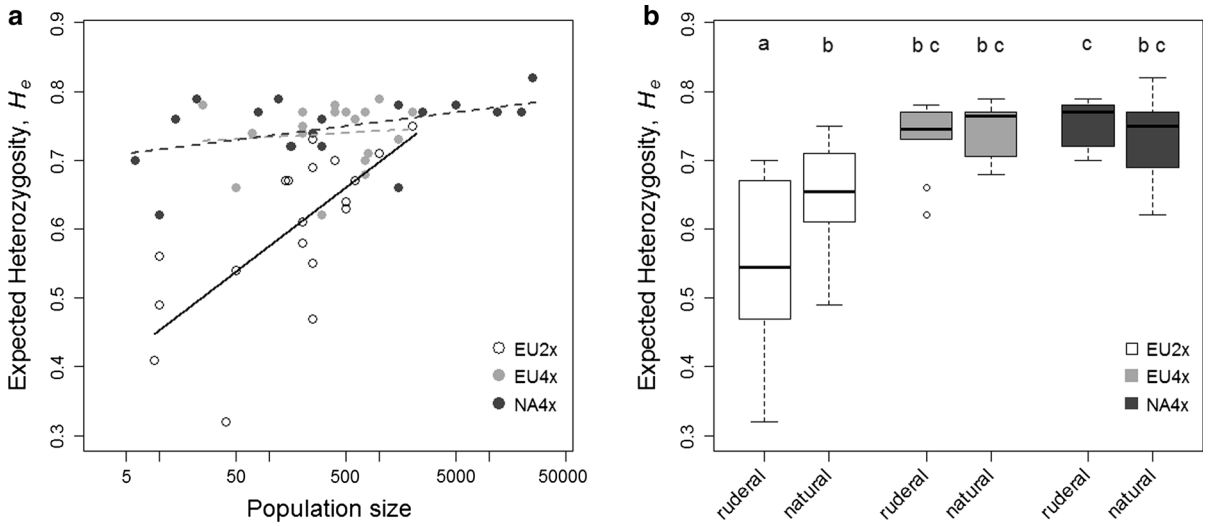


Fig. 4 Genetic diversity in relation to population size and habitat type. **a** Expected heterozygosity (H_e) in relation to the interaction of geo-cytotype (GCT) and population size. *Lines* represent predictions of the respective models as follows: *solid* = EU2x (significant; $F_{1,18} = 18.78$, $P < 0.001$), *light grey dashed* = EU4x (non-significant), *dark grey dashed* = NA4x (non-significant). **b** H_e in relation to the interaction of GCT and habitat. *Boxplot symbols* represent following statistics: *bold line* = median; *box* = interquartile range; *whiskers* = 1.5 times the inter-quartile range or the range of data, whichever is smaller; *points* = outliers. Groups *a*, *b*, and *c* are based on pair-wise comparisons with Tukey post hoc tests (significance level $P < 0.05$). *Colors* in (**a**) and (**b**) correspond to the GCTs [*white* = EU2x (native range, diploid); *light grey* = EU4x (native range, tetraploid); *dark grey* = NA4x (invasive range, tetraploid); see legend]

Influence of population size and habitat on genetic diversity

We observed a significant interaction in the effects of population size and GCT on H_e ($F_{2,50} = 9.04$, $P < 0.001$; Fig. 4a) and A_R ($F_{2,50} = 4.63$, $P < 0.05$). Specifically, we found a positive effect of population size on H_e in EU2x ($F_{1,18} = 18.78$, $P < 0.001$), while there was no significant relationship in any of the two tetraploid GCTs. Allelic richness revealed similar patterns and was positively related to population size in EU2x ($F_{1,18} = 20.94$, $P < 0.01$) and NA4x ($F_{1,16} = 6.99$, $P < 0.05$), but not in EU4x. In addition, we found a significant interaction between habitat and GCT for H_e ($F_{2,50} = 4.31$, $P < 0.05$; Fig. 4b) and A_R ($F_{2,50} = 3.26$, $P < 0.05$). In particular, H_e was significantly smaller in ruderal compared to natural habitats within EU2x ($t = 3.24$, $P < 0.05$), but it did not differ significantly within both of the tetraploid GCTs. Allelic richness showed a similar pattern although the difference among habitats within EU2x was non-significant ($t = 2.16$, $P = 0.28$).

Discussion

Consequences of polyploid formation for genetic diversity

Our results confirmed the hypothesis that, in *C. stoebe*, tetraploids sustain higher genetic diversity than diploids. Since 16 of 115 alleles were unique to the tetraploid cytotype, the high genetic diversity may have resulted from hybridization with a second divergent parental species as suggested by Mráz et al. (2012a). The few alleles unique to diploids (i.e. 2 of 115) suggest that polyploidization occurred on multiple occasions, which aligns with the currently prevalent opinion that the majority of polyploids originated from multiple polyploidization events (Soltis et al. 2014). The PCA that investigated genetic structure across the GCTs showed a strict separation of both cytotypes into two clusters, which supports previous findings of a strong reproductive isolation between diploid and tetraploid *C. stoebe* (Mráz et al. 2012b).

Most studies on polyploid complexes have revealed higher genetic diversity in polyploids than in diploids (e.g. Eliášová et al. 2013 and references therein; but see Ferriol et al. 2014). However, we explicitly investigated this difference with a focus on GCTs. This is particularly important regarding the correlation between genetic diversity and invasion success (Forsman 2014).

In *C. stoebe,* higher genetic diversity in tetraploids may account for broader adaptive capabilities, which may have enabled tetraploids to adapt to novel conditions in the non-native range with a remarkable climatic niche shift towards a drier and more continental climate in NA4x as compared to EU4x (see Treier et al. 2009). In addition, heterozygosity masks recessive deleterious mutations, which may result in lower inbreeding depression compared to diploids under the same level of inbreeding (Eliášová et al. 2013). Such genetic processes may, in concert, have led to higher population growth rates of tetraploids, which Hahn et al. (2012) recently recorded in a common garden study with artificial populations of the three GCTs.

The role of multiple introductions and admixture in genetic diversity of NA4x

In accordance with our hypothesis, genetic diversity between both tetraploid ranges was comparable. Previous population genetic studies on tetraploid *C. stoebe* revealed different outcomes. While Marrs et al. (2008) reported significantly higher expected heterozygosity for microsatellite loci in NA4x than EU4x, chloroplast haplotype diversity was found to be substantially lower (Hufbauer and Sforza 2008). In contrast to Marrs et al. (2008), genetic diversity of NA4x did not exceed that of EU4x, which may be related to different sampling designs. Marrs et al. (2008) analyzed considerably fewer samples per population in EU4x than in NA4x, and they included one mixed-ploidy population and two EU2x populations in their estimation of native genetic diversity without explicitly distinguishing between diploids and tetraploids.

Our Structure analysis supports that high genetic diversity of NA4x is a result of multiple introductions, most likely from different parts of the native range (see also Hufbauer and Sforza 2008; Marrs et al. 2008). As multiple introductions are common, high genetic diversity in NA4x corresponds to numerous studies on invasive plants (e.g. Kelager et al. 2012; Bousset et al. 2013). However, even if current genetic diversity is not reduced in later invasion phases, initial colonization and range expansion may still have involved bottlenecks, as contrasting demographic events (e.g. bottlenecks and admixture) may act simultaneously at different points in space and time (Keller et al. 2012). The considerably higher number of alleles unique to EU4x than to NA4x, along with the significantly higher *DW* of EU4x, at least suggests that rare alleles were not exhaustively carried to the exotic range. The loss of rare alleles shows evidence of bottlenecks in the past (Comps et al. 2001). When populations face bottlenecks, they most likely undergo strong selection towards restoring heterozygosity (Theoharides and Dukes 2007). Moreover, benefits of admixture may rule out disadvantages from introgression of maladapted genotypes in the new range, where local adaptation is rather weak compared to the native range (Verhoeven et al. 2011). We indeed found a substantially higher frequency of admixed individuals in NA4x (27 %) than in EU4x (22 %), but the difference in H_A remained non-significant due to large among-population variation. Nevertheless, comparable genetic diversity of NA4x and EU4x indicates that the ability of polyploids to maintain or restore high genetic diversity may be crucial for colonization of new ranges in *C. stoebe*.

Influences of ploidy level, life history and range dynamics on differentiation

In line with our third hypothesis, differentiation was stronger among diploid than among tetraploid populations. Due to reduced genetic drift in polyploids (Ronfort et al. 1998), higher F_{ST}-values of diploids are common in polyploid complexes (e.g. Eliášová et al. 2013 and references therein). However, in contrast to previous investigations (e.g. Hardy and Vekemans 2001; Eliášová et al. 2013), in our study, ρ_{ST}-statistics also revealed remarkably stronger differentiation among diploids than among tetraploids, which cannot be accounted for by mathematical principles of differences in drift (Meirmans and van Tienderen 2013). In addition, other mechanisms that influence demographic history have to be considered.

Firstly, in *C. stoebe*, tetraploids are, unlike the monocarpic diploids, characterized by increased

longevity (Mráz et al. 2011) and a meta-analysis confirmed weaker differentiation among polycarpic populations due to trans-generational gene flow and less frequent inbreeding (Nybom 2004). Particularly, during events that lead to a short-term reduction of flowering mating partners (e.g. mowing), drift will be reduced in species that do not necessarily have to reproduce sexually every year. Instead polycarpic species may outlast such events as rosettes and can reproduce sexually after re-sprouting in the following vegetation period. Quantitative information about specific disturbance regimes (e.g. mowing frequencies) were not available for our populations, but should be included in more mechanistic investigations of this aspect in future studies.

Secondly, while tetraploids in both ranges recently expanded their range with potentially high gene flow among populations, EU2x show a more stable and scattered distribution, which may hamper gene flow (Mráz et al. 2014). Accordingly, our Structure analyses showed that the vast majority of tetraploid populations were assigned to mixed clusters, while most diploid populations were assigned to a distinct cluster. The distribution of clusters was differentiated along a north-south gradient in EU2x and in NA4x, which corresponds to findings of Mráz et al. (2014) in phenotypic trait variation that was largely explained by latitudinal clines.

Signatures of range dynamics in the differentiation within tetraploid ranges

We found a rather weak differentiation in both ranges, which, in contrast to our expectations, did not differ between NA4x and EU4x (see also Marrs et al. 2008). In North America, low differentiation can be explained by multiple introductions with repeated admixture events and by huge metapopulation sizes, as *C. stoebe* is highly abundant across our invasive study area (Maron et al. 2013). The weak differentiation in EU4x may result from a recent range expansion from its presumed ancestral region in south-eastern Europe towards central Europe (Mráz et al. 2014), mainly into ruderal habitats (Otisková et al. 2014). We indeed recorded a switch from natural sites inhabited by all our investigated populations east of the 19th longitudinal degree to ruderal habitats west of it. This corresponds to the longitudinal gradient in the cluster distribution of EU4x. Within tetraploids, we found a large among-population variation in H_A, whereby populations at the margins of the sampled distribution range tended to be less admixed. In particular, the most obvious geographical signal of our tetraploid structure analysis was that four south-eastern EU4x populations form a common gene pool. On the opposite, the ruderal populations in central Europe were rather admixed (to a similar extent as in NA4x). Thus, recent range expansions in both ranges may have led to similar conditions of ongoing admixture. Human-mediated long distance dispersal may facilitate such high gene flow, with railways and roads serving as the most probable dispersal corridors (Broennimann et al. 2014).

Small-scale genetic structure suggests biparental inbreeding and spatial cytotype segregation

As hypothesized, tetraploids exhibited smaller relationship coefficients at short distances than diploids. Since both cytotypes have comparable pollen and achene dispersal agents (Mráz et al. 2012b; Hahn et al. 2013), and ρ_{ij}-statistics account for differences in drift between ploidy levels, this outcome may have resulted from less frequent inbreeding in tetraploids due to their polycarpic life cycle. In addition, other, mutually non-exclusive mechanisms, e.g. seed or pollen number per lifetime, community diversity or plant density, may influence gene dispersal distances (Zeng et al. 2011).

Nonetheless, the relatedness of individuals significantly decreased with spatial distance displaying a certain level of biparental inbreeding in all GCTs. In the polyploid complex of *Centaurea jacaea*, Hardy and Vekemans (2001) found similar results that coincide with barochory and with pollinators that normally travel short distances. They argued that these modes of gene dispersal seem to maintain mixed-ploidy populations via spatial segregation of cytotypes. Indeed, *C. stoebe* exhibits equal pollen and seed dispersal capabilities, and spatial segregation prevails in mixed-cytotype populations (Mráz et al. 2012b).

Contrasting influences of population size and habitat on genetic diversity between the cytotypes

In line with our hypotheses, H_e increased with population size, and natural populations revealed

higher genetic diversity than ruderal populations in diploids. However, neither population size nor habitat influenced H_e in tetraploids. In the native range, this ability to buffer fluctuating population sizes may contribute to the ecological prevalence of tetraploids in ruderal habitats with high environmental stochasticity as observed by Broennimann et al. (2014) and Otisková et al. (2014). Such ruderal EU4x populations contemporarily face conditions of human-altered habitats, which may result in increased pre-adaptation to highly disturbed habitats of primary invasion (i.e. anthropogenically induced adaptation to invade theory; Hufbauer et al. 2012).

The considerably weaker influence of population size and disturbance regime on genetic diversity in tetraploids than in diploids may particularly result from: (a) reduced drift in polyploids due to mathematical principles of inheritance, (b) recent range expansion in NA4x and EU4x with high connectivity among populations, and (c) increased longevity in polyploids. These processes, in concert, can decelerate genetic depletion through bottlenecks, and thus enhance the probability of tetraploid founder populations surviving and persisting until admixture restores the gene pool. Such a scenario corresponds to the observed lag phase in *C. stoebe* (Broennimann et al. 2014). At the same time, diploids are prone to a higher susceptibility to genetic drift. Under small or fluctuating population sizes, this may ultimately lead to extinction of populations during colonization (Szűcs et al. 2014).

Besides polyploidy, asexual propagation can help to avoid loss of genetic diversity during demographic bottlenecks (Cosendai et al. 2013; Stein et al. 2014). Moreover, there are species that became invasive despite exhibiting strong reductions of genetic diversity in the non-native range (reviewed in Uller and Leimu 2011). However, for the majority of species, mechanisms that maintain high levels of genetic diversity enhance their invasiveness (Forsman 2014), and this should be particularly the case for obligate outcrossers that show no vegetative spread.

Conclusions and perspectives

It is increasingly clear that genetic bottlenecks occur far more frequently during biological invasions than suggested by rather simplistic native versus introduced comparisons of mean genetic diversity at an advanced temporal stage of invasion (Keller et al. 2012). More emphasis should therefore be put on identifying mechanisms, such as polyploidy and longevity, that help founder populations to persist until gene flow and genetic admixture occur. Although several examples of cytotype shifts between native and invasive ranges have been reported (reviewed in te Beest et al. 2011), we are the first to explicitly show how differences in genetic diversity and drift between GCTs may relate to colonization success.

Our results highlight (a) the higher initial genetic diversity of tetraploids than diploids in *C. stoebe*, and (b) the ability of tetraploids to counteract genetic depletion in phases of small or fluctuating population sizes in ruderal habitats. While polycarpic tetraploid founder populations have an enhanced probability to outlast several generations of demographic disequilibrium, monocarpic diploids may be excluded from non-native ranges by lower initial genetic diversity, more frequent inbreeding and stronger drift.

Our analyses help explaining the outstanding invasion success of tetraploid *C. stoebe* on the one hand and the apparent lack of diploids in North America on the other hand. In addition, they provide important insights towards a more mechanistic understanding of the general colonization advantage of polyploids. We are, however, well aware that purely observational studies cannot unequivocally identify drivers of invasions. More colonization genetic studies on polyploid complexes are required to test any generality and limitations in our results. Moreover, identifying the second parental species involved in the origin of allotetraploid *C. stoebe* s.l. may help to understand the relative importance of hybridization in generating ecological and evolutionary change.

Acknowledgments This study was supported by a graduate stipend of the federal state of Saxony-Anhalt (to C.R.), and a DAAD (German Academic Exchange Service) funded exchange scholarship (Grant Number D|12|00534 to C.R.). We are grateful to B. Müller, M. Laudien, I. Geyer and M. Herrmann for their efforts in the lab, and to A. Diaconou, S. Španiel, B. Banan, R.M. Callaway, J. Maron, A. Schaar, C. Herron-Sweet, V. Mrázová, J. Hoheimer, I. Link, F. Rosche, G. Newcombe and A. Vitale for their help during field work. We thank K. Schrieber, C. S. Brown and two anonymous reviewers for valuable comments on previous versions of the manuscript.

Compliance with ethical standard

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent All authors have given formal consent to the publication of this manuscript.

References

Bates D, Maechler M, Bolker B (2014) lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://CRAN.R-project.org/package=lme4

Blanchet É, Penone C, Maurel N et al (2015) Multivariate analysis of polyploid data reveals the role of railways in the spread of the invasive South African Ragwort (*Senecio inaequidens*). Conserv Genet 16:523–533

Bock DG, Caseys C, Cousens RD et al (2015) What we still don't know about invasion genetics. Mol Ecol 24:2277–2297

Bousset L, Pointier J-P, David P, Jarne P (2013) Neither variation loss, nor change in selfing rate is associated with the worldwide invasion of *Physa acuta* from its native North America. Biol Invasions 16:1769–1783

Broennimann O, Mráz P, Petitpierre B et al (2014) Contrasting spatio-temporal climatic niche dynamics during the eastern and western invasions of spotted knapweed in North America. J Biogeogr 41:1126–1136

Bruvo R, Michiels NK, D'Souza TG, Schulenburg H (2004) A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Mol Ecol 13:2101–2106

Clark LV, Jaseniuk M (2011) polysat: an R package for polyploid microsatellite analysis. Mol Ecol Res 11:562–566

Comps B, Gömöry D, Letouzey J et al (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. Genetics 157:389–397

Corn JG, Story JM, White LJ (2006) Impacts of the biological control agent *Cyphocleonus achates* on spotted knapweed, *Centaurea maculosa*, in experimental plots. Biol Control 37:75–81

Cosendai A-C, Wagner J, Ladinig U et al (2013) Geographical parthenogenesis and population genetic structure in the alpine species *Ranunculus kuepferi* (Ranunculaceae). Heredity 110:560–569

Crawley MJ (2014) Statistics: an introduction using R. Wiley, Chichester

Development Core Team R (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Dietz H, Edwards PJ (2006) Recognition that causal processes change during plant invasion helps explain conflicts in evidence. Ecology 87:1359–1367

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. Mol Ecol 17:431–449

Dufresne F, Stift M, Vergilino R, Mable BK (2014) Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Mol Ecol 23:40–69

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Eliášová A, Trávníček P, Mandák B, Münzbergová Z (2013) Autotetraploids of *Vicia cracca* show a higher allelic richness in natural populations and a higher seed set after artificial selfing than diploids. Ann Bot 113:159–170

EUNIS (2008) European Nature Information System (EUNIS) database: habitat types and habitat classifications. European Topic Centre on Biological Diversity, ETC/BD-EEA, Paris

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620

Ferrero V, Barrett SCH, Castro S et al (2015) Invasion genetics of the Bermuda buttercup (*Oxalis pes-caprae*): complex intercontinental patterns of genetic diversity, polyploidy and heterostyly characterize both native and introduced populations. Mol Ecol 24:2143–2155

Ferriol M, Merle H, Garmendia A (2014) Microsatellite evidence for low genetic diversity and reproductive isolation in tetraploid *Centaurea seridis* (Asteraceae) coexisting with diploid *Centaurea aspera* and triploid hybrids in contact zones. Bot J Linn Soc 176:82–98

Forsman A (2014) Effects of genotypic and phenotypic variation on establishment are important for conservation, invasion, and infection biology. PNAS 111:302–307

Fox J, Weisberg S (2010) An R companion to applied regression. SAGE, Thousand Oaks

Hahn MA, Buckley YM, Müller-Schärer H (2012) Increased population growth rate in invasive polyploid *Centaurea stoebe* in a common garden. Ecol Lett 15:947–954

Hahn MA, Lanz T, Fasel D, Müller-Schärer H (2013) Increased seed survival and seedling emergence in a polyploid plant invader. Am J Bot 100:1555–1561

Hardy OJ, Vekemans X (2001) Patterns of allozyme variation in diploid and tetraploid *Centaurea jacea* at different spatial scales. Evolution 55:943–954

Hardy OJ, Vekemans X (2002) spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. Biom J 50:346–363

Hufbauer RA, Sforza R (2008) Multiple introductions of two invasive *Centaurea* taxa inferred from cpDNA haplotypes. Divers Distrib 14:252–261

Hufbauer RA, Facon B, Ravigné V et al (2012) Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. Evol Appl 5:89–101

Hufbauer RA, Rutschmann A, Serrate B et al (2013) Role of propagule pressure in colonization success: disentangling the relative importance of demographic, genetic and habitat effects. J Evol Biol 26:1691–1699

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Kelager A, Pedersen JS, Bruun HH (2012) Multiple introductions and no loss of genetic diversity: invasion history of Japanese Rose, *Rosa rugosa*, in Europe. Biol Invasions 15:1125–1141

Keller SR, Taylor DR (2010) Genomic admixture increases fitness during a biological invasion. J Evol Biol 23:1720–1731

Keller SR, Gilbert KJ, Fields PD, Taylor DR (2012) Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. Mol Ecol 21:4721–4734

Maron JL, Waller LP, Hahn MA et al (2013) Effects of soil fungi, disturbance and propagule pressure on exotic plant recruitment and establishment at home and abroad. J Ecol 101:924–932

Marrs RA, Sforza R, Hufbauer RA (2008) Evidence for multiple introductions of *Centaurea stoebe micranthos* (spotted knapweed, Asteraceae) to North America. Mol Ecol 17:4197–4208

Meirmans PG, Van Tienderen PH (2013) The effects of inheritance in tetraploids on genetic diversity and population divergence. Heredity 110:131–137

Mráz P, Bourchier RS, Treier U et al (2011) Polyploidy in phenotypic space and invastion context: a morphometric study of *Centaurea stoebe* s.l. Int J Plant Sci 172:386–402

Mráz P, Garcia-Jacas N, Gex-Fabry E et al (2012a) Allopolyploid origin of highly invasive *Centaurea stoebe* s.l. (Asteraceae). Mol Phylogenet Evol 62:612–623

Mráz P, Španiel S, Keller A et al (2012b) Anthropogenic disturbance as a driver of microspatial and microhabitat segregation of cytotypes of *Centaurea stoebe* and cytotype interactions in secondary contact zones. Ann Bot 110:615–627

Mráz P, Tarbush E, Müller-Schärer H (2014) Drought tolerance and plasticity in the invasive knapweed *Centaurea stoebe* s.l. (Asteraceae): effect of populations stronger than those of cytotype and range. Ann Bot 114:289–299

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 13:1143–1155

Ochsmann J (2000) Morphologische und molekularsystematische Untersuchungen an der *Centaurea stoebe* L.–Gruppe (Asteraceae–Cardueae) in Europa. Dissertationes Botanicae, vol 324. J. Cramer, Berlin

Otisková V, Koutecký T, Kolář F, Koutecký P (2014) Occurrence and habitat preferences of diploid and tetraploid cytotypes of *Centaurea stoebe* in the Czech Republic. Preslia 86:67–80

Pandit MK, Pocock MJO, Kunin WE (2011) Ploidy influences rarity and invasiveness in plants. J Ecol 99:1108–1115

Pandit MK, White SM, Pocock MJO (2014) The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis. New Phytol 203:697–703

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Ronfort J, Jenczewski E, Bataillon T, Rousset F (1998) Analysis of population structure in autotetraploid species. Genetics 150:921–930

Rosenberg NA (2004) distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Sax DF, Brown JH (2000) The paradox of invasion. Glob Ecol Biogeogr 9:363–371

Schlaepfer DR, Edwards PJ, Widmer A, Billeter R (2008) Phylogeography of native ploidy levels and invasive tetraploids of *Solidago gigantea*. Mol Ecol 17:5245–5256

Schönswetter P, Tribsch A (2005) Vicariance and dispersal in the alpine perennial *Bupleurum stellatum* L. (Apiaceae). Taxon 54:725–732

Shah MA, Callaway RM, Shah T et al (2014) *Conyza canadensis* suppresses plant diversity in its nonnative ranges but not at home: a transcontinental comparison. New Phytol 202:1286–1296

Simberloff D, Martin J-L, Genovesi P et al (2013) Impacts of biological invasions: what's what and the way forward. Trends Ecol Evol 28:58–66

Soltis PS, Soltis DE (2000) The role of genetic and genomic attributes in the success of polyploids. PNAS 97:7051–7057

Soltis DE, Visger CJ, Soltis PS (2014) The polyploidy revolution then…and now: Stebbins revisited. Am J Bot 101:1057–1078

Stein K, Rosche C, Hirsch H, Kindermann A, Köhler J, Hensen I (2014) The influence of forest fragmentation on clonal diversity and genetic structure in *Heliconia angusta*, an endemic understorey herb of the Brazilian Atlantic rain forest. J Trop Ecol 30:199–208

Szűcs M, Melbourne BA, Tuff T, Hufbauer RA (2014) The roles of demography and genetics in the early stages of colonization. Proc R Soc Lond B Biol 281:20141073. doi:10.1098/rspb.2014.1073

te Beest M, Roux JJL, Richardson DM et al (2011) The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot 109:19–45

Theoharides KA, Dukes JS (2007) Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. New Phytol 176:256–273

Treier UA, Broennimann O, Normand S et al (2009) Shift in cytotype frequency and niche space in the invasive plant *Centaurea maculosa*. Ecology 90:1366–1377

Uller T, Leimu R (2011) Founder events predict changes in genetic diversity during human-mediated range expansions. Glob Change Biol 17:3478–3485

Van Puyvelde K, Van Geert A, Triest L (2010) atetra, a new software program to analyse tetraploid microsatellite data: comparison with tetra and tetrasat. Mol Ecol Resour 10:331–334

Verhoeven KJF, Macel M, Wolfe LM, Biere A (2011) Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. Proc R Soc Lond B Biol 278:2–8

Zeng X, Michalski SG, Fischer M, Durka W (2011) Species diversity and population density affect genetic structure and gene dispersal in a subtropical understory shrub. J Plant Ecol 5:270–278

Zorić M, Dodig D, Kobiljski B et al (2012) Population structure in a wheat core collection and genomic loci associated with yield under contrasting environments. Genetica 140:259–275